Published in final edited form as:

J Nonparametr Stat. 2012 January 1; 24(3): 733–752. doi:10.1080/10485252.2012.698281.

Semiparametric Estimation with Recurrent Event Data under Informative Monitoring

Akim Adekpedjou[†] and Edsel A. Peña^{*}

[†]akima@mst.edu. Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla, MO 65409

Abstract

Consider a study where the times of occurrences of a recurrent event for n units are monitored. For the i th unit, $T_{i1}, T_{i2}, \dots$, denote the successive event interoccurrence times and this unit is monitored over a random period $[0, \tau_i]$ with τ_i independent of the T_{ij} s. Over this monitoring

period, $K_i = \max \left\{ k: \sum_{j=1}^k T_{ij} \leq \tau_i \right\}$ is the random number of event occurrences. The T_{ij} s are independent and identically distributed (IID) from an unknown continuous distribution function F and the τ_i s are IID from a distribution function G . A generalized Koziol-Green (GKG) structure wherein $1-G = (1-F)^\beta$ for some $\beta > 0$ is assumed. Under this model, Nelson-Aalen and product-limit type estimators of $\Lambda = -\log(1-F)$ and F are obtained, as well as an estimator of β .

Asymptotic and small-sample properties of these estimators are obtained and the estimator of F is compared to the fully nonparametric estimator in Peña et al. (2001) in terms of their finite-sample and asymptotic efficiency. The performance of the estimators of F are also examined when the GKG structure does not hold through a simulation study.

Keywords

Dependent and informative censoring; generalized Koziol-Green model; Nelson-Aalen estimator; mis-specified model; productlimit estimator; relative efficiency; right-censoring; sum-quota accrual

1. Introduction

In a variety of studies in the reliability/engineering, medical, biological, public health, sociological, and economic settings, the event of interest is recurrent. Examples are the failure of a machine, hospitalization of a patient, occurrence of a tumor, commission of a criminal act, and a change of job. It is of importance to have mathematical and stochastic models for recurrent event data, and the appropriate statistical inference methods for dealing with such models. See, for instance, Proschan (1963), Gill (1981), Sellke (1988), Aalen and Husebye (1991), Therneau and Hamilton (1997), Wang and Chang (1999), Hougaard (2000), Peña et al. (2001), Cook and Lawless (2002), Nelson (2003), Lindqvist (2006), Peña et al. (2007), Stocker and Peña (2007), the books by Cook and Lawless (2007) and Aalen et al. (2008) which contain several real recurrent event data sets, Gjessing et al. (2010), and Adekpedjou et al. (2010) for some works dealing with recurrent events.

In Peña et al. (2001) the nonparametric estimation of the inter-event time distribution when recurrent event data is available was considered. Their model has $\{T_i, n-1\}$ independent and identically distributed (IID) random variables whose common continuous distribution function is $F(t) = P\{T_1 \leq t\}$. The variable T_i could be viewed as the inter-event time between the $(i-1)$ th and i th event occurrences in an experimental unit in a reliability or engineering study or for a subject in a medical or public health study. The calendar time to the n th event

occurrence is $S_n = \sum_{j=1}^n T_j$, with $S_0 = 0$. In Peña et al. (2001) the unit is monitored over a random period $[0, \tau]$ where τ has a distribution function $G(t) = P\{\tau \leq t\}$. It is assumed that τ is independent of the T_j 's, and furthermore, there is no tacit relationship between F and G , only that they belong to the class of continuous distribution functions. In Peña et al. (2001) the data observables utilized in the estimation of F are n IID copies $D_1, D_2, \dots, D_n$ of $D = (K, T_1, T_2, \dots, T_K, \tau - S_K)$, where K is the random number of event recurrences observed over $[0, \tau]$, so that $K = \max\{k \in \{0, 1, 2, \dots\} : S_k \leq \tau\}$. Observe that $\tau - S_K$ is the right-censoring variable for T_{K+1} . Furthermore, note that the components of the random vector $(T_1, T_2, \dots, T_K, T_{K+1})$ are not anymore IID random variables from F ; in particular, the distribution of T_K is *not* anymore F . Indeed, the preceding vector satisfies the sum-quota constraint given by

$\sum_{j=1}^K T_j \leq \tau \leq \sum_{j=1}^{K+1} T_j$. Consequently, this recurrent event model has *both* an informative censoring mechanism as well as a dependent censoring structure. If these two features of the recurrent event data accrual are not properly accounted for in the inferential methods, then erroneous inferences regarding model parameters ensue, as was demonstrated analytically and through simulations in Peña et al. (2001) and Adekpedjou et al. (2010).

In this paper we consider the estimation of the distribution function F when there is an additional structure in the model through a relationship between F and G . The model of interest, which was introduced in Adekpedjou et al. (2010), is a generalization of the so-called Koziol-Green (KG) model introduced in Koziol and Green (1976). This model postulates that F and G are related according to $(1 - G) = (1 - F)^\beta$ for some $\beta > 0$. Henceforth, we shall refer to this as the GKG model. This equation specifies a relationship between the distribution of the inter-event times and the distribution of the length of the monitoring period. With $\Lambda = \Lambda_F = -\log(1 - F)$, the (cumulative) hazard function of F , the GKG condition is equivalent to $\Lambda G = \beta \Lambda_F$, a proportionality relation between the hazards of F and G , analogous to Lehmann-type alternatives in nonparametric rank-based theory. Through this more-structured model, which has only the two parameters F and β , we will be able to assess the loss in efficiency of the fully nonparametric estimator of F in Peña et al. (2001). The ability of this GKG model to serve as a mathematical specimen enabling analytical examination of methods for recurrent events is, perhaps, its *raison d'être* more than any potential it has as a practical model for recurrent events; see Adekpedjou et al. (2010) for discussion of this aspect. This situation is analogous to the single-event setting where the KG model, which is not a truly practical model; cf., Csörgö and Faraway (1998), facilitated the exact calculation of the bias and variance functions of the product-limit estimator (PLE) in Chen et al. (1982) thus enabling the assessment of the goodness of asymptotic approximations as well as the efficiency loss of the PLE when there is more structure in the right-censoring model; cf., Cheng and Lin (1987). In the GKG model, the parameter β was referred to in Adekpedjou et al. (2010) as the monitoring parameter since $\beta/(1 + \beta)$ can be interpreted as the approximate number of right-censored observations among all observations.

The outline of this paper is as follows. In section 2, we introduce relevant stochastic processes and develop the estimators of Λ , F , and β . The asymptotic properties of the estimators will be examined in section 3. Section 4 will examine the impact of the additional model structure in terms of the loss of efficiency of the fully nonparametric estimator of F in

Peña et al. (2001) relative to the estimator of F under the GKG model. In section 5 we provide results from a simulation study. Technical proofs are gathered in an appendix section.

2. Semiparametric Estimation

We begin by defining relevant stochastic processes. The *calendar* time processes are for $i = 1, 2, \dots, n, j = 1, 2, \dots$ given by $N_i^\dagger = \{N_i^\dagger(s) : s \geq 0\}$, $Y_i^\dagger = \{Y_i^\dagger(s) : s \geq 0\}$, and $N_i^\tau = \{N_i^\tau(s) : s \geq 0\}$, where $N_i^\dagger(s) = \sum_{j=1}^\infty I\{S_{ij} \leq s \wedge \tau_i\}$, $Y_i^\dagger(s) = I\{\tau_i \geq s\}$, and $N_i^\tau(s) = I\{\tau_i \leq s\}$. In the sequel, $T_{ij} = S_{ij} - S_{ij-1}$, the inter-event times or gap-times. The $N_i^\dagger$ process determines the event occurrences for the i th subject up to time τ_i ; the $Y_i^\dagger$ process is the at-risk process for the i th subject; while N_i^τ indicates whether the end of observation period has been reached. Note that $Y_i^\dagger(s) = 1 - N_i^\tau(s-)$.

Let $R_i = \{R_i(s) : s \geq 0\}$ be the backward recurrence time process for the i th subject, so that $R_i(s) = s - S_{iN_i^\dagger(s-)}$. Then, the compensator process of $N_i^\dagger$ is $A_i^\dagger = \{A_i^\dagger(s) : s \geq 0\}$ with $A_i^\dagger(s) = \int_0^s Y_i^\dagger(v) \lambda[R_i(v)] dv$. The $\lambda(\cdot)$ is the hazard rate function of F . The martingale process, with respect to the natural filtration $\mathcal{F}_s = \{\mathcal{F}_s : s \geq 0\}$ generated by $\{(N_i^\dagger(s), N_i^\tau(s), Y_i^\dagger(s)) : s \geq 0\}$, $i = 1, 2, \dots, n$, is $M_i^\dagger = \{M_i^\dagger(s) : s \geq 0\}$ with $M_i^\dagger(s) = N_i^\dagger(s) - A_i^\dagger(s)$. Following Peña et al. (2001) (see also Sellke (1988)), we define the doubly-indexed processes $N_i = \{N_i(s, t) : (s, t) \in [0, \infty)^2\}$, $A_i = \{A_i(s, t) : (s, t) \in [0, \infty)^2\}$, and $M_i = \{M_i(s, t) : (s, t) \in [0, \infty)^2\}$, where

$$\begin{aligned} N_i(s, t) &= \int_0^s Z_i(w, t) dN_i^\dagger(w) = \sum_{j=1}^{N_i^\dagger(s)} I(T_{ij} \leq t); \\ A_i(s, t) &= \int_0^s Z_i(w, t) dA_i^\dagger(w) = \int_0^t Y_i(s, w) \lambda(w) dw; \\ M_i(s, t) &= \int_0^s Z_i(w, t) dM_i^\dagger(w), \end{aligned}$$

with $Z_i(s, t) = I\{R_i(s) \geq t\}$ and $Y_i(s, t) = \sum_{j=1}^{N_i^\dagger(s-)} I\{T_{ij} \geq t\} + I\{(s \wedge \tau_i) - S_{iN_i^\dagger(s-)} \geq t\}$, the generalized at-risk process. Note that the components of $\{Y_i(s, t), i = 1, 2, \dots, n\}$ are IID processes. From Proposition 2 in Peña et al. (2001), with $G_s(w) = G(w)I(w < s) + I(w = s)$ and $\bar{G}_s = 1 - G_s$, we have

$$E\{Y_1(s, t)\} \equiv y(s, t) = \bar{F}(t) \bar{G}_s(t) + \bar{F}(t) \int_t^\infty \rho(w - t) dG_s(w), \quad (1)$$

where, with F^{*k} the k th convolution of F , $\rho(t) = \sum_{k=1}^\infty F^{*k}(t)$ is its renewal function. The aggregated processes will be denoted by $N(s, t) = \sum_{i=1}^n N_i(s, t)$, $Y(s, t) = \sum_{i=1}^n Y_i(s, t)$, and $M(s, t) = \sum_{i=1}^n M_i(s, t)$.

2.1. Estimating Λ and F

Let us first assume that β is known and consider the estimation of Λ and F . With $Y_i^\tau(s) = I\{\tau_i \geq s\}$, define the processes $A_i^\tau(s) = \int_0^s Y_i^\tau(w) \beta \lambda(w) dw$, $M_i^\tau(s) = N_i^\tau(s) - A_i^\tau(s)$,

and their respective aggregated processes $A^\tau(s) = \sum_{i=1}^n A_i^\tau(s)$ and $M^\tau(s) = \sum_{i=1}^n M_i^\tau(s) \equiv N^\tau(s) - A^\tau(s)$. With $J(s, w; \beta) = I\{Y(s, w) + \beta Y^\tau(w) > 0\}$, we have

$$\int_0^t \frac{J(s, w; \beta) \{M(s, dw) + M^\tau(dw)\}}{Y(s, w) + \beta Y^\tau(w)} = \int_0^t \frac{J(s, w; \beta) \{N(s, dw) + N^\tau(dw)\}}{Y(s, w) + \beta Y^\tau(w)} - \int_0^t J(s, w; \beta) \Lambda(dw).$$

Using Proposition 1 in Peña et al. (2001),

$$\int_0^t \frac{J(s, w; \beta) \{M(s, dw) + M^\tau(dw)\}}{Y(s, w) + \beta Y^\tau(w)} = \sum_{i=1}^n \int_0^s \frac{J(s, R_i(v); \beta) \{M_i(dv, t) + M_i^\tau(dv)\}}{Y(s, R_i(v)) + \beta Y^\tau(R_i(v))}.$$

Since $J(s, R_i(v); \beta)/[Y(s, R_i(v)) + \beta Y^\tau(R_i(v))]$ is bounded and predictable, it follows that

$$\left\{ \sum_{i=1}^n \int_0^s \frac{J(s, R_i(v); \beta) \{M_i(dv, t) + M_i^\tau(dv)\}}{Y(s, R_i(v)) + \beta Y^\tau(R_i(v))} : s \geq 0 \right\}$$

is a zero-mean square-integrable martingale for every fixed $t \geq 0$. Hence,

$$E \left\{ \int_0^t \frac{J(s, w; \beta) \{N(s, dw) + N^\tau(dw)\}}{Y(s, w) + \beta Y^\tau(w)} \right\} = E \left\{ \int_0^t J(s, w; \beta) \Lambda(dw) \right\}. \quad (2)$$

The moment identity in (2) motivates the Nelson-Aalen type estimator for $\Lambda(t)$, as in Aalen (1978). Under the initial assumption that we know β , an estimator of Λ is given by

$$\widehat{\Lambda}(s, \cdot; \beta) = \int_0^s J(s, w; \beta) \left\{ \frac{N(s, dw) + N^\tau(dw)}{Y(s, w) + \beta Y^\tau(w)} \right\} = \int_0^s \left\{ \frac{N(s, dw) + N^\tau(dw)}{Y(s, w) + \beta Y^\tau(w)} \right\}. \quad (3)$$

The induced product-limit estimator of $\bar{F} = 1 - F$ is then

$$\widehat{\bar{F}}(s, \cdot; \beta) = \prod_{w=0}^s \left\{ 1 - \widehat{\Lambda}(s, dw; \beta) \right\} = \prod_{w=0}^s \left\{ 1 - \frac{N(s, dw) + N^\tau(dw)}{Y(s, w) + \beta Y^\tau(w)} \right\}, \quad (4)$$

where ‘ $\prod$ ’ denotes product integration; cf., Andersen et al. (1993); Fleming and Harrington (1991); Gill and Johansen (1990). This estimator may also be viewed as a Kaplan-Meier type estimator of F in honor of Kaplan and Meier (1958).

However, the initial assumption that β is known is clearly implausible. The random functions in (3) and (4) are not valid estimators when β is unknown. To obtain estimators of Λ and F when β is unknown, we need to develop an estimator of β , say $\widehat{\beta}$. Upon obtaining such an estimator, it will be plugged-in for β in the aforementioned random functions to obtain legitimate estimators of Λ and F .

A referee has suggested an interesting generalization, which potentially could lead to improved estimators of Λ and F . In the developments above, the $N(s, \cdot)$ and $N^\tau(\cdot)$ are

weighted equally, but it appears that this is not necessary. The referee's suggestion amounts to considering as starting point in the derivation the more general martingale process given by

$$\left\{ \sum_{i=1}^n \int_0^s \frac{J^*(s, R_i(v); \beta) \{W(s, R_i(v)) M_i(dv, t) + W^\tau(R_i(v)) M_i^\tau(dv)\}}{W(s, R_i(v)) Y(s, R_i(v)) + W^\tau(R_i(v)) \beta Y^\tau(R_i(v))} : s \geq 0 \right\},$$

where $W(s, \cdot)$ and $W^\tau(\cdot)$ are bounded predictable weight functions and with $J^*(s, R_i(v); \beta) = I\{W(s, R_i(v)) Y(s, R_i(v)) + W^\tau(R_i(v)) \beta Y^\tau(R_i(v)) > 0\}$. We do agree with the referee that, indeed, this weighted approach could lead to better estimators with proper choices of the weight functions. However, we do not pursue this more general approach in this paper since such a more general estimator will also alter the estimator of β arising from the profile likelihood justification presented in the following subsection. See also the remark after Theorem 2.1 for another reason for not pursuing this more general estimator in this paper.

2.2. Estimating β

We now develop a profile likelihood for β whose maximizer will serve as our estimator for β . The likelihood function for (Λ, β) over the time period $[0, s^*]$ for some $s^* \in (0, \infty)$ is given by (cf., Jacod (1975))

$$\begin{aligned} L(s^*; \Lambda, \beta) &= \left\{ \prod_{i=1}^n \prod_{v=0}^{s^*} [Y_i^\dagger(v) \lambda[R_i(v)] dv]^{\Delta N_i^\dagger(v)} [Y_i^\tau(v) \beta \lambda(v) dv]^{\Delta N_i^\tau(v)} \right\} \\ &\times \exp \left\{ - \sum_{i=1}^n \left[\int_0^{s^*} Y_i^\dagger(v) \lambda[R_i(v)] dv + \int_0^{s^*} Y_i^\tau(v) \beta \lambda(v) dv \right] \right\}. \end{aligned} \quad (5)$$

The random function $\widehat{\Lambda}(\cdot; \beta)$ in (3) is a step-function with jumps only at the values of $\{T_{ij}\}$ s and $\{\tau_i\}$ s. Taking these two properties into consideration, when we plug-in this function $\widehat{\Lambda}(\cdot; \beta)$ for $\Lambda(\cdot)$ into the likelihood function in (5), and dropping terms not dependent on β , we get

$$\begin{aligned} L(s^*; \widehat{\Lambda}, \beta) &= \left[\prod_{i=1}^n \left\{ \prod_{j=1}^{N_i^\dagger(s^*)} [\Delta \widehat{\Lambda}(T_{ij}; \beta)]^{\Delta N_i^\dagger(s_{ij})} \right\} \left\{ \beta^{N_i^\dagger(s^*)} [\Delta \widehat{\Lambda}(\tau_i; \beta)]^{\Delta N_i^\tau(\tau_i)} \right\} \right] \times \\ &\exp \left\{ - \int_0^{s^*} \left[\sum_{i=1}^n (Y_i(s^*, w) + \beta Y_i^\tau(w)) \right] \widehat{\Lambda}(dw; \beta) \right\}, \end{aligned} \quad (6)$$

where we used the identity $\int_0^{s^*} Y_i^\dagger(v) \lambda[R_i(v)] dv = \int_0^{s^*} Y_i(s^*, w) \Lambda(dw)$ in the exponential function. Proceeding with the substitution, observe that the exponential function term simply becomes $\exp \left\{ - \int_0^{s^*} [N(s^*, dw) + N^\tau(dw)] \right\}$, which is independent of β . Also, observe that the amount of the jumps of $\widehat{\Lambda}(\cdot; \beta)$ are given by

$$\widehat{\Lambda}(s^*, \Delta T_{ij}; \beta) = \frac{N(s^*, \Delta T_{ij})}{Y(s^*, T_{ij}) + \beta Y^\tau(T_{ij})} \quad \text{and} \quad \Delta \widehat{\Lambda}(\tau_i; \beta) = \frac{\Delta N^\tau(\tau_i)}{Y(s^*, \tau_i) + \beta Y^\tau(\tau_i)}.$$

Thus, substituting these into the first two terms in (6), we obtain the profile likelihood for β given by

$$L_p(s^*; \beta) = \beta^{N^\tau(s^*)} \times \prod_{i=1}^n \left\{ \prod_{v=0}^{s^*} \left\{ \frac{1}{Y(s^*, v) + \beta Y^\tau(v)} \right\}^{N_i^\tau(\Delta v)} \right\} \left[\prod_{v=0}^{s^*} \left\{ \frac{1}{Y(s^*, v) + \beta Y^\tau(v)} \right\}^{N_i(s^*, \Delta v)} \right]. \quad (7)$$

The estimator $\widehat{\beta}$ of β is the maximizer of this profile likelihood function. The logarithm of this function is given by

$$l_p(s^*; \beta) = - \int_0^{s^*} \{ \log [Y(s^*, v) + \beta Y^\tau(v)] \} [N^\tau(dv) + N(s^*, dv)] + N^\tau(s^*) \log \beta.$$

By taking the derivative with respect to β and equating to zero, the resulting estimating equation for $\widehat{\beta}$ is given by

$$\int_0^{s^*} \left\{ \frac{\beta Y^\tau(v)}{Y(s^*, v) + \beta Y^\tau(v)} \right\} [N^\tau(dv) + N(s^*, dv)] = N^\tau(s^*).$$

Observe that the integral in the left-hand side is just a finite sum over the $\tau_j s$ and $T_j s$.

Nevertheless, obtaining $\widehat{\beta}$ from this estimating equation requires iterative methods such as the Newton-Raphson procedure. With $\widehat{\beta}$ at hand, we finally obtain legitimate estimators of $\Lambda(\cdot)$ and $\bar{F}(\cdot)$ even when β is unknown, given respectively by

$$\widehat{\Lambda}(s^*; \cdot) = \int_0^{s^*} \left\{ \frac{N(s^*, dw) + N^\tau(dw)}{Y(s^*, w) + \widehat{\beta} Y^\tau(w)} \right\} \quad \text{and} \quad \widehat{\bar{F}}(s^*; \cdot) = \prod_{w=0}^{s^*} \left\{ 1 - \frac{N(s^*, dw) + N^\tau(dw)}{Y(s^*, w) + \widehat{\beta} Y^\tau(w)} \right\}, \quad (8)$$

which are Nelson-Aalen (NA) type and product-limit (PL) type estimators.

A referee has also perceptively wondered whether the profile likelihood above used in estimating the β is *truly* a profile likelihood in the sense that it satisfies

$$L_p(s^*; \beta) = \max_{\Lambda \in \mathfrak{C}} L(s^*; \Lambda, \beta) \quad (9)$$

where $\mathfrak{C}$ is a family of cumulative hazard functions. We establish that this is indeed the case with $\mathfrak{C}$ being the space of all cumulative hazard functions.

Theorem 2.1 Let $\mathfrak{C}$ be the collection of all cumulative hazard functions on $[0, \infty)$. Then (9) holds. In addition, the maximizing $\Lambda \in \mathfrak{C}$ is the $\widehat{\Lambda}(\cdot; \beta)$ defined in (3).

Proof : Assume that β is fixed. Then the relevant portion of the likelihood function in (5) that involves $\Lambda(\cdot)$ is given by

$$\tilde{L}(s^*; \Lambda) = \left[\prod_{i=1}^n \prod_{v=0}^{s^*} \left\{ [d\Lambda(R_i(v))]^{N_i^\tau(dv)} [\Lambda(dv)]^{N_i^\tau(dv)} \right\} \right] \times \exp \left\{ - \int_0^{s^*} [Y(s^*, w) + \beta Y^\tau(w)] \Lambda(dw) \right\}.$$

To maximize this function over $\mathfrak{C}$, we must choose a $\widehat{\Lambda} \in \mathfrak{C}$ with $d\widehat{\Lambda}(T_{ij}) > 0$ and $d\widehat{\Lambda}(\tau_i) > 0$ for all $j=1, 2, \dots, N_i^+(s^*)$ and $i=1, 2, \dots, n$. Furthermore, since $\Lambda(\cdot)$ s are nondecreasing functions, we must take $\widehat{\Lambda}(\cdot)$ to be a step-function. Thus, the maximizing $\widehat{\Lambda} \in \mathfrak{C}$ is a step-function whose jumps occur only at the T_{ij} s and τ_i s. Denoting by $\{q_k, k=1, 2, \dots, Q\}$ the distinct values of T_{ij} s with $S_{ij} = s^*$ and τ_i s with $\tau_i = s^*$, we must have the maximizing $\widehat{\Lambda}$ satisfying $\widehat{\Lambda}(t) = \sum_{\{k: q_k \leq t\}} \widehat{\gamma}_k$, where $\widehat{\gamma}_k$ is the positive jump at q_k . Also, let d_k be the number of T_{ij} s and τ_i s exactly equal to q_k . Then, the relevant portion of the likelihood evaluated at the $\widehat{\Lambda}$ is given by

$$\tilde{L}(s^*; \widehat{\Lambda}) = \left\{ \prod_{k=1}^Q \widehat{\gamma}_k^{d_k} \right\} \exp \left\{ - \sum_{k=1}^Q [Y(s^*, q_k) + \beta Y^\tau(q_k)] \widehat{\gamma}_k \right\}.$$

It follows that for each $k=1, 2, \dots, Q$, $\widehat{\gamma}_k \equiv \widehat{\gamma}_k(\beta) = d_k / [Y(s^*, q_k) + \beta Y^\tau(q_k)]$. Consequently, using the definition of the d_k s, we have that

$$\widehat{\Lambda}(t; \beta) = \sum_{\{k: q_k \leq t\}} \frac{d_k}{Y(s^*, q_k) + \beta Y^\tau(q_k)} = \int_0^t \frac{N(s^*, dw) + N^\tau(dw)}{Y(s^*, w) + \beta Y^\tau(w)},$$

which is the estimator (3) when β is known. From the derivation of L_P above, it therefore follows that (9) holds. $\parallel$

We remark that this result is also one of our motivations for not considering the more general weighted-type estimators of $\Lambda(\cdot)$, since the resulting estimators we ultimately obtained can be viewed as nonparametric maximum likelihood estimators. Using a weighted-type estimator of $\Lambda(\cdot)$ will lead to general M -type estimators. This certainly is an interesting open problem and calls for further study.

3. Asymptotic Properties

In this section we present asymptotic properties of the estimators of β , Λ , and F . To improve readability and flow of presentation, the proofs of the theorems and corollary in remaining sections are all provided in the Appendix. In the development of our asymptotic results, the sample size n , which represents the number of units, is increasing to infinity. The value $s^* \in (0, \infty)$ is assumed fixed, though we may let this also depend on n . The estimator of β , $\widehat{\beta}_n$ based on the n units will then be based on the profile likelihood $L_P(s^*; \beta)$ associated with the n units and over the $[0, s^*]$. Unless confusion may arise, we suppress showing the dependence on n of our profile likelihood, profile score function, and profile information function. Thus, we will simply write $L_P(s^*; \beta)$, $l_P(s^*; \beta) = \log L_P(s^*; \beta)$, $U_P(s^*; \beta) = l_P(s^*; \beta) / \beta$, and $I_P(s^*; \beta) = -^2 l_P(s^*; \beta) / \beta^2$ for the profile likelihood, log-likelihood, score, and information functions, respectively, where $L_P(s^*; \cdot)$ is defined in (7), but noting that these functions depend on the sample size n . We first list the assumed regularity conditions needed in the theorems, lemmas, and corollary.

- I. The GKG model holds with true parameter values F_0 , a continuous distribution function with associated hazard function Λ_0 , and $\beta_0 \in (0, \infty)$;
- II. Let $s^* \in (0, \infty)$ with $0 < \Lambda_0(s^*)$ and $y_0^\tau(s^*) \equiv y^\tau(s^*; \Lambda_0, \beta_0) > 0$, with

$$y^{\tau}(s; \Lambda, \beta) = \exp\{-\beta \Lambda(s)\} = \bar{F}(s)^{\beta};$$

III. Let $t^* \in (0, \infty)$ with $\Lambda_0(t^*) < \infty$ and $y_0(s^*, t^*) \equiv y(s^*, t^*; \Lambda_0, \beta_0) > 0$, where $y(s, t; \Lambda, \beta) \equiv y(s, t)$ with $y(s, t)$ defined in (1) with $\bar{F}(t) = \exp\{-\Lambda(t)\}$ and $\bar{G}(t) = \exp\{-\beta \Lambda(t)\}$.

The first two results deal with the consistency of the estimators.

Theorem 3.1 For the sequence of estimating equations for β given by

$$\left(U_p(s^*; \beta) = \frac{\partial}{\partial \beta} l_p(s^*; \beta) = 0, n = 1, 2, 3, \dots \right),$$

there exists a sequence of solutions $(\hat{\beta}_n \equiv \hat{\beta}_n(s^*), n = 1, 2, 3, \dots)$ such that, under (F_0, β_0) , $\hat{\beta}_n$ converges in probability to β_0 .

Theorem 3.2 As $n \rightarrow \infty$, $\sup_{t \in [0, t^*]} |\hat{\Lambda}(s^*, t) - \Lambda_0(t)| \xrightarrow{pr} 0$ and $\sup_{t \in [0, t^*]} |\hat{F}(s^*, t) - \bar{F}_0(t)| \xrightarrow{pr} 0$.

Next, we provide results pertaining to the weak convergence properties of the estimators. Theorem 3.3 deals with asymptotic normality of the estimator of β . Theorem 3.4 deals with the weak convergence of $\sqrt{n} [\hat{\Lambda}(s^*, \cdot) - \Lambda_0(\cdot)]$ to a Gaussian process, while Corollary 3.5 states the asymptotic Gaussianity of the estimator of F .

Theorem 3.3 As $n \rightarrow \infty$, $\sqrt{n}(\hat{\beta} - \beta_0) \Rightarrow N(0, [\mathcal{J}_p(s^*; \Lambda_0, \beta_0)]^{-1})$, where

$$\mathcal{J}_p(s^*; \Lambda_0, \beta_0) = \frac{1}{\beta_0} \int_0^{s^*} \frac{y_0^{\tau}(v) y_0(s^*, v)}{y_0(s^*, v) + \beta_0 y_0^{\tau}(v)} \Lambda_0(dv).$$

Theorem 3.4 As $n \rightarrow \infty$, $\left\{ \sqrt{n} [\hat{\Lambda}(s^*, t) - \Lambda_0(t)] : t \in [0, t^*] \right\}$ converges weakly to a zero-mean Gaussian process with variance function defined via

$$\sigma_{\hat{\Lambda}}^2(s^*, t) = \int_0^t \frac{\Lambda_0(dv)}{y_0(s^*, v) + \beta_0 y_0^{\tau}(v)} + \left[\int_0^{s^*} \frac{y_0(s^*, v) y_0^{\tau}(v)}{\beta_0 [y_0(s^*, v) + \beta_0 y_0^{\tau}(v)]} \Lambda_0(dv) \right]^{-1} \left[\int_0^t \frac{y_0^{\tau}(v)}{y_0(s^*, v) + \beta_0 y_0^{\tau}(v)} \Lambda_0(dv) \right]^2.$$

The result concerning the estimator of the survivor function $\bar{F}$ in Corollary 3.5 is a direct consequence of Theorem 3.4 and the Functional Delta Method; cf., pages 109–114 of Andersen et al. (1993).

Corollary 3.5 As $n \rightarrow \infty$, $\left\{ \sqrt{n} [\hat{F}(s^*, t) - \bar{F}_0(t)] : t \in [0, t^*] \right\}$ converges weakly to a zero-mean Gaussian process whose variance function is $\sigma_{\hat{F}}^2(s^*, t) = \bar{F}_0(t)^2 \sigma_{\hat{\Lambda}}^2(s^*, t)$.

4. Efficiency Comparisons

In the absence of a relationship between F and G for the recurrent event model, Peña et al. (2001) obtained an estimator of $\bar{F}(t)$ given by

$$\tilde{\bar{F}}(s^*, t) = \prod_{w=0}^t \left[1 - \frac{N(s^*, dw)}{Y(s^*, w)} \right]. \quad (10)$$

They established that this estimator is asymptotically a Gaussian process with mean function $\bar{F}_0(t)$ and asymptotic variance function

$$AVar \left\{ \tilde{\bar{F}}(s^*, t) \right\} = \frac{\bar{F}_0(t)^2}{n} \int_0^t \frac{\Lambda_0(dw)}{y_0(s^*, w)} \equiv \frac{\bar{F}_0(t)^2}{n} \sigma_{\bar{F}}^2(s^*, t).$$

One of our main goals in introducing the GKG model for this recurrent event setting is to be able to examine the efficiency of the fully nonparametric estimator of F in (10) when viewed in light of a more structured model. Under this GKG model, and first assuming that the

parameter β_0 is known, the estimator of $\bar{F}$ is $\widehat{\bar{F}}(s^*, t; \beta_0)$ in (4), with β replaced by the true value β_0 . This estimator is asymptotically a Gaussian process with mean function $\bar{F}_0(t)$ and asymptotic variance function

$$AVar \left\{ \widehat{\bar{F}}(s^*, t; \beta_0) \right\} = \frac{\bar{F}_0(t)^2}{n} \int_0^t \frac{\Lambda_0(dw)}{y_0(s^*, w) + \beta_0 y_0^T(w)}.$$

A measure of the asymptotic relative efficiency (ARE) of $\tilde{\bar{F}}(s^*, t)$ relative to $\widehat{\bar{F}}(s^*, t; \beta_0)$ is the ratio of their asymptotic variances, that is,

$$ARE \left\{ \tilde{\bar{F}}(s^*, t) : \widehat{\bar{F}}(s^*, t; \beta_0) \right\} = \left\{ \int_0^t \frac{\Lambda_0(dw)}{y_0(s^*, w)} \right\}^{-1} \left\{ \int_0^t \frac{\Lambda_0(dw)}{y_0(s^*, w) + \beta_0 y_0^T(w)} \right\}. \quad (11)$$

Since $y_0(s^*, w)$ and $y_0^T(w)$ are both non-negative, then it immediately follows from the

expression in (11) that $ARE \left\{ \tilde{\bar{F}}(s^*, t) : \widehat{\bar{F}}(s^*, t; \beta_0) \right\} \leq 1$, that is, asymptotically the fully nonparametric estimator cannot be more efficient than the estimator derived under the GKG model under the assumption that the parameter β_0 is known. To obtain a more concrete result regarding the efficiency expression, we take the case where $\bar{F}_0$ is an exponential distribution and we let $s^* \rightarrow \infty$.

Theorem 4.1 Assume the GKG model holds with parameter vector (F_0, β_0) . If $\bar{F}_0(t) =$

$\exp\{-\theta_0 t\}$ for $t \geq 0$ and $s^* \rightarrow \infty$, then $ARE \equiv ARE \left\{ \tilde{\bar{F}}(\infty, t) : \widehat{\bar{F}}(\infty, t; \beta_0) \right\}$ is

$$ARE = \left\{ \int_{\bar{F}_0(t)}^1 \frac{du}{(1+\beta_0)u^{2+\beta_0}} \right\}^{-1} \left\{ \int_{\bar{F}_0(t)}^1 \frac{du}{(1+\beta_0)u^{2+\beta_0} + \beta_0^2 u^{1+\beta_0}} \right\}.$$

In addition, for every $t > 0$, $ARE \left\{ \tilde{F}(\infty, t) : \hat{F}(\infty, t; \beta_0) \right\} \leq (1 + \beta_0) / (1 + \beta_0 + \beta_0^2)$.

The preceding efficiency result, however, is still to be expected since we are comparing the fully nonparametric estimator with the estimator developed under the assumption of β_0 being known. It is of more interest to see if the fully nonparametric estimator is still dominated by the estimator under the GKG model derived under the situation where both F_0 and β_0 are unknown. Clearly, it is not anymore intuitive that we should expect full domination since the estimation of the parameter β_0 will have the resultant effect of also increasing the asymptotic variance of the estimator of $\tilde{F}$. The question is whether, *even* with this increase in variance, the estimator $\hat{F}(s^*, t)$ given in (8) still dominates the fully nonparametric estimator $\tilde{F}(s^*, t)$ when the GKG model holds. The next theorem establishes that this is still indeed the case!

Theorem 4.2 Under the GKG model, for all $(\bar{F}_0, \beta_0)$ with $\beta_0 > 0$, $\tilde{F}(s^*, t)$ is asymptotically dominated by $\hat{F}(s^*, t)$ in the sense that $ARE \left(\tilde{F}(s^*, t) : \hat{F}(s^*, t) \right) \leq 1$.

To acquire a more concrete knowledge of the degree of asymptotic efficiency loss of the fully nonparametric estimator $\tilde{F}(s^*, t)$ relative to $\hat{F}(s^*, t)$, we consider again the case where $\bar{F}_0(t) = \exp(-\theta_0 t)$ for $t > 0$ and with $s^* \rightarrow \infty$. As the ARE is the ratio between $\sigma_{\tilde{\Lambda}}^2(s^*, t)$ and $\sigma_{\hat{\Lambda}}^2(s^*, t)$, we first obtained expressions of these variance functions under the above specifications. Straightforward calculations and expressing the results in terms of β_0 and $\bar{F}_0(t)$, we find that

$$\sigma_{\tilde{\Lambda}}^2(t) = \frac{\beta_0}{1 + \beta_0} \int_{\bar{F}_0(t)}^1 \frac{du}{u^{2 + \beta_0}} = \frac{\beta_0}{(1 + \beta_0)^2} \left[\frac{1 - \bar{F}_0(t)^{1 + \beta_0}}{\bar{F}_0(t)^{1 + \beta_0}} \right]; \quad (12)$$

$$\sigma_{\hat{\Lambda}}^2(t) = \beta_0 \left\{ \int_{\bar{F}_0(t)}^1 \frac{du}{u^{1 + \beta_0} [(1 + \beta_0)u + \beta_0^2]} + \left(\frac{\beta_0^2}{1 + \beta_0} \right) \left[\int_0^1 \frac{u^{\beta_0} du}{(1 + \beta_0)u + \beta_0^2} \right]^{-1} \left[\int_{\bar{F}_0(t)}^1 \frac{du}{[(1 + \beta_0)u + \beta_0^2]} \right]^2 \right\}. \quad (13)$$

The integrals in (13) are not in closed-forms so we obtained their values using the integrate object function in the R Library. From the expressions in (12) and (13), we then obtained the asymptotic relative efficiency of $\tilde{F}(\infty, t)$ relative to $\hat{F}(\infty, t)$ under this exponential inter-

event time model via $ARE \left(\tilde{F}(\infty, t) : \hat{F}(\infty, t) \right) = \sigma_{\tilde{\Lambda}}^2(t) / \sigma_{\hat{\Lambda}}^2(t)$. These ARE-values are plotted in Figure 1 for different values of $\bar{F}_0(t) = \exp(-\theta_0 t)$ and different values of β_0 . Observe that as β_0 increases, the efficiency of the fully nonparametric estimator generally decreases since in this situation the number of events observed per unit tends to decrease, hence $\tilde{F}(\infty, t)$ is deprived of data, while the estimator $\hat{F}(\infty, t)$ is still able to exploit the information contained in the observed $\tau_{\mathcal{F}}$. Note that when β_0 is large, there could be a great loss in efficiency of the fully nonparametric estimator. Furthermore, observe that none of the ARE-values exceeded unity, consistent with the result of Theorem 4.2. An interesting observation also from the

ARE-plots is that, in contrast to the case where β_0 is known, the maximum ARE-value does not anymore occur at the limiting case of $\bar{F}_0(t) \rightarrow 1$. It may be of some mathematical interest to determine, in terms of $\bar{F}_0(t)$ and β_0 , where this maximum occurs, but we leave this as an exercise to the interested reader.

5. Simulation Studies

5.1. When GKG Model Holds

The asymptotic variances of $\hat{\beta}$ and $\hat{\bar{F}}$ both depend on the renewal function of F_0 . Except for the case where F_0 is an exponential distribution or is a gamma distribution with shape parameter 2, there is no explicit closed-form expression for the renewal function; cf., Resnick (1994). As such there is generally no explicit closed-form expression for the

asymptotic variances of $\hat{\beta}$ and $\hat{\bar{F}}$. In order to study the properties of the estimators for non-exponential distributions as well as to also examine their small- to moderate-sample size properties, computer simulation studies were therefore performed. The goals of the numerical study are (i) to examine the impact of sample size and β on the properties of the estimators; (ii) to examine the bias and mean-squared error of the estimators; (iii) to examine the performance of the semiparametric estimator of $\bar{F}$ relative to the fully nonparametric estimator in Peña et al. (2001), and (iv) to assess the quality of the asymptotic approximations to the standard errors.

In the simulation we considered the cases where $\bar{F}_0$ is exponential and Weibull. The scale parameter utilized in both cases was unity, while for the Weibull distribution the shape parameter α were set equal to 0.9 and 2.0. The β -values considered were 0.1, 0.3, 0.5, 0.7, 0.9, 1.0, and 1.5. The sample size n were set to 30, 50, and 100. Each simulation had 10000 replications. In computing the estimate $\hat{\beta}$, we implemented a two-step approach. The *R* object `optimize` was used to obtain a preliminary estimate of $\hat{\beta}$ (see Ihaka and Gentleman (1996)), and this estimate was then fed as seed value to a Newton-Raphson (NR) iteration to obtain the final estimate $\hat{\beta}$. We observed that such an approach eliminates the problem of having seed values for the NR iteration which are too far off from the solution. By using the seed value from the `optimize` step, the NR iteration also converged quickly. The simulation program was coded entirely in *R*.

Table 1 presents the simulated means and standard errors of $\hat{\beta}$ under an exponential $\bar{F}_0$ for different values of β_0 and n . We also provided the values of the asymptotic standard error (ASE) of $\hat{\beta}$, which are obtained by taking the square root of

$$AVar(\hat{\beta}) = \frac{1}{n} \frac{\beta_0}{1+\beta_0} \left[\int_0^1 \frac{u^{\beta_0}}{(1+\beta_0)u + \beta_0^2} du \right]^{-1}.$$

We observe from this table that as the sample size increases, the simulated means of $\hat{\beta}$ for the different values of β_0 get closer to the respective values of β_0 . We also notice that when β_0 is large then there is more discrepancy between the simulated mean of $\hat{\beta}$ and β_0 . This could be partly attributed to the fact that there are less event recurrences compared to when β_0 is smaller. The agreement between the simulated standard error and the asymptotic standard error also improves as the sample size increases and when β_0 decreases.

Figure 2 provides simulated efficiency plots of the estimator $\tilde{F}$ versus the estimator $\hat{F}$ under an exponential F_0 for $n = 30$ and different values of β_0 . The relative efficiency at time t is obtained by taking the ratio of the simulated mean-squared errors of $\tilde{F}(t)$ and of $\hat{F}(t)$. We used the mean-squared errors to take into consideration the biases that may still exist for finite sample sizes. Observe that these plots are in agreement with the asymptotic efficiency plots provided in Figure 1.

Results for the Weibull F_0 are provided in Table 2 and Figure 3. The general observations from these simulation results are similar to those when F_0 is exponential. For estimating β , when β_0 decreases or n increases, then the estimates are closer to their target parameters, which could be explained by the fact that in either case there are more data points available. In this Weibull case, a decrease in the value of the shape parameter also leads to more precision in the estimators, again partly due to the increase in the number of data points available. Figure 3 also empirically confirms that the estimator $\hat{F}$ of $\bar{F}$, which exploits the GKG structure, always dominates the fully nonparametric estimator $\tilde{F}$ in Peña et al. (2001). In the next subsection we demonstrate, however, that if the GKG structure does not hold, then $\hat{F}$ may not have good performance relative to $\tilde{F}$.

5.2. When GKG Model Does Not Hold

Prodded by a referee's suggestion, the performance of the nonparametric estimator $\tilde{F}$ in Peña et al. (2001) was also compared with the performance of the estimator $\hat{F}$ in terms of their MSE when the GKG model does *not* hold. In the first simulation, the true F_0 is a unit exponential distribution, while G_0 is a uniform distribution over $[0, \theta]$, where θ took values in $\{1, 2, 4, 8\}$. In the second simulation, the true F_0 was a Weibull(2, 1) distribution with a G_0 being a unit exponential distribution. For each of the simulation runs, we utilized a sample size of $n = 30$ and 10000 simulation replications. Figure 4 provides the overlaid plots of the simulated efficiency of $\tilde{F}$ relative to $\hat{F}$ plotted for different values of $F_0(t)$ for these two simulation models.

Observe that in these mis-specified models, the GKG-based estimator becomes less efficient for some regions of $F_0(t)$ compared to the fully nonparametric estimator. For the exponential F_0 and uniform G_0 the loss in efficiency of the GKG estimator is not drastic and for most of the values of $F_0(t)$ the simulated efficiency values are close to unity. On the other hand, for the Weibull F_0 and exponential G_0 , the GKG estimator is quite inefficient for large regions of $F_0(t)$ especially when the length of the monitoring period is small.

It is interesting to observe the spikes in the efficiency plots for the cases where the monitoring periods are short, corresponding to $\theta = 1$ for the uniform G_0 and $\theta = 1.2$ for the exponential G_0 . This is a bias-variance phenomenon, with the degree of bias of the estimators playing a leading role. Figure 5 presents the bias plots and the relative efficiency plots for the case where $\theta = 1$ in both simulation models. Observe that the bias of the GKG estimator is highly pronounced compared to the PLE. In the exponential F_0 and uniform G_0 case, the bias functions for both estimators are increasing and almost linear after the value of $F_0(t)$ of about .63. Coincidentally, notice that under this case the probability that the first event time will be right-censored is $P\{T_1 > U\} = E\{\exp(-U)\} = 1 - \exp(-1) = .6322$ and this value is also equal to $F_0(1)$. The almost linear property of the bias functions is a consequence of the fact that the estimators become constants for t -values exceeding the maximum of $\{\min(T_{1i}, \tau_i), i = 1, 2, \dots, n\}$. This maximum value never exceeds 1.0 when θ

$= 1$, so both estimates of $F_0(t)$ for each simulation replication will be constant for $t = 1$, or equivalently, for $F_0(t) = .6322$. For the Weibull F_0 and exponential G_0 , notice that the GKG estimator is negatively biased for small values of $F_0(t)$ and becomes positively biased for larger values of $F_0(t)$. Regions of $F_0(t)$ where the GKG estimator is highly biased makes it inefficient relative to the PLE; while the region where its bias is close to zero is where the GKG estimator becomes better than the PLE, as can be seen in the plots in Figure 5.

We have noted that in the Weibull F_0 and exponential G_0 model, the efficiency loss of the GKG estimator arising from the model mis-specification cannot be ignored. This seems to suggest that a nonparametric estimation approach, when uncertain about the true underlying model, may still be the prudent approach. More interestingly, a two-step approach may be undertaken where a goodness-of-fit test of the GKG model is first performed, and if the model is statistically not rejected, to utilize the GKG estimator; otherwise, utilize the PLE. But, such an approach, as in any so-called preliminary test approach, requires caution because of the double-dipping on the data. Furthermore, research into tests of goodness-of-fit of the GKG model will be needed. Some ideas currently being pursued in this context is a procedure comparing the GKG and the PLE of F_0 to perform a goodness-of-fit test of the GKG model.

Acknowledgments

E. Peña acknowledges support from National Science Foundation (NSF) Grants DMS 0805809 and DMS 1106435, and NIH Grants RR17698 (COBRE) and R01 CA154731. The authors also graciously thank the reviewers for their useful comments and perceptive suggestions on the first version of the manuscript which led to considerable improvements.

Appendix A. Proofs

We provide in this appendix the proofs that the proposed estimators under the GKG model possess consistency and weak convergence properties. In addition, the proofs of the efficiency results are provided here.

A.1. Consistency Proofs

Proof of Theorem 3.1: Define the random function

$$D(s^*; \beta) = \frac{1}{n} \left[l_p(s^*; \beta) - l_p(s^*; \beta_0) \right] \frac{1}{n} \int_0^{s^*} \log \left\{ \left[\frac{\beta}{\beta_0} \right] \left[\frac{Y(s^*, v) + \beta_0 Y^\tau(v)}{Y(s^*, v) + \beta Y^\tau(v)} \right] \right\} N^\tau(dv) - \frac{1}{n} \int_0^{s^*} \log \left\{ \left[\frac{Y(s^*, v) + \beta Y^\tau(v)}{Y(s^*, v) + \beta_0 Y^\tau(v)} \right] \right\} N(s^*, dv). \quad (A1)$$

Also, define the deterministic function

$$d(s^*; \beta) = \int_0^{s^*} \log \left\{ \left[\frac{\beta}{\beta_0} \right] \left[\frac{y_0(s^*, v) + \beta_0 y_0^\tau(v)}{y_0(s^*, v) + \beta y_0^\tau(v)} \right] \right\} y_0^\tau \beta_0 \lambda_0(v) dv - \int_0^{s^*} \log \left\{ \left[\frac{y_0(s^*, v) + \beta y_0^\tau(v)}{y_0(s^*, v) + \beta_0 y_0^\tau(v)} \right] \right\} y_0(s^*, v) \lambda_0(v) dv. \quad (A2)$$

Both of these functions are twice-differentiable with respect to $\beta \in (0, \infty)$. Furthermore, observe that $\hat{\beta}$ is a maximizer of $\beta \mapsto D(s^*; \beta)$. The first derivative of $D(s^*; \beta)$ is the (scaled) profile score function and this is easily seen to be

$$U_p(s^*; \beta) = - \frac{1}{n} \int_0^{s^*} \left[\frac{Y^\tau(v)}{Y(s^*, v) + \beta Y^\tau(u)} \right] [N^\tau(dv) + N(s^*, dv)] + \frac{N^\tau(s^*)}{n} \frac{1}{\beta}. \quad (A3)$$

Taking the derivative of the negative of $U_P(s^*; \beta)$ yields the scaled profile observed information function which turns out to be

$$I_P(s^*; \beta) = \frac{1}{n\beta^2} \left\{ \int_0^{s^*} \left\{ 1 - \left[\frac{\beta Y^\tau(v)}{Y(s^*, v) + \beta Y^\tau(v)} \right]^2 \right\} N^\tau(dv) - \int_0^{s^*} \left[\frac{\beta Y^\tau(v)}{Y(s^*, v) + \beta Y^\tau(v)} \right]^2 N(s^*, dv) \right\}. \quad (A4)$$

From results in Peña et al. (2001) and the Glivenko-Cantelli Theorem, we also have that

$$\sup_{v \in [0, s^*]} \left| \frac{Y(s^*, v)}{n} - y_0(s^*, v) \right| \xrightarrow{pr} 0 \quad \text{and} \quad \sup_{v \in [0, s^*]} \left| \frac{Y(v)}{n} - y_0^\tau(v) \right| \xrightarrow{pr} 0 \quad (A5)$$

Furthermore, since the processes

$$\begin{aligned} W^\tau &= \left\{ W^\tau(s) = \frac{1}{\sqrt{n}} \left[N^\tau(s) - \int_0^s Y^\tau(v) \beta_0 \lambda_0(v) dv \right] : s \in [0, s^*] \right\}; \\ W(s^*) &= \left\{ W(s^*, s) = \frac{1}{\sqrt{n}} \left[N(s^*, s) - \int_0^s Y(s^*, v) \lambda_0(v) dv \right] : s \in [0, s^*] \right\}, \end{aligned}$$

both converge to Gaussian processes on Skorohod's space $\mathcal{D}[0, s^*]$, so that $\|W^\tau\| = O_P(1)$ and $\|W(s^*)\| = O_P(1)$ where $\|\cdot\|$ is the supremum norm $\mathcal{D}[0, s^*]$, then we can conclude that $I_P(s^*; \beta) \xrightarrow{pr} \mathcal{J}_P(s^*; \beta)$, where

$$\mathcal{J}_P(s^*; \beta) = \frac{1}{\beta^2} \left\{ \int_0^{s^*} \left[1 - \left(\frac{\beta y_0^\tau(v)}{y_0(s^*, v) + \beta y_0^\tau(v)} \right)^2 \right] y_0^\tau(v) \beta_0 \lambda_0(v) dv - \int_0^{s^*} \left(\frac{\beta y_0^\tau(v)}{y_0(s^*, v) + \beta y_0^\tau(v)} \right)^2 y_0(s^*, v) \lambda_0(v) dv \right\}.$$

$\beta \mapsto \mathcal{J}_P(s^*; \beta)$ is clearly a continuous mapping. Furthermore, observe that

$$\mathcal{J}_P(s^*; \beta_0) = \frac{1}{\beta_0} \int_0^{s^*} \frac{y_0^\tau(v) y_0(s^*, v)}{y_0(s^*, v) + \beta_0 y_0^\tau(v)} \lambda_0(v) dv. \quad (A6)$$

This is strictly positive since the condition $\Lambda_0(s^*) > 0$ implies the existence of an open subset O of $[0, s^*]$ such that $\lambda_0(s) > 0$ for $s \in O$, which consequently implies that in O both $y_0^\tau(\cdot)$ and $y_0(s^*, \cdot)$ are strictly positive. From the strict positivity of $\mathcal{J}_P(s^*; \beta_0)$, it implies that there exists a neighborhood $\mathfrak{B}_0$ about β_0 such that, for large n , $\beta \mapsto D(s^*; \beta)$ is strictly concave for $\beta \in \mathfrak{B}_0$. We take $\widehat{\beta}$ to be the maximizer of $D(s^*; \beta)$ inside $\mathfrak{B}_0$. We also observe that as $n \rightarrow \infty$, $\sup_{\beta \in \mathfrak{B}_0} |D(s^*; \beta) - d(s^*; \beta)| \xrightarrow{pr} 0$. Taking the second derivative with respect to β of the negative of $d(s^*; \beta)$, we find that $-\left[d^2/d\beta^2\right] d(s^*; \beta) = \mathcal{J}_P(s^*; \beta)$. From the earlier result about $\mathcal{J}_P(s^*; \beta_0)$, this implies that on $\mathfrak{B}_0$, $\beta \rightarrow d(s^*; \beta)$ is strictly concave. Finally, since β_0 is a maximizer of $\beta \rightarrow d(s^*; \beta)$ in $\mathfrak{B}_0$ by noting that the first derivative of $d(s^*; \beta)$ equals zero at $\beta = \beta_0$, it follows that the chosen $\widehat{\beta}$ -sequence will converge in probability to β_0 . This completes the proof of the existence of a consistent sequence of the profile likelihood MLEs. $\square$

For establishing Theorem 3.2, we first develop some preliminary results. We shall denote by $\mathcal{D}[0, s^*]$ the space of functions on $[0, s^*]$ which are right-continuous and with left-hand limits (the cadlag functions) and endow this with Skorohod's metric. We shall then denote

by $\mathfrak{Q} = \mathfrak{D}[0, s^*]^4 \times \mathfrak{R}_+$ and endow this space with the product metric d_P formed from Skorohod's metric for each of the four $\mathfrak{D}[0, s^*]$ and Euclidean metric for $\mathfrak{R}_+$. Also, for $D \in \mathfrak{D}[0, s^*]$, denote by $\|D\|_\infty = \sup_{s \in [0, s^*]} |D(s)|$ and let $d_S: \mathfrak{Q} \times \mathfrak{Q} \rightarrow \mathfrak{R}_+$ with

$$d_S[(U_1, U_2, U_3, U_4, q), (V_1, V_2, V_3, V_4, r)] = \sqrt{\|U_1 - V_1\|_\infty^2 + \|U_2 - V_2\|_\infty^2 + \|U_3 - V_3\|_\infty^2 + \|U_4 - V_4\|_\infty^2 + |q - r|^2}.$$

Letting $Q \equiv Q_n = \left\{ \left(\frac{N^\tau(s)}{n}, \frac{Y^\tau(s^+)}{n}, \frac{N(s^*, s)}{n}, \frac{Y(s^*, s^+)}{n} \right), s \in [0, s^*] \right\}, \widehat{\beta}_n$, observe that $Q \in \mathfrak{Q}$. Also, denote by

$$Q_0 = \left\{ \left(\int_0^s y_0^\tau(v) \beta_0 \lambda_0(v) dv, y_0^\tau(s^*), \int_0^s y_0(s^*, v) \lambda_0(v) dv, y_0(s^*, s^*) \right), s \in [0, s^*] \right\}, \beta_0.$$

Lemma A.1 Under (β_0, F_0) , as $n \rightarrow \infty$, $d_P(Q_n, Q_0) \xrightarrow{pr} 0$.

Proof of Lemma A.1: It suffices to show that $d_S(Q_n, Q_0) \xrightarrow{pr} 0$ since $d_P(\cdot, \cdot) = d_S(\cdot, \cdot)$. By Glivenko-Cantelli theorem, we immediately see that $\|Y^\tau/n - y_0^\tau\|_\infty \xrightarrow{pr} 0$. Using results in Peña et al. (2001), we also have $\|Y(s^*, \cdot)/n - y_0(s^*, \cdot)\|_\infty \xrightarrow{pr} 0$. We have already shown in Theorem 3.1 that $\|\widehat{\beta}_n - \beta_0\|_\infty \xrightarrow{pr} 0$. By the martingale central limit theorem (MCLT) we also have that $\{M^\tau(s)/\sqrt{n}: s \in [0, s^*]\}$ converges weakly to a Gaussian process, from where it follows that $\|N^\tau(\cdot)/n - \int_0^\cdot y_0^\tau(v) \beta_0 \lambda_0(v) dv\| = o_p(1)$. Finally, from results in Peña et al. (2001), $\{M(s^*, s)/\sqrt{n}: s \in [0, \infty)\}$ converges weakly to a Gaussian process, hence $\|(N^\tau(s^*, \cdot)/n - \int_0^{s^*} y_0(s^*, v) \lambda_0(v) dv)\|_\infty = o_p(1)$. Combining these results we obtain the assertion of the lemma. $\parallel$

Proof of Theorem 3.2: Recall that

$$\widehat{\Lambda}(s^*, t) = \int_0^t \frac{I\{\widehat{\beta} Y^\tau(v) + Y(s^*, v) > 0\}}{\widehat{\beta} Y^\tau(v) + Y(s^*, v)} [N^\tau(dv) + N(s^*, dv)], t \in [0, t^*].$$

We may view this as a mapping H from $\mathfrak{Q}$ into $\mathfrak{D}[0, t^*]$ which maps Q into $\widehat{\Lambda}(s^*, \cdot)$. If we could show that this mapping is continuous, it will follow by the Continuous Mapping Theorem and Lemma A.1 that $\|\widehat{\Lambda}(s^*, \cdot) - \Lambda_0(\cdot)\|_\infty = o_p(1)$, since we note that for $t \in [0, t^*]$ we have

$$\int_0^t \frac{I\{\beta_0 y_0^\tau(v) + y_0(s^*, v) > 0\}}{\beta_0 y_0^\tau(v) + y_0(s^*, v)} [y_0^\tau(v) \beta_0 \lambda_0(v) dv + y_0(s^*, v) \lambda_0(v) dv] = \Lambda_0(t).$$

Clearly, the mappings $Q \mapsto \left(\frac{N^\tau(\cdot)}{n} + \frac{N(s^*, \cdot)}{n} \right)$ and $Q \mapsto \left(\widehat{\beta} \frac{Y^\tau(\cdot)}{n} - \frac{Y(s^*, \cdot)}{n} \right)$ are continuous maps, so that to show that the mapping H is continuous it suffices to show that the mapping H^* from $\mathfrak{D}[0, s^*]^2$ into $\mathfrak{D}[0, t^*]$ given by

$$(N, Y) \mapsto \int_0^\cdot \frac{I\{Y(v-) > 0\}}{Y(v-)} N(dv) \equiv \int_0^\cdot \frac{I\{Y(v-) / n > 0\}}{Y(v-) / n} [N(dv) / n]$$

is a continuous map with respect to the sup-metrics in these spaces under appropriate conditions. In the above, N is a non-decreasing nonnegative function, while Y is a non-increasing nonnegative function. Thus, let $\{(N^*, Y^*)\}$ be a sequence in $\mathcal{D}[0, s^*]^2$ such that $ds[(N^*, Y^*) / n, (n_0^*, y_0^*)] \rightarrow 0$ with y_0^* a continuous function and $n_0^*(\cdot) = \int_0^\cdot y_0^*(v) \lambda_0(v) dv$. For a function f , f_- will denote its left-continuous version, that is, $f_-(v) = f(v-) = \lim_{u \uparrow v} f(u)$. We have that

$$\begin{aligned} & \left| \int_0^\cdot \frac{J_-^*}{Y_-^* / n} dN^* / n - \int_0^\cdot \frac{1}{y_0^*} dn_0^* \right| \\ & \leq \left\{ \sup_{t \in [0, t^*]} \left| \frac{J_-^*}{Y_-^* / n} - \frac{1}{y_0^*} \right| \right\} \left[\frac{N^*(t^*)}{n} + \frac{1}{y_0^*(t^*)} \|N^* / n - n_0^*\|_\infty \right] \\ & \leq \|Y^* / n - y_0^*\|_\infty \left[\frac{n_0^*(t^*) + o(1)}{y_0^*(t^*)(y_0^*(t^*) + o(1))} \right] + \left[\frac{1}{y_0^*(t^*)} \right] \|N^* / n - n_0^*\|_\infty. \end{aligned}$$

This upper bound is of order $o(1)$ provided $y_0^*(t^*) > 0$ and $n_0(t^*) = \int_0^{t^*} y_0^*(v) \lambda_0(v) dv < \infty$, which implies the continuity of the mapping H^* at (n_0^*, y_0^*) . For our specific problem, we have that $y_0^*(t) = \beta_0 y_0(t) + y_0(s^*, t)$. Since we assumed that t^* is such that $\Lambda_0(t^*) < \infty$ and $\beta_0 y_0(t^*) + y_0(s^*, t^*) > 0$, then we have established that $d_s[\widehat{\Lambda}(s^*, \cdot), \Lambda_0(\cdot)] \xrightarrow{pr} 0$, establishing the result in the theorem pertaining to the estimator of $\Lambda(\cdot)$. Since it is also well-known that the map $\Lambda \rightarrow \bar{F} = \Pi[1 - d\Lambda]$ is a continuous map on $\mathcal{D}[0, t^*]$ with the sup-norm and where Λ is a hazard function, $\bar{F}$ is a survivor function, and Π denotes the product-integral, then by the Continuous Mapping Theorem and the just-established result, it follows that $d_s[\widehat{\bar{F}}(s^*, \cdot), \bar{F}_0(\cdot)] \xrightarrow{pr} 0$ which establishes the second assertion in the theorem. $\parallel$

A.2. Weak Convergence Proofs

To make notation more compact, let

$$\begin{aligned} \mathbb{M}(\cdot) &= \mathbb{M}_{s^*}(\cdot) = \begin{bmatrix} M(s^*, \cdot; \Lambda_0) \\ M^\tau(\cdot; \Lambda_0, \beta_0) \end{bmatrix}; & \mathbb{N}(\cdot) &= \mathbb{N}_{s^*}(\cdot) = \begin{bmatrix} N(s^*, \cdot) \\ N^\tau(\cdot) \end{bmatrix}; \\ \mathbb{Y}(\cdot) &= \mathbb{Y}_{s^*}(\cdot) = \begin{bmatrix} Y(s^*, \cdot) \\ T^\tau(\cdot) \end{bmatrix}; & \mathbb{A}(\cdot) &= \mathbb{A}_{s^*}(\cdot) = \begin{bmatrix} A(s^*, \cdot; \Lambda_0) \\ A^\tau(\cdot; \Lambda_0, \beta_0) \end{bmatrix}. \end{aligned}$$

In the sequel, $\text{Dg}(\mathbf{a})$ denotes the diagonal matrix formed from the vector $\mathbf{a}$ and $'$ is matrix transpose. To establish Theorem 3.3 and Theorem 3.4, we shall prove a succession of lemmas. Lemma A.2 extends Theorem 1 in Peña et al. (2000).

Lemma A.2 Let $\mathbf{H}$ be a $k \times 2$ matrix of processes where the i th row is of form $[H_{i1}(s^*, w), H_{i2}(w)]$ with $H_{i2}(\cdot)$ a predictable bounded process and $H_{i1}[R_i(\cdot), t]$ a predictable bounded process for each t and i . Furthermore, for each $r \in \{1, 2, \dots, k\}$ and $j \in \{1, 2\}$, there exists a deterministic function $h_{rj}(\cdot)$ satisfying $\|H_{rj}(s^*, \cdot) - h_{rj}(\cdot)\|_\infty \xrightarrow{pr} 0$. Then, as $n \rightarrow \infty$, the

vector-valued process $\{\mathbf{W}(s^*, t) : t \in [0, t^*]\}$ with $\mathbf{W}(s^*, t) = \frac{1}{\sqrt{n}} \int_0^t \mathbf{H}(w) \mathbb{M}(dw)$ converges

weakly to $\mathbf{G}$, a $k \times 1$ zero-mean Gaussian process with covariance matrix function

$$\Xi(s^*, t) = \int_0^t \mathbf{h}(w) \text{Dg} \left[y_0(s^*, w), \beta_0 y_0^\tau(w) \right] \mathbf{h}(w)' \Lambda_0(dw).$$

Proof : Write $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2]$ and $\mathbf{h} = [\mathbf{h}_1, \mathbf{h}_2]$ so $\mathbf{W} = \mathbf{W}_1 + \mathbf{W}_2$, where

$$\begin{aligned} \mathbf{W}_1(s^*, t) &= \frac{1}{\sqrt{n}} \int_0^t \mathbf{H}_1(s^*, w) M(s^*, dw) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{s^*} \mathbf{H}_1(s^*, R_i(v)) I\{R_i(v) \leq t\} M_i^\dagger(dv); \\ \mathbf{W}_2(t) &= \frac{1}{\sqrt{n}} \int_0^t \mathbf{H}_2(v) M^\tau(dv) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{s^*} \mathbf{H}_2(v) I\{v \leq t\} M_i^\tau(dv). \end{aligned}$$

From Peña et al. (2000) we know that $\mathbf{W}_1(s^*, \cdot)$ weakly converges to a zero-mean Gaussian process with variance function $\sum_1(s^*, t) = \int_0^t \mathbf{h}_1(s^*, w)^{\otimes 2} y_0(s^*, w) \Lambda_0(dw)$; while from standard counting process results we also know that $\mathbf{W}_2(\cdot)$ weakly converges to a zero-mean Gaussian process whose variance function is $\sum_2(t) = \int_0^t \mathbf{h}_2(w)^{\otimes 2} \beta_0 y_0^\tau(w) \Lambda_0(dw)$. If we could show that $\mathbf{W}_1$ and $\mathbf{W}_2$ are asymptotically uncorrelated, then the result of the lemma will follow by virtue of the marginal weak convergence to Gaussian processes and since $\Xi(s^*, t) = \sum_1(s^*, t) + \sum_2(t)$. But this asymptotic uncorrelatedness follows by noting that

$$\begin{aligned} & \text{Cov} \left\{ \int_0^{t_1} \mathbf{H}_1(s^*, w) M(s^*, dw), \int_0^{t_2} \mathbf{H}_2(w) M^\tau(dw) \right\} \\ &= \text{Cov} \left\{ \sum_{i=1}^n \int_0^{s^*} \mathbf{H}_1(s^*, v) I\{R_i(v) \leq t_1\} M_i^\dagger(dv), \sum_{i=1}^n \int_0^{s^*} \mathbf{H}_2(v) I\{v \leq t_2\} M_i^\tau(dv) \right\} \\ &= E \left[\sum_{i,j=1}^n \int_0^{s^*} \mathbf{H}_1(s^*, v) \mathbf{H}_2(v)' \langle M_i^\dagger, M_j^\tau \rangle(dv) \right] = 0, \end{aligned}$$

since for every i, j , $\langle M_i^\dagger, M_j^\tau \rangle(dv) = 0$ by the continuity of the $T_i s$ and $\tau_i s$. $\parallel$

Lemma A.3 As $n \rightarrow \infty$, $\sqrt{n}(\widehat{\beta} - \beta_0) = \{\mathcal{J}_p(s^*; \Lambda_0, \beta_0)\}^{-1} \left[\sqrt{n} U_p(s^*; \beta_0) + o_p(1) \right]$, where $U_p(s^*, \beta_0)$ is the profile likelihood score function in (A3) evaluated at $\beta = \beta_0$ and $\mathcal{J}_p(s^*; \beta_0)$ is the profile likelihood information in (A6).

Proof : This is immediate via Taylor's expansion, weak law of large numbers, and the fact that $\widehat{\beta}_n$ converges in probability to β_0 . $\parallel$

Let us define the process $\{U_p(s^*, t; \beta_0) : t \in [0, s^*]\}$, where with $\mathbb{M}(s^*, \cdot) \equiv \mathbb{M}_{s^*}(\cdot)$,

$$U_p(s^*, t; \beta_0) = \frac{1}{n} \int_0^t \mathbf{H}_2(s^*, v) \mathbb{M}(s^*, dv)$$

with

$$\mathbf{H}_2(s^*, v) = \left[\frac{1}{\beta_0} \left(\frac{\beta_0 Y^\tau(v)}{Y(s^*, v) + \beta_0 Y^\tau(v)} \right), -\frac{1}{\beta_0} \left(\frac{Y(s^*, v)}{Y(s^*, v) + \beta_0 Y^\tau(v)} \right) \right].$$

Observe that $U_p(s^*; \beta_0) = U_p(s^*, s^*; \beta_0)$.

Lemma A.4 As $n \rightarrow \infty$, $\{\sqrt{n}U_p(s^*, t; \beta_0) : t \in [0, t^*]\}$ converges weakly to a zero-mean Gaussian process with covariance function $\sigma_2^2(s^*, \cdot; \Lambda_0, \beta_0)$ given by

$$\sigma_2^2(s^*, t; \Lambda_0, \beta_0) = \int_0^t \frac{y_0(v) y_0^\tau(v)}{\beta_0 [y_0(v) + \beta_0 y_0^\tau(v)]} \Lambda_0(dv).$$

In particular, $\sqrt{n}U_p(s^*; \beta_0) = \sqrt{n}U_p(s^*, s^*; \beta_0)$ converges in distribution to a normal variable with mean zero and variance $\mathcal{J}_p(s^*; \beta_0)$.

Proof : Follows directly from Lemma A.2 and straightforward simplification of the variance matrix function. $\parallel$

Proof of Theorem 3.3: That $\sqrt{n}(\widehat{\beta} - \beta_0)$ converges to a normal distribution with mean zero and variance equal to $\mathcal{J}_p(s^*; \Lambda_0, \beta_0)^{-1}$ is now an immediate consequence of Lemma A.3 and Lemma A.4. $\parallel$

Let us also define the process $\{\Lambda_0^*(s^*, t; \beta_0) : t \in [0, t^*]\}$ via

$$\Lambda_0^*(s^*, t; \beta_0) = \int_0^t I\{\beta_0 Y^\tau(v) + Y(s^*, v) > 0\} \Lambda_0(dv),$$

and the deterministic function

$$t \in [0, s^*] \mapsto a(s^*, t; \Lambda_0, \beta_0) = \int_0^t \frac{y_0^\tau(v)}{y_0(s^*, v) + \beta_0 y_0^\tau(v)} \Lambda_0(dv).$$

Lemma A.5 Let $V_n(s^*, t) = \sqrt{n}[\widehat{\Lambda}(s^*, t) - \Lambda_0^*(s^*, t; \beta_0)] + a(s^*, t; \Lambda_0, \beta_0) \sqrt{n}(\widehat{\beta} - \beta_0)$. Then $V_n(s^*, \cdot)$ and $\sqrt{n}U_p(s^*, \cdot; \beta_0)$ are asymptotically independent and $V_n(s^*, \cdot)$ converges weakly to a zero-mean Gaussian process with variance function

$$\sigma_1^2(\cdot; \Lambda_0, \beta_0) = \int_0^\cdot \frac{\Lambda_0(dv)}{y_0(s^*, v) + \beta_0 y_0^\tau(v)}.$$

Furthermore, $V_n(s^*, \cdot)$ and $\sqrt{n}(\widehat{\beta} - \beta_0)$ are asymptotically independent.

Proof : The result follows immediately since by Taylor expansion and using the consistency of $\widehat{\beta}$, we obtain the asymptotic representation for V_n given by

$$V_n(s^*, t) = \frac{1}{\sqrt{n}} \int_0^t \mathbf{H}_1(s^*, v) \mathbb{M}(dv) + o_p(1),$$

with

$$\mathbf{H}_1(s^*, v) = \frac{I\{Y(s^*, v)/n + \beta_0 Y^\tau(v)/n > 0\}}{Y(s^*, v)/n + \beta_0 Y^\tau(v)/n} [1, 1].$$

Coupling this with the representation $\sqrt{n}U_p(s^*, t) = \frac{1}{\sqrt{n}} \int_0^t \mathbf{H}_2(s^*, v) \mathbb{M}(dv)$, then applying Lemma A.4 and Slutsky's Theorem, we find that $[V_n(s^*, \cdot), \sqrt{n}U_p(s^*, \cdot)]'$ converges weakly to a zero-mean Gaussian process with covariance function

$$\sum(s^*, v) = \int_0^v \begin{bmatrix} \mathbf{h}_1(s^*, v) \\ \mathbf{h}_2(s^*, v) \end{bmatrix} Dg[y_0(s^*, v), \beta_0 y_0^\tau(v)] \begin{bmatrix} \mathbf{h}_1(s^*, v) \\ \mathbf{h}_2(s^*, v) \end{bmatrix}' \Lambda_0(dv),$$

where

$$\begin{aligned} \mathbf{h}_1(s^*, t) &= (\beta_0 [y_0(s^*, t) + \beta_0 y_0^\tau(t)])^{-1} (1, 1); \\ \mathbf{h}_2(s^*, t) &= (\beta_0 [y_0(s^*, t) + \beta_0 y_0^\tau(t)])^{-1} (\beta_0 y_0^\tau(t), -y_0(s^*, t)). \end{aligned}$$

Simplifying, we find that

$$\sum(s^*, t) = \int_0^t \frac{1}{y_0(s^*, v) + \beta_0 y_0^\tau(v)} Dg(1, y_0(s^*, v) + \beta_0 y_0^\tau(v)) \Lambda_0(dv).$$

This establishes the first part of the lemma. The second result immediately follows from the fact that $\widehat{\beta}$ is a functional of $U_P(s^*, \cdot)$. $\parallel$

With these preliminaries we are now ready to present the proof of the weak convergence to a Gaussian process of the centered and scaled $\widehat{\Lambda}(s^*, \cdot)$.

Proof of Theorem 3.4: Observe that

$$\sqrt{n} [\widehat{\Lambda}(s^*, t) - \Lambda_0(t)] = V_n(s^*, t) + a(s^*, t; \Lambda_0, \beta_0) \sqrt{n} (\widehat{\beta} - \beta_0) + \sqrt{n} [\Lambda_0^*(s^*, t; \beta_0) - \Lambda_0(t)].$$

By the preceding lemma, it follows that the sum of the first two terms weakly converges to a zero-mean Gaussian process with variance function given by the expression $\sigma_{\widehat{\Lambda}}^2(s^*, t)$ in the statement of the theorem. Thus it suffices to show that the supremum over $t \in [0, t^*]$ of the third term is $o_p(1)$. We have that

$$\sup_{t \in [0, t^*]} |\sqrt{n} [\Lambda_0^*(s^*, t; \beta_0) - \Lambda_0(t)]| \leq \sqrt{n} \Lambda_0(t^*) I\{Y(s^*, t^*) = 0\}.$$

Invoking Markov's Inequality, for $\epsilon > 0$, we get that

$$P\{\sqrt{n} I\{Y(s^*, t^*) = 0\} > \epsilon\} \leq \frac{\sqrt{n} P\{Y(s^*, t^*) = 0\}}{\epsilon}.$$

By the IID assumption over the n units, we have

$$\sqrt{n}P\{Y(s^*, t^*)=0\} = \sqrt{n}[P\{Y_1(s^*, t^*)=0\}]^n \rightarrow 0$$

since $P\{Y_1(s^*, t^*)=0\} < 1$ by virtue of the assumption that $\bar{F}_0(t^*)\beta_0^{t^*+1} > 0$. Thus,

$$\sup_{t \in [0, t^*]} \sqrt{n}[\Lambda_0^*(s^*, t; \beta_0) - \Lambda_0(t)] = o_p(1). \parallel$$

A.3. Efficiency Proofs

Proof of Theorem 4.1: Under the exponential model for F_0 and with $s^* \rightarrow \infty$, we have that

$y_0^\tau(t) = \exp(-\beta_0 \theta_0 t)$ and $y_0(\infty, t) = \left(\frac{1+\beta_0}{\beta_0}\right) \exp\{-(1+\beta_0)\theta_0 t\}$. For the last expression see, for instance, Hollander and Peña (2004) and Adekpedjou et al. (2010). The expression in the statement of the theorem is routinely obtained by making the variable transformation $u = \exp(-\theta_0 w)$ in the integrals in

$$ARE\left(\hat{\bar{F}}(\infty, t) : \hat{\bar{F}}(\infty, t; \beta_0)\right) = \left[\int_0^t \frac{\theta_0 dw}{y_0(\infty, w)}\right]^{-1} \left[\int_0^t \frac{\theta_0 dw}{y_0(\infty, w) + \beta_0 y_0^\tau(w)}\right],$$

then simplifying. The upper bound follows by replacing $u^{1+\beta_0}$ by $u^{2+\beta_0}$. $\parallel$

Proof of Theorem 4.2: Define the measure ν on the Borel measurable space $([0, s^*], \sigma[0, s^*])$ via

$$\nu(A) = \int_A \frac{\Lambda_0(dw)}{y_0(s^*, w) + \beta_0 y_0^\tau(w)}, \forall A \in \sigma[0, s^*].$$

Then we could write $\sigma_\Lambda^2(s^*, t)$ in Theorem 3.4 in terms of ν via

$$\sigma_\Lambda^2(s^*, t) = \nu[0, t] + \beta_0 \left[\int_0^{s^*} y_0(s^*, w) y_0^\tau(w) \nu(dw)\right]^{-1} \left[\int_0^t y_0^\tau(w) \nu(dw)\right]^2.$$

Applying the Cauchy-Schwartz Inequality we get for the last term above

$$\begin{aligned} & \left[\int_0^t y_0^\tau(w) \nu(dw)\right]^2 \\ &= \left[\int_0^{s^*} \frac{I(w \leq t) \sqrt{y_0^\tau(w)}}{\sqrt{y_0(s^*, w)}} \sqrt{y_0^\tau(w) y_0(s^*, w)} \nu(dw)\right]^2 \leq \left[\int_0^{s^*} \frac{I(w \leq t) y_0^\tau(w)}{y_0(s^*, w)} \nu(dw)\right] \left[\int_0^{s^*} y_0(s^*, w) y_0^\tau(w) \nu(dw)\right]. \end{aligned}$$

Consequently,

$$\sigma_\Lambda^2(s^*, t) \leq \nu[0, t] + \beta_0 \int_0^t \frac{y_0^\tau(w)}{y_0(s^*, w)} \nu(dw) = \int_0^t \left[1 + \frac{\beta_0 y_0^\tau(w)}{y_0(s^*, w)}\right] \nu(dw) = \int_0^t \frac{\Lambda_0(dw)}{y_0(s^*, w)} = \sigma_\Lambda^2(s^*, t)$$

completing the proof. $\parallel$

References

- Aalen O. Nonparametric inference for a family of counting processes. *Annals of Statistics*. 1978; 6:701–726.
- Aalen O, Husebye E. Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine*. 1991; 10:1227–1240. [PubMed: 1925154]
- Aalen, OO.; Borgan, Ø.; Gjessing, HK. *Survival and event history analysis: A process point of view, Statistics for Biology and Health*. Springer; New York: 2008.
- Adekpedjou A, Peña E, Quiton J. Estimation and efficiency with recurrent event data under informative monitoring. *Journal of Statistical Planning and Inference*. 2010; 140:597–615. [PubMed: 20161471]
- Andersen, P.; Borgan, O.; Gill, R.; Keiding, N. *Statistical Models Based on Counting Processes*. Springer-Verlag; New York: 1993.
- Chen YY, Hollander M, Langberg NA. Small-sample results for the Kaplan-Meier estimator. *J. Amer. Statist. Assoc.* 1982; 77:141–144.
- Cheng PE, Lin GD. Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazards model. *Statist. Probab. Lett.* 1987; 5:75–80.
- Cook R, Lawless J. Analysis of Repeated Events. *Statistical Methods in Medical Research*. 2002; 11:141–166. [PubMed: 12040694]
- Cook, RM.; Lawless, JF. *The Statistical Analysis of Recurrent Events*. Springer; New York: 2007.
- Csörgö S, Faraway JJ. The paradoxical nature of the proportional hazards model of random censorship. *Statistics*. 1998; 31:67–78.
- Fleming, T.; Harrington, D. *Counting Processes and Survival Analysis*. Wiley; New York: 1991.
- Gill RD. Testing with replacement and the product-limit estimator. *The Annals of Statistics*. 1981; 9:853–860.
- Gill RD, Johansen S. A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* 1990; 18:1501–1555.
- Gjessing HK, Røysland K, Peña EA, Aalen OO. Recurrent events and the exploding Cox model. *Lifetime Data Anal.* 2010; 16:525–546. [PubMed: 20625827]
- Hollander M, Peña E. Nonparametric Methods in Reliability. *Statistical Science*. 2004; 19:644–651. [PubMed: 16710444]
- Hougaard, P. *Analysis of Multivariate Survival Data*. Springer; New York: 2000.
- Ihaka R, Gentleman R. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*. 1996; 5:299–314.
- Jacod J. Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Geb.* 1975; 31:235–253.
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 1958; 53:457–481.
- Koziol JA, Green SB. A Cramér-von Mises statistic for randomly censored data. *Biometrika*. 1976; 63:465–474.
- Lindqvist BH. On the statistical modeling and analysis of repairable systems. *Statist. Sci.* 2006; 21:532–551.
- Nelson, WB. *Recurrent events data analysis for product repairs, disease recurrences, and other applications*, Vol. 10 of ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics (SIAM); Philadelphia, PA: 2003.
- Peña E, Strawderman R, Hollander M. Nonparametric estimation with recurrent event data. *J. Amer. Statist. Assoc.* 2001; 96:1299–1315.
- Peña E, Slate E, Gonzalez JR. Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*. 2007; 137:1727–1747. [PubMed: 19823592]
- Peña, EA.; Strawderman, RL.; Hollander, M. *Recent advances in reliability theory (Bordeaux, 2000)* Stat. Ind. Technol. Birkhäuser Boston; Boston, MA: 2000. A weak convergence result relevant in recurrent and renewal models; p. 493-514.

- Proschan F. Theoretical explanation of observing decreasing failure rate. *Technometrics*. 1963; 5:375–383.
- Resnick, S. *Adventures in Stochastic Processes*. Birkhauser; Boston: 1994.
- Sellke T, Gupta S, Berger J. Weak convergence of the Aalen estimator for a censored renewal process. *Statistical Decision Theory and Related Topics IV*. 1988; 2:183–194.
- Stocker R, Peña EA. A General Class of Parametric Models for Recurrent Event Data. *Technometrics*. 2007; 49:210–220.
- Therneau T, Hamilton S. rhDNase as an example of recurrent event analysis. *Statistics in Medicine*. 1997; 16:2029–2047. [PubMed: 9308130]
- Wang MC, Chang SH. Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association*. 1999; 94:146–153.

Figure 1.

Asymptotic relative efficiency of the fully nonparametric estimator in Peña et al. (2001) relative to the estimator under the GKG model under an exponential $\bar{F}_0$. Each curve is for a specific value of β_0 . From the top curve to the bottom curve, the values of β_0 are $\{.1, .3, .5, .7, .9, 1.0, 1.5\}$.

Figure 2.

Simulated efficiency of $\tilde{F}$ relative to $\hat{F}$ under an exponential $\tilde{F}_0$ based on 10000 replications for different values of β and for $n = 30$. From the top curve to the bottom curve the β -values are .1, .3, .5, .7, .9, 1.0, and 1.5.

Figure 3.

Simulated efficiency of $\tilde{F}$ relative to $\hat{F}$ under a Weibull $\bar{F}_0$ based on 10000 replications for different values of β , for $n = 50$, and for shape parameter $\alpha = 2$. From the top curve to the bottom curve the β -values are .1, .3, .5, .7, .9, 1.0, and 1.5.

Figure 4.

Simulated efficiency plots of $\tilde{F}$ relative to $\hat{F}$ under (i) a unit exponential $\bar{F}_0$ and when G_0 is a uniform distribution over $[0, \theta]$ with $\theta \in \{1, 2, 4, 8\}$, and (ii) under a Weibull(2, 1) F_0 and an exponential(θ) G_0 with $\theta \in \{1.2, .1, .5, .25\}$. All simulation runs had 10000 replications.

Figure 5. Bias and relative efficiency plots of the PLE and GKG estimators of $\tilde{F}$ based on 10000 simulation replications for (i) unit exponential F_0 and unit uniform G_0 , and (ii) a Weibull(2, 1) F_0 and a unit exponential G_0 .

Table 1

Simulated means and standard errors (SE) of $\widehat{\beta}$ under $\bar{F}_0(t) = \exp(-t)$ for different values of β_0 and n . The number of replications in the simulation is 10000. Also provided are the asymptotic standard errors (ASE).

β_0	$n = 30$			$n = 50$			$n = 100$		
	Mean	SE	ASE	Mean	SE	ASE	Mean	SE	ASE
0.1	.1017	.0316	.0302	.1014	.0245	.0234	.1005	.0169	.0165
0.3	.3085	.0824	.0770	.3048	.0624	.0596	.3018	.0430	.0422
0.5	.5159	.1420	.1270	.5119	.1038	.0983	.5045	.0708	.0695
0.7	.7308	.2093	.1815	.7176	.1518	.1406	.7075	.1032	.0994
0.9	.9466	.2827	.2406	.9213	.1964	.1864	.9093	.1356	.1318
1.0	1.0544	.3184	.2720	1.0294	.2298	.2107	1.0172	.1579	.1490
1.5	1.6087	.5611	.4447	1.5615	.3963	.3445	1.5316	.2572	.2436

Table 2

Simulated means and standard errors of $\widehat{\beta}$ under $\bar{F}_0(t) = \exp(-t^\alpha)$ for $\alpha \in \{.9, 2\}$ and different values of β_0 and n . The number of replications in the simulation is 10000.

β_0	$n = 30, \alpha = 2$		$n = 50, \alpha = 2$		$n = 30, \alpha = .9$		$n = 50, \alpha = .9$	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
0.1	.1028	.0377	.1012	.0283	.1017	.0313	.1010	.0237
0.3	.3106	.0973	.3061	.0724	.3069	.0797	.3043	.0611
0.5	.5242	.1674	.5147	.1211	.5158	.1371	.5092	.1017
0.7	.7384	.2381	.7204	.1720	.7278	.2006	.7129	.1466
0.9	.9558	.3209	.9299	.2312	.9415	.2686	.9220	.1968
1.0	1.0675	.3624	1.0378	.2616	1.0478	.3066	1.0318	.2296
1.5	1.6405	.6388	1.5851	.4357	1.6099	.5564	1.5654	.3834