

A penalized likelihood approach for efficiently estimating a partially linear additive transformation model with current status data*

Yan Liu and Minggen Lu[†]

*School of Community Health Sciences, University of Nevada, Reno,
Reno, NV, USA*
e-mail: yliu23@unr.edu; minggenl@unr.edu

Christopher S. McMahan

*School of Mathematical and Statistical Sciences, Clemson University,
Clemson, SC, USA*
e-mail: mcmaha2@clemson.edu

Abstract: Current status data are commonly encountered in medical and epidemiological studies in which the failure time for study units is the outcome variable of interest. Data of this form are characterized by the fact that the failure time is not directly observed but rather is known relative to an observation time, i.e., the failure times are either left- or right-censored. Due to its structure, the analysis of such data can be challenging. To circumvent these challenges and to provide for a flexible modeling construct which can be used to analyze current status data, herein a partially linear additive transformation model is proposed. In the formulation of this model, constrained B -splines are employed to model the monotone transformation function and nonparametric covariate effects. To provide for more efficient estimators, a penalization technique is used to regularize the estimation of all unknown functions. An easy to implement hybrid algorithm is developed for model fitting, and a simple and consistent estimator of the large-sample variance-covariance matrix for regression parameter estimators is proposed. It is shown theoretically that the proposed estimators of the finite-dimensional regression coefficients are root- n consistent, asymptotically normal, and achieve the semiparametric information bound, while the estimators of the nonparametric components attain the optimal rate of convergence. The finite-sample performance of the proposed methodology is evaluated through extensive numerical studies and is further demonstrated through the analysis of human papillomavirus (HPV) data.

Keywords and phrases: B -spline, current status data, isotonic regression, partially linear additive transformation model, penalized estimation.

Received June 2020.

*This research was supported by the National Institutes of Health grants R01-AI121351, National Science Foundation grant OIA-1826715, and Department of Defense's Office of Naval Research grant N00014-19-1-2295. Christopher S. McMahan was funded by Grant R01 AI121351 from the National Institutes of Health, Grant OIA-1826715 from the National Science Foundation, and Grant N00014-19-1-2295 from the Office of Naval Research.

[†]Corresponding Author

1. Introduction

Current status data are characterized by the fact that the failure/event time cannot be directly measured, but rather is known relative to a single observation time. That is, the failure time is known to be smaller or larger than an observation time leading to left- or right-censored data, respectively. Data of this structure commonly arise as a part of medical and epidemiological studies, among many others, in which study participants are observed only once due to destructive/invasive testing, resource limitations, etc. For example, as part of the National Health and Nutrition Examination Survey (NHANES), participants were examined for the presence of the human papillomavirus (HPV). In this study, HPV onset time could not be directly observed but was known relative to the participant's age at the examination time.

For the regression analysis of current status data, several parametric and semiparametric methods have been proposed. In particular, many of these efforts have focused on the proportional hazards (PH) [4] and proportional odds (PO) models [5]; for a review see [11, 19, 26] and the references therein. To provide a more general modeling framework, some researchers have developed estimation techniques under linear transformation models [7, 10, 12, 27, 35, 37, 38]. A potentially prohibitive assumption made in all of the aforementioned works is that the covariate effects on the failure time are "linear". To relax this assumption, Ma and Kosorok [17] and Lu and McMahan [13] presented partially linear transformation and PH models, respectively, for current status data. These works allowed for the estimation of a single nonparametric effect, which could again be limiting. To avoid this potential limitation, Cheng and Wang [2] developed a semiparametric additive transformation model which allows for the estimation of multiple nonparametric regression functions. To our knowledge, the work of these authors is the most general model designed for the analysis of current status data.

Motivated by the work of Cheng and Wang [2], herein we seek to develop a flexible, easy to implement, and computationally efficient methodology which can be used to conduct the regression analysis of current status data under a partially linear additive transformation model. The proposed model makes use of constrained B -splines to approximate all unknown functions, i.e., the transformation and regression functions. In order to provide for more efficient estimates and to largely circumvent the need to carefully specify the basis function representation, a penalization strategy is used to regularize the estimation of the unknown functions. The novelty of our approach resides in the carefully constructed form of the proposed model which leads to the development of a computationally efficient and easy to implement hybrid algorithm for model fitting. The uniqueness of this algorithm is that it determines all penalty parameters as a part of the model fitting process; thus obviating the need to conduct a computationally expensive grid search to identify the same. Theoretically, it is shown that the proposed estimators of the regressions coefficients, which rely on the penalization and spline approximation technique, are asymptotically normal and efficient, i.e., they attain the semiparametric information

bound. Moreover, we establish that the estimators of the nonparametric components achieve the optimal rate of convergence for appropriately chosen order of the smoothing parameters and knots of spline spaces. Moreover, the estimator of the variance-covariance matrix of the estimated regression coefficients is shown to be asymptotically consistent and its finite-sample performance is also evaluated. Further, through numerical studies we demonstrate that the proposed methodology has better inferential and computational characteristics than the method developed by Cheng and Wang [2] and Lu and McMahan [13], the only existing, to our knowledge, competing approaches that are the most directly comparable.

The remainder of this manuscript is organized as follows. In Section 2, we develop a penalized partial linear additive transformation model for current status data. In Section 3, we outline the details of our model and the hybrid algorithm. In Section 4, we present the large-sample properties of the penalized estimators and the estimator of the variance-covariance matrix of the estimated regression coefficients. In Section 5, we summarize the results of a simulation study that was designed to examine the finite-sample performance of our approach. In Section 6, we further illustrate our methodology by applying it to HPV data collected as a part of the NHANES study. In Section 7, we conclude with a summary discussion. Technical details and proofs of the asymptotic properties are relegated to Appendix A and the additional results referenced in Sections 5 and 6 are available in Appendices B and D. Finally, the simulation results for comparison between the proposed penalized approach and the regression spline method of Lu and McMahan [13] are presented in Appendix C.

2. Model

Consider a study which examines n subjects for a failure time of interest. Let T_i denote the time that the i th subject experiences the failure, for $i = 1, \dots, n$. To provide for modeling flexibility, we assume that T_i obeys a partially linear additive transformation model which is given by

$$\eta(T_i) = -\mathbf{Z}_i^\top \boldsymbol{\beta} - \sum_{j=1}^J \varphi_j(W_{ij}) + \varepsilon_i, \quad (1)$$

where $\eta(\cdot)$ is an unknown monotone increasing transformation function on $(0, \infty)$ with $\eta(0) = -\infty$ and ε_i is a random variable having cumulative distribution function (CDF) $H(\cdot)$. In the above specification, $\boldsymbol{\beta}$ is a q -dimensional vector of regression coefficients corresponding to the covariate $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iq})^\top$ and $\varphi_j(\cdot)$ is an unknown smooth regression function taking the covariate W_{ij} as its argument, for $j = 1, \dots, J$. Throughout, we assume that $\varepsilon_i \perp \varepsilon_j$, for $i \neq j$, and $\varepsilon_i \perp \mathbf{X}_i$, where \perp denotes statistical independence and $\mathbf{X}_i = (\mathbf{Z}_i^\top, W_{i1}, \dots, W_{iJ})^\top$. This formulation provides for a very general modeling construct that holds several popular models as special cases. For example, specifying $H(\cdot)$ to be the CDF of the extreme value distribution causes (1) to reduce to the standard PH

model, while taking $H(\cdot)$ to be the CDF of the standard logistic distribution results in the PO model.

Under the assumed model, we note that the conditional CDF of T_i , given the covariate vector \mathbf{X}_i , can be expressed as

$$F(t|\mathbf{X}_i) = g^{-1} \left\{ \mathbf{z}_i^\top \boldsymbol{\beta} + \eta(t) + \sum_{j=1}^J \varphi_j(W_{ij}) \right\}, \quad (2)$$

where $g(\cdot)$ is a known, smooth, and strictly increasing link function on $(0, 1)$. Though equivalent to (1), the representation provided in (2) is more convenient for reasons that will shortly become apparent. Here $g(\cdot)$ assumes the role of $H(\cdot)$ with respect to determining the final form of the proposed model, e.g., specifying $g(x) = \log\{-\log(1-x)\}$ provides for the PH model and $g(x) = \log\{x/(1-x)\}$ results in the PO model.

A key feature of current status data is that the failure times (i.e., T_i , for $i = 1, \dots, n$) are not directly observed but rather are known relative to observation times, i.e., T_i is either left- or right-censored indicating that T_i occurred before or after the observation time, respectively. To mathematically frame the structure of this data, let Y_i denote the observation time for the i th individual and let $\Delta_i = I(T_i < Y_i)$ denote the corresponding censoring indicator, where $I(\cdot)$ is the usual indicator function. Note, $\Delta_i = 1(0)$ represents the event that the i th individual's failure time is left (right)-censored. Thus, the observed data consist of $\{(Y_i, \Delta_i, \mathbf{X}_i)\}_{i=1}^n$. Following the works of [16, 17, 18], we propose to estimate the unknown parameters via the following penalized log-likelihood

$$\begin{aligned} l_{\boldsymbol{\lambda}}(\boldsymbol{\tau}) = & \sum_{i=1}^n [\Delta_i \log F(Y_i|\mathbf{X}_i) + (1 - \Delta_i) \log\{1 - F(Y_i|\mathbf{X}_i)\}] \\ & - \frac{\lambda_0^2}{2} \mathcal{J}^2(\eta) - \sum_{j=1}^J \frac{\lambda_j^2}{2} \mathcal{J}^2(\varphi_j), \end{aligned}$$

where $\boldsymbol{\tau} = (\boldsymbol{\beta}, \eta, \varphi_1, \dots, \varphi_J)$ represents the collection of unknown model parameters, $\boldsymbol{\lambda} = (\lambda_0, \dots, \lambda_J)^\top$ is a vector of smoothing parameters, such that $\lambda_j \geq 0$ for all j , and

$$\mathcal{J}^2(\eta) = \int \{\eta^{(r)}(t)\}^2 dt, \quad \mathcal{J}^2(\varphi_j) = \int \{\varphi_j^{(r)}(W_j)\}^2 dW_j.$$

In the penalties above, $\eta^{(r)}(\cdot)$ and $\varphi_j^{(r)}(\cdot)$ denote the r th derivative of $\eta(\cdot)$ and $\varphi_j(\cdot)$, respectively, $r \geq 1$. This penalized log-likelihood is derived under the assumption that the covariates are time-invariant and that the failure and censoring times are conditionally independent, given the covariates. These assumptions are common among the literature, e.g., see [1, 2, 17] and the references therein.

3. Estimation

3.1. Penalized spline log-likelihood

In general, estimating β , $\eta(\cdot)$ and $\varphi_j(\cdot)$ directly from the penalized log-likelihood can be a tumultuous task. Thus, motivated by the works of [11, 17, 33], we propose to strike a balance between modeling flexibility and complexity by approximating all unknown functions in (2) through the use of constrained B -splines functions. In particular,

$$\eta(\cdot) \approx \eta_n(\cdot | \gamma) = \sum_{k=1}^{p_0} \gamma_k b_{0k}(\cdot) \quad \text{and} \quad \varphi_j(\cdot) \approx \varphi_{nj}(\cdot | \alpha_j^*) = \sum_{k=1}^{p_j} \alpha_{jk}^* b_{jk}^*(\cdot),$$

where $b_{0k}(\cdot)$ and $b_{jk}^*(\cdot)$ are B -spline basis functions with unknown coefficients $\gamma = (\gamma_1, \dots, \gamma_{p_0})^\top$ and $\alpha_j^* = (\alpha_{j1}^*, \dots, \alpha_{jp_j}^*)^\top$, respectively, $1 \leq j \leq J$. Once a knot set is chosen and the degree of the spline function is specified these basis functions are uniquely determined; for further discussion see Schumaker [21]. To insure monotonicity of $\eta_n(\cdot | \gamma)$ we require that $\gamma_1 \leq \dots \leq \gamma_{p_0}$; see Theorem 5.9 of Schumaker [21]. It is important to note that other techniques of enforcing the monotonicity constraint have been proposed but often lead to complex and unreliable model fitting strategies, e.g., see Cheng and Wang [2].

As with all additive models, $\varphi_j(\cdot)$ is only identifiable up to an additive constant. Thus, sum-to-zero constraints are imposed such that $\sum_{i=1}^n \varphi_{nj}(W_{ji} | \alpha_j^*) = 0$, for $1 \leq j \leq J$. For implementation purposes, the sum-to-zero constraints can be expressed as $\mathbf{B}_j^{*\top} \alpha_j^* = 0$, where $\mathbf{B}_j^* = \sum_{i=1}^n \mathbf{B}_{ji}^*$, $\mathbf{B}_{ji}^* = \{b_{j1}^*(W_{ji}), \dots, b_{jp_j}^*(W_{ji})\}^\top$, and $\alpha_j^* = (\alpha_{j1}^*, \dots, \alpha_{jp_j}^*)^\top$. Motivated by Wood [30], the model is reparameterized such that the constrained set of parameters α_j^* can be uniquely determined by an unconstrained reduced set $\alpha_j = (\alpha_{j1}, \dots, \alpha_{jp_j-1})^\top$. In particular, let \mathbf{Q}_j denote a $p_j \times (p_j - 1)$ semi-orthogonal matrix whose columns are orthogonal to \mathbf{B}_j^* . Obviously, $\mathbf{B}_j^{*\top} \mathbf{Q}_j \alpha_j = 0$. Thus, for any $\alpha_j^* \in \mathbb{R}^{p_j}$ satisfying the sum-to-zero constraint there exists $\alpha_j \in \mathbb{R}^{p_j-1}$ such that $\alpha_j^* = \mathbf{Q}_j \alpha_j$. Note, \mathbf{Q}_j can easily be found via a QR-decomposition of \mathbf{B}_j^* ; see Wood [30] for further details. Thus, the log-likelihood under the proposed spline model is obtained by replacing $F(Y_i | \mathbf{X}_i)$ in the penalized log-likelihood function by

$$F_n(Y_i | \mathbf{X}_i) = g^{-1} \left\{ \mathbf{Z}_i^\top \beta + \mathbf{M}_i^\top \gamma + \sum_{j=1}^J \mathbf{B}_{ji}^\top \alpha_j \right\}, \tag{3}$$

where $\mathbf{M}_i = \{b_{01}(Y_i), \dots, b_{0p_0}(Y_i)\}^\top$ and $\mathbf{B}_{ji} = \mathbf{Q}_j^\top \mathbf{B}_{ji}^*$, for $1 \leq i \leq n$. Thus, the set of unknown parameters to be estimated are given by $\theta = (\beta^\top, \gamma^\top, \alpha^\top)^\top$, where $\alpha = (\alpha_1^\top, \dots, \alpha_J^\top)^\top$.

Under this strategy, the penalized spline log-likelihood function is given by

$$l_{n,\lambda}(\theta) = \sum_{i=1}^n [\Delta_i \log F_n(Y_i | \mathbf{X}_i) + (1 - \Delta_i) \log \{1 - F_n(Y_i | \mathbf{X}_i)\}]$$

$$-\frac{\lambda_0^2}{2}\boldsymbol{\gamma}^\top\mathbf{D}_0\boldsymbol{\gamma}-\sum_{j=1}^J\frac{\lambda_j^2}{2}\boldsymbol{\alpha}_j^\top\mathbf{D}_j\boldsymbol{\alpha}_j, \quad (4)$$

where \mathbf{D}_0 and \mathbf{D}_j are band diagonal matrices whose (k, l) th elements are given by $\int b_k^{(r)}(y)b_l^{(r)}(y)dy$ and $\int b_{jk}^{(r)}(w_j)b_{jl}^{(r)}(w_j)dw_j$, respectively. Cubic splines are well-known for their numerical stability and are widely used in semiparametric regression in literature; e.g., see [3, 15, 11], among many others. For this reason, cubic splines are used to approximate the unknown functions in our numerical experiments and data application, and hence r is set to be 2. The algorithm for finding \mathbf{D}_0 and \mathbf{D}_j is discussed in Wood [30] and is implemented in the `smooth-Con` function in the R package `mgcv`. The penalized spline maximum likelihood estimator of $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta}} l_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta})$, where the maximization is completed subject to the monotonicity constraints, i.e., $\gamma_1 \leq \dots \leq \gamma_{p_0}$. Accordingly, the spline penalized estimator of $\boldsymbol{\tau}$ is defined as $\hat{\boldsymbol{\tau}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\eta}, \hat{\varphi}_1, \dots, \hat{\varphi}_J)^\top$ with $\hat{\eta}(\cdot) = \eta_n(\cdot | \hat{\boldsymbol{\gamma}})$ and $\hat{\varphi}_j(\cdot) = \varphi_{n_j}(\cdot | \hat{\boldsymbol{\alpha}}_j)$. The theoretical properties and numerical performance of $\hat{\boldsymbol{\tau}}$ are discussed in Section 4 and Section 5, respectively.

The proposed penalized spline estimation approach has several appealing features when compared to existing methods. In particular, under the smoothness assumption, η can be well approximated by a monotone spline which can be written as a linear combination of spline basis functions with a much smaller number of coefficients to estimate, for instance, $n^{1/3}$ instead of n , which is required by fully nonparameteric estimation methods like Ma and Kosorok [17]. Further, the performance of existing spline techniques including Cheng and Wang [2] and Lu and McMahan [13] is intrinsically tied to the specification of these basis functions. Regretfully, as with most spline based procedures, to effectively implement these techniques one must perform a grid search over different basis function representations, which can be computationally expensive, especially when there are multiple unknown functions to be estimated. On the other hand, via the proposed penalized approach we are able to regularize the estimation of these parameters to address the potential issue of over-fitting. These two features acting in concert provide an estimation framework which is both more accurate and more efficient. Moreover, as discussed in Section 3.2, our formulation leads to the development of a computationally efficient and easy to implement hybrid algorithm for model fitting. All of these features lead to a methodology that is of practical interest to practitioners who are left to analyze data of this form.

3.2. Parameter estimation

To complete model fitting in a computationally efficient manner, herein we develop a novel hybrid algorithm which can be used to simultaneously identify both the penalized spline estimators and the smoothing parameters via a nested iterative process. The inner process is aimed at identifying the value of $\boldsymbol{\theta}$ that maximizes (4) for a fixed value of $\boldsymbol{\lambda}$, while adhering to the monotonicity constraints, i.e., the inner step identifies $\tilde{\boldsymbol{\theta}}_{\boldsymbol{\lambda}} = \operatorname{argmax}_{\boldsymbol{\theta}} l_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta})$ s.t. $\gamma_1 \leq \dots \leq \gamma_{p_0}$.

To solve this constrained optimization problem, an iterative algorithm is developed. In each iteration of this algorithm, the unconstrained updates of the parameters, say $\boldsymbol{\theta}^*$, are first determined based on the current parameter value, say $\boldsymbol{\theta}^{(m)}$, via the following Fisher-scoring step

$$\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(m)} + \mathcal{I}_{n,\boldsymbol{\lambda}}^{-1}(\boldsymbol{\theta}^{(m)})\nabla l_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}^{(m)}),$$

where $\nabla l_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}^{(m)})$ and $\mathcal{I}_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}^{(m)})$ are the gradient and the expected negative value of the hessian matrix of (4) evaluated at $\boldsymbol{\theta}^{(m)}$, respectively. Closed form expressions for these quantities are given by

$$\begin{aligned}\nabla l_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}) &= \mathcal{X}^\top \boldsymbol{\Omega}(\boldsymbol{\theta}) \boldsymbol{\Delta}(\boldsymbol{\theta}) - \mathbf{S}\boldsymbol{\theta} \\ \mathcal{I}_{n,\boldsymbol{\lambda}}(\boldsymbol{\theta}) &= \mathcal{X}^\top \boldsymbol{\Omega}(\boldsymbol{\theta}) \mathcal{X} + \mathbf{S},\end{aligned}$$

where $\mathcal{X} = (\mathbf{Z}, \mathbf{M}, \mathbf{B}_1, \dots, \mathbf{B}_J)$, $\mathbf{Z}^\top = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)$, $\mathbf{M}^\top = (\mathbf{M}_1, \dots, \mathbf{M}_n)$, and $\mathbf{B}_j^\top = (\mathbf{B}_{j1}, \dots, \mathbf{B}_{jn})$, and $\mathbf{S} = \sum_{j=0}^J \lambda_j^2 \mathbf{S}_j$, with $\mathbf{S}_0 = \text{diag}\{\mathbf{0}_{q \times q}, \mathbf{D}_0, \mathbf{0}_{r_{20} \times r_{20}}\}$, $r_{20} = p_1 + \dots + p_J - J$, $\mathbf{S}_j = \text{diag}\{\mathbf{0}_{r_{1j} \times r_{1j}}, \mathbf{D}_j, \mathbf{0}_{r_{2j} \times r_{2j}}\}$, $r_{1j} = q + p_0 + \dots + p_{j-1} - (j-1)$, and $r_{2j} = p_{j+1} + \dots + p_J - (J-j)$, $2 \leq j \leq J-1$, and $\mathbf{S}_J = \text{diag}\{\mathbf{0}_{r_{1J} \times r_{1J}}, \mathbf{D}_J\}$, $r_{1J} = q + p_0 + \dots + p_{J-1} - (J-1)$. In the expressions above, the transformed response vector and weight matrix for a specific value of $\boldsymbol{\theta}$ are given by

$$\begin{aligned}\boldsymbol{\Delta}(\boldsymbol{\theta}) &= \left\{ \frac{\Delta_1 - \pi_1}{G'(\mathcal{X}_1^\top \boldsymbol{\theta})}, \dots, \frac{\Delta_n - \pi_n}{G'(\mathcal{X}_n^\top \boldsymbol{\theta})} \right\}^\top \\ \boldsymbol{\Omega}(\boldsymbol{\theta}) &= \text{diag} \left\{ \frac{G'(\mathcal{X}_1^\top \boldsymbol{\theta})^2}{\pi_1(1-\pi_1)}, \dots, \frac{G'(\mathcal{X}_n^\top \boldsymbol{\theta})^2}{\pi_n(1-\pi_n)} \right\},\end{aligned}$$

where \mathcal{X}_i is the i th row of \mathcal{X} , $G'(\cdot)$ is the first derivative of $G(\cdot) = g^{-1}(\cdot)$, and $\pi_i = g^{-1}(\mathcal{X}_i^\top \boldsymbol{\theta})$.

The updated values of γ available in $\boldsymbol{\theta}^*$ do not necessarily satisfy the monotonicity constraints; see Section 2. To enforce these conditions, isotonic regression is implemented to project the unconstrained estimates into the constrained space, i.e., into $\boldsymbol{\Gamma} = \{\boldsymbol{\gamma} : \gamma_1 \leq \dots \leq \gamma_{p_0}\}$. This is accomplished by solving the following minimization problem

$$\boldsymbol{\gamma}^{(m+1)} = \underset{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}}{\text{argmin}} \left\{ \sum_{k=1}^{p_0} \sigma_k^2 (\gamma_k - \gamma_k^*)^2 \right\},$$

where σ_k^2 is the diagonal element of $\mathcal{I}_{n,\boldsymbol{\lambda}}^{-1}(\boldsymbol{\theta}^*)$ corresponding to γ_k , for $k = 1, \dots, p_0$. This problem can be solved using the `pava` function in the R package `iso`, which takes $\boldsymbol{\gamma}^*$ and $(\sigma_1^2, \dots, \sigma_{p_0}^2)^\top$ as inputs. Substituting $\boldsymbol{\gamma}^{(m+1)}$ for $\boldsymbol{\gamma}^*$ in $\boldsymbol{\theta}^*$ one obtains the constrained update $\boldsymbol{\theta}^{(m+1)}$. These two steps are completed in turn until convergence is attained, i.e., the difference between $\boldsymbol{\theta}^{(m)}$ and $\boldsymbol{\theta}^{(m+1)}$ is less than some specified stopping criterion. At the point of convergence of this inner process, one obtains $\boldsymbol{\theta}_\lambda$, i.e., the maximizer of the penalized spline log-likelihood function for a fixed value of $\boldsymbol{\lambda}$. Once the inner process is complete,

the outer process updates the smoothing parameters and the inner process is restarted. To obtain the updated smoothing parameters, say $\boldsymbol{\lambda}^* = (\lambda_0^*, \dots, \lambda_J^*)^\top$, a generalized Fellner-Schall approach [31] is adopted and makes the update as

$$\lambda_j^{*2} = \frac{\text{tr}\{\mathbf{S}^- \mathbf{S}_j\} - \text{tr}\left\{\mathcal{I}_{n,\lambda}^{-1}(\tilde{\boldsymbol{\theta}}_\lambda) \mathbf{S}_j\right\}}{\tilde{\boldsymbol{\theta}}_\lambda^\top \mathbf{S}_j \tilde{\boldsymbol{\theta}}_\lambda} \lambda_j^2, \text{ for } 0 \leq j \leq J,$$

where \mathbf{S}^- is the generalized inverse of \mathbf{S} . Using the updated smoothing parameters (i.e., $\boldsymbol{\lambda}^*$), the inner process is then used to identify $\tilde{\boldsymbol{\theta}}_{\boldsymbol{\lambda}^*}$. This nested iterative process continues until the difference between $\tilde{\boldsymbol{\theta}}_\lambda$ and $\tilde{\boldsymbol{\theta}}_{\boldsymbol{\lambda}^*}$ is less than some specified stopping criterion. It is worthwhile to point out that the update of λ_j is guaranteed to be positive, which is required, as long as $\mathcal{I}_{n,\lambda}(\boldsymbol{\theta}_\lambda)$ is positive definite (see Wood and Fasiolo [31]), which is the case for the proposed model. At convergence of the proposed hybrid algorithm, the final update of $\boldsymbol{\theta}$ is defined as the spline penalized estimator $\hat{\boldsymbol{\theta}}$.

4. Large-sample properties

4.1. Assumptions

In this section we present the asymptotic properties of the penalized spline estimators. Let $\mathcal{T}_0 = [a_0, b_0]$ and $\mathcal{T}_j = [a_j, b_j]$ be supports of Y and W_j , respectively, for $j = 1, \dots, J$. Define \mathcal{M} as a space of monotone splines with degree $r + 1$ on the compact set \mathcal{T}_0 and \mathcal{S}_j as a space of splines with degree $r + 1$ on the compact set \mathcal{T}_j . Let the regression parameter space Φ be a compact set of \mathbb{R}^q and the nonparametric space be $\mathcal{N} = \{(\eta, \varphi_1, \dots, \varphi_J) : \eta \in \mathcal{M}, \varphi_j \in \mathcal{S}_j, \mathcal{J}(\eta) < \infty, \mathcal{J}(\varphi_j) < \infty\}$, where $\mathcal{J}(\cdot)$ is the Sobolev norm for a fixed integer $r \geq 1$ defined in Section 2. Denote by $\boldsymbol{\tau}_0 = (\beta_0, \varphi_{0,0}, \varphi_{0,1}, \dots, \varphi_{0,J})$ with $\varphi_{0,0} = \eta_0$ and $\hat{\boldsymbol{\tau}} = (\hat{\beta}, \hat{\varphi}_0, \hat{\varphi}_1, \dots, \hat{\varphi}_J)$ with $\hat{\varphi}_0 = \hat{\eta}$ the true value and the penalized spline likelihood estimator of $\boldsymbol{\tau}$, respectively. Let $\|\cdot\|_n$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ be empirical norm, L_2 -norm, and uniform norm, respectively. We established the asymptotic properties of $\hat{\boldsymbol{\tau}}$ under the following regularity conditions:

Condition 1: The true parameter β_0 is an interior point of Φ .

Condition 2: The r th derivative of $\varphi_{0,j}$, for $0 \leq j \leq J$, satisfies the Lipschitz condition on \mathcal{T}_j , and $\varphi_{0,0}$ is strictly increasing on \mathcal{T}_0 .

Condition 3: The support of the observation times is an interval within $[l_y, u_y]$ with $0 < l_y < u_y < \inf\{t : F_0(t) = 1\}$, where F_0 is the CDF of the observation times.

Condition 4: The covariate vector \mathbf{Z} has a bounded support and $E(\mathbf{Z}\mathbf{Z}^\top)$ is non-singular.

Condition 5: The first derivative of the link function $Q(\cdot) \equiv g^{-1}(\cdot)$ is bounded away from 0 and infinite, and the second derivative of $Q(\cdot)$ is bounded. Further, there exists some constant $0 < \varepsilon_0 < 1$ such that $\varepsilon_0 \leq Q(x) \leq 1 - \varepsilon_0$.

Condition 6: The joint density of $(Y, \mathbf{Z}, \mathbf{W}, \Delta)$ is bounded away from 0 and infinite.

Condition 7: For any $\beta \neq \beta_0$, $\Pr(\mathbf{Z}^\top \beta \neq \mathbf{Z}^\top \beta_0) > 0$ and $E\{\varphi_j(W_j)\} = 0$, for $1 \leq j \leq J$.

Condition 8: The smoothing parameters are specified such that

$$\lambda_j = O_p(n^{-r/(2r+1)}) \quad \text{and} \quad \lambda_j^{-1} = O_p(n^{r/(2r+1)}), \quad \text{for } 0 \leq j \leq J.$$

Condition 9: The ratio of maximum and minimum spacings of the knots for \mathcal{S}_j is uniformly bounded in n , for $0 \leq j \leq J$ with $\mathcal{S}_0 \equiv \mathcal{M}$. Further, the dimension of sum space $\mathcal{S}_+ = \mathcal{S}_0 + \dots + \mathcal{S}_J$ is $O(q_n)$ with $q_n = n^{r/(1+2r)}$.

Condition 10: The efficient information $\mathcal{I}(\beta_0)$ defined in (5) is non-singular.

It is noted that condition 1 is a standard assumption in semiparametric estimation. The smoothness assumption of $\varphi_{0,j}$ in condition 2 guarantees that the unknown nonparametric components can be well approximated by B -splines. Further, under the smooth condition, the upper bound of entropy for Soblev space is lower than that for a collection of monotone functions; see Examples of 19.10 and 19.11 of van der Vaart [29] for more information. It follows that the penalized estimator of φ_0 achieves faster rate of convergence compared to that of nonparametric alternative proposed by Ma and Kosorok (2005). Conditions 3–6 are necessary in entropy calculation used to derive the rate of convergence of $\hat{\tau}$ and the asymptotic normality of $\hat{\beta}$. Condition 7 is required to establish the identifiability of the model. Condition 8 is a typical assumption used to derive the asymptotic properties of the penalized estimators. Condition 9 specifies the appropriate order of dimensions of spline spaces to derive the optimal rate of convergence of $\hat{\tau}$. Finally, condition 10 is common for the asymptotic normality and the efficiency of $\hat{\beta}$.

4.2. Efficient information calculation

The log-likelihood of $(Y, \Delta, \mathbf{Z}, \mathbf{W})$ for τ is given by

$$\ell(\tau) = \Delta \log Q(\zeta) + (1 - \Delta)\{1 - Q(\zeta)\},$$

where $\zeta = \mathbf{Z}^\top \beta + \eta(Y) + \sum_{j=1}^J \varphi_j(W_j)$. Define $l(t) = q(t)/[Q(t)\{1 - Q(t)\}]$, where $q(t)$ is the first derivative of $Q(t)$. The score function for β takes the form

$$\dot{\ell}_\beta(\tau) = \{\Delta - Q(\zeta)\} l(\zeta) \mathbf{Z}.$$

Differentiating the log-likelihood $\ell(\tau)$ along the parametric submodels $\eta_t = \eta + ta$ for $\mathcal{J}(a) < \infty$ at $t = 0$ and $\varphi_{s_j} = \varphi_j + s_j h_j$ for $\mathcal{J}(h_j) < \infty$ at $s_j = 0$, for $1 \leq j \leq J$, yields the score operator for η

$$\dot{\ell}_\eta(\tau)[a] = \{\Delta - Q(\zeta)\} l(\zeta) a(Y)$$

and the score operator for φ_j

$$\dot{\ell}_{\varphi_j}(\tau)[h_j] = \{\Delta - Q(\zeta)\} l(\zeta) h_j(W_j),$$

respectively. Let $\mathcal{H}_0 = \{a : \mathcal{J}(a) < \infty\}$ and $\mathcal{H}_j = \{h_j : E\{h_j(\cdot)\} = 0, \mathcal{J}(h_j) < \infty\}$, for $1 \leq j \leq J$. In the sequel, to simplify notation, define $\varphi_0(\cdot) = \eta(\cdot)$ and $h_0(\cdot) = a(\cdot)$. For $\mathbf{h}_j = (h_{j,1}, \dots, h_{j,q})^\top \in \mathcal{H}_j^q$, $0 \leq j \leq J$, and $\mathbf{h} = \mathbf{h}_0 + \dots + \mathbf{h}_J$, define $\dot{\ell}_{\varphi_j}(\boldsymbol{\tau})[\mathbf{h}_j] = (\dot{\ell}_{\varphi_j}(\boldsymbol{\tau})[h_{j,1}], \dots, \dot{\ell}_{\varphi_j}(\boldsymbol{\tau})[h_{j,q}])^\top$ and $\dot{\ell}_\varphi(\boldsymbol{\tau})[\mathbf{h}] = \dot{\ell}_{\varphi_0}(\boldsymbol{\tau})[\mathbf{h}_0] + \dots + \dot{\ell}_{\varphi_J}(\boldsymbol{\tau})[\mathbf{h}_J]$. By applying the similar arguments to those in the proof of Lemma 5 in Lu and McMahan [13], we conclude that there exist unique functions $\mathbf{h}_1^*, \dots, \mathbf{h}_J^*$ that minimize the distance $\|\dot{\ell}_\beta(\boldsymbol{\tau}) - \sum_{j=0}^J \dot{\ell}_{\varphi_j}(\boldsymbol{\tau})[\mathbf{h}_j]\|_2^2$, for $\mathbf{h}_j \in \mathcal{H}_j^q$. Further, it can be shown that \mathbf{h}_j^* is bounded and smooth on \mathcal{T}_j^q ; see Lemma S.1 of Cheng et al. [3] for more information. Let $\mathbf{h}^* = \mathbf{h}_0^* + \dots + \mathbf{h}_J^*$. Define the efficient information of β at $\boldsymbol{\tau} = \boldsymbol{\tau}_0$ as

$$\mathcal{I}(\beta_0) = E \{ \dot{\ell}_\beta^*(\boldsymbol{\tau}_0) \}^{\otimes 2} \equiv E \left\{ \dot{\ell}_\beta(\boldsymbol{\tau}) - \dot{\ell}_\varphi(\boldsymbol{\tau})[\mathbf{h}^*] \right\}^{\otimes 2}. \quad (5)$$

4.3. Asymptotic results

Theorem 1 (Consistency and rate of convergence) *Under conditions 1–9, the penalized estimator $\hat{\beta}$ is asymptotically consistent for β_0 , $\|\hat{\varphi}_j\|_\infty = O_p(1)$, $\mathcal{J}(\hat{\varphi}_j) = O_p(1)$, $\|\hat{\varphi}_j - \varphi_{0,j}\|_\infty = o_p(1)$, and $\|\hat{\varphi}_j - \varphi_{0,j}\|_2 = O_p(n^{-r/(1+2r)})$, for $0 \leq j \leq J$.*

Theorem 2 (Asymptotic normality and efficiency) *Under conditions 1–10,*

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}\{0, \mathcal{I}^{-1}(\beta_0)\}, \text{ as } n \rightarrow \infty.$$

The proofs of these results are relegated to Appendix A. It is worthwhile to point out that if the smoothing parameters and the dimensions of spline spaces are chosen to be of the appropriate order (see conditions 8 and 9), our functional estimators are uniformly bounded and consistent on compact sets, i.e., uniform convergence can be achieved on boundaries for penalized spline estimators, and attain the optimal rate of convergence, $O_p(n^{-r/(1+2r)})$, which is the best attainable rate of convergence in the context of semiparametric regression; see Stone [23] for more information. Moreover, $\hat{\beta}$ is efficient in the semiparametric sense since it achieves the information bound.

4.4. Variance estimation

As discussed in Section 4.2, the least favorable direction $(\mathbf{h}_0^*, \dots, \mathbf{h}_J^*)$ is the one that minimizes the distance $\|\dot{\ell}_\beta(\boldsymbol{\tau}) - \sum_{j=0}^J \dot{\ell}_{\varphi_j}(\boldsymbol{\tau})[\mathbf{h}_j]\|_2^2$, for $\mathbf{h}_j \in \mathcal{H}_j^q$. Since the least favorable direction does not have a closed form, it is very demanding to estimate it directly, and so is the information matrix $\mathcal{I}(\beta_0)$. To address the issue, we adopt the least-square approach proposed by Zhang et al. [36]. In particular, we use B -splines to approximate $\mathbf{h}_j^* = (h_{j,1}^*, \dots, h_{j,q}^*)^\top$, $0 \leq j \leq J$, i.e., for $1 \leq l \leq q$,

$$h_{0,l}^*(\cdot) \approx h_{n0,l}^*(\cdot) = \sum_{k=1}^{p_0} \delta_{0l,k} b_{0k}(\cdot)$$

and

$$h_{j,l}^*(\cdot) \approx h_{n,j,l}^*(\cdot) = \sum_{k=1}^{p_j-1} \delta_{jl,k} b_{jk}(\cdot),$$

up to some constant. Here the sum-to-zero constraints are applied to $h_{n,j,l}^*(\cdot)$, for $1 \leq j \leq J$ and $1 \leq l \leq q$. The constraint only affects $h_{n,j,l}^*(\cdot)$ vertically to ensure its mean to be zero. Let

$$(\hat{\mathbf{h}}_0^*, \dots, \hat{\mathbf{h}}_J^*) = \min_{\mathbf{h}_j \in \mathcal{H}_{n_j}^q} \|\dot{\ell}_\beta(\hat{\boldsymbol{\tau}}) - \sum_{j=0}^J \dot{\ell}_{\varphi_j}(\hat{\boldsymbol{\tau}})[\mathbf{h}_j]\|_n^2,$$

where \mathcal{H}_{n_j} is the approximated B -spline space \mathcal{H}_j . For $\hat{\mathbf{h}}_j^* = (\hat{h}_{j,1}^*, \dots, \hat{h}_{j,q}^*)^\top$, $j = 0, \dots, J$, write $\hat{h}_{j,l}^*(\cdot) = \sum_{k=1}^{p_j} \hat{\delta}_{jl,k} b_{jk}(\cdot)$, for $1 \leq l \leq q$ and then define

$$\widehat{\mathcal{I}}_n = \mathbb{P}_n \left\{ \dot{\ell}_\beta(\hat{\boldsymbol{\tau}}) - \sum_{j=0}^J \dot{\ell}_{\varphi_j}(\hat{\boldsymbol{\tau}})[\hat{\mathbf{h}}_j^*] \right\}^2, \tag{6}$$

where \mathbb{P}_n is the empirical measure. In the above specification, $\hat{\boldsymbol{\delta}}_l = (\hat{\boldsymbol{\delta}}_{0,l}^\top, \dots, \hat{\boldsymbol{\delta}}_{J,l}^\top)^\top$ for $\hat{\boldsymbol{\delta}}_{j,l} = (\hat{\delta}_{jl,1}, \dots, \hat{\delta}_{jl,p_j})^\top$, for $j = 0, 1, \dots, J$, minimize

$$\mathbb{P}_n \left\{ \dot{\ell}_{\beta,l}(\hat{\boldsymbol{\tau}}) - \sum_{j=0}^J \dot{\ell}_{\varphi_j}(\hat{\boldsymbol{\tau}})[h_{n,j,l}^*] \right\}^2,$$

where $\dot{\ell}_{\beta,l}(\hat{\boldsymbol{\tau}})$ is the l th element of $\dot{\ell}_\beta(\hat{\boldsymbol{\tau}})$. By the standard least-square calculation,

$$\hat{\boldsymbol{\delta}}_l = (\mathbb{P}_n \mathbf{G}^{\otimes 2})^{-1} \mathbb{P}_n \left\{ \dot{\ell}_{\beta,l}(\hat{\boldsymbol{\tau}}) \mathbf{G} \right\},$$

where $\mathbf{G} = (\mathbf{G}_0^\top, \dots, \mathbf{G}_J^\top)^\top$ with $\mathbf{G}_j = (\dot{\ell}_{\varphi_j}(\hat{\boldsymbol{\tau}})[b_{j1}], \dots, \dot{\ell}_{\varphi_j}(\hat{\boldsymbol{\tau}})[b_{jp_j}])^\top$, for $j = 0, 1, \dots, J$. Thus,

$$\begin{aligned} \widehat{\mathcal{I}}_n &= \mathbb{P}_n \left[\dot{\ell}_\beta(\hat{\boldsymbol{\tau}}) - \mathbb{P}_n \left\{ \dot{\ell}_\beta(\hat{\boldsymbol{\tau}}) \mathbf{G}^\top \right\} (\mathbb{P}_n \mathbf{G}^{\otimes 2})^{-1} \mathbf{G} \right]^{\otimes 2} \\ &= \mathbb{P}_n \left\{ \dot{\ell}_\beta(\hat{\boldsymbol{\tau}}) \right\}^{\otimes 2} - \mathbb{P}_n \left\{ \dot{\ell}_\beta(\hat{\boldsymbol{\tau}}) \mathbf{G}^\top \right\} (\mathbb{P}_n \mathbf{G}^{\otimes 2})^{-1} \mathbb{P}_n \left\{ \mathbf{G} \dot{\ell}_\beta^\top(\hat{\boldsymbol{\tau}}) \right\}. \end{aligned}$$

Clearly, by the formula of the inverse of block matrix, $\widehat{\mathcal{I}}_n^{-1}$ is the upper $q \times q$ submatrix of the inverse of the observed information of (4) for $\boldsymbol{\beta}$ without the penalized term. It can be shown that $\widehat{\mathcal{I}}_n$ is a consistent estimator of $\mathcal{I}(\boldsymbol{\beta}_0)$. For the purposes of conducting large-sample inference, we proposed to estimate the variance-covariance matrix of $\boldsymbol{\beta}$ by $\widehat{\mathcal{I}}_{n,\boldsymbol{\lambda}}$, which is the upper $q \times q$ submatrix of

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{\mathcal{X}}^\top \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}) \boldsymbol{\mathcal{X}} + \mathbf{S})^{-1} \boldsymbol{\mathcal{X}}^\top \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}) \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}^\top \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}) \boldsymbol{\mathcal{X}} + \mathbf{S})^{-1}.$$

Using this estimator one can conduct standard Wald type inference about the regression coefficients. Theorem 3 shows that $\widehat{\mathcal{I}}_{n,\boldsymbol{\lambda}}$ and $\widehat{\mathcal{I}}_n$ are asymptotically equivalent, and hence $\mathcal{I}(\boldsymbol{\beta}_0)$ can be consistently estimated by $\widehat{\mathcal{I}}_{n,\boldsymbol{\lambda}}$.

Theorem 3 (Variance estimation) Under conditions 1–10, $\widehat{\mathcal{I}}_{n,\lambda}$ is a consistent estimator of $\mathcal{I}(\beta_0)$.

Moreover, the simulation study presented in Section 5 indicates that the proposed variance estimation technique performs well in practical situations.

5. Simulation Study

To assess the finite-sample performance of the proposed methodology the following simulation study was conducted. In this study, the true failure time T was generated according to the following model

$$g_\alpha\{F(t|\mathbf{X})\} = Z_1\beta_1 + Z_2\beta_2 + \eta(t) + \varphi_1(W_1) + \varphi_2(W_2), \quad (7)$$

where $W_j \sim \mathcal{U}[-1, 1]$, for $j = 1, 2$, $Z_1 \sim \text{Bernoulli}(0.5)$ and $Z_2 \sim \mathcal{N}(0, 1)$. To provide for a broad range of examples, the link function $g_\alpha(\cdot)$ was assumed to belong to a family of links indexed by a parameter α as:

$$g_\alpha(u) = \begin{cases} \log \frac{(1-u)^{-\alpha}-1}{\alpha} & \text{if } \alpha > 0, \\ \log\{-\log(1-u)\} & \text{if } \alpha = 0. \end{cases} \quad (8)$$

Proceeding in this fashion allows us to capture several common survival models, i.e., $\alpha = 0$ leads to a PH model, while $\alpha = 1$ corresponds to a PO model.

To examine different forms of the unknown functions and regression coefficients three scenarios were considered. In Scenario 1 (S1) the three functions were specified as $\eta(t) = \log(2t)$, $\varphi_1(w) = \exp(w + 0.5) - \{\exp(1.5) - \exp(-0.5)\}/2$, and $\varphi_2(w) = 2\sin(-\pi w)$, and the regression coefficients were chosen to be $\beta_1 = 0.5$ and $\beta_2 = -0.5$; in Scenario 2 (S2) the three functions were specified as $\eta(t) = \log\{1.5t - \log(1 + 1.5t)\}$, $\varphi_1(w) = 2\sin(-\pi w)$, and $\varphi_2(w) = 4w^2 - 4/3$, and the regression coefficients were set to be $\beta_1 = 0.5$ and $\beta_2 = 0.5$; in Scenario 3 (S3) the three functions were specified as $\eta(t) = \log\{\log(1 + t/10) + \sqrt{t}/10\}$, $\varphi_1(w) = 4w^2 - 4/3$, and $\varphi_2(w) = \exp(w + 0.5) - \{\exp(1.5) - \exp(-0.5)\}/2$, and the regression coefficients were set to be $\beta_1 = -0.5$ and $\beta_2 = -0.5$. In each of S1-S3, we consider values of $\alpha \in \{0, 0.5, 1\}$. To create current status data, observation times (Y) were sampled from an exponential distribution with a mean of 2, 2, and 1 under S1, S2, and S3, respectively, and the censoring indicator was determined as $\Delta = I(T < Y)$. These specifications were made to yield a variety of right-censoring rates; for example when $\alpha = 0(1)$, we have right censoring rates of 27% (36%), 44% (51%), and 77% (81%) under S1, S2, and S3, respectively. For each simulation configuration, this data generating process was repeated 1000 times to create 1000 independent data sets consisting of $n = 200$ or 400 observations.

The proposed methodology was used to analyze each of the simulated data sets. For purposes of comparison, we also implement the method of Cheng and Wang [2]. To compute both estimators, cubic B -splines were used to approximate all unknown functions with interior knots being placed at equally-spaced

quantiles of the covariate values. For each data set, the proposed approach made use of a single knot set consisting of $\lceil n^{1/3} \rceil$ interior knots. In contrast, the competing estimator was computed using multiple knot set configurations with the number of interior knots ranging from $\lceil n^{1/3} \rceil - 2$ to $\lceil n^{1/3} \rceil + 2$, with the final unpenalized estimator being selected based on the BIC criterion; for further discussion see Cheng and Wang [2]. The estimator of the variance-covariance matrix of $\hat{\beta}$ developed in Section 4 was used to construct 95% Wald confidence intervals for the estimated regression coefficients, while the observed information approach proposed in Cheng and Wang [2] was applied for inference on the regression parameter estimators obtained from the competing approach.

Tables 1 and 2 summarize the results of estimating the regression coefficients across all considered simulation configurations for both the proposed penalized spline method and the competing approach with $n = 200$ and 400 , respectively. Provided in this summary are the empirical bias, sample standard deviation, and mean squared error of the 1000 point estimates, as well as the average of the 1000 estimated standard errors and empirical coverage probability associated with 95% confidence intervals. From these results it is apparent that the proposed approach works well, i.e., the penalized spline estimators of the regression coefficients exhibit little bias, the mean squared errors of estimates decrease as sample size increases, the standard deviations of the point estimates are in agreement with the average estimated standard errors, and the empirical coverage probabilities are at their nominal level. The latter two findings tend to suggest that the proposed method of estimating the variance-covariance matrix works well and that large-sample inference is possible even for relatively small sample sizes. Further, the proposed penalized spline estimator performed as good if not better than the competing technique across all considered simulation configurations even though this competing estimator was chosen as the best from amongst multiple candidates. In fact, the performance of this competing technique varied dramatically (results not shown) across the different knot set configurations, i.e., it was found to be sensitive to the specification of the number and placement of the basis functions. This sensitivity resulted in the necessity to conduct a grid search over the different knot set configurations. This search resulted in an average model fitting time that was between approximately 20 and 100 times slower than the proposed approach.

One of the primary goals of this study was to assess the performance of the proposed method with respect to estimating the unknown functions. To this end, Figures 1 and 2 summarize the estimates of the unknown functions under S2 with all considered values of α for both approaches when $n = 200$ and 400 , respectively. This summary includes the 0.025, 0.5, and 0.975 point-wise quantiles of the estimates. Four figures providing analogous results under S1 and S3 when $n = 200$ and 400 , respectively, are presented in Appendix B (i.e., Figures 4, 5, 6 and 7). These results suggest that the proposed methodology can be used to accurately estimate all unknown functions; i.e., very little bias can be observed when one compares the median of the functional estimates to the true underlying function. The same can not be said for the competing technique.

TABLE 1

Summary of the results that were obtained using the competing technique (Cheng and Wang[2]) and the proposed penalized spline estimator when $\alpha = 0$ (PH model), $\alpha = 0.5$, and $\alpha = 1$ (PO model) under $n = 200$. Presented results include the empirical bias (Bias), sample standard deviation (SD), average of the 1000 estimated standard errors (SE), mean squared error (MSE) of the 1000 point estimates, as well as empirical coverage probability associated with 95% confidence intervals (CP) and average model fitting time (TIME; measured in Seconds).

$n = 200$		Cheng and Wang						Penalized Spline Estimator					
		Bias	SD	SE	MSE	CP	TIME	Bias	SD	SE	MSE	CP	TIME
S1($\alpha = 0.0$)	β_1	0.075	0.467	0.394	0.224	92.2	871	-0.008	0.364	0.320	0.133	92.6	13
	β_2	-0.117	0.288	0.222	0.097	91.3		-0.004	0.223	0.173	0.050	90.2	
S2($\alpha = 0.0$)	β_1	0.096	0.437	0.382	0.200	93.3	821	0.014	0.363	0.326	0.131	94.8	15
	β_2	0.085	0.231	0.210	0.061	94.1		0.003	0.204	0.173	0.041	94.0	
S3($\alpha = 0.0$)	β_1	-0.060	0.457	0.415	0.212	94.5	436	-0.028	0.410	0.384	0.169	94.6	7
	β_2	-0.062	0.242	0.220	0.062	93.5		-0.043	0.228	0.202	0.054	92.6	
S1($\alpha = 0.5$)	β_1	0.038	0.420	0.403	0.177	94.3	827	0.002	0.367	0.364	0.134	96.1	13
	β_2	-0.056	0.234	0.215	0.058	94.6		0.006	0.198	0.191	0.039	95.5	
S2($\alpha = 0.5$)	β_1	0.073	0.465	0.415	0.221	92.6	831	0.027	0.400	0.376	0.160	94.3	17
	β_2	0.075	0.243	0.220	0.065	94.1		0.011	0.210	0.197	0.044	95.1	
S3($\alpha = 0.5$)	β_1	-0.020	0.502	0.466	0.252	93.9	454	0.013	0.473	0.442	0.224	93.5	8
	β_2	-0.055	0.266	0.245	0.074	94.1		-0.037	0.253	0.230	0.065	94.0	
S1($\alpha = 1.0$)	β_1	0.009	0.479	0.443	0.229	93.3	825	-0.000	0.426	0.408	0.181	94.2	16
	β_2	-0.056	0.246	0.232	0.064	94.2		-0.006	0.221	0.212	0.049	94.8	
S2($\alpha = 1.0$)	β_1	0.025	0.478	0.453	0.229	94.7	828	0.000	0.416	0.417	0.173	95.6	18
	β_2	0.049	0.245	0.238	0.062	95.0		-0.003	0.223	0.216	0.050	95.0	
S3($\alpha = 1.0$)	β_1	-0.016	0.562	0.513	0.316	94.6	443	0.023	0.530	0.493	0.281	95.2	8
	β_2	-0.037	0.298	0.266	0.090	93.7		-0.022	0.275	0.255	0.076	94.2	

In particular, the approach of Cheng and Wang [2] provides estimates that are far more biased and variable than our approach. These findings reinforce the assertion that the proposed penalized spline estimator performs better with respect to both estimation and inference than this existing procedure.

To further examine the performance of our approach, we conducted a second simulation study to compare our method to the proposal of Lu and McMahan [13]. It is important to note that this existing technique was developed for the partially linear Cox PH model; i.e., this approach can only accommodate a single unknown function of a covariate. Thus, in this study comparisons were made under the partially linear Cox PH model. In particular, three data generating scenarios and two sample sizes ($n = 200$ and $n = 400$) were considered. Under each setting 1000 data sets were created and analyzed by the two techniques. More details on the specific simulation settings are provided in Appendix C. In addition, the simulation results are summarized and presented in Table 4 and Figures 8 and 9 in Appendix C. In conclusion, both of these techniques can accurately estimate the regression coefficients and unknown functions and draw

TABLE 2

Summary of the results that were obtained using the competing technique (Cheng and Wang[2]) and the proposed penalized spline estimator when $\alpha = 0$ (PH model), $\alpha = 0.5$, and $\alpha = 1$ (PO model) under $n = 400$. Presented results include the empirical bias (Bias), sample standard deviation (SD), average of the 1000 estimated standard errors (SE), mean squared error (MSE) of the 1000 point estimates, as well as empirical coverage probability associated with 95% confidence intervals (CP) and average model fitting time (TIME; measured in Seconds).

$n = 400$	Cheng and Wang						Penalized Spline Estimator						
	Bias	SD	SE	MSE	CP	TIME	Bias	SD	SE	MSE	CP	TIME	
S1($\alpha = 0.0$)	β_1	0.038	0.280	0.246	0.080	92.1	1176	-0.019	0.229	0.214	0.053	93.5	20
	β_2	-0.042	0.149	0.136	0.024	93.4		0.023	0.126	0.115	0.016	92.1	
S2($\alpha = 0.0$)	β_1	0.037	0.252	0.244	0.065	95.2	1188	0.006	0.227	0.222	0.051	94.7	18
	β_2	0.032	0.147	0.131	0.023	93.1		-0.010	0.129	0.118	0.017	93.2	
S3($\alpha = 0.0$)	β_1	-0.029	0.296	0.270	0.088	92.6	610	-0.018	0.270	0.256	0.073	93.8	9
	β_2	-0.035	0.156	0.143	0.025	93.3		-0.022	0.144	0.134	0.021	93.3	
S1($\alpha = 0.5$)	β_1	0.006	0.288	0.285	0.083	93.2	1181	-0.022	0.251	0.250	0.063	94.8	17
	β_2	-0.031	0.153	0.147	0.024	92.5		0.014	0.136	0.131	0.019	94.5	
S2($\alpha = 0.5$)	β_1	0.035	0.297	0.274	0.089	93.9	1206	0.008	0.265	0.258	0.070	94.4	20
	β_2	0.023	0.152	0.144	0.023	94.6		-0.014	0.139	0.134	0.019	95.5	
S3($\alpha = 0.5$)	β_1	-0.022	0.308	0.307	0.095	95.9	653	0.005	0.294	0.298	0.086	95.8	6
	β_2	-0.015	0.166	0.161	0.028	94.1		-0.006	0.157	0.154	0.025	94.1	
S1($\alpha = 1.0$)	β_1	0.017	0.328	0.299	0.107	93.6	1187	0.011	0.298	0.281	0.089	93.3	18
	β_2	-0.032	0.169	0.155	0.029	94.0		0.008	0.154	0.146	0.024	94.5	
S2($\alpha = 1.0$)	β_1	0.012	0.327	0.307	0.107	94.2	1191	-0.008	0.292	0.289	0.085	95.4	20
	β_2	0.039	0.172	0.161	0.031	92.8		0.004	0.154	0.150	0.024	93.9	
S3($\alpha = 1.0$)	β_1	-0.035	0.360	0.341	0.131	94.5	634	-0.001	0.338	0.333	0.114	95.6	11
	β_2	-0.027	0.187	0.177	0.036	94.1		-0.013	0.176	0.172	0.031	94.7	

reliable inference. That being said, our approach performs as good if not better than this existing technique. In particular, it is observed that our method yields a smaller bias, variability, and mean squared error for the regression coefficients and attains less variability in the estimated functions. More importantly, our approach is far more computational efficient than the approach of Lu and McMahan [13]. That is, our approach is able to complete model fitting approximately 15 to 70 times faster than this existing technique.

6. Application to HPV data

To further illustrate our approach, the proposed methodology is now used to analyze HPV data which was collected as a part of the National Health and Nutrition Examination Survey (NHANES). HPV is one of the most common sexually transmitted diseases (STD) in the United States and persistent infections may lead to cancer (e.g., cervical cancer) and genital warts. Thus, given these potential sequelae and the asymptomatic nature with which this disease

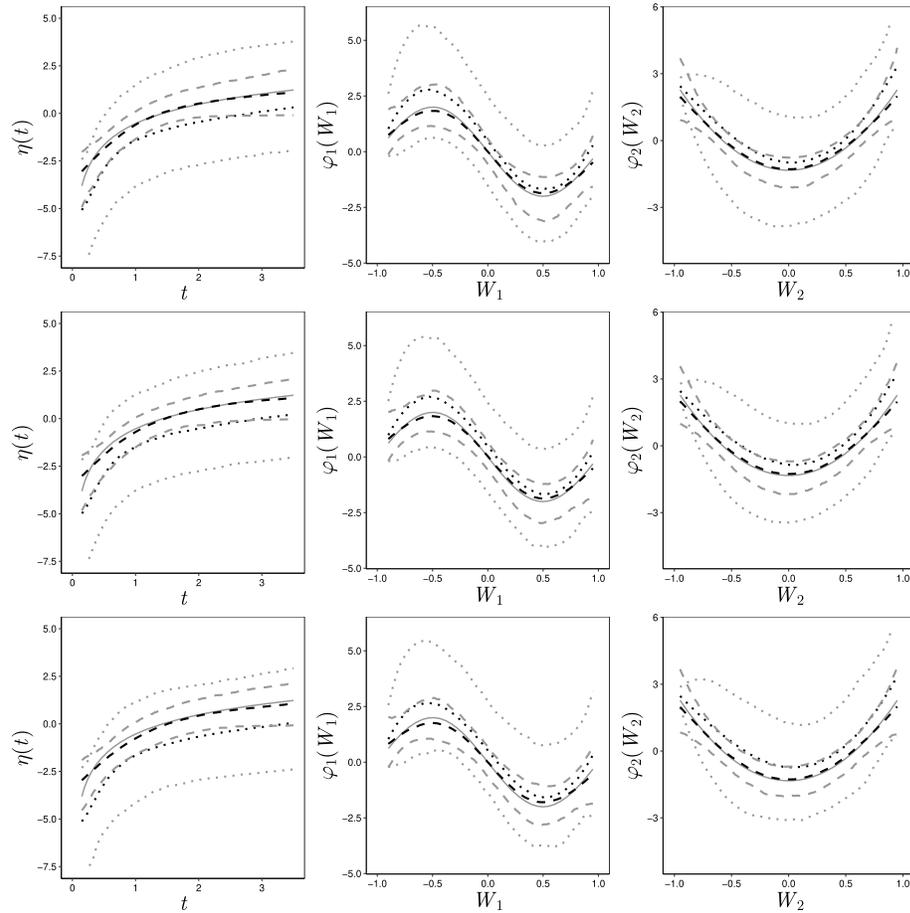


FIG 1. Estimates of the unknown functions under scenario $S2$ when $n = 200$. From top to bottom, the panels correspond to $\alpha = 0, 0.5, 1.0$, respectively. The solid grey curves correspond to the true value of the functions. Dashed lines are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the penalized spline approach. Dotted curves are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the approach of Cheng and Wang [2].

presents, it is essential to monitor the prevalence of this disease within the population of the United States, and elsewhere.

To this end, we consider analyzing the most recent data released by the NHANES study containing HPV test results (2015-2016). In total, there are 2840 participants and the left-censoring rate is approximately 41.7%. Given that HPV is an STD, we define the HPV onset time relative to the reported time of first sexual intercourse for each participant. In addition to HPV status, several covariate were also available; namely, gender ($Z_1 = 1$ if male), race ($Z_2 = 1$ if Mexican American, $Z_3 = 1$ if Other Hispanic, $Z_4 = 1$ if Non-Hispanic Black

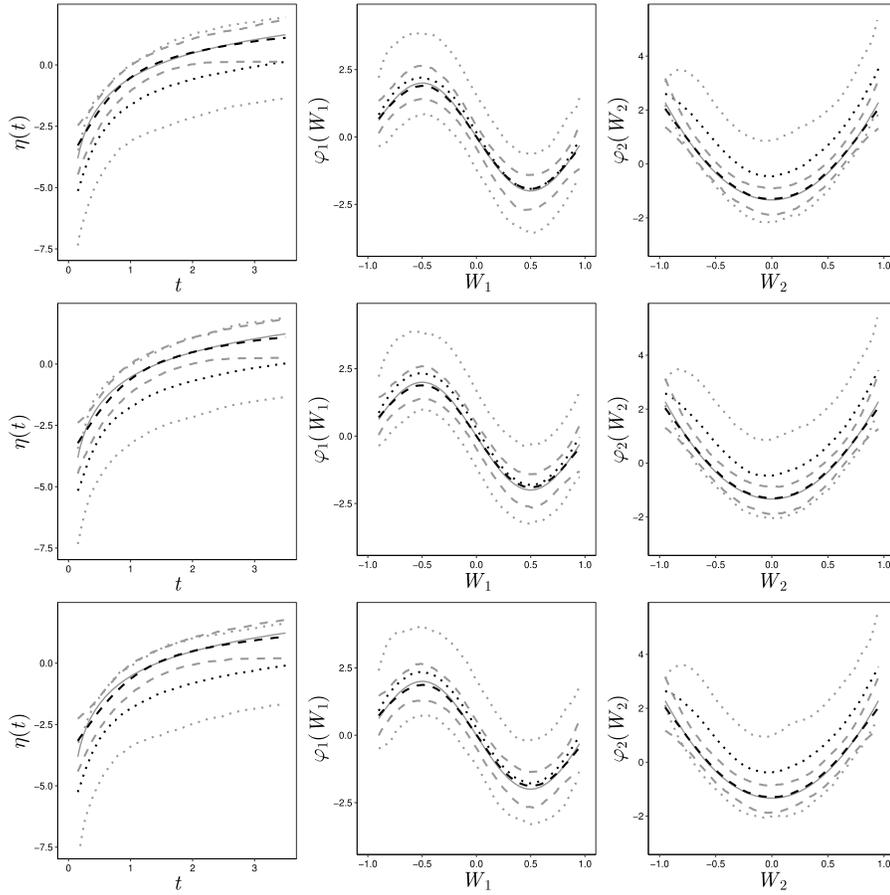


FIG 2. Estimates of the unknown functions under scenario S2 when $n = 400$. From top to bottom, the panels correspond to $\alpha = 0, 0.5, 1.0$, respectively. The solid grey curves correspond to the true value of the functions. Dashed lines are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the penalized spline approach. Dotted curves are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the approach of Cheng and Wang [2].

and $Z_5 = 1$ if Other; with Non-Hispanic White as the reference), ratio of family income to poverty (W_1) and the participant’s age at first sexual intercourse (W_2). To identify how the continuous covariates should enter into the final model (i.e., W_1 and W_2), the model was first fit allowing for nonlinear effects in these two variables. From this fit, there was very little evidence of nonlinearity in W_1 ; see Figure 13 in Appendix D. Thus, the HPV onset time is related to risk factors through the following model

$$g_\alpha\{F(t|\mathbf{X})\} = Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3 + Z_4\beta_4 + Z_5\beta_5 + W_1\beta_6 + \eta(t) + \varphi(W_2),$$

where $g_\alpha(u)$ is defined as in (8). In fitting the proposed model, cubic B -splines

TABLE 3

Human Papillomavirus data: Presented results include the parameter estimates, estimated standard errors (SE) and 95% Wald type confidence intervals (95% C.I.). The results were obtained from the penalized spline method and two competing approaches under both the PH and PO models.

	Parameter	PH Model			PO Model		
		Estimate	SE	95% C.I.	Estimate	SE	95% C.I.
Penalized Approach	Male	-0.025	0.060	(-0.142, 0.092)	-0.032	0.079	(-0.187, 0.124)
	Mexican American	-0.290	0.096	(-0.478, -0.102)	-0.361	0.121	(-0.598, -0.124)
	Other Hispanic	0.085	0.098	(-0.108, 0.277)	0.135	0.131	(-0.121, 0.392)
	Non-Hispanic Black	0.356	0.077	(0.204, 0.508)	0.503	0.108	(0.291, 0.715)
	Other race	-0.183	0.105	(-0.389, 0.022)	-0.237	0.131	(-0.493, 0.020)
	Income	-0.078	0.020	(-0.117, -0.040)	-0.101	0.026	(-0.152, -0.050)
Cheng and Wang	Male	0.139	0.056	(0.028, 0.249)	0.017	0.079	(-0.138, 0.173)
	Mexican American	0.257	0.085	(0.090, 0.424)	-0.268	0.121	(-0.505, -0.032)
	Other Hispanic	0.350	0.095	(0.163, 0.536)	0.178	0.131	(-0.079, 0.435)
	Non-Hispanic Black	0.591	0.078	(0.439, 0.743)	0.586	0.109	(0.373, 0.799)
	Other race	0.271	0.102	(0.072, 0.470)	-0.319	0.132	(-0.577, -0.061)
	Income	-0.068	0.019	(-0.105, -0.030)	-0.112	0.026	(-0.163, -0.062)
Lu and McMahan	Male	-0.019	0.060	(-0.136, 0.098)			
	Mexican American	-0.272	0.095	(-0.458, -0.087)			
	Other Hispanic	0.095	0.097	(-0.094, 0.284)			
	Non-Hispanic Black	0.370	0.077	(0.219, 0.522)			
	Other race	-0.153	0.107	(-0.363, 0.057)			
	Income	-0.082	0.020	(-0.120, -0.043)			

with a knot set consisting of $\lceil n^{1/3} \rceil = 15$ interior knots were used to approximate all unknown functions. Model fitting was completed under both PH ($\alpha = 0$) and PO ($\alpha = 1$) models using the hybrid algorithm outlined in Section 4. The same variance-covariance estimation strategies as those used in Section 5 were used to conduct inference for the regression coefficients.

Table 3 summarizes the regression parameter estimates, along with their estimated standard errors and corresponding 95% confidence intervals. Results obtained from both the PH and PO models are very similar. The results suggest that gender is not a significant factor, Mexican Americans have a significantly lower and Non-Hispanic Blacks have a significantly higher hazard of developing HPV when compared to Non-Hispanic Whites, while Other Hispanic and Other race do not. The ratio of family income to poverty has a significantly negative linear effect. Figure 3 provides the estimates of $\eta(\cdot)$ and $\varphi(\cdot)$, along with 95% point-wise confidence intervals. The lower and upper limits of the 95% point-wise confidence intervals were obtained as the 0.025 and 0.975 point-wise quantiles of 1000 bootstrap estimates, respectively. The estimated functions obtained under the PH model show similar features to those obtained under the PO model. The estimate of $\varphi(\cdot)$ clearly exhibits a nonlinear pattern, i.e., age at first sexual

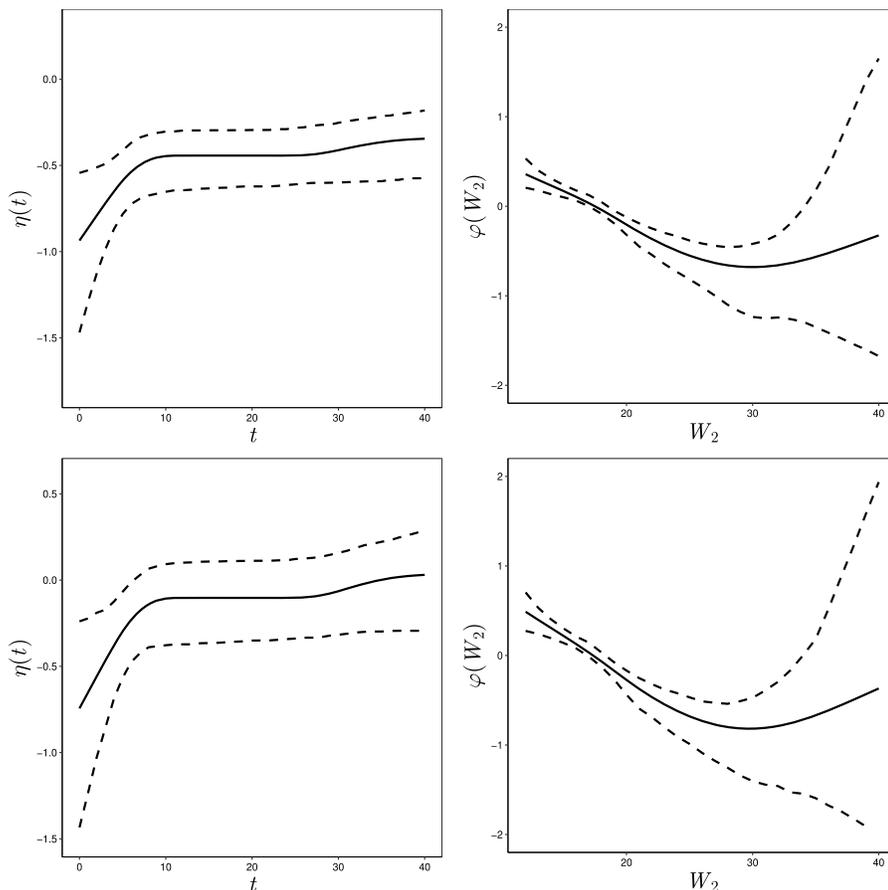


FIG 3. Penalized functional estimates for $\eta(\cdot)$ and $\varphi(\cdot)$. First and second row correspond to the PH and PO model, respectively. The solid black curve corresponds to the estimated function and the dashed curves correspond to the 0.025 and 0.975 quantiles of 1000 estimated functions based on bootstrap samples.

intercourse has a nonlinear impact on the hazard of contracting HPV. Note, the confidence interval for $\varphi(\cdot)$ becomes wider at the right end which is attributable to the small number (1.48%) of individuals whose age at first sexual intercourse exceeds 30; see Figure 10 in Appendix D for a histogram of these ages.

For comparative purposes, we also fit the models proposed by Cheng and Wang [2] and Lu and McMahan [13] to the HPV data. To achieve the best performance for these competing approaches, for each of the functions $\eta(\cdot)$ and $\varphi(\cdot)$, we conducted a grid search to identify the optimal number of interior knots where the candidate models had knot set configurations with the number of interior knots ranging from 1 to 20. The final model for each of these competing techniques was selected from among these candidates based on the BIC criterion.

The results from these competing approaches are also included in Table 3. From these results, one will first note that the estimates of the regression coefficients and the associated 95% confidence intervals obtained from the proposed method and that of Lu and McMahan [13] exhibit a large degree of similarity under PH model, but differ quite a bit from those obtained from the approach of Cheng and Wang [2]. Note, for our approach the results obtained under the PH and PO models do not contradict each other. However, for the approach of Cheng and Wang [2] the results under the PH and PO model do contradict each other. Second, to complete a model fitting, the proposed approach took 4 seconds, however due to the need to conduct a grid search to determine the appropriate number of interior knots, the approach of Cheng and Wang [2] took approximately 4 hours, and Lu and McMahan [13] took around 3 minutes. The estimates of $\eta(\cdot)$ and $\varphi(\cdot)$ obtained by the two competing approaches are summarized in Figures 11 and 12 in Appendix D.

Finally, to select the best model for these data, we note that under our general model the 2 candidates (i.e., the PH or PO model) are special cases which correspond to the specification of α . For this reason, a common approach used to select from amongst these models is to simply examine which possess a larger likelihood value [34, 39]. In order to account for the penalization in the proposed approach, we proceed similarly but make use of a modified BIC criterion. The modification involves replacing the usual degrees of freedom with an expression that quantifies the effective degrees of freedom that results due to penalization. In particular, following the work of [8] (i.e., §6.8.3) and [32], the effective degrees of freedom is given by $\text{trace}\{(\mathcal{X}^\top \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}) \mathcal{X}) \mathcal{I}_{n,\lambda}^{-1}(\hat{\boldsymbol{\theta}})\}$, where $\mathcal{I}_{n,\lambda}(\hat{\boldsymbol{\theta}}) = \mathcal{X}^\top \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}}) \mathcal{X} + \mathbf{S}$ is the negative of the expected Hessian matrix of the penalized log likelihood (4). Expressions for \mathcal{X} , $\boldsymbol{\Omega}(\hat{\boldsymbol{\theta}})$, and \mathbf{S} are provided in Section 3.2. As a result, the effective degrees of freedom for the PH and PO models are 14.58 and 14.45, respectively. Accordingly the modified BIC is 3771.9 and 3769.6 for the PH and PO model. Thus, the PO model is slightly preferred for this considered data set.

7. Discussion

In this work we developed a computationally efficient penalized approach for fitting a flexible partially linear additive transformation model to current status data. The proposed procedure demonstrates appealing properties, both numerical and theoretical. The hybrid algorithm developed for model fitting is easy to implement and robust to initialization. A simple variance-covariance estimation procedure was established and was shown to provide reliable Wald-type inference. In addition to attractive finite-sample performance, the estimators of the nonparametric components are shown to attain the optimal rate of convergence and we prove that estimators of regression coefficients are asymptotically normal and efficient in the semiparametric sense. To further disseminate this work, code (written in R) which implements all aspects of this research has been prepared and is available upon request from the corresponding author.

Given the successes of this approach, future work will be directed at generalizing this methodology to allow for the analysis of interval-censored data. A key obstacle in the development of this methodology will be the model fitting strategy, which will be far more complicated due to the complex nature of interval-censored data. Another direction for future research could involve further extending this work to allow for multiple end points, both for current status and interval censored data. Moreover, this penalized spline technique could be used to facilitate the estimation of other survival models, such as semiparametric generalized probit models for informative current status data [6]. Lastly, and potentially most challenging, would be the development of goodness-of-fit tests that could be used to guide the specification of the link function. This is a notoriously hard problem in the transformation model setting, and is only made harder when one considers censored data, and/or additive models.

Appendix A: Proofs

A.1. Notations and technical lemmas

For a probability measure P and a class of measurable functions \mathcal{F} , $H(\varepsilon, \mathcal{F}, L_2(P))$ is the ε -entropy for $L_2(P)$ -metric. In view of standard spline estimation and Lemma A1 of Lu et al. [14], there exist a monotone B -spline $\varphi_{n,0} \in \mathcal{S}_0$ and B -splines $\varphi_{n,j} \in \mathcal{S}_j$ such that $\|\varphi_{0,0} - \varphi_{n,0}\|_\infty = O(\rho_n)$ and $\|\varphi_{0,j} - \varphi_{n,j}\|_\infty = O(\rho_n)$, for $\rho_n = n^{-r/(1+2r)}$, $r \geq 1, j = 1, \dots, J$. Write $\boldsymbol{\tau}_n = (\boldsymbol{\beta}_0, \varphi_{n,0}, \dots, \varphi_{n,J})$. Let $\hat{\zeta}$, ζ_n , and ζ_0 denote ζ evaluated at $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}$, $\boldsymbol{\tau} = \boldsymbol{\tau}_n$, and $\boldsymbol{\tau} = \boldsymbol{\tau}_0$, respectively. Define $\bar{Q}(\zeta) = \{Q(\zeta) + Q(\zeta_n)\}/2$. In the sequel, let $\mathcal{J}(\zeta) = \sum_{j=0}^J \mathcal{J}(\varphi_j)$ and $\mathcal{J}^2(\zeta, \boldsymbol{\lambda}) = \sum_{j=0}^J \lambda_j^2 \mathcal{J}^2(\varphi_j)$ with $\varphi_0(\cdot) = \eta(\cdot)$. Define

$$\mathcal{G} = \left\{ \mathbf{Z}^T \boldsymbol{\beta} + \sum_{j=0}^J \varphi_j : \boldsymbol{\beta} \in \boldsymbol{\Phi}, \varphi_j \in \mathcal{S}_j, \mathcal{J}(\varphi_j) < \infty \right\}.$$

The following are key results to derive the consistency and rate of convergence of $\hat{\boldsymbol{\tau}}$. Lemma 4 is Lemma 8.4 of van de Geer [29], and Lemma 5 is a generalization of Lemma 7.2 of Mammen and van de Geer [18] to additive models.

Lemma 4 *Suppose that U_1, \dots, U_n are independent random variables with expectation 0 satisfying the uniformly sub-Gaussian condition, i.e.,*

$$\max_{i=1, \dots, n} K^2 \{E \exp(U_i^2/K^2) - 1\} \leq \sigma_0^2$$

for some fixed positive constants K and σ_0 and that

$$H(\varepsilon, \mathcal{F}, \|\cdot\|_n) \leq A\varepsilon^{-\alpha},$$

for all $\varepsilon > 0$ and some constants $A > 0$ and $0 < \alpha < 2$. Further, assume $\sup_{f \in \mathcal{F}} \|f\|_n \leq R$ for some constant $R > 0$. Then for some constant $C > 0$

depending on A, α, R, K and σ_0 , we have for all $T \geq C$,

$$P \left(\sup_{f \in \mathcal{F}} \frac{n^{1/2} \|Uf\|_n}{\|f\|_n^{1-\frac{\alpha}{2}}} \geq T \right) \leq C \exp(-T^2/C^2).$$

Lemma 5 For every probability measure P and every $\varepsilon > 0$,

$$\sup_P H \left(\varepsilon, \left\{ \frac{\bar{Q}^{1/2}(\zeta)}{1 + \mathcal{J}(\zeta)} : \zeta \in \mathcal{G} \right\}, L_2(P) \right) \leq C \left(\frac{1}{\varepsilon} \right)^{1/r},$$

for some constant $C > 0$.

A.2. Proof of Lemma 5

The proof follows similar arguments to those in the proof of Lemma 7.2 of Mammen and van de Geer [29]. For any function $f(\cdot)$ defined on $[x_1, x_2]$ with $\mathcal{J}(f) < \infty$, by a Taylor expansion, for $x \in [x_1, x_2]$,

$$f(x) = \sum_{k=0}^{r-1} \frac{f^{(k)}(x_1)}{k!} (x - x_1)^k + \int_{x_1}^x \frac{f^{(r)}(u)}{(r-1)!} (x - u)^{r-1} du \equiv f_1(x) + f_2(x).$$

In view of the Cauchy-Schwarz inequality, we have $\sup_{x \in [x_1, x_2]} |f_2(x)| \leq C_0 \mathcal{J}(f)$ for $C_0 > 0$ and $\mathcal{J}(f) = \mathcal{J}(f_2)$. Without loss of generality, assume $0 < C_0 < 1$, i.e., $\sup_{x \in [x_1, x_2]} |f_2(x)| \leq \mathcal{J}(f)$. Thus, write $\zeta = \zeta_1 + \zeta_2$, where $\zeta_1 = \mathbf{Z}^\top \boldsymbol{\beta} + \sum_{j=0}^J \varphi_{1,j}$ and $\zeta_2 = \sum_{j=0}^J \varphi_{2,j}$ with $\mathcal{J}(\varphi_j) = \mathcal{J}(\varphi_{2,j})$, for $j = 0, 1, \dots, J$. Clearly, $\mathcal{J}(\zeta) = \mathcal{J}(\zeta_2)$. Let \mathcal{H} be a class of measurable functions h with $\|h\|_\infty \leq M$ and $\mathcal{J}(h) \leq M$ for a constant $M > 0$. According to Lemma 2.4 of van de Geer [29], for any $\varepsilon > 0$, the entropy number $H(\varepsilon, \mathcal{H}, \|\cdot\|_\infty)$ is of the order $(1/\varepsilon)^{1/r}$, and hence the uniform entropy number of \mathcal{H} for L_2 -norm is also of the order $(1/\varepsilon)^{1/r}$, i.e., for some constant $C > 0$,

$$\sup_P H(\varepsilon, \mathcal{H}, L_2(P)) \leq C \left(\frac{1}{\varepsilon} \right)^{1/r}. \tag{9}$$

For a fixed bounded function h , by Example 3.7.4d of van de Geer [29], the class of uniformly bounded functions $\zeta_1 + h$ is a Vapnik-Chervonenkis subgraph class of index bounded by $q + rJ + 2$, and hence, Lemma 19.15 of van der Vaart [28] applies and concludes

$$\sup_P H(\varepsilon, \{\zeta_1 + h\}, L_2(P)) \leq C \log \left(\frac{1}{\varepsilon} \right). \tag{10}$$

Under the assumptions $\|Q'\|_\infty \leq M$ and $\varepsilon_0 \leq Q(x) \leq 1 - \varepsilon_0$ for $M > 0$ and $0 < \varepsilon_0 < 1$, we have

$$\left| \bar{Q}^{1/2}(\zeta) - \bar{Q}^{1/2}(\tilde{\zeta}) \right| \leq \frac{1}{2} \varepsilon_0^{-1/2} M |\zeta - \tilde{\zeta}|,$$

i.e., $\bar{Q}^{1/2}(\cdot)$ is Lipschitz. Thus, we read

$$\sup_P H\left(\varepsilon, \left\{\overline{Q}^{1/2}(\zeta_1 + h)\right\}, L_2(P)\right) \leq C_1 \log\left(\frac{1}{\varepsilon}\right), \tag{11}$$

for some constant $C_1 > 0$. Write

$$\nu(\zeta) = \left\lfloor \frac{1}{\{1 + \mathcal{J}(\zeta)\}\varepsilon} \right\rfloor \varepsilon,$$

where $\lfloor \cdot \rfloor$ is the floor function. In view of $1/2 < a \lfloor 1/a \rfloor < 1$ for $0 < a < 1$ and $\|\zeta_2\|_\infty \leq \mathcal{J}(\zeta)$, we read $\mathcal{J}(\nu(\zeta)\zeta_2) \leq 1$ and $\|\nu(\zeta)\zeta_2\|_\infty \leq 1$. Thus, the entropy result (9) holds for $\{\nu(\zeta)\zeta_2\}$, i.e.,

$$\sup_P H(\varepsilon, \{\nu(\zeta)\zeta_2\}, L_2(P)) \leq C_2 \left(\frac{1}{\varepsilon}\right)^{1/r}, \tag{12}$$

for some constant $C_2 > 0$. Then, for $\zeta = \zeta_1 + \zeta_2$, by entropy results (11) and (12), there exist γ_k for $1 \leq k \leq \exp\{C_2(1/\varepsilon)^{1/r}\}$ such that

$$\|\nu(\zeta)\zeta_2 - \gamma_k\|_n \leq \varepsilon$$

and ρ_l for $1 \leq l \leq (1/\varepsilon)^{C_1}$ such that

$$\left\| \overline{Q}^{1/2}\left(\zeta_1 + \frac{\gamma_k}{\nu(\zeta)}\right) - \overline{Q}^{1/2}\left(\rho_l + \frac{\gamma_k}{\nu(\zeta)}\right) \right\|_n \leq \varepsilon.$$

Thus, in view of the fact that $\overline{Q}^{1/2}(\cdot)$ is Lipschitz and bounded, together with $1/2 < a \lfloor 1/a \rfloor < 1$ for $0 < a < 1$, we have

$$\begin{aligned} & \left\| \frac{\overline{Q}^{1/2}(\zeta_1 + \zeta_2)}{1 + \mathcal{J}(\zeta)} - \overline{Q}^{1/2}\left(\rho_l + \frac{\gamma_k}{\nu(\zeta)}\right) \nu(\zeta) \right\|_n \\ & \leq \frac{1}{1 + \mathcal{J}(\zeta)} \left\| \overline{Q}^{1/2}(\zeta_1 + \zeta_2) - \overline{Q}^{1/2}\left(\zeta_1 + \frac{\gamma_k}{\nu(\zeta)}\right) \right\|_n \\ & \quad + \overline{Q}^{1/2}\left(\zeta_1 + \frac{\gamma_k}{\nu(\zeta)}\right) \left| \frac{1}{1 + \mathcal{J}(\zeta)} - \nu(\zeta) \right| \\ & \quad + \nu(\zeta) \left\| \overline{Q}^{1/2}\left(\zeta_1 + \frac{\gamma_k}{\nu(\zeta)}\right) - \overline{Q}^{1/2}\left(\rho_l + \frac{\gamma_k}{\nu(\zeta)}\right) \right\|_n \\ & \leq C_3\varepsilon + C_4\varepsilon + \varepsilon, \end{aligned}$$

for some constants $C_3, C_4 > 0$. The Lemma follows.

A.3. Proof of Theorem 1 (Consistency and rate of convergence)

For observations $(\Delta_1, Y_1, \mathbf{X}_1), \dots, (\Delta_n, Y_n, \mathbf{X}_n)$, the log-likelihood function for τ can be written as

$$l_n(\tau; Q) = \mathbb{P}_n[\Delta \log\{Q(\zeta)\} + (1 - \Delta) \log\{1 - Q(\zeta)\}].$$

In view of the concavity of logarithm function, together with $\bar{Q}(\zeta_n) = Q(\zeta_n)$, we have

$$\begin{aligned} l_n(\hat{\tau}; \bar{Q}) - l_n(\tau_n; \bar{Q}) &= \mathbb{P}_n \left\{ \Delta \log \frac{\bar{Q}(\hat{\zeta})}{Q(\zeta_n)} + (1 - \Delta) \log \frac{1 - \bar{Q}(\hat{\zeta})}{1 - Q(\zeta_n)} \right\} \\ &\geq \frac{1}{2} \mathbb{P}_n \left\{ \Delta \log \frac{Q(\hat{\zeta})}{Q(\zeta_n)} + (1 - \Delta) \log \frac{1 - Q(\hat{\zeta})}{1 - Q(\zeta_n)} \right\}. \end{aligned}$$

Then, by the definition of $\hat{\tau}$,

$$\frac{1}{4} \mathcal{J}^2(\hat{\zeta}, \boldsymbol{\lambda}) - \frac{1}{4} \mathcal{J}^2(\zeta_n, \boldsymbol{\lambda}) \leq l_n(\hat{\tau}; \bar{Q}) - l_n(\tau_n; \bar{Q}). \quad (13)$$

On the other hand, by the inequality $\log(x) < 2(\sqrt{x} - 1)$ for $x > 0$,

$$l_n(\hat{\tau}; \bar{Q}) - l_n(\tau_n; \bar{Q}) \quad (14)$$

$$\begin{aligned} &\leq 2\mathbb{P}_n \left\{ \Delta \left(\sqrt{\frac{\bar{Q}(\hat{\zeta})}{Q(\zeta_0)}} - 1 \right) + (1 - \Delta) \left(\sqrt{\frac{1 - \bar{Q}(\hat{\zeta})}{1 - Q(\zeta_0)}} - 1 \right) \right\} \\ &\quad + 2\mathbb{P}_n \Delta \left(\sqrt{\frac{\bar{Q}(\hat{\zeta})}{Q(\zeta_n)}} - \sqrt{\frac{\bar{Q}(\hat{\zeta})}{Q(\zeta_0)}} \right) \\ &\quad + 2\mathbb{P}_n (1 - \Delta) \left(\sqrt{\frac{1 - \bar{Q}(\hat{\zeta})}{1 - Q(\zeta_n)}} - \sqrt{\frac{1 - \bar{Q}(\hat{\zeta})}{1 - Q(\zeta_0)}} \right) \equiv I_1 + I_2 + I_3. \end{aligned} \quad (15)$$

Write $U = \Delta - E(\Delta|\mathbf{X})$. We have

$$I_1 = 2\mathbb{P}_n U \left(\sqrt{\frac{\bar{Q}(\hat{\zeta})}{Q(\zeta_0)}} - 1 \right) - 2\mathbb{P}_n U \left(\sqrt{\frac{1 - \bar{Q}(\hat{\zeta})}{1 - Q(\zeta_0)}} - 1 \right) - h_n^2(\hat{\zeta}, \zeta_0),$$

where

$$h_n^2(\hat{\zeta}, \zeta_0) = \left\| \sqrt{\bar{Q}(\hat{\zeta})} - \sqrt{Q(\zeta_0)} \right\|_n^2 + \left\| \sqrt{1 - \bar{Q}(\hat{\zeta})} - \sqrt{1 - Q(\zeta_0)} \right\|_n^2.$$

Next we show that $I_2 = I_3 = O_p(\rho_n^2) \left\{ 1 + \mathcal{J}(\hat{\zeta}) \right\}^{1/(2r)}$. Under the assumption $\varepsilon_0 \leq Q(x) \leq 1 - \varepsilon_0$ for $0 < \varepsilon_0 < 1$, the class

$$\mathcal{F}_1 = \left\{ \frac{1}{1 + \mathcal{J}(\zeta)} \left(\sqrt{\frac{\bar{Q}(\zeta)}{Q(\zeta_n)}} - \sqrt{\frac{\bar{Q}(\zeta)}{Q(\zeta_0)}} \right), \zeta \in \mathcal{G} \right\}$$

is uniformly bounded. Applying Lemma 5 yields

$$H(\varepsilon, \mathcal{F}_1, \|\cdot\|_n) \leq C \left(\frac{1}{\varepsilon} \right)^{1/r},$$

for some constant $C > 0$. Further, the sub-Gaussian condition holds for binary indicator Δ . It concludes from Lemma 4 that

$$\begin{aligned} & \mathbb{P}_n \Delta \left(\sqrt{\frac{\overline{Q}(\widehat{\zeta})}{Q(\zeta_n)}} - \sqrt{\frac{\overline{Q}(\widehat{\zeta})}{Q(\zeta_0)}} \right) \\ &= O_p(n^{-1/2}) \|Q(\zeta_n) - Q(\zeta_0)\|_n^{1-1/(2r)} \{1 + \mathcal{J}(\widehat{\zeta})\}^{1/(2r)}. \end{aligned}$$

Here we use the fact

$$\left\| \sqrt{\frac{\overline{Q}(\zeta)}{Q(\zeta_n)}} - \sqrt{\frac{\overline{Q}(\zeta)}{Q(\zeta_0)}} \right\|_n \leq C \|Q(\zeta_n) - Q(\zeta_0)\|_n.$$

Since $Q'(\cdot)$ is assumed to be bounded and $\|\zeta_n - \zeta_0\|_\infty = O(\rho_n)$, we obtain $I_2 = O_p(\rho_n^2) \{1 + \mathcal{J}(\widehat{\zeta})\}^{1/(2r)}$. The same result can also be read for I_3 . Combining (13) and (14) with the results of I_2 and I_3 yields

$$\begin{aligned} \frac{1}{4} \mathcal{J}^2(\widehat{\zeta}, \boldsymbol{\lambda}) + h_n^2(\widehat{\zeta}, \zeta_0) &\leq 2\mathbb{P}_n U \left(\sqrt{\frac{\overline{Q}(\widehat{\zeta})}{Q(\zeta_0)}} - 1 \right) - 2\mathbb{P}_n U \left(\sqrt{\frac{1 - \overline{Q}(\widehat{\zeta})}{1 - Q(\zeta_0)}} - 1 \right) \\ &+ \frac{1}{4} \mathcal{J}^2(\zeta_n, \boldsymbol{\lambda}) + O_p(\rho_n^2) \{1 + I(\widehat{\zeta})\}^{1/(2r)}. \end{aligned} \tag{16}$$

Next, applying the similar arguments to those for deriving I_2 and I_3 yields

$$\begin{aligned} & \mathbb{P}_n U \left(\sqrt{\frac{\overline{Q}(\widehat{\zeta})}{Q(\zeta_0)}} - 1 \right) \\ &= O_p(n^{-1/2}) \left\| \sqrt{\overline{Q}(\widehat{\zeta})} - \sqrt{Q(\zeta_0)} \right\|_n^{1-1/(2r)} \{1 + \mathcal{J}(\widehat{\zeta})\}^{1/(2r)} \end{aligned} \tag{17}$$

and

$$\begin{aligned} & \mathbb{P}_n U \left(\sqrt{\frac{1 - \overline{Q}(\widehat{\zeta})}{1 - Q(\zeta_0)}} - 1 \right) \\ &= O_p(n^{-1/2}) \left\| \sqrt{1 - \overline{Q}(\widehat{\zeta})} - \sqrt{1 - Q(\zeta_0)} \right\|_n^{1-1/(2r)} \{1 + \mathcal{J}(\widehat{\zeta})\}^{1/(2r)}. \end{aligned} \tag{18}$$

Thus, combing (17) and (18) with (16), we derive the basic inequality

$$\begin{aligned} & \frac{1}{4} \mathcal{J}^2(\widehat{\zeta}, \boldsymbol{\lambda}) + h_n^2(\widehat{\zeta}, \zeta_0) \\ & \leq O_p(n^{-1/2}) \left\{ h_n^2(\widehat{\zeta}, \zeta_0) \vee O_p(\rho_n^2) \right\}^{(2r-1)/(4r)} \{1 + \mathcal{J}(\widehat{\zeta})\}^{1/(2r)} \\ & + \frac{1}{4} \mathcal{J}^2(\zeta_n, \boldsymbol{\lambda}), \end{aligned} \tag{19}$$

which implies either

$$\frac{1}{4}\mathcal{J}^2(\widehat{\zeta}, \boldsymbol{\lambda}) + h_n^2(\widehat{\zeta}, \zeta_0) \leq \frac{1}{4}\mathcal{J}^2(\zeta_n, \boldsymbol{\lambda}) \quad (20)$$

or

$$\begin{aligned} & \frac{1}{4}\mathcal{J}^2(\widehat{\zeta}, \boldsymbol{\lambda}) + h_n^2(\widehat{\zeta}, \zeta_0) \\ & \leq \left\{ O_p(n^{-1/2})h_n^{1-1/(2r)}(\widehat{\zeta}, \zeta_0) \vee O_p(\rho_n^2) \right\} \left\{ 1 + \mathcal{J}(\widehat{\zeta}) \right\}^{1/(2r)}. \end{aligned} \quad (21)$$

Clearly, in view of the assumption of the order of λ_j , $j = 0, \dots, J$, (20) implies $h_n(\widehat{\zeta}, \zeta_0) = O_p(\rho_n)$ and $\mathcal{J}(\widehat{\zeta}) = O_p(1)$. Inequality (21) implies either

$$\frac{1}{4}\mathcal{J}^2(\widehat{\zeta}, \boldsymbol{\lambda}) + h_n^2(\widehat{\zeta}, \zeta_0) \leq O_p(\rho_n^2) \left\{ 1 + \mathcal{J}(\widehat{\zeta}) \right\}^{1/(2r)} \quad (22)$$

or

$$\frac{1}{4}\mathcal{J}^2(\widehat{\zeta}, \boldsymbol{\lambda}) + h_n^2(\widehat{\zeta}, \zeta_0) \leq O_p(n^{-1/2})h_n^{1-1/(2r)} \left\{ 1 + \mathcal{J}(\widehat{\zeta}) \right\}^{1/(2r)}. \quad (23)$$

In view of the assumption of the order of λ_j , (22) implies $\mathcal{J}(\widehat{\zeta}) \leq O_p(1)\{1 + \mathcal{J}(\widehat{\zeta})\}^{1/(2r)}$, and hence $\mathcal{J}(\widehat{\zeta}) = O_p(1)$. As a result, it concludes from (22) that $h_n(\widehat{\zeta}, \zeta_0) = O_p(\rho_n)$. Further, (23) implies

$$h_n(\widehat{\zeta}, \zeta_0) \leq O_p(\rho_n) \left\{ 1 + \mathcal{J}(\widehat{\zeta}) \right\}^{1/(1+2r)}. \quad (24)$$

Combing (23) and (24) yields $\mathcal{J}^2(\widehat{\zeta}, \boldsymbol{\lambda}) \leq O_p(\rho_n^2) \left\{ 1 + \mathcal{J}(\widehat{\zeta}) \right\}^{2/(1+2r)}$. It concludes from the assumption of the order of λ_j that $\mathcal{J}(\widehat{\zeta}) = O_p(1)$, and hence, in view of (24), $h_n(\widehat{\zeta}, \zeta_0) = O_p(\rho_n)$. In either case, $\mathcal{J}(\widehat{\zeta}) = O_p(1)$ and $h_n(\widehat{\zeta}, \zeta_0) = O_p(\rho_n)$, and in turn this implies

$$\|Q(\widehat{\zeta}) - Q(\zeta_0)\|_n = O_p(\rho_n).$$

Here we use the triangle inequality and $\|Q(\zeta_n) - Q(\zeta_0)\|_n = O_p(\rho_n)$. Applying the similar arguments to those in the proof of Lemma 5, we can establish

$$\sup_P H \left(\varepsilon, \left\{ \frac{Q(\zeta) - Q(\zeta_0)}{1 + \mathcal{J}(\zeta)}, \zeta \in \mathcal{G} \right\}, L_2(P) \right) \leq C \left(\frac{1}{\varepsilon} \right)^{1/r}.$$

Thus, the class $\{Q(\zeta) - Q(\zeta_0)\}/\{1 + \mathcal{J}(\zeta)\}$ for $\zeta \in \mathcal{G}$ is a Donsker class, and hence is a Glivenko-Cantelli class. Since $\|Q(\widehat{\zeta}) - Q(\zeta_0)\|_n = o_p(1)$, together with $\mathcal{J}(\widehat{\zeta}) = O_p(1)$, it follows from Glivenko-Cantelli theorem that

$$\|Q(\widehat{\zeta}) - Q(\zeta_0)\|_2 = \|Q(\widehat{\zeta}) - Q(\zeta_0)\|_n + o_p(1) \left\{ 1 + \mathcal{J}(\widehat{\zeta}) \right\} = o_p(1),$$

i.e., $Q(\widehat{\zeta})$ converges to $Q(\zeta_0)$ in L_2 -norm. Thus, Lemma 19.24 of van der Vaart [28] applies and yields that

$$\|Q(\zeta) - Q(\zeta_0)\|_2 = \|Q(\zeta) - Q(\zeta_0)\|_n + o_p(n^{-1/2}) \left\{1 + \mathcal{J}(\widehat{\zeta})\right\} = O_p(\rho_n).$$

Under the assumption $|Q'(\zeta)| > \eta_0 > 0$ for all ζ in a neighborhood of ε_0 , we obtain the rate of convergence for $\widehat{\zeta}$, i.e.,

$$\|\widehat{\zeta} - \zeta_0\|_2 = O_p(\rho_n).$$

Then, by Lemma 3.1 of Stone [25],

$$\|\mathbf{Z}^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_2^2 + \left\| \sum_{j=0}^J (\widehat{\varphi}_j - \varphi_{0,j})(W_j) \right\|_2^2 \leq O_p(1) \|\widehat{\zeta} - \zeta_0\|_2^2 = O_p(\rho_n^2).$$

Thus, $\|\mathbf{Z}^\top(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_2 = O_p(\rho_n)$, and hence $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 = O_p(\rho_n)$ results from the non-singularity assumption of $E(\mathbf{Z}\mathbf{Z}^\top)$. Further, by Lemma 1 of Stone [24], we have $\|\widehat{\varphi}_j - \varphi_{0,j}\|_2 = O_p(\rho_n)$, $j = 0, \dots, J$. The rate of convergence for $\widehat{\boldsymbol{\tau}}$ follows. By Lemma 7.3 of Murphy and van der Vaart [20], $\|\widehat{\varphi}_j\|_\infty \leq C \{\mathcal{J}(\widehat{\varphi}_j) + \|\widehat{\varphi}_j\|_2\}$, $j = 0, \dots, J$, for some constant $C > 0$. Since $\mathcal{J}(\widehat{\varphi}_j) = O_p(1)$ and $\|\widehat{\varphi}_j\|_2 \leq \|\widehat{\varphi}_j - \varphi_{0,j}\|_2 + \|\varphi_{0,j}\|_2 = O_p(1)$, $\|\widehat{\varphi}_j\|_\infty = O_p(1)$ follows. Finally, by Cauchy-Schwarz inequality and $J(\widehat{\varphi}_j) = O_p(1)$, the first derivative of $\widehat{\varphi}_j(\cdot)$ is uniformly bounded in probability, and hence is uniformly equicontinuous in probability on the compact set \mathcal{T}_j . The uniform convergence of $\widehat{\varphi}(\cdot)$ follows from the Arzel-Ascoli theorem.

Proof of Theorem 2 (Asymptotic normality and efficiency)

We first show that $\widehat{\boldsymbol{\tau}}$ satisfies the efficient score equations $\mathbb{P}_n \ell_{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\tau}}) = 0$, up to a $o_p(n^{-1/2})$ term. Differentiating the penalized log-likelihood function $l_{\boldsymbol{\lambda}}(\widehat{\boldsymbol{\beta}} + \mathbf{t}, \widehat{\varphi}_0 - \mathbf{t}^\top \mathbf{h}_0^*, \dots, \widehat{\varphi}_J - \mathbf{t}^\top \mathbf{h}_J^*)$ at $\mathbf{t} = \mathbf{0}$ with \mathbf{h}_j^* satisfying $\mathcal{J}(\mathbf{h}_j^*) < \infty$, for $0 \leq j \leq J$, yields

$$\mathbb{P}_n \ell_{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\tau}}) + \sum_{j=0}^J \lambda_j^2 \int \widehat{\varphi}_j^{(r)}(u_j) \mathbf{h}_j^{*(r)}(u_j) du_j = 0. \tag{25}$$

The second term in (25) can be bounded by $\sum \lambda_j^2 \mathcal{J}(\widehat{\varphi}_j) \mathcal{J}(\mathbf{h}_j^*)$ as a result of Cauchy-Schwarz inequality. In view of $\mathcal{J}(\widehat{\varphi}_j) = O_p(1)$, together with the assumption of the order of λ_j , we derive the stationary equation

$$\mathbb{P}_n \ell_{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\tau}}) = \mathbb{P}_n \left\{ \Delta - Q(\widehat{\zeta}) \right\} l(\widehat{\zeta}) \mathbf{D} = o_p(n^{-1/2}), \tag{26}$$

where $\mathbf{D} = \mathbf{Z} - \sum_{j=0}^J \mathbf{h}_j^*$. Let D_k be k th element of the vector \mathbf{D} , for $1 \leq k \leq q$. Write

$$\begin{aligned} \mathbb{P}_n \ell_{\boldsymbol{\beta}}^*(\widehat{\boldsymbol{\tau}}) &= \mathbb{P}_n \{ \Delta - Q(\zeta_0) \} l(\widehat{\zeta}) \mathbf{D} - \mathbb{P}_n (\widehat{\zeta} - \zeta_0) q(\zeta_0) l(\zeta_0) \mathbf{D} \\ &\quad - \mathbb{P}_n \left\{ Q(\widehat{\zeta}) - Q(\zeta_0) - (\widehat{\zeta} - \zeta_0) q(\zeta_0) \right\} l(\zeta_0) \mathbf{D} \end{aligned}$$

$$\begin{aligned}
 & -\mathbb{P}_n \left\{ Q(\widehat{\zeta}) - Q(\zeta_0) \right\} \left\{ l(\widehat{\zeta}) - l(\zeta_0) \right\} \mathbf{D} \\
 \equiv & A_1 - A_2 - A_3 - A_4.
 \end{aligned}$$

Because the second derivative of $Q(\cdot)$ is assumed to be bounded, $l(\cdot)$ is Lipschitz. Further, since both $\Delta - Q(\zeta_0)$ and D_k are fixed bounded functions, it concludes from Theorem 2.4 of van de Geer [29] and Theorem 9.23 of Kosorok [9] that the entropy for uniform norm of the class of uniformly bounded functions $\{\Delta - Q(\zeta_0)\} \{l(\zeta) - l(\zeta_0)\} D_k$ with $\|\zeta - \zeta_0\|_2 \leq M$, $\|\zeta\|_\infty \leq M$, and $\mathcal{J}(\zeta) \leq M$ is of order $(1/\varepsilon)^{1/r}$, and hence the class is a Donsker class. Since $l(\cdot)$ is Lipschitz, we have

$$\left\| \{\Delta - Q(\zeta_0)\} \{l(\widehat{\zeta}) - l(\zeta_0)\} D_k \right\|_2 = O_p(1) \left\| l(\widehat{\zeta}) - l(\zeta_0) \right\|_2 = o_p(1).$$

Further, according to Theorem 1, $\{\Delta - Q(\zeta_0)\} \{l(\widehat{\zeta}) - l(\zeta_0)\} D_k$ is within the above class. Thus, it concludes from Lemma 19.24 of van der Vaart [28] that

$$(\mathbb{P}_n - P) \{\Delta - Q(\zeta_0)\} \{l(\widehat{\zeta}) - l(\zeta_0)\} D_k = o_p(n^{-1/2}),$$

and hence

$$A_1 = \mathbb{P}_n \{\Delta - Q(\zeta_0)\} l(\zeta_0) \mathbf{D} + o_p(n^{-1/2}) = \mathbb{P}_n \ell_{\beta}^*(\tau_0) + o_p(n^{-1/2}). \tag{27}$$

Here we use the fact that $E \{\Delta - Q(\zeta_0)\} \{l(\widehat{\zeta}) - l(\zeta_0)\} D_k = 0$. Similarly, by Theorem 2.4 of van der Geer [29], together with the fact that $q(\zeta_0)l(\zeta_0)\mathbf{D}$ is bounded, the class of uniformly bounded functions $(\zeta - \zeta_0)q(\zeta_0)l(\zeta_0)\mathbf{D}$ with $\|\zeta - \zeta_0\|_2 \leq M$, $\|\zeta\|_\infty \leq M$, and $\mathcal{J}(\zeta) \leq M$ is a Donsker class. It concludes from Theorem 1 and Lemma 19.24 of van der Vaart [28] that

$$\mathbb{P}_n(\widehat{\zeta} - \zeta_0)q(\zeta_0)l(\zeta_0)\mathbf{D} = E(\widehat{\zeta} - \zeta_0)q(\zeta_0)l(\zeta_0)\mathbf{D} + o_p(n^{-1/2}).$$

In view of the definition of $(\mathbf{h}_0^*, \dots, \mathbf{h}_J^*)$, we have $E \ell_{\beta}^*(\tau_0) \ell_{\varphi_j}(\tau_0) [h_j(W_j)] = 0$ for any $h_j(W_j)$, $j = 0, \dots, J$. Thus, $E(\widehat{\varphi}_j - \varphi_{0,j})(W_j)q(\zeta_0)l(\zeta_0)\mathbf{D} = 0$. Therefore,

$$E(\widehat{\zeta} - \zeta_0)q(\zeta_0)l(\zeta_0)\mathbf{D} = E[\{\Delta - Q(\zeta_0)\} l(\zeta_0)\mathbf{D}]^{\otimes 2} (\widehat{\beta} - \beta_0).$$

It follows that

$$A_2 = \mathcal{J}(\beta_0)(\widehat{\beta} - \beta_0) + o_p(n^{-1/2}). \tag{28}$$

As shown above, the class of uniformly bounded functions $\zeta - \zeta_0$ with $\|\zeta\|_\infty \leq M$ and $\mathcal{J}(\zeta) \leq M$ is a Donsker class. It concludes from Theorem 1 and Lemma 19.24 of van der Vaart [28] that $\|\widehat{\zeta} - \zeta_0\|_n = \|\widehat{\zeta} - \zeta_0\|_2 + o_p(n^{-1/2})$, which implies that $\|\cdot\|_2$ and $\|\cdot\|_n$ share the same rate of convergence. Then, under condition 5, a Taylor expansion yields

$$|A_3| \leq O_p(1) \|\widehat{\zeta} - \zeta_0\|_n^2 = O_p(1) \|\widehat{\zeta} - \zeta_0\|_2^2 + o_p(n^{-1/2}) = o_p(n^{-1/2}). \tag{29}$$

Further, since both $Q(\cdot)$ and $l(\cdot)$ are Lipschitz, it follows from Theorem 1 that

$$|A_4| \leq O_p(1) \|\widehat{\zeta} - \zeta_0\|_n^2 = o_p(n^{-1/2}). \tag{30}$$

Therefore, combing the stationary equation (26) with (27)–(30) yields

$$n^{1/2} \mathcal{J}(\beta_0)(\widehat{\beta} - \beta_0) = n^{1/2} \mathbb{P}_n \ell_{\beta}^*(\tau_0) + o_p(1).$$

The asymptotic normality of $\widehat{\beta}$ follows immediately from the standard central limit theorem.

A.4. Proof of Theorem 3

Define $h_{+,l}^* = \sum_{j=0}^J h_{j,l}^*$ and $\widehat{h}_{+,l}^* = \sum_{j=0}^J \widehat{h}_{j,l}^*$. In addition, let $\dot{\ell}_{\varphi}(\tau, h_+) = \sum_{j=0}^J \dot{\ell}_{\varphi_j}(\tau)[h_j]$ for $h_{+,l} = \sum_{j=0}^J h_{j,l}$, $\kappa_l(\tau, h_{+,l}) = \left\{ \dot{\ell}_{\beta,l}(\tau) - \dot{\ell}_{\varphi}(\tau)[h_{+,l}] \right\}^2$. Because $h_{+,l}^*(\cdot)$ is bounded and smooth, according to standard spline approximation result, there exists $h_{n+,l}^* \in \mathcal{S}_+$ such that $\|h_{+,l}^* - h_{n+,l}^*\|_{\infty} = o(1)$. As shown in the proof of Theorem 2, the class of uniformly bounded functions ζ with $\|\tau - \tau_0\|_2 \leq M$, $J(\zeta) \leq M$, and $\|\zeta\|_{\infty} \leq M$ for $M > 0$ is Donsker. Further, $\kappa_l(\tau, h_{+,l})$ is uniformly bounded and Lipschitz with respect to ζ , and hence, in view of Donsker preservation theorem (e.g. Corollary 9.32 of Kosorok [9]), for any $h_{+,l} \in \mathcal{H}_+ \equiv \mathcal{H} + \dots + \mathcal{H}$, the class

$$\mathcal{F}_2 = \{ \kappa_l(\tau, h_{+,l}) : \|\tau - \tau_0\|_2 \leq M, J(\zeta) \leq M, \|\zeta\|_{\infty} \leq M \}$$

is also Donsker. In view of Theorem 1 and the Glivenko-Cantelli theorem, we have

$$(\mathbb{P}_n - P)\kappa_l(\widehat{\tau}, h_{+,l}) = o_p(1) \quad \text{and} \quad (\mathbb{P}_n - P)\kappa_l(\widehat{\tau}, h_{+,l}^*) = o_p(1). \quad (31)$$

Further, it follows from the continuous mapping theorem and dominated convergence theorem, together with the consistency of $\widehat{\tau}$ that

$$\begin{aligned} P \{ \kappa_l(\widehat{\tau}, h_{+,l}) - \kappa_l(\tau_0, h_+) \} &= o_p(1), \\ P \{ \kappa_l(\widehat{\tau}, h_{+,l}^*) - \kappa_l(\tau_0, h_{+,l}^*) \} &= o_p(1). \end{aligned} \quad (32)$$

Combining (31) and (32) yields

$$\mathbb{P}_n \{ \kappa_l(\widehat{\tau}, h_{+,l}) - \kappa_l(\widehat{\tau}, h_{+,l}^*) \} = P \{ \kappa_l(\tau_0, h_+) - \kappa_l(\tau_0, h_{+,l}^*) \} + o_p(1).$$

By the characteristic of the least favorable direction $h_{+,l}^*$, for any $\varepsilon > 0$ and any $h_{+,l} \in \mathcal{H}_+$ with $|h_{+,l} - h_{+,l}^*| \geq \varepsilon$, $P \{ \kappa_l(\tau_0, h_{+,l}) - \kappa_l(\tau_0, h_{+,l}^*) \} > 0$, and hence we have

$$\mathbb{P}_n \{ \kappa_l(\widehat{\tau}, h_{+,l}) - \kappa_l(\widehat{\tau}, h_{+,l}^*) \} > 0 \quad (33)$$

with probability converging to 1. As shown in the proof of Theorem 2, the class of uniformly bounded functions $\{ \Delta - Q(\zeta) \} l(\zeta)$ for $\|\tau - \tau_0\| \leq M$, $\|\zeta\|_{\infty} \leq M$, and $J(\zeta) \leq M$ is a Donsker class. Further, in view of bracketing entropy calculation of spline in Shen and Wong [22], the class of uniformly bounded functions $z_l - h_{+,l}$ for $h_{+,l} \in \mathcal{S}_+$, $\|h_{+,l} - h_{+,l}^*\|_{\infty} \leq M$ is also a Donsker class. The Donsker

preservation theorem (e.g., Corollary 9.32 of Kosorok [9]) applies and yields that the class of uniformly bounded functions $\kappa_l(\boldsymbol{\tau}, h_{+,l})$ with $\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\|_2 \leq M$, $J(\zeta) \leq M$, $\|\zeta\|_\infty \leq M$, $h_{+,l} \in \mathcal{S}_+$, and $\|h_{+,l} - h_{+,l}^*\|_2 \leq M$ is a Donsker class, hence a Glivenko-Cantelli class. Thus, it concludes from the continuous mapping theorem and dominated convergence theorem, together with the consistency of $\widehat{\boldsymbol{\tau}}$ and $\|h_{n+,l}^* - h_{+,l}^*\|_\infty = o(1)$ that $\mathbb{P}_n \left\{ \kappa_l(\widehat{\boldsymbol{\tau}}, h_{n+,l}^*) - \kappa_l(\widehat{\boldsymbol{\tau}}, h_{+,l}^*) \right\} = o_p(1)$, and hence

$$\mathbb{P}_n \left\{ \kappa_l(\widehat{\boldsymbol{\tau}}, \widehat{h}_{+,l}^*) - \kappa_l(\widehat{\boldsymbol{\tau}}, h_{+,l}^*) \right\} = \mathbb{P}_n \left\{ \kappa_l(\widehat{\boldsymbol{\tau}}, \widehat{h}_{+,l}^*) - \kappa_l(\widehat{\boldsymbol{\tau}}, h_{n+,l}^*) \right\} + o_p(1). \quad (34)$$

Thus, by definition of $\widehat{h}_{+,l}^*$, $|\widehat{h}_{+,l}^* - h_{+,l}^*| \geq \varepsilon$ implies that $\mathbb{P}_n \left\{ \kappa_l(\widehat{\boldsymbol{\tau}}, \widehat{h}_{+,l}^*) - \kappa_l(\widehat{\boldsymbol{\tau}}, h_{+,l}^*) \right\} \leq 0$ with probability converging to 1 from (34). Thus, it concludes from (33) that

$$P(|\widehat{h}_{+,l}^* - h_{+,l}^*| \geq \varepsilon) \leq P \left[\mathbb{P}_n \left\{ \kappa_l(\widehat{\boldsymbol{\tau}}, \widehat{h}_{+,l}^*) - \kappa_l(\widehat{\boldsymbol{\tau}}, h_{+,l}^*) \right\} \leq 0 \right] \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

The consistency of $\widehat{h}_{+,l}^*$ follows, and hence $|\widehat{h}_{+,l}^*| \leq |\widehat{h}_{+,l}^* - h_{+,l}^*| + |h_{+,l}^*| \leq o_p(1) + O(1) = O_p(1)$, i.e., $\|\widehat{h}_{+,l}^*\|_\infty = O_p(1)$. Since Theorem 2.4 of van de Geer [29] is also valid for high dimensions (see van de Geer [29], p. 187 for further detail), the class of uniformly bounded functions $\kappa_l(\boldsymbol{\tau}, h_{+,l})$ with $\|\boldsymbol{\tau} - \boldsymbol{\tau}_0\|_2 \leq M$, $J(\zeta) \leq M$, $\|\zeta\|_\infty \leq M$, $\|h_{+,l} - h_{+,l}^*\|_2 \leq M$, $J(h) \leq M$, and $\|h_{+,l}\|_\infty \leq M$ is Donsker. It concludes from the continuous-mapping theorem and dominated convergence theorem, together with the consistency of $\widehat{\boldsymbol{\tau}}$ and $\widehat{\mathbf{h}}^*$ yield $\widehat{\mathcal{F}}_n = \mathcal{F}(\boldsymbol{\beta}_0) + o_p(1)$. The consistency of $\widehat{\mathcal{F}}_{n,\lambda}$ follows immediately from the assumption of the order of $\boldsymbol{\lambda}$.

Appendix B: Results of simulation study under scenario S1 and S3

In this section, figures 4, 5, 6 and 7 summarize the estimates of the unknown functions under scenario S1 and S3 for $n = 200$ and 400. In particular, each figure presents the 0.025, 0.5, and 0.975 point-wise quantiles of the estimates for all considered values of α (i.e., 0, 0.5, 1.0) obtained under our proposed approach and Cheng and Wang's approach [2].

Appendix C: Results of simulation study under partially linear Cox PH model

In this simulation study, we compare the finite sample performance of our proposed penalized approach and the approach of Lu and McMahan [13]. Note, Lu and McMahan's method was developed for a partially linear Cox PH model, and hence the comparison is made under the following model specifying the true failure time

$$F(t|\mathbf{X}) = 1 - \exp[-\exp\{Z_1\beta_1 + Z_2\beta_2 + \eta(t) + \varphi(W)\}], \quad (35)$$

where $Z_1 \sim \text{Bernoulli}(0.5)$, $Z_2 \sim \mathcal{N}(0, 1)$, and $W \sim \mathcal{U}[-1, 1]$. We consider three

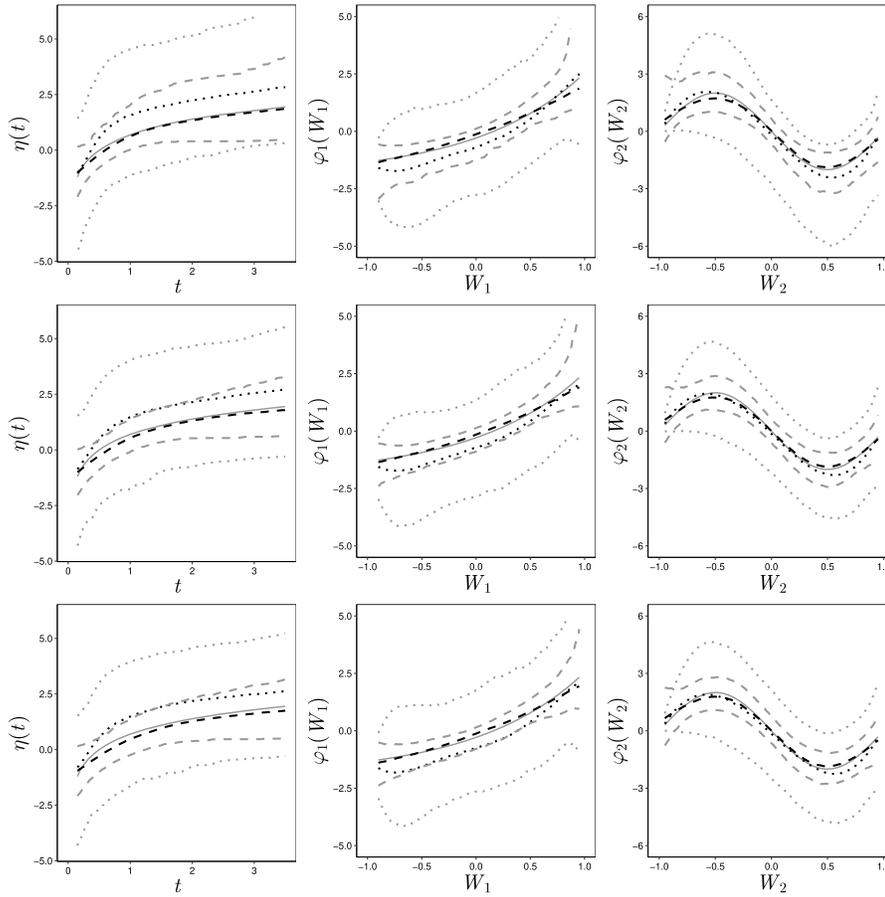


FIG 4. Estimates of the unknown functions under scenario S1 when $n = 200$. From top to bottom, the panels correspond to $\alpha = 0, 0.5, 1.0$, respectively. The solid grey curves correspond to the true value of the functions. Dashed lines are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the penalized spline approach. Dotted curves are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the approach of Cheng and Wang [2].

scenarios. In particular, in Scenario 1 (S1) the two functions were specified as $\eta(t) = \log(2t)$, $\varphi(w) = 2w$, and the regression coefficients were $\beta_1 = 0.5$ and $\beta_2 = -0.5$; in Scenario 2 (S2) the two functions were $\eta(t) = \log\{1.5t - \log(1 + 1.5t)\}$, $\varphi(w) = 2 \sin(-\pi w)$, and the regression coefficients were $\beta_1 = 0.5$ and $\beta_2 = 0.5$; and in Scenario 3 (S3) the two functions were $\eta(t) = \log\{\log(1+t/10) + \sqrt{t}/10\}$, $\varphi(w) = 4w^2 - 4/3$, and the regression coefficients were $\beta_1 = -0.5$ and $\beta_2 = -0.5$. To create current status data, observation times (Y) were simulated from an exponential distribution with a mean of 2, 2, and 1 under S1, S2, and S3, respectively, and the censoring indicator was determined as $\Delta = I(T < Y)$.

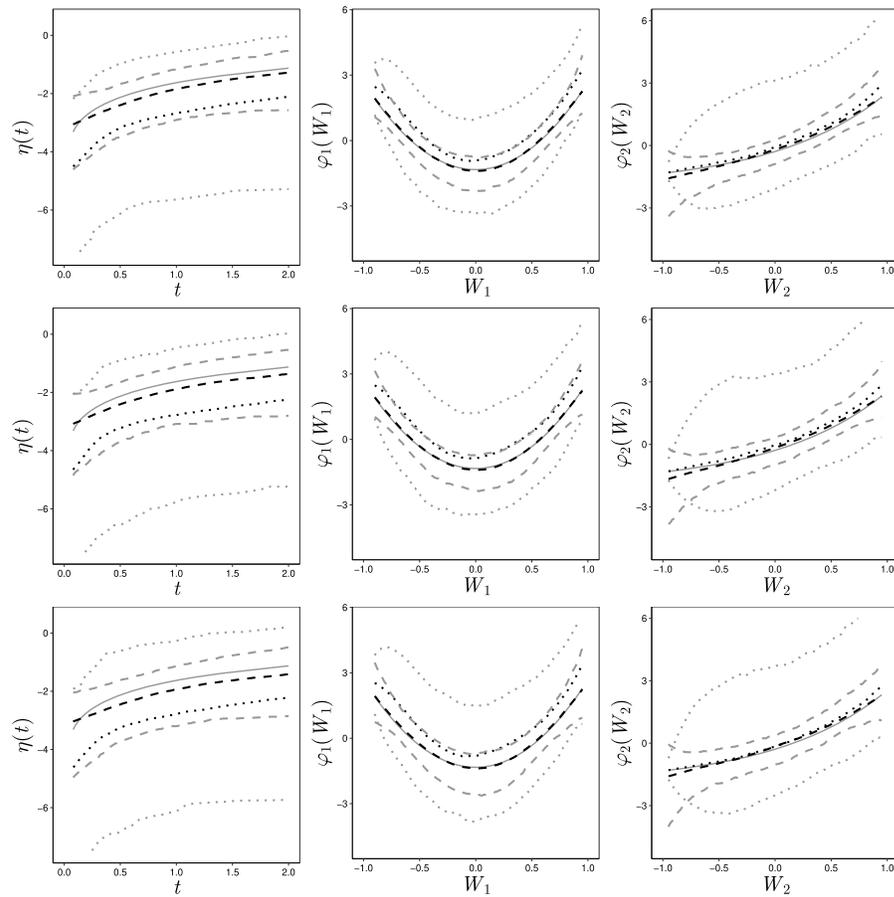


FIG 5. Estimates of the unknown functions under scenario S3 when $n = 200$. From top to bottom, the panels correspond to $\alpha = 0, 0.5, 1.0$, respectively. The solid grey curves correspond to the true value of the functions. Dashed lines are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the penalized spline approach. Dotted curves are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the approach of Cheng and Wang [2].

These specifications yield a variety of right-censoring rates; for example, under S1, S2, and S3 the right censoring rates are 22%, 43%, and 80%, respectively. For each simulation configuration, this data generating process was repeated 1000 times to create 1000 independent data sets consisting of $n = 200$ and $n = 400$ observations. We use $\lceil n \rceil^{1/3}$ inner knots for our approach. In contrast, for the competing approach, estimators were computed using multiple knot set configurations with the number of interior knots of each function ranging from $\lceil n^{1/3} \rceil - 4$ to $\lceil n^{1/3} \rceil + 4$, then the final unpenalized estimator is selected based on the BIC criterion. Table 4 summarizes the estimation of the regression coef-

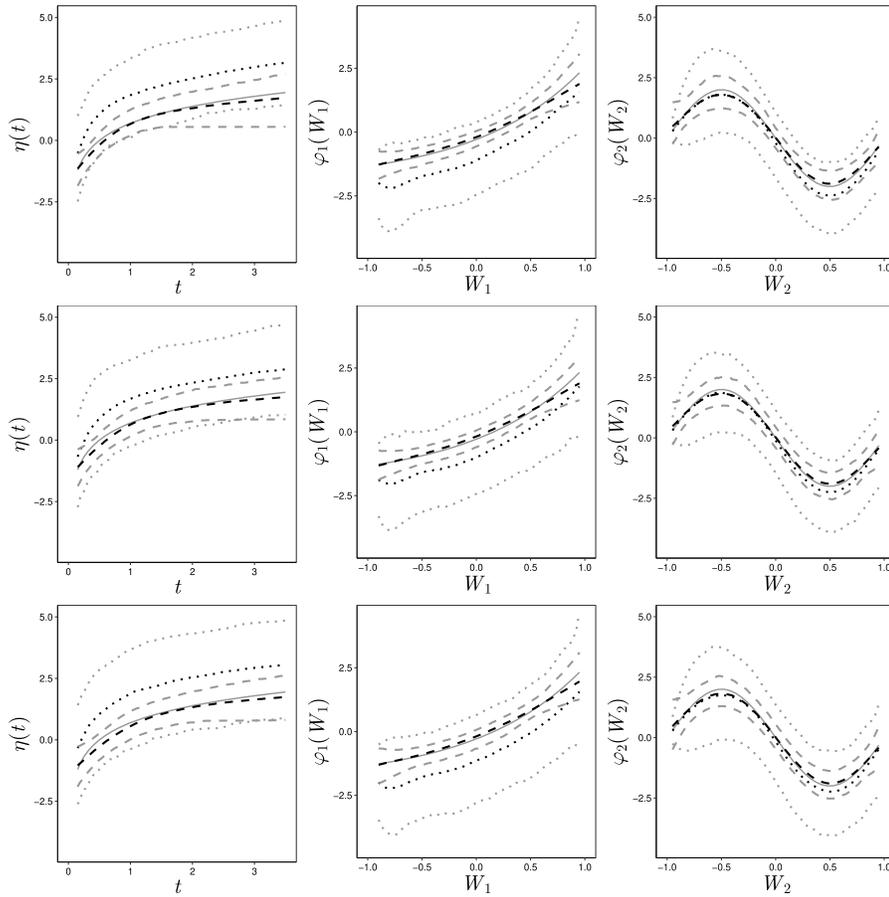


FIG 6. Estimates of the unknown functions under scenario S1 when $n = 400$. From top to bottom, the panels correspond to $\alpha = 0, 0.5, 1.0$, respectively. The solid grey curves correspond to the true value of the functions. Dashed lines are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the penalized spline approach. Dotted curves are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the approach of Cheng and Wang [2].

ficients for all simulation configurations, while Figures 8 and 9 summarize the estimates of the two unknown functions when $n = 200$ and $n = 400$, respectively.

Appendix D: Additional results for HPV data from two competing approaches

Figure 10 presents the histogram of age at first sexual intercourse. Figure 11 summarizes the estimates of $\eta(\cdot)$ and $\varphi(\cdot)$ (i.e., nonlinear function of age at first sexual intercourse) under both PH (first row) and PO (second row) models,

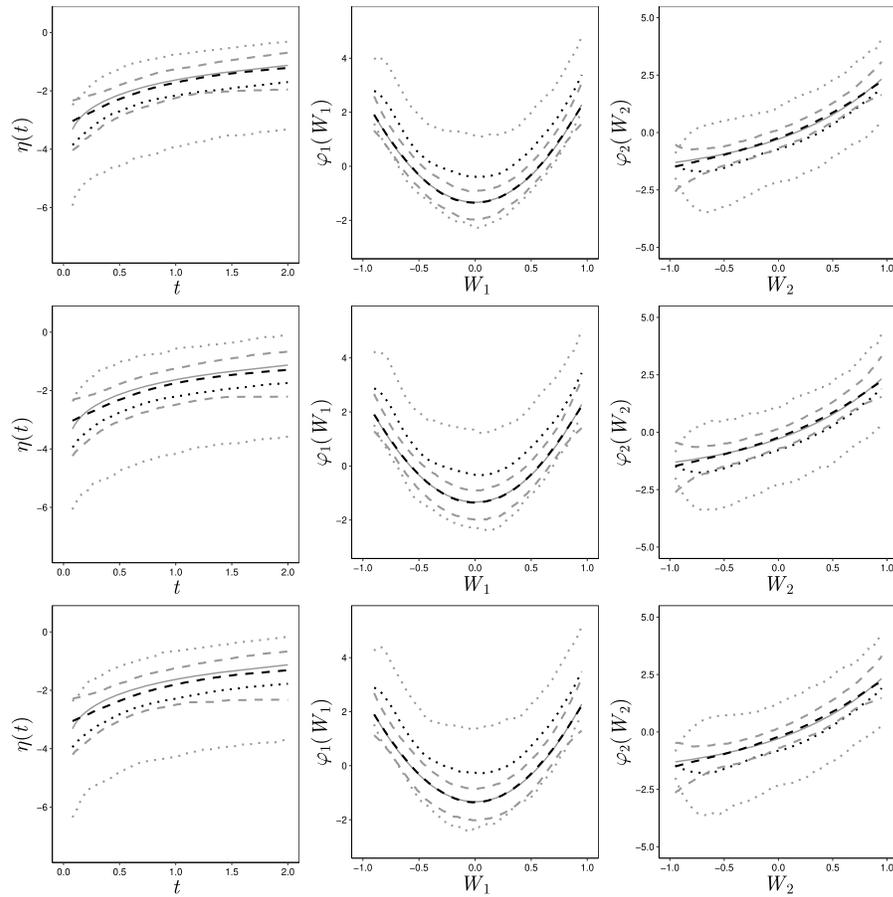


FIG 7. Estimates of the unknown functions under scenario S3 when $n = 400$. From top to bottom, the panels correspond to $\alpha = 0, 0.5, 1.0$, respectively. The solid grey curves correspond to the true value of the functions. Dashed lines are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the penalized spline approach. Dotted curves are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the approach of Cheng and Wang [2].

along with 95% point-wise confidence intervals, using the approach by Cheng and Wang [2]. Figure 12 presents the estimates of $\eta(\cdot)$ and $\varphi(\cdot)$ under PH model using the approach by Lu and McMahan [13]. Figure 13 summarizes the estimates of $\eta(\cdot)$, $\varphi_1(\cdot)$ and $\varphi_2(\cdot)$.

References

- [1] BANERJEE, M. (2012). *Current status data in the twenty-first century: Some interesting developments*. CRC Press Taylor & Francis Group.

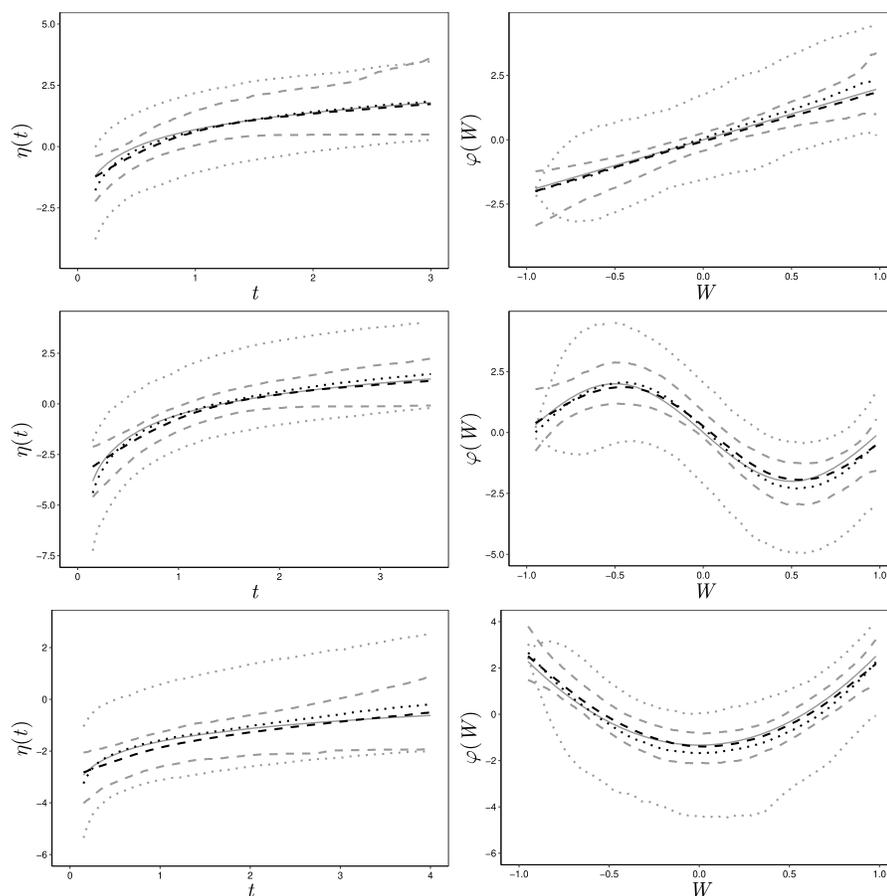


FIG 8. Estimates of the unknown functions for sample size $n = 200$. From top to bottom, the panels correspond to scenarios $S1$, $S2$ and $S3$, respectively. The solid grey curves correspond to the true value of the functions. Dashed lines are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the penalized spline approach. Dotted curves are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the approach of Lu and McMahan [13]

[MR3014204](#)

- [2] CHENG, G. and WANG, X. (2011). Semiparametric additive transformation model under current status data. *Electronic Journal of Statistics* **5** 1735–1764. [MR2870149](#)
- [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] COX, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society* **34** 187–220. [MR0341758](#)

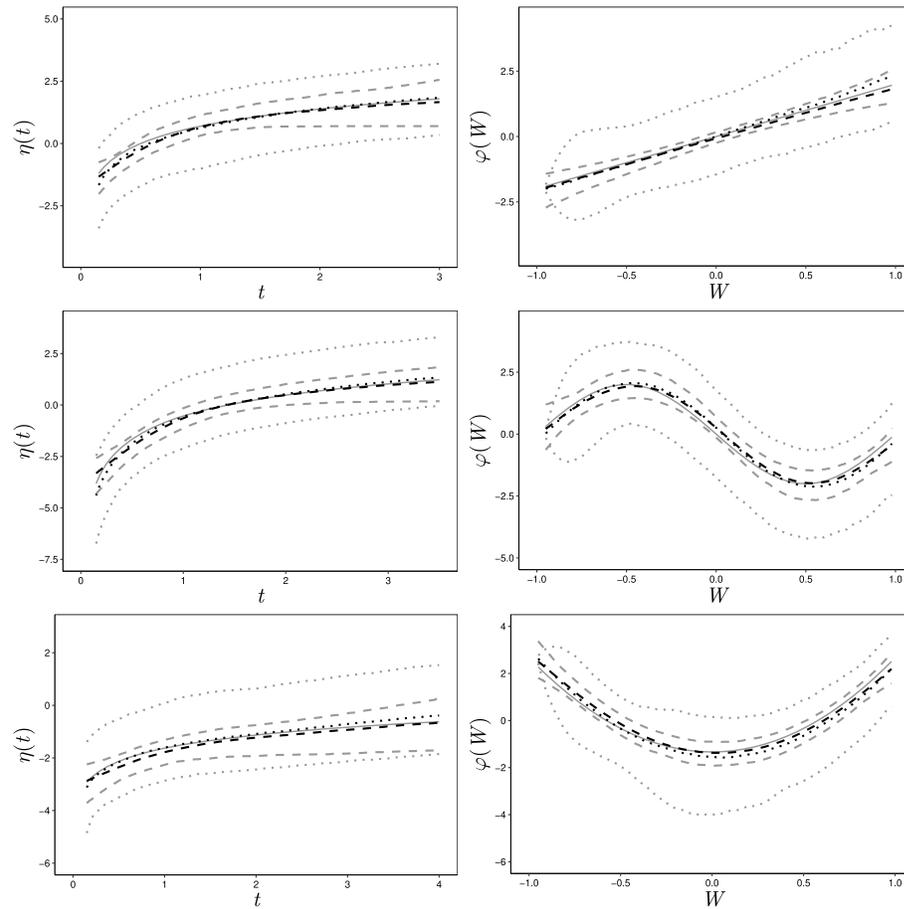


FIG 9. Estimates of the unknown functions for sample size $n = 400$. From top to bottom, the panels correspond to scenarios $S1$, $S2$ and $S3$, respectively. The solid grey curves correspond to the true value of the functions. Dashed lines are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the penalized spline approach. Dotted curves are the medians (black), 0.025 and 0.975 quantiles (grey) of the estimates obtained from the approach of Lu and McMahan [13]

- [5] DABROWSKA, D. M. and DOKSUM, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association* **83** 744–749. [MR0963802](#)
- [6] DU, M., DU, T. and SUN, J. (2019). Semiparametric probit model for informative current status data. *Statistics in Medicine* **38** 219–2227. [MR3939813](#)
- [7] GU, M., SUN, L. and ZUO, G. (2005). A baseline-free procedure for transformation models under interval censorship. *Lifetime Data Analysis* **11** 473–488. [MR2213500](#)

TABLE 4

Summary of the results that were obtained using the competing technique (Lu and McMahan[13]) and the proposed penalized spline estimator under partially linear Cox PH model. Presented results include the empirical bias (Bias), sample standard deviation (SD), average of the 1000 estimated standard errors (SE), mean squared error (MSE) of the 1000 point estimates, as well as empirical coverage probability associated with 95% confidence intervals (CP) and average model fitting time (TIME; measured in Seconds).

		Lu and McMahan						Penalized Spline Estimator					
		Bias	SD	SE	MSE	CP	TIME	Bias	SD	SE	MSE	CP	TIME
S1($n = 200$)	β_1	0.051	0.331	0.341	0.112	95.8	475	0.017	0.319	0.296	0.102	95.1	7
	β_2	-0.054	0.176	0.190	0.034	97.0		-0.015	0.177	0.163	0.031	94.0	
S2($n = 200$)	β_1	0.048	0.325	0.336	0.107	96.2	335	0.002	0.305	0.306	0.093	95.6	10
	β_2	0.040	0.178	0.184	0.033	95.5		-0.003	0.171	0.163	0.029	93.4	
S3($n = 200$)	β_1	0.004	0.400	0.397	0.160	93.3	153	0.048	0.373	0.367	0.141	94.6	4
	β_2	-0.056	0.224	0.212	0.053	93.6		-0.017	0.207	0.191	0.043	94.0	
S1($n = 400$)	β_1	0.031	0.221	0.227	0.050	94.8	783	-0.006	0.209	0.202	0.044	94.0	9
	β_2	-0.032	0.123	0.126	0.016	95.7		0.008	0.119	0.110	0.014	93.8	
S2($n = 400$)	β_1	0.031	0.229	0.226	0.053	95.0	382	0.002	0.220	0.211	0.048	94.5	14
	β_2	0.025	0.121	0.124	0.015	95.5		-0.005	0.118	0.113	0.014	93.9	
S3($n = 400$)	β_1	-0.029	0.268	0.265	0.072	93.3	71	0.005	0.252	0.254	0.063	94.4	5
	β_2	-0.037	0.145	0.141	0.022	94.1		-0.009	0.134	0.132	0.018	94.8	

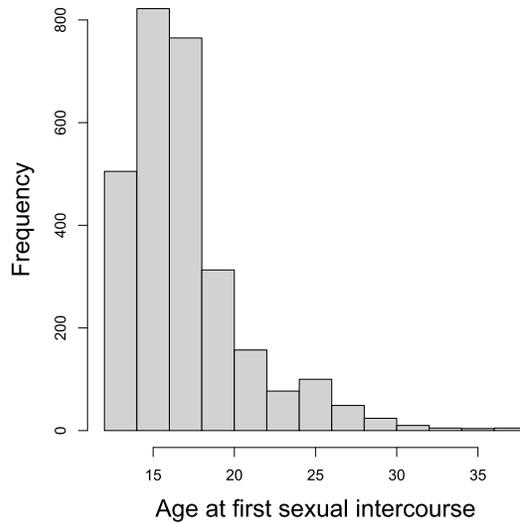


FIG 10. Histogram of age at first sexual intercourse.

[8] HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized additive models* **43**. CRC press. [MR1082147](#)
 [9] KOSOROK, M. R. (2008). *Introduction to empirical processes and semi-*

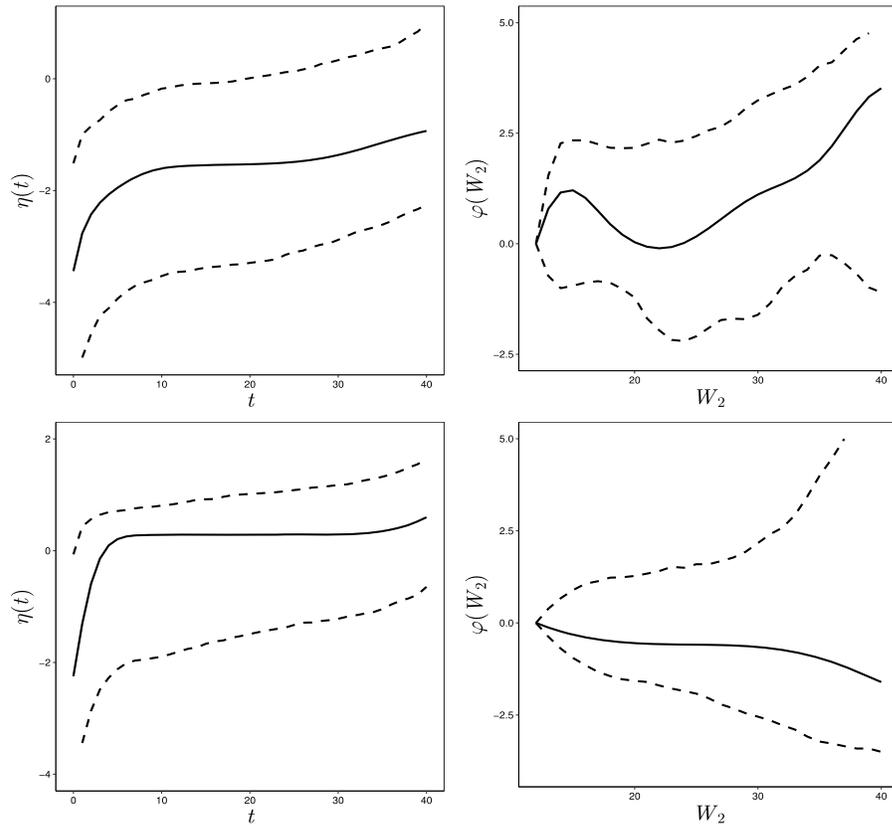


FIG 11. Functional estimates obtained from the approach by Cheng and Wang [2] for $\eta(\cdot)$ and $\varphi(\cdot)$. First and second row correspond to the PH and PO model, respectively. The solid black curve corresponds to the estimated function and the dashed curves correspond to the 0.025 and 0.975 quantiles of 1000 estimated functions based on bootstrap samples.

parametric inference. Springer. [MR2724368](#)

- [10] LI, S., DU, T. and SUN, J. (2020). Regression analysis of misclassified current status data. *Journal of Nonparametric Statistics* **32** 1–19. [MR4077750](#)
- [11] LU, M. and LI, C. S. (2017). Penalized estimation for proportional hazards models with current status data. *Statistics in Medicine* **36** 4893–4907. [MR3734481](#)
- [12] LU, M., LIU, Y. and LI, C. S. (2020). Efficient estimation of a linear transformation model for current status data via penalized splines. *Statistical methods in medical research* **29** 3–14. [MR4055118](#)
- [13] LU, M. and McMAHAN, C. S. (2018). A partially linear proportional hazards model for current status data. *Biometrics* **74** 1240–1249. [MR3908142](#)
- [14] LU, M., ZHANG, Y. and HUANG, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*

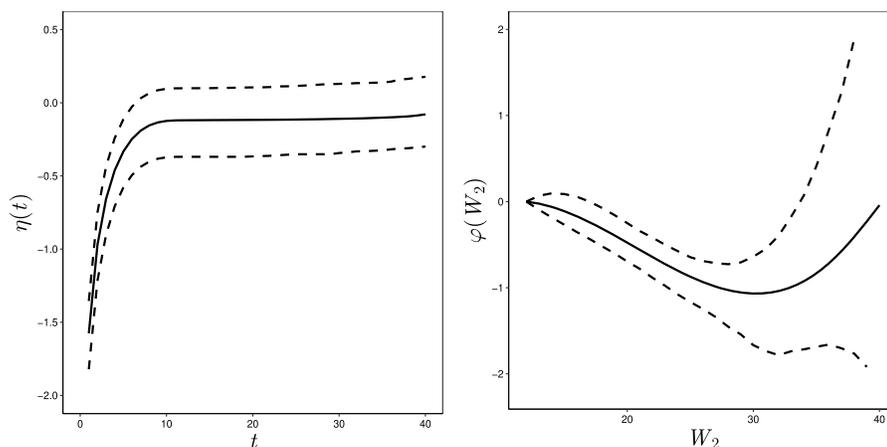


FIG 12. Functional estimates obtained from the approach by Lu and McMahan [13] for $\eta(\cdot)$ and $\varphi(\cdot)$. First and second row correspond to the PH and PO model, respectively. The solid black curve corresponds to the estimated function and the dashed curves correspond to the 0.025 and 0.975 quantiles of 1000 estimated functions based on bootstrap samples.

94 705–718. [MR2410018](#)

- [15] LU, M., ZHANG, Y. and HUANG, J. (2009). Semiparametric estimation methods for panel count data using monotone B-splines. *Journal of the American Statistical Association* **104** 1060–1070. [MR2750237](#)
- [16] MA, S. (2009). Cure model with current status data. *Statistica Sinica* **19** 233–249. [MR2487887](#)
- [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 DER GEER, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *The Annals of Statistics* **25** 1014–1035. [MR1447739](#)
- [19] MCMAHAN, C. S., WANG, L. and TEBBS, J. M. (2013). Regression analysis for current status data using the EM algorithm. *Statistics in Medicine* **32** 4452–4466. [MR3118368](#)
- [20] MURPHY, S. A. and VAN DER VAART, A. W. (1999). Observed information in semi-parametric models. *Bernoulli* **5** 381–412. [MR1693616](#)
- [21] SCHUMAKER, L. (1981). *Spline functions: basic theory*. Cambridge University Press. [MR2348176](#)
- [22] SHEN, X. and WONG, W. H. (1994). Convergence rate of sieve estimates. *The Annals of Statistics* 580–615. [MR1292531](#)
- [23] STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The annals of statistics* 1040–1053. [MR0673642](#)
- [24] STONE, C. J. (1985). Additive regression and other nonparametric models. *The annals of Statistics* **13** 689–705. [MR0790566](#)

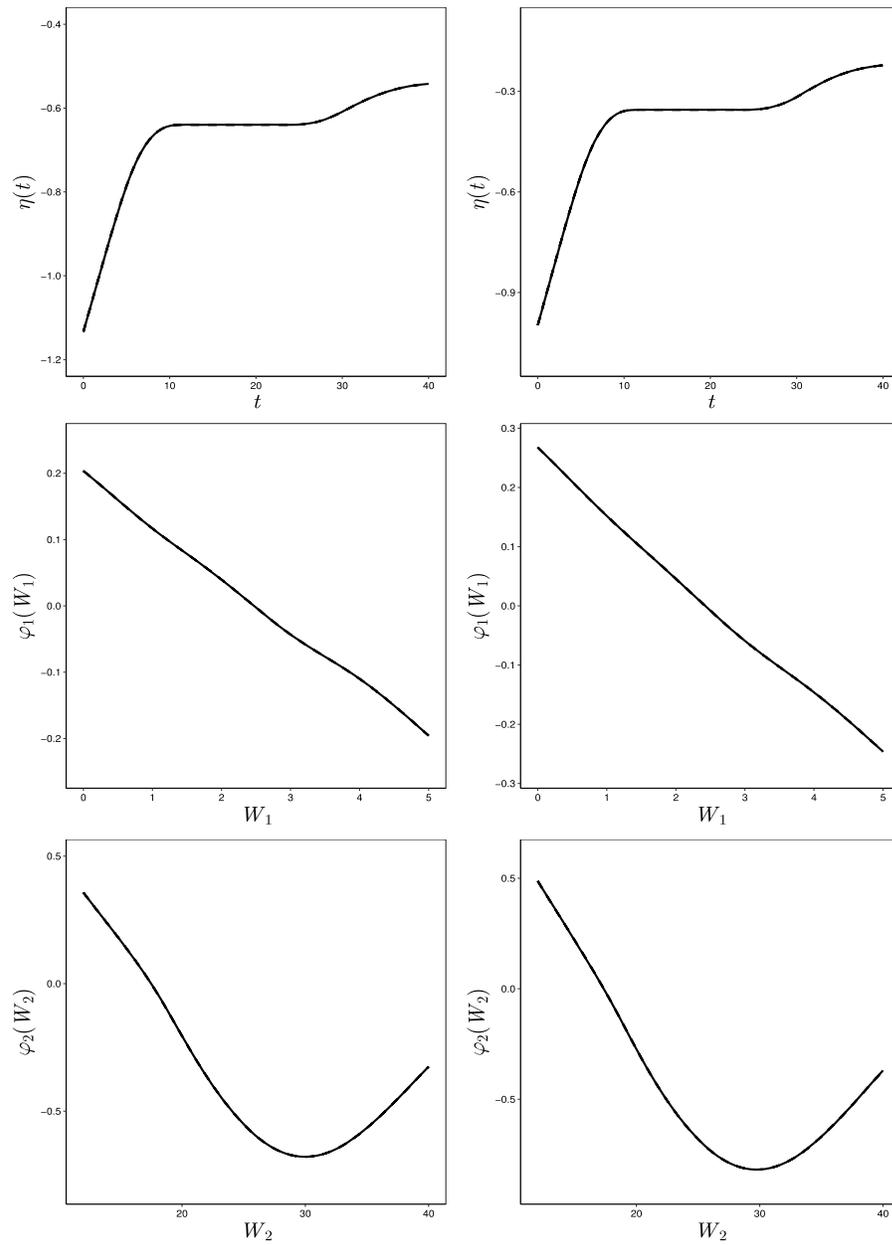


FIG 13. Functional estimates obtained from the proposed penalized approach for $\eta(\cdot)$, $\varphi_1(\cdot)$ and $\varphi_2(\cdot)$. Left and right row correspond to the PH and PO model, respectively. The solid black curve corresponds to the estimated function.

- [25] STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *The Annals of Statistics* 118–171. [MR1272079](#)
- [26] SUN, J. (2007). *The statistical analysis of interval-censored failure time data*. Springer Science & Business Media. [MR2287318](#)
- [27] SUN, J. and SUN, L. (2005). Semiparametric linear transformation models for current status data. *Canadian Journal of Statistics* **33** 85–96. [MR2155000](#)
- [28] VAN DE GEER, S. A. (2000). *Empirical Processes in M-estimation* **6**. Cambridge university press. [MR0937256](#)
- [29] VAN DER VAART, A. W. (2000). *Asymptotic statistics* **3**. Cambridge university press. [MR1652247](#)
- [30] WOOD, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC. [MR2206355](#)
- [31] 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](#)
- [32] WOOD, S. N., PYA, N. and SÄFKEN, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* **111** 1548–1563. [MR3601714](#)
- [33] XIANG, D. and WAHBA, G. (1997). Approximate smoothing spline methods for large data sets in the binary case. In *Proceedings of the American Statistical Association Joint Statistical Meetings, Biometrics Section* 94–99.
- [34] ZENG, D. and LIN, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 507–564. [MR2370068](#)
- [35] ZENG, D., MAO, L. and LIN, D. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika* **103** 253–271. [MR3509885](#)
- [36] 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](#)
- [37] ZHANG, Z., SUN, L., ZHAO, X. and SUN, J. (2005). Regression analysis of interval-censored failure time data with linear transformation models. *Canadian Journal of Statistics* **33** 61–70. [MR2287318](#)
- [38] ZHANG, Z. and ZHAO, Y. (2013). Empirical likelihood for linear transformation models with interval-censored failure time data. *Journal of Multivariate Analysis* **116** 398–409. [MR3049912](#)
- [39] ZHOU, J., ZHANG, J. and LU, W. (2017). An Expectation Maximization algorithm for fitting the generalized odds-rate model to interval censored data. *Statistics in medicine* **36** 1157–1171. [MR3621015](#)