

Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data

BY LYNN M. JOHNSON

Department of Statistical Science, Cornell University, Ithaca, New York 14853, U.S.A.
lms86@cornell.edu

AND ROBERT L. STRAWDERMAN

Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14853, U.S.A.
rls54@cornell.edu

SUMMARY

This paper extends the induced smoothing procedure of [Brown & Wang \(2006\)](#) for the semi-parametric accelerated failure time model to the case of clustered failure time data. The resulting procedure permits fast and accurate computation of regression parameter estimates and standard errors using simple and widely available numerical methods, such as the Newton–Raphson algorithm. The regression parameter estimates are shown to be strongly consistent and asymptotically normal; in addition, we prove that the asymptotic distribution of the smoothed estimator coincides with that obtained without the use of smoothing. This establishes a key claim of [Brown & Wang \(2006\)](#) for the case of independent failure time data and also extends such results to the case of clustered data. Simulation results show that these smoothed estimates perform as well as those obtained using the best available methods at a fraction of the computational cost.

Some key words: Censoring; Convex optimization; Multivariate survival data; Rank regression.

1. INTRODUCTION

The need to analyze failure time data, possibly subject to right-censoring, arises in a number of fields, including medicine, economics, epidemiology, demography and engineering. Semiparametric regression models are commonly used for characterizing the relationship between failure time and covariates, with the Cox proportional hazards regression model ([Cox, 1972](#)) being used almost exclusively in practice. The accelerated failure time model (e.g. [Kalbfleisch & Prentice, 2002](#)) provides a useful but infrequently used alternative to the Cox proportional hazards model. Letting \bar{T}_i and X_i respectively denote the failure time and vector of observed covariates for subject i ($i = 1, \dots, n$), the accelerated failure time model specifies that $\log \bar{T}_i = X_i' \beta + \epsilon_i$, where the error terms are independent and identically distributed with an unspecified distribution. The regression coefficient β has a nice interpretation and a variety of simple estimators are available when $\bar{T}_1, \dots, \bar{T}_n$ are fully observed. The infrequent use of this model in applications involving censored failure time data may be largely attributed to the computational challenges that arise in both regression parameter and covariance matrix estimation.

In the presence of censoring, the observed data for subject i can be described by the triplet (T_i, Δ_i, X_i) where $T_i = \min(\bar{T}_i, C_i)$, $\Delta_i = I(\bar{T}_i \leq C_i)$, and C_i denotes the censoring time for

subject i . [Tsiatis \(1990\)](#) proposes to estimate β using a weighted estimating equation of the form

$$W_n^*(\beta) = \sum_{i=1}^n w_i(\beta) \Delta_i \left[X_i - \frac{\sum_{j=1}^n X_j I\{e_j(\beta) \geq e_i(\beta)\}}{\sum_{j=1}^n I\{e_j(\beta) \geq e_i(\beta)\}} \right], \quad (1)$$

where $e_i(\beta) = \log(T_i) - X_i' \beta$ and $w_i(\cdot)$ are nonnegative weight functions ($i = 1, \dots, n$). Due to the fact that β appears in this expression only inside indicator functions, $W_n^*(\beta)$ is not a continuous function of β and a solution to $W_n^*(\beta) = 0$ typically does not exist. Parameter estimates may instead be obtained by minimizing $\|W_n^*(\beta)\|$, where $\|v\|$ denotes $(v'v)^{1/2}$ for a vector v . However, this minimization problem may admit several solutions and, because $W_n^*(\beta)$ is not necessarily monotone in β , the resulting set of minimizers is not even guaranteed to be convex. Hence, despite the existence of a consistent and asymptotically normal sequence of generalized solutions (e.g. [Tsiatis, 1990](#)), identifying this sequence can be challenging in practice.

[Fyngenson & Ritov \(1994\)](#) show that using the Gehan weight function $w_i(\beta) = \sum_{j=1}^n I\{e_j(\beta) \geq e_i(\beta)\}$ ($i = 1, \dots, n$) leads to the monotone estimating equation

$$W_n(\beta) = \sum_{i=1}^n \sum_{j=1}^n \Delta_i (X_i - X_j) I\{e_i(\beta) - e_j(\beta) \leq 0\}. \quad (2)$$

Recognizing that $W_n(\beta)$ is the gradient of the convex objective function

$$O_n(\beta) = \sum_{i=1}^n \sum_{j=1}^n \Delta_i \{e_j(\beta) - e_i(\beta)\} I\{e_i(\beta) - e_j(\beta) \leq 0\}, \quad (3)$$

a regression parameter estimate may be obtained by minimizing $O_n(\beta)$ with respect to β . The resulting set of solutions is convex and thus easier to locate than in the general case. However, even in this comparatively nice setting, the associated lack of smoothness continues to present computational challenges. Numerous methods have been proposed for finding parameter estimates derived from (2) and (3). The most promising methods to date utilize linear programming techniques (e.g. [Jin et al., 2003](#)). However, while such methods can be implemented with relative ease, the computational burden can be high, especially with large datasets.

Estimating the covariance matrix of the regression parameter estimate obtained under the accelerated failure time model remains a challenging problem. [Fyngenson & Ritov \(1994\)](#) show that the regression parameter estimate derived from (3) is asymptotically normal with a covariance matrix that involves the hazard function of the unspecified error distribution. Direct estimation of the covariance matrix thus requires an estimate of this hazard function. [Tsiatis \(1990\)](#) suggests kernel-based estimation, whereas [Fyngenson & Ritov \(1994\)](#) suggest a form of numerical differentiation. Both have proven to be unstable choices in the presence of censored data and several authors have since tackled this problem in other ways; see, for example, [Jones \(1997\)](#) and [Jin et al. \(2003\)](#). [Jin et al. \(2003\)](#) propose to randomly reweight the Gehan log-rank objective function (3) and then minimize the resulting perturbed objective function. Repeating this process a large number of times, the covariance matrix may then be estimated using the empirical covariance matrix of these parameter estimates. This interesting and useful approach eliminates the need to estimate the indicated hazard function. However, the computationally intensive nature of this procedure quickly becomes unwieldy, particularly with large datasets. [Huang \(2002\)](#), [Strawderman \(2005\)](#) and [Jin et al. \(2006a\)](#) propose useful alternatives in three related problems.

Several authors have recently proposed useful smoothing methods for nonsmooth estimating equations arising in the accelerated failure time model; see, for example, [Brown & Wang \(2005; 2006\)](#), [Heller \(2007\)](#) and [Song et al. \(2007\)](#). Each of these smoothing methods leads to a

continuously differentiable objective or estimating function that can be dealt with using standard numerical methods. Of direct relevance to this paper are [Brown & Wang \(2006\)](#) and [Heller \(2007\)](#). Building on [Brown & Wang \(2005\)](#), [Brown & Wang \(2006\)](#) propose the use of induced smoothing for the Gehan estimating equation (2). This method, described in more detail in § 2, involves solving the equation $E_Z\{W_n(\beta + \Gamma_n Z)\} = 0$, where $W_n(\cdot)$ is given in (2), Z is a continuous, mean zero normal random vector independent of all of the data, and Γ_n is a sequence of matrices converging to zero with elements $\Gamma_{n,ij} = O_p(n^{-1/2})$. The smoothed estimating equation $E_Z\{W_n(\beta + \Gamma_n Z)\}$ reduces to

$$\tilde{W}_n(\beta) = \sum_{i=1}^n \sum_{j=1}^n \Delta_i(X_i - X_j) \Phi \left\{ \frac{e_j(\beta) - e_i(\beta)}{r_{n,ij}} \right\}, \quad (4)$$

where $r_{n,ij}^2 = (X_i - X_j)' \Sigma_n (X_i - X_j)$, $\Sigma_n = \Gamma_n^2$, and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. In a related vein, [Heller \(2007\)](#) directly approximates the indicator function $I(u \leq 0)$ in $W_n(\beta)$ with $1 - \Upsilon(u/h)$, where $\Upsilon(\cdot)$ denotes a local distribution function satisfying certain conditions and the fixed scalar parameter h is used to control the accuracy of approximation. The resulting estimating equation,

$$W_n^{**}(\beta) = \sum_{i=1}^n \sum_{j=1}^n \Delta_i(X_i - X_j) \Upsilon \left\{ \frac{e_j(\beta) - e_i(\beta)}{h} \right\}, \quad (5)$$

has the same structure as (4). In fact, upon taking $\Upsilon(\cdot)$ to be the standard normal distribution function $\Phi(\cdot)$, (5) is essentially a special case of (4), utilizing a fixed bandwidth h in place of the covariate-dependent bandwidth $r_{n,ij}$. [Heller \(2007\)](#) also proposes a robust version of (5) having a bounded influence function. A potential difference between (4) and (5) lies in the ability of the former to employ a smoothing parameter that respects the scaling and covariance structure of the solution sequence. [Brown & Wang \(2006\)](#) claim but do not prove that the sequence of solutions obtained under (4) has the same asymptotic distribution as that obtained in the absence of smoothing. [Heller \(2007\)](#) proves that the solution sequence obtained under (5) is consistent and asymptotically normal, provided that h satisfies $nh \rightarrow \infty$ and $nh^4 \rightarrow 0$ as $n \rightarrow \infty$. Interestingly, [Heller \(2007\)](#) further proves that (2) and (5) are asymptotically equivalent but does not establish the equivalence result posited in [Brown & Wang \(2006\)](#).

The problem of regression parameter estimation under the accelerated failure time model with correlated survival data has also been considered. For example, [Lin & Wei \(1992\)](#), [Lee et al. \(1993\)](#) and [Jin et al. \(2006b\)](#) consider the setting in which failure times are grouped into clusters, such that observations within a cluster may be correlated but observations in distinct clusters may be considered independent. Each proposes a marginal method for rank-based estimation of regression parameters, avoiding the need to model the correlation structure among observations. [Jin et al. \(2006b\)](#) also devise a suitable extension of the resampling procedure proposed in [Jin et al. \(2003\)](#). [Pan \(2001\)](#) and [Zhang & Peng \(2007\)](#) instead propose frailty models, handling the dependence among failure times within a cluster using an additive cluster level random effect; see also [Strawderman \(2006\)](#) for related work in the case of a recurrent event outcome. These various methods suffer from estimation and computational challenges that equal or exceed those experienced in the case of independent failure time data. However, to our knowledge, the validity and utility of smoothing methods like those developed in [Brown & Wang \(2006\)](#) and [Heller \(2007\)](#) have not been investigated in connection with the clustered data problem.

This paper extends the smoothing procedure of [Brown & Wang \(2006\)](#) to the problem of marginal estimation of the regression parameter in the presence of clustered data. We prove that the resulting estimator is consistent and asymptotically normal in both the independent and

correlated data settings. We further establish the equivalence of these limiting distributions with those arising in the unsmoothed case, providing rigorous justification of the equivalence claim made in [Brown & Wang \(2006\)](#) for the case of independent failure times and its extension to the setting of clustered data. Several possible methods of covariance matrix estimation are evaluated, among them a generalization of the [Brown & Wang \(2006\)](#) procedure and a modification of the resampling procedure due to [Jin et al. \(2006b\)](#). A useful consequence of developing the extended [Brown & Wang \(2006\)](#) estimator is an easy-to-compute sandwich estimator that avoids the need for resampling. The proposed methods substantially ease the computational burden of previously proposed methods for parameter and covariance matrix estimation.

2. METHODOLOGY AND KEY RESULTS

2.1. Notation and assumptions

Consider a random sample of n independent clusters with K_i members in the i th cluster. Let \bar{T}_{ik} and C_{ik} denote the failure time and censoring time for the k th member of the i th cluster, and let X_{ik} denote the corresponding $p \times 1$ vector of covariates. We assume that $(\bar{T}_{i1}, \dots, \bar{T}_{iK_i})'$ and $(C_{i1}, \dots, C_{iK_i})'$ are independent conditional on the covariates $(X_{i1}, \dots, X_{iK_i})'$. Let the survival data for the k th member of i th cluster be denoted $W_{ik} = (\log T_{ik}, \Delta_{ik}, X_{ik})'$, where $T_{ik} = \min(\bar{T}_{ik}, C_{ik})$ and $\Delta_{ik} = I(\bar{T}_{ik} \leq C_{ik})$.

We assume that the marginal distribution of T_{ik} follows the accelerated failure time model

$$\log \bar{T}_{ik} = X'_{ik}\beta_0 + \epsilon_{ik},$$

where β_0 is a $p \times 1$ vector of unknown regression parameters contained in a compact subset \mathbb{B} of \mathbb{R}^p and $(\epsilon_{i1}, \dots, \epsilon_{iK_i})'$ ($i = 1, \dots, n$) are independent random error vectors. Within each cluster i , the error terms $\epsilon_{i1}, \dots, \epsilon_{iK_i}$ may be correlated; however, as in [Jin et al. \(2006b, § 4\)](#), we assume that these error terms are exchangeable with a common, unknown marginal distribution. That is, for any $i, j = 1, \dots, n$ and $K \leq \min(K_i, K_j)$, the vectors $(\epsilon_{i1}, \dots, \epsilon_{iK})'$ and $(\epsilon_{j1}, \dots, \epsilon_{jK})'$ have the same distribution. Evidently, the case of independent failure time data follows as a special case of the above model upon setting $K_i = 1$ for all i .

2.2. Estimation for clustered data using the Gehan weight

Let $e_{ik}(\beta) = \log(T_{ik}) - X'_{ik}\beta$. Under the assumptions of [§ 2.1](#), the relevant extension of [\(3\)](#) to the clustered data setting may be written

$$L_n(\beta) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{k=1}^{K_i} \sum_{j=1}^n \sum_{l=1}^{K_j} \Delta_{ik} \{e_{jl}(\beta) - e_{ik}(\beta)\} I\{e_{ik}(\beta) - e_{jl}(\beta) \leq 0\}; \quad (6)$$

see, for example, [Jin et al. \(2006b, § 4\)](#). Observe that $L_n(\beta)$ is a continuous convex function for $\beta \in \mathbb{B}$ and thus differentiable almost everywhere. The derivative of the objective function with respect to β , or $S_n(\beta) = \nabla L_n(\beta)$, is the discontinuous function

$$S_n(\beta) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{k=1}^{K_i} \sum_{j=1}^n \sum_{l=1}^{K_j} \Delta_{ik} (X_{ik} - X_{jl}) I\{e_{ik}(\beta) - e_{jl}(\beta) \leq 0\}. \quad (7)$$

Let $\hat{\beta}_n = \operatorname{argmin}_{\beta \in \mathbb{B}} L_n(\beta)$. The solution to this minimization problem may not be unique; however, the convexity of $L_n(\beta)$ implies that the set of minimizers on \mathbb{B} is convex (e.g. [Fygenson & Ritov, 1994](#)). The lack of smoothness makes minimization of $L_n(\beta)$ computationally challenging, particularly with multiple covariates. However, under regularity conditions to

be described later, the results of Jin et al. (2006b, Theorem 5) imply that there exists a sequence of solutions that is strongly consistent for β_0 and, in addition, such that $n^{1/2}(\hat{\beta}_n - \beta_0)$ converges in distribution to a $N(0, A^{-1}\Omega A^{-1})$ random vector, where $\Omega = \lim_{n \rightarrow \infty} \text{var}\{n^{1/2}S_n(\beta_0)\}$ and $A = \nabla S_0(\beta_0)$ for $S_0(\beta) = \lim_{n \rightarrow \infty} S_n(\beta)$. An explicit formula for A is provided in (A1). In addition to the numerical challenges that arise in computing the solution $\hat{\beta}_n$, variance estimation is difficult due to the dependence on A and the fact that $S_n(\beta)$ is not differentiable in β .

2.3. Induced smoothing for clustered data

Brown & Wang (2005) propose an induced smoothing method for approximating discontinuous but monotone estimating functions using continuously differentiable functions. Assuming independent failure time observations, Brown & Wang (2006) apply this smoothing method to the problem of estimating the regression parameter in the accelerated failure time model, using (4) in place of (2). As shown below, the extension of this methodology to the problem of estimating β in the clustered data setting under the assumptions of § 2.1 is straightforward.

Let Z be a $N(0, I_p)$ random vector independent of the data, where I_p denotes the $p \times p$ identity matrix. Let Γ be a $p \times p$ matrix such that $\|\Gamma\| = O(1)$ and $\Gamma^2 = \Sigma$, where Σ is some symmetric, positive definite matrix. Then, similarly to Brown & Wang (2005, 2006), a smoothed score function may be constructed by adding the random perturbation $n^{-1/2}\Gamma Z$ to the argument of the score function $S_n(\beta)$ in (7) and then taking the expectation with respect to Z . Specifically, with $\tilde{S}_n(\beta) = E_Z\{S_n(\beta + n^{-1/2}\Gamma Z)\}$, an easy calculation shows that

$$\tilde{S}_n(\beta) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{k=1}^{K_i} \sum_{j=1}^n \sum_{l=1}^{K_j} \Delta_{ik} (X_{ik} - X_{jl}) \Phi \left[n^{1/2} \left\{ \frac{e_{jl}(\beta) - e_{ik}(\beta)}{r_{ikjl}} \right\} \right], \tag{8}$$

where $r_{ikjl}^2 = (X_{ik} - X_{jl})' \Sigma (X_{ik} - X_{jl})$. With $K_i = K_j = 1$ for $i, j = 1, \dots, n$, this estimating equation reduces to (4). Alternatively, one might work directly with the smoothed objective function $\tilde{L}_n(\beta) = E_Z\{L_n(\beta + n^{-1/2}\Gamma Z)\}$. Let $\phi(\cdot)$ denote the standard normal density function. Then, using standard results for normal random variables and integration by parts, we have

$$\tilde{L}_n(\beta) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{k=1}^{K_i} \sum_{j=1}^n \sum_{l=1}^{K_j} \Delta_{ik} \left[\{e_{jl}(\beta) - e_{ik}(\beta)\} H_{ikjl}^{(n)}(\beta) + \frac{r_{ikjl}}{n^{1/2}} h_{ikjl}^{(n)}(\beta) \right], \tag{9}$$

where

$$H_{ikjl}^{(n)}(\beta) = \Phi \left[n^{1/2} \left\{ \frac{e_{jl}(\beta) - e_{ik}(\beta)}{r_{ikjl}} \right\} \right], \quad h_{ikjl}^{(n)}(\beta) = \phi \left[n^{1/2} \left\{ \frac{e_{jl}(\beta) - e_{ik}(\beta)}{r_{ikjl}} \right\} \right]. \tag{10}$$

A straightforward calculation shows that $\nabla \tilde{L}_n(\beta) = \tilde{S}_n(\beta)$.

Let $\tilde{\beta}_n = \text{argmin}_{\beta \in \mathbb{B}} \tilde{L}_n(\beta)$. The smoothed objective function, $\tilde{L}_n(\beta)$, is convex and continuously differentiable and standard numerical methods can be used to efficiently compute $\tilde{\beta}_n$. Alternatively, $\tilde{\beta}_n$ can be found as the multivariate root of $\tilde{S}_n(\beta)$. The asymptotic results, summarized below and proved in the Appendix, also imply that inference for $\tilde{\beta}_n$ is straightforward.

THEOREM 1. *Let $\Sigma = \Gamma^2$ be any symmetric and positive definite matrix with $\|\Gamma\| < \infty$. Under conditions A1–A4 of the Appendix, $\tilde{\beta}_n$ is a strongly consistent estimator of β_0 .*

THEOREM 2. *Let $\Sigma = \Gamma^2$ be any symmetric and positive definite matrix with $\|\Gamma\| < \infty$. Under conditions A1–A6 of the Appendix, $n^{1/2}(\tilde{\beta}_n - \beta_0)$ converges in distribution to $N(0, \Psi)$, where $\Psi = A^{-1}\Omega A^{-1}$, $\Omega = \lim_{n \rightarrow \infty} \text{var}\{n^{1/2}S_n(\beta_0)\}$ and $A = \nabla S_0(\beta_0)$ is defined in (A1).*

The above results provide theoretical justification for the proposed smoothing procedure when estimating regression parameters under the marginal accelerated failure time model with clustered failure time data. Importantly, the matrices A and Ω in Theorem 2 are defined in terms of the limiting behaviour (7), demonstrating that the limiting distribution of $n^{1/2}(\tilde{\beta}_n - \beta_0)$ coincides with that for $n^{1/2}(\hat{\beta}_n - \beta_0)$, where $\hat{\beta}_n$ is obtained via the unsmoothed objective function (6). Since justification for the independent data case follows directly from the above theorems upon setting $K_i = 1$ ($i = 1, \dots, n$), Theorems 1 and 2 also provide rigorous justification for the claims made in Brown & Wang (2006).

Remark 1. The above results hold for a general smoothing matrix Γ that satisfies certain minimal conditions. Brown & Wang (2006) propose an iterative procedure for estimating β_0 , in which $\Sigma = \Gamma^2$ is updated at each iteration using successive estimates of Ψ . One implementation of this procedure in the clustered data setting is provided in § 3.2.

Remark 2. The smoothing bandwidth employed in (8) and (9) is $O(n^{-1/2})$, where n denotes the number of independent clusters. In the absence of clustering, Heller (2007) recommends the choice $h = \hat{\sigma}n^{-0.26}$ in (5), where $\hat{\sigma}$ is an estimate of the residual variance obtained using a minimizer of the unsmoothed equation (3). The selection $h = O(n^{-0.26})$ is motivated as that which provides “the quickest rate of convergence while satisfying the bandwidth constraint $nh^4 \rightarrow 0$ ”. In asymptotic terms, Theorem 2 suggests that such oversmoothing is unnecessary.

3. METHODS OF VARIANCE ESTIMATION

3.1. The sandwich variance estimator

The sandwich form of the covariance matrix of $n^{1/2}(\tilde{\beta}_n - \beta_0)$ in Theorem 2 suggests a natural estimator provided that suitable estimates of both A and Ω can be found. In the independent data case, Brown & Wang (2006) suggest estimating A with $\tilde{A}_n = \nabla \tilde{S}_n(\tilde{\beta}_n)$; Theorems 1 and 2 imply that this remains a consistent estimator in the clustered data setting. Brown & Wang (2006) further suggest several estimates of Ω , including the asymptotic variance of $n^{1/2}S_n(\beta_0)$ provided in Jin et al. (2003) and an estimator of Ω based on the U-statistic structure of the estimating function (4). However, neither estimator of Ω properly accounts for the correlation between observations within a cluster. Lee et al. (1993) show that the asymptotic variance of $n^{1/2}S_n(\beta_0)$ in the clustered data case can be consistently estimated via $\hat{\Omega}_n = \hat{\Omega}_n(\hat{\beta}_n)$, where

$$\hat{\Omega}_n(\beta) = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{K_i} \sum_{l=1}^{K_i} \{\hat{\xi}_{ik}(\beta)\}^{\otimes 2},$$

$v^{\otimes 2} = vv'$ for any vector v , and

$$\begin{aligned} \hat{\xi}_{ik}(\beta) = & \sum_{j=1}^n \sum_{l=1}^{K_j} \left(\frac{\Delta_{ik}}{n} (X_{ik} - X_{jl}) I\{e_{ik}(\beta) < e_{jl}(\beta)\} \right. \\ & \left. - \frac{\Delta_{jl}}{n} \left[\frac{\sum_{r=1}^n \sum_{s=1}^{K_r} (X_{ik} - X_{rs}) I\{e_{rs}(\beta) \geq e_{jl}(\beta)\}}{\sum_{m=1}^n \sum_{k=1}^{K_m} I\{e_{mk}(\beta) \geq e_{jl}(\beta)\}} \right] I\{e_{ik}(\beta) \geq e_{jl}(\beta)\} \right). \end{aligned}$$

Conditions A1–A5 of the Appendix ensure that $\hat{\Omega}_n$ is a consistent estimator of Ω ; with the addition of condition A6, Ψ can be consistently estimated in the clustered data setting using

$$\hat{\Psi}_n = \tilde{A}_n^{-1} \hat{\Omega}_n \tilde{A}_n^{-1}. \tag{11}$$

3.2. The Brown & Wang (2006) procedure for clustered data

As suggested in Brown & Wang (2006), an iterative procedure can be used to simultaneously estimate the regression parameters and their covariance matrix. Denoting $\tilde{A}_n(\beta) = \nabla \tilde{S}_n(\beta)$, the proposed procedure consists of the following steps in the clustered data setting:

Step 1. Set $i = 0$ and initialize $\hat{\Sigma}_{(0)}$ such that $\|\hat{\Sigma}_{(0)}\| = O(1)$; for example, $\hat{\Sigma}_{(0)} = I_p$.

Step 2. Set $i = i + 1$ and solve $\tilde{S}_n(\beta) = 0$ for $\tilde{\beta}_{(i)}$ using $\Gamma = (\hat{\Sigma}_{(i-1)})^{1/2}$ in equation (8).

Step 3. Using $\tilde{\beta}_{(i)}$, calculate $\tilde{A}_{(i)} = \tilde{A}_n(\tilde{\beta}_{(i)})$ and $\hat{\Omega}_{(i)} = \hat{\Omega}(\tilde{\beta}_{(i)})$.

Step 4. Compute $\hat{\Sigma}_{(i)} = \tilde{A}_{(i)}^{-1} \hat{\Omega}_{(i)} \tilde{A}_{(i)}^{-1}$.

Step 5. Repeat steps 2–4 until convergence of both $\tilde{\beta}_{(i)}$ and $\hat{\Sigma}_{(i)}$ is achieved to a specified tolerance.

In our experience, convergence of this algorithm typically occurs with relatively few iterations, the value of $\hat{\Sigma}_{(*)}$ at convergence being very close to $\hat{\Psi}_n$ in (11).

Remark 3. The above procedure makes use of a data-dependent smoothing parameter. The proofs of Theorems 1 and 2 assume that the matrix Γ is known; however, since $\|\hat{\Gamma}_{(*)} - (A^{-1}\Omega A^{-1})^{1/2}\| = O_p(n^{-1/2})$, replacing Γ by $\hat{\Gamma}_{(*)}$ does not alter these asymptotic results.

3.3. The resampling variance estimator

Jin et al. (2006b, § 4) propose a useful resampling method for estimating Ψ in the presence of correlated data. This method, which can be motivated by the conditional multiplier central limit theorem (e.g. Martinussen & Scheike, 2006, p. 43), involves randomly reweighting the Gehan log-rank objective function (6) and then minimizing the resulting perturbed objective function. Specifically, let

$$L_n^*(\beta) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{k=1}^{K_i} \sum_{j=1}^n \sum_{l=1}^{K_j} Z_i Z_j \Delta_{ik} \{e_{jl}(\beta) - e_{ik}(\beta)\} I\{e_{ik}(\beta) - e_{jl}(\beta) \leq 0\},$$

where Z_1, \dots, Z_n are independent positive random variables with $E(Z_i) = \text{var}(Z_i) = 1$ ($i = 1, \dots, n$). Let $\hat{\beta}_n^* = \text{argmin}_{\beta \in \mathbb{B}} L_n^*(\beta)$. Jin et al. (2006b, Theorem 5) prove that, conditional on the data $\{W_{ik}; k = 1, \dots, K_i; i = 1, \dots, n\}$, the limiting distribution of $n^{1/2}(\hat{\beta}_n^* - \hat{\beta}_n)$ converges almost surely to the limiting distribution of $n^{1/2}(\hat{\beta}_n - \beta_0)$. Thus, the distribution of $\hat{\beta}_n$ can be approximated by repeatedly generating random samples Z_1, \dots, Z_n and then minimizing $L_n^*(\beta)$ to obtain realizations of $\hat{\beta}_n^*$. The covariance matrix of $\hat{\beta}_n$ can be approximated directly by the empirical covariance matrix of the realizations of $\hat{\beta}_n^*$.

Jin et al. (2006b, § 4) work directly with the unsmoothed Gehan objective function and utilize linear programming methods in combination with resampling in order to obtain regression parameter and covariance matrix estimates. Specifically, linear programming is used to minimize $L_n(\beta)$, obtaining the estimated regression parameter $\hat{\beta}_n$; it is then applied repeatedly in minimizing each of the realizations of $L_n^*(\beta)$ generated for the purposes of covariance matrix estimation. The use of linear programming methods can be avoided by randomly reweighting the smoothed objective function $\tilde{L}_n(\beta)$ in (9). Such an approach allows for standard numerical methods to be used for minimization, resulting in the potential for computational savings with larger datasets.

With Z_1, \dots, Z_n defined as above, let

$$\tilde{L}_n^*(\beta) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{k=1}^{K_i} \sum_{j=1}^n \sum_{l=1}^{K_j} Z_i Z_j \Delta_{ik} \left[\{e_{jl}(\beta) - e_{ik}(\beta)\} H_{ikjl}^{(n)}(\beta) + \frac{r_{ikjl}}{n^{1/2}} h_{ikjl}^{(n)}(\beta) \right],$$

where $H_{ikjl}^{(n)}(\beta)$ and $h_{ikjl}^{(n)}(\beta)$ are defined in (10), and define $\tilde{\beta}_n^* = \operatorname{argmin}_{\beta \in \mathbb{B}} \tilde{L}_n^*(\beta)$. Theorems 1 and 2 imply that an argument identical to the one given in Jin et al. (2006b, Theorem 5) can be used to show that the limiting distribution of $n^{1/2}(\tilde{\beta}_n^* - \tilde{\beta}_n)$ converges almost surely to the limiting distribution of $n^{1/2}(\tilde{\beta}_n - \beta_0)$. The covariance matrix of $\tilde{\beta}_n$ can then be approximated exactly as described above, using the simulated realizations of $\tilde{\beta}_n^*$ in place of the $\hat{\beta}_n^*$ s.

4. SIMULATION STUDY

Simulation studies were carried out to assess the performance of $\tilde{\beta}_n$ as well as to evaluate the covariance matrix estimators described in § 3. The proposed simulation studies are modelled after that described in Jin et al. (2006b, § 5), allowing for a direct comparison between their simulation results and those to be summarized below.

Specifically, for each cluster, we use the algorithm of Johnson (1987, § 10·1) to generate two failure times from the bivariate Gumbel distribution

$$F(t_1, t_2) = F_1(t_1)F_2(t_2) [1 + \theta \{1 - F_1(t_1)\} \{1 - F_2(t_2)\}],$$

where $-1 \leq \theta \leq 1$, $F_k(\cdot)$ is the cumulative distribution function for an exponential random variable with hazard function $\lambda_k = \exp(\beta_1 X_{1k} + \beta_2 X_{2k})$, X_{1k} is $\operatorname{Ber}(0.5)$, and X_{2k} is standard normal truncated at ± 2 ($k = 1, 2$). All covariates are generated independently and the correlation between \bar{T}_1 and \bar{T}_2 is $\theta/4$. The resulting failure time model is a special case of the accelerated failure time model of § 2·1 with true regression parameters $\beta_1 = 1$ and $\beta_2 = 0.5$. Censoring times are independently generated from a $\operatorname{Un}(0, \tau)$ distribution, where τ is selected to achieve a desired level of censoring. Similarly to Jin et al. (2006b, § 5), we consider the cases $\theta = 0$ and $\theta = 1$, 50 clusters of size two, and censoring percentages of 0%, 25% and 50%.

Two different estimation methods are considered. Method 1 refers to the iterative method of § 3·2 for simultaneously estimating the regression parameters and covariance matrix. Method 2 refers to estimating the regression parameter by minimizing the smoothed objective function (9) with the fixed choice $\Sigma = I_2$. Within Method 2, we consider estimating the covariance matrix using the resampling-based variance estimate of § 3·3 and also using the sandwich variance estimator (11). All simulations were conducted in R and use the routine `nlm` for optimization (R Development Core Team, 2005); the simulation code is available upon request.

Table 1 summarizes the results of two simulation studies. Each row of the table is based on the same 1000 simulated datasets. In the first simulation study, the semiparametric accelerated failure time model of § 2·1 is fitted using the covariates X_{1k} and X_{2k} . We report the results for the estimation of the regression parameters $\beta_1 = 1$ and $\beta_2 = 0.5$ and associated standard errors using Methods 1 and 2. The second simulation study repeats the first simulation study, fitting a model that uses the covariates $X_{1k}^* = X_{1k}$ and $X_{2k}^* = X_{2k}/500$. The underlying failure time model is identical to that used in the first simulation study, the true regression parameters now being $\beta_1^* = \beta_1 = 1$, $\beta_2^* = 500$ and $\beta_2 = 250$. However, in contrast to the first simulation study, the magnitudes of X_{2k}^* and β_2^* are quite different from those of X_{1k}^* and β_1^* . The results for β_1^* , not shown, are very similar to those reported in Table 1 for β_1 ; hence, we only report the results for β_2^* . The intent of the second study is to investigate the impact of using the fixed smoothing

Table 1. Simulation results based on 1000 replications and $n = 50$ pairs for regression parameter and standard error estimates obtained using the induced smoothing methodology. The results for β_1 and β_2 are based on an accelerated failure time model that depends on covariates X_1 and X_2 ; the results for β_2^* are based on an accelerated failure time model that depends on covariates $X_1^* = X_1$ and $X_2^* = X_2/500$

Regression parameter	θ	Censoring %	Method 1			Method 2			
			RBIAS	RSE	RSEE1	RBIAS	RSE	RSEE2A	RSEE2B
$\beta_1 = 1$	0	0	0.32	0.25	0.25	0.21	0.25	0.25	0.25
		25	3.91	0.28	0.28	3.41	0.28	0.28	0.28
		50	4.30	0.35	0.36	2.22	0.34	0.36	0.35
	1	0	0.91	0.26	0.25	0.78	0.25	0.25	0.25
		25	1.58	0.27	0.28	1.12	0.27	0.27	0.27
		50	5.52	0.37	0.36	3.38	0.36	0.36	0.35
$\beta_2 = 0.5$	0	0	0.74	0.28	0.26	0.78	0.28	0.26	0.26
		25	1.89	0.30	0.30	1.58	0.30	0.30	0.30
		50	5.13	0.38	0.38	3.40	0.38	0.38	0.38
	1	0	1.23	0.26	0.26	1.20	0.26	0.26	0.26
		25	1.99	0.30	0.30	1.69	0.28	0.30	0.30
		50	4.40	0.40	0.38	2.70	0.38	0.38	0.38
$\beta_2^* = 250$	0	0	0.28	0.27	0.27	0.00	0.27	0.26	0.29
		25	2.99	0.29	0.30	1.96	0.29	0.29	0.35
		50	5.77	0.38	0.38	2.76	0.37	0.38	0.46
	1	0	5.80	0.26	0.27	2.40	0.26	0.26	0.30
		25	1.56	0.29	0.29	5.60	0.29	0.29	0.35
		50	5.25	0.37	0.38	2.20	0.36	0.37	0.46

RBIAS, $1000 \times$ absolute relative bias; RSE, empirical standard error, relative to parameter; RSEE1, standard error relative to parameter, with standard error estimate obtained using iterative method of § 3.2 with $\hat{\Sigma}_{(0)} = I_2$; RSEE2A, standard error relative to parameter, with standard error estimate obtained using the resampling procedure of § 3.3 with 500 random reweightings and regression parameters estimated using the induced smoothing procedure of § 2.3 with the fixed choice $\Sigma = I_2$; RSEE2B, standard error relative to parameter, with standard error estimate based on (11) and regression parameters estimated as described for RSEE2A.

parameter $\Sigma = I_2$ versus the data-dependent smoothing parameter of § 3.2, a choice that better reflects the covariance structure and scaling of the regression parameter.

Considering only β_1 and β_2 , the relative biases are observed to be small, comparable in magnitude, and generally increase with the censoring percentage. In addition, estimates obtained using Method 1 frequently exhibit greater bias than those obtained using Method 2, with no apparent reduction in standard error. The standard error estimates for β_1 and β_2 are accurate and similar across all estimation methods. Remarkably, the results reported here are also comparable to those summarized in the right panel of Table 1 in Jin et al. (2006b) for the Gehan weight function, where no smoothing is employed.

Turning to the comparison of results for β_2 and β_2^* , biases generally follow the patterns described above. In addition, all methods of standard error estimation perform well, though some evidence of inflation in the relative standard error RSEE2B is now present. Overall, the results suggest that the smoothing parameter has a minimal impact on the bias or actual standard error of the regression parameter estimates. However, given the relative accuracy of both RSEE1 and RSEE2A, the discrepancy observed in RSEE2B suggests that the scaling of the problem, hence choice of smoothing parameter, can adversely impact the accuracy of (11).

On the basis of these results, we recommend using Method 1 as described in § 3.2; in comparison with the simulation-based methodology of Jin et al. (2006b), it requires far less computational effort with no evidence of penalty in bias or accuracy of standard error estimation.

5. CONCLUDING REMARKS AND FURTHER EXTENSIONS

The attractive nature of the induced smoothing procedure, in both computational and theoretical terms, stems largely from the convexity of the Gehan-weighted objective function (9). The asymptotic results obtained in this paper make significant use of this convexity. A minor extension of these results can also be used to justify an alternative smoothing methodology for the bounded influence estimator introduced in Heller (2007). Variations on this smoothing methodology may facilitate simpler and more stable estimation procedures for accelerated failure time frailty models; see, for example, Pan (2001); Strawderman (2006) and Zhang & Peng (2007).

The use of the Gehan weight function in (2) has frequently been criticized for the inefficiency of the resulting estimator. The selection of an alternative weight function may result in efficiency improvements at the expense of monotonicity, resulting in weaker asymptotic statements and increased computational challenges. To counteract these drawbacks, Jin et al. (2003) propose to use the Gehan estimator as a starting point for successively solving a sequence of convex optimization problems derived from (1). Jin et al. (2006b) extend these results to the setting of multivariate failure time data. The resulting class of estimation procedures is computationally stable and yields a consistent and asymptotically normal sequence of estimators with reasonably general weight functions. However, it does not lend itself to a simple method of variance estimation. Use of the resampling method described in § 3.3 is recommended for this purpose but only amplifies the required computational effort. Jin et al. (2006a) propose a strongly related class of procedures for the Buckley–James estimator. Starting from the Gehan estimator, Strawderman (2005) demonstrates how one may instead use one-step estimation to achieve the same goal and introduces an alternative simulation-based method of variance computation that requires no additional optimization. The results of this paper show that the induced smoothing methodology provides an asymptotically valid and computationally convenient starting point for each of these other methods of estimation. In addition, the methodology itself can be directly incorporated as part of the iterative methods developed in Jin et al. (2003, 2006a, b); the asymptotic results of this paper guarantee that their results also remain valid for the corresponding smoothed version.

A direct extension of this smoothing methodology is available for general weight functions. However, it lacks the same computational convenience due to important structural differences between the Gehan-weighted estimating equation and those used for general weight functions.

ACKNOWLEDGEMENT

The authors thank the referees, associate editor and the editor for helpful comments. This work was supported by a grant from the U.S. National Institutes of Health.

APPENDIX

Proofs

We impose the following regularity conditions:

1. Condition A1. The parameter space \mathbb{B} containing β_0 is a compact subset of \mathbb{R}^p .
2. Condition A2. $\sum_{k=1}^{K_i} \|X_{ik}\| + K_i$ is bounded almost surely by a nonrandom constant ($i = 1, \dots, n$).

3. Condition A3. The assumptions of § 2.1 hold with $\text{var}(\epsilon_{11}) < \infty$.
4. Condition A4. The matrix $A = \nabla S_0(\beta_0)$, where $S_0(\beta) = \lim_{n \rightarrow \infty} S_n(\beta)$, exists and is nonsingular.
5. Condition A5. Let $f_0(\cdot)$ denote the marginal density associated with model error term ϵ_{11} and let $\lambda_0(\cdot)$ denote its corresponding hazard function. Then, $f_0(\cdot)$ and $f'_0(\cdot)$ are bounded functions on \mathbb{R} with

$$\int_{\mathbb{R}} \left\{ \frac{f'_0(t)}{f_0(t)} \right\}^2 f_0(t) dt < \infty.$$

6. Condition A6. The marginal distribution of C_{rs} is absolutely continuous and has a bounded density $g_{rs}(\cdot)$ on \mathbb{R} ($r = 1, \dots, n; s = 1, \dots, K_r$).

As indicated in the statements of Theorems 1 and 2, $\Sigma = \Gamma^2$ is assumed to be a symmetric and positive definite matrix with $\|\Gamma\| < \infty$. Conditions A1, A2, A4, A5 and A6 are standard and ensure consistency and asymptotic normality of the unsmoothed Gehan estimator (Tsiatis, 1990; Ying, 1993; Jin et al., 2006b, for example). Condition A3 implies $|\text{cov}(\epsilon_{ik}, \epsilon_{il})| \leq \text{var}(\epsilon_{11})$ ($i = 1, \dots, n; k, l = 1, \dots, K_i$); hence, the covariances between all error terms within a cluster are bounded.

The proof of Theorem 1 relies on the following pair of lemmas, both of which hold under conditions A1–A3. The proof of Lemma 1 is a direct consequence of the strong law of large numbers for U-statistics and results in Andersen & Gill (1982, Theorem II.1). The proof of Lemma 2 relies on this result and properties of the normal cumulative distribution and density functions. These proofs are available in a technical report.

LEMMA 1. $\sup_{\beta \in \mathbb{B}} |L_n(\beta) - L_0(\beta)| \rightarrow 0$ almost surely, where $L_0(\beta)$ is convex for $\beta \in \mathbb{B}$.

LEMMA 2. $\sup_{\beta \in \mathbb{B}} |\tilde{L}_n(\beta) - L_0(\beta)| \rightarrow 0$ almost surely, where $L_0(\cdot)$ is defined in Lemma 1.

Proof of Theorem 1. Lemmas 1 and 2, respectively, establish the uniform almost sure convergence of $L_n(\beta)$ and $\tilde{L}_n(\beta)$ to the convex function $L_0(\beta)$ for $\beta \in \mathbb{B}$. By condition A4, $L_0(\beta)$ is strictly convex at β_0 and β_0 is a unique minimizer. The respective minimizers $\hat{\beta}_n$ and $\tilde{\beta}_n$ of $L_n(\beta)$ and $\tilde{L}_n(\beta)$ thus converge almost surely to β_0 (Andersen & Gill, 1982, Corollary II.2). \square

The next lemma is required in order to prove Theorem 2; an abbreviated proof of this result and also Theorem 2 are provided below, with expanded versions of these arguments available in a technical report. A fact used in proving Lemma 3 is that condition A4, in conjunction with (A1), implies that the probability that $X_{1k} \neq X_{2l}$ for at least one (k, l) pair must be positive.

LEMMA 3. Under A1–A6 and as $n \rightarrow \infty$, $\|\nabla \tilde{S}_n(\beta_0) - A\| \rightarrow 0$ almost surely, where $A = \nabla S_0(\beta_0)$,

$$\nabla S_0(\beta_0) = \frac{1}{2} \sum_{k=1}^{K_1} \sum_{l=1}^{K_2} \int_{-\infty}^{\infty} E\{(X_{1k} - X_{2l})(X_{1k} - X_{2l})' \bar{G}_{1k}(u) \bar{G}_{2l}(u)\} \{f_0^2(u) + f'_0(u) \bar{F}_0(u)\} du, \quad (\text{A1})$$

$\bar{G}_{rs}(\cdot)$ denotes the survivor function of $\log C_{rs} - X'_{rs} \beta_0$ and $\bar{F}_0(s) = \int_s^{\infty} f_0(u) du$ for every s .

Proof of Lemma 3. Using calculations analogous to those found in Fyngenson & Ritov (1994), it can be shown that $E\{S_n(\beta)\} = S_0(\beta) + O(n^{-1})$, where the $O(\cdot)$ term holds uniformly on $\beta \in \mathbb{B}$ and

$$S_0(\beta) = E \left[\sum_{k=1}^{K_1} \sum_{l=1}^{K_2} \{M_{1k,2l}(\beta) - M_{2l,1k}(\beta)\} \right] \quad (\text{A2})$$

for

$$M_{ab,cd}(\beta) = \int_{-\infty}^{\infty} \bar{G}_{ab}(u) \bar{G}_{cd}\{u + (\beta_0 - \beta)'(X_{ab} - X_{cd})\} \bar{F}_0\{u + (\beta_0 - \beta)'(X_{ab} - X_{cd})\} f_0(u) du.$$

The outer expectation in (A2) is understood to be taken over the joint distribution of the covariates. Evidently, $S_0(\beta_0) = 0$. Conditions A1–A6 permit us to differentiate (A2) directly; doing so and evaluating the result at $\beta = \beta_0$, we obtain (A1) (Fyngenson & Ritov, 1994, p. 737).

Recalling notation from § 2.1, let the survival data for cluster i be denoted $\mathcal{W}_i = \{W_{ik}, k = 1, \dots, K_i\}$. The smoothed equation (8) may then be written as

$$\tilde{S}_n(\beta) = \frac{1}{c_n} \frac{1}{n^2} \sum_{i=1}^n \tilde{h}_\beta(\mathcal{W}_i, \mathcal{W}_i) + \frac{1}{\binom{n}{2}} \sum_{i < j} \tilde{\psi}_\beta(\mathcal{W}_i, \mathcal{W}_j),$$

where $c_n = 1 - n^{-1}$, $\tilde{\psi}_\beta(\mathcal{W}_i, \mathcal{W}_j) = (1/2)\{\tilde{h}_\beta(\mathcal{W}_i, \mathcal{W}_j) + \tilde{h}_\beta(\mathcal{W}_j, \mathcal{W}_i)\}$ and

$$\tilde{h}_\beta(\mathcal{W}_i, \mathcal{W}_j) = \sum_{k=1}^{K_i} \sum_{l=1}^{K_j} \Delta_{ik}(X_{ik} - X_{jl}) \Phi \left[n^{1/2} \left\{ \frac{e_{jl}(\beta) - e_{ik}(\beta)}{r_{ikjl}} \right\} \right].$$

Differentiating this representation with respect to β , setting $\beta = \beta_0$, and then using both the strong law of large numbers for independent observations and for U-statistics, we find

$$\nabla \tilde{S}_n(\beta_0) \rightarrow \frac{1}{2} \sum_{k=1}^{K_1} \sum_{l=1}^{K_2} E\{(X_{1k} - X_{2l})(X_{1k} - X_{2l})' (\mathcal{A}_{1k,2l} + \mathcal{A}_{2l,1k})\} \tag{A3}$$

almost surely as $n \rightarrow \infty$. The random variable $\mathcal{A}_{ab,cd}$ is defined to be zero with probability one if $X_{ab} = X_{cd}$; otherwise, one may write

$$\mathcal{A}_{ab,cd} = \int_{-\infty}^{\infty} \bar{G}_{ab}(u) f_0(u) \xi_{cd}(u) du + \lim_{n \rightarrow \infty} \left(\frac{m_n}{\pi} \right)^{1/2} \int_{-\infty}^{\infty} \tau(w) e^{-m_n w^2} dw,$$

where $m_n = n/(2r_{abcd}^2)$, $r_{abcd}^2 > 0$, $\tau(w) = \int_{-\infty}^{\infty} \bar{G}_{ab}(u) f_0(u) \{\xi_{cd}(w+u) - \xi_{cd}(u)\} du$ and $\xi_{cd}(s) = f_0(s) \bar{G}_{cd}(s) + \bar{F}_0(s) g_{cd}(s)$. Under conditions A1–A6, $\tau(\cdot)$ is integrable, continuous and bounded on \mathbb{R} with $\tau(0) = 0$ and the second term on the right-hand side therefore vanishes (e.g. Kanwal, 1998, p. 11). Using the resulting formula for $\mathcal{A}_{ab,cd}$ and integration by parts, it can now be shown that

$$\mathcal{A}_{1k,2l} + \mathcal{A}_{2l,1k} = \int_{-\infty}^{\infty} \bar{G}_{1k}(u) \bar{G}_{2l}(u) \{f_0^2(u) + f_0'(u) \bar{F}_0(u)\} du.$$

Substituting this last result into (A3), we observe agreement with (A1), proving the result. □

Proof of Theorem 2. Using notation introduced in § 2.2, we have that $A^{-1} n^{1/2} S_n(\beta_0)$ is asymptotically normal with mean zero and variance $A^{-1} \Omega A^{-1}$ under assumptions A1–A5 (Jin et al., 2006b, Theorem 5). Suppose that

$$n^{1/2}(\tilde{\beta}_n - \beta_0) + A^{-1} n^{1/2} S_n(\beta_0) \rightarrow 0 \tag{A4}$$

in probability. Then, $n^{1/2}(\tilde{\beta}_n - \beta_0) \rightarrow N(0, A^{-1} \Omega A^{-1})$ in distribution, establishing the desired asymptotic result as well as the equality of the limiting distributions of the smoothed and unsmoothed estimators.

To prove that (A4) holds, we can make use of Theorem 3 of Arcones (1998). Using notation from Arcones (1998), define $G_n(\beta) = n \tilde{L}_n(\beta)$ for all β in \mathbb{B} . For $n \geq 1$, define the sequence of $p \times 1$ random vectors $\eta_n = n^{1/2} S_n(\beta_0)$ and sequences of nonsingular, symmetric $p \times p$ matrices $M_n = n^{1/2} I_p$ and $V_n = (1/2)A$. The required result (A4) becomes

$$M_n(\tilde{\beta}_n - \beta_0) + \frac{1}{2} V_n^{-1} \eta_n \rightarrow 0. \tag{A5}$$

in probability. The result (A5) follows directly from Arcones (1998, Theorem 3) provided that conditions A1–A6 are sufficient to ensure that the following regularity conditions hold:

1. Condition B1. $G_n(\beta)$ is convex and $\tilde{\beta}_n$ is a sequence satisfying $G_n(\tilde{\beta}_n) \leq \inf_{\beta \in \mathbb{B}} G_n(\beta) + o_p(1)$.
2. Condition B2. $\eta_n = O_p(1)$, $\liminf_{n \rightarrow \infty} \inf_{|\beta|=1} \beta' V_n \beta > 0$ and $\limsup_{n \rightarrow \infty} \sup_{|\beta|=1} \beta' V_n \beta < \infty$.
3. Condition B3. For each $\beta \in \mathbb{R}^p$, $G_n(\beta_0 + M_n^{-1} \beta) - G_n(\beta_0) - \beta' \eta_n - \beta' V_n \beta = o_p(1)$.

Conditions B1 and B2 are immediate consequences of conditions A1–A6. For example, condition B1 follows because $G_n(\beta)$ is easily shown to be convex, $\tilde{\beta}_n = \operatorname{argmin}_{\beta \in \mathbb{B}} \tilde{L}_n(\beta) = \operatorname{argmin}_{\beta \in \mathbb{B}} G_n(\beta)$, $G_n(\beta)$ is continuous, and \mathbb{B} is compact. In addition, conditions A1–A5 are sufficient to ensure that $\eta_n = n^{1/2} S_n(\beta_0)$ converges in distribution, hence $O_p(1)$ as required. Since $V_n = (1/2)A$ is a positive definite matrix for every n , condition B2 is also satisfied.

It remains to establish condition B3. Using the definitions of M_n , $G_n(\cdot)$, $\tilde{L}_n(\cdot)$ and $\tilde{S}_n(\cdot)$, and a Taylor series expansion, we may write

$$G_n(\beta_0 + M_n^{-1}\beta) - G_n(\beta_0) - \beta' \{n^{1/2} \tilde{S}_n(\beta_0)\} - \frac{1}{2} \beta' \{\nabla \tilde{S}_n(\beta_n^*)\} \beta = o_p(1), \tag{A6}$$

where $\|\beta_n^* - \beta_0\| \leq \|M_n^{-1}\beta\|$. The triangle inequality, Lemma 3, and the fact that $\{\tilde{S}_n\}$ form a sequence of bounded, continuously differentiable functions implies that we can replace $\nabla \tilde{S}_n(\beta_n^*)$ in (A6) by the matrix A without altering this result. Therefore, if

$$n^{1/2} \|\tilde{S}_n(\beta_0) - S_n(\beta_0)\| \rightarrow 0 \tag{A7}$$

in probability, the definitions of V_n and η_n imply that condition B3 holds. To see that (A7) holds, write

$$\tilde{S}_n(\beta_0) - S_n(\beta_0) = \int_{\mathbb{R}^p} \{S_n(\beta_0 + n^{-1/2}u) - S_n(\beta_0)\} \psi(u) du,$$

where $\psi(\cdot)$ denotes the pdf of ΓZ .

Let Θ be a fixed matrix such that $\|\Theta\| \leq M$ for some $M < \infty$ and define for suitable u the function $K_n(u; \beta_0, \Theta) = \|S_n(\beta_0 + n^{-1/2}u) - S_n(\beta_0) - n^{-1/2}\Theta u\|$. Then, since $\int_{\mathbb{R}^p} u \psi(u) du = 0$, the triangle inequality implies

$$\|\tilde{S}_n(\beta_0) - S_n(\beta_0)\| \leq \int_{\|u\| \leq \epsilon_n} K_n(u; \beta_0, \Theta) \psi(u) du + \int_{\|u\| > \epsilon_n} K_n(u; \beta_0, \Theta) \psi(u) du \tag{A8}$$

for any $\epsilon_n > 0$. The result (A7) therefore holds if we can find $\epsilon_n > 0$ such that both integrals on the right-hand side of (A8) converge in probability to zero.

Following Ying (1993, Theorem 2) and Jin et al. (2006b, Theorem 5), the matrix A satisfies

$$\sup_{\|b - \beta_0\| \leq d_n} \frac{\|S_n(b) - S_n(\beta_0) - A(b - \beta_0)\|}{1 + n^{1/2}\|b - \beta_0\|} = o_p(n^{-1/2}) \tag{A9}$$

for any positive sequence $d_n \rightarrow 0$. Suppose $\epsilon_n = o(n^{1/2})$. Then, taking $b = \beta_0 + n^{-1/2}u$, $d_n = n^{-1/2}\epsilon_n$ and $\Theta = A$, (A9) implies

$$\sup_{u \leq \epsilon_n} \frac{\|K_n(u; \beta_0, A)\|}{1 + \|u\|} = o_p(n^{-1/2}). \tag{A10}$$

An easy calculation, in combination with (A10), now shows that the first integral on the right-hand side of the inequality in (A8) converges in probability to zero, even if $\epsilon_n \rightarrow \infty$. With regard to the second term on the right-hand side of (A8), we may use the definition of $K_n(\cdot; \beta_0, A)$ and the triangle inequality to write $n^{1/2} \int_{\|u\| > \epsilon_n} K_n(u; \beta_0, A) \psi(u) du \leq Q_3 + Q_4$, where

$$Q_3 = \left\{ \sup_{\|u\| > \epsilon_n} \|S_n(\beta_0 + n^{-1/2}u) - S_n(\beta_0)\| \right\} n^{1/2} \int_{\|u\| > \epsilon_n} \psi(u) du, \quad Q_4 = \|A\| \int_{\|u\| > \epsilon_n} \|u\| \psi(u) du.$$

For all $\beta \in \mathbb{B}$, $\|S_n(\beta)\| \leq Q$ for some constant $Q < \infty$ by condition A2; hence, $Q_3 \leq 2Qn^{1/2}p(\|\Gamma Z\| > \epsilon_n)$. Letting $\epsilon_n \rightarrow \infty$, it follows that $n^{1/2}p(\|\Gamma Z\| > \epsilon_n) \rightarrow 0$ as $n \rightarrow \infty$. Similarly, $\int_{\|u\| > \epsilon_n} \|u\| \psi(u) du \rightarrow 0$. Thus, provided that $n, \epsilon_n \rightarrow \infty$, the bounds Q_3 and Q_4 , hence, the second integral on the right-hand side of the inequality in (A8), also converge in probability to zero. Since we can select a sequence $\epsilon_n = o(n^{1/2})$ such that both $n, \epsilon_n \rightarrow \infty$, it follows that (A8) converges in probability to zero as $n \rightarrow \infty$, establishing (A7) and concluding the proof. \square

REFERENCES

- ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–20.
- ARCONES, M. A. (1998). Asymptotic theory for M -estimators over a convex kernel. *Economet. Theory* **14**, 387–422.
- BROWN, B. M. & WANG, Y.-G. (2005). Standard errors and covariance matrices for smoothed rank estimators. *Biometrika* **92**, 149–58.
- BROWN, B. M. & WANG, Y.-G. (2006). Induced smoothing for rank regression with censored survival times. *Statist. Med.* **26**, 828–36.
- COX, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc. B* **34**, 187–220.
- FYGENSON, M. & RITOV, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.* **22**, 732–46.
- HELLER, G. (2007). Smoothed rank regression with censored data. *J. Am. Statist. Assoc.* **102**, 552–59.
- HUANG, Y. (2002). Calibration regression of censored lifetime medical cost. *J. Am. Statist. Assoc.* **97**, 318–27.
- JIN, Z., LIN, D. Y., WEI, L. J. & YING, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–53.
- JIN, Z., LIN, D. Y. & YING, Z. (2006a). On least squares regression with censored data. *Biometrika* **93**, 147–62.
- JIN, Z., LIN, D. Y. & YING, Z. (2006b). Rank regression analysis of multivariate failure time data based on marginal linear models. *Scand. J. Statist.* **33**, 1–23.
- JOHNSON, M. E. (1987). *Multivariate Statistical Simulation*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.
- JONES, M. P. (1997). A class of semiparametric regressions for the accelerated failure time model. *Biometrika* **84**, 73–84.
- KALBFLEISCH, J. D. & PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd Ed. Hoboken, New Jersey: Wiley-Interscience.
- KANWAL, R. P. (1998). *Generalized Functions: Theory and Technique*. Boston: Birkhäuser.
- LEE, E. W., WEI, L. J. & YING, Z. (1993). Linear regression analysis for highly stratified failure time data. *J. Am. Statist. Assoc.* **88**, 557–65.
- LIN, J. S. & WEI, L. J. (1992). Linear regression analysis for multivariate failure time observations. *J. Am. Statist. Assoc.* **87**, 1091–97.
- MARTINUSSEN, T. & SCHEIKE, T. H. (2006). *Dynamic Regression Models for Survival Data*. *Statistics for Biology and Health*. New York: Springer.
- PAN, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Anal.* **7**, 55–64.
- R DEVELOPMENT CORE TEAM (2005). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- SONG, X., MA, S., HUANG, J. & ZHOU, X. (2007). A semiparametric approach for the nonparametric transformation survival model with multiple covariates. *Biostatistics* **6**, 197–211.
- STRAWDERMAN, R. L. (2005). The accelerated gap times model. *Biometrika* **92**, 647–66.
- STRAWDERMAN, R. L. (2006). A regression model for dependent gap times. *Int. J. Biostatistics* **2**, Iss. 1, Article 1.
- TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–72.
- YING, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76–99.
- ZHANG, J. & PENG, Y. (2007). An alternative estimation method for the accelerated failure time frailty model. *Comp. Statist. Data Anal.* **51**, 4413–23.

[Received July 2008. Revised December 2008]