

NONPARAMETRIC ESTIMATION WITH RECURRENT EVENT DATA

BU- 1568-M

May 2001

E.A. Peña
Robert L. Strawderman
and
M. Hollander

Keywords: correlated recurrent times, EM algorithm, frailty model, informative censoring, Lexis diagram, Martingale

Abstract:

The problem of non-parametric estimation for the distribution function governing the time to occurrence of a recurrent event in the presence of censoring is considered. We derive Nelson-Aalen and Kaplan-Meier-type estimators for the distribution function, and establish their respective finite-sample and asymptotic properties. We allow for random observation periods for each subject under study and explicitly account for the informative sum-quota nature of the data accrual scheme. These allowances complicate technical matters considerably and in particular invalidate the direct use of Martingale methods. Consistency and weak convergence of our estimators are obtained by extending an approach due to Sellke (1988), who considered a single renewal process (i.e., recurrent events on a single subject) observed over an infinite time period. A useful feature of the present analysis is that strong parallels are drawn to the usual "single-event" setting, providing a natural route towards developing extensions that involve covariates (e.g., weighted log-rank tests, Cox-type regression and frailty models). Another advantage is that we obtain explicit, closed-form expressions for the asymptotic variances for these estimators. This enables for instance the characterization of the efficiency loss resulting from employing only the first, possibly right-censored, observation per subject. An interesting feature of these results is the prominent role of the renewal function. Finally, we discuss the case of correlated recurrence times, propose an estimator in the case where the within-unit interoccurrence times follow a gamma frailty model, and compare the performance of our estimators to an estimator recently proposed by Wang and Chang (1999).

CORNELL UNIVERSITY BIOMETRICS DEPARTMENT
TECHNICAL REPORT BU-1568-M

Nonparametric Estimation with Recurrent Event Data

E. A. Peña*[†] R. L. Strawderman[‡] M. Hollander[§]

May 30, 2001

Abstract

The problem of nonparametric estimation for the distribution function governing the time to occurrence of a recurrent event in the presence of censoring is considered. We derive Nelson-Aalen and Kaplan-Meier-type estimators for the distribution function, and establish their respective finite-sample and asymptotic properties. We allow for random observation periods for each subject under study and explicitly account for the informative sum-quota nature of the data accrual scheme. These allowances complicate technical matters considerably and in particular invalidate the direct use of martingale methods. Consistency and weak convergence of our estimators are obtained by extending an approach due to Sellke (1988), who considered a single renewal process (i.e., recurrent events on a single subject) observed over an infinite time period. A useful feature of the present analysis is that strong parallels are drawn to the usual “single-event” setting, providing a natural route towards developing extensions that involve covariates (e.g., weighted log-rank tests, Cox-type regression and frailty models). Another advantage is that we obtain explicit, closed-form expressions for the asymptotic variances for these estimators. This enables for instance the characterization of the efficiency loss resulting from employing only the first, possibly right-censored, observation per subject. An interesting feature of these results is the prominent role of the renewal function. Finally, we discuss the case of correlated recurrence times, propose an estimator in the case where the within-unit interoccurrence times follow a gamma frailty model, and compare the performance of our estimators to an estimator recently proposed by Wang and Chang (1999).

KEYWORDS: Correlated recurrent times; EM algorithm; Frailty model; Informative censoring; Lexis diagram; Martingale; Sum-quota sampling; Weak convergence.

*The first two authors contributed equally to this manuscript.

[†]E. A. Peña is Professor in the Department of Statistics, University of South Carolina, Columbia, SC 29208. His research is partially supported by NIH Grant 1 R01 GM56182.

[‡]R. L. Strawderman is Associate Professor in the Departments of Biometrics and Statistics, Cornell University, Ithaca, NY, 14853.

[§]M. Hollander is Robert O. Lawton Distinguished Professor and Chairman of the Department of Statistics, Florida State University, Tallahassee, FL 36303. His research is partially supported by NIH Grant 1 R01 GM56183.

1 Introduction

Data involving recurrent events arise in a wide variety of settings, including public health, biomedicine, psychology, psychiatry, engineering, sociology and economics. Examples of such events in the health and biomedical sciences are drug abuse of teenagers or adults, recurrent hospitalization of patients with chronic diseases, bouts with migraine headaches, epileptic seizures, and episodes of hypoglycemia in diabetics. In psychiatric studies the onset of depression and dementia are instances of recurring events; in the engineering and reliability settings, the breakdown of mechanical or electronic systems, computer software crashes, stoppages of nuclear power plants, and warranty claims for manufactured products are all examples of recurrent phenomena. Examples in sociology and economics include serious disagreements in a marriage, onset of labor strikes, and auto insurance claims. The development of stochastic models and statistical methods appropriate for analyzing recurrent data is therefore of considerable importance.

The problem of nonparametric estimation and inference for recurrent event data has, among others, been by considered Gill (1980, 1981), Lin, Wei and Ying (1998), McClean and Devine (1995), Soon and Woodroffe (1996), Sellke (1988), Vardi (1982ab), and Wang and Chang (1999). Gill (1981) dealt with the problem of nonparametric inference for renewal processes in a life-testing setting. Assuming a fixed observation window, he derives analogs of the Nelson-Aalen and Kaplan-Meier estimators for the event interoccurrence time distribution and establishes consistency and weak convergence using empirical process techniques similar to those used in Gill (1980), where related results are established in a considerably more general setting. Sellke (1988) considered the problem of establishing weak convergence of a Nelson-Aalen-type estimator in the case where a single renewal process is observed over an infinite time period. Vardi (1982b) presented an algorithm for obtaining the maximum likelihood estimator of the distribution function when the underlying event times are arithmetic under a data accrual scheme in which the first event time for each subject is a left-truncated observation; see also McClean and Devine (1995) for related results. Soon and Woodroffe (1996) extended Vardi's algorithm to allow for the situation where the event time distribution is not arithmetic, and prove the consistency of their estimator under the assumption that the length of the observation period converges to some fixed interval. Wang and Chang (1999) considered a weighted moment estimator for the event time distribution that ignores the last censored observation on all cases experiencing at least one event, and argue for consistency and weak convergence of their estimator using empirical process techniques.

The data accrual and censoring schemes to be considered in this paper are straightforward. In particular, except in Section 5, we assume that event interoccurrence times represent independent and identically distributed observations from a fixed distribution function, and that each subject is observed for a random length of time. Save the independence assumption, the situation being considered is identical to that in Wang and Chang (1999), and is appropriate e.g. when the initial occurrence of the event is also the criterion for admission in the study. It

also differs from that Vardi (1982) and Soon and Woodroffe (1996), who impose the simplifying assumption that the renewal processes under study are stationary (e.g., Resnick, 1994, §3.9). The i.i.d. model is actually quite reasonable in many engineering and reliability settings, and has been the subject of extensive studies in the probability and stochastic processes literature. Nevertheless, our assumptions are acknowledged as being restrictive in the biomedical context, and here we point out at least three interesting avenues of generalization: (i) correlated interoccurrence times through the use of frailty-type models; (ii) the dependence of the interoccurrence time distribution on relevant covariates; and (iii) allowing the distribution of the first event occurrence to be different from the succeeding interoccurrence time distributions, but without imposing that it be the stationary distribution. This paper provides a useful starting point for developing appropriate likelihood-based estimation procedures for each of these generalizations; the case of correlated recurrence times is considered in further detail in Section 5.

The seemingly straightforward nature of the data accrual scheme belies some important technical issues that differentiate the single event and recurrent event settings. For example, since the observation period for a given subject is random, the frequency of event occurrences is a random variable whose distribution depends on the event time distribution. This remains true even when the distribution of the random termination time does not depend on the event time distribution. Thus, aside from the event times themselves, the number of events during the observation period is also informative concerning the distribution of interest. Moreover, the last observation for each subject is always right-censored, and depends both on the length of the observation interval and on previous event times for that subject. Consequently censoring is in general informative. These features prevent the usual tools from martingale theory from being used in studying the asymptotic behavior of estimators that use all of the available information. An excellent example of this type of data is the gastroenterology data presented in Aalen and Husebye (1991). Other papers that deal with related problems under this basic data accrual scheme include Prentice, Williams and Peterson (1981), Andersen and Gill (1982) and Keiding, Andersen and Fledelius (1998). Still other papers dealing with recurrent data and motivated by engineering and reliability problems are those by Kalbfleisch, Lawless and Robinson (1991), Lawless and Nadeau (1995), and Hu and Lawless (1996).

Perhaps the most important difference between the present paper and Wang and Chang (1999) is the way in which the right censored observations are handled. The moment estimator of Wang and Chang (1999) only utilizes the right-censored observation for a subject if there are no complete observations for that subject, hence ignores information. The present paper provides asymptotic theory for the nonparametric maximum likelihood estimator (NPMLE) of the underlying interoccurrence time distribution, and utilizes the right-censored observations on all subjects. This is accomplished by embedding the problem into a nonstandard counting process framework that employs both calendar time and duration time. In addition to allowing for easy likelihood construction, this approach allows us to employ tools from martingale theory that would be otherwise unavailable. The idea of considering a bivariate time scale is borrowed

from Sellke (1988); however, the problem being considered here differs in a fundamental way. In particular, Sellke (1988) considers the observation of a single renewal process, and studies the asymptotics of the resulting Nelson-Aalen estimator of the interarrival time distribution as the length of a deterministic observation period becomes unbounded. In contrast, we desire to estimate the event time distribution based on n independently censored realizations of a renewal process. The observation period for each process is governed by a censoring distribution that can have finite support (i.e., the length of the observation period can be, and usually is, bounded) and the asymptotics in question relate to the number of subjects under study, not time on observation. These changes complicate technical matters significantly. Our work here complements that of Gill (1980), who considers nonparametric estimation for a class of semi-Markov renewal processes. Careful inspection shows that his results also apply to the present problem; however, consistency and weak convergence are proved under stronger and less familiar technical conditions. Restricting attention to estimation of the interoccurrence distribution from recurrent event data makes the problem and corresponding solution more transparent, and in particular places inference for the interoccurrence time distribution in terms that effectively parallel the single event setting. For example, in addition to establishing consistency and weak convergence of our proposed estimators, we obtain closed-form expressions for the limiting variance of these estimators that properly takes informative stopping into account. The renewal function associated with interoccurrence time distribution plays a central role in these results. These various contributions fill what we perceive to be an important void in the current theory of estimation for recurrent event data, and the formulation to be presented allows for reasonably straightforward extensions to both regression and frailty models for recurrent event data.

We now outline the contents of this paper. In Section 2 we define relevant processes and summarize important results that will be used throughout the remainder of this paper. Derivations of our estimators, as well as corresponding finite sample and asymptotic properties, are facilitated through this stochastic process formulation and a general weak convergence result in a related paper by Peña, Strawderman and Hollander (PSH) (2000). Section 3 considers nonparametric estimation of the hazard function Λ and the survival function \bar{F} . An estimator for Λ is motivated using method-of-moments, and product-integration is used to obtain an estimator for \bar{F} . These estimators turn out to be natural generalizations of the Nelson-Aalen and the product-limit estimators, as has also been seen in related recurrent models, cf., Gill (1980, 1981), Andersen and Gill (1982), Vardi (1982b), Sellke (1988), and Keiding, *et al.* (1998). We establish that these estimators are NPMLs, and derive their asymptotic properties; in particular, bias, variance, uniform consistency and weak convergence are all established under conditions that parallel those required for the single event setting (e.g. Andersen, Borgan, Gill and Keiding, 1993). In Section 4, we evaluate the consequences of ignoring (i) all but the first and (ii) the right-censored observations, approaches common to the medical literature involving recurrent event data (cf., Aalen and Husebye, 1991). In Section 5, we evaluate the consistency of the product limit estimator in the case where the interoccurrence times are correlated within each unit and propose a new estimator for the marginal survivor function. Finally, in Sections 6 and 7 we respectively

summarize the results of a simulation study and illustrate these estimation procedures using the gastroenterology data of Aalen and Husebye (1991). The simulation study compares biases and mean-squared errors of our estimators to those of an estimator recently proposed in Wang and Chang (1999).

2 Stochastic Process Formulation

2.1 Introduction

In order to formally describe the problem of interest let us introduce some notation. We suppose that n independent units (e.g., subjects) are available in the study. For the i th unit, we denote the successive interoccurrence times of the recurrent event of interest by $\{T_{ik}, k = 1, 2, \dots\}$. The interoccurrence times are assumed to be independent and identically distributed (i.i.d.) nonnegative random variables with a common absolutely continuous distribution function $F(t) = \mathbf{P}\{T_{ij} \leq t\}$ and associated renewal function $\rho(t) = \sum_{k=1}^{\infty} F^{*k}(t)I\{t \geq 0\}$, where F^{*k} denotes the k th convolution of F with itself. The cumulative hazard function associated with $F(\cdot)$ is given by $\Lambda(t) = \int_0^t \lambda(s)ds$, where $\lambda(s) = f(s)/\bar{F}(s)$, $f(s) = dF(s)/ds$, and $\bar{F}(s) = 1 - F(s)$ is the survival function. We assume that monitoring of the i th unit ceases at a possibly random time τ_i , where $\tau_1, \tau_2, \dots, \tau_n$ are i.i.d. with a common distribution function $G(w) = \mathbf{P}\{\tau_i \leq w\}$. We also assume that the τ_i 's and the T_{ij} 's are mutually independent. We allow for the possibility that G is degenerate, as would be the case if the observation window for all units were restricted to be of the same length; hence, for F continuous our results contain those of Gill (1981) as a special case. Finally, we will assume that $E[\rho(\tau)] < \infty$. This condition basically requires that the average number of events observed is finite. Note, however, that such a condition does not necessarily imply $G(\cdot)$ has finite support since it is satisfied in the trivial case where T_{ij} and τ represent mutually independent exponential random variables.

For each $i = 1, \dots, n$, let $S_{i0} = 0$ and $S_{ij} = \sum_{l=1}^j T_{il}, j = 1, 2, \dots$. The number of event occurrences for the i th unit is

$$K_i = \max\{k \in \{0, 1, \dots\} : S_{ik} \leq \tau_i\}, \quad (1)$$

and the observable random variables for the i th unit are

$$(K_i, \tau_i, T_{i1}, T_{i2}, \dots, T_{iK_i}, \tau_i - S_{iK_i}). \quad (2)$$

Specification of the last observation time $\tau_i - S_{iK_i}$ is redundant since it is determined by the other observable variables. It is retained, however, to emphasize the fact that this variable right-censors T_{iK_i+1} . Assuming the data take the form (2) for $i = 1, 2, \dots, n$, we rigorously develop and obtain the properties of various nonparametric estimators of F . We will require some further notation to facilitate this analysis, and present this in the next section.

2.2 Counting Processes and Martingales

In this section we recast the mathematical setup into a stochastic process setting. We let $(\Omega, \mathcal{F}, \mathbf{P})$ denote the probability space on which all random entities are defined. Before proceeding with the appropriate stochastic process formulation, we first describe an intuitively appealing way of defining the stochastic processes and filtration. This construction turns out to have an important flaw that prevents the direct use of martingale methods. It also serves to introduce some useful notation, and allows us to gain crucial understanding into where some of the important technical issues arise.

For $j = 1, 2, \dots$ and $i = 1, 2, \dots, n$ define $C_{ij} = \tau_i - S_{ij-1}$,

$$N_{ij}(t) = I\{T_{ij} \leq t, T_{ij} \leq C_{ij}\} \quad \text{and} \quad Y_{ij}(t) = I\{T_{ij} \geq t, C_{ij} \geq t\},$$

where the processes $N_{ij}(t)$ and $Y_{ij}(t)$ are defined for $t \geq 0$, and we assume $N_{ij}(0) = 0$ and $Y_{ij}(0) = 1$ almost surely. Furthermore, with \mathcal{F}_0 being the σ -field containing all information available at time zero including τ_i , define $\mathcal{F}_{ijt} = \sigma\{(N_{ij}(s), Y_{ij}(s+)) : 0 \leq s \leq t\} \vee \mathcal{F}_0$. Since T_{ij} and C_{ij} are independent, $N_{ij}(\cdot)$ and $Y_{ij}(\cdot)$ are just the usual counting and at-risk processes associated with a failure time T_{ij} possibly right-censored by C_{ij} . As a consequence,

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s) d\Lambda(s), \quad t \geq 0, \quad (3)$$

is a local square-integrable martingale with respect to the filtration $\{\mathcal{F}_{ijt}, t \geq 0\}$ (cf., Fleming and Harrington, 1991). Finally, for the i th unit, we may define aggregated processes

$$N_i(t) = \sum_{j=1}^{\infty} N_{ij}(t), \quad Y_i(t) = \sum_{j=1}^{\infty} Y_{ij}(t), \quad \text{and} \quad M_i(t) = \sum_{j=1}^{\infty} M_{ij}(t), \quad (4)$$

and the associated filtration $\{\mathcal{F}_{it} = \bigvee_{j=1}^{\infty} \mathcal{F}_{ijt}, t \geq 0\}$.

The time scale under consideration in the above construction refers solely to “gap times” between successive events, and hence $\{\mathcal{F}_t, t \geq 0\}$ really represents a “gap time” filtration. Although this is clearly the natural time scale from the point of view of making inference about the interoccurrence time distribution, the stochastic processes and the filtration just defined do not, in fact, correspond to the way the recurrent event data are truly observed. This proves to be the main source of difficulty with the present construction, as $M_i(t)$ need not be a martingale with respect to the filtration $\{\mathcal{F}_{it} : t \geq 0\}$, and thus $\mathbf{M}(t) = (M_1(t), \dots, M_n(t))$ need not be a vector of martingales with respect to the filtration $\mathbf{F} = \{\mathcal{F}_t : t \geq 0\}$ for $\mathcal{F}_t = \bigvee_{i=1}^n \mathcal{F}_{it}$. To see where the martingale property breaks down, notice first that

$$\mathbf{E}\{M_{ij}(t+s)|\mathcal{F}_t\} = \mathbf{E}[\mathbf{E}\{M_{ij}(t+s)|\mathcal{F}_{it} \vee \mathcal{F}_t\}|\mathcal{F}_t] = \mathbf{E}[\mathbf{E}\{M_{ij}(t+s)|\mathcal{F}_{it}\}|\mathcal{F}_t].$$

However,

$$\mathbf{E}\{M_{ij}(t+s)|\mathcal{F}_{it}\} = \mathbf{E}[\mathbf{E}\{M_{ij}(t+s)|\mathcal{F}_{ijt} \vee \bigvee_{k \neq j} \mathcal{F}_{ikt}\}|\mathcal{F}_{it}] \neq \mathbf{E}[\mathbf{E}\{M_{ij}(t+s)|\mathcal{F}_{ijt}\}|\mathcal{F}_{it}] = M_{ij}(t),$$

since, for $k > j$, $M_{ij}(t+s)$ is not independent of \mathcal{F}_{ikt} by virtue of the fact that C_{ik} depends on T_{i1}, \dots, T_{ij} . Thus, in general, $\mathbf{E}\{M_{ij}(t+s)|\mathcal{F}_{it}\}$ need not equal $M_{ij}(t)$, implying that $\{M_i(t) : t \geq 0\}$ is not an \mathbf{F} -martingale. Although it is not difficult to come up with a concrete example illustrating the above situation, we leave this task to the reader.

As shown in Sellke (1988), the stochastic processes one should consider turn out to be those indexed by both gap time and calendar time. These processes arise by first considering the processes that keep track of event occurrences as calendar time progresses. We begin by defining the following processes, which consider calendar time only. For $i = 1 \dots n$ and $s \geq 0$ set

$$N_i^\dagger(s) = \sum_{j=1}^{\infty} I\{S_{ij} \leq s, S_{ij} \leq \tau_i\} \quad \text{and} \quad Y_i^\dagger(s) = I\{\tau_i \geq s\},$$

and let $\mathbf{G} = \{\mathcal{G}_s : s \geq 0\}$ be a filtration such that $\{(N_i^\dagger(s), Y_i^\dagger(s)) : s \geq 0\}$ is \mathbf{G} -adapted. Moreover, let $A_i^\dagger(s) = \int_0^s Y_i^\dagger(v) \lambda(v - S_{iN_i^\dagger(v-)}) dv$, and suppose $M_i^\dagger(s) = N_i^\dagger(s) - A_i^\dagger(s)$, $s \geq 0$ is a local square-integrable \mathbf{G} -martingale with predictable quadratic covariation process $\langle M_i^\dagger, M_{i'}^\dagger \rangle(s) = A_i^\dagger(s) I\{i = i'\}$. In particular, this holds if \mathbf{G} is taken to be the natural filtration generated by $\{(N_i^\dagger(s), Y_i^\dagger(s)) : s \geq 0, i = 1, 2, \dots, n\}$, i.e., $\mathcal{G}_s = \mathcal{F}_s^\dagger = \mathcal{F}_0 \vee \sigma\{(N_i^\dagger(v), Y_i^\dagger(v+)) : v \leq s, i = 1, 2, \dots, n\}$, with \mathcal{F}_0 being the σ -field containing all information at time zero. In fact, as implied in Gill (1980), if $\lim_{s \rightarrow \infty} \mathbf{E}\{A_i^\dagger(s)\} < \infty$, $\{M_i^\dagger(s), s \geq 0\}$ is a square integrable \mathbf{G} -martingale. The assumption that $\mathbf{E}\{\rho(\tau)\} < \infty$ ensures this as $\mathbf{E}\{A_i^\dagger(s)\} = \mathbf{E}\{N_i^\dagger(s)\} = \mathbf{E}\{\sum_{j=1}^{\infty} I\{S_{ij} \leq s \wedge \tau_i\}\} = \mathbf{E}\{\rho(\tau \wedge s)\}$.

We now introduce appropriate processes that are indexed by calendar time s and gap time t . These are the basic processes considered in PSH (2000; see also Sellke, 1988) and provide the crucial connection between the gap time formulation and that based on calendar time. For $i = 1, 2, \dots, n$, let $Z_i(s, t) = I\{s - S_{iN_i^\dagger(s-)} \leq t\}$. For a given unit (e.g., individual) and calendar time s , this indicates whether at most t time units have elapsed since the time of the last event. Note that for fixed t , $Z_i(\cdot, t)$ is a \mathbf{G} -adapted, bounded, and has left-continuous paths, hence is a \mathbf{G} -predictable bounded process. For $s, t \geq 0$ we may thus define

$$N_i(s, t) = \int_0^s Z_i(v, t) dN_i^\dagger(dv); \tag{5}$$

$$A_i(s, t) = \int_0^s Z_i(v, t) A_i^\dagger(dv); \tag{6}$$

$$M_i(s, t) = \int_0^s Z_i(v, t) M_i^\dagger(dv) = N_i(s, t) - A_i(s, t). \tag{7}$$

Evidently, for the i th unit, $N_i(s, t)$ counts the number of observed events occurring over the calendar period $[0, s]$ whose event times (i.e., interoccurrence times) are at most t . The notation would suggest that $A_i(s, t)$ is the compensator of $N_i(s, t)$ and therefore $M_i(s, t)$ is a martingale. This is true, in a sense to be discussed later. However, we first provide an alternative form for (6), which as written does not have the “usual” multiplicative form of a counting process

compensator (cf. Fleming and Harrington, 1991). Towards this end, define

$$Y_i(s, t) = \sum_{j=1}^{N_i^\dagger((s \wedge \tau_i)-)} I\{T_{ij} \geq t\} + I\{(s \wedge \tau_i) - S_{iN_i^\dagger((s \wedge \tau_i)-)} \geq t\}, i = 1, 2, \dots, n. \quad (8)$$

Then, it can be shown that (PSH, 2000, Proposition 1)

$$A_i(s, t) = \int_0^t Y_i(s, w) \lambda(w) dw, \quad (9)$$

which is of the usual multiplicative intensity form. Notice that $Y_i(s, t)$ counts the number of observed events on calendar period $[0, s]$ whose event times were at least t . Moreover, for each fixed t , the s -indexed stochastic process $Y_i(\cdot, t)$ has left-continuous sample paths and is therefore \mathbf{G} -predictable. These facts, combined with the relation (9), show that $Y_i(s, t)$ serves as the basic “at-risk” process corresponding to the i^{th} doubly-indexed process $N_i(s, t)$.

It should also be apparent that the processes $N_i(s, t)$ and $Y_i(s, t)$ are related to the original processes $N_i(t)$ and $Y_i(t)$ defined in (4). In particular, notice that $N_i(s, t) = N_i((s \wedge \tau_i), t)$ for each $s \geq 0, t \geq 0$; thus, as $s \rightarrow \infty$, $N_i(s, t) \xrightarrow{\text{a.S.}} N_i(\tau_i, t) \equiv N_i(t)$. Moreover, $N_i^\dagger((s \wedge \tau_i)-) \xrightarrow{\text{a.S.}} N_i^\dagger(\tau_i-)$ $\stackrel{\text{a.S.}}{\equiv} K_i$ as $s \rightarrow \infty$; hence, $Y_i(s, t) \xrightarrow{\text{a.S.}} \sum_{j=1}^{K_i} I\{T_{ij} \geq t\} + I\{\tau_i - S_{iK_i} \geq t\} \equiv Y_i(t)$ as $s \rightarrow \infty$.

PSH (2000) develop key properties of the aforementioned processes in further detail. The following important result is one that we shall use repeatedly, and connects stochastic integrals over calendar time to those over gap times. The proof is a direct consequence of the argument used to prove Proposition 1 of PSH (2000) and is therefore omitted.

Proposition 1 *Let $R_i(v) = v - S_{iN_i^\dagger(v-)}$, and let $H_i(s, R_i(v))$ be a bounded \mathbf{G} -predictable process. Then, for $Q_i(s, t)$ given by one of $N_i(s, t)$, $A_i(s, t)$, or $M_i(s, t)$ (and Q_i^\dagger defined analogously),*

$$\int_0^s H_i(s, R_i(v)) Q_i(dv, t) \equiv \int_0^s H_i(s, R_i(v)) Z_i(v, t) Q_i^\dagger(dv) \equiv \int_0^t H_i(s, w) Q_i(s, dw).$$

A trivial consequence of these results is $\int_0^s H_i(s, R_i(v)) M_i(dv, t) = \int_0^t H_i(s, w) M_i(s, dw)$; (7) and (9) then imply $\int_0^s H_i(s, R_i(v)) M_i(dv, t) = \int_0^t H_i(s, w) N_i(s, dw) - \int_0^t H_i(s, w) Y_i(s, w) \lambda(w) dw$. With this basic framework in place, we may now define the aggregated processes

$$N(s, t) = \sum_{i=1}^n N_i(s, t), \quad A(s, t) = \sum_{i=1}^n A_i(s, t), \quad \text{and} \quad M(s, t) = \sum_{i=1}^n M_i(s, t). \quad (10)$$

By (9), $A(s, t) = \sum_{i=1}^n A_i(s, t) = \int_0^t Y(s, w) \lambda(w) dw$, where $Y(s, t) = \sum_{i=1}^n Y_i(s, t)$; consequently,

$$M(s, t) = \sum_{i=1}^n M_i(s, t) = N(s, t) - A(s, t) = N(s, t) - \int_0^t Y(s, w) \lambda(w) dw \quad (11)$$

has the usual martingale form. In fact, the following holds under the assumptions of this paper.

Proposition 2 For each fixed $t \geq 0$, the aggregated process $\{M(s, t), s \geq 0\}$ is a square-integrable \mathbf{G} -martingale with predictable quadratic variation process

$$\langle M(\cdot, t), M(\cdot, t) \rangle(s) = A(s, t) = \int_0^t Y(s, w) \lambda(w) dw < \infty.$$

Importantly, $M(s, t)$ is not a martingale process in t for any fixed s . Nevertheless, in light of earlier results, stochastic integrals of the form $\int_0^t H(s, w) M(s, dw)$ can still be handled using martingale methods. In particular, by Proposition 1,

$$\sum_i \int_0^t H_i(s, w) M_i(s, dw) = \sum_i \int_0^s H_i(s, R_i(v)) M_i(dv, t),$$

where the right-hand side is, for each fixed $t \geq 0$, a square-integrable \mathbf{G} -martingale process.

For reasons to be explained in Section 3, the relationships between the various processes just established are of limited utility in the case where F is to be modeled parametrically. However, they prove to be essential in developing nonparametric estimators for \bar{F} .

2.3 Expectation of the At-Risk Processes

In order to study the finite sample and asymptotic properties in the nonparametric setting, we require the expectation of the at-risk processes $Y_i(s, t)$ and $Y_i(t)$. We recall that the aggregated at-risk processes for the i th unit can be expressed as follows:

$$Y_i(s, t) = \sum_{j=1}^{N_i^\dagger((s \wedge \tau_i)-)} I\{T_{ij} \geq t\} + I\{(s \wedge \tau_i) - S_{iN_i^\dagger((s \wedge \tau_i)-)} \geq t\} \quad (12)$$

$$Y_i(t) = \sum_{j=1}^{K_i} I\{T_{ij} \geq t\} + I\{\tau_i - S_{iK_i} \geq t\}, \quad (13)$$

where K_i is as defined in (1). Evidently, since $Y_1(\cdot), \dots, Y_n(\cdot)$ are independent and identically distributed processes and $Y_i(t) = Y_i(\infty, t)$, $i = 1 \dots n$, it suffices to consider $Y_1(s, t)$, $s, t \geq 0$. Let $\tau_i(s) \equiv s \wedge \tau_i$; then, if τ_i has distribution function $G(\cdot)$, the distribution function of $\tau_i(s)$ is $G_s(w) = G(w)I\{w < s\} + I\{w \geq s\}$. The following result is analogous to Lemma 1 of Gill (1981) and is proved in the Appendix using a renewal argument; it is helpful to recall here that $\rho(t)$ denotes the renewal function associated with the event interoccurrence time distribution.

Proposition 3 Suppose $F(0) = 0$. Then, for each $s, t \geq 0$ and with $\bar{G}_s(t) = 1 - G_s(t)$,

$$\begin{aligned} \mathbf{E}\{Y_1(s, t)\} = y(s, t) &\equiv \bar{F}(t)\bar{G}_s(t-) \{1 + \mathbf{E}\{\rho(\tau(s) - t) | \tau(s) \geq t\}\} \\ &= \bar{F}(t)\bar{G}_s(t-) + \bar{F}(t) \int_t^\infty \rho(w - t) dG_s(w). \end{aligned}$$

Furthermore, $y(s, t) = 0$ for $s < t$, $y(s, t) > 0$ if $t \in \{u : F(u-) < 1, G(u-) < 1, u \leq s\}$, and $\lim_{s \rightarrow \infty} y(s, t) = y(t)$, where $y(t) = \mathbf{E}\{Y_1(t)\} = \bar{F}(t)\bar{G}(t-) + \bar{F}(t) \int_t^\infty \rho(w - t) dG(w)$ and $y(0) < \infty$.

Notice that the form of the expected at-risk process $y(t)$ directly generalizes that under the usual single-event setting, and in particular demonstrates how the extra information available from having multiple events per unit increases the effective number at risk. The presence of the renewal function is quite natural here, as $\rho(t)$ simply equals the expected number of events on $[0, t]$ for one unit. The following Glivenko-Cantelli-type theorem shows we may estimate $y(s, t)$ via $n^{-1}Y(s, t)$:

Proposition 4 *For each $s \geq 0$ $\sup_{t \in [0, \infty]} \left| \frac{1}{n} \sum_{i=1}^n Y_i(s, t) - y(s, t) \right| \rightarrow 0$ a.s. as $n \rightarrow \infty$.*

The fact that this holds uniformly in $t \geq 0$ is a consequence of the condition $E[\rho(\tau)] < \infty$ and the uniform law of large numbers of Hoffman-Jorgensen (1994, §9.15); for details see PSH (2000, Proposition 2).

Finally, in order to examine the bias that occurs if right-censored observations are not included in the analysis, we shall also need $\mathbf{E}\{Y_1^*(\infty, t)\}$, where $Y_1^*(\infty, t) = \sum_{j=1}^{K_1} I\{T_{1j} \geq t\}$ denotes the at-risk process that ignores the right-censored observation. Recognizing that $Y_1^*(\infty, t) = Y_1(\infty, t) - I\{\tau_1 - S_{1K_1} \geq t\}$, the following result provides the requisite expectation.

Proposition 5 *Suppose $t \geq 0$ and $F(0) = 0$. Then,*

$$\mathbf{E}\{Y_1^*(\infty, t)\} = y^*(\infty, t) \equiv y(\infty, t) - \mathbf{P}\{\tau_1 - S_{1K_1} \geq t\}, \quad (14)$$

where $\mathbf{P}\{\tau_1 - S_{1K_1} \geq t\} = \bar{G}(t-) \mathbf{E}\left\{ \bar{F}(\tau_1) + \int_0^{\tau_1 - t} \bar{F}(\tau_1 - w) d\rho(w) \mid \tau_1 \geq t \right\}$.

2.4 A General Weak Convergence Result

To establish the weak convergence of various martingale transforms we will come across, we shall need a result due to PSH (2000, Theorem 1). Theorem 1 of PSH (2000) actually assumes the length of the observation period on the calendar time scale (i.e., s) to be finite. Careful inspection of their proof shows that this assumption is unnecessary, and that with minor changes the results continue to hold as the calendar time $s \rightarrow \infty$. For convenience, a more general version of their result is thus presented below.

With n indexing sample size, consider a sequence of processes of the form

$$\mathbf{W}^{(n)}(s, t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \mathbf{H}_i^{(n)}(s, w) dM_i^{(n)}(s, dw), \quad (15)$$

where $M_i^{(n)}(s, t)$ is defined analogously to (7) and $\mathbf{H}_i^{(n)}(s, t) = (H_{1i}^{(n)}(s, t), \dots, H_{pi}^{(n)}(s, t))^t$ is, for each n , a vector-valued process such that (i) $H_{ji}^{(n)}(\cdot, \cdot)$ is almost surely bounded; and (ii) $\mathbf{H}_i^{(n)}(s, t)$ is \mathbf{G} -predictable for each t . As described in Section 2.2, $\mathbf{W}^{(n)}(s, t)$ is not a martingale transform since we are integrating $dM_i^{(n)}(s, dw)$ over gap times w , fixing calendar time s , rather

than integrating over calendar time s , fixing w . Despite this fact, the following general weak convergence result still holds.

Theorem 1 *Let $S = [0, s^*]$, $s^* \leq \infty$ and $y(s, t)$ be defined as in Proposition 3. Fix $s \in S$ and suppose the following conditions are satisfied for $t, t_1, t_2 \in [0, t^*]$, where $t^* < \infty$:*

(a) *The processes $\{\mathbf{H}_i(v, w) : 0 \leq v \leq s; 0 \leq w \leq t^*\}$ are left-continuous and bounded, and there exists a continuous bounded deterministic function $\mathbf{h}(v, w)$ on $[0, s] \times [0, t^*]$ such that*

$$\max_{1 \leq i \leq n} \sup_{0 \leq w \leq t^*} |\mathbf{H}_i(s, w) - \mathbf{h}(s, w)| \xrightarrow{\text{pr}} 0; \quad (16)$$

(b) $\inf_{w \in [0, t^*]} y(s, w) > 0$;

(c) *The matrix function $\Sigma(s, t) = \int_0^t \mathbf{h}(s, w)^{\otimes 2} y(s, w) \lambda(w) dw$ is such that for each $t_1, t_2 \in (0, t^*]$ with $t_1 < t_2$, $0 < \det\{\Sigma(s, t_2) - \Sigma(s, t_1)\} < \infty$;*

(d) *For each t , $\|\mathbf{V}^{(n)}(s, t) - \Sigma(s, t)\| \xrightarrow{\text{pr}} 0$ as $n \rightarrow \infty$, where*

$$\mathbf{V}^{(n)}(s, t) = \frac{1}{n} \sum_{i=1}^n \int_0^t [\mathbf{H}_i^{(n)}(s, w)]^{\otimes 2} Y_i^{(n)}(s, w) \lambda(w) dw.$$

Then, as $n \rightarrow \infty$, the process $\{\mathbf{W}^{(n)}(s, t) : t \in [0, t^]\}$ converges weakly on Skorohod's space $\mathcal{D}[0, t^*]$ to a mean zero Gaussian process $\{\mathbf{W}^{(\infty)}(s, t) : t \in [0, t^*]\}$ with covariance function*

$$\text{Cov}\{\mathbf{W}^{(\infty)}(s, t_1), \mathbf{W}^{(\infty)}(s, t_2)\} = \begin{bmatrix} \Sigma(s, t_1) & \Sigma(s, t_1) \\ \Sigma(s, t_1) & \Sigma(s, t_2) \end{bmatrix} \text{ for } t_1 \leq t_2.$$

Conditions (a)-(d) ensure convergence of the finite dimensional distributions and tightness, and in particular establish that the limit process is a multidimensional Gaussian process with a continuous variance function. However, despite apparent similarities, the above result is not a simple consequence of Rebolledo's martingale central limit (Rebolledo, 1980; see also Andersen *et al*, 1993, Theorem II.5.1). Moreover, it is important to note here that the above result does not establish that $\{\mathbf{W}^{(n)}(s, t) : t \in [0, t^*]\}$ converges weakly as a process in both s and t ; rather, it only establishes weak convergence in t , given a fixed value of $s \in S$.

3 Nonparametric Estimation

In this section we derive nonparametric estimators for the hazard function Λ and the interoccurrence distribution F . We first establish Nelson-Aalen and Kaplan-Meier-type estimators using the method of moments, and then motivate why these are respectively NPMLs for $\Lambda(\cdot)$ and $\bar{F}(\cdot)$. In what follows, we allow for the possibility that $s = \infty$.

3.1 Method-of-Moments Estimators

Let $M(s, t)$ be defined as in (11). Then, with $H_i(s, t) \equiv 1$, $i = 1, 2, \dots, n$, results presented in Section 2.2 imply that $\left\{M(v, t) = N(v, t) - \int_0^t Y(v, w)\lambda(w)dw : 0 \leq v \leq s\right\}$, is a square-integrable martingale for each fixed $t \geq 0$ (PSH, 2000, Proposition 3). Let $J(v, w) = I\{Y(v, w) > 0\}$. With the usual convention that $0/0 = 0$, we may thus write

$$\int_0^t \frac{J(s, w)}{Y(s, w)} M(s, dw) = \int_0^t \frac{J(s, w)}{Y(s, w)} N(s, dw) - \int_0^t J(s, w)\Lambda(dw). \quad (17)$$

Let us examine the left-hand side of (17). By Proposition 1,

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

where we recall the notation $R_i(v) = v - S_{iN_i^+(v-)}$, $i = 1, 2, \dots, n$. Since the integrand is a bounded predictable process by virtue of the left-continuity of $Y_i(s, t)$ in both s and t , we have that $\left\{\sum_{i=1}^n \int_0^s [J(s, R_i(v))/Y(s, R_i(v))] M_i(dv, t) : 0 \leq s < \infty\right\}$ is a zero-mean square-integrable martingale for each fixed $t \geq 0$. Hence,

$$\mathbf{E} \left\{ \int_0^t \frac{J(s, w)}{Y(s, w)} N(s, dw) \right\} = \mathbf{E} \left\{ \int_0^t J(s, w)\Lambda(dw) \right\}. \quad (18)$$

This moment identity suggests a Nelson-Aalen-type estimator of $\Lambda(t)$; specifically, for a given study period $[0, s]$, set

$$\hat{\Lambda}(s, t) = \int_0^t \frac{J(s, w)}{Y(s, w)} N(s, dw), \quad 0 \leq t < \infty. \quad (19)$$

Since $\bar{F}(t) = \prod_{w \leq t} [1 - \Lambda(dw)]$, a product-limit estimator for \bar{F} is then easily obtained:

$$\hat{\hat{F}}(s, t) = \prod_{w \leq t} \left\{ 1 - \hat{\Lambda}(s, dw) \right\} = \prod_{w \leq t} \left\{ 1 - \frac{N(s, \Delta w)}{Y(s, w)} \right\}, \quad (20)$$

where $\hat{\hat{F}}(s, t) = \hat{\hat{F}}(s, B_{n,s})$ for $t > B_{n,s} = \inf\{w : Y(s, w) = 0\}$. It is evident that $\hat{\Lambda}(s, t)$ and $\hat{\hat{F}}(s, t)$ are left-continuous in s and right-continuous in t . These properties are crucial, and allow us to exploit stochastic integration theory for obtaining finite sample and asymptotic properties in the sequel.

REMARK: By (19) and results from Section 2.2,

$$\lim_{s \rightarrow \infty} \hat{\Lambda}(s, t) = \int_0^t [J(w)/Y(w)] dN(w) \equiv \hat{\Lambda}(t). \quad \text{a.s.} \quad (21)$$

Then,

$$\hat{\hat{F}}(\infty, t) \equiv \hat{\hat{F}}(t) = \prod_{w \leq t} \left\{ 1 - \hat{\Lambda}(dw) \right\} = \prod_{w \leq t} \left\{ 1 - \frac{N(\Delta w)}{Y(w)} \right\}, \quad (22)$$

where $\hat{\hat{F}}(t) = \hat{\hat{F}}(B_n)$ for $t > B_n = \inf\{w : Y(w) = 0\}$. These estimators continue to remain applicable even if ties are present. They also constitute a special case of those considered by Gill (1980), and generalize those derived in Gill (1981).

3.2 $\hat{\Lambda}(s, t)$ & $\hat{F}(s, t)$ are NPMLEs

Following Jacod (1975) (see also Section II.7 of Andersen *et al.*, 1993), define

$$L(s) = \prod_{v \in [0, s]} \prod_{i=1}^n \left\{ \left[\lambda \left(v - S_{iN_i^\dagger(v-)} \right) \right]^{N_i^\dagger(\Delta v)} \exp \left(- \int_0^s Y_i^\dagger(w) \lambda \left(w - S_{iN_i^\dagger(w-)} \right) dw \right) \right\}. \quad (23)$$

This general partial likelihood process over the study period $[0, s]$ is proportional to the relevant point process likelihood for the problem being considered. From this, we obtain the associated log-likelihood process

$$l(s) = \sum_{i=1}^n \int_0^s \left\{ \log \left(\lambda \left(v - S_{iN_i^\dagger(v-)} \right) \right) N_i^\dagger(dv) - Y_i^\dagger(v) \lambda \left(v - S_{iN_i^\dagger(v-)} \right) dv \right\}. \quad (24)$$

This log-likelihood does not correspond to a multiplicative intensity model for the intensity process. Moreover, it is difficult to maximize (24) with respect to $\Lambda(\cdot)$ directly. However, similarly to the single event setting, it is intuitively clear that nonparametric estimation of $\Lambda(t)$ for a fixed t should involve only those gap times that are less than or equal t .

Let us therefore consider a modified log-likelihood process based on the information that has accrued in the study (or real) time $[0, s]$ and duration (or gap) time $[0, t]$ space. Specifically, let

$$l(s, t) = \sum_{i=1}^n \int_0^s Z_i(v, t) \left\{ \log \left(\lambda \left(v - S_{iN_i^\dagger(v-)} \right) \right) N_i^\dagger(dv) - Y_i^\dagger(v) \lambda \left(v - S_{iN_i^\dagger(v-)} \right) dv \right\}, \quad (25)$$

where we recall that $Z_i(v, t) = I\{v - S_{iN_i^\dagger(v-)} \leq t\}$. Notice that (25) is just a gap-time-restricted version of (24). Using the fact that $M_i^\dagger(dv) = N_i^\dagger(dv) - Y_i^\dagger(v) \lambda \left(v - S_{iN_i^\dagger(v-)} \right) dv$ and applying Proposition 1, (25) is easily shown to be equivalent to

$$l(s, t) = \sum_{i=1}^n \int_0^t \log \lambda(w) N_i(s, dw) - Y_i(s, w) \lambda(w) dw. \quad (26)$$

The restricted likelihood (25) establishes an interesting and useful link between (24) and (26). For example, from (26), it can be seen directly that (19) and hence (20) are NPMLEs based on the information collected to calendar time s . Moreover, (21) and (22) are respectively NPMLEs, since $\hat{\Lambda}(t)$ maximizes the log-likelihood function

$$l(\infty, t) = \sum_{i=1}^n \int_0^t \log \lambda(w) N_i(dw) - Y_i(w) \lambda(w) dw. \quad (27)$$

It is not difficult to see that $l(\infty, t)$ may also be derived by considering all gap times as independent, possibly right-censored observations from the interoccurrence distribution $F(\cdot)$, as is implied in the discussions of Aalen and Husebye (1991), Andersen *et al.* (1993, Example X.1.8), Gill (1983), and Keiding *et al.* (1998, §5.1).

At this point we digress for a moment and point out an interesting distinction between the parametric and nonparametric settings. In the case where $\lambda(\cdot)$ is parametrically specified, parametric

maximum likelihood estimation can be carried out based on either (24) or (26); see Andersen *et al.* (1993, Example X.1.9) for more discussion regarding the use of (24). Above, we have established that (26) and (25) are equivalent, and it can thus be seen that inference based on (25) will generally be less efficient than that based on (24). Passing to the limit (i.e., as $s \rightarrow \infty$) should not alter this ordering of efficiency, and hence estimators based on (24) for $s = \infty$ should be more efficient than those based on (27).

3.3 Bias and Variance

By following the derivation of the bias of the Nelson-Aalen estimator when the data consist of right-censored times to the first event occurrence, (e.g., see Fleming and Harrington, 1991), the bias of $\hat{\Lambda}(s, t)$ in (19) may be directly obtained. Let $\Lambda^*(s, t) = \int_0^t J(s, w)\Lambda(dw)$. Then, since $\mathbf{E}\{\hat{\Lambda}(s, t) - \Lambda^*(s, t)\} = 0$, $\text{Bias}\{\hat{\Lambda}(s, t)\} = \mathbf{E}\{\Lambda^*(s, t) - \Lambda(t)\} = -\int_0^t [\mathbf{P}\{Y_1(s, w) = 0\}]^n \Lambda(dw)$. Evidently $\text{Bias}\{\hat{\Lambda}(s, t)\} \leq 0$. Furthermore, since $\pi(s, t) = \mathbf{P}\{Y_1(s, t) = 0\}$ is increasing in t for fixed s , $\text{Bias}\{\hat{\Lambda}(s, t)\} \geq -[\mathbf{P}\{Y_1(s, t) = 0\}]^n \Lambda(t)$. The bias of (19) as an estimator for $\Lambda(\cdot)$ therefore satisfies the familiar inequality

$$-[\pi(s, t)]^n \Lambda(t) \leq \text{Bias}\{\hat{\Lambda}(s, t)\} \leq 0,$$

and converges to zero exponentially fast for (s, t) such that $\pi(s, t) > 0$.

For the estimator in (20), a representation analogous to that of the usual product-limit estimator (e.g., see Chapter 3 of Fleming and Harrington, 1991) can be established. In particular,

$$\frac{\hat{\bar{F}}(s, t) - \bar{F}(t)}{\bar{F}(t)} = -\int_0^t \frac{\hat{\bar{F}}(s, w-)}{\bar{F}(w)} \frac{J(s, w)}{Y(s, w)} M(s, dw) + \int_0^t \frac{\hat{\bar{F}}(s, w-)}{\bar{F}(w)} I\{Y(s, w) = 0\} \Lambda(dw). \quad (28)$$

Using Proposition 1,

$$\int_0^t \frac{\hat{\bar{F}}(s, w-)}{\bar{F}(w)} \frac{J(s, w)}{Y(s, w)} M(s, dw) = \sum_{i=1}^n \int_0^s \frac{\hat{\bar{F}}(s, R_i(v)-)}{\bar{F}(R_i(v))} \frac{J(s, R_i(v))}{Y(s, R_i(v))} dM_i(dv, t).$$

Since the integrand is a bounded and left-continuous (hence predictable) process, the process on the right-hand side is, for fixed t , a zero-mean martingale in s . Consequently,

$$\text{Bias}\{\hat{\bar{F}}(s, t)\} = \mathbf{E}\{\hat{\bar{F}}(s, t) - \bar{F}(t)\} = \bar{F}(t) \mathbf{E} \left\{ \int_0^t \frac{\hat{\bar{F}}(s, w-)}{\bar{F}(w)} I\{Y(s, w) = 0\} \Lambda(dw) \right\}.$$

The same arguments used to establish bounds for the bias of the usual product-limit estimator (e.g., see Fleming and Harrington, 1991, pp. 98-99) now lead to

$$0 \leq \text{Bias}\{\hat{\bar{F}}(s, t)\} \leq [1 - \bar{F}(t)] \mathbf{P}\{Y(s, t) = 0\} = [1 - \bar{F}(t)] [\pi(s, t)]^n.$$

The estimator $\hat{\bar{F}}$ is therefore positively biased; however, this bias converges to zero exponentially fast for (s, t) such that $\pi(s, t) > 0$.

We also seek approximations to the finite-sample variances of $\hat{\Lambda}(s, t)$ and $\hat{F}(s, t)$ for fixed $s \in (0, \infty)$ and $t \in [0, t^*]$, where $t^* \in (0, \infty)$ is such that $\Lambda(t^*) < \infty$ and $\pi(s, t^*) > 0$. Define

$$\sigma_{\text{NA}}^2(s, t) = \mathbf{E} \left\{ [\hat{\Lambda}(s, t) - \Lambda^*(s, t)]^2 \right\} = \mathbf{E} \left\{ \left(\int_0^t \frac{J(s, w)}{Y(s, w)} M(s, dw) \right)^2 \right\}.$$

Since the bias of $\hat{\Lambda}(s, t)$ vanishes exponentially fast, it follows by direct analogy to the usual survival analysis setting (e.g., see Fleming and Harrington, 1991, pp. 92-94) that $\mathbf{Var}\{\hat{\Lambda}(s, t)\} \approx \sigma_{\text{NA}}^2(s, t)$. Employing Proposition 1 and known results for martingale process,

$$\begin{aligned} \sigma_{\text{NA}}^2(s, t) &= \mathbf{E} \left\{ \left(\sum_{i=1}^n \int_0^s \frac{J(s, R_i(v))}{Y(s, R_i(v))} M_i(dv, t) \right)^2 \right\} \\ &= \mathbf{E} \left\{ \left\langle \sum_{i=1}^n \int_0^s \frac{J(s, R_i(v))}{Y(s, R_i(v))} M_i(dv, t) \right\rangle (s) \right\} = \mathbf{E} \left\{ \int_0^t \frac{J(s, w)}{Y(s, w)} \Lambda(dw) \right\}. \end{aligned}$$

As in the case of the Nelson-Aalen estimator for right-censored data (e.g., see Fleming and Harrington, 1991, p. 94), one can estimate this variance using

$$\hat{\sigma}_{\text{NA}}^2(s, t) = \int_0^t \frac{J(s, w)}{Y(s, w)^2} \left[\frac{Y(s, w) - N(s, \Delta w)}{Y(s, w) - 1} \right] N(s, dw). \quad (29)$$

Using the representation in (28), similar arguments to Fleming and Harrington (1991, p. 104) also show that one may estimate the variance of $\hat{F}(s, t)$ by

$$\hat{\sigma}_{\text{PL}}^2(s, t) = \hat{F}(s, t)^2 \int_0^t \frac{N(s, dw)}{Y(s, w)(Y(s, w) - N(s, \Delta w))}. \quad (30)$$

These estimators are identical in form to the estimators for the usual Nelson-Aalen and product-limit estimators for right-censored data; however, it is important recognize that the at-risk processes are necessarily more complex. For example, the computational form of $\hat{\sigma}_{\text{NA}}^2(s, t)$ when there are no tied interoccurrence times is

$$\hat{\sigma}_{\text{NA}}^2(s, t) = \sum_{i=1}^n \sum_{j=1}^{N_i^\dagger(s^-)} \frac{I\{T_{ij} \leq t\}}{Y(s, t_{ij})^2}, \quad (31)$$

where $Y(s, T_{ij}) = \sum_{k=1}^n \left\{ \sum_{k=1}^{N_k^\dagger(s^-)} I\{T_{ik} \geq T_{ij}\} + I\{(s \wedge \tau_i) - S_{i, N_i^\dagger(s^-)} \geq T_{ij}\} \right\}$.

3.4 Uniform Consistency and Weak Convergence

Next we consider limiting properties of our estimators, specifically consistency and weak convergence. The following theorem demonstrates uniform consistency of $\hat{\Lambda}(s, t)$ and $\hat{F}(s, t)$, and is proved in the Appendix.

Theorem 2 Fix $s \in [0, \infty]$, and let $y(s, t) = \bar{F}(t)\bar{G}_s(t-)[1 + \mathbf{E}\{\rho((\tau \wedge s) - t)|(\tau \wedge s) \geq t\}]$ be the function in Proposition 3. If $t^* \in (0, \infty)$ is such that $y(s, t^*) > 0$ and $\Lambda(t^*) < \infty$, then, as $n \rightarrow \infty$, $\sup_{0 \leq t \leq t^*} |\hat{\Lambda}(s, t) - \Lambda(t)| \xrightarrow{\text{pr}} 0$ and $\sup_{0 \leq t \leq t^*} |\hat{\bar{F}}(s, t) - \bar{F}(t)| \xrightarrow{\text{pr}} 0$.

As an aside, we point out that the usual technique of employing Lengart's inequality to establish uniform consistency for the Nelson-Aalen and product-limit estimators in standard right-censored models is not possible here because the processes of interest with respect to the duration time t are not submartingales. The next theorem presents weak convergence results for the estimators; it is also proved in the appendix.

Theorem 3 Under the same conditions as Theorem 2, as $n \rightarrow \infty$,

(i) the process $\{V(s, t) = \sqrt{n}[\hat{\Lambda}(s, t) - \Lambda(t)] : t \in [0, t^*]\}$ converges weakly on $\mathcal{D}[0, t^*]$ to a zero-mean Gaussian process $\{V^\infty(s, t) : t \in [0, t^*]\}$ whose covariance function is

$$\mathbf{Cov}\{V^\infty(s, t_1), V^\infty(s, t_2)\} = d(s, \min(t_1, t_2)) \quad \text{for } t_1, t_2 \in [0, t^*],$$

where $d(s, t) = \int_0^t \frac{\Lambda(dw)}{y(s, w)}$; and

(ii) the process $\{W(s, t) = \sqrt{n}[\hat{\bar{F}}(s, t) - \bar{F}(t)] : t \in [0, t^*]\}$ converges weakly on $\mathcal{D}[0, t^*]$ to a zero-mean Gaussian process $\{W^\infty(s, t) : t \in [0, t^*]\}$ with covariance function

$$\mathbf{Cov}\{W^\infty(s, t_1), W^\infty(s, t_2)\} = \bar{F}(t_1)\bar{F}(t_2)d(s, \min(t_1, t_2)).$$

For purposes of constructing confidence intervals or confidence bands for Λ or \bar{F} , as well as for testing hypotheses concerning these functions, possible estimators of the asymptotic variances of $\hat{\Lambda}(s, t)$ and $\hat{\bar{F}}(s, t)$ are those presented in (29) and (30), respectively. Confidence bands for Λ or \bar{F} may then be constructed using standard techniques such as in Hall and Wellner (1980); see also Nair (1984) and Hollander and Peña (1989).

To assess the quality of the normal approximation suggested by our weak convergence result for $\hat{\bar{F}}(t) = \hat{\bar{F}}(\infty, t)$, we performed a simulation study under the independence interoccurrence times model. The true parameters for simulation model were $F = \text{EXP}(\theta)$ and $G = \text{EXP}(\eta)$, with $\theta = 6$ and $\eta = 1$. On the basis of 1000 replications, the means, standard errors, and sampling distribution of $\sqrt{n}[\hat{\bar{F}}(t) - \bar{F}(t)]$ for $t \in \{.10, .20, .30, .40\}$ and $n \in \{20, 50, 80\}$ were obtained. Table 1 provides a summary of the results of this simulation. In this table we have also indicated the theoretical asymptotic standard errors under this specific simulation model. Using the results of Theorem 3(ii), this asymptotic standard error is simply

$$\left\{ \frac{\theta\eta}{(\theta + \eta)^2} \exp\{-(\theta - \eta)t\} [1 - \exp\{-(\theta + \eta)t\}] \right\}^{\frac{1}{2}}.$$

Looking at the means, we note a slight positive bias, with this bias decreasing as n increases. Notice also that the simulated standard errors are quite close to the theoretical values except

when t is large and n is small. Examining the simulated sampling distributions, which are presented as boxplots in Figure 1, the normal approximation is satisfactory for $n \geq 20$, as well as for n small but with t not too large. For this exponentially distributed T_{ij} 's and τ_i 's and for the chosen simulation parameter values, these simulation results provide additional empirical support to the theoretical results.

4 Consequences of Ignoring Some Observations

A common approach to dealing with recurrent event data in the medical literature has been to restrict attention to the first, possibly right-censored observation of each renewal process, or to simply delete the last censored observation and base the analysis only on the complete observations (Aalen and Husebye, 1991). A bit of reflection shows that the former analysis should be inefficient and that the latter analysis is biased towards shorter interoccurrence times. Below we respectively characterize the loss in efficiency and the form of the bias. For simplicity, we consider only the case where $s \rightarrow \infty$; analogous results are easily shown to hold for any $s < \infty$.

Consider first the problem of estimating \bar{F} nonparametrically using only the possibly right-censored times of first event occurrences. In this case, we are in a standard right-censored data setting, and the asymptotic variance of the cumulative hazard process (i.e., analogous to $d(\infty, t)$ of Theorem 3) is simply

$$d^\dagger(t) = \int_0^t \frac{\Lambda(dw)}{\bar{F}(w)\bar{G}(w-)}$$

(e.g., see Andersen *et al.*, 1993, Thm. IV.1.2). The effect of taking into account the other event interoccurrence times for the units is thus contained in the function $1 + \mathbf{E}\{\rho(\tau - t) | \tau \geq t\}$ which appears in the denominator of the integrand in $d(t)$. Since this function is bounded below by 1, its effect is to decrease the asymptotic variance of the Nelson-Aalen and product-limit estimators relative to estimators using only the first observations for each unit. Indeed, the asymptotic relative efficiency of $\hat{\bar{F}}(t)$ to $\tilde{\bar{F}}(t)$ at time t is

$$\text{ARE} = \left\{ \int_0^t \frac{\Lambda(dw)}{\bar{F}(w)\bar{G}(w-)} \right\} \left\{ \int_0^t \frac{\Lambda(dw)}{\bar{F}(w)\bar{G}(w-)[1 + \mathbf{E}\{\rho(\tau - w) | \tau \geq w\}]} \right\}^{-1}.$$

Define the function $m : \mathfrak{R}_+ \rightarrow \mathfrak{R}_+$ via

$$m(t; F, G) = \inf_{w \leq t} \mathbf{E}\{\rho(\tau - w) | \tau \geq w\} = \inf_{w \leq t} \left\{ \bar{G}(w-)^{-1} \int_{[w, \infty)} \rho(u - w) dG(u) \right\}.$$

Then, it is easy to see that $\text{ARE} \geq 1 + m(t; F, G)$, with equality if and only if $\mathbf{E}\{\rho(\tau - w) | \tau \geq w\}$ is constant in w . Suppose, for example, that the T_{ij} 's are distributed according to $F = \text{EXP}(\theta)$ and the τ_i 's are distributed according to $G = \text{EXP}(\eta)$. Then, $\mathbf{E}\{\rho(\tau - w) | \tau \geq w\} = \theta/\eta$ is constant in w , and consequently $\text{ARE} = 1 + \theta/\eta$. One may interpret this last result as

saying that the loss in efficiency becomes increasingly more severe as the expected length of the observation period becomes large relative to the expected interoccurrence times.

Let us now consider the effects of ignoring the right-censored observation for each unit. Let the “at-risk” process be given by $Y^*(t) = \sum_{i=1}^n Y_i^*(\infty, t) = \sum_{i=1}^n \sum_{j=1}^{N_i^\dagger(\tau_i^-)} I\{T_{ij} \geq t\}$, and define Nelson-Aalen and product-limit type estimators of $\Lambda(t)$ and $\bar{F}(t)$ as

$$\check{\Lambda}(t) = \int_0^t \frac{N(dw)}{Y^*(w)} \quad \text{and} \quad \check{\bar{F}}(t) = \prod_{w=0}^t [1 - \check{\Lambda}(dw)].$$

Define $y^*(t) = \mathbf{E}\{Y_1^*(\infty, t)\} = y(t) - \mathbf{P}\{\tau - S_{N^+(\tau^-)} \geq t\}$ and

$$B(t) = \int_0^t \left[\frac{y(w) - y^*(w)}{y^*(w)} \right] \Lambda(dw) = \int_0^t \left[\frac{\mathbf{P}\{\tau - S_{N^+(\tau_i^-)} \geq w\}}{y^*(w)} \right] \Lambda(dw), \quad (32)$$

and suppose t^* satisfies $y^*(t^*) > 0$. Then, $\sup_{0 \leq t \leq t^*} |\check{\Lambda}(t) - \Lambda(t) - B(t)| = o_p(1)$. Since the integrand in (32) is strictly positive, it follows that $\check{\Lambda}(t)$ is inconsistent for $\Lambda(t)$. Moreover, it is clear that analogous results hold for the estimator $\check{\bar{F}}(t)$. In particular, since $B(\cdot)$ is continuous, then by virtue of the product-integral representation, we have

$$\sup_{0 \leq t \leq t^*} |\check{\bar{F}}(t) - \bar{F}(t) \exp\{-B(t)\}| = o_p(1).$$

It follows that $\check{\bar{F}}(t)$ will be inconsistent for $\bar{F}(t)$, and in fact underestimates it for $t \leq t^*$.

5 Correlated Recurrence Time Data

The product limit estimators (20) and (22) are valid assuming that the interoccurrence times represent an i.i.d. sample from some underlying distribution function F . This assumption is clearly restrictive in biomedical applications, and one obvious generalization allowing association between interoccurrence times is a frailty model. Specifically, suppose that for a given unit or subject, there exists a random variable Z with distribution H_Z such that given $Z = z$, the event interoccurrence times $\{T_j, j \geq 1\}$ are i.i.d. with some distribution $F(\cdot|z)$. In the case where Z is not observed, the goal of the analysis is generally to estimate the marginal survivor function $\bar{F}(\cdot) = \int \bar{F}(\cdot|z) dH_Z(z)$.

The product-limit estimator, which was developed under the i.i.d. model, is expected to yield a biased estimator of the marginal interoccurrence time distribution in such a setting. As a referee has pointed out, a unit with many short recurrence times will contribute more terms to the likelihood function (i.e., see Section 3.2). Thus, the likelihood will tend to emphasize shorter recurrence times, causing problems in situations where this does not occur completely at random (i.e., outside the i.i.d. setting). Indeed, the problem is more serious, for in Section 5.1 we shall demonstrate that the product-limit estimator is inconsistent in the case where the within-unit recurrence times satisfy a multiplicative frailty model, and derive the asymptotic

bias of the product-limit estimator in the case where the within-unit recurrence times follow a mixed Poisson process. In Section 5.2 we provide an estimator of the marginal interoccurrence time distribution under the assumption that the frailty variable Z follows a gamma distribution. Finally, in Section 5.3 we briefly discuss an estimator recently proposed by Wang and Chang (1999) that is valid for both i.i.d. and correlated recurrence times. We compare the performance of the product-limit estimator to these other estimators under various degrees of correlation in Section 6.

5.1 Inconsistency of the PLE

A specific type of model that results in correlated within-unit interoccurrence times is a multiplicative frailty model; see, for instance, Vaupel (1990), Heckman and Singer (1984), Oakes (1989, 1991), Andersen, Borgan, Gill and Keiding (1993) and Murphy (1995). In this model it is postulated that there exists for each unit an unobservable positive-valued frailty Z_i such that, *conditionally on* $Z_i = z$, the interoccurrence times T_{i1}, T_{i2}, \dots are i.i.d. with common conditional survivor function

$$\bar{F}(t|Z_i = z) = [\bar{F}_0(t)]^z = \exp \left\{ -z \int_0^t \lambda_0(u) du \right\}, \quad (33)$$

where $\lambda_0(\cdot)$ is the hazard function associated with $\bar{F}_0(\cdot)$. The frailties Z_1, Z_2, \dots, Z_n are assumed to be i.i.d. from an unknown distribution function H_Z . Under this model, the marginal survivor function of T_{ij} is

$$\bar{F}(t) = \mathbf{E}\{\bar{F}_0^{Z_1}(t)\} = \psi[\Lambda_0(t)] \quad (34)$$

where $\psi(u) = \mathbf{E}\{\exp(-uZ_1)\}$ is the Laplace transform of Z_1 and $\Lambda_0(t) = -\log[\bar{F}_0(t)]$ is the cumulative hazard function of \bar{F}_0 . The marginal cumulative hazard function under the multiplicative frailty model can be written

$$\Lambda(t) = - \int_0^t \frac{\psi'(\Lambda_0(u))}{\psi(\Lambda_0(u))} \lambda_0(u) du = \int_0^t \frac{\mathbf{E}[Z_1 \bar{F}_0^{Z_1}(u)]}{\mathbf{E}[\bar{F}_0^{Z_1}(u)]} \lambda_0(u) du,$$

both representations following immediately from the definition of $\psi(\cdot)$.

In both non- and semi-parametric settings the survivor function \bar{F}_0 , hence \bar{F} , are left unspecified. The following result characterizes the asymptotic behavior of the product-limit estimator in the case where the recurrence times satisfy a multiplicative frailty model. For convenience we focus on the limiting behavior of the Nelson-Aalen estimator and consider the case where $s = \infty$ only.

Theorem 4 *Let $t \leq t^*$ where t^* satisfies $\mathbf{E}[n^{-1}Y(t^*)] > 0$ and $\Lambda(t^*) < \infty$. Then, under the frailty model (33), $\hat{\Lambda}(t) \xrightarrow{\text{pr}} \int_0^t \gamma(w) \lambda_0(w) dw$ as $n \rightarrow \infty$, where*

$$\gamma(t) = \frac{\mathbf{E} \left\{ Z_1 \bar{F}_0(t)^{Z_1} \left[1 + \bar{G}(t-)^{-1} \int_t^\infty \rho(w-t | \bar{F}_0^{Z_1}) dG(w) \right] \right\}}{\mathbf{E} \left\{ \bar{F}_0(t)^{Z_1} \left[1 + \bar{G}(t-)^{-1} \int_t^\infty \rho(w-t | \bar{F}_0^{Z_1}) dG(w) \right] \right\}}, \quad (35)$$

and $\rho(\cdot|\bar{F}_0^{Z_1})$ is the renewal function associated with the conditional survivor function $\bar{F}(\cdot|Z_1) = \bar{F}_0(\cdot)^{Z_1}$. Moreover, $\hat{\Lambda}(t)$ is consistent for $\Lambda(t)$ if and only if

$$\frac{\mathbf{E}[Z_1 \exp\{-Z_1 \Lambda_0(t)\} \Psi(t, Z_1)]}{\mathbf{E}[\exp\{-Z_1 \Lambda_0(t)\} \Psi(t, Z_1)]} = \frac{\mathbf{E}[Z_1 \exp\{-Z_1 \Lambda_0(t)\}]}{\mathbf{E}[\exp\{-Z_1 \Lambda_0(t)\}]}, \quad (36)$$

where $\Psi(t, Z) = 1 + \bar{G}^{-1}(t-) \int_t^\infty \rho(w - t|\bar{F}_0^Z) dG(w)$.

The proof of this theorem is presented in the appendix. Because the renewal function uniquely determines the interoccurrence time distribution, a necessary and sufficient condition for the recurrence times to be i.i.d. is that $\rho(w|\bar{F}_0^{Z_1})$ is constant as a function of Z_1 . This occurs if Z_1 has a degenerate distribution function, in which case (36) clearly holds. We conjecture, but have not yet been able to prove, that a necessary condition for (36) is that $\rho(w|\bar{F}_0^z)$ is a constant function of z . The following corollary establishes that the product-limit estimator is inconsistent when $\bar{F}_0(t) = \exp\{-\theta t\}$, or equivalently, when the within-unit recurrence times denote the waiting times of a mixed Poisson process and the renewal function $\rho(w|\bar{F}_0^z)$ is a linear function of z .

Corollary 1 *Suppose $\bar{F}_0(t) = \exp\{-\theta t\}$ for $\theta > 0$. Let $\psi(u) = \mathbf{E}\{\exp(-uZ_1)\}$ denote the Laplace transform of Z_1 , and suppose that Z_1 is not degenerate. Then, with $r_t = \bar{G}^{-1}(t-)\theta \int_t^\infty (w-t)dG(w)$ and $\gamma^*(t) = -\psi'(\theta t)/\psi(\theta t)$, $\gamma(t) = [\gamma^*(t) + r_t \frac{\psi''(\theta t)}{\psi(\theta t)}] / [1 + r_t \gamma^*(t)]$ for $t \leq t^*$. Moreover, $\hat{\Lambda}(t)$ is not a consistent estimator of $\Lambda(t)$ for $t \leq t^*$.*

An exact formula for the asymptotic bias of the product-limit estimator may be obtained by specifying the frailty distribution H and the censoring distribution G . Below we consider choices pertinent to our simulation study in Section 6. Corollary 2 will allow us to compute the asymptotic bias and compare it to the simulated bias under the different levels of association between within-unit interoccurrence times considered in Section 6.

Corollary 2 *Suppose $\bar{F}_0(t) = \exp\{-\theta t\}$ for $\theta > 0$, and suppose further that the frailty distribution is Gamma(α, α) so that the density function is $h(z) = \alpha^\alpha z^{\alpha-1} \exp\{-\alpha z\}/\Gamma(\alpha)$, and $\bar{G}(t) = \exp\{-\eta t\}$. Then the asymptotic bias of the PLE $\hat{F}(t)$ for t such that $y(t) > 0$ is*

$$\text{AsyBias}(t; \alpha, \theta, \eta) = - \left(\frac{\theta t}{\alpha + \theta t} \right) \left(\frac{\theta}{\theta + \eta} \right) \left(\frac{\alpha}{\alpha + \theta t} \right)^\alpha. \quad (37)$$

The bias formula (37) is informative. For example, as $\alpha \rightarrow \infty$ (i.e., the independence model), the limiting bias converges to zero, as expected. For $\alpha < \infty$, we see that $\hat{F}(t)$ underestimates the true marginal distribution, with the bias becoming more severe as the level of association increases (i.e., α decreases). Interestingly, as the mean period of observation per unit increases (i.e., η decreases, which also implies an increase in the observed number of recurrences per unit) the magnitude of the asymptotic bias increases.

5.2 Estimation under a Gamma Frailty Model

A common and convenient choice of frailty distribution is a gamma distribution with shape and scale parameters set equal to an unknown parameter α ; cf., Corollary 2. With this choice the common marginal survivor function (34) simplifies to

$$\bar{F}(t) = \left[\frac{\alpha}{\alpha + \Lambda_0(t)} \right]^\alpha. \quad (38)$$

As alluded to in the previous section, the parameter α controls the degree of association between interoccurrence times within a unit. Specifically, as α increases (decreases), association decreases (increases); letting $\alpha \rightarrow \infty$, we obtain a model with independent interoccurrence times in which the T_{ij} 's have a common survivor function of \bar{F}_0 . In fact, $Z_i \xrightarrow{PF} 1$ for $i = 1 \dots n$ as $\alpha \rightarrow \infty$.

We consider here the case for general $s \leq \infty$. When the gamma frailty model holds, then conditional on $Z_i = z_i$, the compensator of $N_i^\dagger(s)$ is $A_i^\dagger(s|z_i) = z_i \int_0^s Y_i^\dagger(v) \lambda_0[R_i(v)] dv$, $i = 1 \dots n$. Therefore, analogously to Section 3.2, the conditional likelihood over $[0, s]$ is

$$L(s|z_1, z_2, \dots, z_n) = \prod_{i=1}^n \left[\left\{ \prod_{v \in [0, s]} [z_i \lambda_0[R_i(v)]]^{N_i^\dagger(\Delta v)} \right\} \exp \left\{ -z_i \int_0^s Y_i^\dagger(v) \lambda_0[R_i(v)] dv \right\} \right].$$

The conditional NPMLE of $\Lambda_0(t)$ based on the preceding likelihood is easily shown to be

$$\hat{\Lambda}_0(s, t|z_1, z_2, \dots, z_n) = \sum_{i=1}^n \int_0^t \frac{N_i(s, dw)}{z_i Y_i(s, w)}.$$

Since Z_1, Z_2, \dots, Z_n are unobservable, neither $L(s|z_1, z_2, \dots, z_n)$ nor $\hat{\Lambda}_0(s, t|z_1, z_2, \dots, z_n)$ can be computed. Multiplying $L(s|z_1, z_2, \dots, z_n)$ by the joint density of (Z_1, Z_2, \dots, Z_n) yields

$$L(s, z_1, z_2, \dots, z_n) = \prod_{i=1}^n \left[\frac{\alpha^\alpha}{\Gamma(\alpha)} z_i^{\alpha + N_i^\dagger(s) - 1} \times \exp \left\{ -z_i \left(\alpha + \int_0^s Y_i^\dagger(v) \lambda_0(R_i(v)) dv \right) \right\} \left(\prod_{v \leq s} Y_i^\dagger(v) \lambda_0(R_i(v)) \right)^{N_i^\dagger(\Delta v)} \right].$$

Therefore, given α , $\Lambda_0(\cdot)$, and $\{(N_i^\dagger(v), Y_i^\dagger(v)) : v \in [0, s], i = 1, 2, \dots, n\}$, it follows that Z_1, Z_2, \dots, Z_n are independent, with Z_i having a gamma distribution with shape parameter $\alpha + N_i^\dagger(s)$ and scale parameter $\alpha + \int_0^s Y_i^\dagger(v) d\Lambda_0[R_i(v)]$. Hence

$$\mathbf{E} \left\{ Z_i | \alpha, \Lambda_0(\cdot), \{(N_i^\dagger(v), Y_i^\dagger(v)) : v \in [0, s], i = 1, 2, \dots, n\} \right\} = \frac{\alpha + N_i^\dagger(s)}{\alpha + \int_0^s Y_i^\dagger(v) d\Lambda_0[R_i(v)]}.$$

Integrating out z_1, z_2, \dots, z_n in $L(s, z_1, z_2, \dots, z_n)$, we obtain the likelihood function

$$L_F(s) = \prod_{i=1}^n \frac{\Gamma(\alpha + N_i^\dagger(s))}{\Gamma(\alpha)} \left[\frac{\alpha}{\alpha + \int_0^s Y_i^\dagger(w) \lambda_0(R_i(w)) dw} \right]^{\alpha + N_i^\dagger(s)} \left(\prod_{v \leq s} \left[\frac{Y_i^\dagger(v) \lambda_0(R_i(v))}{\alpha} \right]^{N_i^\dagger(\Delta v)} \right).$$

Given $\Lambda_0(\cdot)$, this likelihood can be maximized to obtain an estimate of α . However, joint maximization of this likelihood with respect to α and $\Lambda_0(\cdot)$ is more difficult. Fortunately, the properties demonstrated above, together with ideas in Nielsen, Gill, Andersen and Sorensen (1992), suggest using the EM algorithm (Dempster, Laird and Rubin, 1977) for obtaining maximum likelihood estimates of α and $\Lambda_0(\cdot)$. We describe one such implementation below.

Given the data in (2), denote the distinct ordered observed interoccurrence times by $t_{(1)} < t_{(2)} < \dots < t_{(Q)}$, and the number of observed failures at $t_{(l)}$ until calendar time s for unit i by

$$d_{(l)i}(s) = N_i(s, \Delta t_{(l)}) \equiv N_i(s, t_{(l)}) - N_i(s, t_{(l)}^-).$$

Furthermore, denote the estimates of the baseline hazard probabilities at $t_{(l)}$ by $\hat{\lambda}_{(l)}(s)$, $l = 1, 2, \dots, Q$, and the corresponding cumulative hazard estimate

$$\hat{\Lambda}_0(s, t) = \sum_{\{l: t_{(l)} \leq t\}} \hat{\lambda}_{(l)}(s). \quad (39)$$

The **E-Step** of the algorithm proceeds as follows: Given estimates $\hat{\alpha}$ and $\hat{\lambda}_{(l)}(s)$, $l = 1, 2, \dots, Q$, obtain estimates of the unobserved frailties via

$$\hat{Z}_i = \frac{\hat{\alpha} + N_i^\dagger(s)}{\hat{\alpha} + \sum_{l=1}^Q \hat{\lambda}_{(l)}(s) Y_i(s, t_{(l)})}, \quad i = 1, 2, \dots, n.$$

The **M-Step** of the algorithm requires two steps. First, given \hat{Z}_i , $i = 1, 2, \dots, n$, obtain the updated estimates

$$\hat{\lambda}_{(l)}(s) = \frac{\sum_{i=1}^n d_{(l)i}(s)}{\sum_{i=1}^n \hat{Z}_i Y_i(s, t_{(l)})}, \quad l = 1, 2, \dots, Q.$$

Then, obtain an updated estimate $\hat{\alpha}$ of α by maximizing the log-profile likelihood function

$$l(s; \xi) = n \frac{\xi \log \xi}{1 - \xi} + \sum_{i=1}^n \left[\sum_{j=1}^{N_i^\dagger(s)} \log \{ \xi + (1 - \xi)(j - 1) \} \right] - \left(N_i^\dagger(s) + \frac{\xi}{1 - \xi} \right) \log \left\{ \xi + (1 - \xi) \left(\sum_{l=1}^Q \hat{\lambda}_{(l)}(s) Y_i(s, t_{(l)}) \right) \right\} + C, \quad (40)$$

with respect to $\xi = \alpha / (1 + \alpha)$, where C is a constant independent of ξ (i.e., α).

This alternating **E-step** and **M-step** iterative process is terminated when the values of $\hat{\alpha}$, $\hat{\lambda}_{(l)}(s)$, $l = 1, 2, \dots, Q$, and \hat{Z}_i , $i = 1, 2, \dots, n$, have stabilized. Upon convergence, the common marginal survivor function of the interoccurrence times is estimated via

$$\tilde{F}(s, t) = \left[\frac{\hat{\alpha}}{\hat{\alpha} + \hat{\Lambda}_0(s, t)} \right]^{\hat{\alpha}}, \quad (41)$$

where $\hat{\Lambda}_0(s, t)$ is defined in (39). One may view this estimator as directly generalizing $\hat{F}(s, t)$ to the gamma frailty model. The i.i.d. model is achieved by letting $\alpha \rightarrow \infty$, which forces the

Z_i 's to converge in probability to unity. Thus, \hat{F} and \tilde{F} are expected to be close under the i.i.d. model. However, these two estimators need not coincide under the i.i.d. model due to the need to estimate α , whose estimate may be finite in a given sample. The fact that α must be estimated indicates, at least intuitively, that there will be a consequent loss in efficiency when \tilde{F} is applied to data generated under an i.i.d. model. Under the i.i.d. model and as the sample size n increases, we expect that $\hat{\alpha}$ will converge in probability to $\alpha = \infty$ and $\hat{F} - \tilde{F}$ will converge in probability to zero. Nevertheless, further work is needed to understand both the finite sample and asymptotic behavior of $\tilde{F}(s, t)$. The asymptotic behavior of this estimator can potentially be characterized by extending the results of Murphy (1995); see also Parner (1998). For now, we content ourselves by investigating its finite sample behavior through simulation in Section 6, where we also compare it to \hat{F} and another estimator to be described in the next section.

5.3 Estimator of Wang and Chang (1999)

Wang and Chang (1999) proposed an estimator of the common marginal survivor function in the case where within-unit interoccurrence times are correlated. The correlation structure considered by Wang and Chang (1999) is quite general and contains, in particular, both the i.i.d. and multiplicative (hence gamma) frailty models as a special case. Their estimator, taking the weights (a_i 's, in their notation) to all be unity, is given by

$$\hat{S}(t) = \prod_{\{T_k \in \mathcal{T}_d(\infty); T_k \leq t\}} \left[1 - \frac{d^*(T_k)}{R^*(T_k)} \right] \quad (42)$$

where, with $K_i^* = I\{K_i = 0\} + K_i I\{K_i > 0\}$, $d^*(t) = \sum_{i=1}^n \left\{ [K_i^*]^{-1} \sum_{j=1}^{K_i} I\{T_{ij} = t\} \right\}$; $R^*(t) = \sum_{i=1}^n [K_i^*]^{-1} \left\{ \sum_{j=1}^{K_i} I\{T_{ij} \geq t\} + I\{\tau_i - S_{iK_i} \geq t\} I\{K_i = 0\} \right\}$, and $\mathcal{T}_d(s) =$ distinct elements of $\mathcal{T}(s)$ with $\mathcal{T}(s) = \left\{ T_{ij} : j = 1, 2, \dots, N_i^\dagger(s \wedge \tau_i); i = 1, 2, \dots, n \right\}$. This estimator removes the bias noted for the product-limit estimator when recurrence times are correlated within units. However, when applied to i.i.d. interoccurrence times, the estimator $\hat{S}(t)$ is not expected to perform as well as $\tilde{F}(t)$, especially with regard to efficiency.

6 Simulation Results

Simulation studies were performed to examine the small to moderate sample size properties of the estimator $\hat{F}(t) = \hat{F}(\infty, t)$ under the renewal model for which it was developed. In particular, biases and root-mean-square-errors (RMSEs) for this estimator, as well as the Wang-Chang estimator $\hat{S}(t)$ and the gamma frailty model NPMLE $\tilde{F}(t) = \tilde{F}(\infty, t)$, were obtained under the independent interoccurrence times model. In addition, biases and RMSEs of all three estimators were also obtained under a gamma frailty model with $\bar{F}_0(t|\theta) = \exp\{-\theta t\}$ and $\bar{G}(t|\eta) = \exp\{-\eta t\}$.

The simulations were carried out in a combination of S-Plus and Fortran. Simulated biases and standard errors of $\hat{F}(t)$ (labelled IIDPLE), $\hat{S}(t)$ (labelled WCPLE), and $\tilde{\tilde{F}}(t)$ (labelled FRMLE) were obtained for equally-spaced values $t_l, l = 1, 2, \dots, 40$ of duration time t over the interval $[0, 2]$ based on $m = 1000$ replications. In order to use our EM algorithm for computing $\tilde{\tilde{F}}(t)$ there was a need to provide initial estimates for $\hat{Z}_i, 1 \dots n$. To automate this process in the simulation runs, we simply set $\hat{Z}_i = Z_i, i = \dots n$ in the initial E-step, and then proceeded with the algorithm as described earlier. The estimate of α , obtained by maximizing the profile likelihood (40) with respect to $\xi = \alpha/(1 + \alpha)$, is obtained using a combination of golden section search and successive parabolic interpolation (Brent, 1973). This procedure provided a fast and robust method of determining $\hat{\alpha}$. In Table 2, we present summary statistics pertaining to the estimation of $\xi = \alpha/(1 + \alpha)$ in our simulation runs.

The simulation parameters were combinations of $n \in \{20, 50, 80\}$, $\theta \in \{3, 6\}$, $\eta = 1$, and $\alpha \in \{\infty, 6, 2\}$. The value of $\alpha = \infty$ corresponds to the independent interoccurrence times model. Results obtained are reported for $\theta = 6$; results for $\theta = 3$ show similar patterns. Table 2 reports summary data for the estimated values of α obtained for the different sample sizes and true α values. Figure 2 compares the performance of the three estimators for $n = 50$. Figure 3 depicts the effect of the degree of association on the performance of the estimators. Figure 4, 6, and 7 respectively illustrate the effect of sample size on the performance of $\hat{F}(t)$, $\hat{S}(t)$ and $\tilde{\tilde{F}}(t)$. Finally, Figure 5 compares the simulated bias of the PLE and the asymptotic bias expression given in Corollary 2.

Comparison of Estimators: By examining Figure 2 and Figure 3, it is evident that the Wang-Chang estimator \hat{S} has negligible bias for the three values of α considered. The estimator $\tilde{\tilde{F}}$, which is also theoretically appropriate in each of these cases, is seen to have a slight bias in the right tail. This tail bias is likely due to the fact (41) is almost surely positive in the presence of censoring; for example, if we let $\hat{\alpha} \rightarrow \infty$ the resulting estimator converges to $\exp\{-\hat{\Lambda}(t)\}$, which can never equal zero. Under the independence model ($\alpha = \infty$), the estimator \hat{F} also has negligible bias; in fact, the simulated bias is almost always nonnegative, consistent with the theoretical results. However, when association is introduced in the model (i.e., for $\alpha = 6$ and $\alpha = 2$), the bias of \hat{F} becomes severe, increasing as the association among the within-unit interoccurrence times becomes stronger. Moreover, the simulated bias functions of \hat{F} for $\alpha = 6$ and $\alpha = 2$ are now both negative, which is expected in light of Corollary 2.

When we examine the RMSE functions in Figure 2, we immediately notice that \hat{F} dominates both $\tilde{\tilde{F}}$ and \hat{S} under the independence model ($\alpha = \infty$). The latter result is not surprising since \hat{S} is not expected to be optimal under the independence model, a fact we noted earlier. The fact that \hat{F} dominates $\tilde{\tilde{F}}$ is, similarly to before, a likely result of the positivity of \hat{F} . Under the gamma frailty model, we see that $\tilde{\tilde{F}}$ outperforms \hat{S} , leading to a smaller RMSE, hence trading bias for efficiency. In contrast, \hat{F} performs very poorly compared to either \hat{S} or $\tilde{\tilde{F}}$, especially for shorter duration times and strong levels of association. Indeed, the inconsistency of the estimator \hat{F} under the frailty model renders it an unacceptable estimator for this case.

Evidently, by examining Figure 3, the performance of all of these estimators degrades as the level of association among the within-unit interoccurrence times increases.

Effect of Sample Size: Figure 4 presents plots of the simulated bias and RMSE functions as the sample size n changes for the estimator \hat{F} . We immediately notice the detrimental impact of strong association between the within-unit interoccurrence times. For $\alpha = 2$ or 6 , there is negligible improvement in RMSE when n is increased, with the performance of \hat{F} degrading as the association increases. This occurs because the RMSE is dominated by the bias component, which is negligibly affected by the change in sample size and an obvious indicator of its inconsistency under the gamma frailty model (see Corollary 1). In Figure 5, we overlaid the plots of the theoretical limiting bias function (37) and the simulated bias functions of \hat{F} for $n \in \{20, 50, 80\}$ and $\alpha \in \{2, 6\}$ for $\theta = 6$ and $\eta = 1$. The close agreement between the theoretical and empirical functions for large n shows that the bias observed here is real.

REMARK: The product-limit estimator is not guaranteed to be inconsistent in the presence of any dependence. For example, if the correlation of within-unit interoccurrence times T_{ij_1} and T_{ij_2} approaches zero as $|j_1 - j_2|$ increases (e.g., a weakly-mixing model), it is possible for consistency to be achieved; see Cai and Roussas (1998) for related results.

Figures 6 and 7 respectively depict the simulated bias and RMSE functions for \hat{S} and \tilde{F} as the sample size changes. From these plots, one can discern that increasing the sample size improves the performance of both estimators. These results seem to indicate that the bias of the MLE under the frailty model diminishes as the sample size increases, indicating the promising possibility that this estimator is consistent when the interoccurrence times obey the gamma frailty model.

7 Application: Small Bowel Motility

We apply the various estimators discussed in the preceding sections to data from a study concerning small bowel motility (Husebye, Skar, Aalen and Osnes, 1990). A description of the data, as well as the data itself, is presented in Aalen and Husebye (1991). The main object of their analysis is to estimate the mean length of the Migratory Motor Complex (MMC) period (i.e., the mean interoccurrence time), and two approaches are considered: the first based on a variance component model, and the second using an intensity-based formulation with a gamma frailty component and a parametrically specified hazard (i.e., Weibull). The renewal assumption (i.e., i.i.d. interoccurrence times) for each subject in this data set needs to be formally verified; however, Aalen and Husebye (1991, p. 1229) stated that the “consecutive MMC periods for each individual appear (to be) approximate renewal processes.” Statistical methods for checking this renewal assumption in the presence of recurrent data of the type considered here do not appear to be available. The various nonparametric estimators discussed here should prove useful in developing such validation procedures.

To serve as a concrete illustration, we computed the IIDPLE $\hat{\tilde{F}}(t)$, the WCPLLE $\hat{S}(t)$, and the FRMLE $\tilde{\tilde{F}}(t)$ for these data. The resulting estimates of the interoccurrence time survivor function are presented in Figure 8. For aesthetic purposes, we have connected the points by lines instead of presenting as step-functions. The similarity of these survivor function estimates is striking. Based on the results of our simulation study, one may infer that such similarity would only be expected in the i.i.d. setting (exactly or approximately so). Thus, we regard the excellent agreement among the three estimates as providing support for Aalen and Husebye's assertion that the renewal assumption is valid. The estimate of the frailty parameter α used for computing $\tilde{\tilde{F}}(t)$ is $\hat{\alpha} = 10.18$, indicating weak association among the within-subject interoccurrence times; for comparison, the estimate obtained by Aalen and Husebye (1991, Table III) under the Weibull hazard assumption is $\tilde{\alpha} = 6.85 = 1/0.146$. These results provide additional support for this conjecture.

Using these estimates, the corresponding estimates of the mean MMC period length (in minutes) are $\mu[\hat{\tilde{F}}(\cdot)] = 104.1$; $\mu[\hat{S}(\cdot)] = 106.0$, and $\mu[\tilde{\tilde{F}}(\cdot)] = 105.5$. For comparison, the estimate of the mean MMC period length obtained under Aalen and Husebye's (1991) variance-component model is 106.8 minutes, while that from their Weibull-intensity parametric model is 107.7 minutes. For these data, we see that the point estimates for the mean MMC period length largely remain unaffected by the method of analysis.

References

- [1] O. Aalen and E. Husebye. Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine*, 10:1227–1240, 1991.
- [2] P. Andersen, O. Borgan, R. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, 1993.
- [3] P. Andersen and R. Gill. Cox’s regression model for counting processes: a large sample study. *Annals of Statistics*, 10:1100–1120, 1982.
- [4] R. Brent. *Algorithms for Minimization Without Derivatives*. Prentice-Hall, New York, 1973.
- [5] Z. Cai and G. Roussas. Kaplan-Meier estimator under association. *J. Multivariate Anal.*, 67(2):318–348, 1998.
- [6] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood estimation from incomplete data via the em algorithm (with discussion). *J. Roy. Statist. Soc. B*, 39:1–38, 1977.
- [7] T. Fleming and D. Harrington. *Counting Processes and Survival Analysis*. Wiley, New York, 1991.
- [8] R. Gill. Nonparametric estimation based on censored observations of a markov renewal process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 53:97–116, 1980.
- [9] R. Gill. Discussion of papers by Helland and Kurtz. *Bulletin of the International Statistical Institute*, 50:239–243, 1983.
- [10] R. D. Gill. Testing with replacement and the product-limit estimator. *The Annals of Statistics*, 9:853–860, 1981.
- [11] W. Hall and J. Wellner. Confidence bands for a survival curve from censored data. *Biometrika*, 67:133–143, 1980.
- [12] J. Heckman and B. Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52:271–320, 1984.
- [13] J. Hoffmann-Jorgensen. *Probability With a View Towards Statistics II*. Chapman Hall, New York, 1994.
- [14] M. Hollander and E. Pena. Families of confidence bands for the survival function under the general random censorship model and the Koziol-Green model. *Canadian Journal of Statistics*, 17:59–74, 1989.
- [15] J. Hu and J. Lawless. Estimation of rate and mean functions from truncated recurrent event data. *Journal of the American Statistical Association*, 91:300–310, 1996.

- [16] J. Jacod. Multivariate point processes: Predictable projection, radon-nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Geb.*, 31:235–253, 1975.
- [17] J. D. Kalbfleisch, J. F. Lawless, and J. A. Robinson. Methods for the analysis and prediction of warranty claims. *Technometrics*, 33(3):273–285, 1991.
- [18] N. Keiding, C. Andersen, and P. Fledelius. The Cox regression model for claims data in non-life insurance. *Astin Bulletin*, 28:95–118, 1998.
- [19] J. Lawless and C. Nadeau. Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168, 1995.
- [20] D. Lin, L. Wei, and Z. Ying. Accelerated failure time models for counting processes. *Biometrika*, 85(3):605–618, 1998.
- [21] S. McClean and C. Devine. A nonparametric maximum likelihood estimator for incomplete renewal data. *Biometrika*, 82(4):791–803, 1995.
- [22] S. Murphy. Asymptotic theory for the frailty model. *The Annals of Statistics*, 23(1):182–198, 1995.
- [23] V. Nair. Confidence bands for survival functions with censored data: A comparative study. *Technometrics*, 26:265–275, 1984.
- [24] G. Nielsen, R. Gill, P. Andersen, and T. Sorensen. A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.*, 19:25–43, 1992.
- [25] D. Oakes. Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.*, 84:487–493, 1989.
- [26] D. Oakes. *Frailty models for multiple event times*, chapter In *Survival Analysis: State of the Art*, pages 371–379. Kluwer, Dordrecht, (j. p. klein and p. k. goel, (eds.) edition, 1991.
- [27] E. Parner. Asymptotic theory for the correlated gamma frailty model. *The Annals of Statistics*, 26(1):183–214, 1998.
- [28] E. Peña, R. Strawderman, and M. Hollander. A weak convergence result relevant in recurrent models. *Recent Advances in Reliability Theory (Bordeaux 2000)*, (N. Limnios and M. Nikulin, eds.), pages 493–515, 2000.
- [29] R. Prentice, B. Williams, and A. Peterson. On the regression analysis of multivariate failure time data. *Biometrika*, 68:373–379, 1981.
- [30] R. Rebolledo. Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete*, 51:269–286, 1980.
- [31] S. Resnick. *Adventures in Stochastic Processes*. Birkhauser, Boston, 1994.

- [32] T. Sellke. Weak convergence of the aalen estimator for a censored renewal process. In *Statistical Decision Theory and Related Topics IV* (eds., S. Berger and J. Berger), 2:183–194, 1988.
- [33] G. Soon and M. Woodroffe. Nonparametric estimation and consistency for renewal processes. *Journal of Statistical Planning and Inference*, 53:171–195, 1996.
- [34] Y. Vardi. Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, 10:616–620, 1982a.
- [35] Y. Vardi. Nonparametric estimation in renewal processes. *The Annals of Statistics*, 10:772–785, 1982b.
- [36] J. Vaupel. *Kindred lifetimes: frailty models in population genetics*, volume In *Convergent Issues in Genetics and Demography*, pages 156–170. Oxford Univ. Press, (j. adams, d. a. lam, a. i. herman and p. e. smouse, eds.) edition, 1990.
- [37] M. C. Wang and S. H. Chang. Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association*, 94:146–153, 1999.

8 Appendix: Proofs and Technical Details

Proof of Proposition 2:

Under the assumptions of this paper it is automatic that $\{M(s, t), s \geq 0\}$ is a local square integrable martingale for every fixed $t \geq 0$. By Corollary 4 of Protter (1990, p. 67), if the limiting expected quadratic variation is finite (i.e., if $\lim_{s \rightarrow \infty} \mathbf{E}\{[M(\cdot, t), M(\cdot, t)](s)\} < \infty$), then $\{M(s, t), s \geq 0\}$ is a square integrable martingale for every fixed $t \geq 0$. Notice that

$$\lim_{s \rightarrow \infty} \mathbf{E}\{[M(\cdot, t), M(\cdot, t)](s)\} = \lim_{s \rightarrow \infty} \mathbf{E}\{A(s, t)\} = \lim_{s \rightarrow \infty} \mathbf{E}\{N(s, t)\} = \mathbf{E}\{N(\tau, t)\}.$$

Since $\mathbf{E}\{A(s, t)\} = \mathbf{E}\{(M(\cdot, t), M(\cdot, t))(s)\}$, the desired result holds provided $\mathbf{E}\{N(\tau, t)\} < \infty$. However, $E[N(\tau, t)] \leq nE[N_1^\dagger(\tau)] = nE[\sum_{j=1}^\infty I\{S_{1j} \leq \tau\}] = nE[\rho(\tau)]$, the last equality resulting from the definition of the renewal function and the independence of τ and $\{T_{11}, T_{12}, \dots\}$. Since $E[\rho(\tau)] < \infty$ by assumption, the result holds. \parallel

Proof of Proposition 3: We establish the result for $y(t)$ using a renewal argument; the result for $y(s, t)$ is easily obtained by replacing τ_1 with $\tau_1(s) = s \wedge \tau_1$ and $G(\cdot)$ by $G_s(\cdot)$. Let $H_t(v) = \mathbf{E}\{Y_1(t) | \tau_1 = v\}$. Integrating with respect to the realization of T_{11} , we have

$$H_t(v) = \int_0^v \mathbf{E}\{Y_1(t) | \tau_1 = v, T_{11} = w\} dF(w) + \int_v^\infty \mathbf{E}\{Y_1(t) | \tau_1 = v, T_{11} = w\} dF(w).$$

If $w > v$, then $K_1 = 0$, so $Y_1(t) = I\{v \geq t\}$, hence $\mathbf{E}\{Y_1(t) | \tau_1 = v, T_{11} = w\} = I\{v \geq t\}$. On the other hand, if $w \leq v$, then $\mathbf{E}\{Y_1(t) | \tau_1 = v, T_{11} = w\} = I\{w \geq t\} + H_t(v - w)$. Consequently,

$$H_t(v) = I\{v \geq t\} \bar{F}(v) + \int_0^v \{I\{w \geq t\} + H_t(v - w)\} dF(w)$$

from which it is easily seen that $H_t(\cdot)$ satisfies the renewal equation

$$H_t(v) = \bar{F}(t) I\{v \geq t\} + \int_0^v H_t(v - w) dF(w). \quad (43)$$

Since $\bar{F}(t) I\{v \geq t\}$ is a bounded function in v , then for $v \geq 0$ the unique locally bounded solution of (43) is (e.g. Resnick, 1994, Thm. 3.5.1)

$$H_t(v) = \bar{F}(t) I\{v \geq t\} + \int_0^v \bar{F}(t) I\{v - w \geq t\} d\rho(w) = \bar{F}(t) [I\{v \geq t\} + \rho(v - t)],$$

the last relation following from the fact that $\rho(0) = F(0) = 0$. Averaging out τ_1 , we obtain

$$y(t) = \int_0^\infty H_t(v) dG(v) = \bar{F}(t) \bar{G}(t-) + \bar{F}(t) \int_{[t, \infty)} \rho(v - t) dG(v).$$

The stated properties of $y(s, t)$ are easily obtained from the formula for $y(s, t)$ given in the statement of the result and the fact that $\lim_{s \rightarrow \infty} G_s(u) = G(u)$. That $y(0) < \infty$ follows from the fact that \bar{F} and \bar{G} are survivor functions (hence bounded above by 1) and the assumption that $E[\rho(\tau)] = \int_0^\infty \rho(w) dG(w) < \infty$. \parallel

Proof of Proposition 5:

Set $J_t(v) = \mathbf{E}\{I\{\tau_1 - S_{1K_1} \geq t\} | \tau_1 = v\}$. Integrating with respect to T_{11} , we obtain the renewal equation $J_t(v) = \int_0^v J_t(v-w) dF(w) + I\{v \geq t\} \bar{F}(v)$. The locally bounded unique solution to this renewal equation is (cf., Resnick (1994)) $J_t(v) = I\{v \geq t\} \bar{F}(v) + \int_0^{v-t} \bar{F}(v-w) d\rho(w)$. Averaging over τ_1 (i.e., integrating with respect to $G(v)$), we thus obtain

$$\begin{aligned} \mathbf{E}\{I\{\tau_1 - S_{1K_1} \geq t\}\} &= \mathbf{E}\{J_t(\tau_1)\} = \int_t^\infty \left\{ \bar{F}(v) + \int_0^{v-t} \bar{F}(v-w) d\rho(w) \right\} dG(v) \\ &= \bar{G}(t-) \mathbf{E} \left\{ \bar{F}(\tau_1) + \int_0^{\tau_1-t} \bar{F}(\tau_1-w) d\rho(w) | \tau_1 \geq t \right\}. \end{aligned}$$

The second result now follows directly from the fact that $Y_1^*(t) = Y_1(t) - I\{\tau_1 - S_{1K_1} \geq t\}$ and Proposition 3. \parallel

Proof of Theorem 2:

To prove the first result, it suffices to show that $\sup_{0 \leq t \leq t^*} |\hat{\Lambda}(s, t) - \Lambda^*(s, t)| \xrightarrow{\text{PR}} 0$; this follows because $|\hat{\Lambda}(s, t) - \Lambda(t)| \leq |\hat{\Lambda}(s, t) - \Lambda^*(s, t)| + |\Lambda^*(s, t) - \Lambda(t)|$ and $\sup_{0 \leq t \leq t^*} |\Lambda^*(s, t) - \Lambda(t)| \xrightarrow{\text{PR}} 0$ as $n \rightarrow \infty$. We have

$$|\hat{\Lambda}(s, t) - \Lambda^*(s, t)| \leq \left| \frac{1}{n} \int_0^t \left[\frac{J(s, w)}{[Y(s, w)/n]} - \frac{1}{y(s, w)} \right] M(s, dw) \right| + \left| \frac{1}{n} \int_0^t \frac{M(s, dw)}{y(s, w)} \right|. \quad (44)$$

Considering the first term of (44), we obtain

$$\begin{aligned} &\sup_{0 \leq t \leq t^*} \left| \frac{1}{n} \int_0^t \left[\frac{J(s, w)}{[Y(s, w)/n]} - \frac{1}{y(s, w)} \right] M(s, dw) \right| \\ &\leq \left\{ \sup_{0 \leq t \leq t^*} \left| \frac{J(s, t)}{[Y(s, t)/n]} - \frac{1}{y(s, t)} \right| \right\} \left\{ \frac{N(s, t^*)}{n} + \frac{1}{n} \int_0^{t^*} Y(s, w) \lambda(w) dw \right\}. \end{aligned}$$

Proposition 4 and the WLLN show that the right-hand side is $o_p(1)O_p(1)$; thus,

$$\sup_{0 \leq t \leq t^*} \left| \frac{1}{n} \int_0^t \left[\frac{J(s, w)}{[Y(s, w)/n]} - \frac{1}{y(s, w)} \right] M(s, dw) \right| = o_p(1).$$

By Theorem 1 of PSH (2000), the process $\left\{ \frac{1}{\sqrt{n}} \int_0^t \frac{1}{y(s, w)} M(s, dw) : t \in [0, t^*] \right\}$ converges weakly to a zero-mean Gaussian process $\{V^\infty(s, t) : t \in [0, t^*]\}$, so that for each $a > 0$ and as $n \rightarrow \infty$, and by applying the continuous mapping theorem,

$$\mathbf{P} \left\{ \sup_{0 \leq t \leq t^*} \left| \frac{1}{\sqrt{n}} \int_0^t \frac{1}{y(s, w)} M(s, dw) \right| > a \right\} \rightarrow \mathbf{P} \left\{ \sup_{0 \leq t \leq t^*} |V^\infty(s, t)| > a \right\}.$$

Fix an arbitrary $\epsilon > 0$. Choose a $B > 0$ such that $\mathbf{P}\{\sup_{0 \leq t \leq t^*} |V^\infty(s, t)| \geq B\} \leq \epsilon/3$. Also, let n_1 be such that whenever $n \geq n_1$,

$$\left| \mathbf{P} \left\{ \sup_{0 \leq t \leq t^*} \left| \frac{1}{\sqrt{n}} \int_0^t \frac{1}{y(s, w)} M(s, dw) \right| \geq B \right\} - \mathbf{P} \left\{ \sup_{0 \leq t \leq t^*} |V^\infty(s, t)| > B \right\} \right| \leq \frac{\epsilon}{3}.$$

Let $n_2 = \lceil B^2/\epsilon^2 \rceil + 1$. Then, for $n \geq \max\{n_1, n_2\}$, we have

$$\begin{aligned} \mathbf{P} \left\{ \sup_{0 \leq t \leq t^*} \left| \frac{1}{n} \int_0^t \frac{1}{y(s, w)} M(s, dw) \right| \geq \epsilon \right\} &\leq \mathbf{P} \left\{ \sup_{0 \leq t \leq t^*} |V^\infty(s, t)| \geq B \right\} + \\ &\left| \mathbf{P} \left\{ \sup_{0 \leq t \leq t^*} \left| \frac{1}{\sqrt{n}} \int_0^t \frac{1}{y(s, w)} M(s, dw) \right| \geq \sqrt{n}\epsilon \right\} - \mathbf{P} \left\{ \sup_{0 \leq t \leq t^*} |V^\infty(s, t)| \geq B \right\} \right| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Hence $\sup_{0 \leq t \leq t^*} \left| \frac{1}{n} \int_0^t \frac{1}{y(s, w)} M(s, dw) \right| = o_p(1)$, establishing the uniform consistency of $\hat{\Lambda}(s, t)$.

To prove uniform consistency of $\hat{F}(s, t)$, we first recall the product-integral representation of \hat{F} and \bar{F} , which are $\hat{F}(s, t) = \prod_{w=0}^t [1 - \hat{\Lambda}(s, dw)]$ and $\bar{F}(t) = \prod_{w=0}^t [1 - \Lambda(dw)]$. Using the Duhamel equation (Andersen *et al.*, 1993, Thm. II.6.2), we obtain

$$\frac{\hat{F}(s, t) - \bar{F}(t)}{\bar{F}(t)} = \int_0^t \frac{\hat{F}(s, w-)}{\bar{F}(w)} [\hat{\Lambda}(s, dw) - \Lambda(dw)], \quad t \in [0, t^*]. \quad (45)$$

Applying the integration-by-parts formula on the right-hand side of (45) and noting that $\hat{\Lambda}(s, 0) = \Lambda(0) = 0$, we obtain

$$\int_0^t \frac{\hat{F}(s, w-)}{\bar{F}(w)} [\hat{\Lambda}(s, dw) - \Lambda(dw)] = \frac{\hat{F}(s, t)}{\bar{F}(t)} [\hat{\Lambda}(s, t) - \Lambda(t)] - \int_0^t [\hat{\Lambda}(s, w) - \Lambda(w)] d_w \left\{ \frac{\hat{F}(s, w)}{\bar{F}(w)} \right\},$$

where d_w means that integration is being taken with respect to w . Consequently,

$$\sup_{0 \leq t \leq t^*} \left| \hat{F}(s, t) - \bar{F}(t) \right| \leq \sup_{0 \leq t \leq t^*} \left| \hat{\Lambda}(s, t) - \Lambda(t) \right| \left\{ 1 + \int_0^{t^*} \left| d_w \left\{ \frac{\hat{F}(s, w)}{\bar{F}(w)} \right\} \right| \right\}.$$

Now,

$$\int_0^{t^*} \left| d_w \left\{ \frac{\hat{F}(s, w)}{\bar{F}(w)} \right\} \right| \leq \int_0^{t^*} \left[\frac{1}{\bar{F}(w)} \hat{F}(s, dw) + \frac{\hat{F}(s, w)}{\bar{F}(w)^2} F(dw) \right] \leq \left\{ \frac{1}{\bar{F}(t^*)} + \frac{\Lambda(t^*)}{\bar{F}(t^*)} \right\},$$

which is finite since $\Lambda(t^*) < \infty$, and hence $\bar{F}(t^*) > 0$. Therefore, since $\sup_{0 \leq t \leq t^*} |\hat{\Lambda}(s, t) - \Lambda(t)| = o_p(1)$, it now follows that $\sup_{0 \leq t \leq t^*} |\hat{F}(s, t) - \bar{F}(t)| = o_p(1)$, completing the proof of the theorem.

||

Proof of Theorem 3:

Since

$$\sqrt{n} [\hat{\Lambda}(s, t) - \Lambda(t)] = \frac{1}{\sqrt{n}} \int_0^t \frac{J(s, w)}{[Y(s, w)/n]} M(s, dw) + \sqrt{n} \int_0^t [J(s, w) - 1] \Lambda(dw), \quad (46)$$

it suffices to show that the first term on the right-hand side converges weakly to $V^\infty(s, \cdot)$ and that $\sup_{0 \leq t \leq t^*} |\sqrt{n} \int_0^t [J(s, w) - 1] \Lambda(dw)| = o_p(1)$. We prove the latter first. Since for each $t \in [0, t^*]$, $\{Y(s, t) = 0\} \subseteq \{Y(s, t^*) = 0\}$ because $Y(s, t)$ is nonnegative and nonincreasing in t ,

we have $\sup_{0 \leq t \leq t^*} \left| \sqrt{n} \int_0^t [J(s, w) - 1] \Lambda(dw) \right| \leq \sqrt{n} \Lambda(t^*) I\{Y(s, t^*) = 0\}$. By Markov's Inequality, and using the facts that $\Lambda(t^*) < \infty$ and $\pi(s, t^*) < 1$ as implied by $y(s, t^*) > 0$, then for each $0 < \epsilon < 1$, $\mathbf{P} \{ \sqrt{n} \Lambda(t^*) I\{Y(s, t^*) = 0\} > \epsilon \} \leq \sqrt{n} \Lambda(t^*) [\pi(s, t^*)]^n / \epsilon \rightarrow 0$ as $n \rightarrow \infty$. Therefore, $\sup_{0 \leq t \leq t^*} \left| \sqrt{n} \int_0^t [J(s, w) - 1] \Lambda(dw) \right| = o_p(1)$.

Let us now consider the first term on the right-hand side of (46). For this purpose, let

$$H_i(s, t) = H(s, t) = \frac{J(s, t)}{Y(s, t)/n}, \quad i = 1, 2, \dots, n, \quad \text{and} \quad h(s, t) = \frac{1}{y(s, t)} I\{t \leq s\}.$$

Note that $H(s, t)$ is left-continuous in s and t and adapted, hence predictable, and is also bounded by unity. Also, since $y(s, t) \geq y(s, t^*)$ whenever $t \in [0, t^*]$ and $y(s, t^*) > 0$, then $h(s, \cdot)$ is a bounded function on $[0, t^*]$. Furthermore, since $\sup_{0 \leq t \leq t^*} \left| \frac{1}{n} Y(s, t) - y(s, t) \right| \xrightarrow{\text{a.s.}} 0$ by Proposition 4, $\sup_{0 \leq t \leq t^*} |H(s, t) - h(s, t)| \xrightarrow{\text{pr}} 0$. Also, we have that

$$\sigma^2(s, t) = \int_0^t h(s, w)^2 y(s, w) \lambda(w) dw = \int_0^t \frac{\lambda(w)}{y(s, w)} dw = d(s, t)$$

is finite for each $t \in [0, t^*]$ because $y(s, t^*) > 0$. Consequently, by Theorem 1 of PSH (2000), $\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t H(s, w) M_i(s, dw) : t \in [0, t^*] \right\}$ converges weakly on $\mathcal{D}[0, t^*]$ to the process $\{V^\infty(s, t) : t \in [0, t^*]\}$ stated in the statement of the theorem. This completes the proof of part (i) of the theorem.

To prove part (ii) we first note that $\sqrt{n} \left\{ \left[\hat{F}(s, t) - \bar{F}(t) \right] / \bar{F}(t) \right\}$ equals

$$-\sqrt{n} \int_0^t \frac{\hat{F}(s, w-)}{\bar{F}(w)} \frac{J(s, w)}{Y(s, w)} M(s, dw) + \sqrt{n} \int_0^t \frac{\hat{F}(s, w-)}{\bar{F}(w)} I\{Y(s, w) = 0\} \Lambda(dw). \quad (47)$$

Since $\sup_{0 \leq t \leq t^*} \left| \sqrt{n} \int_0^t \frac{\hat{F}(s, w-)}{\bar{F}(w)} I\{Y(s, w) = 0\} \Lambda(dw) \right| \leq \sqrt{n} \frac{\Lambda(t^*)}{\bar{F}(t^*)} I\{Y(s, t^*) = 0\}$, and because the upper bound converges in probability to zero by Markov's inequality, the second term on the right-hand side of (47) is $o_p(1)$. The first term on the right-hand side of (47) can be expressed as

$$\sqrt{n} \int_0^t \frac{\hat{F}(s, w-)}{\bar{F}(w)} \frac{J(s, w)}{Y(s, w)} M(s, dw) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t H(s, w) M_i(s, dw)$$

where now $H_i(s, t) = H(s, t) = \frac{\hat{F}(s, t-)}{\bar{F}(t)} \frac{nJ(s, t)}{Y(s, t)}$, $i = 1, 2, \dots, n$. Since $\hat{F}(s, t-)$ is left-continuous in both s and t , $H(s, t)$ is a bounded predictable process. If we can show that

$$\sup_{0 \leq t \leq t^*} \left| \frac{\hat{F}(s, t-)}{\bar{F}(t)} \frac{nJ(s, t)}{Y(s, t)} - \frac{1}{y(s, t)} \right| \xrightarrow{\text{pr}} 0, \quad (48)$$

then part (ii) of the theorem will follow from Theorem 1 of PSH (2000). We may write

$$\begin{aligned} \sup_{0 \leq t \leq t^*} \left| \frac{\hat{F}(s, t-)}{\bar{F}(t)} \frac{nJ(s, t)}{Y(s, t)} - \frac{1}{y(s, t)} \right| &\leq \frac{1}{\bar{F}(t^*)} \sup_{0 \leq t \leq t^*} \left| \frac{J(s, t)}{[Y(s, t)/n]} - \frac{1}{y(s, t)} \right| + \\ &\quad \frac{1}{y(s, t^*) \bar{F}(t^*)} \sup_{0 \leq t \leq t^*} \left| \hat{F}(s, t-) - \bar{F}(t) \right|. \end{aligned}$$

Proposition 4, Theorem 2, the assumed continuity of $\bar{F}(\cdot)$, and the fact that $\bar{F}(t^*) > 0$ and $y(s, t^*) > 0$ now imply that (48) holds. This completes the proof of the theorem. \parallel

Proof of Theorem 4:

Let $M_i(s, t|Z_i) = N_i(s, t) - \int_0^t Y_i(s, w) Z_i \lambda_0(w) dw$ for $i = 1, 2, \dots, n$. Then,

$$\hat{\Lambda}(s, t) = \sum_{i=1}^n \int_0^t \left[\frac{J(s, w)}{Y(s, w)} \right] M_i(s, dw|Z_i) + \int_0^t \left[\frac{\sum_{i=1}^n Z_i Y_i(s, w)}{Y(s, w)} \right] \lambda_0(w) dw. \quad (49)$$

Using Proposition 1, the first term on the RHS of (49) becomes

$$\frac{1}{n} \sum_{i=1}^n \int_0^s \left[\frac{J(s, R_i(v))}{Y(s, R_i(v))/n} \right] I\{R_i(v) \leq t\} M_i^\dagger(dv|Z_i), \quad (50)$$

where $M_i^\dagger(s|Z_i) = N_i^\dagger(s) - \int_0^s Y_i^\dagger(v) Z_i \lambda_0(R_i(v)) dv$, $i = 1, 2, \dots, n$. Conditional on Z_1, Z_2, \dots, Z_n , $(M_1^\dagger(s), M_2^\dagger(s), \dots, M_n^\dagger(s))$ are orthogonal square integrable zero-mean martingales with (conditional) predictable quadratic covariation processes

$$\langle M_{i_1}^\dagger(\cdot|Z_{i_1}), M_{i_2}^\dagger(\cdot|Z_{i_2}) \rangle(s) = I\{i_1 = i_2\} \int_0^s Y_{i_1}^\dagger(v) Z_{i_1} \lambda_0(R_{i_1}(v)) dv.$$

By stochastic integration theory, (50) is a zero-mean martingale under the frailty model. Hence, by the martingale central limit theorem,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^s \left[\frac{J(s, R_i(v))}{Y(s, R_i(v))/n} \right] I\{R_i(v) \leq t\} M_i^\dagger(dv|Z_i) = O_p(1).$$

Consequently the first term on the RHS of (49) is $o_p(1)$. The second term on the RHS of (49) may be written $\int_0^t \left[\frac{\frac{1}{n} \sum_{i=1}^n Z_i Y_i(s, w)}{\frac{1}{n} \sum_{i=1}^n Y_i(s, w)} \right] \lambda_0(w) dw$. Using Proposition 4 and analogous arguments, we have

$$\sup_{t \in [0, \infty)} \left| \frac{1}{n} \sum_{i=1}^n Y_i(s, t) - \mathbf{E}[Y_1(s, t)] \right| \xrightarrow{\text{pr}} 0 \quad \text{and} \quad \sup_{t \in [0, \infty)} \left| \frac{1}{n} \sum_{i=1}^n Z_i Y_i(s, t) - \mathbf{E}[Z_1 Y_1(s, t)] \right| \xrightarrow{\text{pr}} 0.$$

Consequently, $\int_0^t \left\{ \left[\frac{\sum_{i=1}^n Z_i Y_i(s, w)}{Y(s, w)} \right] \lambda_0(w) dw \right\} \xrightarrow{\text{pr}} \int_0^t \gamma(s, w) \lambda_0(w) dw$ as $n \rightarrow \infty$, where $\gamma(s, t) = \mathbf{E}[Z_1 Y_1(s, t)] / \mathbf{E}[Y_1(s, t)]$. Letting $s \rightarrow \infty$ establishes the first assertion in the theorem. To obtain the stated form of $\gamma(t)$, note that $\mathbf{E}\{Z_1 Y_1(t)\} = \mathbf{E}\{Z_1 \mathbf{E}[Y_1(t)|Z_1]\} = \mathbf{E}\{Z_1 y(t|\bar{F}_0^{Z_1})\}$. By Proposition 3 with $\bar{F} = \bar{F}_0^{Z_1}$, we have $y(t|\bar{F}_0^{Z_1}) = \bar{F}_0(t)^{Z_1} \left[\bar{G}(t-) + \int_t^\infty \rho(w - t|\bar{F}_0^{Z_1}) dG(w) \right] = \bar{F}_0(t)^{Z_1} \left[\bar{G}(t-) + \int_t^\infty \rho(w - t|\bar{F}_0^{Z_1}) dG(w) \right]$. Also, $\mathbf{E}\{Y_1(t)\} = \mathbf{E}\{\mathbf{E}[Y_1(t)|Z_1]\} = \mathbf{E}\{y(t|\bar{F}_0^{Z_1})\}$. Expression (35) for $\gamma(t)$ in the statement of the theorem is now obtained upon substitution.

Finally, since $\gamma(t)$, $\lambda_0(t)$, and $\mathbf{E}[Z_1 \bar{F}_0^{Z_1}(t)] / \mathbf{E}[\bar{F}_0^{Z_1}(t)]$ are all positive, continuous functions of t , it follows that $\Lambda(t) - \int_0^t \gamma(u) \lambda_0(u) du \equiv 0$ for all $t \leq t^*$ if and only if the stated consistency condition holds except possibly on a countable set of points with Lebesgue measure zero. \parallel

Proof of Corollary 1:

When $\bar{F}_0(t) = \exp(-\theta t)$, one immediately obtains $\rho(w|\bar{F}_0^{Z_1}) = \theta Z_1 w$. Substituting this into (35) and simplifying yields

$$\gamma(t) = \frac{-\bar{G}(t-)\psi'(s_t) + \psi''(s_t)\bar{G}(t-)r_t}{\bar{G}(t-)\psi(s_t) + \psi'(s_t)\bar{G}(t-)r_t},$$

where $s_t = \theta t$ and r_t is defined in the statement of the corollary. Dividing both numerator and denominator by $\bar{G}(t-)\psi(s_t)$ yields the stated result.

Since Z_1 is not degenerate and $\gamma^*(t) = \mathbf{E}[Z_1 \bar{F}_0^{Z_1}(t)]/\mathbf{E}[\bar{F}_0^{Z_1}(t)]$ we must only show that $\gamma(t) \neq \gamma^*(t)$ for $t \leq t^*$ in order to prove inconsistency. We shall prove this by showing that $\gamma(t) = \gamma^*(t)$ for $t \leq t^*$ if and only if Z_1 has a degenerate distribution function. Since $t \leq t^*$, it follows that $r_t > 0$. Assuming this to be the case, it is straightforward to prove that $\gamma(t) = \gamma^*(t)$ if and only if $\psi''(\lambda_0 t) - [\psi'(\theta t)]^2/\psi(\theta t) = 0$ for $t \leq t^*$. Without loss of generality we assume the mean of Z_1 is 1. Then, under the initial conditions $\psi(0) = 1$ and $\psi'(0) = -1$, the unique solution to this differential equation is simply $\psi(\theta t) = \exp\{-\theta t\}$. Recalling that $\psi(s) = \mathbf{E}[\exp\{-sZ_1\}]$ is the Laplace transform of Z , we see that $\psi(\theta t) = \exp\{-\theta t\}$ if and only if Z has a degenerate distribution function at $Z = 1$, proving the corollary. \parallel

Proof of Corollary 2:

Since $\bar{F}_0(t) = \exp(-\theta t)$, we may use Corollary 1 to find $\gamma(t)$. Making the substitution $\bar{G}(t) = \exp(-\eta t)$ yields $r_t = \theta/\eta$ upon simplification. Moreover, since $Z_1 \sim \text{Gamma}(\alpha, \alpha)$, we have that $\psi(\theta t) = \{\alpha/(\alpha + \theta t)\}^\alpha$. Thus, $\gamma(t) = [\gamma_1(t) + \frac{\theta}{\eta}\gamma_2(t)] / [1 + \frac{\theta}{\eta}\gamma_1(t)]$ for $\gamma_1(t) = -\psi'(\theta t)/\psi(\theta t) = \alpha/(\alpha + \theta t)$ and $\gamma_2(t) = \psi''(\theta t)/\psi(\theta t) = \alpha(\alpha + 1)/(\alpha + \theta t)$. Upon simplification we obtain $\gamma(t) = \left(\frac{\alpha}{\alpha + \theta t}\right) \left\{1 + \frac{1}{\alpha} \left[\left\{ \left(\frac{\alpha}{\alpha + \theta t}\right) \frac{\theta}{\eta} \right\} / \left\{ 1 + \left(\frac{\alpha}{\alpha + \theta t}\right) \frac{\theta}{\eta} \right\} \right] \right\}$. Since $\lambda_0(t) = \theta$, it follows by Theorem 4 that $\hat{\Lambda}(t) \xrightarrow{\text{pr}} \theta \int_0^t \gamma(w)dw$. A suitable change-of-variables yields $\theta \int_0^t \gamma(w)dw = \alpha \int_\alpha^{\alpha + \theta t} dv/v + (\alpha\theta/\eta) \int_\alpha^{\alpha + \theta t} dv/[v(v + \alpha\theta/\eta)]$, implying that $\hat{\Lambda}(t) \xrightarrow{\text{pr}} \Lambda_F(t)$, where $\Lambda_F(t) = -\log\left(\frac{\alpha}{\alpha + \theta t}\right)^\alpha - \log\left\{\left(\frac{\alpha}{\alpha + \theta t}\right) / \left(\frac{\alpha(1 + \theta/\eta)}{\alpha(1 + \theta/\eta) + \theta t}\right)\right\}$. Since $\hat{F}(t) = \prod_{w=0}^t [1 - \hat{\Lambda}(dw)]$ and because $\Lambda_F(t)$ is continuous, it follows that

$$\hat{F}(t) \xrightarrow{\text{pr}} \exp\{-\Lambda_F(t)\} = \left(\frac{\alpha}{\alpha + \theta t}\right)^\alpha \left\{ \left(\frac{\alpha}{\alpha + \theta t}\right) / \left(\frac{\alpha(1 + \theta/\eta)}{\alpha(1 + \theta/\eta) + \theta t}\right) \right\}.$$

Under the gamma frailty model being considered, the true marginal survivor function of the interoccurrence times is $\bar{F}(t) = \left(\frac{\alpha}{\alpha + \theta t}\right)^\alpha$. Hence, the limiting bias of $\hat{F}(t)$ is given by

$$\left(\frac{\alpha}{\alpha + \theta t}\right)^\alpha \left\{ \left[\left(\frac{\alpha}{\alpha + \theta t}\right) / \left(\frac{\alpha(1 + \theta/\eta)}{\alpha(1 + \theta/\eta) + \theta t}\right) \right] - 1 \right\} = -\left(\frac{\theta t}{\alpha + \theta t}\right) \left(\frac{\theta}{\theta + \eta}\right) \left(\frac{\alpha}{\alpha + \theta t}\right)^\alpha,$$

completing the proof. \parallel

		$n = 20$			$n = 50$			$n = 80$		
t	$\bar{F}(t)$	Mean	SE	ASE	Mean	SE	ASE	Mean	SE	ASE
.10	.5488	.0149	.1959	.1934	.0118	.1942	.1934	.0069	.1874	.1934
.20	.3012	.0160	.1888	.1842	.0141	.1867	.1842	.0056	.1734	.1842
.30	.1653	.0140	.1653	.1548	.0030	.1575	.1548	.0013	.1486	.1548
.40	.0907	.0126	.1382	.1248	.0009	.1228	.1248	.0003	.1283	.1248

Table 1: Simulated means and standard errors, together with the asymptotic standard errors, of $\sqrt{n}[\hat{F}(t) - \bar{F}(t)]$ for sample sizes $n \in \{20, 50, 80\}$, $\theta = 6$, $\eta = 1$, and for different values of t , under the i.i.d. interoccurrence times model with $F = \text{EXP}(\theta)$ and $G = \text{EXP}(\eta)$. For each n , 1000 replications were performed.

True α	2			6			∞		
n	20	50	80	20	50	80	20	50	80
True ξ	.667	.667	.667	.856	.856	.856	1.00	1.00	1.00
Mean of $\hat{\xi}$.718	.687	.681	.880	.871	.868	.987	.992	.992
Median of $\hat{\xi}$.717	.685	.680	.889	.874	.869	.9995	.9996	.9997
Observed S.E. of $\hat{\xi}$.125	.076	.062	.092	.060	.047	.029	.016	.013

Table 2: Characteristics of the simulation runs for $\theta = 6$ and $\eta = 1$ with regards to the estimator $\hat{\alpha}$ under the gamma frailty model. The reparametrization is $\xi = \alpha/(1 + \alpha)$ and $\hat{\xi} = \hat{\alpha}/(1 + \hat{\alpha})$. Each combination of α and n had 1000 replications.

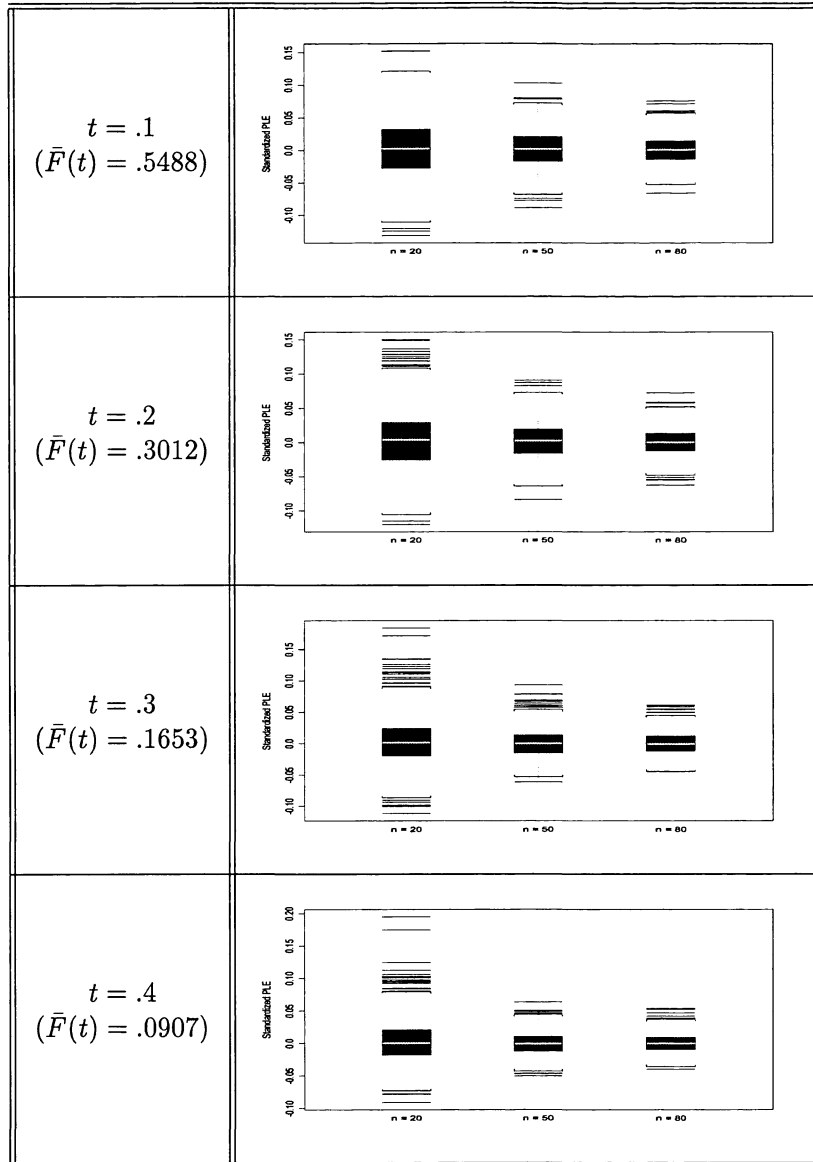


Figure 1: Boxplots of the simulated sampling distributions of $\sqrt{n}[\hat{F}(t) - \bar{F}(t)]$ for sample sizes $n \in \{20, 50, 80\}$, $\theta = 6$, $\eta = 1$, and for different values of t , under the model with $F = \text{EXP}(\theta)$ and $G = \text{EXP}(\eta)$.

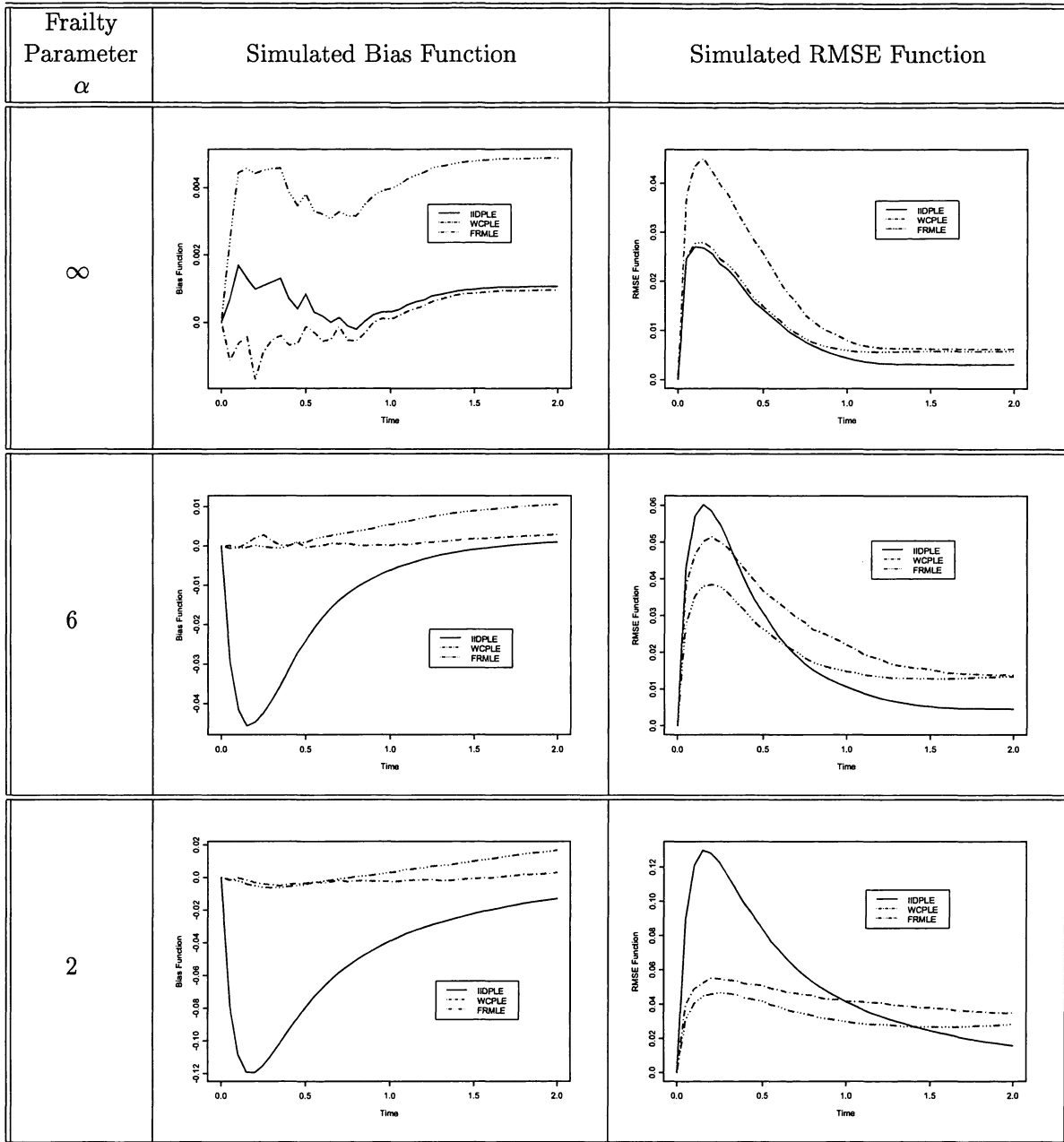


Figure 2: Simulated biases and root-mean-squared errors of the estimators $\hat{F}(\infty, t)$ (IID-
 PLE), $\hat{S}(t)$ (WCPLE), and under the frailty model, the maximum likelihood estimator $\tilde{F}(\infty, t)$
 (FRMLE). The simulation parameters were $n = 50$, $\theta = 6$, $\eta = 1$, and with 1000 replications.

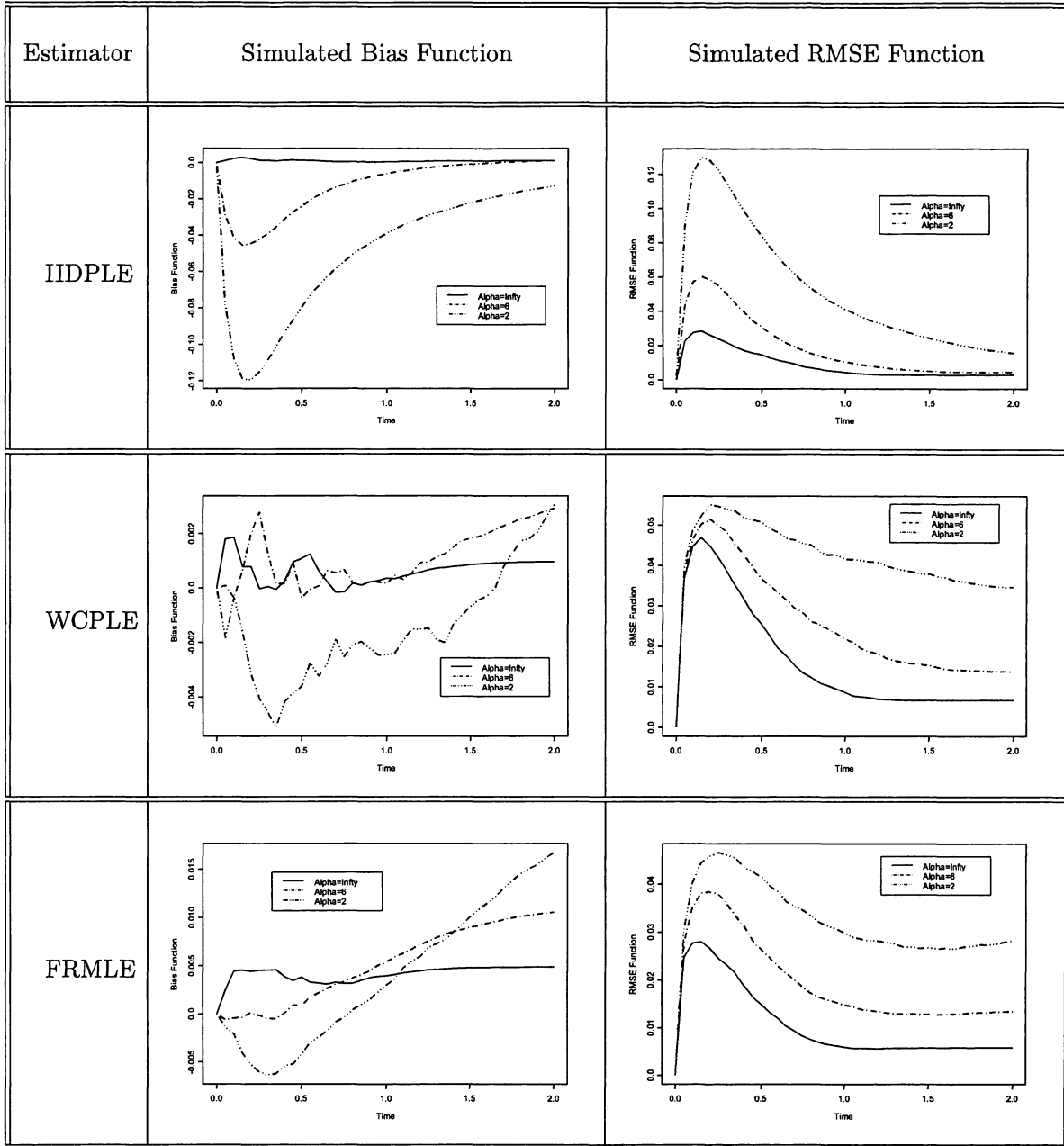


Figure 3: Simulated biases and root-mean-squared errors of the estimators $\hat{F}(\infty, t)$ (IIDPLE), $\hat{S}(t)$ (WCPL), and the frailty model maximum likelihood estimator $\tilde{F}(\infty, t)$ (FRMLE). The simulation parameters were $n = 50$, $\theta = 6$, $\eta = 1$, and with 1000 replications.

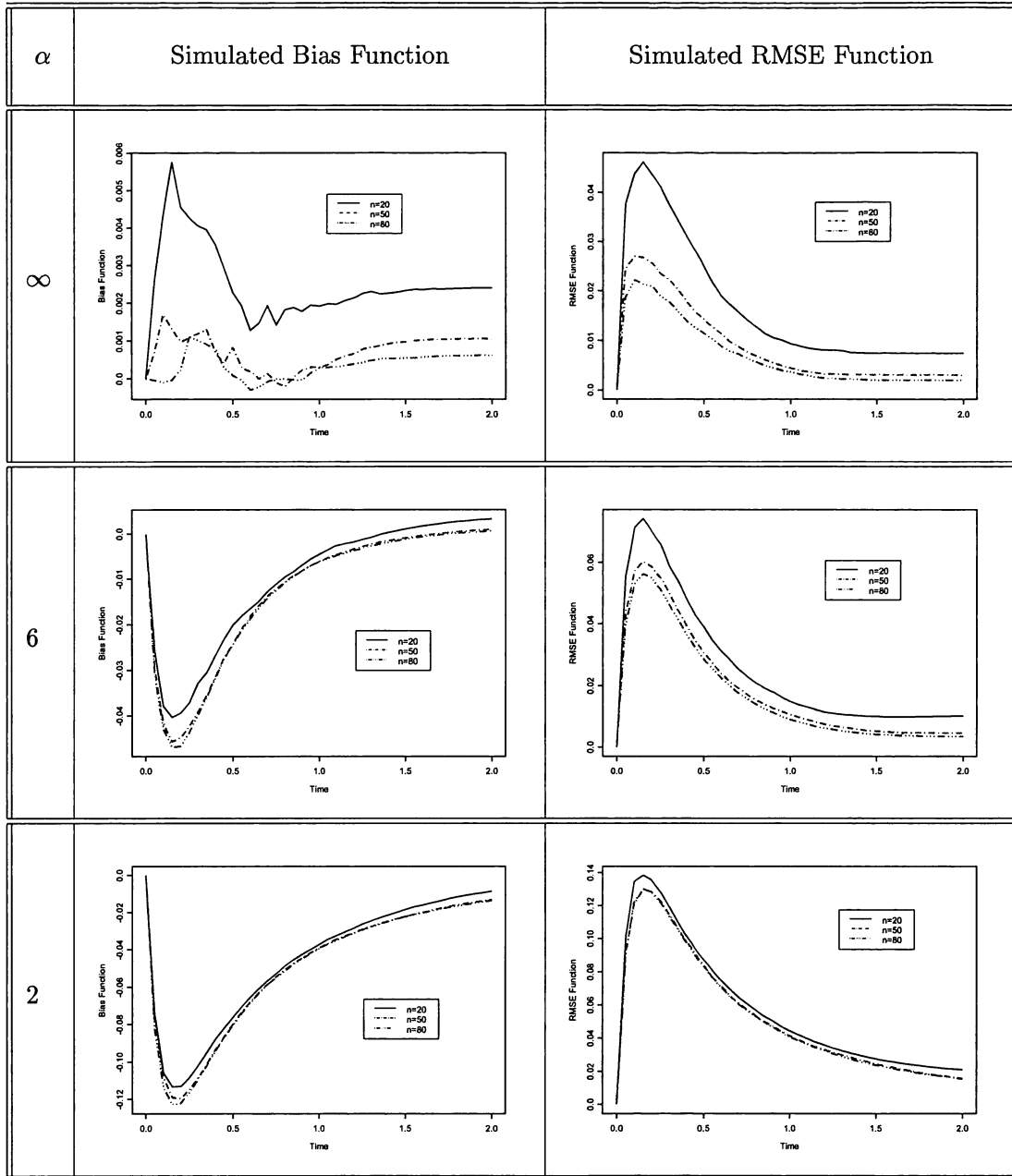


Figure 4: Simulated biases and root-mean-squared errors of the estimator $\hat{F}(\infty, t)$ (IIDPLE) overlaid as the sample size changes. The simulation parameters were $\theta = 6$, $\eta = 1$, and with 1000 replications.

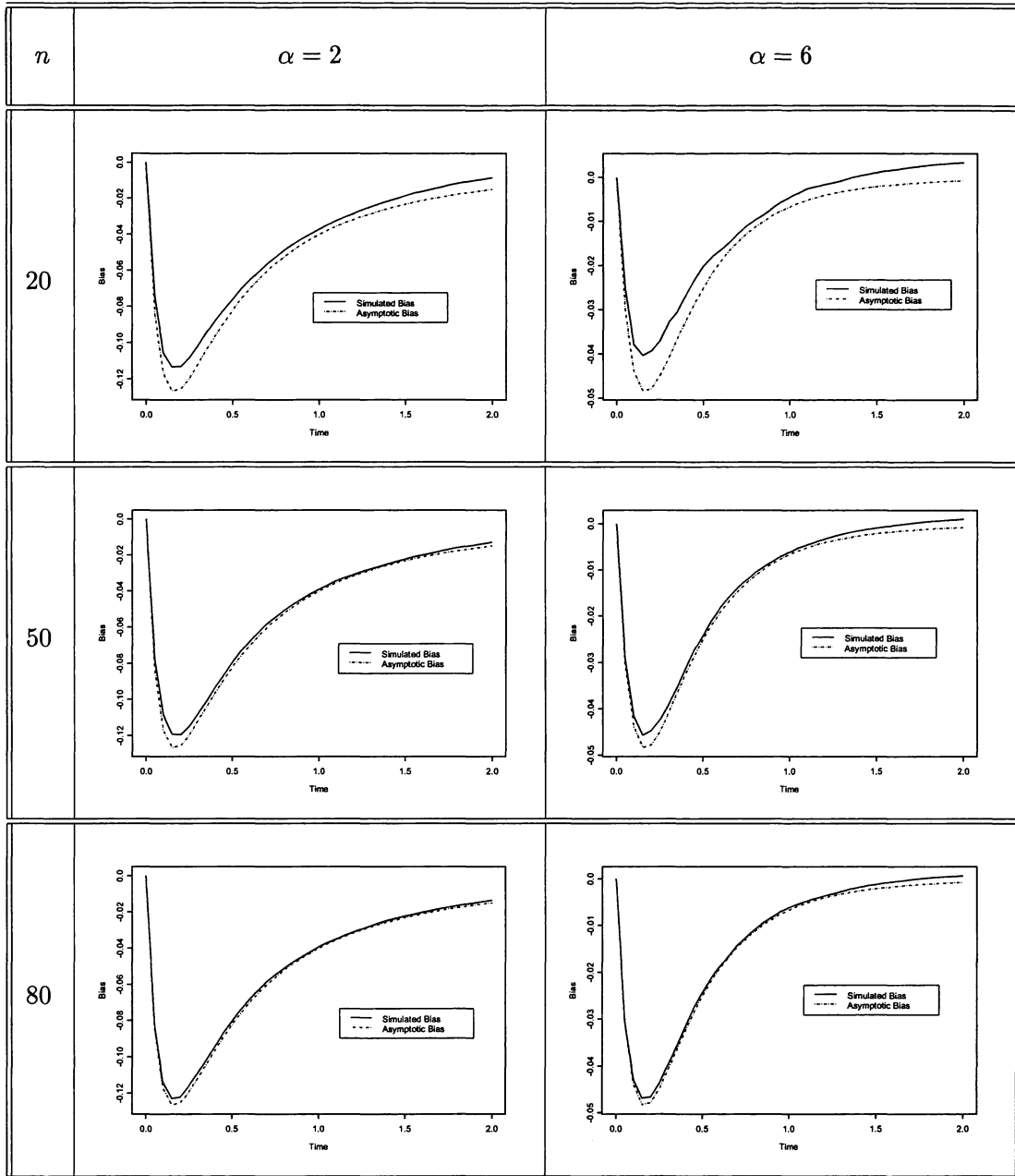


Figure 5: Overlaid plots of the simulated bias function and limiting bias expression in (37) for the estimator $\hat{F}(\infty, t)$ under the gamma frailty model. The simulation parameters were $\theta = 6$, $\eta = 1$, and with 1000 replications.

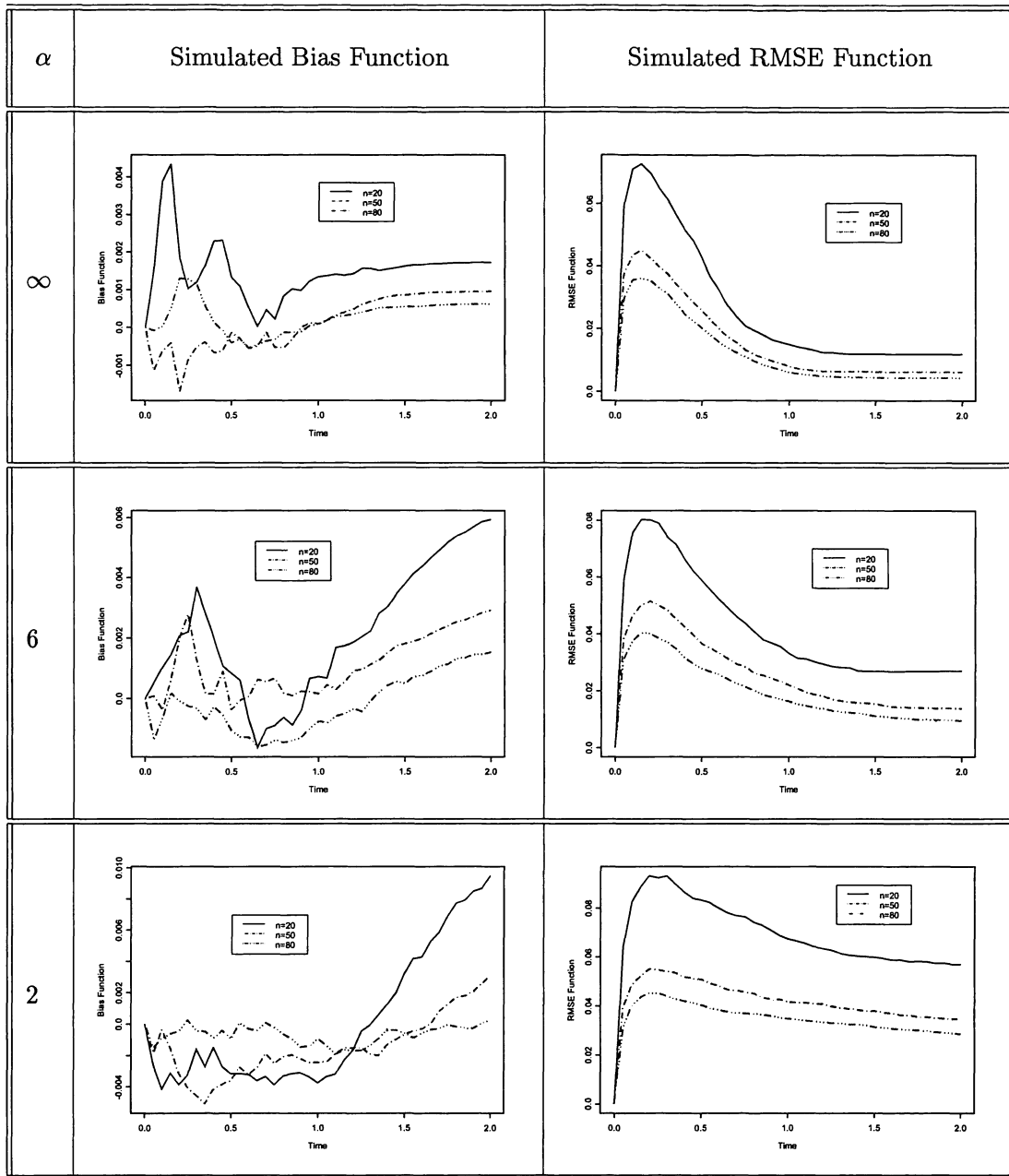


Figure 6: Simulated biases and root-mean-squared errors of the estimator $\hat{S}(t)$ (WCPLE) overlaid as the sample size changes. The simulation parameters were $\theta = 6$, $\eta = 1$, and with 1000 replications.

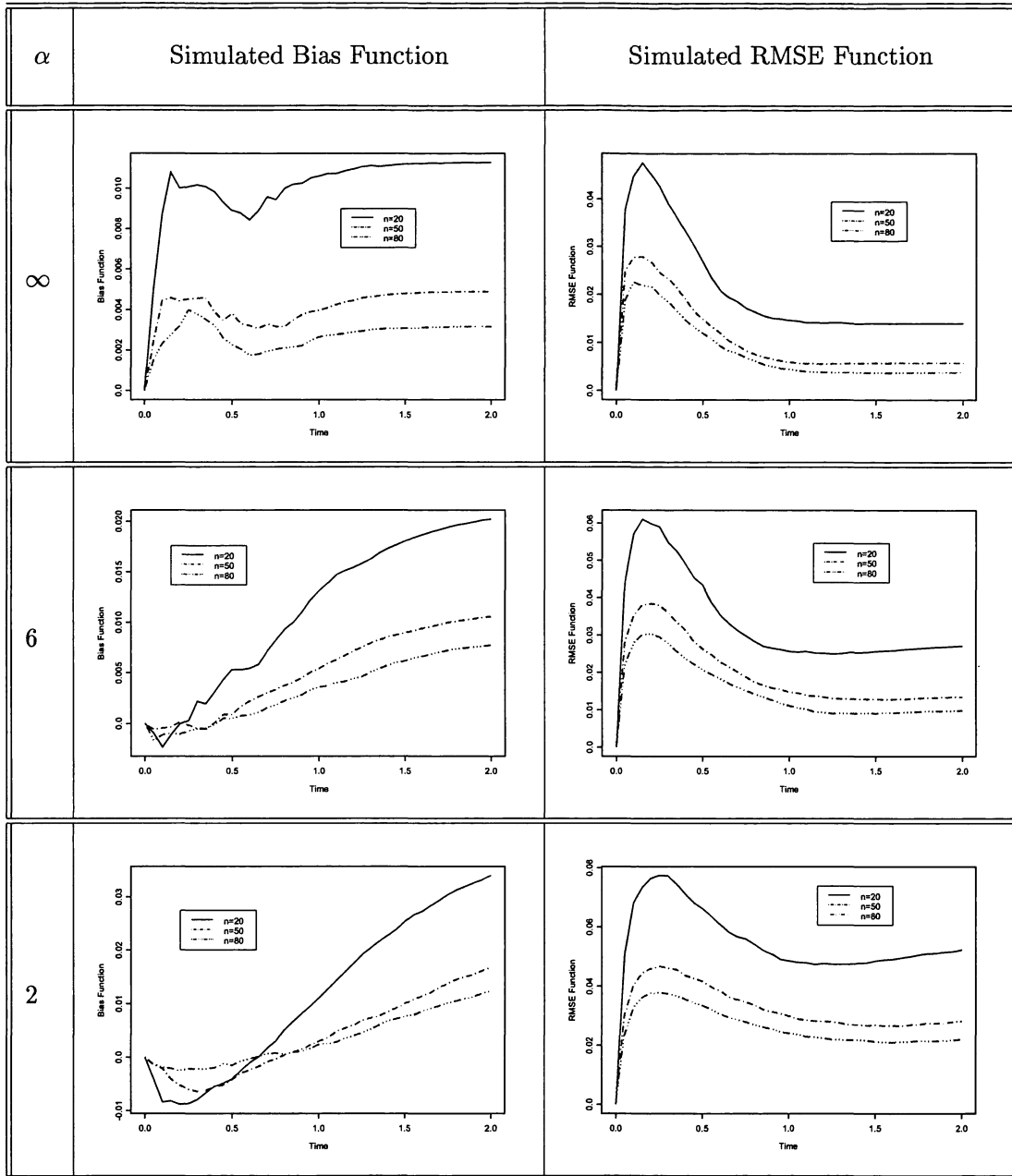


Figure 7: Simulated biases and root-mean-squared errors of the estimator $\tilde{F}(\infty, t)$ (FRMLE) overlaid as the sample size changes. The simulation parameters were $\theta = 6$, $\eta = 1$, and with 1000 replications.

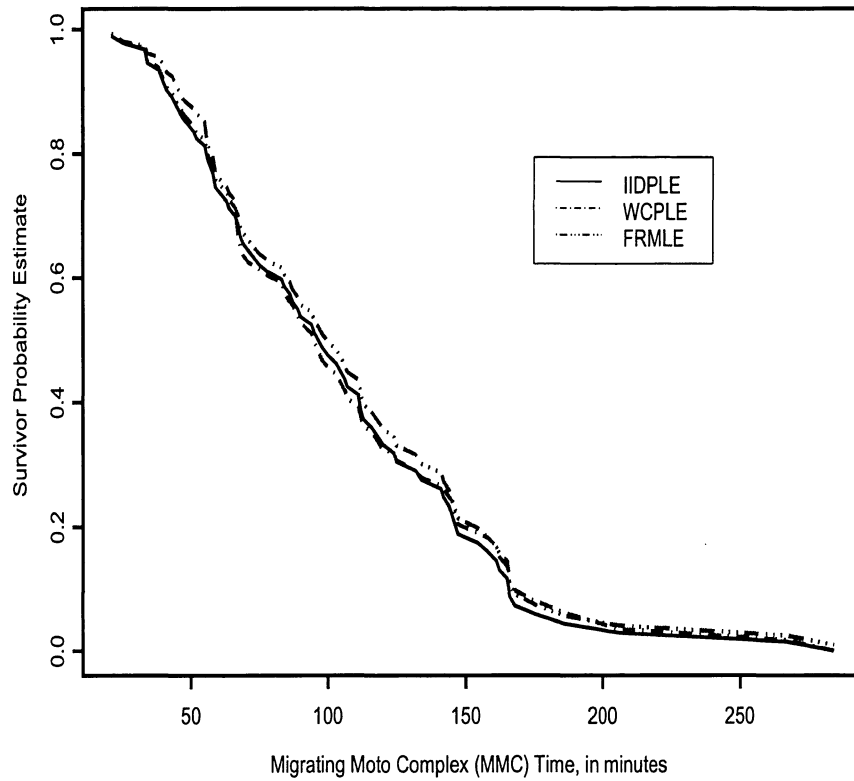


Figure 8: Plots of the three survivor function estimators for the MMC data set. The maximum likelihood estimate of the frailty parameter α under the gamma frailty model is $\hat{\alpha} = 10.17562$, or, equivalently, $\hat{\alpha}/(1 + \hat{\alpha}) = 0.9105$.