Pseudo full likelihood estimation for prospective survival analysis with a general semiparametric shared frailty model: asymptotic theory

David M. Zucker[1]

Department of Statistics, Hebrew University, Mt. Scopus, Jerusalem 91905, Israel

mszucker@mscc.huji.ac.il

Malka Gorfine

Faculty of Industrial Engineering and Management, Technion, Technion City, Haifa 32000, Israel, and *Department of Mathematics, Bar-Ilan University, Ramat-Gan, 52900, Israel*

gorfinm@ie.technion.ac.il

Li Hsu

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024, USA

lih@fhcrc.org

August 29, 2007

[1]To whom correspondence should be addressed. Phone: +972-2-588-1291. Fax: +972-2-588-3549.

Abstract

In this work we present a simple estimation procedure for a general frailty model for analysis of prospective correlated failure times. Earlier work showed this method to perform well in a simulation study. Here we provide rigorous large-sample theory for the proposed estimators of both the regression coefficient vector and the dependence parameter, including consistent variance estimators.

Key words: Correlated failure times; EM algorithm; Frailty model; Prospective family study; Survival analysis.

1 Introduction

Many epidemiological studies involve failure times that are clustered into groups, such as families or schools. Unobserved characteristics shared by members of the same cluster (e.g. genetic information or unmeasured shared environmental exposures) could influence time to the studied event. Frailty models express within-cluster dependence through a shared unobservable random effect. Estimation in the frailty model has received much attention under various frailty distributions, including gamma (Gill, 1985, 1989; Nielsen et al., 1992; Klein 1992, among others), positive stable (Hougaard, 1986; Fine et al., 2003), inverse Gaussian, compound Poisson (Henderson and Oman, 1999) and log-normal (McGilchrist, 1993; Ripatti and Palmgren, 2000; Vaida and Xu, 2000, among others). Hougaard (2000) provides a comprehensive review of the properties of the various frailty distributions. In a frailty model, the parameters of interest typically are the regression coefficients, the cumulative baseline hazard function, and the dependence parameters in the random effect distribution.

Since the frailties are latent covariates, the Expectation-Maximization (EM) algorithm is a natural estimation tool, with the latent covariates estimated in the E-step and the likelihood maximized in the M-step after substituting in the estimated latent quantities. Gill (1985), Nielsen et al. (1992) and Klein (1992) discussed EM-based maximum likelihood estimation for the semiparametric gamma frailty model. One problem with the EM algorithm is that variance estimates for the estimated parameters are not readily available (Louis, 1982; Gill, 1989; Nielsen et al., 1992; Andersen et al., 1997). It has been suggested (Gill, 1989; Nielsen et al, 1992) that a nonparametric information calculation could yield consistent variance estimators. Parner (1998), building on Murphy (1994, 1995), proved the consistency and asymptotic normality of the maximum likelihood estimator in the gamma frailty model. Parner also presented a consistent estimator of the

limiting covariance matrix of the estimator, based on inverting a discrete observed information matrix. He noted that since the dimension of the observed information matrix grows with the number of observed survival times, inverting the matrix is practically infeasible for a large data set with many distinct failure times. He therefore suggested an alternate approach to estimating the covariance, based on solving a discrete version of a second order Sturm-Liouville equation, along the lines of Bickel (1985). This covariance estimator requires less computational effort, but still is not so simple to implement.

We (Gorfine et al., 2006) developed a new method that can handle any parametric frailty distribution with finite moments. Nonconjugate frailty distributions can be handled by a simple univariate numerical integration over the frailty distribution. Our new method possesses a number of desirable properties: a non-iterative procedure for estimating the cumulative hazard function; consistency and asymptotic normality of the parameter estimates; a direct consistent covariance estimator; and easy computation and implementation. The method was found to perform well in a simulation study and the results are very similar to those of the EM-based method. Indeed, on a dataset-by-dataset basis, the correlation between our estimator and the EM estimator was found to be 95% for the covariate regression parameter and 98-99% for the within-cluster dependence parameter. The purpose of the current paper is to present in detail the theoretical justification for the method.

Our technical approach resembles that of Bagdonavicius and Nikulin (1999) and Dabrowska (2006a, 2006b). These works, however, dealt with a univariate data context, whereas we deal with a clustered data context. Dabrowska works with a transformation model with unknown transformation. She discusses the univariate gamma frailty model, but assumes that the shape parameter of the frailty distribution is known. Indeed, as discussed in Dabrowska (2006a, pp. 147-148), identifiability problems arise in the univariate

gamma frailty model with unknown shape parameter when an unknown transformation is involved. In fact, even when the transformation is known, if there are no covariate effects on the hazard rate (i.e., in the model (1) below, the regression parameter vector $\boldsymbol{\beta}$ is equal to zero), the shape parameter cannot be identified from univariate data (Lancaster and Nickell, 1980). In our setting, there is no unknown transformation, and we have clustered data. In this case, the shape parameter is identifiable irrespective of whether $\boldsymbol{\beta}$ is zero or nonzero. In our work, we are specifically interested in estimating the shape parameter, which expresses the within-cluster dependence. In genetic research and other contexts, this cluster dependence parameter is itself of significant scientific interest, because it provides insight into the impact of genetic and environmental factors on the disease incidence. Dabrowska (2006b) discusses a one-step method for converting a consistent estimator into a semiparametric efficient estimator. In principle, this approach could be applied to our estimator as well. In our simulations, however, we found that our estimator was comparable in efficiency to the full nonparametric MLE. Thus, although our estimator is not theoretically semiparametric efficient, in practical terms it closely approaches semiparametric efficiency.

The plan of the paper is as follows. Section 2 presents the estimation procedure. Section 3 presents the consistency and asymptotic normality results, along with the covariance estimator for the parameter estimates. Section 4 presents a simulation study. Section 5 presents the technical conditions required for our theoretical results and the proofs of these results. The proofs are patterned after Zucker (2005), but with a number of significant differences, which are described at the beginning of Section 5.

2 The Proposed Approach

Consider n families, with family i containing m_i members, $i = 1, \ldots, n$. Following Parner (1998, p. 187), we regard m_i as a random variable over $\{1, \ldots, m\}$ for some m, and build up the remainder of the model conditional on m_i. Let T_{ij}^0 and C_{ij} denote the failure and censoring times, respectively, for individual ij. The observed follow-up time is $T_{ij} = \min(T_{ij}^0, C_{ij})$, and the failure indicator is $\delta_{ij} = I(T_{ij}^0 \leq C_{ij})$. On each individual, we observe a p-vector of covariates $\mathbf{Z}_{ij}$. In addition, we associate with family i an unobservable family-level covariate W_i, the "frailty", which induces dependence among family members. The conditional hazard function for individual ij, given the family frailty W_i, is taken to be

$$\lambda_{ij}(t) = W_i \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) \qquad i = 1, \ldots, n \quad j = 1, \ldots, m_i \tag{1}$$

where λ_0 is an unspecified conditional baseline hazard and $\boldsymbol{\beta}$ is a p-vector of unknown regression coefficients. This is an extension of the Cox (1972) proportional hazards model, with the hazard function for an individual in family i multiplied by W_i. Conditional on W_i, the individuals within a family are assumed independent. We also assume that, given $\mathbf{Z}_{ij}$ and W_i, the censoring is independent and noninformative for W_i and $(\boldsymbol{\beta}, \Lambda_0)$ (Andersen et al., 1993, Sec. III.2.3). We assume further that the frailty W_i is independent of $\mathbf{Z}_{ij}$ and has a density $f(w; \theta)$, where θ is an unknown parameter. For simplicity we assume that θ is a scalar, but the development extends readily to the case where θ is a vector. Finally, we assume that for any given family, there is a positive probability of at least two failures. This condition is necessary to ensure identifiability of the model; see Nielsen et al. (1992, Sec. 4, end).

Let τ be the end of the observation period. The full likelihood of the data then can

be written as

$$\begin{aligned} L = & \prod_{i=1}^{n} \int \prod_{j=1}^{m_i} \{\lambda_{ij}(T_{ij})\}^{\delta_{ij}} S_{ij}(T_{ij}) f(w) dw \\ = & \prod_{i=1}^{n} \prod_{j=1}^{m_i} \{\lambda_0(T_{ij}) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\}^{\delta_{ij}} \prod_{i=1}^{n} \int w^{N_{i.}(\tau)} \exp\{-wH_{i.}(\tau)\} f(w) dw, \end{aligned} \tag{2}$$

where $N_{ij}(t) = \delta_{ij} I(T_{ij} \leq t)$, $N_{i.}(t) = \sum_{j=1}^{m_i} N_{ij}(t)$, $H_{ij}(t) = \Lambda_0(T_{ij} \wedge t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})$, $a \wedge b = \min\{a, b\}$, $\Lambda_0(\cdot)$ is the baseline cumulative hazard function, $S_{ij}(\cdot)$ is the conditional survival function of subject ij, and $H_{i.}(t) = \sum_{j=1}^{m_i} H_{ij}(t)$. The log-likelihood is given by

$$l = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \delta_{ij} \log\{\lambda_0(T_{ij}) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\} + \sum_{i=1}^{n} \log \left\{ \int w^{N_{i.}(\tau)} \exp\{-wH_{i.}(\tau)\} f(w) dw \right\}.$$

The normalized scores (log-likelihood derivatives) for $(\beta_1, \ldots, \beta_p)$ are given by

$$U_r = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \delta_{ij} Z_{ijr} - \frac{1}{n} \sum_{i=1}^{n} \frac{\left[\sum_{j=1}^{m_i} H_{ij}(T_{ij}) Z_{ijr}\right] \int w^{N_{i.}(\tau)+1} \exp\{-wH_{i.}(\tau)\} f(w) dw}{\int w^{N_{i.}(\tau)} \exp\{-wH_{i.}(\tau)\} f(w) dw} \tag{3}$$

for $r = 1, \ldots, p$. The normalized score for θ is

$$U_{p+1} = \frac{1}{n} \sum_{i=1}^{n} \frac{\int w^{N_{i.}(\tau)} \exp\{-wH_{i.}(\tau)\} f'(w) dw}{\int w^{N_{i.}(\tau)} \exp\{-wH_{i.}(\tau)\} f(w) dw}$$

where $f'(w) = \frac{d}{d\theta} f(w)$. Let $\boldsymbol{\gamma} = (\boldsymbol{\beta}^T, \theta)$ and $\mathbf{U}(\boldsymbol{\gamma}, \Lambda_0) = (U_1, \ldots, U_p, U_{p+1})^T$. To obtain estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$, we propose to substitute an estimator of Λ_0, denoted by $\hat{\Lambda}_0$, into the equations $\mathbf{U}(\boldsymbol{\gamma}, \Lambda_0) = 0$.

Let $Y_{ij}(t) = I(T_{ij} \geq t)$ and let $\mathcal{F}_t$ denote the entire observed history up to time t, that is

$$\mathcal{F}_t = \sigma\{N_{ij}(u), Y_{ij}(u), \mathbf{Z}_{ij}, i = 1, \ldots, n; j = 1, \ldots, m_i; 0 \leq u \leq t\}.$$

Then, as discussed by Gill (1992) and Parner (1998), the stochastic intensity process for $N_{ij}(t)$ with respect to $\mathcal{F}_t$ is given by

$$\lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) Y_{ij}(t) \psi_i(\boldsymbol{\gamma}, \Lambda_0, t-), \tag{4}$$

where

$$\psi_i(\boldsymbol{\gamma}, \Lambda_0, t) = \mathrm{E}(W_i|\mathcal{F}_t).$$

Using a Bayes theorem argument and the joint density (2) with observation time restricted to $[0, t)$, we obtain

$$\psi_i(\boldsymbol{\gamma}, \Lambda, t) = \phi_{2i}(\boldsymbol{\gamma}, \Lambda, t)/\phi_{1i}(\boldsymbol{\gamma}, \Lambda, t),$$

where

$$\phi_{ki}(\boldsymbol{\gamma}, \Lambda_0, t) = \int w^{N_{i\cdot}(t)+(k-1)} \exp\{-wH_{i\cdot}(t)\} f(w) dw, \quad k = 1, \ldots, 4.$$

Given the intensity model (4), in which $\exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\psi_i(\boldsymbol{\gamma}, \Lambda_0, t-)$ may be regarded as a time dependent covariate effect, a natural estimator of Λ_0 is a Breslow (1974) type estimator along the lines of Zucker (2005). For given values of $\boldsymbol{\beta}$ and θ we estimate Λ_0 as a step function with jumps at the observed failure times τ_k, $k = 1, \ldots, K$, with

$$\Delta\hat{\Lambda}_0(\tau_k) = \frac{d_k}{\sum_{i=1}^{n} \psi_i(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1}) \sum_{j=1}^{m_i} Y_{ij}(\tau_k) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})} \tag{5}$$

where d_k is the number of failures at time τ_k. Note that given the intensity model (4), the estimator of the kth jump depends on $\hat{\Lambda}_0$ up to and including time τ_{k-1}. By this approach, we avoid complicating the iterative optimization process with a further iterative scheme for estimating the cumulative hazard. This feature, however, does not necessarily translate into a computational advantage relative to the EM-method, because $\psi_i(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})$ has to be computed at every jump. Bagdonavicius and Nikulin (1999) proposed a similar estimator in a univariate survival context, for a model which they called the "generalized proportional hazards model," which includes univariate frailty-type models.

3 Asymptotic Properties

Let $\boldsymbol{\gamma}^\circ = (\boldsymbol{\beta}^{\circ T}, \theta^\circ)^T$ with $\boldsymbol{\beta}^\circ$, θ° and $\Lambda_0^\circ(t)$ denoting the respective true values of $\boldsymbol{\beta}$, θ and $\Lambda_0(t)$, and let $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\beta}}^T, \hat{\theta})^T$. We assume the technical conditions listed in Section 4.1.

In Section 4.3, we establish the following results.

A. $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ converges almost surely to $\Lambda_0(t, \boldsymbol{\gamma})$ uniformly in t and $\boldsymbol{\gamma}$.

B. $\mathbf{U}(\boldsymbol{\gamma}, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}))$ converges almost surely uniformly in t and $\boldsymbol{\gamma}$ to a limit $\mathbf{u}(\boldsymbol{\gamma}, \Lambda_0(\cdot, \boldsymbol{\gamma}))$.

C. There exists a unique consistent root to $\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) = \mathbf{0}$.

In Section 4.4, we show that $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)$ is asymptotically normally distributed. We accomplish this by analyzing in turn each of the terms in the following decomposition:

$$\begin{aligned} \mathbf{0} = & \ \mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) \\ = & \ \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) + [\mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)] \\ & + [\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ))]. \end{aligned} \tag{6}$$

We show further that the covariance matrix of $\hat{\boldsymbol{\gamma}}$ can be consistently estimated by a sandwich estimator of the following form:

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\gamma}}) = \mathbf{D}^{-1}(\hat{\boldsymbol{\gamma}})\{\hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}}) + \hat{\mathbf{G}}(\hat{\boldsymbol{\gamma}}) + \hat{\mathbf{C}}(\hat{\boldsymbol{\gamma}})\}\mathbf{D}^{-1}(\hat{\boldsymbol{\gamma}})^T. \tag{7}$$

The matrix $\mathbf{D}$ consists of the derivatives of the U_r's with respect to the parameters $\boldsymbol{\gamma}$. $\mathbf{V}$ is the asymptotic covariance matrix of $\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)$, $\mathbf{G}$ is the asymptotic covariance matrix of $[\mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)]$, and $\mathbf{C}$ is the asymptotic covariance matrix between $\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)$ and $[\mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)]$. The term $\mathbf{G} + \mathbf{C}$ reflects the added variance resulting from the need to estimate the cumulative hazard function. All these matrices are defined explicitly in Section 4.4.

4 Simulation Study for the Gamma Frailty Case

In Gorfine et al. (2006), we presented a simulation study comparing our method to the EM method under the gamma frailty distribution with expectation 1 and variance θ. Here

we extend the simulation study by considering larger θ values and family sizes larger than two.

Gorfine et al. describes in detail the steps of the EM-based algorithm, as given in Nielsen et al. (1992), in parallel with the corresponding steps in our procedure. We refer the reader to Gorfine et al. for details.

The setup for the simulation study, which is patterned after Hsu et al. (2004), is as follows. We worked with a sample size of 300 families, with a common family size of 2 or 5. We generated for each family a common frailty value W from the gamma distribution with scale and shape parameters both equal to θ^{-1}, and for each individual a single covariate Z from the standard normal distribution. Conditional on W, the survivor function was taken to be

$$S(t|Z,W) = \exp\{-W\exp(\beta Z)(0.01t)^{4.6}\}.$$

We took the censoring distribution to be $N(60, 15^2)$. The β values examined were $\beta^\circ = ln(2)$ and $\beta^\circ = ln(3)$, leading to censoring levels of approximately 85% and 80%, respectively. The censoring distribution was chosen in order to generate an appropriate mean age at onset and age-of-onset distribution, similar to what is often observed for late onset diseases. With censoring distributed according to $N(130, 15^2)$ the respective censoring levels are approximately 35% and 30%. The θ values examined were $\theta^\circ = 2$ and $\theta^\circ = 4$.

Tables 1-2 present the simulation results for the two estimation techniques, based on 1,000 replications. For our method, we compare the mean estimated standard error based on our theoretical formula with the empirical standard error, and provide the empirical coverage rate of the 95% Wald-type confidence interval. For the EM-based method, we report only the empirical standard error. In addition, the empirical correlation between the EM-based estimators and our estimators is presented. The additional simulation

results confirm our earlier findings. Both estimation techniques perform very well in terms of bias. Also, for our method, fairly good agreement was observed between the estimated and the empirical standard error, although some differences were seen in some cases. The high values of the correlations implies similarity between the two estimation techniques not only on an average basis, but actually on a data set by data set basis.

5 Technical Conditions and Proofs

5.1 Introductory Remarks

This section presents the technical conditions we assume for the asymptotic results and the proofs of these results. Some details have been omitted for the sake of brevity. These details are provided in an expanded version of this paper which is available at the *Front for the Mathematics ArXiv* under **Statistics**, publication number: math.ST/0602253.

The general pattern of the argument follows that of Zucker (2005), but with some significant changes. Our estimator for the cumulative hazard is based on the formula

$$\hat{\Lambda}_0(t) = \int_0^t \frac{n^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} dN_{ij}(s)}{n^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} \psi_i(\boldsymbol{\gamma}, \hat{\Lambda}_0, s-) Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})} \ .$$

The quantity $\psi_i(\boldsymbol{\gamma}, \hat{\Lambda}_0, s-)$ involves terms of the form $\hat{\Lambda}_0(s - \wedge T_{ij})$, i.e. it involves $\hat{\Lambda}_0$ values at T_{ij} as well as at $s-$. By contrast, the corresponding integrand in Zucker's (2005) estimator involves only $\hat{\Lambda}(s-)$. (The estimator of Bagdonavicius and Nikulin (1999) is similar to that of Zucker (2005) in this respect.) This difference in the structure of the estimators entails the need for substantial extensions to the argument.

In particular, Zucker's consistency proof for the cumulative hazard estimate makes use of a result of the form $\sup_{\boldsymbol{\beta},t,c} |A_0(\boldsymbol{\beta}, t, c) - a_0(\boldsymbol{\beta}, t, c)| \to 0$ a.s., where $A_0(\boldsymbol{\beta}, t, c)$ is a certain empirical process, $a_0(\boldsymbol{\beta}, t, c)$ is its expectation, and the supremum is over $\boldsymbol{\beta} \in \mathcal{B}, t \in [0, \tau]$, and $c \in [0, \Lambda_{max}]$. In our consistency proof, we need the more complex result given in (20)

below, whose proof requires a sophisticated argument. In the asymptotic normality proof, a number of extra steps are required, relative to Zucker's proof, to deal with the middle term in the decomposition (6). In particular, we need to introduce the decomposition of $\hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t)$ given in (25) below, and the interchange of integrals that is carried out right after introducing this decomposition. Furthermore, unlike in Zucker (2005), the first two terms in the decomposition (6) are correlated, so that extra development is needed to deal with the correlation (Step III of the asymptotic normality proof). The structure of the derivative matrix of the score function vector is more complex than in Zucker (2005). Finally, in contrast with Zucker (2005), we use mainly the classical CLT for sums of iid's rather than the martingale CLT.

We take note here that since $\boldsymbol{\beta}$ and $\mathbf{Z}_{ij}$ are bounded, there exists a constant $\nu > 0$ such that

$$\nu^{-1} \leq \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) \leq \nu. \tag{8}$$

This fact is used repeatedly in our proofs.

We also introduce here some basic definitions. Recall that

$$\psi_i(\boldsymbol{\gamma}, \Lambda, t) = \frac{\int w^{N_i(t)+1} e^{-H_{i\cdot}(t)w} f(w) dw}{\int w^{N_i(t)} e^{-H_{i\cdot}(t)w} f(w) dw},$$

with $H_{i\cdot}(t) = H_{i\cdot}(t, \boldsymbol{\gamma}, \Lambda) = \sum_{j=1}^{m_i} \Lambda(T_{ij} \wedge t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})$ (here we define $H_{i\cdot}$ so as to allow dependence on a general $\boldsymbol{\gamma}$ and Λ, which will often not be explicitly indicated in the notation). We define (for $0 \leq r \leq m$ and $h \geq 0$)

$$\psi^*(r, h) = \frac{\int w^{r+1} e^{-hw} f(w) dw}{\int w^r e^{-hw} f(w) dw}. \tag{9}$$

We further define $\psi^*_{min}(h) = \min_{0 \leq r \leq m} \psi^*(r, h)$ and $\psi^*_{max}(h) = \max_{0 \leq r \leq m} \psi^*(r, h)$. In (9), the numerator and denominator are bounded above since W is assumed to have finite $(m+2)$-th moment. Also, since W is nondegenerate, the numerator and denominator are

strictly positive. Thus $\psi^*_{max}(h)$ is finite and $\psi^*_{min}(h)$ is strictly positive. The following result can be proved by elementary calculus (details in the expanded version).

Lemma 1: The function $\psi^*(r,h)$ is decreasing in h. Hence for all $\boldsymbol{\gamma} \in \mathcal{G}$ and all t,

$$\psi_i(\boldsymbol{\gamma}, \Lambda, t) \leq \psi^*_{max}(0), \tag{10}$$

$$\psi_i(\boldsymbol{\gamma}, \Lambda, t) \geq \psi^*_{min}(m\nu\Lambda(t)). \tag{11}$$

5.2 Technical Conditions

In deriving the asymptotic properties of $\hat{\boldsymbol{\gamma}}$ we make the following assumptions:

1. The random vectors $(m_i, T^0_{i1}, \ldots, T^0_{im_i}, C_{i1}, \ldots, C_{im_i}, \mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{im_i}, W_i)$, $i = 1, \ldots, n$, are independent and identically distributed.

2. There is a finite maximum follow-up time $\tau > 0$, with $\mathrm{E}[\sum_{j=1}^{m_i} Y_{ij}(\tau)] = y^* > 0$ for all i.

3. (a) Conditional on $\mathbf{Z}_{ij}$ and W_i, the censoring is independent and noninformative of W_i and $(\boldsymbol{\beta}, \Lambda_0)$.

   (b) W_i is independent of $\mathbf{Z}_{ij}$ and of m_i.

4. The frailty random variable W_i has finite moments up to order $(m+2)$.

5. $\mathbf{Z}_{ij}$ is bounded.

6. The parameter $\boldsymbol{\gamma}$ lies in a compact subset $\mathcal{G}$ of $\mathbb{R}^{p+1}$ containing an open neighborhood of $\boldsymbol{\gamma}^\circ$.

7. There exist $B > 0$ and $\bar{h} > 0$ (independent of θ) such that, for all $h \geq \bar{h}$, we have $\psi^*_{min}(h) \geq Bh^{-1}$.

8. The baseline hazard function $\lambda_0^\circ(t)$ is bounded over $[0, \tau]$ by some fixed (but not necessarily known) constant λ_{max}.

9. The function $f'(w;\theta) = (d/d\theta)f(w;\theta)$ is absolutely integrable.

10. The censoring distribution has at most finitely many jumps on $[0, \tau]$.

11. For any given family, there is a positive probability of at least two failures.

12. The matrix $[(\partial/\partial\boldsymbol{\gamma})\mathbf{U}(\boldsymbol{\gamma}, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}))]|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^\circ}$ is invertible with probability going to 1 as $n \to \infty$.

In regard to Assumption 7, the assumption is satisfied if *either one* of the following two conditions holds.

a. There exist $b(\theta) > 0$ and $C(\theta) > 0$ such that

$$\sup_\theta \left| \frac{f(w;\theta)}{C(\theta)w^{(b(\theta)-1)}} - 1 \right| \to 0 \quad \text{as } h \to 0,$$

with $b(\theta)$ bounded from below over θ.

b. We have $\lim_{w\to 0} \sup_\theta f(w;\theta) = 0$, and there exists $a > 0$ independent of θ such that $f(w;\theta)$ is increasing in w over $w \in [0, a]$.

These conditions cover a wide range of frailty distributions, including popular choices such as the gamma, inverse Gaussian, and lognormal.

5.3 Preliminary Lemmas

Lemma 2: Define $\bar{\Lambda} = 1.03e^{m\sigma}\bar{h}/(m\nu)$, with $\sigma = 1.01m\nu^2/(By^*)$, with $\bar{h}$ and B as above. Then, with probability one, there exists n' such that, for all $t \in [0, \tau]$ and $\boldsymbol{\gamma} \in \mathcal{G}$,

$$\hat{\Lambda}_0(t, \boldsymbol{\gamma}) \leq \bar{\Lambda} \quad \text{for } n \geq n'. \tag{12}$$

Thus, $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ is naturally bounded, with no need to impose an upper bound artificially.

Proof: To simplify the writing below, we will suppress the argument $\boldsymbol{\gamma}$ in $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$. Recall

$$\Delta\hat{\Lambda}_0(\tau_k) = \left[\sum_{i=1}^{n} \psi_i(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1}) \sum_{j=1}^{m_i} Y_{ij}(\tau_k) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})\right]^{-1},$$

where we now take $d_k = 1$ since the survival time distribution is assumed continuous. Using Lemma 1 and (8), we have

$$\Delta\hat{\Lambda}_0(\tau_k) \leq n^{-1}\nu\psi^*_{min}(m\nu\hat{\Lambda}(\tau_{k-1}))^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} Y_{ij}(\tau)\right]^{-1}.$$

By the strong law of large numbers, there exists with probability one some n^* such that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} Y_{ij}(\tau) \geq 0.999y^* \quad \text{for } n \geq n^*. \tag{13}$$

We thus have, for $n \geq n^*$,

$$\Delta\hat{\Lambda}_0(\tau_k) \leq n^{-1}\left(\frac{1.01\nu}{y^*}\right)\psi^*_{min}(m\nu\hat{\Lambda}(\tau_{k-1}))^{-1}. \tag{14}$$

Given this result, the desired conclusion can be obtained via simple technical manipulations, detailed in the expanded version of this paper.

Lemma 3: We have $\sup_{s\in[0,\tau]} |\hat{\Lambda}_0(s, \boldsymbol{\gamma}^\circ) - \hat{\Lambda}_0(s-, \boldsymbol{\gamma}^\circ)| \to 0$ as $n \to \infty$, as an immediate consequence of Lemma 2 and (14).

5.4 Consistency

We now show the almost sure consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0$. The argument is built on Claims A-C of Section 3, which we prove below. Our argument follows Zucker (2005, Appendix A.3).

Claim A: $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ converges a.s. to some function $\Lambda_0(t, \boldsymbol{\gamma})$ uniformly in t and $\boldsymbol{\gamma}$.

Proof: Whenever a functional norm is written below, the relevant uniform norm is intended. We define $\Lambda_{max} = \max(\bar{\Lambda}, \lambda_{max}\tau)$, $h_{max} = m\nu\Lambda_{max}$, and $\psi^{**}(r, h) = \psi^*(r, h \wedge$

h_{max}). It is easy to see that $\psi^{**}(r,h)$ is Lipschitz continuous in h, uniformly in r. Recall that $\psi_i(\boldsymbol{\gamma},\Lambda,t) = \psi^*(N_i(t), H_{i\cdot}(t,\boldsymbol{\gamma},\Lambda))$. Lemma 2 implies that $H_{i\cdot}(t,\boldsymbol{\gamma},\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma})) \leq h_{max}$ for all $t \in [0,\tau]$ and $\boldsymbol{\gamma} \in \mathcal{G}$. Hence $\psi_i(\boldsymbol{\gamma},\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma}),t) = \psi^{**}(N_i(t), H_{i\cdot}(t,\boldsymbol{\gamma},\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma})))$.

Now define, for a general function Λ,

$$\Xi_n(t,\boldsymbol{\gamma},\Lambda) = \int_0^t \frac{n^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} dN_{ij}(s)}{n^{-1}\sum_{i=1}^n \sum_{j=1}^{m_i} \psi^{**}(N_i(s-), H_{i\cdot}(s-,\boldsymbol{\gamma},\Lambda))Y_{ij}(s)\exp(\boldsymbol{\beta}^T\mathbf{Z}_{ij})}$$

and

$$\Xi(t,\boldsymbol{\gamma},\Lambda) = \int_0^t \frac{\mathrm{E}[\sum_{j=1}^{m_i} \psi^*(N_i(s-), H_{i\cdot}(s-,\boldsymbol{\gamma}^\circ,\Lambda_0^\circ))Y_{ij}(s)\exp(\boldsymbol{\beta}^{\circ T}\mathbf{Z}_{ij})]}{\mathrm{E}[\sum_{j=1}^{m_i} \psi^{**}(N_i(s-), H_{i\cdot}(s-,\boldsymbol{\gamma},\Lambda))Y_{ij}(s-)\exp(\boldsymbol{\beta}^T\mathbf{Z}_{ij})]}\lambda_0^\circ(s)ds.$$

By definition, $\hat{\Lambda}_0(t,\boldsymbol{\gamma})$ satisfies the equation

$$\hat{\Lambda}_0(t,\boldsymbol{\gamma}) = \Xi_n(t,\boldsymbol{\gamma},\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma})). \tag{15}$$

Next, define

$$q_{\boldsymbol{\gamma}}(s,\Lambda) = \frac{\mathrm{E}[\sum_{j=1}^{m_i} \psi^*(N_i(s-), H_{i\cdot}(s-,\boldsymbol{\gamma}^\circ,\Lambda_0^\circ))Y_{ij}(s)\exp(\boldsymbol{\beta}^{\circ T}\mathbf{Z}_{ij})]}{\mathrm{E}[\sum_{j=1}^{m_i} \psi^{**}(N_i(s-), H_{i\cdot}(s-,\boldsymbol{\gamma},\Lambda))Y_{ij}(s)\exp(\boldsymbol{\beta}^T\mathbf{Z}_{ij})]}\lambda_0^\circ(s).$$

This function is uniformly bounded by $B^* = [\psi^*_{max}(0)/\psi^*_{min}(h_{max})]\lambda_{max}$. Moreover, by the Lipschitz continuity of $\psi^{**}(r,h)$ with respect to h, it satisfies a Lipschitz-like condition of the form $|q_{\boldsymbol{\gamma}}(s,\Lambda_1) - q_{\boldsymbol{\gamma}}(s,\Lambda_2)| \leq K \sup_{0\leq u\leq s}|\Lambda_1(u) - \Lambda_2(u)|$. Hence, by mimicking the argument of Hartman (1973, Theorem 1.1), we find that the equation $\Lambda(t) = \Xi(t,\boldsymbol{\gamma},\Lambda)$ has a unique solution, which we denote by $\Lambda_0(t,\boldsymbol{\gamma})$. The claim then is that $\hat{\Lambda}_0(t,\boldsymbol{\gamma})$ converges almost surely (uniformly in t and $\boldsymbol{\gamma}$) to $\Lambda_0(t,\boldsymbol{\gamma})$.

Define $\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma})$ to be a modified version of $\hat{\Lambda}_0(t,\boldsymbol{\gamma})$ defined by linear interpolation between the jumps. Lemma 3 implies that, with probability one,

$$\sup_{t,\boldsymbol{\gamma}} |\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma}) - \hat{\Lambda}_0(t,\boldsymbol{\gamma})| \to 0, \tag{16}$$

and thus

$$\sup_{t,\boldsymbol{\gamma}} |\Xi_n(t,\boldsymbol{\gamma},\tilde{\Lambda}_0(t,\boldsymbol{\gamma})) - \Xi_n(t,\boldsymbol{\gamma},\hat{\Lambda}_0(t,\boldsymbol{\gamma}))| \to 0. \tag{17}$$

Lemma 2 shows that the family $\mathcal{L} = \{\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\gamma}),\ n \geq n'\}$ is uniformly bounded. We can show further that $\mathcal{L}$ is equicontinuous, using arguments similar to those of Zucker (2005). The first step is to note that, with $\bar{N}(t) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} N_{ij}(t)$, we have $\bar{N}(t) \to \mathrm{E}[N_i(t)]$ as $n \to \infty$ uniformly in t with probability one. From this we can obtain the following result: with probability one, for any $\epsilon > 0$ there exists $n''(\epsilon)$ such that for all t and u with $u < t$,

$$\hat{\Lambda}_0(t, \boldsymbol{\gamma}) - \hat{\Lambda}_0(u, \boldsymbol{\gamma}) \leq B^*(t-u) + \frac{\epsilon}{2} \quad \text{for all } n \geq n''(\epsilon). \tag{18}$$

Moreover, $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ is Lipschitz continuous in $\boldsymbol{\gamma}$, uniformly in $\boldsymbol{\gamma}$ and t. The equicontinuity follows. Given that $\mathcal{L}$ is a.s. uniformly bounded and equicontinuous, the Arzela-Ascoli theorem implies that it is (almost surely) a relatively compact set in $C([0, \tau] \times \mathcal{G})$.

Next, define

$$\begin{aligned} A(\boldsymbol{\gamma}, \Lambda, s) &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \psi^{**}(N_i(s-), H_{i\cdot}(s-, \boldsymbol{\gamma}, \Lambda)) Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}), \\ a(\boldsymbol{\gamma}, \Lambda, s) &= \mathrm{E}\left[\sum_{j=1}^{m_i} \psi^{**}(N_i(s-), H_{i\cdot}(s-, \boldsymbol{\gamma}, \Lambda)) Y_{ij}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) \right]. \end{aligned}$$

For any fixed continuous Λ, the functional strong law of large numbers of Andersen and Gill (1982, Appendix III) implies that

$$\sup_{s, \boldsymbol{\gamma}} |A(\boldsymbol{\gamma}, \Lambda, s) - a(\boldsymbol{\gamma}, \Lambda, s)| \to 0 \quad \text{a.s.} \tag{19}$$

Here we need the following more complex result:

$$\sup_{s, \boldsymbol{\gamma}} |A(\boldsymbol{\gamma}, \tilde{\Lambda}^{(n)}, s) - a(\boldsymbol{\gamma}, \tilde{\Lambda}^{(n)}, s)| \to 0 \quad \text{a.s.} \tag{20}$$

The proof of (20) is involved; we give the details in Section 4.5 below. In outline form, the proof involves two steps: (1) showing that, for any given $\epsilon > 0$, we can define an appropriate *finite* class $\mathcal{L}_\epsilon^*$ of functions Λ such that $\tilde{\Lambda}^{(n)}$ can be suitably approximated by some member of the class; (2) applying the result (19), which will hold uniformly over the finite class.

Given (20) and the a.s. uniform convergence of $\bar{N}(t)$ to $\mathrm{E}[N_i(t)]$, we can infer that

$$\sup_{t,\boldsymbol{\gamma}} |\Xi_n(t,\boldsymbol{\gamma},\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma})) - \Xi(t,\boldsymbol{\gamma},\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma}))| \to 0 \quad \text{a.s.} \tag{21}$$

This result is obtained by adapting the argument of Aalen (1976, Lemma 6.1).

From (15), (16), (17), and (21) it follows that any limit point of $\{\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma})\}$ must satisfy the equation $\Lambda = \Xi(t,\boldsymbol{\gamma},\Lambda)$. Since $\Lambda_0(t,\boldsymbol{\gamma})$ is the unique solution of this equation, it is the unique limit point of $\{\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma})\}$. Thus $\{\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma})\}$ is a sequence in a compact set with unique limit point $\Lambda_0(t,\boldsymbol{\gamma})$. Hence $\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma})$ converges a.s. uniformly in t and $\boldsymbol{\gamma}$ to $\Lambda_0(t,\boldsymbol{\gamma})$. In view of (16), the same holds of $\hat{\Lambda}_0(t,\boldsymbol{\gamma})$, which is the desired result. Note that $\Lambda_0(\cdot,\boldsymbol{\gamma}^\circ) = \Lambda_0^\circ(\cdot)$. Indeed, if we plug Λ_0° into the expression for $\Xi(t,\boldsymbol{\gamma}^\circ,\Lambda)$, the expectation terms cancel, and so we are left with the integral of $\lambda_0^\circ(s)$. Thus, Λ_0° is the solution to the equation $\Lambda = \Xi(t,\boldsymbol{\gamma}^\circ,\Lambda)$.

Claim B: With $\mathbf{u}(\boldsymbol{\gamma},\Lambda_0(\cdot,\boldsymbol{\gamma})) = \mathrm{E}[\mathbf{U}(\boldsymbol{\gamma},\Lambda_0(\cdot,\boldsymbol{\gamma}))]$, we have $\mathbf{U}(\boldsymbol{\gamma},\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma})) \to \mathbf{u}(\boldsymbol{\gamma},\Lambda_0(\cdot,\boldsymbol{\gamma}))$ uniformly in $\boldsymbol{\gamma} \in \mathcal{G}$ with probability one.

Proof: As in Zucker (2005).

Claim C: There exists a unique consistent root to $\mathbf{U}(\hat{\boldsymbol{\gamma}},\hat{\Lambda}_0(\cdot,\hat{\boldsymbol{\gamma}})) = \mathbf{0}$.

Proof: We apply Foutz's (1977) consistency theorem for maximum likelihood type estimators. The following conditions must be established:

F1. $\partial\mathbf{U}(\boldsymbol{\gamma},\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma}))/\partial\boldsymbol{\gamma}$ exists and is continuous in an open neighborhood about $\boldsymbol{\gamma}^\circ$.

F2. The convergence of $\partial\mathbf{U}(\boldsymbol{\gamma},\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma}))/\partial\boldsymbol{\gamma}$ to its limit is uniform in open neighborhood of $\boldsymbol{\gamma}^\circ$.

F3. $\mathbf{U}(\boldsymbol{\gamma}^\circ,\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma}^\circ)) \to \mathbf{0}$ as $n \to \infty$.

F4. The matrix $-[\partial\mathbf{U}(\boldsymbol{\gamma},\hat{\Lambda}_0(\cdot,\boldsymbol{\gamma}))/\partial\boldsymbol{\gamma}]|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^\circ}$ is invertible with probability going to 1 as

$n \to \infty$. (In Foutz's paper, the matrix in question is symmetric, and so he stated the condition in terms of positive definiteness. But his proof, which is based on the inverse function theorem, shows that the basic condition needed is invertibility.)

It is easily seen that Condition F1 holds. Given Assumptions 2, 4, and 5, Condition F2 follows from the previously-cited functional law of large numbers. As for Condition F3, Claim B says that $\mathbf{U}(\boldsymbol{\gamma}, \Lambda_0(\cdot, \boldsymbol{\gamma}))$ converges a.s. uniformly to $\mathbf{u}(\boldsymbol{\gamma}, \Lambda_0(\cdot, \boldsymbol{\gamma})) = \mathrm{E}[\mathbf{U}(\boldsymbol{\gamma}, \Lambda_0(\cdot, \boldsymbol{\gamma}))]$. We noted already that $\Lambda_0(\cdot, \boldsymbol{\gamma}^\circ) = \Lambda_0(\cdot)$. Thus we need only show that $\mathrm{E}[\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0)] = \mathbf{0}$. Since $\mathbf{U}$ is a score function derived from a classical iid likelihood, this result follows from classical likelihood theory. Condition F4 has been assumed in Assumption 12. With Conditions F1-F4 established, the result follows.

5.5 Asymptotic Normality

To show that $\hat{\boldsymbol{\gamma}}$ is asymptotically normally distributed, we write

$$\begin{aligned}
\mathbf{0} = \;& \mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) \\
= \;& \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) + [\mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)] \\
& + [\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ))]
\end{aligned}$$

In the following we consider each of the terms of the right-hand side of the equation.

Step I

We can write $\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) = n^{-1} \sum_{i=1}^n \boldsymbol{\xi}_i$, where $\boldsymbol{\xi}_i$ is a $(p+1)$-vector with r-th element, $r = 1, \ldots, p$, given by

$$\xi_{ir} = \sum_{j=1}^{m_i} \delta_{ij} Z_{ijr} - \frac{\left[\sum_{j=1}^{m_i} H_{ij}(\tau) Z_{ijr}\right] \int w^{N_{i\cdot}(\tau)+1} \exp\{-w\{H_{i\cdot}(\tau)\} f(w;\theta) dw}{\int w^{N_{i\cdot}(\tau)} \exp\{-w H_{i\cdot}(\tau)\} f(w;\theta) dw}$$

and $(p+1)$-th element given by

$$\xi_{i(p+1)} = \frac{\int w^{N_{i\cdot}(\tau)} \exp\{-w H_{i\cdot}(\tau)\} f'(w;\theta) dw}{\int w^{N_{i\cdot}(\tau)} \exp\{-w H_{i\cdot}(\tau)\} f(w;\theta) dw}.$$

Thus $\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)$ is the mean of the iid mean-zero random vectors $\boldsymbol{\xi}_i$. It hence follows from the central limit theorem that $n^{\frac{1}{2}}\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)$ is asymptotically mean-zero multivariate normal. To estimate the covariance matrix, let $\boldsymbol{\xi}_i^*$ be the counterpart of $\boldsymbol{\xi}_i$ with estimates of $\boldsymbol{\gamma}$ and Λ_0 substituted for the true values. Then an empirical estimator of the covariance matrix is given by $\hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}}) = n^{-1}\sum_{i=1}^n \boldsymbol{\xi}_i^* \boldsymbol{\xi}_i^{*T}$. This is a consistent estimator of the covariance matrix since $\hat{\Lambda}_0(t, \boldsymbol{\gamma})$ converges to $\Lambda_0(t, \boldsymbol{\gamma})$ a.s. uniformly in t and $\boldsymbol{\gamma}$ (Claim A), and $\hat{\boldsymbol{\gamma}}$ is a consistent estimator of $\boldsymbol{\gamma}^\circ$ (Claim C).

Step II

Let $\hat{U}_r = U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0)$, $r = 1, \ldots, p$, and $\hat{U}_{p+1} = U_{p+1}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0)$ (in this segment of the proof, when we write $(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0)$ the intent is to signify $(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ))$. First order Taylor expansion of $\hat{U}_r$ about Λ_0°, $r = 1, \ldots, p+1$, gives

$$
\begin{aligned}
& n^{1/2}\{U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)\} \\
& \quad = n^{-1/2}\sum_{i=1}^n \sum_{j=1}^{m_i} Q_{ijr}(\boldsymbol{\gamma}^\circ, \Lambda^\circ, T_{ij})\{\hat{\Lambda}_0(T_{ij}, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(T_{ij})\} + o_p(1), \qquad (22)
\end{aligned}
$$

where

$$
\begin{aligned}
Q_{ijr}(\boldsymbol{\gamma}^\circ, \Lambda^\circ, T_{ij}) \;=\; & -\left\{ \frac{\phi_{2i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)}{\phi_{1i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)} R_{ij}^* Z_{ijr} - \frac{\phi_{3i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)}{\phi_{1i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)} R_{ij}^* \sum_{j=1}^{m_i} H_{ij}(T_{ij}) Z_{ijr} \right. \\
& \left. + \frac{\phi_{2i}^2(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)}{\phi_{1i}^2(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)} R_{ij}^* \sum_{j=1}^{m_i} H_{ij}(T_{ij}) Z_{ijr} \right\}
\end{aligned}
$$

for $r = 1, \ldots, p$, and

$$
Q_{ij(p+1)}(\boldsymbol{\gamma}^\circ, \Lambda^\circ, T_{ij}) = R_{ij}^* \left\{ \frac{\phi_{2i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)\phi_{1i}^{(\theta)}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)}{\phi_{1i}^2(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)} - \frac{\phi_{2i}^{(\theta)}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)}{\phi_{1i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, \tau)} \right\},
$$

with $R_{ij}^* = \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij})$ and

$$
\phi_{ki}^{(\theta)}(\boldsymbol{\gamma}, \Lambda_0, t) = \int w^{N_{i\cdot}(t)+(k-1)} \exp\{-wH_{i\cdot}(t)\} f'(w) dw, \quad k = 1, 2.
$$

The validity of the approximation (22) can be seen by an argument similar to that used in connection with (24) below.

Given the intensity process (4), the process

$$M_{ij}(t) = N_{ij}(t) - \int_0^t \lambda_0(u) \exp(\boldsymbol{\beta}^{\circ T}\mathbf{Z}_{ij}) Y_{ij}(u) \psi_i(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, u-) du$$

is a mean zero martingale with respect to the filtration $\mathcal{F}_t$. Also, by Lemma 3, we have that $\sup_{s\in[0,\tau]} |\hat{\Lambda}_0(s, \boldsymbol{\gamma}^\circ) - \hat{\Lambda}_0(s-, \boldsymbol{\gamma}^\circ)|$ converges to zero. Thus, replacing $s-$ by s we obtain the following approximation, uniformly over $t \in [0, \tau]$:

$$\begin{aligned} \hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t) &\approx \frac{1}{n} \int_0^t \{\mathcal{Y}(s, \Lambda_0^\circ)\}^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} dM_{ij}(s) \\ &+ \frac{1}{n} \int_0^t \left[\{\mathcal{Y}(s, \hat{\Lambda}_0)\}^{-1} - \{\mathcal{Y}(s, \Lambda_0^\circ)\}^{-1}\right] \sum_{i=1}^n \sum_{j=1}^{m_i} dN_{ij}(s), \end{aligned} \tag{23}$$

where

$$\mathcal{Y}(s, \Lambda) = \frac{1}{n} \sum_{i=1}^n \psi_i(\boldsymbol{\gamma}^\circ, \Lambda, s) \sum_{j=1}^{m_i} Y_{ij}(s) \exp(\boldsymbol{\beta}^{\circ T}\mathbf{Z}_{ij}).$$

Now let $\mathcal{W}(s, r) = \{\mathcal{Y}(s, \Lambda_0^\circ + r\Delta)\}^{-1}$ with $\Delta = \hat{\Lambda}_0 - \Lambda_0^\circ$. Define $\dot{\mathcal{W}}$ and $\ddot{\mathcal{W}}$ as the first and second derivative of $\mathcal{W}$ with respect to r, respectively. Then, computing the necessary derivatives and carrying out a first order Taylor expansion of $\mathcal{W}(s, r)$ around $r = 0$ evaluated at $r = 1$ with Lagrange remainder (Abramowitz and Stegun, 1972, p. 880), we get

$$\begin{aligned} &\{\mathcal{Y}(s, \hat{\Lambda}_0)\}^{-1} - \{\mathcal{Y}(s, \Lambda_0^\circ)\}^{-1} = \dot{\mathcal{W}}(s, 0) + \frac{1}{2}\ddot{\mathcal{W}}(s, \tilde{r}(s)) \\ &= -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} \left[\frac{R_{i\cdot}(s)\eta_{1i}(0, s)}{\{\mathcal{Y}(s, \Lambda_0^\circ)\}^2} - \frac{1}{2} h_i(\tilde{r}(s), s) \right] \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) \{\hat{\Lambda}_0(T_{ij} \wedge s) - \Lambda_0^\circ(T_{ij} \wedge s)\}, \end{aligned} \tag{24}$$

where $R_{ij}(u) = \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) Y_{ij}(u)$, $R_{i\cdot}(u) = \sum_{j=1}^{m_i} R_{ij}(u)$, $\tilde{r}(s) \in [0, 1]$,

$$\eta_{1i}(r, s) = \frac{\phi_{3i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ + r\Delta, s)}{\phi_{1i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ + r\Delta, s)} - \left\{ \frac{\phi_{2i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ + r\Delta, s)}{\phi_{1i}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ + r\Delta, s)} \right\}^2,$$

and $h_i(r, s)$ is as defined in Section 4.6 below, and shown there to be $o(1)$ uniformly in r and s.

Let $\eta_{1i}(s) = \eta_{1i}(0, s)$. Plugging (24) into (23) we get

$$\begin{aligned}
\hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t) \approx\ & n^{-1} \int_0^t \{\mathcal{Y}(s, \Lambda_0^\circ)\}^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} dM_{ij}(s) \\
& - n^{-2} \int_0^t \sum_{k=1}^n \sum_{l=1}^{m_k} \frac{I(T_{kl} > s) R_{k\cdot}(s) \eta_{1k}(s)}{\{\mathcal{Y}(s, \Lambda_0^\circ)\}^2} \exp(\boldsymbol{\beta}^T \mathbf{Z}_{kl}) \{\hat{\Lambda}_0(s) - \Lambda_0^\circ(s)\} \sum_{i=1}^n \sum_{j=1}^{m_i} dN_{ij}(s) \\
& - n^{-2} \int_0^t \sum_{k=1}^n \sum_{l=1}^{m_k} \frac{I(T_{kl} \le s) R_{k\cdot}(s) \eta_{1k}(s)}{\{\mathcal{Y}(s, \Lambda_0^\circ)\}^2} \exp(\boldsymbol{\beta}^T \mathbf{Z}_{kl}) \{\hat{\Lambda}_0(T_{kl}) - \Lambda_0^\circ(T_{kl})\} \sum_{i=1}^n \sum_{j=1}^{m_i} dN_{ij}(s) \\
& + n^{-2} \int_0^t \sum_{k=1}^n \sum_{l=1}^{m_k} \frac{1}{2} h_k(\tilde{r}(s), s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{kl}) \{\hat{\Lambda}_0(T_{kl}) - \Lambda_0^\circ(T_{kl})\} \sum_{i=1}^n \sum_{j=1}^{m_i} dN_{ij}(s). \qquad (25)
\end{aligned}$$

The third term of the above equation can be written, by interchanging the order of integration, as

$$\begin{aligned}
& n^{-2} \sum_{k=1}^n \sum_{l=1}^{m_k} \sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^t \frac{R_{k\cdot}(s) \eta_{1k}(s)}{\{\mathcal{Y}(s, \Lambda_0^\circ)\}^2} \exp(\boldsymbol{\beta}^T \mathbf{Z}_{kl}) \left[\int_0^s \{\hat{\Lambda}_0(u) - \Lambda_0^\circ(u)\} d\tilde{N}_{kl}(u) \right] dN_{ij}(s) \\
& \qquad = \int_0^t \{\hat{\Lambda}_0(s) - \Lambda_0^\circ(s)\} \sum_{i=1}^n \sum_{j=1}^{m_i} \Omega_{ij}(s, t) d\tilde{N}_{ij}(s),
\end{aligned}$$

where $\tilde{N}_{ij}(t) = I(T_{ij} \le t)$ and

$$\Omega_{ij}(s, t) = n^{-2} \int_s^t \{\mathcal{Y}(u, \Lambda_0^\circ)\}^{-2} R_{i\cdot}(u) \eta_{1i}(u) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) \sum_{k=1}^n \sum_{l=1}^{m_k} dN_{kl}(u).$$

Hence we get

$$\begin{aligned}
\hat{\Lambda}_0(t, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t) &= n^{-1} \int_0^t \{\mathcal{Y}(s, \Lambda_0^\circ)\}^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} dM_{ij}(s) \\
&- \int_0^t \{\hat{\Lambda}_0(s, \boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(s)\} \sum_{i=1}^n \sum_{j=1}^{m_i} \{\delta_{ij} \Upsilon(s) + \Omega_{ij}(s, t) + o(n^{-1})\} d\tilde{N}_{ij}(s)
\end{aligned}$$

where

$$\Upsilon(s) = n^{-2} \{\mathcal{Y}(s, \Lambda_0^\circ)\}^{-2} \sum_{k=1}^n \sum_{l=1}^{m_k} I(T_{kl} > s) R_{k\cdot}(s) \eta_{1k}(s) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{kl}).$$

The $o(n^{-1})$ is uniform in t (see Section 4.6) and will be dominated by Ω and Υ, which are of order n^{-1}. Hence the $o(n^{-1})$ term can be ignored.

An argument similar to that of Yang and Prentice (1999) and Zucker (2005) now yields the martingale representation

$$\hat{\Lambda}_0(t,\boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(t) \;\approx\; \frac{1}{n\hat{p}(t)} \int_0^t \frac{\hat{p}(s-)\sum_{i=1}^n \sum_{j=1}^{m_i} dM_{ij}(s)}{\mathcal{Y}(s,\Lambda_0^\circ)}, \tag{26}$$

where

$$\hat{p}(t) = \prod_{s \le t} \left[1 + \sum_{i=1}^n \sum_{j=1}^{m_i} \{\delta_{ij}\Upsilon(s) + \Omega_{ij}(s,t)\} d\tilde{N}_{ij}(s) \right].$$

Based on (22), we can write

$$U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) \approx n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \int_0^\tau Q_{ijr}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ, s)\{\hat{\Lambda}_0(s,\boldsymbol{\gamma}^\circ) - \Lambda_0^\circ(s)\} d\tilde{N}_{ij}(s).$$

Plugging the martingale representation (26) into the above equation and carrying out some more algebra (again involving an interchange of integrals) gives

$$\begin{aligned} & U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) \\ & \qquad \approx\; n^{-1} \int_0^\tau \pi_r(s, \boldsymbol{\gamma}^\circ, \Lambda_0^\circ) \frac{\hat{p}(s-)\sum_{k=1}^n \sum_{l=1}^{m_k} dM_{kl}(s)}{\mathcal{Y}(s,\Lambda_0^\circ)}, \end{aligned} \tag{27}$$

where

$$\pi_r(s,\boldsymbol{\gamma},\Lambda_0) = n^{-1} \int_s^\tau \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} Q_{ijr}(\boldsymbol{\gamma},\Lambda_0,t) d\tilde{N}_{ij}(t)}{\hat{p}(t)}.$$

Therefore, $n^{1/2}[\mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot,\boldsymbol{\gamma}^\circ)) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ(\cdot,\boldsymbol{\gamma}^\circ))]$ is asymptotically mean zero multivariate normal with covariance matrix that can be consistently estimated by

$$G_{rl}(\hat{\boldsymbol{\gamma}}) = n^{-1} \int_0^\tau \pi_r(s,\hat{\boldsymbol{\gamma}},\hat{\Lambda}_0)\pi_l(s,\hat{\boldsymbol{\gamma}},\hat{\Lambda}_0)\{\hat{p}(s-)\}^2 \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} dN_{ij}(s)}{\{\mathcal{Y}(s,\hat{\Lambda}_0)\}^2}$$

for $r, l = 1, \ldots, p+1$.

Step III

We now examine the sum of $\mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)$ and $\mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot,\boldsymbol{\gamma}^\circ)) - \mathbf{U}(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)$. From (27), we have

$$U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot,\boldsymbol{\gamma}^\circ)) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) \approx n^{-1} \int_0^\tau \alpha_r(s) \sum_{k=1}^n \sum_{l=1}^{m_k} dM_{kl}(s) = \frac{1}{n} \sum_{k=1}^n \mu_{kr},$$

where $\alpha_r(s)$ is the limiting value of $\pi_r(s, \boldsymbol{\gamma}^\circ, \Lambda_0^\circ)\hat{p}(s-)/\mathcal{Y}(s, \Lambda_0^\circ)$ and μ_{kr} is defined as

$$\mu_{kr} = \int_0^\tau \alpha_r(s) \sum_{l=1}^{m_k} dM_{kl}(s).$$

Arguments in Yang and Prentice (1999, Appendix A) can be used to show that $\hat{p}(s-)$ has a limit. Also, clearly $\mathrm{E}[\mu_{kr}] = 0$.

We thus have

$$U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) + [U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)] \approx \frac{1}{n}\sum_{i=1}^{n}(\xi_{ir} + \mu_{ir}),$$

which is a mean of n iid random variables. Hence $n^{1/2}\{U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ) + [U_r(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) - U_r(\boldsymbol{\gamma}^\circ, \Lambda_0^\circ)]\}$ is asymptotically normally distributed. The covariance matrix may be estimated by $\hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}}) + \hat{\mathbf{G}}(\hat{\boldsymbol{\gamma}}) + \hat{\mathbf{C}}(\hat{\boldsymbol{\gamma}})$, where

$$\hat{C}_{rl}(\hat{\boldsymbol{\gamma}}) = \frac{1}{n}\sum_{i=1}^{n}(\xi_{ir}^*\mu_{il}^* + \xi_{il}^*\mu_{ir}^*), \quad r, l = 1, \ldots, p+1,$$

with

$$\mu_{ir}^* = \int_0^\tau \frac{\pi_r(s, \hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0)\hat{p}(s-)}{\mathcal{Y}(s, \hat{\Lambda}_0)} \sum_{j=1}^{m_i} d\hat{M}_{ij}(s)$$

and

$$\hat{M}_{ij}(t) = N_{ij}(t) - \int_0^t \exp(\hat{\boldsymbol{\beta}}^T\mathbf{Z}_{ij})Y_{ij}(u)\psi_i(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0, u-)d\hat{\Lambda}_0(u).$$

Step IV

First order Taylor expansion of $\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}}))$ about $\boldsymbol{\gamma}^\circ = (\boldsymbol{\beta}^{\circ T}, \theta^\circ)^T$ gives

$$\mathbf{U}(\hat{\boldsymbol{\gamma}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\gamma}})) = \mathbf{U}(\boldsymbol{\gamma}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}^\circ)) + \mathbf{D}(\boldsymbol{\gamma}^\circ)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)^T + o_p(1),$$

where

$$D_{ls}(\boldsymbol{\gamma}) = \partial U_l(\boldsymbol{\gamma}, \hat{\Lambda}_0(\cdot, \boldsymbol{\gamma}))/\partial\gamma_s$$

for $l, s = 1, \ldots, p+1$, with $\gamma_{p+1} = \theta$.

For $l, s = 1, \ldots, p$ we have

$$\begin{aligned} D_{ls}(\boldsymbol{\gamma}) &= -n^{-1} \sum_{i=1}^{n} \left\{ \frac{\phi_{2i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} \sum_{j=1}^{m_i} Z_{ijl} \frac{\partial \hat{H}_{ij}(T_{ij})}{\partial \beta_s} \right. \\ &\quad \left. - \left[\frac{\phi_{3i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} - \frac{\phi_{2i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} \right] \sum_{j=1}^{m_i} \hat{H}_{ij}(T_{ij}) Z_{ijl} \frac{\partial \hat{H}_{i.}(\tau)}{\partial \beta_s} \right\}, \end{aligned} \tag{28}$$

$$\frac{\partial \hat{H}_{ij}(\tau_k)}{\partial \beta_s} = \frac{\partial \hat{\Lambda}_0(T_{ij} \wedge \tau_k)}{\partial \beta_s} \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) + \hat{\Lambda}_0(T_{ij} \wedge \tau_k) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}) Z_{ijs}$$

and

$$\begin{aligned} \frac{\partial \Delta \hat{\Lambda}_0(\tau_k)}{\partial \beta_s} &= -d_k \left\{ \sum_{i=1}^{n} \frac{\phi_{2i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})} R_{i.}(\tau_k) \right\}^{-2} \\ &\quad \sum_{i=1}^{n} \left[\left\{ \frac{\phi_{2i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})}{\phi_{1i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})} - \frac{\phi_{3i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})} \right\} \frac{\partial \hat{H}_{i.}(\tau_{k-1})}{\partial \beta_s} R_{i.}(\tau_k) \right. \\ &\quad \left. + \frac{\phi_{2i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau_{k-1})} \sum_{j=1}^{m_i} R_{ij}(\tau_k) Z_{ijs} \right]. \end{aligned}$$

For $l = 1, \ldots, p$ we have

$$\begin{aligned} D_{l(p+1)}(\boldsymbol{\gamma}) &= -n^{-1} \sum_{i=1}^{n} \left\{ \frac{\phi_{2i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} \sum_{j=1}^{m_i} Z_{ijl} \frac{\partial \hat{H}_{ij}(T_{ij})}{\partial \theta} \right. \\ &\quad + \left[\frac{\phi_{2i}^{(\theta)}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} - \frac{\phi_{2i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau) \phi_{1i}^{(\theta)}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} \right. \\ &\quad \left. \left. + \left\{ \frac{\phi_{2i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} - \frac{\phi_{3i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} \right\} \frac{\partial \hat{H}_{i.}(\tau)}{\partial \theta} \right] \sum_{j=1}^{m_i} \hat{H}_{ij}(T_{ij}) Z_{ijl} \right\} \end{aligned} \tag{29}$$

and

$$D_{(p+1)l}(\boldsymbol{\gamma}) = n^{-1} \sum_{i=1}^{n} \left\{ \frac{\phi_{1i}^{(\theta)}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau) \phi_{2i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} - \frac{\phi_{2i}^{(\theta)}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} \right\} \frac{\partial \hat{H}_{i.}(\tau)}{\partial \beta_l}. \tag{30}$$

Finally,

$$\begin{aligned} D_{(p+1)(p+1)}(\boldsymbol{\gamma}) &= n^{-1} \sum_{i=1}^{n} \left\{ \frac{\phi_{1i}^{(\theta,\theta)}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} - \left[\frac{\phi_{1i}^{(\theta)}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0)} \right]^2 \right. \\ &\quad \left. + \left[\frac{\phi_{1i}^{(\theta)}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau) \phi_{2i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}^2(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} - \frac{\phi_{2i}^{(\theta)}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)}{\phi_{1i}(\boldsymbol{\gamma}, \hat{\Lambda}_0, \tau)} \right] \frac{\partial \hat{H}_{i.}(\tau)}{\partial \theta} \right\} \end{aligned} \tag{31}$$

where

$$\phi_{1i}^{(\theta,\theta)}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau) = \int w^{N_{i.}(\tau)} \exp\{-w\hat{H}_{i.}(\tau)\} \frac{d^2 f(w)}{d\theta^2} dw,$$

$$\frac{\partial \hat{H}_{ij}(\tau_k)}{\partial \theta} = \frac{\partial \hat{\Lambda}_0(T_{ij} \wedge \tau_k)}{\partial \theta} \exp(\boldsymbol{\beta}^T \mathbf{Z}_{ij}),$$

and

$$\begin{aligned}
\frac{\partial \Delta\hat{\Lambda}_0(\tau_k)}{\partial \theta} &= -d_k \left\{ \sum_{i=1}^{n} \frac{\phi_{2i}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})}{\phi_{1i}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})} R_{i.}(\tau_k) \right\}^{-2} \\
&\quad \sum_{i=1}^{n} R_{i.}(\tau_k) \left[\frac{\phi_{2i}^{(\theta)}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})}{\phi_{1i}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})} - \frac{\phi_{2i}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})\phi_{1i}^{(\theta)}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})}{\phi_{1i}^2(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})} \right. \\
&\quad \left. + \frac{\partial \hat{H}_{i.}(\tau_{k-1})}{\partial \theta} \left\{ \frac{\phi_{2i}^2(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})}{\phi_{1i}^2(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})} - \frac{\phi_{3i}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})}{\phi_{1i}(\boldsymbol{\gamma},\hat{\Lambda}_0,\tau_{k-1})} \right\} \right].
\end{aligned}$$

Step V

Combining the results above we get that $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^\circ)$ is asymptotically zero-mean normally distributed with a covariance matrix that can be consistently estimated by

$$\hat{\mathbf{D}}^{-1}(\hat{\boldsymbol{\gamma}})\{\hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}}) + \hat{\mathbf{G}}(\hat{\boldsymbol{\gamma}}) + \hat{\mathbf{C}}(\hat{\boldsymbol{\gamma}})\}\hat{\mathbf{D}}^{-1}(\hat{\boldsymbol{\gamma}})^T.$$

5.6 Proof of (20)

The goal is to prove that

$$\sup_{s,\boldsymbol{\gamma}} |A(\boldsymbol{\gamma},\tilde{\Lambda}^{(n)},s) - a(\boldsymbol{\gamma},\tilde{\Lambda}^{(n)},s)| \to 0 \quad \text{a.s.} \tag{32}$$

This involves several steps.

First, it is easy to see that there exists a constant κ (independent of $\boldsymbol{\gamma}$ and s) such that

$$\sup_{s,\boldsymbol{\gamma}} |A(\boldsymbol{\gamma},\Lambda_1,s) - A(\boldsymbol{\gamma},\Lambda_2,s)| \leq \kappa \|\Lambda_1 - \Lambda_2\|, \tag{33}$$

$$\sup_{s,\boldsymbol{\gamma}} |a(\boldsymbol{\gamma},\Lambda_1,s) - a(\boldsymbol{\gamma},\Lambda_2,s)| \leq \kappa \|\Lambda_1 - \Lambda_2\|. \tag{34}$$

Next, for any fixed continuous Λ, the functional strong law of large numbers of Andersen and Gill (1982, Appendix III) implies that, with probability one,

$$\sup_{s,\boldsymbol{\gamma}} |A(\boldsymbol{\gamma}, \Lambda, s) - a(\boldsymbol{\gamma}, \Lambda, s)| \to 0. \tag{35}$$

Now, given $\epsilon > 0$, define the sets $\{t_j^{(\epsilon)}\}$, $\{\boldsymbol{\gamma}_k^{(\epsilon)}\}$, and $\{\Lambda_l^{(\epsilon)}\}$ to be finite partition grids of $[0, \tau]$, $\mathcal{G}$, and $[0, \Lambda_{max}]$, respectively, with distance of no more than ϵ between grid points. Define $\mathcal{L}_\epsilon^*$ to be the set of functions of t and $\boldsymbol{\gamma}$ defined by linear interpolation through vertices of the form $(t_j^{(\epsilon)}, \boldsymbol{\gamma}_k^{(\epsilon)}, \Lambda_l^{(\epsilon)})$.

Obviously $\mathcal{L}_\epsilon^*$ is a finite set. Hence, in view of (35), there exists a probability-one set of realizations Ω_ϵ for which

$$\sup_{s\in[0,\tau],\boldsymbol{\gamma}\in\mathcal{G},\Lambda\in\mathcal{L}_\epsilon^*} |A(\boldsymbol{\gamma}, \Lambda, s) - a(\boldsymbol{\gamma}, \Lambda, s)| \to 0. \tag{36}$$

Define

$$\Omega^{**} = \bigcap_{\ell=1}^{\infty} \Omega_{1/\ell}$$

and $\Omega_0 = \Omega^* \cap \Omega^{**}$, with Ω^* as defined earlier. Clearly $\Pr(\Omega_0) = 1$. From now on, we restrict attention to Ω_0.

Now let $\epsilon > 0$ be given. Choose $\ell > \epsilon^{-1}$. In view of (18) and (36), we can find for any $\omega \in \Omega_0$ a suitable positive integer $\bar{n}(\epsilon, \omega)$ such that, whenever $n \geq \bar{n}(\epsilon, \omega)$,

$$|\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\gamma}) - \tilde{\Lambda}_0^{(n)}(u, \boldsymbol{\gamma})| \leq B^*(t - u) + \frac{\epsilon}{2} \quad \forall t, u, \tag{37}$$

$$\sup_{s\in[0,\tau],\boldsymbol{\gamma}\in\mathcal{G},\Lambda\in\mathcal{L}_{1/\ell}^*} |A(\boldsymbol{\gamma}, \Lambda, s) - a(\boldsymbol{\gamma}, \Lambda, s)| \leq \epsilon. \tag{38}$$

Next, let $\bar{\Lambda}_0^{(n)}$ denote the function defined by linear interpolation through $(t_j^{(\epsilon)}, \boldsymbol{\gamma}_k^{(\epsilon)}, \bar{\Lambda}_{jk}^{(\epsilon)})$, where $\bar{\Lambda}_{jk}^{(\epsilon)}$ is the element of $\{\Lambda_l^{(\epsilon)}\}$ that is closest to $\tilde{\Lambda}_0^{(n)}(t_j^{(\epsilon)}, \boldsymbol{\gamma}_k^{(\epsilon)})$. It is clear that

$$|\bar{\Lambda}_0^{(n)}(t_j^{(\epsilon)}, \boldsymbol{\gamma}_k^{(\epsilon)}) - \tilde{\Lambda}_0^{(n)}(t_j^{(\epsilon)}, \boldsymbol{\gamma}_k^{(\epsilon)})| \leq \epsilon \quad \forall j, k.$$

Using (37) and the Lipschitz continuity of $\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ (which follows from the corresponding property of $\hat{\Lambda}_0(t,\boldsymbol{\gamma})$), we thus obtain

$$\sup_{t,\boldsymbol{\gamma}} |\bar{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma}) - \tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\gamma})| \leq B^{**}\epsilon$$

for a suitable fixed constant B^{**} (depending on B^* and C^*). Combining this with (38) and (34), we obtain

$$\sup_{s,\boldsymbol{\gamma}} |A(\boldsymbol{\gamma},\tilde{\Lambda}^{(n)},s) - a(\boldsymbol{\gamma},\tilde{\Lambda}^{(n)},s)| \leq (2\kappa B^{**}+1)\epsilon \quad \text{for all } n \geq \bar{n}(\epsilon,\omega).$$

Since ϵ was arbitrary, the desired conclusion (32) follows, and the proof is thus complete.

5.7 Definition and behavior of $h_i(r,s)$

The quantity $h_i(r,s)$ appearing in (24) is given by

$$\begin{aligned} h_i(r,s) &= \frac{2R_{i\cdot}(s)\eta_{1i}(r,s)}{\{\mathcal{Y}(s,\Lambda_0^\circ + r\Delta)\}^3}\frac{1}{n}\sum_{l=1}^{n} R_{l\cdot}(s)\eta_{1l}(r,s)\sum_{j=1}^{m_i}\exp(\boldsymbol{\beta}^T\mathbf{Z}_{lj})\Delta(T_{lj}\wedge s) \\ &\quad -\frac{R_{i\cdot}(s)\eta_{2i}(r,s)}{\{\mathcal{Y}(s,\Lambda_0^\circ + r\Delta)\}^2}\sum_{j=1}^{m_i}\exp(\boldsymbol{\beta}^T\mathbf{Z}_{ij})\Delta(T_{ij}\wedge s) \end{aligned}$$

where $\Delta(T_{ij}\wedge s) = \hat{\Lambda}_0(T_{ij}\wedge s) - \Lambda_0^o(T_{ij}\wedge s)$ and

$$\eta_{2i}(r,s) = 2\left\{\frac{\phi_{2i}(\boldsymbol{\gamma}^\circ,\Lambda_0^\circ + r\Delta,s)}{\phi_{1i}(\boldsymbol{\gamma}^\circ,\Lambda_0^\circ + r\Delta,s)}\right\}^3 + \frac{\phi_{4i}(\boldsymbol{\gamma}^\circ,\Lambda_0^\circ + r\Delta,s)}{\phi_{1i}(\boldsymbol{\gamma}^\circ,\Lambda_0^\circ + r\Delta,s)} - 3\frac{\phi_{2i}(\boldsymbol{\gamma}^\circ,\Lambda_0^\circ + r\Delta,s)\phi_{3i}(\boldsymbol{\gamma}^\circ,\Lambda_0^\circ + r\Delta,s)}{\{\phi_{1i}(\boldsymbol{\gamma}^\circ,\Lambda_0^\circ + r\Delta,s)\}^2}.$$

For all $i = 1,\ldots,n$ and $s \in [0,\tau]$, we have $0 \leq R_{i\cdot}(s) \leq m\nu$, where ν is as in (8). Moreover, for $k = 1,\ldots,4$, we have

$$\mathrm{E}[W_i^{r_{min}+(k-1)}\exp\{-W_i m e^{\boldsymbol{\beta}^T Z}\Lambda_0^\circ(\tau)\}] \leq \phi_{ki}(\boldsymbol{\gamma}^\circ,\Lambda_0^\circ,s) \leq \mathrm{E}[W_i^{r_{max}+(k-1)}]$$

where $r_{\max} = \arg\max_{1\leq r\leq m}\mathrm{E}(W_i^r)$, $r_{\min} = \arg\min_{1\leq r\leq m}\mathrm{E}(W_i^r)$. Hence, η_{1i} and η_{2i} are bounded. In addition, the the proof of Lemma 2 show that $\mathcal{Y}(s,\Lambda^\circ + r\Delta)$ is uniformly bounded away from zero for n sufficiently large. Finally, in the consistency proof we obtained $\|\Delta\| = o(1)$. Therefore $h_i(r,s)$ is $o(1)$ uniformly in r and s.

6 Acknowledgements

We thanks the referees for their helpful comments, and for calling our attention to the work of Dabrowska (2006a, 2006b).

7 References

Aalen, O. O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3**, 15-27.

Abramowitz, M. and Stegun, I. A. (Eds.) (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing* New York: Dover.

Andersen, P. K., Borgan, O, Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Berlin: Springer-Verlag.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.

Andersen, P. K., Klein, J. P., Knudsen, K. M. and Palacios, R. T. (1997). Estimation of variance in Cox's regression model with shared gamma frailty. *Biometrics* **53**, 1475-1484.

Bickel, P. (1985). Efficient testing in a class of transformation models. *Bull. Int. Statist. Inst.*, **51**, 53-81, Meeting 23, Amsterdam.

Bagdonavicius, V. B., and Nikulin, M. S. (1999). Generalized proportional hazards model based on modified partial likelihood *Lifetime Data Analysis*, **5**, 329-350.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, **30**, 89-99.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc.* **B 34**, 187-220.

Dabrowska, D. (2006a). Estimation in a class of semi-parametric transformation models. In *Optimality: Second Erich L. Lehmann Symposium, Institute of Mathematical Statistics Lecture Notes and Monographs Series Vol. 49* (J. Roho, ed.). Beachwood, OH: Institute of Mathematical Statistics.

Dabrowska, D. (2006b). Information bounds and efficient estimation in a class of censored transformation models. Technical report. Available at arXiv:math.ST/0608088.

Fine, J. P., Glidden D. V. and Lee, K. (2003). A simple estimator for a shared frailty regression model. *J. R. Statist. Soc.* **B 65**, 317-329.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equation. *J. Amer. Statist. Ass.* **72**, 147-148.

Gill, R. D. (1985). Discussion of the paper by D. Clayton and J. Cuzick. *J. R. Statist. Soc.* **A 148**, 108-109.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the Von Mises method (Part 1). *Scand. J. Statist.* **16**, 97-128.

Gill, R. D. (1992). Marginal partial likelihood. *Scand. J. Statist.* **79**, 133-137.

Gorfine, M., Zucker, D. M., and Hsu, L. (2006). Prospective survival analysis with a general semiparametric shared frailty model - a pseudo full likelihood approach. *Biometrika* **93**, 735-741.

HARTMAN, P. (1973). *Ordinary Differential Equations,* 2nd ed. (reprinted, 1982), Boston: Birkhauser.

HENDERSON, R. AND OMAN, P. (1999). Effect of frailty on marginal regression estimates in survival analysis. *J. R. Statist. Soc.* **B 61**, 367-379.

HOUGAARD, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387-396.

HOUGAARD, P. (2000). *Analysis of Multivariate Survival data.* New York: Springer.

KLEIN, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM Algorithm. *Biometrics* **48**, 795-806.

LANCASTER, T., AND NICKELL, S. J. (1980). The analysis of re-employment probabilities for the unemployed. *Journal of the Royal statistical Society, Series A* **143**, 141-165.

LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. R. Statis. Soc.* **B 44**, 226-233.

MCGILCHRIST, C. A. (1993). REML estimation for survival models with frailty. *Biometrics* **49**, 221-225.

MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712-731.

MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182-198.

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sorensen, T. I. (1992). A counting process approach to maximum likelihood estimation of frailty models. *Scand. J. Statist.* **19**, 25-43.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183-214.

Ripatti, S. and Palmgren J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016-1022.

Vaida, F. and Xu, R. H. (2000). Proportional hazards model with random effects. *Stat. in Med.* **19**, 3309-3324.

Yang, S. and Prentice, R. L. (1999). Semiparametric inference in the proportional odds regression model. *J. Amer. Statist. Ass.* **94**, 125-136.

Zucker, D. M. (2005). A pseudo partial likelihood method for semi-parametric survival regression with covariate errors. *J. Amer. Statist. Ass.* **100**, 1264-1277.

Table 1: *Simulation results for family size 2. A: Empirical mean. B: Empirical standard deviation. C: Estimated Standard deviation. D: Coverage rate. E: Correlation.*

				$\hat{\beta}$		$\hat{\theta}$	
		censoring					
θ	β	%		Our approach	EM algorithm	Our approach	EM algorithm
2	ln(2)	35	A	0.692	0.689	1.978	1.969
			B	0.248	0.253	0.268	0.308
			C	0.242	-	0.242	-
			D	95.6	-	96.3	-
			E	0.952		0.989	
		85	A	0.699	0.693	1.942	1.942
			B	0.479	0.481	0.897	0.936
			C	0.487	-	0.919	-
			D	96.6	-	95.0	-
			E	0.952		0.989	
	ln(3)	30	A	1.102	1.078	1.985	1.961
			B	0.255	0.266	0.265	0.259
			C	0.231	-	0.279	-
			D	96.9	-	96.1	-
			E	0.951		0.982	
		80	A	1.099	1.088	1.921	1.870
			B	0.465	0.466	0.800	0.810
			C	0.443	-	0.797	-
			D	94.2	-	96.3	-
			E	0.957		0.993	
4	ln(2)	50	A	0.680	0.676	3.955	3.958
			B	0.309	0.309	0.522	0.521
			C	0.415	-	0.705	-
			D	93.4	-	97.6	-
			E	0.995		0.997	
		90	A	0.685	0.673	3.842	3.824
			B	0.538	0.535	1.486.	1.479
			C	0.705	-	1.601	-
			D	91.0	-	92.7	-
			E	0.992		0.998	
	ln(3)	45	A	1.083	1.080	3.955	3.957
			B	0.309	0.309	0.517	0.515
			C	0.405	-	0.760	-
			D	92.0	-	97.8	-
			E	0.992		0.997	
		85	A	1.086	1.075	3.868	3.843
			B	0.530	0.522	1.345	1.336
			C	0.699	-	1.523	-
			D	92.2	-	94.1	-
			E	0.987		0.997	

Table 2: *Simulation results for family size equals 5. A: Empirical mean. B: Empirical standard deviation. C: Estimated Standard deviation. D: Coverage rate. E: Correlation.*

				$\hat{\beta}$		$\hat{\theta}$	
		censoring					
θ	β	%		Our approach	EM algorithm	Our approach	EM algorithm
2	ln(2)	35	A	0.693	0.693	2.001	2.001
			B	0.129	0.129	0.187	0.186
			C	0.134	-	0.171	-
			D	95.4	-	97.0	-
			E	0.997		0.998	
		85	A	0.698	0.698	1.978	1.977
			B	0.283	0.281	0.386	0.385
			C	0.335	-	0.498	-
			D	97.2	-	95.9	-
			E	0.995		0.999	
	ln(3)	30	A	1.098	1.097	2.001	2.002
			B	0.129	0.129	0.183	0.183
			C	0.153	-	0.221	-
			D	98.0	-	98.0	-
			E	0.996		0.998	
		80	A	1.104	1.104	1.984	1.983
			B	0.270	0.269	0.361	0.360
			C	0.304	-	0.429	-
			D	96.2	-	96.9	-
			E	0.993		0.999	
4	ln(2)	50	A	0.696	0.696	3.996	3.996
			B	0.152	0.151	0.369	0.369
			C	0.159	-	0.428	-
			D	96.4	-	98.0	-
			E	0.997		0.998	
		90	A	0.712	0.709	3.957	3.955
			B	0.307	0.304	0.715	0.714
			C	0.372	-	0.803	-
			D	94.8	-	97.5	-
			E	0.995		0.999	
	ln(3)	45	A	1.102	1.101	3.995	3.996
			B	0.150	0.149	0.363	0.363
			C	0.210	-	0.401	-
			D	98.0	-	94.5	-
			E	0.997		0.998	
		85	A	1.117	1.115	3.979	3.977
			B	0.291	0.288	0.671	0.671
			C	0.346	-	0.689	-
			D	94.2	-	98.0	-
			E	0.993		0.999	