Monotone spline-based least squares estimation for panel count data with informative observation times

Shirong Deng¹, Li Liu¹, and Xingqiu Zhao*²

¹ School of Mathematics and Statistics, Wuhan University, Wuhan, China
² Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China

Received 2 October 2014; revised 14 March 2015; accepted 21 April 2015

This article discusses the statistical analysis of panel count data when the underlying recurrent event process and observation process may be correlated. For the recurrent event process, we propose a new class of semiparametric mean models that allows for the interaction between the observation history and covariates. For inference on the model parameters, a monotone spline-based least squares estimation approach is developed, and the resulting estimators are consistent and asymptotically normal. In particular, our new approach does not rely on the model specification of the observation process. The proposed inference procedure performs well through simulation studies, and it is illustrated by the analysis of bladder tumor data.

Keywords: Informative observation process; Least squares estimation; Monotone B-splines; Panel count data; Semiparametric mean models.

1 Introduction

In many longitudinal follow-up studies, each subject may be observed at several distinct times and only the numbers of events between two adjacent times are available. It may be impossible to record the exact event times because of too expensive examination cost or too frequent occurrence of the events. Moreover, the observation times may vary from subject to subject. Such complex data are called panel count data, which often occur in many fields such as demographic studies, industrial reliability, and clinical trials; see Kalbfleisch and Lawless (1985), Gaver and O’Muircheartaigh (1987), Thall and Lachin (1988), and Sun and Kalbfleisch (1995). For such data, important information includes the observation times, counts of recurrent events, censoring or follow-up times, and covariates related to the study for each study subject. A typical example is bladder tumor data that will be discussed below.

For the analysis of panel count data, Sun and Zhao (2013) provide a relatively complete review for the analysis of panel count data wherein more references can be found. The existing research mainly focuses on the assumption that the underlying observation process and recurrent event process are independent completely or conditional on covariates (e.g., Kalbfleisch and Lawless, 1985; Thall and Lachin, 1988; Sun and Kalbfleisch, 1995; Staniswalis et al., 1997; Sun and Wei, 2000; Wellner and Zhang, 2000, 2007; Zhang, 2002; Hu et al., 2003, 2009; Lu et al., 2007, 2009; He et al., 2008). However, this assumption may be violated in practice. Such an example is a set of panel count data arising from a bladder cancer follow-up study conducted by the Veterans Administration Cooperative Urological Research Group (Byar, 1980). Many patients had multiple recurrences of new tumors during the study. The patients in the thiotepa group had significantly more clinical visits than those in the placebo group since thiotepa needed be instilled in the bladder (Lu et al., 2009; Li et al., 2010; Zhao and

*Corresponding author: e-mail: xingqiu.zhao@polyu.edu.hk, Phone: +852-2766-6921, Fax: +852-2362-9045

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
Tong, 2011). The patients in the thiotepa group received more medical treatment so that these patients had the lower tumor recurrence rate. This indicates that the number of clinical visits may contain some information about the tumor recurrence rate—that is, the observation process is informative.

In contrast, there is limited research on the analysis of panel count data with informative observation times. Li et al. (2010) incorporated the observation history to the mean models of the underlying recurrent event process that follows a class of semiparametric transformation models. Zhao and Tong (2011) proposed a joint modeling approach through an unobserved frailty variable and a completely unspecified link function to characterize the correlation between the underlying recurrent event process and the observation times. A commonly used assumption in their analysis mentioned above is that the observation process follows a Poisson-type with a proportional intensity function. However, the fit of the Poisson model may be inadequate when the observation process displays underdispersion or overdispersion. For example, Zhao et al. (2013) tested the Poisson process assumption about the observation process for the bladder tumor data and concluded that the Poisson process model cannot be acceptable for the observation process in the data. In addition, the relation between the observation and recurrent event processes may vary with some covariates. For example, in the bladder cancer study, patients who received the thiotepa treatment may make clinical visits more often than those in the placebo group and thus may have less superficial bladder tumors, which means that the correlation between the observation times and the tumor recurrent process as mentioned above may be different for different treatment groups. Motivated by the characteristics of the bladder cancer data, we propose a new class of semiparametric regression models by incorporating the interaction between the observation history and some covariates to the mean model of the recurrent event process, where the observation process model is completely unspecified.

The remainder of this article is organized as follows. We begin in Section 2 by introducing notation and describing models for panel count data. In Section 3, a spline-based least squares method is proposed for estimation of regression parameters and the baseline mean function, and the asymptotic properties of the proposed estimators are presented in Appendix. In Section 4, we present some simulation results to assess the finite-sample performance of the proposed inference procedure. In Section 5, the proposed approach is illustrated through the analysis of a dataset from a bladder tumor study. Some concluding remarks are made in Section 6.

## 2 Panel count mean models

Consider a study involving \( n \) independent subjects who may experience some recurrent events and let \( N_i(t) \) denote the total number of occurrences of the event of interest up to time \( t \) for subject \( i \), \( i = 1, \ldots, n \) for \( 0 \leq t \leq \tau \), where \( \tau \) is a known constant time point. Also suppose that for subject \( i \), \( N_i(t) \) is observed only at discrete potential time points \( 0 < T_{K_{i,1}} < T_{K_{i,2}} < \cdots < T_{K_{i,K_{i}}} \), where the total number of observations \( K_i \) is an integer-valued random variable. In general, not every subject can be followed until \( \tau \) and there exists a follow-up time \( C_i \) for subject \( i \). That is, \( N_i(T_{K_{i,j}}) \) is observed only if \( T_{K_{i,j}} \leq C_i \leq \tau \). Define \( H_i(t) = \sum_{j=1}^{K_i} I(T_{K_{i,j}} \leq t) \) to be the counting process that records the number of observations for subject \( i \) up to time \( t \), where \( I(\cdot) \) is the indicator function. Then define \( \tilde{H}_i(t) = H_i(t \wedge C_i) \) as the actual observation process for subject \( i \), where \( a \wedge b = \min(a, b) \). Let \( \mathbf{X}_i = (X_{i1}, \ldots, X_{in})' \) denote a \( p \)-dimensional vector of covariates that may not depend on \( t, i = 1, \ldots, n \). Define \( \mathcal{F}_i(t) = \{H_i(s) : 0 \leq s < t \} \) as the observation history just before \( t \).

By relaxing the model assumption on the observation process, we assume that given \( \mathbf{X}_i, \mathcal{F}_i(t) \), and the covariate \( W_i \), which is allowed to be a component of the vector \( \mathbf{X}_i \), the mean function of \( N_i(t) \) has the form

\[
\mu_i(t) = \mu_0(t) \exp\{\beta' \mathbf{X}_i + \alpha' \phi(\mathcal{F}_i(t), W_i)\},
\]

(1)
where \( \mu_0(t) \) is an unspecified smooth and nondecreasing nonnegative function of \( t \); \( \beta \) and \( \alpha \) are \( p \)-dimensional and \( q \)-dimensional vectors of unknown regression coefficients, respectively; and \( \phi(\cdot) \) is a \( q \)-dimensional vector of known functions of the counting process \( H_i(t) \) up to time \( t \) and the covariates \( W_i \), representing the interaction between the observation history and some covariates. In particular, for panel count data arising from the longitudinal follow-up clinical studies with different treatments, \( W_i \)'s can be considered as the treatment indicators, and thus \( \alpha \) represents the effect of the interaction between the frequency of observation times and the treatment group on the underlying recurrent event process. If \( \alpha = 0 \), then model (1) reduces to the model considered by Sun and Wei (2000), Zhang (2002), and Wellner and Zhang (2007) for regression analysis of panel count data. In fact, our modeling approach is different from the existing approaches. Here, the possible effect of the observation process is directly incorporated into the conditional model about the underlying recurrent event process, and no additional model assumption is needed for the observation process.

For making valid inference, we assume that

\[
E[N_i(t)|X_i, H_i(s), 0 \leq s \leq t, C_i] = E[N_i(t)|X_i, F_i, C_i],
\]

which means that the mean of the recurrent event process valued at time point \( t \) only depends on the observation history before \( t \), conditional on the covariates and the censoring time. Let \( \mathbf{O} = (K, \bar{T}_K, \bar{N}_K, \bar{H}_K, \mathbf{X}, \mathbf{C}) \), with \( \bar{T}_K = (T_{K,1}, \ldots, T_{K,K}) \), \( \bar{N}_K = (N(T_{K,1}), \ldots, N(T_{K,K})) \), \( \bar{H}_K = (H(T_{K,1}), \ldots, H(T_{K,K})) \). Throughout this article, we will assume that we observe \( n \) i.i.d. copies, \( \mathbf{O}_1, \ldots, \mathbf{O}_n \) of \( \mathbf{O} \). The goal is to estimate the unknown nondecreasing function \( \mu_0(t) \) and the regression parameters \( \beta \) and \( \alpha \).

### 3 Estimation procedure

For inference about model (1), we denote \( \tilde{\mu}(t) = \log(\mu(t)) \) and, using the generalized least square estimation approach given in Hu et al. (2009), define the following objective function:

\[
L_n(\beta, \alpha, \tilde{\mu}) = \sum_{i=1}^{n} \left( \frac{N_i(T_{K,i}) - \exp[\tilde{\mu}(T_{K,i}) + \beta' X_i + \alpha' \phi(F_{T_{K,i}}, W_i)]}{\xi_i(T_{K,i})} \right)^2 = \sum_{i=1}^{n} \int_{0}^{\tau} \left[ N_i(t) - \exp[\tilde{\mu}(t) + \beta' X_i + \alpha' \phi(F_{t}, W_i)] \right]^2 d\bar{H}_i(t),
\]

where \( \xi_i(t) = I(C_i \geq t) \).

To make inference about \( \mu_0(t) \), we propose to use B-splines to approximate its logarithm \( \mu_0(t) = \log(\mu_0(t)) \). For a finite closed interval \([0, \tau]\), let \( \mathcal{I} = (t_i)_{i=1}^{m+2l+1} \), with

\[
0 = t_1 = \cdots = t_l < t_{l+1} < \cdots < t_{m+l} < t_{m+l+1} = \cdots = t_{m+2l} = \tau,
\]

be a sequence of knots that partition \([0, \tau]\) into \( m + 1 \) subintervals and \( m_n = O(n^\nu) \), for \( 0 < \nu < 1/2 \).

Let \( \Psi_{l,\mathcal{I}} \) be the class linearly spanned by the B-splines basis functions \( \{B_{il}, 1 \leq i \leq q_n, q_n = m_n + l\} \) with order \( l \) and knots \( \mathcal{I} \), that is

\[
\Psi_{l,\mathcal{I}} = \left\{ \sum_{i=1}^{q_n} \gamma_i B_{il} : \gamma_i \in \mathbb{R}, i = 1, \ldots, q_n \right\}.
\]
We now define a subclass of $\Psi_{I, X}$, as $\Phi_{I, X} = \{\sum_{j=1}^{q} Y_j B_j : \gamma_1 \leq \cdots \leq \gamma_{q_n}\}$. According to the variation-diminishing properties of B-splines (Schumaker, 1981), $\Phi_{I, X}$ is a class of nondecreasing splines on $[0, \tau]$. Then, we can approximate the smooth monotone function $\tilde{\mu}_0(t)$ by $\sum_{j=1}^{q_n} Y_j B_j(t)$ and estimate the coefficients $\gamma_1 \leq \cdots \leq \gamma_{q_n}$ and regression parameters $\beta$ and $\alpha$ jointly through minimizing the approximated expression $L_n(\beta, \alpha, \tilde{\mu})$ subject to nondecreasing constraints.

Using the spline-based sieve approximation $\mu_n(t) = \sum_{i=1}^{q_n} \gamma_i B_i(t)$ of $\mu_0(t)$, $L_n(\beta, \alpha, \tilde{\mu})$ in Eq. (2) can be approximated by

$$L_n(\beta, \alpha, \gamma) = \sum_{i=1}^{q_n} \int_0^\tau \left[ N_i(t) - \exp[\gamma' B_i(t) + \beta' X_i + \alpha' \phi(F_{ij}, W_j)] \right]^2 d\tilde{H}_i(t),$$

(3)

where $\gamma = (\gamma_1, \cdots, \gamma_{q_n})'$ and $B_i(t) = (B_{i1}(t), \cdots, B_{iq_n}(t))'$.

Let $\hat{\beta}_n$, $\hat{\alpha}_n$, $\hat{\gamma}_n$ be the values that minimize $L_n(\beta, \alpha, \gamma)$ in Eq. (3) with the constraint $\gamma_1 \leq \cdots \leq \gamma_{q_n}$.

Then the monotone spline estimator for $\mu_0(t)$ is $\hat{\mu}_n(t) = \hat{\gamma}_n B(t)$. Thus $\mu_0(t)$ can be estimated by $\hat{\mu}_n(t) = \exp[\hat{\gamma}_n B(t)]$. The spline estimation problem can be formulated as the linear inequality constrained minimization problem $\min_{\gamma \in R^{q+n}} L_n(\eta)$, where $\eta = (\beta', \alpha' \gamma')'$ with $\gamma \in \Theta_n = \{\gamma : \gamma_1 \leq \cdots \leq \gamma_{q_n}\}$.

Jamshidian (2004) proposed a generalized gradient projection (GP) algorithm for optimizing a nonlinear objective function with linear inequality constraints, based on the generalized Euclidean metric $\|x\| = x^T V x$ with $V$ being a positive definite matrix and possibly varying from iteration to iteration. Zhang and Jamshidian (2004) applied the GP algorithm to large-scale nonparametric maximum-likelihood estimation problems by choosing $V = D_v$, the matrix containing only the diagonal elements of the negative Hessian matrix $U$, in order to avoid the storage problem in updating $U$. However, this will increase the number of iterations and thereby the computing time. Lu et al. (2007, 2009) used the generalized GP algorithm in Zhang and Jamshidian (2004) with $V = U$ directly because the dimension of unknown parameter space is usually small in their applications due to the use of polynomial splines, which would also substantially reduce the number of iterations. Here we consider using the generalized GP algorithm for the monotone polynomial spline estimation with $V$ unequal to the negative Hessian matrix $U$.

Let $\nabla L_n(\eta)$ be the negative gradient of $L_n(\eta)$ with respect to $\eta$ and

$$V = \sum_{i=1}^{n} \int_0^\tau \exp[\eta Z_{ij}(t)]^2 Z_{ij}^2(t) d\tilde{H}_i(t),$$

which is a positive definite matrix with $Z_{ij}(t) = (X_i, \phi(F_{ij}, W_j), B_i(t))'$.

Furthermore, $n^{-1}(2V)$ has the same limit as the negative Hessian matrix of $-L_n(\eta)$.

Let $A = \{I_1, I_2, \cdots, I_m\}$ denote the index set of active constraints, that is $\gamma_{I_j} = \gamma_{I_{j+1}}$, for $j = 1, \cdots, m$, during the numerical computation. $A$ is allowed to be empty when $m = 0$. We define an $m$ by $p + q + q_n$ working matrix corresponding to $A$ as

$$A = \begin{pmatrix}
0 & \cdots & 0 & 0 & -1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & -1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & 0 & -1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & 0 & -1 & 1 & \cdots & 0 & 0 & \cdots & 0 & 0 \\
\end{pmatrix}.$$
that is, the \( j \)-th row (\( j = 1, \ldots, m \)) consists of \( (p + q + q_n) \) components with its \( (p + q + i_j) \)-th and 
\((p + q + i_j + 1)\)-th elements equal to \(-1\) and \(1\), respectively, and the remaining components zero. The 
generalized GP algorithm is summarized as follows.

Start with a feasible initial value \( \eta \in \mathbb{R}^{p+q} \times \Theta_{p,q} \), and cycle through the following steps until conver-
gence.

S0. (Computing the feasible search direction) Calculate
\[
d = (I - V^{-1} A^T (AV^{-1} A^T)^{-1} A) V^{-1} \nabla L_n(\eta).
\]
When there is no active constraint, take \( d = V^{-1} \nabla L_n(\eta) \).

S1. (Forcing the updated \( \eta \) to fulfill the constraints) If the resulted direction \( d \) is not nondecreasing
in its components, compute
\[
\varphi = \min_{i \in A, d_i \neq d_{i+1}} \left( -\frac{\gamma_{i+1} - \gamma_i}{d_{i+1} - d_i} \right).
\]
Doing so guarantees that \( \gamma_{i+1} + \varphi d_{i+1} \geq \gamma_i + \varphi d_i \), for \( i = 1, \ldots, q_n \).

S2. (Step-halving line search) Find the smallest integer \( k \) starting from 0 such that
\[
\| \nabla L_n(\eta + (1/2)^k d) \| < \| \nabla L_n(\eta) \|.
\]

S3. (Updating the solution) If \( \varphi > (1/2)^k \), replace \( \eta \) by \( \tilde{\eta} = \eta + (1/2)^k d \) and check the stopping
criterion (S5).

S4. (Updating the active constraint set) If \( \varphi \leq (1/2)^k \), in addition to replace \( \eta \) by \( \tilde{\eta} = \eta + \varphi d \),
modify \( A \) by adding indexes of all the newly active constraints to \( A \) and accordingly modify
the working matrix \( A \).

S5. (Checking the stopping criterion) If \( \| d \| \geq \varepsilon \) for a small \( \varepsilon > 0 \), go to S0, otherwise, compute
the Lagrange multipliers \( \lambda = (AV^{-1} A^T)^{-1} AV^{-1} \nabla L_n(\eta) \).
(i) If \( \lambda_i \leq 0 \) for all \( i \in A \), set \( \tilde{\eta} = \eta \) and stop.
(ii) If at least one \( \lambda_i > 0 \), for \( i \in A \), remove the index corresponding to the largest \( \lambda_i \) from \( A \),
and update \( A \) and go to S0.

Let \( \hat{\theta}_n = (\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) \) and \( \theta_0 \) be the true value of \( \theta \). Under some regularity conditions stated in
Appendix, \( \hat{\theta}_n \) is a consistent estimator of \( \theta_0 \) and \( \sqrt{n} \left( \frac{\hat{\beta}_n - \beta_0}{\sigma_n} \right) \) converges in distribution to \( N(0, \Sigma) \),
where \( \Sigma \) is given in Theorem 3 of Appendix.

4 Simulation study

In this section, we conducted a simulation study to assess the finite-sample properties of the proposed
estimators. We considered the situation where there were two covariates and for each subject \( i \), \( X_{1i}, X_{2i}, \) and \( X_{2i} \) were generated from Bernoulli distribution with success probability \( 0.5 \) and the uniform
distribution over interval \((-1, 1)\), respectively. The follow-up time \( C_i \) was generated from the uniform
distribution over interval \( (\tau/2, \tau) \) with \( \tau = 6 \). Given the covariate \( X_i = (X_{1i}, X_{2i})' \) and \( C_i \), two setups
for the observation process \( H_i(t) \) were considered as follows:

(i) The number of observation times \( m_i \) was assumed to follow the Poisson distribution with mean
\((2C_i/\tau) \exp(-0.5X_{1i} + 0.5X_{2i}) \) and the observation times \( T_{m1}, \ldots, T_{m_m} \) were taken to be
the order statistics of a random sample of size \( m_i \) from the uniform distribution over \((0, C_i)\)
(ii) The number of observation times \( n_i \) was assumed to follow the uniform distribution over \{1, 2, 3, 4, 5, 6\} and the observation times \((T_{m,i}, \ldots, T_{m,m_i})\) were generated in the same way as in setup (i).

Then, given \( X_i, m_i \), and the observation times \((T_{m,i}, \ldots, T_{m,m_i})\), we generated recurrent event counts \( N_{m_i} = (N_i(T_{m,i}), \ldots, N_i(T_{m,m_i})) \) from a Poisson process by taking

\[
N_i(T_{m,i}) = N_i(T_{m,i}) + [N_i(T_{m,i} + 1) - N_i(T_{m,i})] + \cdots + [N_i(T_{m,m_i}) - N_i(T_{m,m_i} - 1)],
\]

with \( N_i(T_{m,i}) \sim \text{Poisson}(\exp[\tilde{\mu}_0(T_{m,i}) + \beta_0 X_i + \alpha_0 H_i(T_{m,i} - 1) W_i]) \), and

\[
N_i(T_{m,m_i}) - N_i(T_{m,m_i} - 1) \sim \text{Poisson}(\exp[\tilde{\mu}_0(T_{m,m_i}) + \beta_0 X_i + \alpha_0 H_i(T_{m,m_i} - 1) W_i])
- \exp[\tilde{\mu}_0(T_{m,m_i} - 1) + \beta_0 X_i + \alpha_0 H_i(T_{m,m_i} - 1) W_i]),
\]

for \( j \geq 2 \).

We took \( \beta_0 = (-0.5, 0.5) \); \( \alpha_0 = 0, 0.3, \text{or } 0.5 \); and \( \tilde{\mu}_0(t) = \sqrt{t} \) or \( \log(t + 1) \), and used the cubic B-splines to compute the spline estimators. To choose the number of interior knots, we partitioned the range of \( v \), \((0, 0.5)\) to 20 equally subintervals and chose \( u \) to be the partition points. For each value of \( v \), we took \( m_n = n^v \) as the number of interior knots. To determine locations of knots, we considered two commonly used data-driven methods. One is the equally spaced knots given by \( T_{min} + k(T_{max} - T_{min})/(m_n + 1), \) \( k = 0, 1, \ldots, m_n + 1 \), where \( T_{min} \) and \( T_{max} \) denote the minimum and maximum values of distinct observation times, respectively. Another is the partitions corresponding to quantiles of the observation times, that is, the \( k/\text{fn} + 1 \) quantiles \( (k = 0, 1, \ldots, m_n + 1) \) of the distinct observation times as the knots. The value of \( v \) that minimizes the Bayesian information criteria (BIC) was selected. Here \( \text{BIC} = 2 \log(L) + \log(n)(q_n + 3) \), where \( L = L_n(\tilde{\beta}, \tilde{\alpha}, \tilde{\mu}_0) \) as defined in Eq. (2) and \( q_n \) is the number of the B-spline basis functions. We carried out simulations for the different situations of the number and placement of knots with \( W = X_i, \alpha = 0.3, \tilde{\mu}_0(t) = \log(t + 1) \), and \( n = 100 \), and found that the estimation results are very similar and the method is insensitive to the selection of number and placement of knots, where the value of \( v \) was selected by the BIC as \( 1/8 \). Thus in the following, we took the number of interior knots as \( n^{1/8} \) and used the equally spaced knots.

To initialize the algorithm, we choose \( \gamma = (1, 2, \ldots, q_n) \), while \( \beta \) and \( \alpha \) were all generated from the uniform distribution over interval \((-0.5, 0.5)\). Tables 1 and 2 present the simulation results on estimation of \( \tilde{\beta}_0 \) and \( \tilde{\alpha}_0 \) under Poisson and non-Poisson observation processes with sample size \( n = 100 \) or \( 200 \) for \( \tilde{\mu}_0(t) = \sqrt{t} \) or \( \log(t + 1) \), respectively. The tables include the estimated bias (BIAS) given by the average of the estimates minus the true value, the sample standard deviation error of estimates (SSE), the mean of the bootstrap standard errors of the estimates (BSE), and the bootstrap 95% coverage probability (CP) obtained from 1000 independent runs. Here we used 100 replications in bootstrap to estimate the standard errors. It can be seen from the tables that the proposed estimators are unbiased for different situations considered, which means that our estimation approach does not rely on the Poisson distributional assumption about the observation process.

Also, the SSE and BSE are quite close to each other and become smaller as the sample size increases, which indicates that proposed bootstrap variance estimation procedure provides reasonable estimates. In addition, the 95% bootstrap CP is consistent with the nominal level, which suggests that the normal approximation seems to be appropriate.

To estimate the smooth function \( \tilde{\mu}_0(t) \), we considered cubic B-splines and took \( m_n = n^v \) with \( v = 1/8 \). For a given number of interior knots \( m_n \), the equally spaced knots are chosen. Figure 1 shows the estimation results of \( \tilde{\mu}_0(t) = \log(t + 1) \) for simulated panel count data with Poission and non-Poisson observation processes, \( \phi(F, W) = H(t)X_i \) and \( \alpha = 0.3 \). In the figure, the solid line represents the real curve of \( \tilde{\mu}_0(t) \), and the point line and the dotted line represent the B-spline
Table 1  Estimation results of \((\beta'_0, \alpha'_0)'\) for simulated panel count data with Poisson and non-Poisson observation processes, \(\phi(F_t, W) = H(t-)X_1\), and \(\tilde{\mu}_0(t) = \sqrt{t}\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(\alpha)</th>
<th>(\hat{\beta}_1) BIAS</th>
<th>(\hat{\beta}_1) SSE</th>
<th>(\hat{\beta}_1) BSE</th>
<th>(\hat{\beta}_1) CP</th>
<th>(\hat{\beta}_2) BIAS</th>
<th>(\hat{\beta}_2) SSE</th>
<th>(\hat{\beta}_2) BSE</th>
<th>(\hat{\beta}_2) CP</th>
<th>(\hat{\alpha}) BIAS</th>
<th>(\hat{\alpha}) SSE</th>
<th>(\hat{\alpha}) BSE</th>
<th>(\hat{\alpha}) CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0</td>
<td>P</td>
<td>-0.0019</td>
<td>0.1531</td>
<td>0.1523</td>
<td>0.9470</td>
<td>0.0092</td>
<td>0.1359</td>
<td>0.1279</td>
<td>0.9280</td>
<td>-0.0080</td>
<td>0.1204</td>
<td>0.1305</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
<td>-0.0011</td>
<td>0.1370</td>
<td>0.1305</td>
<td>0.9300</td>
<td>0.0050</td>
<td>0.0980</td>
<td>0.0934</td>
<td>0.9370</td>
<td>-0.0013</td>
<td>0.0411</td>
<td>0.0393</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>P</td>
<td>0.0101</td>
<td>0.1541</td>
<td>0.1501</td>
<td>0.9470</td>
<td>0.0056</td>
<td>0.1283</td>
<td>0.1232</td>
<td>0.9450</td>
<td>-0.0064</td>
<td>0.0990</td>
<td>0.1108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
<td>0.0020</td>
<td>0.1262</td>
<td>0.1234</td>
<td>0.9410</td>
<td>0.0035</td>
<td>0.0745</td>
<td>0.0741</td>
<td>0.9510</td>
<td>0.0006</td>
<td>0.0316</td>
<td>0.0295</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>P</td>
<td>-0.0028</td>
<td>0.1627</td>
<td>0.1536</td>
<td>0.9380</td>
<td>0.0143</td>
<td>0.1328</td>
<td>0.1207</td>
<td>0.9260</td>
<td>-0.0035</td>
<td>0.0978</td>
<td>0.1056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
<td>0.0037</td>
<td>0.1169</td>
<td>0.1188</td>
<td>0.9470</td>
<td>-0.0005</td>
<td>0.0684</td>
<td>0.0757</td>
<td>0.9740</td>
<td>-0.0019</td>
<td>0.0525</td>
<td>0.0585</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>P</td>
<td>-0.0117</td>
<td>0.1095</td>
<td>0.1057</td>
<td>0.9400</td>
<td>0.0025</td>
<td>0.0909</td>
<td>0.0897</td>
<td>0.9410</td>
<td>-0.0035</td>
<td>0.0807</td>
<td>0.0799</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
<td>0.0046</td>
<td>0.0931</td>
<td>0.0921</td>
<td>0.9540</td>
<td>0.0043</td>
<td>0.0674</td>
<td>0.0657</td>
<td>0.9470</td>
<td>-0.0008</td>
<td>0.0278</td>
<td>0.0275</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>P</td>
<td>-0.0047</td>
<td>0.1069</td>
<td>0.1069</td>
<td>0.9490</td>
<td>0.0077</td>
<td>0.0909</td>
<td>0.0864</td>
<td>0.9380</td>
<td>-0.0024</td>
<td>0.0638</td>
<td>0.0637</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
<td>-0.0027</td>
<td>0.0913</td>
<td>0.0866</td>
<td>0.9370</td>
<td>0.0024</td>
<td>0.0565</td>
<td>0.0525</td>
<td>0.9260</td>
<td>0.0012</td>
<td>0.0216</td>
<td>0.0206</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>P</td>
<td>0.0015</td>
<td>0.1101</td>
<td>0.1077</td>
<td>0.9520</td>
<td>0.0045</td>
<td>0.0899</td>
<td>0.0845</td>
<td>0.9430</td>
<td>-0.0030</td>
<td>0.0606</td>
<td>0.0602</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
<td>-0.0004</td>
<td>0.0850</td>
<td>0.0847</td>
<td>0.9420</td>
<td>0.0027</td>
<td>0.0494</td>
<td>0.0535</td>
<td>0.9710</td>
<td>0.0007</td>
<td>0.0214</td>
<td>0.0390</td>
</tr>
</tbody>
</table>

Notes: P and NP represent observation times generated from Poisson and non-Poisson processes, respectively.
Table 2. Estimation results of \((\hat{\beta}_0, \alpha'_0)'\) for simulated panel count data with Poisson and non-Poisson observation processes, \(\phi(F_t, W) = H(t-)X_1\), and \(\tilde{\mu}_0(t) = \log(t + 1)\).

<table>
<thead>
<tr>
<th>n</th>
<th>(\alpha)</th>
<th>(\hat{\beta}_1)</th>
<th>(\hat{\beta}_2)</th>
<th>(\hat{\alpha})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BIAS</td>
<td>SSE</td>
<td>BSE</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>P</td>
<td>0.0020</td>
<td>0.1835</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
<td>-0.0015</td>
<td>0.1595</td>
</tr>
<tr>
<td>0.3</td>
<td>P</td>
<td>0.0017</td>
<td>0.1867</td>
<td>0.1813</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>-0.0015</td>
<td>0.1476</td>
<td>0.1460</td>
</tr>
<tr>
<td>0.5</td>
<td>P</td>
<td>-0.0033</td>
<td>0.1928</td>
<td>0.1843</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>-0.0009</td>
<td>0.1518</td>
<td>0.1385</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>P</td>
<td>-0.0072</td>
<td>0.1254</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
<td>-0.0053</td>
<td>0.1117</td>
</tr>
<tr>
<td>0.3</td>
<td>P</td>
<td>0.0005</td>
<td>0.1317</td>
<td>0.1283</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>-0.0016</td>
<td>0.1039</td>
<td>0.1037</td>
</tr>
<tr>
<td>0.5</td>
<td>P</td>
<td>0.0018</td>
<td>0.1288</td>
<td>0.1289</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>0.0004</td>
<td>0.0991</td>
<td>0.0985</td>
</tr>
</tbody>
</table>

Notes: P and NP represent observation times generated from Poisson and non-Poisson processes, respectively.
Figure 1  Estimates of $\tilde{\mu}_0(t) = \log(t + 1)$ for simulated panel count data with (A) Poisson and (B) non-Poisson observation processes, $\phi(F_t, W) = H(t^-)X_1$, and $\alpha = 0.3$.

estimates of $\tilde{\mu}_0(t)$ for the sample size $n = 100$ and $n = 200$, respectively. From the figure, one can see that the B-spline estimate is close to its real curve with the moderate sample size and especially closer as the sample size increases for all the situations considered here, indicating that the B-spline estimator for $\tilde{\mu}_0(t)$ performs well.

To compare our estimators with the maximum spline pseudo-likelihood estimators (MSPLE) and maximum spline likelihood estimators (MSLE) in Lu et al. (2009), we considered the same setup as
above except that for given \( X_i, m_i \), and \( \{T_{m_i,1}, \ldots, T_{m_i,m_i}\} \), the recurrent event counts were generated from a mixed Poisson process by taking

\[
N_i(T_{m_i,j}) = N_i(T_{m_i,1}) + \{N_i(T_{m_i,2}) - N_i(T_{m_i,1})\} + \cdots + \{N_i(T_{m_i,j}) - N_i(T_{m_i,j-1})\},
\]

with \( N_i(T_{m_i,1})|Q_i \sim \text{Poisson}(Q_i \exp(\mu_0(T_{m_i,1}) + \beta_0 X_i + \alpha_0 H_i(T_{m_i,1}) W_i)) \), and

\[
[N_i(T_{m_i,j}) - N_i(T_{m_i,j-1})]|Q_i \sim \text{Poisson}(Q_i \left[ \exp(\mu_0(T_{m_i,j}) + \beta_0 X_i + \alpha_0 H_i(T_{m_i,j}) W_i) + \right. \exp(\mu_0(T_{m_i,j-1}) + \beta_0 X_i + \alpha_0 H_i(T_{m_i,j-1}) W_i))],
\]

for \( j \geq 2 \) with \( Q_i \) generating from the gamma distribution with mean 1 and variance 0.1. The simulation results are shown in Table 3. From the table, we have the following findings: (i) With both Poisson and non-Poisson observation processes, the proposed estimator, MSPLE, and MSLE in Lu et al. (2009) are approximately unbiased when \( \alpha = 0 \). (ii) When \( \alpha \neq 0 \), the proposed estimator is approximately unbiased while the MSPE and MSLE in Lu et al. (2009) yield biased estimates and the biases could be larger as \( \alpha \) diverges from 0. This means that our proposed estimation procedure seems to be more robust.

We also conducted some sensitivity analysis to evaluate the performance of the proposed estimators for \( \beta \) when the interaction term of the recurrent event process and observation process was misspecified. Specifically, we generated the recurrent event counts from a mixed Poisson process with

\[
\phi(\mathcal{F}, W) = H(t^-) X_1 \text{ and } \mu_0(t) = \log(t + 1).
\]

Table 3 Comparison of estimates for \( \beta_0 \), based on generated data from mixed Poisson recurrent event process with \( \phi(\mathcal{F}, W) = H(t^-) X_1 \) and \( \mu_0(t) = \log(t + 1) \).

<table>
<thead>
<tr>
<th>n</th>
<th>( \alpha )</th>
<th>Method</th>
<th>Poisson observation</th>
<th>Non-Poisson observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_1 )</td>
<td>( \hat{\beta}_2 )</td>
<td>( \hat{\beta}_1 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BIAS</td>
<td>SSE</td>
<td>BIAS</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>Proposed</td>
<td>-0.0031</td>
<td>0.2213</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSPL</td>
<td>0.0023</td>
<td>0.1968</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>Proposed</td>
<td>-0.0078</td>
<td>0.2271</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSPL</td>
<td>0.2372</td>
<td>0.2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSLE</td>
<td>0.2702</td>
<td>0.1816</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>Proposed</td>
<td>0.0002</td>
<td>0.1584</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSPL</td>
<td>0.0038</td>
<td>0.1348</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSLE</td>
<td>0.0031</td>
<td>0.1194</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>Proposed</td>
<td>-0.0076</td>
<td>0.1704</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSPL</td>
<td>0.2563</td>
<td>0.1447</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSLE</td>
<td>0.2902</td>
<td>0.1287</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>Proposed</td>
<td>-0.0098</td>
<td>0.1848</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSPL</td>
<td>0.4699</td>
<td>0.1783</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSLE</td>
<td>0.5343</td>
<td>0.1696</td>
</tr>
</tbody>
</table>

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. www.biometrical-journal.com
process as above with $\mu_0(t) = \log(t + 1)$ and the interaction term $\phi(F_i, W_i) = (H_i(t) - 1)X_{1i}$, where $X_{1i}, X_{2i}, C_i$ and the observation times were generated from the same setup as given above. Here we take $\beta_0 = (-1, 1)$. We considered the interaction term $\phi$ as misspecified by three possible forms: (i) $\phi^{(1)}(F_i, W_i) = (H_i(t) - 0.75)X_{1i}$, (ii) $\phi^{(2)}(F_i, W_i) = \log[\exp(H_i(t) + 0.5) + 1]X_{1i}$, and (iii) $\phi^{(3)}(F_i, W_i) = \sqrt{[H_i(t) - 1]^2 + 1.2X_{1i}}$. We applied the proposed estimation procedure to the true and misspecified models by using the generated data from the true model. The simulation results are summarized in Table 4. It can be seen from the table that the estimates for $\beta$ are still approximately unbiased for the misspecified situations considered here.

5 Application

In this section, we applied our proposed method to analyze the bladder cancer data. There were 116 subjects with superficial bladder tumors, and they were randomized into one of three treatment groups: placebo, thiotepa, and pyridoxine. Following Sun and Wei (2000), we restricted our attention to the placebo and thiotepa groups with respective sizes of 47 and 38. For each patient, all the clinical visits and the numbers of new tumors between clinical visits were recorded. In addition, two baseline covariates were considered and they were the number of initial tumors and the size of the largest initial tumor.

For the analysis, we define the response process $N_i(t)$ and observation process $H_i(t)$ as the cumulative tumor numbers and the accumulated observation numbers of patient $i$ up to time $t$, respectively. Further, for patient $i$, define $X_{1i}$ to be equal to 1 if the $i$-th patient was given the thiotepa treatment and 0 otherwise, $X_{2i}$ the number of initial tumors and $X_{3i}$ the size of the largest initial tumor, $i = 1, \ldots, 85$. Take $H^*(t) = (H(t) - 8)/8$ and assume that $\{N_i(t)\}$ can be described by model (1) with $\phi(F_i, W_i) = H^*(t)X_{1i} + H^*(t)X_{2i}$, meaning that the relation between the recurrence rate of bladder tumors and the observation times may vary with different treatments and different number of initial tumors, that is,

$$
E\{N_i(t)|X_{1i}, X_{2i}, X_{3i}, F_i\} = \exp(\mu(t) + \beta_1'X_{1i} + \beta_2'X_{2i} + \beta_3'X_{3i} + \alpha_1' H^*(t)X_{1i} + \alpha_2' H^*(t)X_{2i}).
$$

Here, we took the last visit time of patient $i$ as $C_i$ in the analysis. For estimation of $\mu(t)$, we use the cubic B-spline approximation by taking the number of interior knots $m_n$ as $n^v$ with $v = 1/8$ and the equally spaced knots.

The application of the estimation procedure proposed in the previous sections gave $\hat{\beta}_1 = -1.7143, \hat{\beta}_2 = 0.1061, \hat{\beta}_3 = 0.0473, \hat{\alpha}_1 = -0.6100$, and $\hat{\alpha}_2 = 0.3070$ with the bootstrap standard errors (100 replications) being 0.3205, 0.0465, 0.0335, 0.1060, and 0.1079, which correspond to $p$-values of $<0.0000, 0.0225, 0.1584, <0.0000$, and 0.0044, respectively, based on the asymptotic results of the estimators. Here $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ represent the estimated regression coefficients corresponding to the treatment indicator, the number of initial tumors, and the size of the largest initial tumor, respectively, while $\hat{\alpha}_1$ and $\hat{\alpha}_2$ represent the estimated effects of the interaction between the observation process and the treatment indicator and the interaction between the observation process and the number of initial tumors, respectively, on the tumor recurrence rate. These results indicate that the tumor recurrence process and the interaction between the observation process and the thiotepa treatment indicator are significantly negatively correlated. The reason for this may be that the patients in the thiotepa group needed more clinical visits, had tumors removed, and received treatment, which reduced the tumor occurrence rate. In addition, the interaction between the observation process and the number of initial tumors significantly positively influences the tumor recurrence process, which can be explained by the reason that patients with more initial tumors tend to have more clinical visits and more new tumors will arise due to the more serious illness. Furthermore, the thiotepa treatment significantly reduces the occurrence rate of the bladder tumors, and the number of initial tumors has a significantly positive effect on the tumor recurrence rate. However, the recurrence rate of the bladder tumors does not seem
Table 4  Sensitivity analysis for the misspecification of the interaction term $\phi$ with $\tilde{\mu}_0(t) = \log (t + 1)$ and $\beta_0 = (-1, 1)$.

<table>
<thead>
<tr>
<th>Obs</th>
<th>$n$</th>
<th>$\alpha$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>BIAS</td>
<td>BIAS$^1$</td>
<td>BIAS$^2$</td>
</tr>
<tr>
<td>P</td>
<td>100</td>
<td>0</td>
<td>-0.0146</td>
<td>-0.0140</td>
<td>-0.0167</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>-0.0152</td>
<td>-0.0073</td>
<td>-0.0411</td>
<td>-0.0741</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-0.0051</td>
<td>-0.0087</td>
<td>-0.0413</td>
<td>-0.0959</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>-0.0085</td>
<td>-0.0036</td>
<td>-0.0089</td>
<td>-0.0085</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>-0.0037</td>
<td>-0.0059</td>
<td>-0.0387</td>
<td>-0.0629</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-0.0007</td>
<td>-0.0046</td>
<td>-0.0387</td>
<td>-0.0794</td>
</tr>
<tr>
<td>NP</td>
<td>100</td>
<td>0</td>
<td>-0.0027</td>
<td>-0.0070</td>
<td>-0.0035</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>-0.0010</td>
<td>0.0002</td>
<td>-0.0119</td>
<td>-0.0260</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-0.0013</td>
<td>-0.0008</td>
<td>-0.0162</td>
<td>-0.0323</td>
</tr>
<tr>
<td>200</td>
<td>0</td>
<td>-0.0012</td>
<td>-0.0032</td>
<td>-0.0025</td>
<td>-0.0012</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>-0.0010</td>
<td>-0.0009</td>
<td>-0.0130</td>
<td>-0.0262</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>-0.0002</td>
<td>0.0009</td>
<td>-0.0152</td>
<td>-0.0314</td>
</tr>
</tbody>
</table>

Notes: P and NP represent observation times generated from Poisson and non-Poisson processes, respectively. BIAS, the estimated bias of the parameter with the true interaction term $H$; BIAS$^1$, BIAS$^2$, and BIAS$^3$, the estimated biases of the parameter with misspecified interaction terms $H^{(1)}$, $H^{(2)}$, and $H^{(3)}$, respectively.
to be significantly related to the size of the largest initial tumor. These conclusions are consistent with those made in Lu et al. (2009), Li et al. (2010), and Zhao and Tong (2011), among others. For example, Lu et al. (2009) analyzed the bladder tumor data, and obtained that the number of initial tumors has a positive effect on the tumor recurrence rate with $p$-values of 0.012 and 0.01 and the thiotepa treatment significantly reduces the recurrence of tumors with $p$-values of 0.015 and 0.019 by MSPLE and MSLE, respectively. Compared to the models in Lu et al. (2009), Li et al. (2010), and Zhao and Tong (2011), our fitted model may provide more information about the correlation between the tumor recurrence rate and observation times over treatment groups and also could be useful to estimate the future recurrence rate based on the observation history.

6 Concluding remarks

This article studied a conditional model for the underlying recurrent event process of the panel count data that allows for the interaction between the informative observation times and covariates, leaving the distributional form of the observation process to be arbitrary. We proposed to use the easy implemented monotone spline-based least squares estimation method to estimate the regression parameters and the unknown smooth monotone function in the model simultaneously, and also established the asymptotical results including consistency, rate of convergence, and asymptotic normality for the estimators. As demonstrated by simulation studies, our proposed model and inference procedure are more flexible and applicable since they can overcome the underdispersion or overdispersion problem resulting from the model specification for the observation process.

In the proposed estimation approach, weights have not been considered. Following the discussions made in Hu et al. (2009), our proposed objection function in Eq. (2) may be related to the MSPLE and MSLE (Lu et al., 2009) by taking some specific weight functions. Further studies are needed to develop monotone spline-based weighted least squares methods and provide weight selection procedure for finding optimal weight function in practical applications.

Motivated by Li et al. (2010), our proposed models can also be extended to a class of transformation models

$$E(N_i(t)|X_i, W_i, F_i) = G(\mu_0(t) + \beta' X_i + \alpha' \phi(F_i, W_i)),$$

with a given monotone smooth function $G$. Then for making inference on these models, the same procedure as presented in this article can be used to obtain the estimators for the regression parameters $\beta$ and $\alpha$ and the nonparametric monotone function $\mu_0(t)$, and the asymptotic properties of the spline-based estimators could be established by using the similar arguments.

Acknowledgment The authors would like to thank the editor, Professor Mauro Gasparini, associate editor, and two reviewers for their constructive and insightful comments and suggestions that greatly improved the article. Deng’s research is partly supported by the Natural Science Foundation of China (nos. 11101315, 11401443). Liu’s research is partly supported by the Natural Science Foundation of China (nos. 11101315, 11171262), and Zhao’s research is partly supported by the Research Grant Council of Hong Kong (nos. 504011, 503513), Natural Science Foundation of China (no. 11371299), and The Hong Kong Polytechnic University.

Conflict of interest

The authors have declared no conflict of interest.

Appendix: Asymptotic results

To establish the asymptotic properties of the estimators, we need the following regularity conditions.

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

www.biometrical-journal.com
C1. The maximum spacing of the knots satisfies $\Delta = \max_{j=1, \ldots, m} |t_j - t_{j-1}| = O(n^{-\nu})$.

C2. The parameter spaces of $(\beta', \alpha')$, $R$ is bounded and convex on $\mathbb{R}^{p+q}$, and the true parameter $\theta_0 = (\beta_0, \alpha_0, \mu_0) \in \mathcal{K}$, where $\mathcal{K}$ is the interior of $\mathcal{R}$, and

$$\mathcal{F}_r \equiv \left\{ \mu : [0, \infty) \to \mathbb{R} | \mu \text{ is monotone and } |\mu^{(k)}(s) - \mu^{(k)}(t)| \leq M |s - t|^\zeta \right\},$$

where $k$ is a nonnegative integer, $\zeta \in (0, 1]$ such that $r = k + \zeta > 0.5$, $M$ is a positive constant and $f^{(k)}$ is the $k$-th derivative of function $f$.

C3. $N_i(\tau)(i = 1, \ldots, n)$ are bounded by a constant, $h(\cdot)$ has a bounded total variation, and there exist a positive integer $M_1$, such that $P(|X| \leq M_1) = 1$, that is, the covariate vector is uniformly bounded.

C4. There exists a positive integer $M_2$ such that $P(K \leq M_2) = 1$, that is, the number of the observation is finite.

C5. If with probability $1$, $h_1(X + h_2\phi(\mathcal{F}_r, W) + h_3(t) = 0$ for $h_1 \in \mathbb{R}^p$, $h_2 \in \mathbb{R}^q$ and some deterministic function $h_3$, then $h_1 = 0, h_2 = 0, h_3 = 0$.

Next, we introduce more notation. Let

$$\mathcal{F} = \left\{ f : [0, \infty) \to \mathbb{R} \left| \|f\|_2 = \left[ E \left\{ \int_0^\tau |f(t)|^2dH(t) \right\} \right]^{1/2} < \infty \right\}. $$

Let $Z = (Z(t, W) \equiv \phi(\mathcal{F}_r, W), 0 \leq t \leq \tau)$ represent a $q$-dimensional bounded random process indexed by $t$. Here, without loss of generality, we assume that $W$ is one-dimensional. Define

$$\mathcal{G} = \left\{ z(t, w) : [0, \tau] \times [-M_1, M_1] \to \mathcal{M} \right\},$$

where $\mathcal{M}$ is a bounded set on $\mathbb{R}^q$, and for function $f(x, z, t) : [-M_1, M_1]^p \times \mathcal{G} \times [0, \tau] \to \mathbb{R}$, define

$$\|f\|_2 = \left[ E \left\{ \sum_{j=1}^K |f(X, Z(T_{K,j}, W), T_{K,j})|^2\xi(T_{K,j}) \right\} \right]^{1/2}.$$

Set $M_n(g) = n^{-1}L_n(\theta) = P_{\theta}m_{\theta}(O)$, where $\theta = (\beta, \alpha, \mu), g(x, z, t) = \exp[\beta'x + \alpha'z(t, w) + \tilde{\mu}(t)]$ and

$$m_{\gamma}(O) = \sum_{j=1}^K \left[ N(T_{K,j}) - g(x, z, T_{K,j}, W), T_{K,j}) \right]^2\xi(T_{K,j}),$$

and $M(g) = Pm_\theta(O)$, where $Pf$ and $P_{\theta}f$ represent $\int f dP$ and $n^{-1}\sum_{n=1}^f(0)_{\theta}$, respectively.

Since $\mathcal{F}$ is a Hilbert space, and $\mathcal{F}_r \subset \mathcal{F}$, by the Hilbert Projection Theorem (Stakgold, 1998, p. 288), for $x_j \in \mathcal{F}$, there is a unique $a^*_j \in \mathcal{F}_r, s.t. (x_j - a^*_j) \perp \mathcal{F}_r$, for $j = 1, \ldots, p$. Let $z_j(t, w)$ be the $l$-th component of $h(\mathcal{F}_r, W)$, $l = 1, \ldots, q$. Then for $z_j(t, w) \in \mathcal{F}_r$, there is a unique $b^*_j(t) \in \mathcal{F}_r, s.t. (z_j - b^*_j) \perp \mathcal{F}_r$, for $l = 1, \ldots, q$. Let $a^* = (a^*_1, \ldots, a^*_p)'$ and $b^* = (b^*_1, \ldots, b^*_q)'$. Then we need the following condition.

C6. $E\left[ \int_0^\tau \left( \frac{X - ax'}{\phi(\mathcal{F}_r, W) - b^*(t)} \right)^{\otimes 2}d\tilde{H}(t) \right]$ is nonsingular, where $a^{\otimes 2} = d'a$ for a vector $a$. 

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim www.biometrical-journal.com
Here, Condition C1 is similar to those required by Stone (1986) and Zhou et al. (1998); Condition C2 is a common assumption in nonparametric smoothing estimation problems; Conditions C3 and C4 are mild and easily justified in many applications; Condition C5 is needed to establish the identifiability of the model; Condition C6 is a technical condition used in the proof for consistency of estimators.

Let \( \hat{\theta}_n = (\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) \) be the estimator of \( \theta_0 \). For \( \theta_1 = (\beta_1, \alpha_1, \mu_1) \), \( \theta_2 = (\beta_2, \alpha_2, \mu_2) \), define

\[
\rho(\theta_1, \theta_2) = \left\{ \| \beta_1 - \beta_2 \|^2 + \| \alpha_1 - \alpha_2 \|^2 + \| \mu_1 - \mu_2 \|^2 \right\}^{1/2}.
\]

Then the asymptotic properties of the estimators \( \hat{\theta}_n \) are summarized as follows.

**Theorem 1 (Consistency).** Under conditions C1–C4 and C6, \( \rho(\hat{\theta}_n, \theta_0) \to 0 \), almost surely.

**Theorem 2 (Rate of Convergence).** Under conditions C1–C6,

\[
\rho(\hat{\theta}_n, \theta_0) = O_p(n^{-\min\{\nu, \frac{1}{2}\}}).
\]

**Remark 1.** When \( \nu = 1/(1 + 2r) \), \( n^{-\min\{\nu, \frac{1}{2}\}} = n^{-\frac{1}{1+2r}} \), we conclude from Stone (1980, 1982) that the rate of convergence of the estimator \( \hat{\mu}_n \) is the optimal rate in nonparametric regression.

To state the asymptotic normality, we define

\[
A = E \left\{ \sum_{j=1}^{K} \xi(T_{K,j}) g_0(X, Z(T_{K,j}, W), T_{K,j}) \times \right.
\]

\[
\left. \times a^{-2}(K, T_{K,j}) c(X, Z(T_{K,j}, W), K, T_{K,j}) \right\}^2
\]

and

\[
B = E \left\{ \sum_{j=1}^{K} \sum_{j=1}^{K} \xi(T_{K,j}) \xi(T_{K,j}) b(X, Z(T_{K,j}, W), T_{K,j}) b(X, Z(T_{K,j}, W), T_{K,j}) \times \right.
\]

\[
\left. \times g_0(X, Z(T_{K,j}, W), T_{K,j}) g_0(X, Z(T_{K,j}, W), T_{K,j}) a^{-1}(K, T_{K,j}) a^{-1}(K, T_{K,j}) \times \right.
\]

\[
\left. \times c(X, Z(T_{K,j}, W), K, T_{K,j}) c(X, Z(T_{K,j}, W), K, T_{K,j}) \right\},
\]

where \( g_0(X, Z(T_{K,j}, W), T_{K,j}) = \exp \{ \beta_0 X + \alpha_0 \phi(F_{T_{K,j}}, W) + \mu_0(T_{K,j}) \} \),

\[
a(K, T_{K,j}) = E \left\{ \xi(T_{K,j}) g_0(X, Z(T_{K,j}, W), T_{K,j}) \big| K, T_{K,j} \right\},
\]

\[
b(X, Z(T_{K,j}, W), T_{K,j}) = N(T_{K,j}) - g_0(X, Z(T_{K,j}, W), T_{K,j}),
\]

\[
c(X, Z(T_{K,j}, W), K, T_{K,j}) = \left( \phi(F_{T_{K,j}}, W) \right) a(K, T_{K,j}) - E \left\{ \left( \phi(F_{T_{K,j}}, W) \right) \times \right.
\]

\[
\left. \times \xi(T_{K,j}) g_0(X, Z(T_{K,j}, W), T_{K,j}) \big| K, T_{K,j} \right\}.
\]
Theorem 3 (Asymptotic normality). Under conditions C1–C6, \( \sqrt{n} \left( \hat{\alpha}_n - \alpha_0 \right) \) converges in distribution to \( N(0, \Sigma) \) with \( \Sigma = A^{-1} B (A^{-1})' \).

Proof of Theorem 1. According to Lemma 5 in Stone (1985), for \( \hat{\mu}_0 \in \mathcal{F}_r \), there exist a \( \hat{\mu}_n \in \Phi_{l, I} \) with order \( l \geq k + 1 \) and knots \( I \) such that \( \| \hat{\mu}_n - \hat{\mu}_0 \|_\infty = O(n^{-r}) \). Let \( g_n(x, z, t) = \exp[\beta_0'x + \alpha_0' z(t, w) + \hat{\mu}_n(t)] \), \( \hat{g}_n(x, z, t) = \exp[\hat{\beta}_n'x + \hat{\alpha}_n' z(t, w) + \hat{\mu}_n(t)] \), and \( g_0(x, z, t) = \exp[\beta_0'x + \alpha_0' z(t, w) + \hat{\mu}_0(t)] \). By using the arguments similar to those used in the proof of Theorem 1 of Lu et al. (2007), we can show that \( \| \hat{g}_n - g_0 \|_2^2 = O(n^{-r} + n^{-\frac{r}{2}}) \) and hence \( \| \log \hat{g}_n - \log g_0 \|_2^2 \leq O(n^{-r} + n^{-\frac{r}{2}}) \). Also note that

\[
\| \log \hat{g}_n - \log g_0 \|_2^2 = \| (\hat{\beta}_n - \beta_0)' x + (\hat{\alpha}_n - \alpha_0)' z + (\hat{\mu}_n - \mu_0) \|_2^2 = \\
= \| (\hat{\beta}_n - \beta_0)' x - a' \|_2^2 + (\hat{\alpha}_n - \alpha_0)' z + (\hat{\mu}_n - \mu_0) \|_2^2 + \\
+ (\hat{\beta}_n - \beta_0)' a' + (\hat{\alpha}_n - \alpha_0)' b' + (\hat{\mu}_n - \mu_0) \|_2^2 = \\
= \| (\hat{\beta}_n - \beta_0)' x - a' \|_2^2 + (\hat{\alpha}_n - \alpha_0)' z - b' \|_2^2 + \\
+ (\hat{\beta}_n - \beta_0)' a' + (\hat{\alpha}_n - \alpha_0)' b' \|_2^2.
\]

By C6, we can get \( \| \hat{\beta}_n - \beta_0 \| \rightarrow 0 \), and \( \| \hat{\alpha}_n - \alpha_0 \| \rightarrow 0 \) from the first term of the right-hand side of the above equality, and thus it follows that \( \| \hat{\mu}_n - \mu_0 \| \rightarrow 0 \). Therefore, \( \| \hat{\mu}_n - \mu_0 \|_2 = \| \exp(\hat{\mu}_n) - \exp(\mu_0) \|_2 \rightarrow 0 \) by Taylor expansion. This completes the proof of Theorem 1. \( \Box \)

Proof of Theorem 2. Let \( \mathcal{F}_n = \Phi_{l, I} \) and \( \Theta_n = \mathcal{R}^0 \times \mathcal{F}_n \). It is easy to see that \( \Theta_n \subseteq \Theta_{n+1} \subseteq \cdots \subseteq \Theta_r = \mathcal{R}^0 \times \mathcal{F}_r \) for \( n \geq 1 \). Note that the sieve estimator \( \hat{\theta}_n = (\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n) \) is the minimizer of \( L_n(\hat{\theta}; \Theta) \) over the sieve space \( \Theta_n \).

For each \( n \), \( g_n(x, z, t) = \exp[\beta_0'x + \alpha_0' z(t, w) + \mu_0(t)] \) with \( \theta_n = (\beta_0, \alpha_0, \mu_0) \in \Theta_n \). Then for any \( \eta > 0 \), let

\[
\mathcal{F}_{n, \eta} = \{ g = \exp(\beta'x + \alpha' z + \mu) : \theta = (\beta, \alpha, \mu) \in \Theta_n \text{ and } \eta < \| g - g_0 \|_2 \leq \eta \}.
\]

Similar to Lemma A.2 in Huang (1999, p. 1557), for any \( \varepsilon \leq \eta \),

\[
\log N(\| \cdot \|_2 ; \mathcal{F}_{n, \eta}, \eta) \leq c_4 q_n \log(\eta/\varepsilon),
\]

for a constant \( c_4 \). Thus, for \( \varepsilon > 0 \), there exists a set of brackets \( \{ [g_{i, l}, g_{i, r}] \} \), \( i = 1, \ldots, (\varepsilon) \), such that, for each \( g \in \mathcal{F}_n \), there is a \( \theta = (g_{i, l}, g_{i, r}) \) with \( g_{i, l}(x, z, t) \leq g(x, z, t) \leq g_{i, r}(x, z, t) \), for all \( x, t \in [0, \tau] \) and \( z \in \mathcal{Z} \), and \( \| g_{i, l} - g_{i, r} \|_2 \leq \varepsilon^2 \).

Next, consider the class \( \mathcal{M}_{n, \eta} = \{ m_n(O) : g \in \mathcal{F}_{n, \eta} \} \), where

\[
m_n(O) = \sum_{j=1}^{K} N(T_{K,j}) - g(X, Z(T_{K,j}, W), T_{K,j})^2 \xi(T_{K,j}).
\]

For \( i = 1, \ldots, (\varepsilon) \), define

\[
m_i(O) = \sum_{j=1}^{K} \left( \left( \| g_{i, l} \|_2 + \| g_{i, r} \|_2 \right)^2 \right) - N(T_{K,j}) g_i(X, Z(T_{K,j}, W), T_{K,j}) +
\]

\[+ 2N(T_{K,j}) g_n(X, Z(T_{K,j}, W), T_{K,j}) - g_n^2(X, Z(T_{K,j}, W), T_{K,j}) \xi(T_{K,j}).
\]
\[ m_i^c(O) = \sum_{j=1}^{K} \left[ \|g_j\| \wedge \|g_j^c\| \right]^2 (X, Z(T_{K,j}, W), T_{K,j}) - 2N(T_{K,j})g_j'(X, Z(T_{K,j}, W), T_{K,j}) + 2N(T_{K,j})g_n(X, Z(T_{K,j}, W), T_{K,j}) - g^2_n(X, Z(T_{K,j}, W), T_{K,j}) \right] \xi(T_{K,j}), \]

where \( a \vee b = \min\{a, b\} \) and \( a \wedge b = \max\{a, b\} \). Then, it is easy to show that \( \|m_i^c(O) - m_i^1(O)\|_{P,B}^2 \leq c_5\varepsilon^2 \) with a positive constant \( c_5 \), where \( \| \cdot \|_{P,B} \) is the “Bernstein norm” defined by \( \|f\|_{P,B} = \left[ 2P(\varepsilon|f| - 1 - |f|) \right]^{1/2} \) (see van der Vaart and Wellner, 1996, p. 324). Thus \( \{m_i^1(O), m_i^1(O)\}, i = 1, \cdots, (\frac{2}{\varepsilon})^{2q_n} \) is the set of brackets for \( M_{n,q} \), which implies that

\[
\log N_{\varepsilon}(\varepsilon, M_{n,q}, \| \cdot \|_{P,B}) \leq c_4 q_n \log(\eta/\varepsilon).
\]

Moreover, by some calculations, we can verify that \( \|m_i(O) - m_i(O)\|_{P,B} \leq c_6\eta^2 \) for any \( g \in F_{n,q} \) by C3 and C4. Therefore, by Lemma 3.4.3 of van der Vaart and Wellner (1996), we obtain

\[
E\|n^{1/2}(\mathbb{P} - P)\|_{M_{n,q}} \leq c_7 f_{\|}\|\|\|_F \|_\varepsilon M_{n,q}, \| \cdot \|_{P,B} \left\{ 1 + \frac{\tilde{f}_{\|}\|\|\|_F \|_\varepsilon M_{n,q}, \| \cdot \|_{P,B}}{\eta^{2n^{1/2}}} M_3 \right\},
\]  

(A1)

where \( M_3 \) is a constant, \( \|n^{1/2}(\mathbb{P} - P)\|_F = \sup_{g \in F} |n^{1/2}(\mathbb{P} - P)g| \), and

\[
\tilde{f}_{\|}\|\|\|_F \|_\varepsilon M_{n,q}, \| \cdot \|_{P,B} = \int_0^\eta \{ 1 + \log N_{\varepsilon}(\varepsilon, M_{n,q}, \| \cdot \|_{P,B}) \}^{1/2} d\varepsilon \leq c_8 q_n^{1/2}\eta.
\]

The right-hand side of Eq. (A1) yields \( \phi_\eta(\eta) = c_9(q_n^{1/2}\eta + q_n/n^{1/2}) \). It is easy to see that \( \phi_\eta(\eta)/\eta \) is decreasing in \( \eta \), and

\[
r^2_n\phi_\eta \left( \frac{1}{r_n} \right) = r_nq_n^{1/2} + r_n^2q_n/n^{1/2} \leq 2n^{1/2},
\]

for \( r_n = \sqrt{\frac{1}{2n}} \) and \( 0 < v < 1/2 \).

Note that

\[
P[m_q(O) - m_{q_n}(O)] =
\]

\[
= P \left[ \int_0^t \left\{ [N(t) - g(X, Z(t, W), t))^2 - [N(t) - g_n(X, Z(t, W), t)]^2 \right\} \xi(t) dH(t) \right] =
\]

\[
= E \left[ \int_0^t (g_n - g)(X, Z(t, W), t)[2N(t) - (g + g_n)(X, Z(t, W), t)]\xi(t) dH(t) \right] \geq
\]

\[
\geq E \left[ \int_0^t (g - g_n)^2(X, Z(t, W), t)\xi(t) dH(t) \right] \geq
\]

\[
\geq \|g - g_n\|_{2}^2.
\]
Furthermore, in the proof of Theorem 4.1, we have that $\|\hat{g}_n - g_0\|_2$ converges to zero in probability. Thus, by Theorem 3.4.1 of van der Vaart and Wellner (1996),

$$n^{-\frac{1}{2r}}\|\hat{g}_n - g_0\|_2 = O_p(1).$$

Since $\|\hat{\mu}_n - \mu_0\|_\infty = O(n^{-vr})$, then $\|g_n - g_0\|_2 = O_p(n^{-vr})$, and so

$$\|\hat{g}_n - g_0\|_2 = O_p(n^{-vr}) + O_p(n^{-\frac{1}{2r}}) = O_p(n^{-\min\{vr, \frac{1}{2r}\}}).$$

Therefore by the similar arguments as those used in the proof of consistency of $\hat{\beta}_n$, $\hat{\alpha}_n$, and $\hat{\mu}_n$, we can get the rate of convergence of $\hat{\mu}_n$, $\hat{\beta}_n$, and $\hat{\alpha}_n$ as stated in Theorem 4.2. The choice of $v = 1/(1 + 2r)$ yields the rate of convergence of $r/(1 + 2r)$, which completes the proof.

**Proof of Theorem 3.** Let $\mathcal{H} \equiv \{(h_1, h_2, h_3) : (h_1', h_2') \in \mathcal{R}, h_3 \in \mathcal{F}_r, \|h_1\| \leq 1, \|h_2\| \leq 1, \|h_3\|_\infty \leq 1\}$. We define a sequence of maps $S_n$ mapping a neighborhood of $\hat{\theta}_0 = (\beta_0, \alpha_0, \mu_0)$, denoted by $\mathcal{U}$, in the parameter space for $\hat{\theta} = (\beta, \alpha, \mu)$ into $l^\infty(\mathcal{H})$ as

$$S_n(\hat{\theta})[h_1, h_2, h_3] = n^{-1} \frac{d}{d\epsilon} L_n(\beta + \epsilon h_1, \alpha + \epsilon h_2, \mu + \epsilon h_3) \bigg|_{\epsilon = 0} =$$

$$= -2 n^{-1} \sum_{i=1}^{n} \int^{\tau}_{0} [N_i(t) - \exp \{\beta' H_i + \alpha' \phi(\mathcal{F}_r, W_i) + \mu(t)\}] \times$$

$$\times \exp \{\beta' H_i + \alpha' \phi(\mathcal{F}_r, W_i) + \mu(t)\} \times$$

$$\times [h_1 X_i + h_2 e(\mathcal{F}_r, W_i) + h_3(t)] d\tilde{H}_i(t) \equiv$$

$$P_n \psi(\hat{\theta}; O)[h_1, h_2, h_3],$$

where

$$\psi(\hat{\theta}; O)[h_1, h_2, h_3] = -2 \int^{\tau}_{0} [N(t) - \exp \{\beta' X + \alpha' \phi(\mathcal{F}_r, W) + \mu(t)\}] \times$$

$$\times \exp \{\beta' X + \alpha' \phi(\mathcal{F}_r, W) + \mu(t)\} \times$$

$$\times [h_1 X + h_2 e(\mathcal{F}_r, W) + h_3(t)] d\tilde{H}(t).$$

Correspondingly, we define the limit map $S : \mathcal{U} \longrightarrow l^\infty(\mathcal{H})$ as

$$S(\hat{\theta})[h_1, h_2, h_3] = P\psi(\hat{\theta}; O)[h_1, h_2, h_3],$$

where $l^\infty(\mathcal{H})$ is the space of bounded functionals on $\mathcal{H}$ under the supremum norm $\|f\|_\infty = \sup_{h \in \mathcal{H}} |f(h)|$.

To derive the asymptotic normality of the estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\alpha}_n, \hat{\mu}_n)$, motivated by the proof of Theorem 3.3.1 of van der Vaart and Wellner (1996), we first need to verify the following four conditions.

(i) $\sqrt{n} (S_n - S(\hat{\theta}_n)) - \sqrt{n} (S_n - S(\hat{\theta}_0)) = o_p(1)$.

(ii) $\sqrt{n} (S_n - S(\hat{\theta}_0))$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H})$.

(iii) $S(\hat{\theta}_0) = 0$ and $S_n(\hat{\theta}_n) = o_p(n^{-1/2})$. 

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim www.biometrical-journal.com
(iv) \( \hat{\theta} \mapsto S(\hat{\theta}) \) is Fréchet-differentiable at \( \hat{\theta}_0 \) with a continuous derivative \( \dot{S}(\hat{\theta}_0) \) and

\[
\sqrt{n} \left( S(\hat{\theta}_n) - S(\hat{\theta}_0) \right) - \sqrt{n} \dot{S}(\hat{\theta}_0) \left( \hat{\theta}_n - \hat{\theta}_0 \right) = o_p(1).
\]

Note that

\[
\sqrt{n}(S_n - S)(\hat{\theta}_n)[h_1, h_2, h_3] - \sqrt{n}(S_n - S)(\hat{\theta}_0)[h_1, h_2, h_3] =
\]

\[
= \sqrt{n}(P_n - P) \left( \psi(\hat{\theta}_n; O)[h_1, h_2, h_3] - \psi(\hat{\theta}_0; O)[h_1, h_2, h_3] \right).
\]

For some \( \delta > 0 \), define

\[
\mathcal{F}_\delta = \left\{ \psi(\tilde{\theta}; O)[h_1, h_2, h_3] - \psi(\tilde{\theta}_0; O)[h_1, h_2, h_3]: \rho(\tilde{\theta}, \tilde{\theta}_0) < \delta, (h_1, h_2, h_3) \in \mathcal{H} \right\}.
\]

It is easy to see that \( \mathcal{H} \) is a Donsker class base on the argument in pages 154–157 of van der Vaart and Wellner (1996). Since

\[
\left| \psi(\tilde{\theta}; O)[h_1, h_2, h_3] \right| =
\]

\[
= \left| -2 \int_0^1 [N(t) - \exp(\beta'X + \alpha'F(t, W) + \tilde{\mu}(t))] \exp(\beta'X + \alpha'F(t, W) + \tilde{\mu}(t)) \times
\]

\[
\times [h_1'X + h_2'\phi(F(t, W) + h_3(t)]d\tilde{H}(t) \right| \leq
\]

\[
\leq M_1 \|h_1\| + M_2 \|h_2\| + M_3 \|h_3\|_\infty,
\]

for constants \( M_1, M_2, M_3 \), which means that \( \psi(\tilde{\theta}; O) \) is a bounded Lipschitz functional with respect to \( \mathcal{H} \); thus \( \mathcal{F}_\delta \) is a Donsker class for \( \delta > 0 \). For any \( \tilde{\theta}_1 = (\tilde{\beta}_1, \tilde{\alpha}_1, \tilde{\mu}_1) \) and \( \tilde{\theta}_2 = (\tilde{\beta}_2, \tilde{\alpha}_2, \tilde{\mu}_2) \), it is easy to verify that

\[
P \left[ \left| \psi(\tilde{\theta}_1; O)[h_1, h_2, h_3] - \psi(\tilde{\theta}_2; O)[h_1, h_2, h_3] \right|^2 \leq c_{10} \rho^2(\tilde{\theta}_1, \tilde{\theta}_2), \right.
\]

for a constant \( c_{10} \). Thus condition (i) holds by Lemma 13.3 of Kosorok (2008).

Condition (ii) is also satisfied since \( \left| \psi(\tilde{\theta}_1; O)[h_1, h_2, h_3] : (h_1, h_2, h_3) \in \mathcal{H} \right| \) is a Donsker class.

Clearly, \( S(\hat{\theta}_0) = 0 \). For \( h_3 \in \mathcal{F}_\delta \), let \( h_{3n} \) be the B-spline function approximation of \( h_3 \) with \( \|h_{3n} - h_3\|_\infty = O(n^{-r}) \), then we have \( S_n(\hat{\theta}_n)[h_1, h_2, h_{3n}] = 0 \).

Thus, for \( (h_1, h_2, h_{3n}) \in \mathcal{H} \),

\[
n^2 S_n(\hat{\theta}_n)[h_1, h_2, h_{3n}] =
\]

\[
= n^2 \left[ S_n(\hat{\theta}_n)[h_1, h_2, h_{3n}] - S_n(\hat{\theta}_n)[h_1, h_{3n}] \right] =
\]

\[
= n^2 \left( P_n - P \right) \psi(\hat{\theta}_n; O)[h_1, h_2, h_{3n}] - n^2 \left( P_n - P \right) \psi(\hat{\theta}_0; O)[h_1, h_2, h_{3n}] +
\]

\[
- \left\{ n^2 \left( P_n - P \right) \psi(\hat{\theta}_0; O)[h_1, h_{3n}] - n^2 \left( P_n - P \right) \psi(\hat{\theta}_0; O)[h_1, h_{3n}] \right\} +
\]

\[
+ n^2 P \left[ \psi(\hat{\theta}_0; O)[h_1, h_2, h_{3n}] - \psi(\hat{\theta}_0; O)[h_1, h_2, h_{3n}] \right]
\]

\[
= Q_{1n} - Q_{2n} + Q_{3n} + Q_{4n}.
\]
It follows from condition (i) that both $Q_{1n}$ and $Q_{2n}$ are $o_p(1)$. Note that
\[
|Q_{1n}| = \left| 2n^{\frac{1}{2}} \int_0^T \left[ \exp \{ \hat{\beta}'_n X + \hat{\alpha}'_n \phi(F_t, W) + \hat{\mu}_n(t) \} - N(t) \right] \times \exp \{ \hat{\beta}'_n X + \hat{\alpha}'_n \phi(F_t, W) + \hat{\mu}_n(t) \} (h_3 - h_{3n})(t) d\tilde{H}(t) \right| \leq
\]
\[
\leq c_{11} n^{\frac{1}{2}} P \int_0^T \left[ \exp \{ \hat{\beta}'_n X + \hat{\alpha}'_n \phi(F_t, W) + \hat{\mu}_n(t) \} \times \exp \{ \beta'_0 X + \alpha'_0 \phi(F_t, W) + \mu_0(t) \} (h_3 - h_{3n})(t) d\tilde{H}(t) \right| \leq
\]
\[
= c_{11} n^{\frac{1}{2}} P \int_0^T \exp \{ f'_n - f_0 \} (h_3 - h_{3n})(t) d\tilde{H}(t) \leq
\]
\[
\leq c_{12} n^{\frac{1}{2}} \rho (\hat{\beta}_n, \hat{\theta}_0) \| h_{3n} - h_3 \|_\infty \leq n^{\frac{1}{2}} O(n^{-1/2}) O(n^{-1/2}) = o_p(1)
\]
for constants $c_{11}$ and $c_{12}$, where $f'_0 = \beta'_0 X + \alpha'_0 \phi(F_t, W) + \mu_0$, $\hat{f}_n = \hat{\beta}'_n X + \hat{\alpha}'_n \phi(F_t, W) + \hat{\mu}_n$, and $f_n^* = (1 - \xi) f_0 + \xi \hat{f}_n$ with $0 \leq \xi \leq 1$. Furthermore, $Q_{3n}$ is also $o_p(1)$ since
\[
P \left[ \psi (\hat{\theta}_n; O) | h_1, h_2, h_3 \right] - \psi (\hat{\theta}_0; O) | h_1, h_2, h_3 \right] \right|\right]^2 =
\[
= P \left[ \int_0^T \left[ N(t) \times \exp \{ \beta'_0 X + \alpha'_0 \phi(F_t, W) + \mu_0(t) \} (h_3 - h_{3n})(t) d\tilde{H}(t) \right] \right]^2 \leq
\]
\[
\leq c_{13} \| h_{3n} - h_3 \|_\infty^2 \rightarrow 0, \quad n \rightarrow \infty,
\]
for a constant $c_{13}$. Thus $S_{n, \hat{\beta}_n \hat{\theta}_n}(\hat{\theta}_n) = o_p(n^{-1/2})$.

For the proof of condition (iv), if the smoothness of $S(\hat{\theta})$, the Fréchet differentiability holds and the derivative of $S(\hat{\theta})$ at $(\hat{\theta}_0)$, denoted by $\hat{S}(\hat{\theta}_0)$, is a map from the space $\{ (\hat{\theta} - \hat{\theta}_0) : \hat{\theta} \in \mathcal{U} \}$ to $L^\infty(\mathcal{H})$ and
\[
\hat{S}(\hat{\theta}_0)(\hat{\theta} - \hat{\theta}_0) | h_1, h_2, h_3 =
\[
= \frac{dS(\beta_0 + \epsilon (\beta - \beta_0), \alpha_0 + \epsilon (\alpha - \alpha_0), \mu_0 + \epsilon (\bar{\mu} - \bar{\mu}_0)) | h_1, h_2, h_3}{d\epsilon}_{\epsilon=0} =
\]
\[
= 2P \int_0^T \left[ \left[ \beta'_0 X + \alpha'_0 \phi(F_t, W) + \mu_0(t) \right] - \left[ \beta_0 X + \alpha_0 \phi(F_t, W) + \mu_0(t) \right] \times \left[ \left[ \beta'_0 X + \alpha'_0 \phi(F_t, W) + \mu_0(t) \right] - \left[ \beta_0 X + \alpha_0 \phi(F_t, W) + \mu_0(t) \right] \times \right. \times \left. [h_1 X + h_2 h(F_t, W) + h_3(t)] d\tilde{H}(t) \right] \right.
\]
\[
\times \left[ \left[ \beta'_0 X + \alpha'_0 \phi(F_t, W) + \mu_0(t) \right] - \left[ \beta_0 X + \alpha_0 \phi(F_t, W) + \mu_0(t) \right] \times \right. \times \left. [h_1 X + h_2 h(F_t, W) + h_3(t)] d\tilde{H}(t) \right] \equiv
\]
\[
= \sigma_1 (h_1, h_2, h_3)' (\beta - \beta_0) + \sigma_2 (h_1, h_2, h_3)' (\alpha - \alpha_0) + \int_0^T (\bar{\mu} - \bar{\mu}_0)(t) d\sigma_3 (h_1, h_2, h_3),
\]
where
\[
\sigma_1(h_1, h_2, h_3) = 2P\left\{ \int_0^T g_0^2(X, Z(t, W), t)[h'_1X + h'_2\phi(F_t, W) + h_3(t)]Xd\tilde{H}(t) \right\},
\]
\[
\sigma_2(h_1, h_2, h_3) = 2P\left\{ \int_0^T g_0^2(X, Z(t, W), t)[h'_1X + h'_2\phi(F_t, W) + h_3(t)]\phi(F_t, W)d\tilde{H}(t) \right\},
\]
and
\[
\sigma_3(h_1, h_2, h_3)(t) = 2P\left\{ \int_0^T g_0^2(X, Z(s, W), s)[h'_1X + h'_2\phi(F_s, W) + h_3(s)]d\tilde{H}(s) \right\},
\]
with \( g_0(X, Z(t, W), t) = \exp \{ \beta_0X + \alpha_0\phi(F_t, W) + \mu_0(t) \} \).

Thus, condition (iv) follows from
\[
|\sqrt{n}[S(\hat{\theta}_n) - S(\theta_0) - S(\hat{\theta}_n - \hat{\theta}_0)][h_1, h_2, h_3]|
\]
\[
= 2\sqrt{n}P\int_0^T \left\{ (\exp(\hat{f}_n) - \exp(f_0))\exp(\hat{f}_n) - \exp(2f_0)(\hat{f}_n - f_0) \right\} \times \]
\[
\times [h'_1X + h'_2\phi(F_t, W) + h_3(t)]d\tilde{H}(t)
\]
\[
= 2\sqrt{n}P\int_0^T \left\{ \left[ \exp(f_0)(\hat{f}_n - f_0) + \frac{\exp(f_0)}{2}(\hat{f}_n - f_0)^2 + o_p(\|\hat{f}_n - f_0\|^2) \right]\exp(\hat{f}_n) - \exp(2f_0)(\hat{f}_n - f_0) \right\} \times \]
\[
\times [h'_1X + h'_2\phi(F_t, W) + h_3(t)]d\tilde{H}(t)
\]
\[
\leq c_{14}\sqrt{n}\left[ \rho^2(\hat{\theta}_n, \theta_0) + o_p\left( \rho^2(\hat{\theta}_n, \theta_0) \right) \right] = \]
\[
= O_p(\nu^{\frac{1}{2}} - (1 - \nu)) + o_p(1) = o_p(1),
\]
for a constant \( c_{14} \). Therefore, by conditions (i) to (iv), we have
\[
\sqrt{n}S(\hat{\theta}_n)(\hat{\theta}_n - \theta_0)[h_1, h_2, h_3] =
\]
\[
= \sigma_1(h_1, h_2, h_3) + \sigma_2(h_1, h_2, h_3) + \sigma_3(h_1, h_2, h_3) + \int_0^\tau \sqrt{n}(\hat{\mu}_n - \mu_0)(t)d\sigma_3(h_1, h_2, h_3)(t) =
\]
\[
= -\sqrt{n}(S_n - S)(\hat{\theta}_0)[h_1, h_2, h_3] + o_p(1). \quad (A2)
\]
Next, we will derive the asymptotic normality of estimators from Eq. (A2). Rewrite

$$\sigma_3(h_1, h_2, h_3)(t) =$$

$$= 2E\left\{ \sum_{j=1}^{K} \xi(T_{K,j})I(T_{K,j} \leq t)g_0^2(X, Z(T_{K,j}, W), T_{K,j})[h'_jX + h'_2\phi(F_{T_{K,j}}, W) + h_3(T_{K,j})] \right\} =$$

$$= 2E\left\{ \sum_{j=1}^{K} I(T_{K,j} \leq t) \left[ E\left\{ \xi(T_{K,j})g_0^2(X, Z(T_{K,j}, W), T_{K,j}) \right| K, T_{K,j} \right] h_3(T_{K,j}) +
+ E\left\{ \xi(T_{K,j})g_0^2(X, Z(T_{K,j}, W), T_{K,j})[h'_jX + h'_2\phi(F_{T_{K,j}}, W)] \right| K, T_{K,j} \right\} \right\}. $$

Thus, we can take

$$h_3(T_{K,j}) = -a^{-1}(K, T_{K,j})E\left\{ \xi(T_{K,j})g_0^2(X, Z(T_{K,j}, W), T_{K,j})[h'_jX + h'_2\phi(F_{T_{K,j}}, W)] \right| K, T_{K,j} \right\} =$$

$$= -\left( \begin{pmatrix} h_1 \\
\h_2 \end{pmatrix} \right)' a^{-1}(K, T_{K,j})E\left\{ g_0^2(X, Z(T_{K,j}, W), T_{K,j}) \xi(T_{K,j}) \left( \frac{X}{\phi(F_{T_{K,j}}, W)} \right) \right| K, T_{K,j} \right\}$$

such that $\sigma_3(h_1, h_2, h_3)(t) \equiv 0$. Furthermore, for this $h_3$, we have

$$\left( \sigma_1(h_1, h_2, h_3)\sigma_2(h_1, h_2, h_3) \right) = 2A \left( \begin{pmatrix} h_1 \\
\h_2 \end{pmatrix} \right).$$

Thus, it follows from Eq. (A1) that

$$\left( \begin{pmatrix} \sigma_1(h_1, h_2, h_3) \\
\sigma_2(h_1, h_2, h_3) \end{pmatrix} \right)' \sqrt{n} \left( \begin{pmatrix} \hat{\beta}_n - \beta_0 \\
\hat{\alpha}_n - \alpha_0 \end{pmatrix} \right) = -\sqrt{n}(P_n - P)\psi(\theta_0; \Omega)[h_1, h_2, h_3] + o_p(1).$$

Thus, $2\left( \begin{pmatrix} h_1 \\
\h_2 \end{pmatrix} \right)' A \sqrt{n} \left( \begin{pmatrix} \hat{\beta}_n & \hat{\alpha}_n \end{pmatrix} - \begin{pmatrix} \beta_0 & \alpha_0 \end{pmatrix} \right) \rightarrow N(0, \sigma^2)$ with

$$\sigma^2 = [\psi(\theta_0; \Omega)[h_1, h_2, h_3]]^2 = 4 \left( \begin{pmatrix} h_1 \\
\h_2 \end{pmatrix} \right)' B \left( \begin{pmatrix} h_1 \\
\h_2 \end{pmatrix} \right).$$

Therefore,

$$\sqrt{n} \left( \begin{pmatrix} \hat{\beta}_n - \beta_0 \\
\hat{\alpha}_n - \alpha_0 \end{pmatrix} \right) \rightarrow N(0, A^{-1} B(A^{-1})').$$

This completes the proof of Theorem 3. $\square$

References