GENERALIZED LEAST SQUARES ESTIMATION OF THE MEAN FUNCTION OF A COUNTING PROCESS BASED ON PANEL COUNTS

X. Joan Hu[1], Stephen W. Lagakos[2] and Richard A. Lockhart[1]

[1]*Simon Fraser University and* [2]*Harvard School of Public Health*

Abstract: This paper considers nonparametric estimation of the mean function of a counting process based on periodic observations, i.e., panel counts. We present estimators derived through minimizing a class of generalized sums of squares subject to a monotonicity constraint. We establish consistency of the estimators and provide procedures to implement them with various weight functions. For specific weight functions, they reduce to the estimator given in Sun and Kalbfleisch (1995), and are closely related to the nonparametric maximum likelihood estimator studied in Wellner and Zhang (2000). With other weight functions, the proposed estimators provide alternatives that can have better efficiency in non-Poisson situations than previous approaches. Simulations are used to examine the finite-sample performance of the proposed estimators.

Key words and phrases: Isotonic regression, monotonicity constraint, periodic observations.

1. Introduction

Suppose that the number of events that occur until time $t > 0$ is $N(t)$, and we wish to estimate the marginal expectation $E\{N(t)\}$ without specifying a probability model for the counting process $\{N(t), t > 0\}$. See Lawless and Nadeau (1995) Lin, Wei, Yang and Ying (2000) and Wellner and Zhang (2000) for examples of this setup.

When $\{N(t), t > 0\}$ is a continuously-observed Poisson process subject to right-censoring, the Nelson-Aalen estimator (cf., Andersen, Borgan, Gill and Keiding (1992, Chap. IV)) is commonly used for estimating the cumulative intensity function of the counting process, which is the same as its mean function $\Lambda_0(t) = E\{N(t)\}$ (e.g., Lawless (1995)). Lawless and Nadeau (1995) and Lin et al. (2000) indicate that the Nelson-Aalen estimator is a consistent estimator for the marginal mean function of a counting process regardless of the underlying probability model.

When the process for an individual is only observed at a finite set of time points (that is, the observations are panel counts), Sun and Kalbfleisch (1995)

present an estimator of the mean function based on isotonic regression, and Wellner and Zhang (2000) derive a maximum pseudo-likelihood estimator (NPMPLE) and the nonparametric maximum likelihood estimator (NPMLE) under the assumption that the counting process is Poisson. Wellner and Zhang note that their NPMPLE is the same as the Sun-Kalbfleisch estimator, demonstrate that the NPMLE is more efficient than the NPMPLE via simulation, and show that both their estimators are consistent without the Poisson assumption. The efficiency of the NPMLE for non-Poisson counts has not been explored. Hu and Lagakos (2007) consider two weighted least squares estimators for the mean of an arbitrary response process in a general incomplete data setting. Their estimators reduce to, or are closely related to, some commonly-used estimators with certain data structures, but are not necessarily monotone with panel counts.

This paper proposes a general class of nonparametric estimators for the mean function $\Lambda_0(\cdot)$ based on periodic observations, without specifying an underlying probability model. The estimators arise by minimizing a generalized weighted sum of squares under a monotonicity constraint. We establish consistency of the general estimator and provide procedures to implement it with various weights. One choice of weight function leads to the Sun-Kalbfleisch estimator and thus also the Wellner-Zhang NPMPLE. Another leads to an estimator that is closely related to the Wellner-Zhang NPMLE. Via simulation, we study the finite sample performance of the proposed estimators in various situations, and compare them to the Sun-Kalbfleisch estimator and the Wellner-Zhang NPMLE. With some other specific weight functions, the proposed approach provides alternatives to the NPMLE and NPMPLE that can have higher efficiency in situations with non-Poisson counts.

Without loss of generality, we take $N(0) \equiv 0$ and thus $\Lambda_0(0) = 0$. We assume that the counting process and the observation mechanism are independent. Section 2 presents a generalized least squares estimator for the mean function $\Lambda_0(\cdot)$ subject to the constraint that $\Lambda_0(\cdot)$ is monotone, and then establishes consistency of the estimator. Section 3 provides procedures to implement the estimator, and Section 4 examines the specific estimators that result from several choices of weight function. Section 5 presents the results of a simulation study comparing the proposed estimators with those of Sun and Kalbfleisch (1995) and Wellner and Zhang (2000). Section 6 discusses the results and extensions.

2. A Generalized Least Squares Monotonic Estimator

Hu and Lagakos (2007) show that two commonly-used nonparametric estimators, the sample mean of available data and the Nelson-Aalen estimator, can be derived as a solution to a weighted sum of squares criterion. This, along with the monotonicity of the mean function $\Lambda_0(\cdot)$, leads us to explore a criterion

for defining nonparametric estimators through minimizing a generalized sum of squares subject to the monotonicity constraint. In the following, we introduce a generalized sum of squares criterion, and establish consistency of the estimator obtained through its constrained maximization.

Let $\mathbf{T} = (T_{K,1}, \ldots, T_{K,K})'$, with $T_{K,j}$ denoting the jth of K observation times during time period $[0, \tau]$, with $0 < \tau < \infty$, where τ is fixed and K can be random. Let $\mathbf{N} = \big(N(T_{K,1}), \ldots, N(T_{K,K})\big)'$ and $\mathbf{\Lambda} = \big(\Lambda(T_{K,1}), \ldots, \Lambda(T_{K,K})\big)'$. Assume $\{\mathbf{X}_i : i = 1, \ldots, n\}$ is a set of i.i.d. samples of $\mathbf{X} = \big(K, \mathbf{T}, \mathbf{N}\big)$. We consider the generalized sum of squares with a symmetric weight function $w(u, v)$ for $u, v \in \{T_{k,1}, \ldots, T_{K,K}\}$,

$$L_n(\Lambda; w) = \sum_{i=1}^{n} m(\Lambda; w | \mathbf{X}_i), \tag{2.1}$$

where

$$m(\Lambda; w | \mathbf{X}) = \big(\mathbf{N} - \mathbf{\Lambda}\big)' \mathbf{W} \big(\mathbf{N} - \mathbf{\Lambda}\big)$$
$$= \sum_{j=1}^{K} \sum_{l=1}^{K} w(T_{K,j}, T_{K,l}) \big\{N(T_{K,j}) - \Lambda(T_{K,j})\big\} \big\{N(T_{K,l}) - \Lambda(T_{K,l})\big\}, \tag{2.2}$$

and where $\mathbf{W}$ is the $K \times K$ symmetric matrix with (j, l) entry $w(T_{K,j}, T_{K,l})$. The weight function is deterministic given $(K, \mathbf{T})$. Let $\hat{\Lambda}_n(\cdot; w)$ be the minimum point of $L_n(\Lambda; w)$ over all nondecreasing functions defined in $[0, \tau]$, that is,

$$\hat{\Lambda}_n(\cdot; w) = \mathrm{argmin}_{\Lambda \in \mathcal{F}} L_n(\Lambda; w) \tag{2.3}$$

with $\mathcal{F} = \big\{\Lambda(\cdot) : \text{ nondecreasing in } [0, \tau]\big\}$. Following Wellner and Zhang (2000), we define a measure μ in $[0, \tau]$ based on the distribution of $(K, \mathbf{T})$:

$$\mu(B) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(T_{K,j} \in B | K = k),$$

where B is a Borel set in $[0, \tau]$. For $\Lambda, \Lambda' \in \mathcal{F}$, define

$$d(\Lambda, \Lambda') = \Big[\int \{\Lambda(t) - \Lambda'(t)\}^2 d\mu(t) \Big]^{\frac{1}{2}}. \tag{2.4}$$

The following theorem establishes the consistency of $\hat{\Lambda}_n(\cdot; w)$ with respect to the metric d.

Condition A: $\Lambda_0(t) = \mathrm{E}\{N(t)\}$ and $c(t, s) = \mathrm{Cov}\{N(t), N(s)\}$ exist for $t, s \in [0, \tau]$, $\Lambda_0(\cdot)$ is strictly increasing with $\Lambda_0(0) = 0$, and $\mathrm{E}\{\mathbf{N}'\mathbf{N}\} < \infty$.

Condition B: The $K \times K$ weight matrix $\mathbf{W} = \left(w(T_{K,j}, T_{K,l}) \right)$ can be expressed as $\mathbf{W} = A^{'}A$ with nonsingular A, and $\mathrm{E}\{||A||^2\} < \infty$ and $\mathrm{E}\{||A^{-1}||^2\} < \infty$, where $||A||^2 = \sum a_{ij}^2 = tr(A^{'}A)$.

Condition C: $\mu(\{\tau\}) > 0$ and $\mathrm{E}\{K^2\} < \infty$.

Theorem 1. *Assume Conditions $A-C$. For a given weight function w, $d(\hat{\Lambda}_n(\cdot; w),$ $\Lambda_0(\cdot)) \to 0$ almost surely as $n \to \infty$.*

A proof of the theorem is outlined in Appendix A. Following the arguments in Wellner and Zhang (2000), we can extend the result to the situations with $\mu(\{\tau\}) = 0$. With some additional conditions on the observation mechanism, we may also show that the convergence rate of $\hat{\Lambda}_n(\cdot; w)$ is at least $n^{1/3}$.

3. Implementation of Estimator $\hat{\Lambda}_n(\cdot; w)$

We begin by introducing an estimator closely related to $\hat{\Lambda}_n(\cdot; w)$ with an arbitrary weight function, then show how this can be used to find $\hat{\Lambda}_n(\cdot; w)$.

3.1. Preparation

Denote the set of distinct values of the observation times $\{T_{K_i,j}, j=1,\ldots,K_i,$ $i = 1,\ldots,n\}$ by $\mathcal{T}_n = \{s_l : l = 1,\ldots,J_n\}$, with $0 = s_0 < \cdots < s_{J_n} \leq \tau$. Let $\underset{\sim}{\mathbf{N}} = \left(N(s_1),\ldots,N(s_{J_n}) \right)^{'}$ with $\underset{\sim}{\mathbf{N}}_i$ denoting the realization from subject i, $\underset{\sim}{\mathbf{\Lambda}} = \left(\Lambda(s_1),\ldots,\Lambda(s_{J_n}) \right)^{'}$, and $\Delta_i = \mathrm{diag}\{\delta_i(s_l) : s_l \in \mathcal{T}_n\}$, where $\delta_i(s_l) = 1$ or 0 depending on whether or not $N_i(s_l)$ is observed. Further, let $\underset{\sim}{\mathbf{W}}_i$ be the $J_n \times J_n$ matrix with $(l,l^{'})$ entry $w_i(s_l, s_{l'}) = w(s_l, s_{l'})$ if both $l, l^{'}$ belong to $\{k : \text{there exists a } j \text{ such that } T_{K_i,j} = s_k\}$, and zero otherwise. The objective function in (2.1) is then expressible as

$$L_n(\Lambda; w) = \sum_{i=1}^{n} \left(\underset{\sim}{\mathbf{N}}_i - \underset{\sim}{\mathbf{\Lambda}} \right)^{'} \Delta_i^{'} \underset{\sim}{\mathbf{W}}_i \Delta_i \left(\underset{\sim}{\mathbf{N}}_i - \underset{\sim}{\mathbf{\Lambda}} \right). \qquad (3.1)$$

The estimator $\hat{\Lambda}_n(\cdot; w)$ defined in (2.3) is thus an element in the class $\{\Lambda : \Lambda \in \mathcal{F}, \underset{\sim}{\mathbf{\Lambda}} = \hat{\underset{\sim}{\mathbf{\Lambda}}}(w)\}$ with $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w) = \mathrm{argmin}_{\underset{\sim}{\mathbf{\Lambda}} \in \mathcal{C}_+^{J_n}} L_n(\Lambda; w)$ for a given weight function $w(\cdot)$. Here $\mathcal{C}_+^J$ is the cone in $\mathcal{R}^J = \{y = (y_1,\ldots,y_J)^{'} : y_j \in \mathcal{R}\}$ defined by $\mathcal{C}_+^J = \{y \in \mathcal{R}^J : 0 \leq y_1 \leq \ldots \leq y_J\}$. A natural choice for $\hat{\Lambda}_n(\cdot; w)$ is the step function, starting with 0 at $t = 0$ and having jumps at s_l with size equal to the lth component of $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w)$ for $l = 1,\ldots,J_n$. For a given weight function $w(\cdot)$, we

can construct $\hat{\Lambda}_n(\cdot; w)$ through $\hat{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$. Thus, in the following we focus on how to obtain $\hat{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$.

The objective function $L_n(\Lambda; w)$ in (3.1) can be decomposed into the sum of two terms,

$$L_n(\Lambda; w) = \sum_{i=1}^{n} \left(\underset{\sim}{\mathbf{N}}_i - \bar{\underset{\sim}{\boldsymbol{\Lambda}}} \right)' \Delta_i' \underset{\sim}{\mathbf{W}}_i \Delta_i \left(\underset{\sim}{\mathbf{N}}_i - \bar{\underset{\sim}{\boldsymbol{\Lambda}}} \right) + \left(\bar{\underset{\sim}{\boldsymbol{\Lambda}}} - \underset{\sim}{\boldsymbol{\Lambda}} \right)' B_n(w) \left(\bar{\underset{\sim}{\boldsymbol{\Lambda}}} - \underset{\sim}{\boldsymbol{\Lambda}} \right) \quad (3.2)$$

with

$$B_n(w) = \sum_{i=1}^{n} \Delta_i' \underset{\sim}{\mathbf{W}}_i \Delta_i \text{ and } \bar{\underset{\sim}{\boldsymbol{\Lambda}}}(w) = \left(\sum_{i=1}^{n} \Delta_i' \underset{\sim}{\mathbf{W}}_i \Delta_i \right)^{-1} \sum_{i=1}^{n} \Delta_i' \underset{\sim}{\mathbf{W}}_i \Delta_i \underset{\sim}{\mathbf{N}}_i. \quad (3.3)$$

Note that $\bar{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$ minimizes $L_n(\Lambda; w)$ in (3.1). It is easy to show that, conditional on the observation times and for a given weight function $w(\cdot)$, $\bar{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$ is an unbiased estimator of $\underset{\sim}{\boldsymbol{\Lambda}}_0 = (\Lambda_0(s_1), \ldots, \Lambda_0(s_{J_n}))'$, provided the inverse in (3.3) exists. The conditional covariance matrix of $\bar{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$ is $B_n(w)^{-1}(\sum \Delta_i' \underset{\sim}{\mathbf{W}}_i \Delta_i \text{Var}(\underset{\sim}{\mathbf{N}}) \Delta_i' \underset{\sim}{\mathbf{W}}_i \Delta_i) B_n(w)^{-1}$. In fact, $\bar{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$ gives a consistent estimator of $\Lambda_0(t)$ for $t \in \mathcal{T} = \lim_{n\to\infty} \mathcal{T}_n$, provided that the probability of an observation at $t \in \mathcal{T}$ is positive (Hu and Lagakos (2007)).

The bar estimator $\bar{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$ does not always belong to $\mathcal{C}_+^{J_n}$ and thus is not the same as $\hat{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$ in general. We present procedures to compute $\hat{\underset{\sim}{\boldsymbol{\Lambda}}}(w) \in \mathcal{C}_+^{J_n}$. From (3.2), the estimator $\hat{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$ can be viewed as the projection of $\bar{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$ into $\mathcal{C}_+^{J_n}$ with respect to the metric ρ: $\rho_w(a, b)^2 = (a - b)' B_n(w)(a - b)$ for $a, b \in \mathcal{R}^{J_n}$.

3.2. Procedures to implement $\hat{\underset{\sim}{\boldsymbol{\Lambda}}}(w)$

We consider an application of the result given in Wellner and Zhan (1997) and Wellner and Zhang (2000), which is restated as Lemma 2 in Appendix B. Let $\phi(\underset{\sim}{\boldsymbol{\Lambda}}) = L_n(\Lambda; w)/2$. The estimator can be obtained by jointly solving the following equation and inequalities in $\underset{\sim}{\boldsymbol{\Lambda}} = (\Lambda(s_1), \ldots, \Lambda(s_{J_n}))' \in \mathcal{C}_+^{J_n}$:

$$\underset{\sim}{\boldsymbol{\Lambda}}' \nabla \phi(\underset{\sim}{\boldsymbol{\Lambda}}) = \sum_{s \in \mathcal{T}_n} \Lambda(s) \frac{\partial \phi(\underset{\sim}{\boldsymbol{\Lambda}})}{\partial \Lambda(s)} = 0, \quad (3.4)$$

$$(\overbrace{0, \ldots, 0}^{j-1}, 1, \ldots, 1)' \nabla \phi(\underset{\sim}{\boldsymbol{\Lambda}}) = \sum_{s \geq s_j; s \in \mathcal{T}_n} \frac{\partial \phi(\underset{\sim}{\boldsymbol{\Lambda}})}{\partial \Lambda(s)} \geq 0, \quad s_j \in \mathcal{T}_n. \quad (3.5)$$

Here we have

$$\nabla\phi(\underset{\sim}{\mathbf{\Lambda}}) = -\sum_{i=1}^{n}\Delta_i'\underset{\sim}{\mathbf{W}}_i\Delta_i\Big(\underset{\sim}{\mathbf{N}}_i - \underset{\sim}{\mathbf{\Lambda}}\Big) = -B_n(w)\big(\bar{\underset{\sim}{\mathbf{\Lambda}}}(w) - \underset{\sim}{\mathbf{\Lambda}}\big),\qquad (3.6)$$

with $B_n(w)$ and $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w)$ defined in (3.3). For the special case where $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w) \in \mathcal{C}_+^{J_n}$, $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w) = \bar{\underset{\sim}{\mathbf{\Lambda}}}(w)$. In general, it is not easy to jointly solve (3.4) and (3.5) for $\underset{\sim}{\mathbf{\Lambda}} \in \mathcal{C}_+^{J_n}$.

An alternative approach for evaluating $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w)$ utilizes the ICM algorithm (e.g., Jongbloed (1998)), which is restated for our application as Lemma 3 in Appendix B. Specifically, the sequence $\big\{\underset{\sim}{\mathbf{\Lambda}}^{(k)}(w) : k = 1, \dots \big\}$ in the algorithm is obtained as

$$\Lambda^{(k)}(s_l) = \max_{v \le l}\min_{r \ge l}\left.\frac{\sum_{j=v}^{r}\Lambda(s_j)\phi_{jj}(\underset{\sim}{\mathbf{\Lambda}}) - \phi_j(\underset{\sim}{\mathbf{\Lambda}})}{\sum_{j=v}^{r}\phi_{jj}(\underset{\sim}{\mathbf{\Lambda}})}\right|_{\underset{\sim}{\mathbf{\Lambda}}=\underset{\sim}{\mathbf{\Lambda}}^{(k-1)}}, \quad l = 1, \dots, J_n, \quad (3.7)$$

where $\underset{\sim}{\mathbf{\Lambda}}^{(0)}$ is an initial estimator, $\phi_j(\underset{\sim}{\mathbf{\Lambda}})$ is the jth component of $\nabla\phi(\underset{\sim}{\mathbf{\Lambda}})$ in (3.6), and $\phi_{jj}(\underset{\sim}{\mathbf{\Lambda}})$ is the (j, j) element of the matrix $\nabla^2\phi(\underset{\sim}{\mathbf{\Lambda}}) = B_n(w)$.

From (3.2), $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w) = \text{argmin}_{\underset{\sim}{\mathbf{\Lambda}}\in\mathcal{C}_+^{J_n}}\phi^*(\underset{\sim}{\mathbf{\Lambda}})$ with $\phi^*(\underset{\sim}{\mathbf{\Lambda}}) = (\bar{\underset{\sim}{\mathbf{\Lambda}}} - \underset{\sim}{\mathbf{\Lambda}})'B_n(w)(\bar{\underset{\sim}{\mathbf{\Lambda}}} - \underset{\sim}{\mathbf{\Lambda}})/2$. It can be computationally easier in some situations to use $\phi^*(\underset{\sim}{\mathbf{\Lambda}})$ instead of $\phi(\underset{\sim}{\mathbf{\Lambda}}) = L_n(\Lambda; w)/2$ to obtain the estimator $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w)$. The kth evaluation $\underset{\sim}{\mathbf{\Lambda}}^{(k)}(w)$ in the iterative procedure is the left derivative of the greatest convex minorant of the cumulative sum diagram

$$\Big(\sum_{r=1}^{j}a_{rr}, \sum_{r=1}^{j}b_r(\underset{\sim}{\mathbf{\Lambda}})\Big)\Big|_{\underset{\sim}{\mathbf{\Lambda}}=\underset{\sim}{\mathbf{\Lambda}}^{(k-1)}}, \quad j = 1, \dots, J_n,$$

with a_{rr} the (r, r) element of the matrix $B_n(w) = \nabla^2\phi(\underset{\sim}{\mathbf{\Lambda}})$, and $b_r(\underset{\sim}{\mathbf{\Lambda}})$ the rth component of the vector $B_n(w)\bar{\underset{\sim}{\mathbf{\Lambda}}}(w)+\big[\text{diag}(B_n(w))-B_n(w)\big]\underset{\sim}{\mathbf{\Lambda}}$. When $B_n(w)$ is diagonal, the estimator $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w)$ is the isotonic regression of $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w)$ with respect to the weights of $B_n(w)$'s diagonal elements (Barlow, Bartholomew, Bremner and Brunk (1972, Chap. 1)). Thus, we view $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w)$ as a generalized isotonic regression of the bar estimator $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w)$ with weight matrix $B_n(w)$.

4. Choices of Weight Function

We now consider several weight functions for the estimator $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w)$. Each determines the random matrix $\mathbf{W}_{K\times K}$ in the expression $m(\Lambda; w|\mathbf{X})$ in (2.2).

4.1. Observed sample mean weight (OSM)

In practice, the sample mean of available data, termed the observed sample mean by Hu and Lagakos (2007), is routinely used to summarize repeated measures with random missing data to indicate trends along the time and differences between treatment groups. This suggests taking $\mathbf{W}$ in (2.2) to be $\mathbf{W}_{OSM} = I_{K \times K}$, the $K \times K$ identity matrix. The matrices $\underset{\sim}{\mathbf{W}}_i$ in (3.1) are then $\Delta_i = \mathrm{diag}\big(\delta_i(s_j) : j = 1, \ldots, J_n\big)$. It is easy to verify that this weight satisfies *Condition B* in Section 2, since $\mathbf{W}_{OSM} = A^{'}A$ with $A = I_{K \times K}$. The bar estimator and the B_n matrix in (3.3) are then

$$\bar{\Lambda}_{OSM}(t) = \frac{\sum_{i=1}^n \delta_i(t) N_i(t)}{\sum_{i=1}^n \delta_i(t)}, \quad t \in \mathcal{T}_n = \{s_1, \ldots, s_{J_n}\}, \tag{4.1}$$

and $B_n(w_{OSM}) = \sum \Delta_i^{'} \Delta_i = \mathrm{diag}\big(n_j : j = 1, \ldots, J_n\big)$ with $n_j = \sum \delta_i(s_j)$, the size of the observations at time s_j. See Hu and Lagakos (2007) for more discussion of $\bar{\Lambda}_{OSM}(t)$ for an arbitrary response process with a more general observation mechanism. For right-censored survival data, it becomes the reduced sample estimator introduced by Kaplan and Meier (1958) as an alternative to the product-limit estimator.

Following the definition given in Barlow et al. (1972, p.9), the estimator $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$ is the isotonic regression of $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$ with weights $\big\{n_j : j = 1, \ldots, J_n\big\}$. Thus $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$ can be obtained directly from $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$ and the given weights by

$$\hat{\Lambda}_{OSM}(s_l) = \max_{v \leq l} \min_{r \geq l} \frac{\sum_{j=v}^r n_j \bar{\Lambda}_{OSM}(s_j)}{\sum_{j=v}^r n_j}, \quad l = 1, \ldots, J_n.$$

The estimator $\hat{\Lambda}_{OSM}(\cdot)$ is the same as the Sun-Kalbfleisch estimator (1995). Wellner and Zhang (2000) note that their NPMPLE is also the same as the Sun-Kalbfleisch estimator.

4.2. Cumulative observed increment weight (COI)

Motivated by the construction of the Kaplan-Meier estimator, Hu and Lagakos (2007) propose an estimator based on the cumulative observed increments (COI), and show that this estimator can be more efficient than the observed sample mean when the observations at different times from the same study individual are correlated. Following their approach, we consider the weight matrix $\mathbf{W}$ in (2.2) as $\mathbf{W}_{COI} = \Omega^{'}\Omega$ with

$$\Omega = \Big(a(T_{K,j}, T_{K,l})\Big)_{K \times K}, \text{and } a(T_{K,j}, T_{K,l}) = \begin{cases} 1 & \text{for } l = j \text{ with } j = 1, \ldots, K, \\ -1 & \text{for } l = j-1 \text{ with } j = 2, \ldots, K, \\ 0 & \text{otherwise.} \end{cases}$$

$$\tag{4.2}$$

It is easy to verify that the weight function satisfies *Condition B* in Section 2. The associated objective function $L_n(\Lambda; w_{COI})$ depends only on the observed increments $\Delta N(T_{K,j}) = N(T_{K,j}) - N(T_{K,j-1})$, $j = 1, \ldots, K$, with $T_{K,0} = 0$. Here the matrix $\underset{\sim}{\mathbf{W}}_i$ in (3.1) becomes $\underset{\sim}{\Omega}_i'\underset{\sim}{\Omega}_i$, where $\underset{\sim}{\Omega}_i$ is the $J_n \times J_n$ matrix with elements (j,l) that are zero or are the same as the corresponding elements of Ω given in (4.2), subject to subject i not having or having observations at s_j and s_l, respectively.

Hu and Lagakos (2007) note that, when using the current weight w_{COI}, $\underset{\sim}{\bar{\mathbf{\Lambda}}}(w_{COI})$ is the nonparametric version of the estimator studied in Lawless and Nadeau (1995) and Lin et al. (2000) for semi-parametric estimation of the mean function $\Lambda_0(\cdot)$ from right-censored counting process data, which is an extension of the Nelson-Aalen estimator of the cumulative intensity function under the Poisson assumption.

The components of $\bar{\mathbf{\Lambda}}(w_{COI})$ are not in general in nondecreasing order when used with periodic observations. The estimator $\hat{\mathbf{\Lambda}}(w_{COI})$, which minimizes the second term of (3.2) with the given weight $\underset{\sim}{\mathbf{\Lambda}} \in \mathcal{C}_+^{J_n}$, is thus more desirable in the current setting. The matrix $B_n(w_{COI}) = \sum \underset{\sim}{\Omega}_i'\underset{\sim}{\Omega}_i$ is not in general diagonal, and thus $\hat{\Lambda}_n(\cdot; w_{COI})$ is a generalized isotonic regression of $\bar{\Lambda}_n(\cdot; w_{COI})$. We can use the procedures in Section 3.2.2 to compute $\hat{\mathbf{\Lambda}}(w_{COI})$ and then to obtain $\hat{\Lambda}_n(\cdot; w_{COI})$.

4.3. Generalized estimating equation weights (GEE)

Motivated by the construction of generalized estimating equations (see Diggle, Liang and Zeger (1994, Chap. 8)) for parametric estimation, consider the weight matrix $\mathbf{W}$ in (2.2) to be the inverse of the covariance matrix of $\mathbf{N} = \left(N(T_{K,1}), \ldots, N(T_{K,K})\right)'$, conditional on the number of observations K and the observation times $\mathbf{T} = \left(T_{K,1}, \ldots, T_{K,K}\right)'$; that is,

$$\mathbf{W}_{GSM} = \left(\text{Cov}\{N(T_{K,j}), N(T_{K,l}) | \mathbf{T}, K\}\right)_{K \times K}^{-1}. \qquad (4.3)$$

The weight matrices $\underset{\sim}{\mathbf{W}}_i$ in (3.1) for evaluating the estimator are now $(\Delta_i \text{Var}(\underset{\sim}{\mathbf{N}}) \Delta_i')^-$, where D^- denotes the Moore-Penrose generalized inverse of D, and $\text{Var}(\underset{\sim}{\mathbf{N}}) = (c(s_j, s_l))$. Another option uses the weight

$$\mathbf{W}_{GCI} = \Omega'\left(\text{Cov}\{\Delta N(T_{K,j}), \Delta N(T_{K,l}) | \mathbf{T}, K\}\right)_{K \times K}^{-1}\Omega, \qquad (4.4)$$

where Ω is defined in (4.2) and $\Delta N(T_{K,j}) = N(T_{K,j}) - N(T_{K,j-1})$. The weight matrices $\mathbf{W}_i$ in (3.1) are then $\mathbf{W}_{GCI,i} = \underset{\sim}{\Omega}_i{}' \left(\underset{\sim}{\Omega}_i \mathrm{Var}(\underset{\sim}{\mathbf{N}}) \underset{\sim}{\Omega}_i{}' \right)^{-} \underset{\sim}{\Omega}_i$.

Using either $\mathbf{W}_{GSM}$ in (4.3) or $\mathbf{W}_{GCI}$ in (4.4) requires the covariance matrix $\mathrm{Var}(\underset{\sim}{\mathbf{N}})$. When this is unknown, as would often occur in practice, it can be replaced by an estimate. Note that, with $\mathrm{Var}(\underset{\sim}{\mathbf{N}})$ substituted by the $J_n \times J_n$ identity matrix, $\mathbf{W}_{GSM}$ and $\mathbf{W}_{GCI}$ reduce to $\mathbf{W}_{OSM}$ (Sec. 4.1) and $\mathbf{W}_{COI}$ (Sec. 4.2), respectively. Below we consider two approximations to the covariance matrix $\mathrm{Var}(\underset{\sim}{\mathbf{N}})$ and the associated estimation procedures, and we compare the resulting estimators with the Wellner-Zhang NPMPLE and NPMLE.

Approximation A: using $\mathbf{Var}(\underset{\sim}{\mathbf{N}}) \approx \mathbf{diag}\big(\Lambda_0(s_j) : j = 1, \ldots, J_n\big)$

If the counting process $\{N(t), t > 0\}$ is Poisson, this approximation is the diagonal matrix of the true covariance matrix, and we plug it into either $\mathbf{W}_{GSM}$ (4.3) or $\mathbf{W}_{GCI}$ (4.4) to obtain a GEE type weight.

Use of this approximation in $\mathbf{W}_{GSM}$ (4.3) corresponds to the weight $\mathbf{W}_{GSM}^* = \mathrm{diag}(1/\Lambda_0(T_{K,j}) : j = 1, \ldots, K)$. Provided $E\{N(t)\} = \Lambda_0(t) > 0$ for $t > 0$, the weight matrix based on this approximation satisfies *Condition B* in Section 2. The matrix $B_n(w) = \nabla^2 \phi(\underset{\sim}{\mathbf{\Lambda}})$ is the diagonal matrix $B_n(w_{GSM}^*) = \mathrm{diag}(n_j/\Lambda_0(s_j) : j = 1, \ldots, J_n)$. Thus, if $\Lambda_0(\cdot)$ were known, the resulting estimator $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w_{GSM}^*)$ would be the isotonic regression of $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w_{GSM}^*)$ with weights $\{n_j/\Lambda_0(s_j) : j = 1, \ldots, J_n\}$. Note that $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w_{GSM}^*)$ is the same as $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$ given by (4.1), and also that $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$ is the isotonic regression of $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$ in Section 4.1, the same as $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w_{GSM}^*)$, with weights $\{n_j : j = 1, \ldots, J_n\}$. The hat estimator with the current weight $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w_{GSM}^*)$ is not, in general, the same as the estimator $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$, which corresponds to the estimator proposed by Sun and Kalbfleisch (1995).

Because the weight $\mathbf{W}_{GSM}^*$ involves $\Lambda_0(\cdot)$, we consider the following iterative algorithm to implement the estimator $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w_{GSM}^*)$: given the $(k-1)$th iteration $\underset{\sim}{\mathbf{\Lambda}}^{(k-1)}$, obtain the kth iteration for the estimator by using the weight $\mathbf{W}_{GSM}^*$ with Λ_0 substituted by $\underset{\sim}{\mathbf{\Lambda}}^{(k-1)}$, for $k = 1, \ldots$ until the sequence converges.

Appendix C shows that the procedure of applying the ICM algorithm (Lemma 3 in Appendix B) to maximize the pseudo-likelihood given in Wellner and Zhang (2000) gives the isotonic regression of $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w_{OSM})$ with weights $\{n_j \bar{\Lambda}_n(s_j; w_{OSM})/ \Lambda_0^2(s_j) : j = 1, \ldots, J_n\}$. As pointed out in Wellner and Zhang (2000), the pro-

cedure should converge to the Wellner-Zhang NPMPLE, which is the same as the Sun-Kalbfleisch estimator (1995); that is, $\hat{\mathbf{\Lambda}}(w_{OSM})$. Since $\bar{\Lambda}_n(s_j; w_{OSM})$ is a consistent estimator of $\Lambda_0(s_j)$, the resulting estimator from the iterative procedure is then close to the estimator $\hat{\mathbf{\Lambda}}(w^*_{GSM})$ for large n. This is confirmed for Poisson counts via simulation in Section 5.1.

Approximation B: using $\mathbf{Var}(\underset{\sim}{\mathbf{N}}) \approx (\sigma_{lj})$ with $\sigma_{lj} = \Lambda_0(s_j)$ for $j \leq l$: $j, l = 1, \ldots, J_n$

Applying the approximation in $\mathbf{W}_{GCI}$ (4.4), the weight in the estimation is $\mathbf{W}^*_{GCI} = \Omega^{'}\mathrm{diag}(1/\Delta\Lambda_0(T_{K,j}) : j = 1, \ldots, K)\Omega$. If $\{N(t), t > 0\}$ is a Poisson process, this approximation uses the true covariance matrix of the increments $\Delta N(T_{K,j}), j = 1, \ldots, K$, where $\mathrm{Cov}\{\Delta N(T_{K,j}), \Delta N(T_{K,l})|\mathbf{T}, K\}$ is $\Delta\Lambda_0(T_{K,j}) = \Lambda_0(T_{K,j}) - \Lambda_0(T_{K,j-1})$ or 0 for $l = j$ or not, respectively. The following additional condition, which holds in many practical situations, ensures that the weight $\mathbf{W}^*_{GCI}$ satisfy *Condition B* in Theorem 1.

Condition D: The observation times are ϵ-separated; that is, there exists a constant $\epsilon > 0$ such that $\mathrm{P}(T_{K,j} - T_{K,j-1} \geq \epsilon : j = 1, \ldots, K) = 1$.

It appears that Wellner and Zhang (2000) need this condition in the asymptotics derivation for their nonparametric maximum likelihood estimator (NPMLE).

Applying Lemma 3 (the ICM algorithm) in Appendix B, we can evaluate $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w^*_{GCI})$ as given in Section 3.2 with weight w^*_{GCI}, which is a generalized isotonic regression of the corresponding bar estimator $\bar{\underset{\sim}{\mathbf{\Lambda}}}(w^*_{GCI})$. Since the weight $\mathbf{W}^*_{GCI}$ involves $\Lambda_0(\cdot)$, an iterative algorithm similar to the one for $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w^*_{GSM})$ is needed to implement $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w^*_{GCI})$.

Appendix D shows the connection of $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w^*_{GCI})$ to the Wellner-Zhang NPMLE (2000). The difference is due to $\underset{\sim}{\mathbf{\Lambda}}$ in the matrix $B_n(w^*_{GCI})$ being fixed at the previous estimate in evaluating $\hat{\underset{\sim}{\mathbf{\Lambda}}}(w^*_{GCI})$, and is treated as unknown in evaluating the Wellner-Zhang NPMLE. Section 5 compares the two estimators for various situations via simulation.

5. Simulation Study

To study the properties of the estimators with the weights discussed in Section 4 in finite-samples, we conducted a simulation with $N(\cdot) = \{N(t), 0 < t \leq 1\}$ being a Poisson or mixed-Poisson process. We used $n = 100$ i.i.d. realizations of $N(\cdot)$, combined with one of the following two observation schemes, each chosen to yield an average of four observations per subject.

Observation Scheme A: Observation times for each individual were generated from a time-homogeneous Poisson process with rate of 4. This scheme simulates a study in which observation times vary among individuals.

Observation Scheme B: The potential observation times are $t_j = 0.05, 0.10$, ..., 0.95, 1.00, with the probability of an observation at time t_j being $1.05 - t_j^{1/4}$, and where the presence or absence of observations at the different times are independent. This observation scheme is intended to simulate a study with pre-scheduled observation times, but where different subjects can have missed visits and the risk of a missed visit increases as the study proceeds.

The following estimators were considered: $\hat{\Lambda}_n(\cdot; w)$ and $\bar{\Lambda}_n(\cdot; w)$, with the OSM, COI, and several GEE type weights. For the GEE type estimators in Section 4.3, we used the following weights: (i) $\mathbf{W}_{GSM}$ in (4.3) and $\mathbf{W}_{GCI}$ in (4.4) with the true covariance matrix, to show the best the GEE weights can achieve, (ii) $\mathbf{W}_{GSM}$ in (4.3) and $\mathbf{W}_{GCI}$ in (4.4) with the covariance matrix replaced by the sample covariance matrix obtained from a random sample with size of 30, denoted respectively by $\mathbf{W}_{GSM2}$ and $\mathbf{W}_{GCI2}$, and (iii) $\mathbf{W}_{GSM}^*$ and $\mathbf{W}_{GCI}^*$ in Section 4.3 with $\Lambda_0(\cdot)$ replaced by its current estimate at each stage in the algorithms.

With the OSM weights, $\hat{\Lambda}_n(\cdot; w)$ corresponds to the Sun-Kalbfleisch estimator (i.e., the Wellner-Zhang NPMPLE estimator), and is denoted by SK. We also computed the Wellner-Zhang NPMLE estimator, denoted by WZMLE.

Our primary program was written in C, and we used Splus functions *runif*, *rpois* and *rgamma* (the Splus generators of uniform, Poisson and Gamma random variables) to generate the random variables needed in the simulation. Iterations were terminated when the largest change in any component of the current estimate of $\Lambda_0(\cdot)$ from the previous estimate is below 10^{-5}. All simulations converged. The results reported in the following are based on 200 repetitions for each of the simulation settings.

5.1. Time-nonhomogeneous Poisson panel counts

The response process $\{N(t), t \in (0, 1]\}$ was generated as a time-nonhomogeneous Poisson process with $\Lambda_0(\cdot) = 6t^\gamma$, where γ was 1/2, 1, or 2. Coupled with the two observation schemes described above, this gave six simulation settings. In each setting, the sample means of the estimators were very close to the true mean functions over time, except occasionally in the right tail where there were fewer observations, confirming that the estimators studied are consistent.

By examining the simulation results, we have the following observations.

Finding 1: The bar estimators $\bar{\Lambda}_n(\cdot; w)$, obtained without a monotonicity constraint, are usually much easier to compute than the hat estimators $\hat{\Lambda}_n(\cdot; w)$.

However, in addition to ensuring monotonicity, the hat estimators $\hat{\Lambda}_n(\cdot; w)$ generally had smaller sample mean squared errors.

Finding 2: The COI weights (e.g., $\mathbf{W}_{COI}$ and $\mathbf{W}_{GCI*}$), based on observed increments of the response process, in general led to estimators with smaller sample mean squared errors than the estimators obtained by using the OSM weights (e.g., $\mathbf{W}_{OSM}$ and $\mathbf{W}_{GSM*}$). This also applies to the greater efficiency seen with the NPMLE estimator compared to the NPMPLE (i.e., SK) estimator, since these can be viewed as being based on observed increments and on observed responses, respectively. When the true or a sample covariance matrix is used in GEE type weights, the resulting COI and OSM estimators, which are the GSM and GCI or GSM2 and GCI2, are very similar.

Finding 3: By taking into account the covariance structure of the response process, the estimators using GEE weights generally have higher efficiency. However, when the covariance or its approximation involves unknown parameters, more computation time is needed to compute the estimates GSM* and GCI* than the corresponding ones with the OSM or COI weight.

The Wellner-Zhang NPMLE performed the best in all the simulated settings. All the evaluations of the SK estimator (i.e., $\hat{\Lambda}_n(\cdot; w)$ with weight $\mathbf{W}_{OSM}$) were very close to those of $\hat{\Lambda}_n(\cdot; w)$ with weight $\mathbf{W}_{GSM*}$. This numerically confirms that in Poisson situations, NPMPLE is the same as the Sun-Kalbfleisch estimator, that the ICM algorithm gives the NPMPLE, and that the GSM* is close to the NPMPLE. The estimates with weights $\mathbf{W}_{GSM}$ or $\mathbf{W}_{GCI}$, using the true covariance matrices $\text{Var}(\underset{\sim}{\mathbf{N}})$ or the sample covariance matrices, were similar, and had better efficiency than the corresponding estimates with $\mathbf{W}_{GSM*}$ or $\mathbf{W}_{GCI*}$. To illustrate the findings, Figure 1 presents the sample means and the sample mean squared errors (pointwise) of the estimates under nonhomogeneous Poisson process responses with $\gamma = 1/2$ and *Observation Scheme A*.

5.2. Mixed-Poisson panel counts

We next took $N(\cdot)$ to be a mixed-Poisson process with conditional intensity function $\lambda(t|\alpha) = 6\alpha t$, where the random effect α was a Gamma random variable with mean 1 and variance $\theta = 1, 2$ or 3. This corresponds to an unconditional mean function $\Lambda_0(t) = 6t$ and variance function $\text{Var}\{N(t)\} = \Lambda_0(t)\{1 + \text{Var}(\alpha)\Lambda_0(t)\}$, and a process with dependent increments (Lawless (1987)). The overdispersion of the simulated counting processes depends on the variance of the random effect $\theta = \text{Var}(\alpha)$. Coupled with the two observation schemes, the three choices of θ gave a total of six settings. Figure 2 presents the pointwise sample means and sample mean squared errors of the estimates evaluated with the data generated from the mixed-Poisson process for $\theta = 2$ with *Observation Scheme B*.
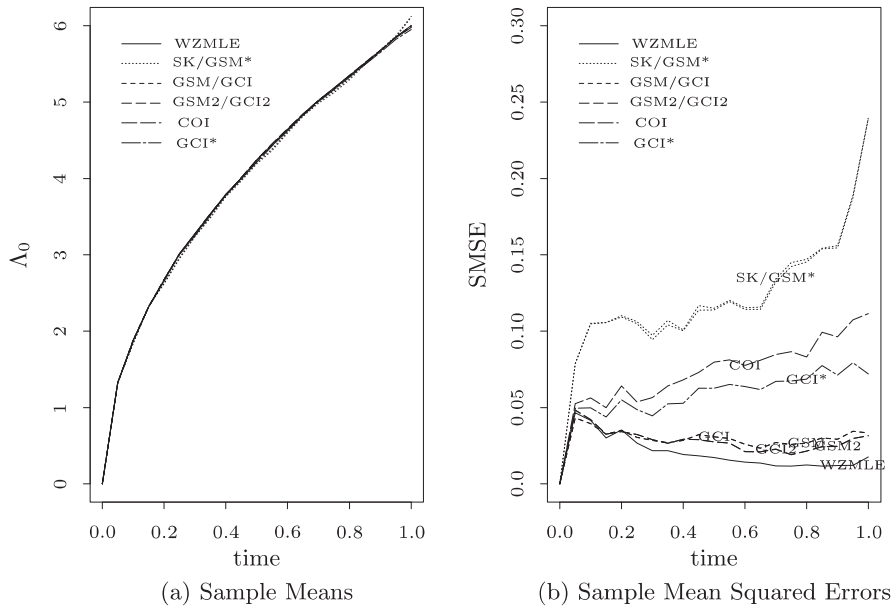
Figure 1. Sample means and sample mean squared errors of estimators with poisson panel counts under observation Scheme A.

Based on the simulation outcome, we had the same general observations with the mixed-Poisson processes as those listed as *Findings* 1−3 in Section 5.1 with time-nonhomogeneous Poisson processes. However, in this non-Poisson setting, the Wellner-Zhang NPMLE was no longer the most efficient estimator. As θ, a measure of the dependence of increments, increased, the proposed estimator $\hat{\Lambda}(\cdot; w)$ with weights $\mathbf{W}_{GSM}$ and $\mathbf{W}_{GCI}$, or with weights $\mathbf{W}_{GSM2}$ and $\mathbf{W}_{GCI2}$, showed higher efficiency. Moreover the estimate, with GEE weight and either of Approximation A or B, which uses a Poisson covariance matrix to approximate the true covariance, did not lead to much efficiency improvement over the OSM or COI estimates. This indicates the need to explore alternative approximations.

Unlike the situations in Section 5.1 with Poisson response, the evaluations of SK (that is, NPMPLE and $\hat{\Lambda}(\cdot; w)$ with weight $\mathbf{W}_{OSM}$), and GSM* (i.e., $\hat{\Lambda}_n(\cdot; w)$ with weight $\mathbf{W}_{GSM*}$) were not in close agreement. Although GSM* is consistent in theory, the sample means of the GSM* estimates indicate some bias, especially in the settings with *Observation Scheme B*. This may be due to fewer observations late in the study period.

6. Final Remarks

This paper presents a general class of estimators for the mean function of a counting process based on panel counts. Special cases include the estimator pro-

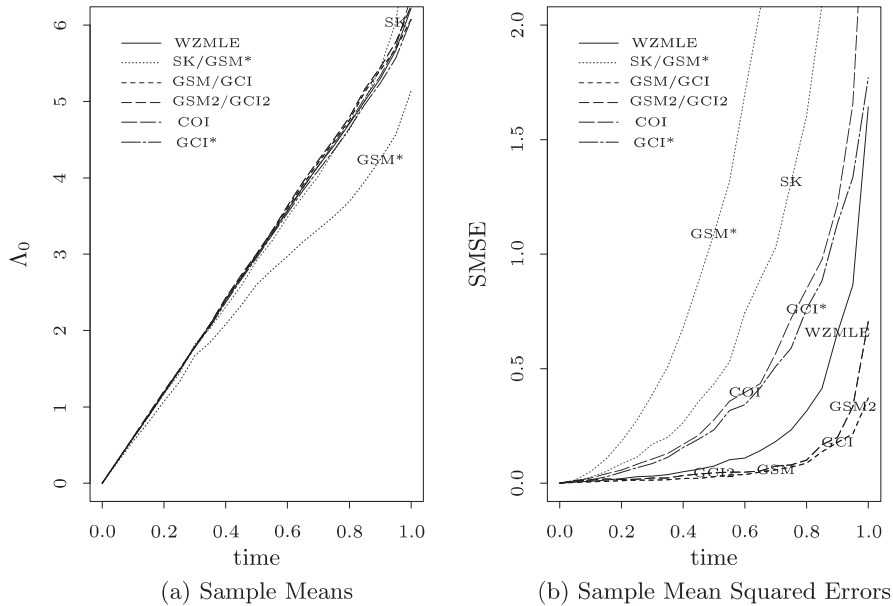(a) Sample Means   (b) Sample Mean Squared Errors

Figure 2. Sample means and sample mean squared errors of estimators with mixed-poisson panel counts ($\theta = 2$) under observation Scheme B.

posed by Sun and Kalbfleisch (1995), and estimators similar to the NPMLE estimator considered by Wellner and Zhang (2000). Simulations suggest that the use of GEE weights can lead to efficiency close to that of the Wellner-Zhang NPMLE for Poisson processes and better efficiency than the Wellner-Zhang NPMLE for non-Poisson processes. With modification, the proposed estimator can be generalized to accommodate nondecreasing processes with jumps that are not necessarily of size one.

Several further investigations would be worthwhile. One of theoretical and practical interest is the determination of asymptotic variance and variance estimation for the estimator $\hat{\Lambda}(\cdot; w)$. In principle, a resampling method could be used. Another is to find the optimal weight for the estimator to achieve the best efficiency in a given situation, and a third is to extend the estimator to the situations where the observation mechanism can depend on the response.

The estimator $\hat{\Lambda}_n(\cdot; w)$ is an M-estimator subject to the monotonicity constraint. As done by Wellner and Zhang (2000), we apply the iterative convex minorant algorithm (ICM) to implement the estimator. This may lead to extensive computation time, and thus faster algorithms would be useful. Finally, the methods can be readily extended to incorporate covariates which, among other things, can be used to assess the dependence of the observation and the response processes.

Acknowledgements

Appendix. Some Technical Details

A. Proof of Theorem 1

We first state a lemma to be used in the proof, then outline the proof of consistency. The approach utilized in the proof is similar to the one in Wellner and Zhang (2000). The following lemma is another version of the one-sided Glivenko-Cantelli Theorem given in Wellner and Zhang (2000) or Ferguson (1996, Sec. 17).

Lemma 1. *Suppose that $\mathcal{U} = \big\{ U(\cdot; \theta) : \theta \in \Theta \big\}$ is a class of measurable functions defined on a probability space $(\mathcal{X}, \mathcal{A}, P)$, where Θ is compact with respect to a metric d and $U(x; \theta)$ is lower semicontinuous in θ for all x. Suppose further that there exists a function $V(x)$ such that $EV(X) < \infty$ and $U(x; \theta) \leq V(x)$ for all $x \in \mathcal{X}$, $\theta \in \Theta$, and for all θ and all sufficiently small $\rho > 0$, $\inf_{\{\phi : d(\phi, \theta) < \rho\}} U(x; \phi)$ is measurable in x. Then if $X_1, \ldots, X_n$ are i.i.d. P with values in $\mathcal{X}$, and $\mathbb{P}_n$ is the empirical measure of the X_i's, almost surely P*

$$\varliminf_{n \to \infty} \inf_{\theta \in \Theta} \mathbb{P}_n U(X; \theta) \geq \inf_{\theta \in \Theta} PU(X; \theta),$$

$$\varliminf_{n \to \infty} \inf_{\theta \in \Theta} \big(\mathbb{P}_n - P\big) U(X; \theta) \geq 0.$$

Moreover, $PU(X; \theta)$ is lower semicontinuous in $\theta \in \Theta$: $\varliminf_{\phi \to \theta} PU(X; \phi) \geq PU(X; \theta)$.

Proof of Theorem 1. Denote by $M(\Lambda; w) = \mathrm{E}\big\{ m(\Lambda; w | \mathbf{X}) \big\}$ the limit of the objective function $L_n(\Lambda; w)/n = M_n(\Lambda; w)$ in (2.1).

We note first that $M(\Lambda; w) \geq M(\Lambda_0; w)$ for $\forall \Lambda \in \mathcal{F}$, and that the equality holds if and only if $\Lambda = \Lambda_0 \ \mu$, since

$$\mathrm{E}\Big\{ m(\Lambda; w | \mathbf{X}) - m(\Lambda_0; w | \mathbf{X}) \big| K, \mathbf{T} \Big\} = \big(\mathbf{\Lambda} - \mathbf{\Lambda}_0\big)^{'} \mathbf{W} \big(\mathbf{\Lambda} - \mathbf{\Lambda}_0\big) \geq 0,$$

and the equality holds if and only if $\Lambda(t) = \Lambda_0(t)$ for all $t = T_{K,j}$, since $\mathbf{W}$ is positive definite (*Condition B*).

We can then show that the sequence of nondecreasing functions $\big\{ \hat{\Lambda}_n(\cdot; w) \big\}$ is bounded in $[0, \tau]$ almost surely and, according to Helly's Selection Theorem, there

is a uniform subsequence of $\{\hat{\Lambda}_n(\cdot;w)(\omega)\}$ $\forall \omega$, denoted by $\{\hat{\Lambda}_{n^*(\omega)}(\cdot;w)(\omega)\}$, that converges. This follows by noting that

$$\varlimsup_{n\to\infty} \mathbb{P}_n \hat{\mathbf{\Lambda}}_n' \mathbf{W} \hat{\mathbf{\Lambda}}_n \leq 2\Big[\varlimsup_{n\to\infty} M_n(\Lambda_0;w|\mathbf{X}) + \varlimsup_{n\to\infty} \mathbb{P}_n \mathbf{N}' \mathbf{W} \mathbf{N}\Big]$$

$$= 2\Big[\mathrm{E}\big\{m(\Lambda_0;w|\mathbf{X})\big\} + \mathrm{E}\big\{\mathbf{N}'\mathbf{W}\mathbf{N}\big\}\Big] < \infty,$$

because $\hat{\Lambda}_n = \mathrm{argmin}_{\Lambda\in\mathcal{F}} M_n(\Lambda;w)$ and by *Conditions A* and *B*. Thus, for $\forall t \in [0,\tau]$,

$$\varlimsup_{n\to\infty} \hat{\Lambda}_n^2(t)\mu\big([t,\tau]\big) \leq \varlimsup_{n\to\infty} \mathbb{P}_n \sum_{j=1}^{K} \mathrm{I}_{[t,\tau]}(T_{K,j})\hat{\Lambda}_n^2(T_{K,j}) \leq \varlimsup_{n\to\infty} \mathbb{P}_n \hat{\mathbf{\Lambda}}_n' \hat{\mathbf{\Lambda}}_n$$

$$\leq \mathrm{E}\big\{\|A^{-1}\|^2\big\} \times \varlimsup_{n\to\infty} \mathbb{P}_n \hat{\mathbf{\Lambda}}_n' \mathbf{W} \hat{\mathbf{\Lambda}}_n$$

is bounded, by *Condition B*, where A is the matrix in the decomposition $\mathbf{W} = A'A$ in *Condition B*. Denote the bound by C, and let $\Lambda^*(\cdot;w)$ be the limit of the subsequence $\{\hat{\Lambda}_{n^*(\omega)}(\cdot;w)\}$. We know that $M(\Lambda^*;w) \geq M(\Lambda_0;w)$. We show that

$$M(\Lambda^*;w) \leq M(\Lambda_0;w), \tag{A.1}$$

and thus $M(\Lambda^*;w) = M(\Lambda_0;w)$, which implies $\Lambda^* = \Lambda_0$ a.e. μ, i.e., all the limits of subsequences of $\{\hat{\Lambda}_n(\cdot;w)\}$ are Λ_0 a.e. μ, and this proves the theorem.

Take

$$\mathcal{F}_\tau = \Big\{\Lambda \in \mathcal{F} : \Lambda(\tau) \leq \Big(\frac{C}{\mu(\{\tau\})}\Big)^{\frac{1}{2}} + 1\Big\},$$

compact in the metric d of (2.4), and $\mathcal{M}_\tau(w) = \big\{m(\Lambda;w|\mathbf{X}) : \Lambda \in \mathcal{F}\big\}$, a class of measurable functions indexed in $\mathcal{F}_\tau$ for a given weight function. We note that, for given w and $\mathbf{X}$, $m(\Lambda;w|\mathbf{X}) \in \mathcal{M}_\tau(w)$ is lower semicontinuous in $\Lambda \in \mathcal{F}_\tau$, since

$$m(\Lambda_1;w|\mathbf{X})^{\frac{1}{2}} \leq m(\Lambda_2;w|\mathbf{X})^{\frac{1}{2}} + \Big[(\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2)'\mathbf{W}(\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2)\Big]^{\frac{1}{2}}$$

and has an integrable envelope, since

$$m(\Lambda;w|\mathbf{X}) \leq 2\Big[\mathbf{N}'\mathbf{W}\mathbf{N} + \mathbf{\Lambda}'\mathbf{W}\mathbf{\Lambda}\Big] \leq 2\Big[\mathbf{N}'\mathbf{W}\mathbf{N} + K\|\Omega\|^2\frac{C}{\mu(\{\tau\})}\Big]$$

using *Conditions A* and *B*. Thus Lemma 1 yields

$$\varliminf_{n\to\infty} \inf_{\Lambda\in\mathcal{F}_\tau} \Big(\mathbb{P}_n - \mathrm{P}\Big)m(\Lambda;w|\mathbf{X}) \geq 0 \tag{A.2}$$

almost surely. Therefore,

$$\varliminf_{n^*\to\infty} \mathbb{P}_{n^*} m(\Lambda;w|\mathbf{X})\big|_{\Lambda=\hat{\Lambda}_{n^*}} \geq \lim_{n^*\to\infty} \mathrm{P}m(\Lambda;w|\mathbf{X})\big|_{\Lambda=\hat{\Lambda}_{n^*}} = M(\Lambda^*;w)$$

almost surely by the Dominated Convergence Theorem. Then (A.1) follows by noting

$$M(\Lambda_0; w) = \lim_{n^* \to \infty} M_{n^*}(\Lambda_0; w) \geq \lim_{n^* \to \infty} M_{n^*}(\hat{\Lambda}_{n^*}; w)$$

by the Strong Law of Large Numbers and the definition of $\hat{\Lambda}_n(\cdot; w)$.

B. Some known results

For the purpose of our application, we restate Theorem 2.1 of Wellner and Zhang (2000) or Lemma 3.1 of Wellner and Zhan (1997). Write $\nabla\phi$ for the gradient of ϕ, $\nabla^2\phi$ for $\nabla(\nabla\phi)'$, and $< \cdot, \cdot >$ for the usual inner product in $\mathcal{R}^J$.

Lemma 2. *Let* $\phi : \mathcal{R}^J \to \mathcal{R} \bigcup \{\infty\}$ *be a continuous convex function,* $\mathcal{K} \subset \mathcal{R}^J$ *be a convex cone, and* $\mathcal{K}_0 = \mathcal{K} \bigcap \phi^{-1}(\mathcal{R})$. *Suppose* $\mathcal{K}_0$ *is nonempty and* ϕ *is differentiable on* $\mathcal{K}_0$. *Then* $\hat{z} \in \mathcal{K}_0$ *satisfies* $\phi(\hat{z}) = \min_{z \in \mathcal{K}} \phi(z)$ *if and only if*

$$< \hat{z}, \nabla\phi(\hat{z}) >= 0 \ and \ < z, \nabla\phi(\hat{z}) >\geq 0 \ for \ all \ z \in \mathcal{K}.$$

It is easy to see that $\hat{z}$ is determined by $\nabla\phi(z)$. The following lemma presents the iterative convex minorant (ICM) algorithm as stated in Jongbloed (1998).

Lemma 3. *If the convex cone* $\mathcal{K}$ *is* $\mathcal{C}^J = \{y \in \mathcal{R}^J : y_1 \leq \ldots \leq y_J\}$, *the sequence* $\{z^{(k)} : k = 1, \ldots\}$ *converges to* $\hat{z} = argmin_{z \in \mathcal{K}}\phi(z)$ *where, for a fixed* $z^{(0)} \in \mathcal{C}^J$, *the lth component of* $z^{(k)}$ *is*

$$z_l^{(k)} = \max_{s \leq l} \min_{r \geq l} \left. \frac{\sum_{j=s}^r z_j \phi_{jj}(z) - \phi_j(z)}{\sum_{j=s}^r \phi_{jj}(z)} \right|_{z=z^{(k-1)}} , \quad l = 1, \ldots, J,$$

with $\phi_j(z) = \partial\phi(z)/\partial z_j$ *and* $\phi_{jj}(z) = \partial^2\phi(z)/\partial z_j^2$. *For* $\mathcal{K} = \mathcal{C}_+^J = \{y \in \mathcal{R}^J : 0 \leq y_1 \leq \ldots \leq y_J\}$, *the negative components of* $\hat{z}$ *should be set to zero.*

See Jongbloed (1998), for example, for a geometric interpretation of the algorithm. Jongbloed (1998) presents a modified iterative convex minorant algorithm to avoid the problem that the ICM algorithm does not necessarily converge globally.

C. Connection of $\hat{\Lambda}_n(w_{GSM}^*)$ with the Wellner-Zhang NPMPLE (2000)

Recall the nonparametric maximum pseudo-likelihood estimator (NPMPLE) considered in Wellner and Zhang (2000) is derived by maximizing

$$\log PL(\Lambda) = n\mathbb{P}_n(\mathbf{N}' \log \mathbf{\Lambda} - \mathbf{1}_K' \mathbf{\Lambda}) = \sum_{i=1}^n \sum_{t:\delta_i(t)=1} N_i(t)log\Lambda(t) - \Lambda(t),$$

where $\mathbf{1}_K$ is the K-dimensional vector with all components 1, and $\log z$ denotes $(\log z_1, \ldots, \log z_k)'$ for $z = (z_1, \ldots, z_k)'$. As indicated in Wellner and Zhang

(2000), $PL(\Lambda)$ is proportional to the pseudo-likelihood, assuming $\{N(t), t > 0\}$ is Poisson and ignoring the dependency of the events within a subject. Note that

$$\nabla \log PL(\Lambda) = \sum_{i=1}^{n} \Delta_i' \left\{ \underset{\sim}{\mathbf{W}}_i(\underset{\sim}{\boldsymbol{\Lambda}}) \Delta_i \underset{\sim}{\mathbf{N}}_i - \mathbf{1}_{J_n} \right\} = B_n(w_{GSM}^*; \underset{\sim}{\boldsymbol{\Lambda}}) \big(\underset{\sim}{\bar{\boldsymbol{\Lambda}}}_n(w_{OSM}) - \underset{\sim}{\boldsymbol{\Lambda}} \big),$$

$$\nabla^2 \log PL(\Lambda) = - \sum_{i=1}^{n} \Delta_i' \mathrm{diag}\big(N_i(s_j) : j = 1, \ldots, J_n\big) \underset{\sim}{\mathbf{W}}_i(\underset{\sim}{\boldsymbol{\Lambda}})^2,$$

with the matrix $\underset{\sim}{\mathbf{W}}_i(\underset{\sim}{\boldsymbol{\Lambda}}) = \mathrm{diag}\big(\delta_i(s_j)/\Lambda(s_j) : j = 1, \ldots, J_n\big)$ and

$$B_n(w_{GSM}^*; \underset{\sim}{\boldsymbol{\Lambda}}) = \sum_{i=1}^{n} \Delta_i' \underset{\sim}{\mathbf{W}}_i(\underset{\sim}{\boldsymbol{\Lambda}}) \Delta_i = \mathrm{diag}\big(\frac{n_j}{\Lambda(s_j)} : j = 1, \ldots, J_n\big).$$

Here $B_n(w_{GSM}^*; \Lambda_0) = B_n(w_{GSM}^*)$ as given in Section 4.3. The solution of $\nabla \log PL(\Lambda) = 0$ is the bar estimator with weight w_{OSM}, $\underset{\sim}{\bar{\boldsymbol{\Lambda}}}(w_{OSM})$. Thus the procedure of applying Lemma 3 in Appendix B with $\phi(\underset{\sim}{\boldsymbol{\Lambda}}) = - \log PL(\Lambda)$ to obtain the Wellner-Zhang NPMPLE is the procedure of getting the isotonic regression of $\underset{\sim}{\bar{\boldsymbol{\Lambda}}}(w_{OSM})$ with weights $\big\{ n_j \bar{\Lambda}_n(s_j)/\Lambda_0^2(s_j) : j = 1, \ldots, J_n \big\}$.

D. Connection of $\hat{\Lambda}_n(w_{GCI}^*)$ with the Wellner-Zhang NPMLE (2000)

The nonparametric maximum likelihood estimator (NPMLE) of Wellner and Zhang (2000) is from the likelihood function based on the current data with the Poisson assumption, the log-transformation of which is

$$\log FL(\Lambda) = n\mathbb{P}\big(\mathbf{N}' \Omega' \log \Omega \boldsymbol{\Lambda} - \Omega \boldsymbol{\Lambda}\big) = \sum_{i=1}^{n} \sum_{t: \delta_i(t) = 1} \Delta N_i(t) \log \Delta \Lambda_i(t) - \Delta \Lambda_i(t)$$

with $\Delta N_i(T_{K_i, j}) = N_i(T_{K_i, j}) - N_i(T_{K_i, j-1})$ for $j = 1, \ldots, K_i$. In our notation,

$$\nabla \log FL(\Lambda) = \sum_{i=1}^{n} \Delta_i' \underset{\sim}{\mathbf{W}}_i(\underset{\sim}{\boldsymbol{\Lambda}}) \Delta_i \underset{\sim}{\mathbf{N}}_i - \Delta_i' \underset{\sim}{\Omega}_i' \mathbf{1}_{J_n} = B_n(w_{GCI}^*; \underset{\sim}{\boldsymbol{\Lambda}}) \big(\underset{\sim}{\bar{\boldsymbol{\Lambda}}}(w_{GCI}^*; \underset{\sim}{\boldsymbol{\Lambda}}) - \underset{\sim}{\boldsymbol{\Lambda}} \big),$$

where $\underset{\sim}{\mathbf{W}}_i(\underset{\sim}{\boldsymbol{\Lambda}}) = \Omega_i' \mathrm{diag}\big(\delta_i(s_j)/e_{ij} : j = 1, \ldots, J_n\big) \Omega_i$, with $0/0 = 0$ and e_{ij} the jth component of $\underset{\sim}{\Omega}_i \Delta_i \underset{\sim}{\boldsymbol{\Lambda}}$, $B_n(w_{GCI}^*; \underset{\sim}{\boldsymbol{\Lambda}}) = \sum \Delta_i' \underset{\sim}{\mathbf{W}}_i(\underset{\sim}{\boldsymbol{\Lambda}}) \Delta_i$ with $B_n(w_{GCI}^*; \underset{\sim}{\boldsymbol{\Lambda}}_0)$ the same as the $B_n(w)$ matrix with weight w_{GCI}^*, and

$$\underset{\sim}{\bar{\boldsymbol{\Lambda}}}(w_{GCI}^*; \underset{\sim}{\boldsymbol{\Lambda}}) = B_n(w_{GCI}^*; \underset{\sim}{\boldsymbol{\Lambda}})^{-1} \sum_{i=1}^{n} \Delta_i' \underset{\sim}{W}_i(\underset{\sim}{\boldsymbol{\Lambda}}) \Delta_i \underset{\sim}{\mathbf{N}}_i.$$

We see that $\bar{\underset{\sim}{\Lambda}}(w^*_{GCI}; \underset{\sim}{\Lambda})$ is the solution of $\nabla \log FL(\Lambda) = 0$ with fixed $B_n(w^*_{GCI}; \underset{\sim}{\Lambda})$, which is the same as the bar estimator $\bar{\underset{\sim}{\Lambda}}(w^*_{GCI})$ if $B_n(w^*_{GCI}; \underset{\sim}{\Lambda})$ is evaluated at $\underset{\sim}{\Lambda} = \underset{\sim}{\Lambda}_0$.

Moreover, note that $B^*_n(\underset{\sim}{\Lambda}) = \nabla \big(\nabla \log FL(\Lambda) \big)'$ is

$$B_n(w^*_{GCI}; \underset{\sim}{\Lambda}) - \nabla B_n(w^*_{GCI}; \underset{\sim}{\Lambda}) \big(\bar{\underset{\sim}{\Lambda}}(w^*_{GCI}; \underset{\sim}{\Lambda}) - \underset{\sim}{\Lambda} \big)' - B_n(w^*_{GCI}; \underset{\sim}{\Lambda}) \nabla \bar{\underset{\sim}{\Lambda}}_n(w^*_{GCI}; \underset{\sim}{\Lambda})'.$$

Using Lemma 3 in Appendix B with $\phi(\underset{\sim}{\Lambda}) = -\log FL(\Lambda)$, the kth iteration for the NPMLE is the left derivative of the greatest convex minorant of the cumulative sum of diagram

$$\left. \Big(\sum_{r=1}^{j} a^*_{rr}(\underset{\sim}{\Lambda}), \sum_{r=1}^{j} b^*_r(\underset{\sim}{\Lambda}) \Big) \right|_{\underset{\sim}{\Lambda} = \underset{\sim}{\Lambda}^{(k-1)}}, \quad j = 1, \ldots, J_n,$$

with a^*_{rr} the (r,r) element of the matrix $B^*_n(\underset{\sim}{\Lambda})$, and $b^*_r(\underset{\sim}{\Lambda})$ the rth component of the vector

$$B_n(w^*_{GCI}; \underset{\sim}{\Lambda}) \bar{\underset{\sim}{\Lambda}}_n(w^*_{GCI}; \underset{\sim}{\Lambda}) + \big\{ \mathrm{diag}(B^*_n(\underset{\sim}{\Lambda})) - B_n(w^*_{GCI}; \underset{\sim}{\Lambda}) \big\} \underset{\sim}{\Lambda}.$$

The difference between $B^*_n(\underset{\sim}{\Lambda})$ and $B_n(w^*_{GCI}; \underset{\sim}{\Lambda})$ leads to the difference between $\hat{\underset{\sim}{\Lambda}}(w^*_{GCI})$ and the Wellner-Zhang NPMLE (2000).

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1992). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions: The Theory and Applications of Isotonic Regression.* Wiley, New York.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Clarendon Press, Oxford.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* Chapman & Hall, London.

Hu, X. J. and Lagakos, S. W. (2007). Nonparametric estimation of the mean function of a stochastic process with missing observations. *Lifetime Data Anal.* **13**, 51-73.

Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.* **28**, 161-183.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Lawless, J. F. (1987). Negative binomial regression models. *Canad. J. Statist.* **15**, 209-226.

Lawless, J. F. (1995). The analysis of recurrent events for multiple subjects. *Appl. Statist.* **44**, 487-498.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics* **37**, 158-168.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B* **62**, 711-730.

Sun, J. and Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica* **5**, 279-290.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J. Amer. Statist. Assoc.* **92**, 945-959.

Wellner, J. A. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779-814.

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

E-mail: joanh@stat.sfu.ca

Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, U.S.A.

E-mail: lagakos@biostat.harvard.edu

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.

E-mail: lockhart@stat.sfu.ca