Biometrika (2007), **94**, 3, *pp*. 705–718
doi:10.1093/biomet/asm057

Advance Access publication 5 August 2007
Printed in Great Britain

Estimation of the mean function with panel count data using monotone polynomial splines

BY MINGGEN LU, YING ZHANG
Department of Biostatistics, The University of Iowa,
200 Hawkins Drive, C22 GH,
Iowa City, Iowa 52242, U.S.A.
minggen-lu@uiowa.edu ying-j-zhang@uiowa.edu

AND JIAN HUANG
Department of Statistics and Actuarial Science, The University of Iowa,
221 Schaeffer Hall, Iowa City, Iowa 52242, U.S.A.
jian-huang@uiowa.edu

SUMMARY

We study nonparametric likelihood-based estimators of the mean function of counting processes with panel count data using monotone polynomial splines. The generalized Rosen algorithm, proposed by Zhang & Jamshidian (2004), is used to compute the estimators. We show that the proposed spline likelihood-based estimators are consistent and that their rate of convergence can be faster than $n^{1/3}$. Simulation studies with moderate samples show that the estimators have smaller variances and mean squared errors than their alternatives proposed by Wellner & Zhang (2000). A real example from a bladder tumour clinical trial is used to illustrate this method.

Some key words: Counting process; Empirical process; Isotonic regression; Maximum likelihood estimator; Maximum pseudolikelihood estimator; Monotone polynomial spline; Monte Carlo.

1. INTRODUCTION

This article considers the estimation of the mean function of counting processes with panel count data using monotone polynomial splines. In many long-term clinical trials or epidemiological studies, the subjects are observed at several time-points during the study. The only available information is the number of recurrent events occurring before each observation time, the exact event times themselves being unknown. The number of observations and the observation times may vary from individual to individual. Such data are referred to as panel count data.

Examples of panel count data are the National Cooperative Gallstone Study Thall & Lachin (1988) and the bladder tumour randomized clinical trial conducted by the Veterans Administration Cooperative Urological Research Group. In the latter, all patients had superficial bladder tumours when they entered the trial, and they were randomly assigned to one of three arms: placebo, pyridoxine or thiotepa. Many patients had multiple recurrences of the tumour, and new tumours were removed at each visit. The goal of the study was to determine the effect of treatment on the frequency of tumour recurrence; see for example Byar et al. (1977), Byar (1980), Wei et al. (1989), Sun & Wei (2000), Wellner & Zhang (2000) and Zhang (2002).

Several papers have considered methods for analyzing panel count data (Kalbfleisch & Lawless (1985); Thall & Lachin (1988); Thall (1988); Lee & Kim (1998); Sun & Kalbfleisch (1995) appear to be the first to study nonparametric estimation of the mean function with panel count data. Their method was based on isotonic regression by pulling the observations together and taking into account the monotonicity of the mean function. Wellner & Zhang (2000) studied nonparametric maximum pseudolikelihood and nonparametric maximum likelihood estimators based on a 'working model' of a nonhomogeneous Poisson process. They showed that the maximum pseudolikelihood estimator was exactly the one proposed by Sun & Kalbfleisch (1995). They also studied the asymptotic properties of the maximum pseudolikelihood and maximum likelihood estimators. The maximum likelihood estimator is more efficient than the maximum pseudolikelihood estimator, but its computation is more difficult.

Many investigators have studied spline estimation of a hazard function or a survival function, but direct modelling of the hazard function using splines may not guarantee the nonnegativity of the estimated function (Anderson & Senthilselvan, 1980; Whittemore & Keller, 1986). To overcome this drawback, Rosenberg (1995) used the spline representation with coefficients expressed in exponential form; Kooperberg et al. (1995) and Cai & Betensky (2003) modeled the log hazard function using linear splines, thereby guaranteeing the nonnegativity of the hazard function.

In our application, the pseudolikelihood and likelihood functions are functions of the cumulative mean function of event numbers. Although we can reparameterize the likelihood functions in terms of the intensity function and use a spline method, such as the one proposed by Kooperberg et al. (1995), there will be unnecessary computational complications in the estimation procedure, especially when the estimation of the mean function is of primary interest. In this article, the monotone cubic I-splines (Ramsay, 1988) are applied to approximate directly the true mean function $\Lambda_0(t)$ of the counting process by

$$\Lambda(t) = \sum_{j=1}^{q_n} \alpha_j I_j(t)$$

subject to $\alpha_j \geqslant 0$, for $j = 1, \ldots, q_n$. The monotonicity of Λ is guaranteed by the nonnegativity constraints on the α_j, $j = 1, \ldots, q_n$.

We express the pseudolikelihood and likelihood functions given in Wellner & Zhang (2000) using the I-spline functions and estimate the spline coefficients using the generalized Rosen algorithm proposed by Zhang & Jamshidian (2004). Our approach has two attractive features: it is much less demanding to compute the spline likelihood estimator than to compute the nonparametric maximum likelihood estimator described by Wellner & Zhang (2000) based on the iterative convex minorant algorithm proposed by Jongbloed (1998) and the spline estimators have a higher convergence rate than their alternatives proposed by Wellner & Zhang (2000) under reasonable smoothness assumptions.

2. MONOTONE SPLINE ESTIMATORS OF THE MEAN FUNCTION

Let $\{\mathcal{N}(t) : t \geqslant 0\}$ be a counting process with mean function $\mathcal{E}\mathcal{N}(t) = \Lambda_0(t)$. The total number of observations K on the counting process is an integer-valued random variable and $T = (T_{K,1}, \ldots, T_{K,K})$ is a sequence of random observation times with $0 < T_{K,1} < \cdots < T_{K,K}$. The cumulative numbers of recurrent events up to these times, $\mathcal{N} = \{\mathcal{N}(T_{K,1}), \ldots, \mathcal{N}(T_{K,K})\}$ with $0 \leqslant \mathcal{N}(T_{K,1}) \leqslant \cdots \leqslant \mathcal{N}(T_{K,K})$, are observed accordingly. The panel count data of the counting process consist of $X = (K, T, \mathcal{N})$. We assume that the number of observations and the sequence of observation times are independent of the underlying process; that is, (K, T) and $\mathcal{N}$ are independent.

Suppose that we observe n independent and identically distributed copies of X, $X_i = (K_i, T_i, \mathcal{N}^{(i)})$ with $T_i = (T^{(i)}_{K_i,1}, \ldots, T^{(i)}_{K_i,K_i})$ and $\mathcal{N}^{(i)} = \{\mathcal{N}^{(i)}(T^{(i)}_{K_i,1}), \ldots, \mathcal{N}^{(i)}(T^{(i)}_{K_i,K_i})\}$, for $i = 1, \ldots, n$. We denote the observed data by $D = (X_1, X_2, \ldots, X_n)$. Wellner & Zhang (2000) proposed two nonparametric estimation methods for the mean function of the counting process. Assuming the underlying counting process to be a nonhomogeneous Poisson process and ignoring the correlations among the cumulative counts, they formulated the log-pseudolikelihood by omitting the parts irrelevant to the mean function Λ:

$$l^{\mathrm{ps}}_n(\Lambda | D) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left\{ \mathcal{N}^{(i)}(T^{(i)}_{K_i,j}) \log \Lambda(T^{(i)}_{K_i,j}) - \Lambda(T^{(i)}_{K_i,j}) \right\}. \tag{1}$$

They also established the log-likelihood for Λ using independence of the increments of $\mathcal{N}(t)$:

$$l_n(\Lambda | D) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left\{ \Delta \mathcal{N}^{(i)}_{K_i,j} \log(\Delta \Lambda_j) \right\} - \sum_{i=1}^{n} \Lambda(T^{(i)}_{K_i,K_i}), \tag{2}$$

where $\Delta \mathcal{N}^{(i)}_{K_i,j} = \mathcal{N}^{(i)}(T^{(i)}_{K_i,j}) - \mathcal{N}^{(i)}(T^{(i)}_{K_i,j-1})$ and $\Delta \Lambda_j = \Lambda(T^{(i)}_{K_i,j}) - \Lambda(T^{(i)}_{K_i,j-1})$, for $j = 1, \ldots, K_i$ and $i = 1, \ldots, n$.

As described in Wellner & Zhang (2000), the nonparametric maximum pseudolikelihood estimator can be computed in one step via the max-min formula, and the computation of the maximum likelihood estimator involves the iterative convex minorant algorithm, which can be computationally demanding when the sample size is large.

For a finite closed interval $[a, b]$, let $\mathcal{I} = \{t_i\}_1^{m_n+2l}$, with

$$a = t_1 = \ldots = t_l < t_{l+1} < \ldots < t_{m_n+l} < t_{m_n+l+1} = \ldots = t_{m_n+2l} = b,$$

be a sequence of knots that partition $[a, b]$ into $m_n + 1$ subintervals $J_i = [t_{l+i}, t_{l+1+i})$, for $i = 0, \ldots, m_n$. Denote by $\varphi_{l,t}$ the class of polynomial splines of order $l \geqslant 1$ with the knot sequence $\mathcal{I}$. For each $s \in \varphi_{l,t}$, s is a polynomial of order l in J_i for $0 \leqslant i \leqslant m_n$, and s is l' times continuously differentiable on $[a, b]$, for $l \geqslant 2$ and $0 \leqslant l' \leqslant l - 2$. A spline for $l = 4$ is a piecewise-cubic polynomial with continuous second-order derivative. As a special case, the spline with $l = 1$ is a step function which is discontinuous at each knot.

In fact, the class $\varphi_{l,t}$ is linearly spanned by the B-spline basis functions $\{B_i, 1 \leqslant i \leqslant q_n\}$; that is, for any $\phi(t) \in \varphi_{l,t}$, there exist $\alpha_1, \ldots, \alpha_{q_n}$ such that $\phi(t) = \sum_{i=1}^{q_n} \alpha_i B_i(t)$, where $q_n = m_n + l$ (Schumaker, 1981). We now define a subclass of $\phi_{l,t}$,

$$\psi_{l,t} = \left\{ \sum_{i=1}^{q_n} \beta_i B_i : 0 \leqslant \beta_1 \leqslant \ldots \leqslant \beta_{q_n} \right\}.$$

According to Theorem 5·9 of Schumaker (1981), $\psi_{l,t}$ is the class of nonnegative and monotone nondecreasing splines on $[a, b]$. The I-splines constructed in Ramsay (1988) are closely related to the B-splines. In fact, an I-spline basis function is a summation of a series of B-spline basis functions (Ramsay, 1988, p. 428). Hence the class $\psi_{l,t}$ can be also linearly spanned by the I-spline basis functions, i.e.

$$\psi_{l,t} = \left\{ \sum_{i=1}^{q_n} \alpha_i I_i : \ \alpha_i \geqslant 0, \ i = 1, \ldots, q_n \right\}.$$

The nonnegativity and monotonicity of the I-splines are guaranteed by the nonnegativity of coefficients (Ramsay, 1988, p. 428). We approximate the smooth monotone mean function $\Lambda_0(t)$ by $\sum_{i=1}^{q_n} \alpha_i I_i(t)$ and estimate the coefficients $\alpha = (\alpha_1, \ldots, \alpha_{q_n})$ through maximizing the approximated pseudolikelihood and likelihood functions subject to nonnegativity constraints.

Let $\hat{\alpha}_p^{\text{ps}}$ for $p = 1, \ldots, q_n$ be the spline coefficients that maximize

$$l_n^{\text{ps}}(\alpha | D) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left[\mathcal{N}^{(i)}(T_{K_i,j}^{(i)}) \log \left\{ \sum_{p=1}^{q_n} \alpha_p I_p(T_{K_i,j}^{(i)}) \right\} - \sum_{p=1}^{q_n} \alpha_p I_p(T_{K_i,j}^{(i)}) \right], \tag{3}$$

subject to $\alpha_j \geqslant 0$, $j = 1, \ldots, q_n$. Similarly, let $\hat{\alpha}_p$, for $p = 1, \ldots, q_n$, be the spline coefficients that maximize

$$l_n(\alpha | D) = \sum_{i=1}^{n} \sum_{j=1}^{K_i} \left[\Delta \mathcal{N}_{K_i,j}^{(i)} \log \left\{ \sum_{p=1}^{q_n} \alpha_p \Delta I_p(T_{K_i,j}^{(i)}) \right\} \right] - \sum_{i=1}^{n} \sum_{p=1}^{q_n} \alpha_p I_p(T_{K_i,K_i}^{(i)}), \tag{4}$$

with $\Delta I_p(T_{K_i,j}^{(i)}) = I_p(T_{K_i,j}^{(i)}) - I_p(T_{K_i,j-1}^{(i)})$, for $p = 1, \ldots, q_n$, subject to the same constraints as above. We denote the spline pseudolikelihood estimator by $\hat{\Lambda}_n^{\text{ps}}(t) = \sum_{j=1}^{q_n} \hat{\alpha}_j^{\text{ps}} I_j(t)$ and the spline likelihood estimator by $\hat{\Lambda}_n(t) = \sum_{j=1}^{q_n} \hat{\alpha}_j I_j(t)$, respectively.

We use cubic splines to approximate the mean function $\Lambda_0(t)$. In fact, the two estimators proposed by Wellner & Zhang (2000) can be viewed as the special splines of $l = 1$ that use all the distinct observation time-points as knots. The estimators constructed here can therefore be treated as extensions of those proposed by Wellner & Zhang (2000) with respect to the smoothness of estimators. However, the number of coefficients to be estimated is remarkably reduced. As a result, the spline estimators are expected to be less computationally demanding.

Both the spline pseudolikelihood and spline likelihood functions are concave with respect to the unknown coefficients, so that the spline estimation problem is equivalent to a nonlinear convex programming problem subject to linear inequality constraints: the spline estimation problems (3) and (4) can be formulated as the constrained maximization problem

$$\max_{\alpha \in \Theta_\alpha} l(\alpha | X), \tag{5}$$

where $\Theta_\alpha = \{\alpha : \alpha_j \geqslant 0, \ j = 1, \ldots, q_n\}$. Rosen (1960) proposed a generalized gradient method for optimizing an objective function with linear constraints, based on the Euclidean metric. Jamshidian (2004) developed a general algorithm based on the generalized Euclidean metric $\|x\| = x^{\text{T}} W x$, where W is a positive definite matrix and can vary from iteration to iteration. Zhang & Jamshidian (2004) used this algorithm to compute the nonparametric maximum likelihood estimator of a failure function with various types of censored data. They used $W = -D_H$, the diagonal elements of the negative Hessian matrix H, to avoid the possible storage problem in updating H for a large-scale nonparametric maximum likelihood estimation problem.

We shall use the full negative Hessian matrix H because the dimension of the unknown parameter space is usually small in our applications. This substantially reduces the number of iterations. A detailed description of the computational method and the algorithm coded in R can be obtained from the first author.

3. Asymptotic results

We study the asymptotic properties of the spline pseudolikelihood and spline likelihood estimators with the same L_2 metric d as defined in Wellner & Zhang (2000); that is,

$$d(\Lambda_1, \Lambda_2) = \|\Lambda_1 - \Lambda_2\|_2 = \left\{ \int |\Lambda_1(t) - \Lambda_2(t)|^2 d\mu(t) \right\}^{1/2},$$

where

$$\mu(t) = \sum_{k=1}^{\infty} \text{pr}(K = k) \sum_{j=1}^{k} \text{pr}(T_{K,j} \leqslant t | K = k),$$

for any $\Lambda_1, \Lambda_2 \in \mathcal{F}$ with $\mathcal{F} = \{\Lambda : \Lambda \text{ is monotone nondecreasing, } \Lambda(0) = 0\}$.

To study the asymptotic properties of the spline estimators, we need to allocate the knots properly. Let

$$\sigma = t_1 = \ldots t_l < t_{l+1} < \ldots < t_{m_n+l} < t_{m_n+l+1} = \ldots = t_{m_n+2l} = \tau$$

be a sequence of knots with $m_n = O(n^\nu)$ for $0 < \nu < 1/2$. We assume the following conditions.

Condition 1. The maximum spacing of the knots satisfies

$$\Delta = \max_{l+1 \leqslant i \leqslant m_n+l+1} |t_i - t_{i-1}| = O(n^{-\nu}).$$

Moreover, there exists a constant $M > 0$ such that $\Delta/\delta \leqslant M$ uniformly in n, where $\delta = \min_{l+1 \leqslant i \leqslant m_n+l+1} |t_i - t_{i-1}|$.

Condition 2. For some interval $O[T] = [\sigma, \tau]$ with $\sigma > 0$ and $\Lambda_0(\sigma) > 0$, $\text{pr}(\cap_{j=1}^{K} \{T_{K,j} \in [\sigma, \tau]\}) = 1$.

Condition 3. There exists a positive integer M_0 such that $\text{pr}(K \leqslant M_0) = 1$; that is, the number of observations is finite.

Condition 4. The true mean function Λ_0 has a bounded rth derivative in $O[T]$ with $r > 1$. Moreover, the first derivative has a positive lower bound in $O[T]$; that is, there exists a constant $C_0 > 0$ such that $\Lambda_0'(t) \geqslant C_0$ for $t \in O[T]$.

Condition 5. The expectation $E[\exp\{C\mathcal{N}(t)\}]$ is uniformly bounded for $t \in S[T] = \{t : 0 < t < \tau\}$ for some $\tau > 0$. The τ can be viewed as the termination time in a clinical follow-up study.

Condition 6. The observation time-points are γ-separated; that is, there exists a $\gamma > 0$ such that $\text{pr}(T_{K,j} - T_{K,j-1} \geqslant \gamma) = 1$ for all $j = 1, \ldots, K$.

The conditions related to the observation schemes, Conditions 2, 3, and 6, are mild and are easily justified for clinical trials data. Condition 5 is true if the underlying counting process is uniformly bounded or if it is a Poisson or mixed Poisson process. The smoothness assumption of the true mean function Λ_0, Condition 4, is standard in the nonparametric smoothing literature. Condition 1 is similar to those required by Stone (1986) & Zhou et al. (1998).

THEOREM 1. (Consistency) *Suppose that Conditions* 1–5 *hold. Then*

$$d(\hat{\Lambda}_n^{\text{ps}}, \Lambda_0) \to 0,$$

in probability, as $n \to \infty$. *If in addition Condition* 6 *holds, then*

$$d(\hat{\Lambda}_n, \Lambda_0) \to 0,$$

in probability, as $n \to \infty$.

THEOREM 2. (Rate of convergence) *Suppose that Conditions* 1–5 *hold. Then*

$$n^{r/(1+2r)} d(\hat{\Lambda}_n^{\text{ps}}, \Lambda_0) = O_p(1).$$

If in addition Condition 6 *holds, then*

$$n^{r/(1+2r)} d(\hat{\Lambda}_n, \Lambda_0) = O_p(1).$$

The proofs of these theorems are given in the Appendix. Theorem 2 shows that the spline estimators can have a higher rate of convergence than their alternatives studied in Wellner & Zhang (2000), because $r/(1+2r) > 1/3$ when $r > 1$.

4. NUMERICAL RESULTS

4·1. *Simulation studies*

We conduct simulation studies to compare the statistical properties and computational complexities among the spline pseudolikelihood/likelihood estimators and their alternatives studied in Wellner & Zhang (2000). Two Monte Carlo simulation studies designed in Zhang & Jamshidian (2004) are carried out here. In each study, we generate n independent and identically distributed observations $X_i = (K_i, T_i, \mathcal{N}^{(i)})$, for $i = 1, \ldots, n$. For each i, K_i is sampled randomly from a discrete uniform distribution on $\{1, 2, 3, 4, 5, 6\}$. Given K_i, the random panel observation times $T_i = (T^{(i)}_{K_i,1}, \ldots, T^{(i)}_{K_i,K_i})$ are K_i ordered random draws from Un(0,10) and are rounded to the second decimal place. The two simulations differ in the method of generating the panel counts $\mathcal{N}^{(i)} = \{\mathcal{N}^{(i)}(T^{(i)}_{K_i,1}), \ldots, \mathcal{N}^{(i)}(T^{(i)}_{K_i,K_i})\}$, given (K_i, T_i).

Simulation 1. The panel counts are generated from Po$(2t)$; that is,

$$\mathcal{N}^{(i)}(T^{(i)}_{K_i,j}) - \mathcal{N}^{(i)}(T^{(i)}_{K_i,j-1}) \sim \text{Po}\{2(T^{(i)}_{K_i,j} - T^{(i)}_{K_i,j-1})\}, \quad \text{for } j = 1, \ldots, K_i.$$

Simulation 2. The panel counts are generated from a mixed Poisson process. We first generate a random sample $\alpha_1, \ldots, \alpha_n$ from $\{-0{\cdot}4, 0, 0{\cdot}4\}$ with pr$(\alpha_i = -0{\cdot}4)$ = pr$(\alpha_i = 0{\cdot}4)$ = 1/4 and pr$(\alpha_i = 0)$ = 1/2, for $i = 1, \ldots, n$. Given α_i, the panel counts for the ith subject are generated according to Po$\{(2 + \alpha_i)t\}$; that is,

$$\mathcal{N}^{(i)}(T^{(i)}_{K_i,j}) - \mathcal{N}^{(i)}(T^{(i)}_{K_i,j-1}) | \alpha_i \sim \text{Po}\{(2 + \alpha_i)(T^{(i)}_{K_i,j} - T^{(i)}_{K_i,j-1})\}, \quad \text{for } j = 1, \ldots, K_i.$$

This counting process is not a Poisson process unconditionally since the mean function of the process, $E\{\mathcal{N}^{(i)}(t)\} = 2t$, is not equal to the variance function of the process: $\text{var}\{\mathcal{N}^{(i)}(t)\} = 2t + 0.08t^2$.

The cubic I-splines are used in computing the spline estimators. Let $T_{\min}$ and $T_{\max}$ be the two endpoints of the collection of distinct observation time-points in the data. The interval

$[T_{\min}, T_{\max}]$ is equally divided into $m_n + 1$ subintervals, in which m_n is selected as the cubic root of the number of distinct observation times plus 1. Hence the spacing of the knots $\Delta_i = t_i - t_{i-1}$ is proportional to $n^{-1/3}$, for $i = l + 1, \ldots, m_n + l + 1$, and Condition 1 in §3 is automatically satisfied. For each scenario, we generate 1000 Monte Carlo samples with $n = 100$ and $n = 200$.

Figures 1(a) and (b) show the four estimates with $n = 100$ along with the true mean function $\Lambda_0(t) = 2t$ in Simulation 1. It is clear that, while all these estimators converge to the true mean function, the spline estimators are closer to the true mean function than their alternatives. To compare these estimators in detail, we calculate the estimates of the mean function at the time-points $t = 1{\cdot}5, 2{\cdot}0, 2{\cdot}5, \ldots, 9{\cdot}5$. Figures 1(c) and (d) show the squares of the pointwise biases and the pointwise mean squared errors at these time-points for $n = 100$. The biases of all the estimators are clearly negligible compared to the mean squared errors, and the pointwise mean squared errors of the spline estimators are smaller than their alternatives. While the spline likelihood estimator appears to be the most efficient estimator among the four, the spline pseudolikelihood estimator performs almost as well as the nonparametric maximum likelihood estimator. Results for $n = 200$ show the same patterns and the pointwise mean squared errors drop substantially, which supports the asymptotic consistency of these estimators.

Figure 2 shows the corresponding results from Simulation 2, again for $n = 100$. The figures reveal the same patterns as in Simulation 1. These studies also reinforce the conclusion made in Wellner & Zhang (2000) that the likelihood method based on the Poisson process is robust against the underlying counting process. However, the mean squared errors of these estimators are elevated when the Poisson process model is wrongly assumed.

The computing times for the four estimators are summarized in Table 1. The nonparametric maximum pseudolikelihood estimator appears to be the least computationally demanding estimator. However, it is also the least efficient, as shown in Figs. 1(d) and 2(d). The spline likelihood estimator is not only statistically efficient but also computationally efficient: the computation time for the spline likelihood estimator is on average less than 1/12 of that for its alternative. The reduction in computation time for the spline likelihood estimator over its alternative becomes more significant when the sample size increases, as shown in Table 1.

4·2. *A real example: the bladder tumour trial*

The proposed methods are illustrated using the bladder tumour data described in the introduction (Andrews & Herzberg, 1985, pp. 253–60). A total of 116 patients were randomly assigned into one of three treatment groups, 47 to placebo, 31 to pyridoxine and 38 to thiotepa. The number of follow-ups and follow-up times varied greatly from patient to patient.

In this study, the investigators were interested in the efficacy of two treatments, pyridoxine pill and thiotepa installation, in terms of suppressing the recurrence of bladder tumour. The spline pseudolikelihood and spline likelihood estimates of the cumulative mean functions for the three treatment groups are shown in Fig. 3 along with the nonparametric estimates proposed by Wellner & Zhang (2000). The difference between the mean functions of the thiotepa group and the placebo group is quite substantial. Note also the big discrepancy between the pseudolikelihood estimates and the likelihood estimates; this may be because the samples are relatively small, with observations at later times being particularly scarce.

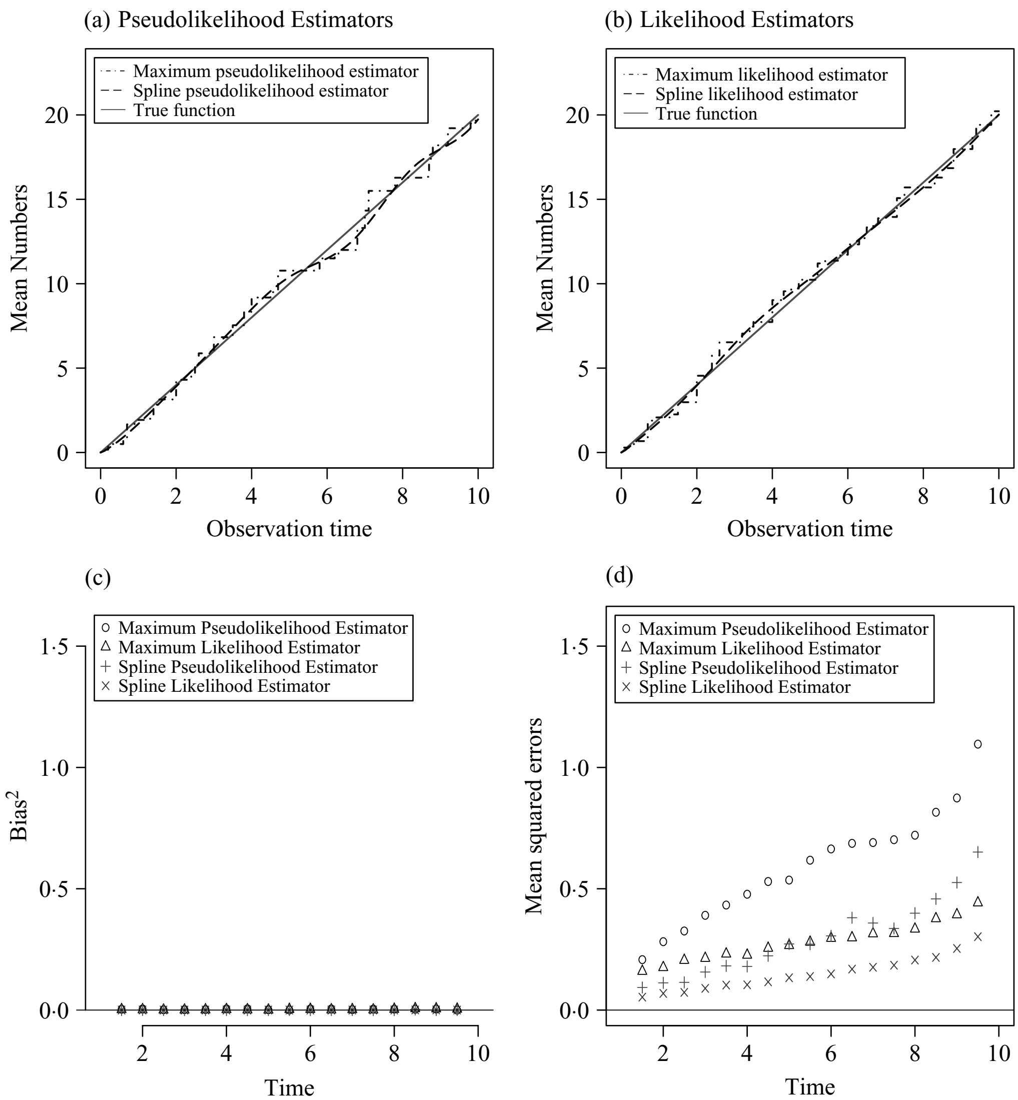

Fig. 1. Simulation study 1, with data from the Poisson process and $n = 100$: (a) maximum pseudolikelihood, and (b) maximum likelihood estimates of the mean function, $\Lambda_0(t) = 2t$, (c) squared biases, and (d) mean squared errors for the four estimators.

5. DISCUSSION

In semiparametric regression problems, joint estimation of the nonparametric component and parametric regression parameters is often a challenging task. For example, in an unpublished technical report of the Department of Statistics at the University of Washington, J. A. Wellner & Y. Zhang considered estimation in the semiparametric proportional mean model with panel count data, namely

$$E\{\mathcal{N}(t)|Z\} = \Lambda_0(t)\exp(\beta_0^{\mathrm{T}} Z), \tag{6}$$

where Z is a vector of covariates and Λ_0 is the baseline mean function. Although the asymptotic properties of the semiparametric maximum pseudolikelihood and the semiparametric maximum likelihood estimators were studied, and the normality of the

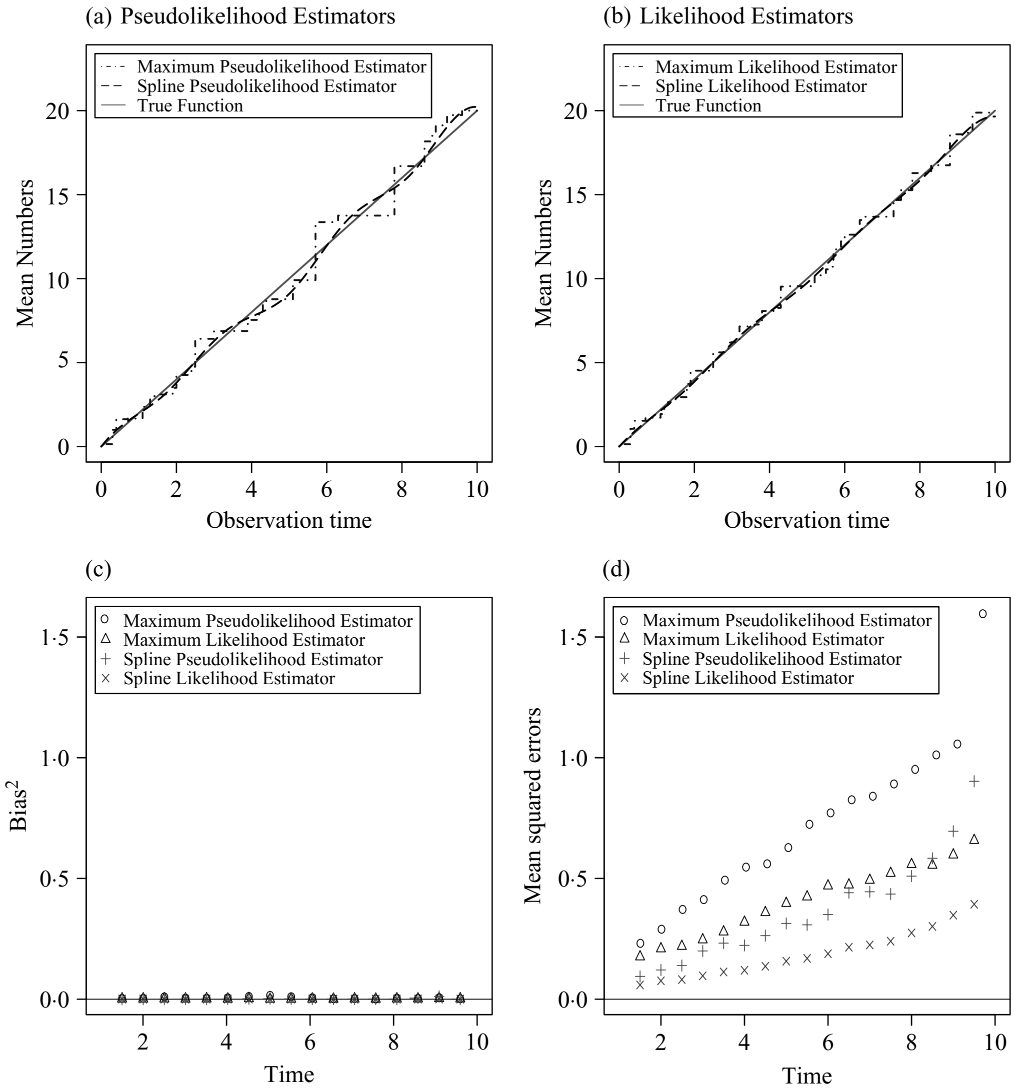

Fig. 2. Simulation study 2, with data from the mixed Poisson process and $n = 100$: (a) maximum pseudolikelihood, and (b) maximum likelihood estimates of the mean function, $\Lambda_0(t) = 2t$, (c) squared biases, and (d) mean squared errors for the four estimators.

estimators of β_0 was established, it was difficult to estimate the asymptotic variances of the estimators directly. The bootstrap procedure that they implemented required a substantial amount of computational effort. It is therefore worthwhile considering computationally more efficient estimators, and the spline estimators appear to be good candidates. The semiparametric model (6) can be reformulated by approximating $\log \Lambda_0(t)$ by a linear combination of the normalized B-splines described in Schumaker (1981), to give

$$E\{\mathcal{N}(t)|Z\} = \Lambda_0(t)\exp(\beta_0^{\mathrm{T}} Z) = \exp\left\{\sum_{i=1}^{q_n} \alpha_i B_i(t) + \beta_0^{\mathrm{T}} Z\right\},$$

subject to the constraints $\alpha_1 \leqslant \cdots \leqslant \alpha_{q_n}$. The joint estimation of the α's and β's can be implemented in a way similar to that described in this article and is expected to be a computationally manageable task.

Table 1. *Comparison of the computing time in seconds among the four estimators of the mean function, based on data generated from the Poisson process or mixed Poisson process with sample size* 100 *or* 200

	Poisson process		Mixed Poisson process	
Estimators	$n = 100$	$n = 200$	$n = 100$	$n = 200$
Maximum pseudolikelihood	0·24	0·66	0·23	0·61
Maximum likelihood	60·34	172·15	58·92	168·43
Spline pseudolikelihood	2·33	4·90	2·24	5·07
Spline likelihood	4·67	11·41	4·59	12·52

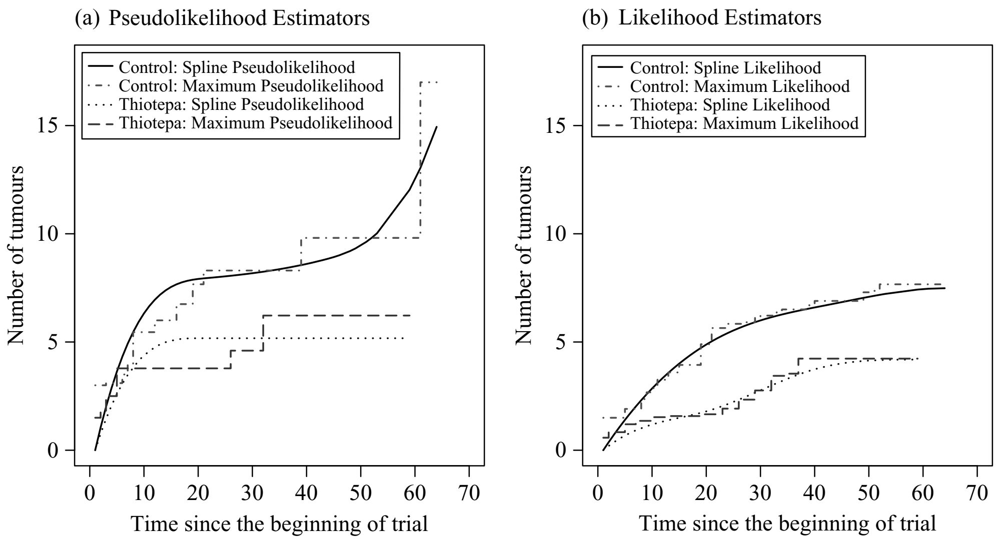

Fig. 3. Bladder tumour data: A comparison of (a) the maximum pseudolikelihood, and (b) maximum likelihood estimates of the mean functions between the control and thiotepa groups.

ACKNOWLEDGEMENT

The authors are grateful to the editor and two referees for their helpful comments and constructive suggestions which led to significant improvement in this paper.

APPENDIX

Proofs

The proofs for two asymptotic results and a technical lemma are sketched here. Modern empirical process theory is the major technical tool to prove the asymptotic results, and the notation used in this section follows that in van der Vaart & Wellner (1996) & Huang (1999). Here we only sketch the proofs for the spline pseudolikelihood estimator, since the proofs for the spline likelihood estimator are very similar.

LEMMA A1. *Suppose g is a monotone nondecreasing function with bounded rth derivative. Then there exists a monotone nondecreasing spline Ag with order $l \geqslant r + 2$ and knot sequence $\mathcal{T}$ satisfying*

$$a = t_1 = \dots t_l < t_{l+1} < \dots < t_{m_n+l} < t_{m_n+l+1} = \dots = t_{m_n+2l} = b$$

such that

$$\|g - Ag\|_\infty \leqslant C|\mathcal{T}|^r \|D^r g\|_\infty$$

where $|\mathcal{T}| = \max_{l+1\leqslant i\leqslant m_n+l+1}(t_i - t_{i-1})$, $D^r g$ *is the rth derivative of g, and* $\|f\|_\infty = \sup_{x\in[a,b]} |f(x)|$.

Proof of Lemma A1. Carl de Boor (2001) constructed a B-spline approximation of g in the form

$$Ag = \sum_{i=1}^{m_n+l} g(\tau_i)B_i,$$

on $[a, b]$, in which $\tau_1 \leqslant \ldots \leqslant \tau_{m_n+l}$ in $[a, b]$ such that $\tau_i = t_{i+l/2}$ if l is even and $\tau_i = (t_{i+(l-1)/2} + t_{i+(l+1)/2})/2$ if l is odd. He showed that $\|g - Ag\|_\infty \leqslant C|\mathcal{T}|^r\|D^r g\|_\infty$ (de Boor, 2001, p. 145). This construction also leads to a monotone spline approximation for a monotone function, since the monotonicity of Ag is guaranteed by the monotonicity among the coefficients $g(\tau_i)$. The proof is complete.

Proof of Theorem 1. A pseudolikelihood function for Λ can be written as

$$\mathcal{M}_n(\Lambda) = l_n^{\mathrm{ps}}(\Lambda|D) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{\mathcal{N}^{(i)}(T_{K_i,j}^{(i)})\log\Lambda(T_{K_i,j}^{(i)}) - \Lambda(T_{K_i,j}^{(i)})\right\}.$$

Let Φ denote the collection of functions h on $O[T]$ whose $(r-1)$th derivative $h^{(r-1)}$, for $r \geqslant 1$, exists and satisfies the Lipschitz conditions

$$|h^{(r-1)}(s) - h^{(r-1)}(t)| \leqslant C|s-t|, \qquad \text{for } s, t \in O[T] = [\sigma, \tau].$$

By Condition 4, the true mean function $\Lambda_0 \in \Phi$. With the knot sequence $\mathcal{T}$ specified in §3, there exists a monotone spline $\Lambda_n \in \psi_{l,t}$ with order $l \geqslant r+2$ and knots $\mathcal{T}$ such that $\|\Lambda_n - \Lambda_0\|_\infty = \sup_{t\in O(T)}|\Lambda_n(t) - \Lambda_0(t)| = O(n^{-\nu r})$ using the result of Lemma A1. This entitles us to modify the monotone spline $\Lambda_n \in \psi_{l,t}$ by adding a constant of order $O(n^{-\nu r})$ such that $\Lambda_n > \Lambda_0$ and $\|\Lambda_n - \Lambda_0\|_\infty = O(n^{-\nu r})$ for large n. Choose a positive monotone $h_n \in \psi_{l,t}$ such that $\|h_n\|_2^2 = O(n^{-\nu r} + n^{-(1-\nu)/2})$. Therefore, for any $\alpha > 0$, $\|\Lambda_n - \Lambda_0 + \alpha h_n\|_2^2 = O(n^{-\nu r} + n^{-(1-\nu)/2})$ and $\inf(\Lambda_n - \Lambda_0 + \alpha h_n) > 0$ for sufficiently large n.

Let $H_n(\alpha) = \mathcal{M}_n(\Lambda_n + \alpha h_n)$. The first and second derivatives of H_n are

$$H_n'(\alpha) = \sum_{i=1}^{n}\sum_{j=1}^{K_i}\left\{\frac{\mathcal{N}^{(i)}(T_{K_i,j}^{(i)})}{\Lambda_n(T_{K_i,j}^{(i)}) + \alpha h_n(T_{K_i,j}^{(i)})} - 1\right\}h_n(T_{K_i,j}^{(i)}),$$

$$H_n''(\alpha) = -\sum_{i=1}^{n}\sum_{j=1}^{K_i}\frac{\mathcal{N}^{(i)}(T_{K_i,j}^{(i)})h_n^2(T_{K_i,j}^{(i)})}{\{\Lambda_n(T_{K_i,j}^{(i)}) + \alpha h_n(T_{K_i,j}^{(i)})\}^2} < 0$$

Thus $H_n'(\alpha)$ is a nonincreasing function. Therefore, to prove Theorem 1, it is sufficient to show that, for any $\alpha_0 > 0$, $H_n'(\alpha_0) < 0$ and $H_n'(-\alpha_0) > 0$ except on an event with probability converging to zero. Then $\hat{\Lambda}_n^{\mathrm{ps}}$ must be between $\Lambda_n - \alpha_0 h_n$ and $\Lambda_n + \alpha_0 h_n$ with probability converging to one, so that $\mathrm{pr}(\|\hat{\Lambda}_n^{\mathrm{ps}} - \Lambda_n\|_2 \leqslant \alpha_0\|h_n\|_2) \to 1$ as $n \to \infty$.

The quantity $H_n'(\alpha_0)/n$ can be written as

$$\mathcal{P}_n\sum_{j=1}^{K}\left\{\frac{\mathcal{N}^{(i)}(T_{K,j})}{\Lambda_n + \alpha_0 h_n} - 1\right\}h_n = (\mathcal{P}_n - P)\sum_{j=1}^{K}\left\{\frac{\mathcal{N}^{(i)}(T_{K,j})}{\Lambda_n + \alpha_0 h_n} - 1\right\}h_n + P\sum_{j=1}^{K}\left\{\frac{\mathcal{N}^{(i)}(T_{K,j})}{\Lambda_n + \alpha_0 h_n} - 1\right\}h_n$$
$$= I_{n1} + I_{n2},$$

where $\mathcal{P}_n$ denotes the empirical measure. With Condition 1, the calculation of Shen & Wong (1994, p. 597) leads to the result that, for $\eta > 0$ and any $\varepsilon \leqslant \eta$,

$$\log N_{[\,]}\{\varepsilon, \psi_{l,t}, L_2(\mu)\} \leqslant cq_n\log(\eta/\varepsilon), \qquad J_{[\,]}\{\eta, \psi_{l,t}, L_2(\mu)\} \leqslant c_0 {q_n}^{1/2}\eta,$$

where $q_n = m_n + l$ is the number of spline base functions.

For each $\varepsilon > 0$, since the bracket number of class $\psi_{l,t}$ is no more than $(\eta/\varepsilon)^{cq_n}$, there exists a set of brackets $\{[\Lambda_i^L, \Lambda_i^R] : i = 1, 2, \ldots, (\eta/\varepsilon)^{cq_n}\}$ such that, for each $\Lambda \in \psi_{l,t}$,

$$\Lambda_i^L(t) \leqslant \Lambda(t) \leqslant \Lambda_i^R(t),$$

for all $t \in O[T]$ and some i, with $d^2(\Lambda_i^R, \Lambda_i^L) = \int \left\{\Lambda_i^R(t) - \Lambda_i^L(t)\right\}^2 d\mu(t) \leqslant \varepsilon^2$.

For any $\eta > 0$, define the class

$$\mathcal{F}_\eta = \left\{ \sum_{j=1}^{K} \left(\frac{\mathcal{N}(T_{K,j})}{\Lambda} - 1 \right) (\Lambda - \Lambda_n) : \ \Lambda \in \psi_{l,t} \text{ and } d(\Lambda, \Lambda_n) \leqslant \eta \right\}.$$

By the Cauchy–Schwartz inequality and Conditions 3–5, we can show that $\mathcal{F}_\eta$ is a Donsker class. Hence, $I_{n1} = O_p(n^{-1/2})$.

Define $m(s) = \dfrac{\Lambda_0}{\Lambda_0 + s\Delta_n}$, where $\Delta_n = \Lambda_n - \Lambda_0 + \alpha_0 h_n$, $0 \leqslant s \leqslant 1$. By Taylor expansion,

$$m(s) = 1 + \left(-\frac{\Delta_n}{\Lambda_0} \right) s + \frac{\Lambda_0 \Delta_n^2}{(\Lambda_0 + \xi\Delta_n)^3} s^2,$$

for some ξ between 0 and 1. Since Λ_0 and Δ_n are bounded on $O[T]$, there exist constants c_1 and c_2 such that

$$c_1 E \sum_{j=1}^{K} \Delta_n^2 \leqslant E \sum_{j=1}^{K} \frac{\Lambda_0 \Delta_n^2}{(\Lambda_0 + \xi\Delta_n)^2} \leqslant c_2 E \sum_{j=1}^{K} \Delta_n^2.$$

Therefore,

$$E \sum_{j=1}^{k} \frac{\Lambda_0 \Delta_n^2}{(\Lambda_0 + \xi\Delta_n)^2} = O(n^{-\nu r} + n^{-(1-\nu)/2}),$$

and hence

$$I_{2n} \leqslant E \sum_{j=1}^{K} (-c_1\Delta_n + c_2\Delta_n^2) h_n \leqslant -\frac{c_1}{2} E \sum_{j=1}^{K} \Delta_n^2 = -\frac{c_1}{2} p_n^{-1},$$

where $p_n^{-1} = n^{-\nu r} + n^{-(1-\nu)/2}$. Since $n^{-\nu r} + n^{-(1-\nu)/2} \geqslant n^{-r/(1+2r)} > n^{-1/2}$, for $0 < \nu < 1/2$, we have

$$H_n'(\alpha_0) \leqslant O_p(n^{-1/2}) - cp_n^{-1} < 0,$$

except on an event with probability converging to zero. The same arguments show that $H_n'(-\alpha_0) > 0$ with probability converging to 1 as $n \to \infty$.

Proof of Theorem 2 . Let $m_\Lambda(x) = \sum_{j=1}^{K} \left\{\mathcal{N}(T_{K,j}) \log \Lambda(T_{K,j}) - \Lambda(T_{K,j})\right\}$ and define $M(\Lambda) = Pm_\Lambda(x)$ and $\mathcal{M}_n(\Lambda) = \mathcal{P}_n m_\Lambda(x)$. Then the log-pseudolikelihood can be written as $n\mathcal{P}_n m_\Lambda(x)$.

For any $\eta > 0$, define the class

$$\mathcal{F}_\eta = \{\Lambda | \Lambda \in \psi_{l,t}, \ d(\Lambda, \Lambda_0) \leqslant \eta\}.$$

By Theorem 1, $\hat{\Lambda}_n^{\text{ps}} \in \mathcal{F}_\eta$ for any $\eta > 0$ and sufficiently large n.

Next, define the class

$$\mathcal{M}_\eta = \{m_\Lambda(x) - m_{\Lambda_0}(x) : \ \Lambda \in \mathcal{F}_\eta\}.$$

With Condition 1, following the calculation of Shen & Wong (1994, p. 597) we can establish that

$$\log N_{[\,]}(\varepsilon, \mathcal{M}_\eta, || \cdot ||_{P,B}) \leqslant cq_n \log(\eta/\varepsilon),$$

where $\| \cdot \|_{P,B}$ is the Bernstein Norm defined as $\|f\|_{P,B} = \{2P(e^{|f|} - 1 - |f|)\}^{1/2}$ by van der Vaart & Wellner (1996). Moreover, some algebraic calculations lead to $||m_\Lambda(X) -$

$m_{\Lambda_0}(X)||^2_{P,B} \leqslant C\eta^2$, for any $m_\Lambda(X) - m_{\Lambda_0}(X) \in \mathcal{M}_\eta$ using Conditions 2–5. Therefore, by Lemma 3·4·3 of van der Vaart & Wellner (1996), we obtain

$$E_P||n^{1/2}(\mathcal{P} - P)||_{\mathcal{M}_\eta} \leqslant CJ_{[\,]}(\eta, \mathcal{M}_\eta, ||\cdot||_{P,B})\left\{1 + \frac{J_{[\,]}(\eta, \mathcal{M}_\eta, ||\cdot||_{P,B})}{\eta^2 n^{1/2}}\right\}, \tag{A1}$$

where

$$J_{[\,]}(\eta, \mathcal{M}_\eta, ||\cdot||_{P,B}) = \int_0^\eta \{1 + \log N_{[\,]}(\varepsilon, \mathcal{F}, ||\cdot||_{P,B})\}^{1/2} d\varepsilon \leqslant c_0 {q_n}^{1/2}\eta.$$

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

$$r_n^2\phi(\frac{1}{r_n}) = r_n {q_n}^{1/2} + r_n^2 q_n/n^{1/2} \leqslant 2n^{1/2},$$

for $r_n = n^{(1-\nu)/2}$ and $0 < \nu < 1/2$. Hence, $n^{(1-\nu)/2} d(\hat{\Lambda}_n^{\text{ps}}, \Lambda_0) = O_p(1)$ by Theorem 3·2·5 of van der Vaart & Wellner (1996). The choice of $\nu = 1/(1+2r)$ yields the rate of convergence of $r/(1+2r)$, which completes the proof.

References

Anderson, J. A. & Senthilselvan, A. (1980). Smooth estimates for the hazard function. *J. R. Statist. Soc.* B **42**, 322–7.

Andrews, D. F. & Herzberg, A. M. (1985). *Data; A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Spring-Verlag.

De boor, C. (2001). *A Practical Guide to Splines*. New York: Spring-Verlag.

Byar, D. P. (1980). The Veterans administration study of chemoprophylaxis for recurrent stage I bladder tumours: comparisons of placebo, pyridoxine, and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology*, Ed. M. Macaluso, P. H. Smith and F. Edsmyr, pp. 363–70. New York: Plenum.

Byar, D. P., Blackard, C. & The Veterans Administration Cooperative Urological Research Group. (1977). Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology* **10**, 556–61.

Cai, T. & Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics* **59**, 570–9.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Ann. Statist.* **27**, 1536–63.

Jamshidian, M. (2004). On algorithms for restricted maximum likelihood estimation. *Comp. Statist. Data Anal.* **45**, 137–57.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comp. Graph. Statist.* **7**, 310–21.

Kalbfleisch, J. D. & Lawless, J. F. (1985). The analysis of panel count data under a Markov assumption. *J. Am. Statist. Assoc.* **80**, 863–71.

Kooperberg, C., Stone, C. J. & Truong, Y. K. (1995). Hazard regression. *J. Am. Statist. Assoc.* **90**, 78–94.

Lee, E. W. & Kim, M. Y. (1998). The analysis of correlated panel count data using a continuous-time Markov model. *Biometrics* **54**, 1638–44.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* **3**, 425–41.

Rosen, J. B. (1960). The gradient projection method for nonlinear programming. *J. SIAM* **8**, 181–217.

Rosenberg, P. S. (1995). Hazard function estimation using B-splines. *Biometrics* **51**, 874–87.

Schumaker, L. (1981). *Spline Functions: Basic Theory*. New York: Wiley.

Shen, X. & Wong, W. H. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22**, 580–615.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14**, 590–606.

Sun, J. & kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica.* **5**, 279–90.

Sun, J. & Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. R. Statist. Soc.* B **62**, 293–302.

Thall, P. F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44**, 197–209.

Thall, P. F. & Lachin, J. M. (1988). Analysis of recurrent events: nonparametric methods for random-interval count data. *J. Am. Statist. Assoc.* **83**, 339–47.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Application to Statistics*. New York: Springer-Verlag.
WEI, L. J., LIN, D. Y. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065–73.
WELLNER, J. A. & ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779–814.
WHITTEMORE, A. S. & KELLER, J. B. (1986). Survival estimation using splines. *Biometrics* **42**, 495–506.
ZHANG, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39–48.
ZHANG, Y. & JAMSHIDIAN, M. (2004). On algorithms for NPMLE of the failure function with censored data. *J. Comp. Graph. Statist.* **13**, 123–40.
ZHOU, S., SHEN, X. & WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26**, 1760–82.

[*Received April* 2006. *Revised January* 2007]