The Annals of Statistics
1995, Vol. 23, No. 1, 182–198

ASYMPTOTIC THEORY FOR THE FRAILTY MODEL[1]

BY S. A. MURPHY

Pennsylvania State University

The frailty model is a generalization of Cox's proportional hazards model which includes a random effect. Nielsen, Gill, Andersen and Sørensen (1992) proposed an EM algorithm to estimate the cumulative baseline hazard and the variance of the random effect. Here the asymptotic distribution of the estimators is given along with a consistent estimator of the asymptotic variance.

0. Introduction. The topic of heterogeneity in the analysis of duration times has received much attention in a wide variety of fields, such as demography [Vaupel (1990)], econometrics [Heckman and Singer (1984)] and statistics [Clayton and Cuzick (1985), Aalen (1988), Nielsen, Gill, Andersen and Sørensen (1992)]. The heterogeneity may be caused by an unobservable covariate as explained by Heckman and Singer or due to block effects as in the Danish twin studies described by Vaupel. An additional source of dependence is due to the subject effect when recording serial events per subject [Oakes (1991)].

A simple model for heterogeneity which is reminiscent of the proportional hazards model is via the concept of frailty as proposed by Vaupel, Manton and Stallard (1979). The frailty is an unobserved random factor applied to the baseline hazard function of the duration time. So if the hazard function of a subject with a frailty value of 1 is given by $\alpha(t)$, then the hazard function of a subject with a frailty value of z is given by $z\alpha(t)$. In this paper the frailty is assumed to follow a gamma distribution with mean 1 and unknown variance. Mathematically this is convenient as the gamma distribution is a conjugate prior. In some settings there is some justification for the assumption that the frailty follows a distribution skewed to the right, such as a lognormal or gamma distribution [Aalen (1988)]. At any rate, as discussed by Manton, Stallard and Vaupel. (1986) often the fit of a model is more sensitive to assumptions on the form of $\alpha(t)$ than on the form of the frailty distribution. Heckman and Singer (1984) give a functional form to α and leave the form of the frailty distribution unspecified. Other settings allow the specification of both the form of α and of the frailty distribution. In this paper, α is not parametrized and, in the manner of Cox's treatment of the proportional hazards model, covariates influence the hazard function proportionally. In other words, the assumption is that conditional on the value of the unobserved frailty, the duration times follow

Received December 1992; revised January 1994.
[1]Research partially supported by NSF Grant DMS-85-05550 while visiting the Mathematical Sciences Research Institute at Berkeley, and by NSF Grant DMS-93-07255.
AMS 1991 *subject classifications*. 62M09, 62G05.
Key words and phrases. Counting processes, survival analysis, frailty model, random effects, nonparametric maximum likelihood.

the usual proportional hazards model. Nielsen, Gill, Andersen and Sørensen (1992) consider maximum likelihood estimation in this model. The estimation is carried out via an EM algorithm and works quite well. Murphy (1994) considers consistency of the maximum likelihood estimators when the model has only two parameters: A, the integrated version of α; and θ, the variance of the gamma-distributed frailty. In this article, the asymptotic distribution of these estimators is given along with estimators of their asymptotic variances. If θ is zero, then the frailty is identically equal to 1, indicating no heterogeneity.

1. The statistical model. In the following this model is placed in a counting process framework. This is convenient not only because of the generality but also because this framework facilitates the recognition that the Fisher information is invertible and that the model is identifiable. However, counting process technology is primarily used in forming means and variances and not in proofs of the weak convergence; there empirical processes are used.

Much of following description is a review of the counting process approach presented in Nielsen, Gill, Andersen and Sørensen (1992). Using their notation, the frailty or random effect is defined on the probability space (Ω', G', P'_θ) and is denoted by $\mathbf{Z} = (Z_1, \dots, Z_n)$. Let $(\Omega'', \{G''_{t\in[0,\tau]}\}, P''_{\alpha\mathbf{z}})$ be a filtered probability space for each $\mathbf{Z} = \mathbf{z}$, so that under $P''_{\alpha\mathbf{z}}$ (i.e., conditionally on $\mathbf{Z} = \mathbf{z}$), the multivariate counting process $\mathbf{N} = (N_i\colon\ i = 1, \dots, n)$ has intensity process $\boldsymbol{\lambda}$ given by

$$\lambda_i(u) = z_i Y_i(u)\alpha(u).$$

The N_i represent the aggregate of the counting processes for group i, so that each N_i can have more than one jump. The members of the ith group share the same frailty, Z_i. The Z_i are assumed to be independent random variables each distributed according to a gamma distribution with mean 1 and variance θ. The Y_i are observable, nonnegative, predictable processes and α is an unknown baseline hazard rate. The goal is to estimate θ and the cumulative baseline hazard $A(t) = \int_0^t \alpha(u)\,du$ based on observation of $(\mathbf{N}, \mathbf{Y})$ only and via maximum likelihood estimation. There are at least two ways to form the likelihood of $(\mathbf{N}, \mathbf{Y})$. The first method is to write the likelihood of $(\mathbf{N}, \mathbf{Y}, \mathbf{Z})$ as the density of $(\mathbf{N}, \mathbf{Y})$ given $\mathbf{Z} = \mathbf{z}$ and the density of $\mathbf{Z}$, and integrate over $\mathbf{z}$. Actually only a partial conditional likelihood of $(\mathbf{N}, \mathbf{Y})$ given $\mathbf{Z} = \mathbf{z}$ is specified, and it is assumed that the remaining term in the conditional likelihood does not involve $\mathbf{z}$ [Nielsen, Gill, Andersen and Sørensen (1992) state this in Assumption 2, "Conditional on $\mathbf{Z} = \mathbf{z}$, censoring is noninformative of $\mathbf{z}$"]. The partial likelihood $(\mathbf{N}, \mathbf{Y})$ given $\mathbf{Z} = \mathbf{z}$ is

$$\prod_{i=1}^{n}\left\{\prod_{t\le\tau}(z_i Y_i(t)\alpha(t))^{\Delta N_i(t)}\exp\left\{-z_i\int_0^\tau Y_i\,dA\right\}\right\}.$$

Multiplying by the density of $\mathbf{Z}$ and then integrating over $\mathbf{z}$ yields the partial likelihood

$$\prod_{i=1}^{n}\frac{\prod_t((1+\theta N_i(t-))Y_i(t)\alpha(t))^{\Delta N_i(t)}}{(1+\theta\int_0^\tau Y_i(t)dA(t))^{1/\theta+N_i(\tau)}}. \tag{1.1}$$

It is also straightforward to see that the distribution of Z_i given $(\mathbf{N}, \mathbf{Y})$ is a gamma distribution with mean

$$\frac{1+\theta N_i(\tau)}{1+\theta\int_0^\tau Y_i\,dA}$$

and variance

$$\theta\frac{1+\theta N_i(\tau)}{(1+\theta\int_0^\tau Y_i\,dA)^2}.$$

A second method of forming the partial likelihood of $(\mathbf{N}, \mathbf{Y})$ is to use the innovation theorem [Bremaud (1981)]; that is, in order to derive the intensity of $\mathbf{N}$ with respect to the observed history (i.e., the product of the trivial sigma field on Ω' with G''), Z_i is replaced by its conditional mean relative to this history. Therefore the intensity of $\mathbf{N}$ is

$$\lambda_i(u)=\frac{1+\theta N_i(u-)}{1+\theta\int_0^{u-}Y_i(s)\,dA(s)}Y_i(u)\alpha(u).$$

The partial likelihood function is then given by

$$\prod_{i=1}^{n}\left\{\prod_{u\le\tau}\left(\frac{1+\theta N_i(u-)}{1+\theta\int_0^{u-}Y_i(s)\,dA(s)}Y_i(u)\alpha(u)\right)^{\Delta N_i(u)}\right.$$
$$\left.\times\exp\left(-\int_0^\tau\frac{1+\theta N_i(t-)}{1+\theta\int_0^{t-}Y_i(s)\,dA(s)}Y_i(t)dA(t)\right)\right\}. \tag{1.2}$$

Since A is continuous, we can use integration by parts to show that equations (1.1) and (1.2) are equivalent. Both are full likelihoods for (θ, A) if the omitted term does not depend on (θ, A) [Nielsen, Gill, Andersen and Sørensen (1992) call this noninformative censoring for the parameter (θ, A)].

The true values of the parameters [say, (θ_0, A_o)] lie in $[0,\infty)\times\{$absolutely continuous cumulative hazards$\}$. However, maximization of the log-likelihood over this parameter space leads to the same difficulties as in estimation of a density function (no maximizer). An effective route out of this difficulty is to extend the parameter space so that the estimator $\hat{A}$ is allowed to be discrete. The parameter space is then $[0,\infty)\times\{$cumulative hazards$\}$. This is the type of extension of parameter spaces which allows one to consider the empirical distribution function as a nonparametric maximum likelihood estimator of a continuous distribution function. To allow for a discrete estimator, replace $\alpha(u)$ by $\Delta A(u)$, the jump of A at the point u, in (1.1) and (1.2). Consideration of L_n yields the result that the maximizer $\hat{A}$ is a step function with positive steps at each jump time of the N_i's. The natural logarithm of (1.1) is given by

$$nL_n(\theta, A) = \sum_{i=1}^{n} \int_0^{\tau} \ln(1 + \theta N_i(u-))\, dN_i(u)$$
$$- (\theta^{-1} + N_i(\tau)) \ln\Big(1 + \theta \int_0^{\tau} Y_i(u)\, dA(u)\Big)$$
$$+ \int_0^{\tau} \ln(Y_i(u) \triangle A(u))\, dN_i(u).$$

If $\theta = 0$, the second term above is defined by its right-hand limit at 0, that is, $\int_0^{\tau} Y_i \, dA$. Nielsen, Gill, Andersen and Sørensen (1992) maximize L_n via the EM algorithm. The EM algorithm solves the score equations set to zero in an iterative fashion. Murphy (1994) proves that the maximum likelihood estimators are consistent.

To calculate the score equations, one might differentiate L_n with respect to the jump sizes of A and θ. However, an equivalent method which is useful as a thinking tool is to consider one-dimensional submodels through the estimators and differentiate at the estimator; that is, set $A_t(\cdot) = \int_0^{\tau} 1 + th_1(u)\, d\hat{A}(u)$ and $\theta_t = th_2 + \hat{\theta}$ for h_1 a function and h_2 a scalar, and differentiate at $t = 0$ to get $S_n(\hat{A}, \hat{\theta})(h_1, h_2)$. Then if $(\hat{A}, \hat{\theta})$ maximizes L_n, $S_n(\hat{A}, \hat{\theta})(h_1, h_2) = 0$ for all (h_1, h_2). The form of S_n is given by $S_n = S_{n1} + S_{n2}$, where

$$S_{n1}(A, \theta)(h_1) = n^{-1} \sum_{i=1}^{n} \int_0^{\tau} h_1 \, dN_i - \frac{1 + \theta N_i(\tau)}{1 + \theta \int_0^{\tau} Y_i dA} \int_0^{\tau} h_1 Y_i \, dA$$

and

$$S_{n2}(A, \theta)(h_2) = h_2 n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \frac{N_i(u-)}{1 + \theta N_i(u-)}\, dN_i(u)$$
$$+ \theta^{-2} \Big(\ln\Big(1 + \theta \int_0^{\tau} Y_i \, dA\Big) - \frac{1 + \theta N_i(\tau)}{1 + \theta \int_0^{\tau} Y_i \, dA} \theta \int_0^{\tau} Y_i \, dA \Big).$$

For $\theta = 0$, the last term is interpretated as its limit as θ approaches zero to get $(\int_0^{\tau} Y_i \, dA)^2/2 - N_i(\tau) \int_0^{\tau} Y_i \, dA$.

Here the class of h is taken to be the space of bounded variation functions cross the reals. Define the norm to be $\|h\|_H = \|h_1\|_v + |h_2|$, where $\|h_1\|_v$ is the absolute value of $h_1(0)$ plus the total variation of h_1 on the interval $[0, \tau]$. Define H_p to be the product space of bounded variation functions on $[0, \tau]$ and real-valued scalars with norm $\|h\|_H = \|h_1\|_v + |h_2| \le p$. If $p = \infty$, then the inequality is strict. In the following, p is assumed finite unless stated otherwise. Define $(A, \theta)(h) = \int_0^{\tau} h_1 \, dA + h_2 \theta$. Then the parameter space Ψ can be considered to be a subset of $l^{\infty}(H_p)$, which is the space of bounded real-valued functions on H_p under the supremum norm $\|U\| = \sup_{h \in H_p} |U(h)|$. The score function S_n is a random map from Ψ to $l^{\infty}(H_p)$ for all finite p.

2. Asymptotic Distribution. Assume that the (N_i, Y_i) are i.i.d. copies of (N, Y), where Y is a.s. left-continuous with right-hand limits and takes on nonnegative values. The process M defined by

$$M(t) = \int_0^t dN(u) - \int_0^t \frac{1+\theta_0 N(u-)}{1+\theta_0 \int_0^{u-} Y(s)\, dA_o(s)} Y(u)\, dA_o(u)$$

is a martingale with respect to the filtration $\sigma\{N(s);\ Y(s),\ s \le t\};\ t \in [0, \tau]$. The variance parameter θ_0 lies in a known interval, say, $[0, K]$. The cumulative baseline hazard A_o is strictly increasing and is continuous on $[0, \tau]$ for $\tau < \infty$. Call the first jump of N, T_1. Convergence in probability (denoted by $\mathscr{P}^*$) and weak convergence is in terms of the outer measure [Pollard (1990)].

THEOREM 1. *Assume the following:*

(a) $\sup_{t\in(0,\tau]} |\hat{A}(t) - A_o(t)| \to_{\mathscr{P}^*} 0$ *and* $|\hat{\theta} - \theta_0| \to_{\mathscr{P}^*} 0$;
(b) *there exists some constant K for which* $\|Y\|_v \le K$ *and* $N(\tau) \le K$ *a.s.;*
(c) $\inf_{u\in(0,\tau]} EY(u) > 0$;
(d) $P[Y(T_1+) \ge 1] > 0$.

Then

$$(\sqrt{n}(\hat{A} - A_o), \sqrt{n}(\hat{\theta} - \theta_0)) \Longrightarrow \mathscr{G}$$

on $l^\infty(H_p)$; $\mathscr{G}$ *is a tight Gaussian process on* $l^\infty(H_p)$ *with mean zero and covariance process*

$$\operatorname{Cov}(\mathscr{G}(h), \mathscr{G}(h')) = \int_0^\tau h_1 \sigma_{(1)}^{-1}(h')\, dA_o + h_2 \sigma_{(2)}^{-1}(h'),$$

where $\sigma = (\sigma_1, \sigma_2)$ *is a continuously invertible linear operator from* H_∞ *onto* H_∞, *with inverse* $\sigma^{-1} = (\sigma_{(1)}^{-1}, \sigma_{(2)}^{-1})$. *The form of* σ *is as follows:*

$$\sigma_1(h)(u) = h_1(u) E(ZY(u)) - E\frac{\theta_0 \int_0^\tau Y h_1\, dA_o}{1+\theta_0 \int_0^\tau Y\, dA_o} ZY(u) - h_2 E\left(\frac{Y(u)}{1+\theta_0\int_0^\tau Y dA_o}\left(\int_0^\tau ZY\, dA_o - N(\tau)\right)\right)$$

and

$$\sigma_2(h) = h_2 E\left(-\frac{\partial^2 L_n(\theta, A)}{\partial \theta^2}\bigg|_{(\theta_0, A_o)}\right) - E\frac{\int_0^\tau h_1 Y\, dA_o}{1+\theta_0\int_0^\tau Y\, dA_o}\left(\int_0^\tau ZY\, dA_o - N(\tau)\right)$$

where the second partial of L_n with respect to θ is given by

$$-\frac{\partial^2 L_n(\theta, A)}{\partial\theta^2}\bigg|_{(\theta_0, A_o)}$$

$$= n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left(\frac{N_i(u-)}{1+\theta_0 N_i(u-)}\right)^2 dN_i(u) - N_i(\tau)\left(\frac{\int_0^\tau Y_i\,dA_o}{1+\theta_0\int_0^\tau Y_i\,dA_o}\right)^2$$

$$+ 2\theta_0^{-3}\left[\ln\left(1+\theta_0\int_0^\tau Y_i\,dA_o\right) - \frac{\theta_0\int_0^\tau dA_o}{1+\theta_0\int_0^\tau Y_i\,dA_o}\right.$$

$$\left. - \frac{1}{2}\left(\frac{\theta_0\int_0^\tau dA_o}{1+\theta_0\int_0^\tau Y_i\,dA_o}\right)^2\right].$$

When $\theta_0 = 0$, the last term above is defined by its limit, which is $\frac{2}{3}(\int_0^\tau Y_i\,dA_o)^3$.

Assumption (c) ensures that N has sufficient activity on the entire interval so as to estimate the parameter A_o. Note that (d) excludes the possibility of N having at most one jump. Some version of this assumption should be necessary because, as pointed out by Nielsen, Gill, Andersen and Sørensen (1992), the model is unidentifiable if all of the N_i have only one jump.

If the censoring is noninformative [Andersen, Borgan, Gill and Keiding (1993)] and we assume that A_o is absolutely continuous, then the partial likelihood is a full likelihood for estimation of (A, θ), and $(\hat{A}, \hat{\theta})$ will be efficient. This is proved via Theorems 3.1 and 3.3 of van der Vaart (1992). The conclusions of Lemmas 1 and 3 in the Appendix are exactly van der Vaart's conditions (2.2), (2.3) and (2.4). Local asymptotic normality is easily verified by using the boundedness of N and Y. Finally, asymptotic linearity of $(\hat{A}, \hat{\theta})$ is proved in the process of proving Theorem 1 [see (2.1) below].

The proof of Theorem 1 is based on the general theorem which has been stated by many people in various forms; the following is from van der Vaart (1992). The parameter space is $\Psi \subset l^\infty(H_p)$ and the score function is a random map S_n: $\Psi \to l^\infty(H_p)$. The true parameter value is ψ_o, and a maximum likelihood estimator is $\hat{\psi}$. The asymptotic version of S_n is S. We have that $S_n(\hat{\psi}) = 0$, $S(\psi_o) = 0$ and $\hat{\psi} - \psi_o = o_{\mathscr{P}^*}(1)$ as elements of $l^\infty(H_p)$. The notation "lin" before a set denotes the set of all finite linear combinations of elements of the set.

THEOREM 2. *Assume the following:*

(a) (*asymptotic distribution of score function*) $\sqrt{n}(S_n(\psi_o) - S(\psi_o)) \Longrightarrow \mathscr{W}$, *where $\mathscr{W}$ is a tight Gaussian process on $l^\infty(H_p)$;*

(b) (*Fréchet differentiability of the asymptotic score*) $S(\hat{\psi}) - S(\psi_o) = -\dot{S}(\psi_o)(\hat{\psi} - \psi_o) + o_{\mathscr{P}^*}(n^{-1/2} \vee \|\hat{\psi} - \psi_o\|)$, *where* $\dot{S}(\psi_o)$: $\mathrm{lin}\{\psi - \psi_o\colon \psi \in \Psi\} \to l^\infty(H_p)$ *is a continuous linear operator;*

(c) (*invertibility*) $\dot{S}(\psi_o)$ *is continuously invertible on its range;*

(d) (*approximation condition*) $\|(S_n - S)(\hat{\psi}) - (S_n - S)(\psi_o)\| = o_{\mathscr{P}^*}(n^{-1/2} \vee \|\hat{\psi} - \psi_o\|)$.

Then,

$$\sqrt{n}(\hat{\psi} - \psi_o) \Longrightarrow \dot{S}(\psi_o)^{-1}\mathscr{W}.$$

This theorem is straightforward to prove. Write

$$\sqrt{n}(S_n(\psi_o) - S(\psi_o)) \\ = \sqrt{n}(S(\psi_o) - S(\hat{\psi})) - \sqrt{n}((S_n - S)(\hat{\psi}) - (S_n - S)(\psi_o)).$$

First the rate of convergence of $(\hat{\psi}-\psi_o)$ follows from the rate of convergence of $(S_n(\psi_o) - S(\psi_o))$ because of the continuous invertibility of $\dot{S}(\psi_o)$ and the Fréchet differentiability condition. This can be seen by taking the norm of both sides and dividing by $\|\hat{\psi} - \psi_o\|$. Use the fact that continuous invertibility of $\dot{S}(\psi_o)$ is equivalent to "$\dot{S}(\psi_o)$ is 1 : 1 on its range and its inverse is a bounded linear operator" and that this implies

$$\inf_{\psi\in\Psi} \frac{\|\dot{S}(\psi_o)(\psi - \psi_o)\|}{\|\psi - \psi_o\|} \geq \varepsilon,$$

for some $\varepsilon > 0$.

Using (b) to solve the above equation for $\dot{S}(\psi_o)(\sqrt{n}(\hat{\psi}-\psi_o))$, we see that we can use Slutsky's theorem [van der Vaart and Wellner (1993)] to prove that $\dot{S}(\psi_o)(\sqrt{n}(\hat{\psi} - \psi_o))$ converges weakly to $\mathscr{W}$. Indeed, we have that

$$\dot{S}(\psi_o)(\sqrt{n}(\hat{\psi} - \psi_o)) = \sqrt{n}(S_n(\psi_o) - S(\psi_o)) + o_{\mathscr{P}^*}(1)$$

as members of $l^\infty(H_p)$. All that is left is to use the continuous mapping theorem to derive the result of the theorem. This is easily done since $\mathscr{W}$ must belong to the closure of the range of $\dot{S}(\psi_o)$, and via the Hahn–Banach theorem we can define a continuous extension of $\dot{S}(\psi_o)^{-1}$ to the closure of the range.

To prove Theorem 1, set $\hat{\psi} = (\hat{A}, \hat{\theta})$, $\psi_o = (A_o, \theta_0)$ and note that the score equation S_n is defined in Section 1. Let $S = (S_1, S_2)$ be the expectation of S_n. In Lemma 1 we verify (a) and (d) for the frailty model. The proof of (a) relies on Pollard's (1984) martingale central limit theorem in $D[0, \tau]$ under the supremum norm metric and the fact that this will imply weak convergence in $l^\infty(H_p)$ for any finite p. The proof of (d) is primarily technical.

The asymptotic distribution of the score function, $\sqrt{n}S_n(A_o, \theta_0)$, is that of a tight Gaussian process in $l^\infty(H_p)$ with

$$\mathrm{Var}(\mathscr{W}(h)) = \int_0^\tau E\biggl[h_1(u) - \frac{\theta_0\int_0^{u-} Y h_1\,dA_o}{1+\theta_0\int_0^{u-} Y\,dA_o} \\ + h_2\biggl(\frac{N(u-)}{1+\theta_0 N(u-)} - \frac{\int_0^{u-} Y\,dA_o}{1+\theta_0\int_0^{u-} Y\,dA_o}\biggr)\biggr]^2 ZY(u)\,dA_o.$$

The above variance can be rewritten as

$$\int_0^\tau \sigma_1(h)(u)\,h_1(u)\,dA_o + \sigma_2(h)\,h_2.$$

So the Fisher information is given by $\sigma = (\sigma_1, \sigma_2)$ and is a linear operator from $L_2(dA_o) \times \mathfrak{R}$ into itself. The classical relationship between the asymptotic variance of the score function (the information for ψ) and the derivative of the score equation, $\dot{S}(\psi_o)$ holds; that is, in Lemma 3 we see that $-\dot{S}(\psi_o)\sqrt{n}(\hat{\psi} - \psi_o)$, evaluated at h, is

$$\int_0^\tau \sigma_1(h)(u)\, d\sqrt{n}(\hat{A} - A_o) + \sigma_2(h)\sqrt{n}(\hat{\theta} - \theta_0).$$

We need to prove continuous invertibility of $\dot{S}(\psi_o)$. Intuitively it is clear that the invertibility of the derivative of the score equation should be closely connected to the invertibility of the Fisher information. The Fisher information is defined in an almost-everywhere sense (dA_o) whereas we will need invertibility everywhere due to the discreteness of $\hat{A}$. However, we will get a start in the proof of invertibility by using the fact that the Fisher information is 1 : 1, that is,

$$\int_0^\tau \sigma_1(h)(u) h_1(u)\, dA_o(u) + \sigma_2(h) h_2 > 0,$$

for all $h = (h_1, h_2) \in L_2(dA_o) \times \mathfrak{R}$ which are nonzero and bounded. In Lemma 2 we prove the above assuming (c) and (d) of Theorem 1. We use the above fact [i.e., σ is 1 : 1 in $L_2(dA_o) \times \mathfrak{R}$] to prove that σ considered as a bounded linear operator from H_∞ to H_∞ is 1 : 1. It turns out that σ is the sum of a continuously invertible linear operator plus a compact operator. Since H_∞ is a Banach space this implies that σ must be continuously invertible. This is a stronger result than continuous invertibility of $\dot{S}(\psi_o)$ and should not be necessary. Lemma 3 gives the details of the above argument in order to verify (b) and (c) of Theorem 2.

Once (b) is verified, we have, for any finite p,

$$\int_0^\tau \sigma_1(h)(u)\, d\sqrt{n}(\hat{A} - A_o) + \sigma_2(h)\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n}(S_n(\psi_o) - S(\psi_o))(h) + o_{\mathscr{P}^*}(1)$$

uniformly in $h \in H_p$. To complete the proof of Theorem 1, all we need to do is verify that the variance of $\dot{S}(\psi_o)^{-1}\mathscr{W}(g)$ is as stated. Note that the continuous invertibility of σ implies that, for finite p, there exists finite q for which $\sigma^{-1}(g) \in H_q$ if $g \in H_p$. To get a weak convergence result for $\hat{\psi}$ at $g \in H_p$, we put $h = \sigma^{-1}(g)$ in the above equation to get

$$\int_0^\tau g_1(u)\, d\sqrt{n}(\hat{A} - A_o) + g_2\sqrt{n}(\hat{\theta} - \theta_0) = \sqrt{n}(S_n(\psi_o) - S(\psi_o))(\sigma^{-1}(g)) + o_{\mathscr{P}^*}(1) \tag{2.1}$$

uniformly in $g \in H_p$. This then shows us that $\dot{S}(\psi_o)^{-1}\mathscr{W}(g)$ is equivalent in distribution to $\mathscr{W}(\sigma^{-1}(g))$ with variance $\int_0^\tau g_1 \sigma_{(1)}^{-1}(g)\, dA_o + g_2 \sigma_{(2)}^{-1}(g)$. This concludes the proof of Theorem 1.

An advantage of proving continuous invertibility of σ is that we can now use the naive approach to estimation of the asymptotic variance of $\hat{\psi} = (\hat{A}, \hat{\theta})$, that is, form the second derivative matrix of L_n by taking derivatives with respect to the jump sizes of $\hat{A}$ and θ (jumps of $\hat{A}$ occur at the jumps of the N_i's). Invert this very large matrix and multiply by -1. Form the vector of $(g_1(u_1), g_1(u_2), \dots, g_2)$, where the u_i's are the locations of jumps of the N_i's. Premultiply and postmultiply the large inverted matrix by this vector to form the estimator of the asymptotic variance of

$$\left(\int_0^\tau g_1(u)\, d\sqrt{n}(\hat{A} - A_o) + g_2\sqrt{n}(\hat{\theta} - \theta_0)\right).$$

This procedure is identical to estimating σ_1 and σ_2, solving $g_1 = \hat{\sigma}_1(h)$ and $g_2 = \hat{\sigma}_2(h)$ and using $\int_0^\tau g_1 \hat{\sigma}^{-1}_{(1)}(g)\, d\hat{A} + g_2 \hat{\sigma}^{-1}_{(2)}(g)$ as the estimator of the asymptotic variance. The estimators of the σ_i's are formed in the obvious fashion:

$$\begin{aligned}\hat{\sigma}_1(h)(u) = h_1(u) n^{-1} \sum_{i=1}^n \hat{Z}_i Y_i(u) - n^{-1} \sum_{i=1}^n \frac{\hat{\theta} \int_0^\tau Y_i h_1\, d\hat{A}}{1 + \hat{\theta}\int_0^\tau Y_1\, d\hat{A}} \hat{Z}_i Y_i(u)\\ - h_2 n^{-1} \sum_{i=1}^n \frac{Y_i(u)}{1 + \hat{\theta}\int_0^\tau Y_i d\hat{A}} \left(\int_0^\tau Y_i\, d\hat{A}\hat{Z}_i - N_i(\tau)\right)\end{aligned}$$

and

$$\hat{\sigma}_2(h) = h_2 \frac{\partial^2 L_n(\theta, A)}{\partial \theta^2}\bigg|_{(\hat{\theta},\hat{A})} - n^{-1} \sum_{i=1}^n \frac{\int_0^\tau h_1 Y_i\, d\hat{A}}{1 + \hat{\theta}\int_0^\tau Y_i\, d\hat{A}} \left(\int_0^\tau Y_i\, d\hat{A}\hat{Z}_i - N_i(\tau)\right),$$

where $\hat{Z}_i = (1 + \hat{\theta} N_i(\tau))/(1 + \hat{\theta}\int_0^\tau Y_i\, d\hat{A})$.

THEOREM 3. *Assume* (a)–(d) *of Theorem* 1. *Then, for* $(g_1, g_2) \in H_p$, *the solution* $h = \hat{\sigma}^{-1}(g)$ *to* $g_1 = \hat{\sigma}_1(h)$, $g_2 = \hat{\sigma}_2(h)$ *exists with probability going to* 1 *as n increases and*

$$\int_0^\tau g_1 \hat{\sigma}^{-1}_{(1)}(g)\, d\hat{A} + g_2 \hat{\sigma}^{-1}_{(2)}(g)$$

converges in probability to $\int_0^\tau g_1 \sigma^{-1}_{(1)}(g)\, dA_o + g_2 \sigma^{-1}_{(2)}(g)$.

This theorem can proved by the following two steps. First $\hat{\sigma}$, which is a continuous linear operator from H_∞ to H_∞, must (with probability going to 1) be one-to-one and onto in order to ensure the existence of $\hat{\sigma}^{-1}(g)$. Next we show that $\sup_{u\in[0,\tau]} |\hat{\sigma}^{-1}_{(1)}(g)(u) - \sigma^{-1}_{(1)}(g)(u)|$ and $|\hat{\sigma}^{-1}_{(2)}(g) - \sigma^{-1}_{(2)}(g)|$ converge to zero in probability. This will imply that $\int_0^\tau g_1 \hat{\sigma}^{-1}_{(1)}(g)\, d\hat{A}$ converges to $\int_0^\tau g_1 \sigma^{-1}_{(1)}(g)\, dA_o$ [using integration by parts and the fact that $\sigma^{-1}_{(1)}(g)$ is of bounded variation].

The proofs of these two steps depend heavily on the fact, from Lemma 3, that σ is continuously invertible and on the nice form of $\hat{\sigma}$, that is, for each n large and with high probability, $\hat{\sigma}$ is the sum of an invertible operator plus a compact operator. The compact operator is compact not only considered as a continuous linear operator under the $\|\cdot\|_H$ norm, but also under the norm $\|\cdot\|_\infty$ defined by $\|h\|_\infty = \sup_{u\in[0,\tau]}|h_1(u)| + |h_2|$. We omit the proof.

APPENDIX

The norm $\|\cdot\|$ on $l^\infty(H_p)$ is equivalent to the larger of the supremum norm on the space of the first component and the absolute value on the space of the second component; for example, $p\|\hat{A} - A_o\|_\infty \vee p|\hat{\theta} - \theta_0| \le \|\hat{\psi} - \psi_o\| \le 2p\|\hat{A} - A_o\|_\infty \vee p|\hat{\theta} - \theta_0|$. In the proofs of the lemmas, this equivalence will be used to replace $\|\hat{\psi} - \psi_o\|$. Another fact that is frequently used is that the expectation of Z given the path of N, Y on $[0,\tau]$ is $(1+\theta_0 N(\tau))/(1+\theta_0\int_0^\tau Y\,dA_o)$.

LEMMA 1. *Assume* (a) *and* (b) *of Theorem* 1. *Then, for any finite p, the following hold:*

(a) $\sqrt{n}(S_n(\psi_o) - S(\psi_o)) \Longrightarrow \mathscr{W}$ *on* $l^\infty(H_p)$*; and*
(b) *for any* δ_n *which decreases to zero as n increases,*

$$\sup_{\|\psi-\psi_o\|<\delta_n} \frac{\|(S_n - S)(\psi) - (S_n - S)(\psi_o)\|}{n^{-1/2}\vee\|\psi-\psi_o\|} = o_{\mathscr{P}^*}(1).$$

PROOF. To prove part (a), note that $\sqrt{n}S_n(\psi_o)(h)$ is equal to $\Phi(Z_n)(h)$, where Φ is a continuous function from $D[0,\tau]\times\Re$ to $l^\infty(H_p)$, defined by

$$\Phi(Z)(h) = Z_1(\tau)h_1(\tau) - Z_1(0)h_1(0) - \int_0^\tau Z_1(u-)dh_1(u) + Z_2h_2$$

and $Z_n = (Z_{n1}, Z_{n2})$,

$$Z_{n1}(t) = n^{-1/2}\sum_{i=1}^n\int_0^\tau\left[g_t(u) - \frac{\theta_0\int_0^{u-}Y_ig_t\,dA_o}{1+\theta_0\int_0^{u-}Y_i\,dA_o}\right]dM_i(u),$$

where $g_t(u)$ is equal to 1 if u is at most t, and zero otherwise. The scalar Z_{n2} is given by

$$n^{-1/2}\sum_{i=1}^n\int_0^\tau\left(\frac{N_i(u-)}{1+\theta_0N_i(u-)} - \frac{\int_0^{u-}Y_i\,dA_o}{1+\theta_0\int_0^{u-}Y_i\,dA_o}\right)dM_i(u).$$

To show that Z_n converges weakly to a continuous Gaussian process in $D[0,\tau]\times\Re$, use Theorem 13 of Pollard [(1984), page 179] to verify the central limit theorem, V3 [Pollard (1984), page 92]. Because the Y_i's and the N_i's are bounded, this is trivial and is omitted. These theorems give asymptotic tightness of Z_n as a process in $D[0,\tau]\times\Re$, under the supremum norm and using the projection sigma field. This translates into weak convergence in outer measure in $D[0,\tau]\times\Re$, under the supremum norm and using the Borel sigma field [Pollard (1990), page 51]. Now use the continuous mapping theorem to get part (a) [Pollard (1990), page 46].

The proof of part (b) relies on functional central limit theorems for

$$n^{-1/2}\sum_{i=1}^{n}[U_iY_i(u)-E(UY(u))] \quad \text{in } l^{\infty}([0,\tau])$$

and

$$n^{-1/2}\sum_{i=1}^{n}[U_iY_i(u)Y_i(v)-E(UY(u)Y(v))] \quad \text{in } l^{\infty}([0,\tau]^2),$$

where the U, U_i's are i.i.d. bounded random variables. The asymptotic tightness is proved via Theorems 1.4.6 and 2.3.4 of van der Vaart and Wellner (1993) and by utilizing the assumption that the Y_i's are left-continuous with right-hand limits with total variation bounded by a constant.

We show that, uniformly over $h \in H_p$ and $\|A-A_o\| < \delta_n$, both $(S_{n1}-S_1)\times(A,\theta)(h)-(S_{n1}-S_1)(A_o,\theta_0)(h)$ and $(S_{n2}-S_2)(A,\theta)(h)-(S_{n2}-S_2)(A_o,\theta_0)(h)$ are $o_{\mathscr{P}^*}(n^{-1/2} \vee \|A-A_o\|_\infty \vee |\theta-\theta_0|)$. We do this by dividing both sums into two parts, to concentrate on the parameters one at a time. To illustrate the proof technique yet not get bogged down in the technical details, we consider here the first term of

$$(S_{n2}-S_2)(A,\theta_0)(h)-(S_{n2}-S_2)(A_o,\theta_0)(h)$$

in the case when θ_0 is nonzero and $A \neq A_o$. This term is given by

$$h_2\left[n^{-1}\sum_{i=1}^{n}\ln\left[\frac{1+\theta_0\int_0^\tau Y_i\,dA}{1+\theta_0\int_0^\tau Y_i\,dA_o}\right]-E\ln\left[\frac{1+\theta_0\int_0^\tau Y\,dA}{1+\theta_0\int_0^\tau Y\,dA_o}\right]\right].$$

In order to show that this term divided by $\|A-A_o\|_\infty$ is $o_{\mathscr{P}^*}(1)$ uniformly in $\|A-A_o\|_\infty < \delta_n$, we first linearize by using the inequality $|\ln(1+x)-x| \le x^2$ for $x > -\frac{1}{2}$. For each i we want to set

$$x = \frac{\theta_0\int_0^\tau Y_i\,d(A-A_o)}{1+\theta_0\int_0^\tau Y_i\,dA_o};$$

this is valid since the fraction goes to zero uniformly over i in probability. The linearization is given by

$$h_2 n^{-1}\sum_{i=1}^{n}\left[\frac{\theta_0\int_0^\tau Y_i\,d(A-A_o)}{1+\theta_0\int_0^\tau Y_i\,dA_o}-E\left[\frac{\theta_0\int_0^\tau Y\,d(A-A_o)}{1+\theta_0\int_0^\tau Y\,dA_o}\right]\right],$$

and the error term is bounded above by

$$\left|h_2 n^{-1}\sum_{i=1}^{n}\left[\frac{\theta_0\int_0^\tau Y_i\,d(A-A_o)}{1+\theta_0\int_0^\tau Y_i\,dA_o}\right]^2\right|+\left|h_2 E\left[\frac{\theta_0\int_0^\tau Y\,d(A-A_o)}{1+\theta_0\int_0^\tau Y\,dA_o}\right]^2\right|.$$

So the error term is bounded above by

$$4p{\theta_0}^2\|A-A_o\|_\infty^2\left[n^{-1}\sum_{i=1}^{n}\|Y_i\|_v^2+E\|Y\|_v^2\right],$$

which divided by $\|A - A_o\|_\infty$ is certainly $o_{\mathscr{P}^*}(1)$ uniformly in $\|A - A_o\|_\infty < \delta_n$ (since $\|Y\|_v$ is a.s. bounded). Consider now the linearization, rewritten as $|h_2\theta_0 \int_0^\tau n^{-1/2} Z_n(u)\, d(A - A_o)(u)|$, where

$$Z_n(u) = n^{-1/2} \sum_{i=1}^{n} \biggl(\frac{Y_i(u)}{1 + \theta_0 \int_0^\tau Y_i \, dA_o} - E\biggl[\frac{Y(u)}{1 + \theta_0 \int_0^\tau Y\, dA_o} \biggr] \biggr).$$

To finish this part of the proof we prove that

$$\text{(A.1)} \qquad \sup_{\|A - A_o\|_\infty < \delta_n} \left| \int_0^\tau Z_n(u) d(A - A_o)(u) \right|$$

converges to zero in probability. To prove this, first note that as mentioned above one can show that Z_n converges weakly to a tight Gaussian process, say, Z_∞ in $l^\infty([0,\tau])$. Define the function ϕ from $l^\infty([0,\tau])$ to $l^\infty(H_q)$ by $\phi(V)(h) = \int_0^\tau V(u)\, dh(u)$. This function is continuous, so by the continuous mapping theorem $\phi(Z_n)$ converges weakly in $l^\infty(H_q)$ to $\phi(Z_\infty)$. Now we note that with probability going to 1, (A.1) is bounded above by a continuous functional of $\phi(Z_n)$, say, k_n, defined by

$$k_n(\phi(V)) = \sup_{\|h\|_\infty < \delta_n, h \in H_q} \left| \int_0^\tau V(u)\, dh(u) \right|.$$

To use the extended continuous mapping theorem [van der Vaart and Wellner (1993)] in order to prove that $k_n(\phi(Z_n)) = o_{\mathscr{P}^*}(1)$, it is sufficient to notice that $k_n(\phi(Z_\infty))$ goes to zero as n increases, that is, we need to verify that the sample paths of $\phi(Z_\infty)$ are uniformly continuous at the point $h = 0$. This is true because, with probability 1, $\phi(Z_\infty)$ has paths which are uniformly continuous with respect to the semimetric defined by its variance function $\rho(h, g) = E[\phi(Z_n)(h - g)]^2$ [see van der Vaart and Wellner (1995)]. However, $\rho(h, g)^2 = E[\int_0^\tau UY - E[UY])\, d(h - g)]^2$, where $U = [1 + \theta_0 \int_0^\tau Y\, dA_o]^{-1}$. Using integration by parts and the fact that the total variation of Y is bounded, one gets that $\sup_{\|h\|_\infty < \delta_n,\ h \in H_q} \rho(h, 0)$ goes to zero as δ_n goes to zero. □

LEMMA 2. *Assume* (c) *and* (d) *of Theorem* 1. *Then, for h_1 a bounded function on $[0, \tau]$ and h_2 a finite real scalar,*

$$\begin{aligned}
&\int_0^\tau \sigma_1(h)(u)\, h_1(u)\, dA_o + \sigma_2(h)\, h_2 \\
&\quad = \int_0^\tau E\biggl[h_1(u) - \frac{\theta_0 \int_0^{u-} Y h_1\, dA_o}{1 + \theta_0 \int_0^{u-} Y\, dA_o} \\
&\qquad\qquad + h_2 \biggl(\frac{N(u-)}{1 + \theta_0 N(u-)} - \frac{\int_0^{u-} Y\, dA_o}{1 + \theta_0 \int_0^{u-} Y\, dA_o} \biggr) \biggr]^2 ZY(u)\, dA_o = 0
\end{aligned}$$

implies that $h_2 = 0$ and $h_1 = 0$ a.s.

PROOF. Note that the left-hand side of the above equation is the variance of $S_{n1}(\psi_o)(h) + S_{n2}(\psi_o)(h)$, so that the theorem states that the score functions are not collinear. This proof is very similar to that used in Murphy (1994) in order to prove identifiability of the frailty model. Indeed, this theorem says that the Fisher information is invertible (maybe not continuously invertible), which functions as a local measure of identifiability.

Define $\tilde{h}_1$ by

$$\tilde{h}_1(t) = E\left[\frac{\theta_0 \int_0^{u-} Y h_1 \, dA_o}{1+\theta_0 \int_0^{u-} Y \, dA_o} - h_2\left(\frac{N(u-)}{1+\theta_0 N(u-)} - \frac{\int_0^{u-} Y \, dA_o}{1+\theta_0 \int_0^{u-} Y \, dA_o}\right)\right]\frac{Y(u)}{EY(u)},$$

for all $u \in [0, \tau]$. Note that $\tilde{h}_1$ is left-continuous with right-hand limits and that $h_1 = \tilde{h}_1$ a.e. (dA_o). It is also useful to note that the set of discontinuities of $\tilde{h}_1$ is countable. The left-continuity of all components involved implies that

$$Y(u)\tilde{h}_1(u) = \left[\frac{\theta_0 \int_0^{u} Y \tilde{h}_1 \, dA_o}{1+\theta_0 \int_0^{u} Y \, dA_o} - h_2\left(\frac{N(u-)}{1+\theta_0 N(u-)} - \frac{\int_0^{u-} Y \, dA_o}{1+\theta_0 \int_0^{u-} Y \, dA_o}\right)\right] Y(u),$$

for every u and a.e. $d\mathscr{P}$. Let T_1 be the first jump of N. Call the set for which the above holds Δ. Intersect Δ with $\{Y(T_1+) \geq 1\}$, $\{Y(T_1) \geq 1\}$ and the set of T_1 not a member of the discontinuities of $\tilde{h}_1$. Since Δ and the last two sets have probability 1, the intersection has positive probability by assumption (d). On this intersection, we have that

$$\tilde{h}_1(T_1) = \frac{\theta_0 \int_0^{T_1} Y \tilde{h}_1 \, dA_o}{1+\theta_0 \int_0^{T_1} Y \, dA_o} + h_2\left(\frac{\int_0^{T_1} Y \, dA_o}{1+\theta_0 \int_0^{T_1} Y \, dA_o}\right)$$

and

$$\tilde{h}_1(T_1+) = \frac{\theta_0 \int_0^{T_1} Y \tilde{h}_1 \, dA_o}{1+\theta_0 \int_0^{T_1} Y \, dA_o} - h_2\left(\frac{1}{1+\theta_0} - \frac{\int_0^{T_1} Y \, dA_o}{1+\theta_0 \int_0^{T_1} Y \, dA_o}\right).$$

Since $\tilde{h}_1(T_1) = \tilde{h}_1(T_1+)$, this will imply that $h_2 = 0$.

We start the proof over again, but this time with $h_2 = 0$. We have

$$\tilde{h}_1(u)Y(u) = \frac{\theta_0 \int_0^{u} Y \tilde{h}_1 \, dA_o}{1+\theta_0 \int_0^{u} Y \, dA_o} Y(u),$$

for all u and a.e. $d\mathscr{P}$. This implies that

$$\tilde{h}_1(u)EY(u) + \theta_0 \tilde{h}_1(u) \int_0^u EY(v)Y(u) \, dA_o = \theta_0 \int_0^u EY(v)Y(u)\tilde{h}_1(v) \, dA_o,$$

for all $u \in [0, \tau]$. If $\tilde{h}_1$ can be shown to be identically equal to 0, then the proof will be done. The supremum and infimum of $\tilde{h}_1$ on $[0, \tau]$ are either attained at a point or attained by evaluating a right-hand limit of $\tilde{h}_1$ at a point. For simplicity assume the former; the proof is similar if the latter holds. Suppose the maximum value of $\tilde{h}_1$ is attained at $t_{\max}$ and $\tilde{h}_1(t_{\max}) \geq 0$. Set $u = t_{\max}$ to get $\tilde{h}_1(t_{\max}) = 0$. Next suppose that the minimum value of $\tilde{h}_1$ is attained at $t_{\min}$ and $\tilde{h}_1(t_{\min}) \leq 0$. Set $u = t_{\min}$ to get $\tilde{h}_1(t_{\min}) = 0$. This gives us that $h_1 = 0$ a.e. (dA_o).

LEMMA 3. *Assume* (a)–(d) *of Theorem* 1. *Then, for any finite p, the following hold:*

(a) *as* $\|\psi - \psi_o\|$ *goes to zero,* $S(\psi) - S(\psi_o) = \dot{S}(\psi_o)(\psi - \psi_o) + o(\|\psi - \psi_o\|)$, *where* $\dot{S}(\psi_o)$: $\text{lin}\{\psi - \psi_o\colon \psi \in \Psi\} \to l^\infty(H_p)$, *is a continuous linear operator;*

(b) $\dot{S}(\psi_o)$ *has a continuous inverse on its range.*

PROOF. From the equations of S_n, we see that $S_1(A,\theta)$ and $S_2(A,\theta)$ must have the following forms:

$$S_1(A,\theta)(h) = \int_0^\tau h_1 E(ZY)\,dA_o - E\left(\frac{1+\theta N(\tau)}{1+\theta\int_0^\tau Y\,dA}\int_0^\tau h_1 Y\,dA\right)$$

and

$$\begin{aligned} &S_2(A,\theta)(h)\\ &\quad = h_2\int_0^\tau E\left(\frac{N(u-)}{1+\theta N(u-)}ZY(u)\right)dA_o\\ &\qquad + h_2 I\{\theta = 0\}E\left[\frac{1}{2}\left(\int_0^\tau Y\,dA\right)^2 - N(\tau)\int_0^\tau Y\,dA\right]\\ &\qquad + h_2 I\{\theta \neq 0\}\theta^{-2}E\left[\ln\left(1+\theta\int_0^\tau Y\,dA\right) - \frac{1+\theta N(\tau)}{1+\theta\int_0^\tau Y\,dA}\theta\int_0^\tau Y\,dA\right]. \end{aligned}$$

It is clear that $-\dot{S}(\psi_o)(\psi_o)(h)$ should be the variance of $\sqrt{n}S_n(\psi_o)(h)$, which is in turn equal to the variance of $\mathscr{W}(h)$ as given in Section 2. So we aim for this in deriving $\dot{S}$. The idea will be first to write $S(A,\theta)$ linearly in $d(A - A_o)$ and $\theta - \theta_0$ plus error terms, then to use integration by parts to write $\dot{S}(\psi_o)(\psi - \psi_o)(h)$ in a similar way to the form of $\text{Var}(\mathscr{W}(h))$ in Section 2.

Using the fact that the expectation of Z given the paths of N and Y on $[0,\tau]$ is $(1+\theta_0 N(\tau))/(1+\theta_0\int_0^\tau Y\,dA_o)$, one can with some algebraic manipulation get

$$\begin{aligned} S_1(A,\theta)(h) = &-E\int_0^\tau h_1 ZY\,d(A - A_o)\\ &+ E\frac{\int_0^\tau h_1 Y\,dA_o}{(1+\theta_0\int_0^\tau Y\,dA_o)}\theta_0\int_0^\tau ZY\,d(A - A_o)\\ &+ (\theta - \theta_0)E\frac{\int_0^\tau h_1 Y\,dA_o}{1+\theta_0\int_0^\tau Y\,dA_o}\left(\int_0^\tau ZY\,dA_o - N(\tau)\right)\\ &+ \text{error}_1(A,\theta)(h) \end{aligned}$$

and

$$
\begin{aligned}
&S_2(A,\theta)(h)\\
&\quad = h_2(\theta_0-\theta)E\bigg[\int_0^\tau\bigg(\frac{N(u-)}{1+\theta_0 N(u-)}\bigg)^2 ZY(u)\,dA_o\\
&\qquad\qquad + \int_0^\tau \frac{2(\int_0^u Y\,dA_o)^2}{(1+\theta_0\int_0^u Y\,dA_o)^3} Y(u)\,dA_o\\
&\qquad\qquad\qquad - N(\tau)\bigg(\frac{\int_0^\tau Y\,dA_o}{1+\theta_0\int_0^\tau Y\,dA_o}\bigg)^2\bigg]\\
&\qquad + h_2 E\frac{\int_0^\tau Y\,d(A-A_o)}{(1+\theta_0\int_0^\tau Y\,dA_o)}\bigg(\int_0^\tau ZY\,dA_o - N(\tau)\bigg) + \mathrm{error}_2(A,\theta)(h).
\end{aligned}
$$

The error terms are very easily shown to satisfy

$$
\sup_{h\in H_p} \frac{|\mathrm{error}_i(A,\theta)(h)|}{\|A-A_o\|_\infty \vee |\theta-\theta_0|} \longrightarrow 0
$$

as $\|A-A_o\|_\infty \vee |\theta-\theta_0|$ goes to zero. This follows from the boundedness of N, Y, A and θ. Note that $S_1(\hat{A},\hat{\theta}) + S_2(\hat{A},\hat{\theta})(h) - \mathrm{error}_1(\hat{A},\hat{\theta})(h) - \mathrm{error}_2(\hat{A},\hat{\theta})(h)$ is equal to $-(\int_0^\tau \sigma_1(h)\,d(\hat{A}-A_o) + \sigma_2(h)(\hat{\theta}-\theta_0))$, where σ_1 and σ_2 are defined in Section 2.

To verify part (b), we must show that, for some p and $\varepsilon > 0$,

$$
\inf_{\psi\in \mathrm{lin}\,\Psi} \frac{\sup_{h\in H_p} |\int_0^\tau \sigma_1(h)\,d\psi_1 + \sigma_2(h)\psi_2|}{2p\|\psi_1\|_\infty \vee p|\psi_2|} > \varepsilon.
$$

To do this we will prove that σ, viewed as a linear operator from H_∞ to H_∞, is onto and continuously invertible. This implies that, for some $q > 0$, $\sigma^{-1}(H_q)$ is contained in H_p. Then we note that the quotient above is no smaller than

$$
\inf_{\psi\in \mathrm{lin}\,\Psi} \frac{\sup_{h\in \sigma^{-1}(H_q)} |\int_0^\tau \sigma_1(h)d\psi_1 + \sigma_2(h)\psi_2|}{2p\|\psi_1\|_\infty \vee p|\psi_2|},
$$

which is equivalent to

$$
\inf_{\psi\in \mathrm{lin}\,\Psi} \frac{\sup_{h\in H_q} |\int_0^\tau h_1\,d\psi_1 + h_2\psi_2|}{2p\|\psi_1\|_\infty \vee p|\psi_2|}.
$$

This, however, is larger than $q/2p$.

Recall that for $h \in H_\infty$ the norm of h is defined by $\|h\|_H = \|h_1\|_v + |h_2|$. It is straightforward to show that σ is a bounded linear operator from H_∞ to H_∞. To prove continuous invertibility, we write σ as the sum of a continuously invertible linear operator plus a compact operator, and we show, using Lemma 2, that σ is one-to-one. Since H_∞ is a Banach space this will imply that σ is continuously invertible with range H_∞ [see Rudin (1973), pages 99–103].

To begin, we show that σ is one-to-one, that is, for $h \in H_\infty$, with $\|h\|_H > 0$, we have that $\|\sigma(h)\|_H > 0$. If this were not the case, then both $\sigma_1(h)(u) = 0$

for all u and $\sigma_2(h) = 0$. By Lemma 2, we get that $h_1(u) = 0$ a.e. (dA_o) and $h_2 = 0$. Now we have that, for $h = (h_1, 0)$,

$$\sigma_1(h)(u) = h_1(u)EZY(u) - E\frac{\theta_0 \int_0^\tau Y h_1\, dA_o}{1 + \theta_0 \int_0^\tau Y\, dA_o} ZY(u) = 0 \quad \text{for all } u.$$

Since $EZY(u) > 0$, we get that h_1 is identically zero. This is of course a contradiction.

Write $\sigma(h)$ as the sum of two linear operators. The first linear operator is $\Sigma(h) = (h_1(u)EZY(u), h_2 \int_0^\tau EK^2(u)ZY(u)\, dA_o)$, where

$$K(u) = \frac{N(u-)}{1 + \theta_0 N(u-)} - \frac{\int_0^u Y\, dA_o}{1 + \theta_0 \int_0^u Y\, dA_o},$$

and the second is $\sigma(h) - \Sigma(h)$. The inverse of Σ is

$$\Sigma^{-1}(h)(u) = \Big(h_1(u)(EZY(u))^{-1},\ h_2\Big(\int_0^\tau EK^2(u)ZY(u)\, dA_o\Big)^{-1}\Big).$$

Because $\inf_{u \in [0,\tau]} EZY(u) > 0$ and $\int_0^\tau EK^2(u)ZY(u)\, dA_o > 0$, Σ^{-1} is a bounded linear operator. This follows from assumptions (b), (c) and (d) of Theorem 1.

All that is left is to show that $\Sigma(h) - \sigma(h)$ is compact. Let $\{h_n\}_{n \geq 1}$ be a sequence in H_1. We must prove that there exists a convergent subsequence of $\sigma(h_n) - \Sigma(h_n)$. Since h_{n1} is of bounded variation, we can write h_{n1} as the difference of increasing functions. Both of these increasing functions are bounded in absolute value by at most 2. This means that we can use Helly's selection theorem to find a pointwise convergent subsequence. Let h^* be the limit of the convergent subsequence of $\{h_n\}_{n \geq 1}$. We must prove that the same subsequence of $\sigma(h_n) - \Sigma(h_n)$ converges to $\sigma(h^*) - \Sigma(h^*)$ in norm. This follows by the dominated convergence theorem and assumption (b). To illustrate this, consider the term

$$E\frac{\theta_0 \int_0^\tau Y h_{1n}\, dA_o}{1 + \theta_0 \int_0^\tau Y\, dA_o} ZY(u)$$

in $\sigma(h_n) - \Sigma(h_n)$. Then

$$\left\| E\frac{\theta_0 \int_0^\tau Y(h_{1n} - h_1^*)\, dA_o}{1 + \theta_0 \int_0^\tau Y\, dA_o} ZY \right\|_v < \theta_0 \int_0^\tau |h_{1n} - h_1^*| E\frac{Y|Z|\, \|Y\|_v}{1 + \theta_0 \int_0^\tau Y\, dA_o}\, dA_o$$

which for the subsequence will converge to zero by the dominated convergence theorem. □

Acknowledgment. The author is indebted to Aad van der Vaart, Mark van der Laan and Joel Zinn for many fruitful conversations.

REFERENCES

AALEN, O. O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine* **7** 1121–1137.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

BREMAUD, P. (1981). *Point Processes and Queues*. Springer, New York.

CLAYTON, D. and CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *J. Roy. Statist. Soc. Ser. A* **148** 82–117.

GREENWOOD, P. E. and WEFELMEYER, W. (1990). Efficiency of estimators for partially specified filtered models. *Stochastic Process. Appl.* **36** 353–370.

HECKMAN, J. and SINGER, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52** 271–320.

MANTON, K. G., STALLARD, E. and VAUPEL, J. W. (1986). Alternate models for the heterogeneity of mortality risks among the aged. *J. Amer. Statist. Assoc.* **81** 635–644.

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

NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. and SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19** 25–44.

OAKES, D. (1991). Frailty models for multiple event times In *Survival Analysis: State of the Art*, (J. P. Klein and P. K. Goel, eds.) 371–379. Kluwer, Dordrecht.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. IMS, Hayward, CA.

RAO, R. R. (1963). The law of large numbers for $D[0,1]$-valued random variables. *Theory Probab. Appl.* **8** 70–74.

RUDIN, W. (1973). *Functional Analysis*. McGraw-Hill, New York.

SELF, S. G. and PRENTICE, R. L. (1986). Incorporating random effects into multivariate relative risk regression models. In *Modern Statistical Methods in Chronic Disease Epidemiology* (S. H. Moolgavkar and R. L. Prentice, eds.) 167–177. Wiley, New York.

VAUPEL, J. W. (1990). Kindred lifetimes: frailty models in population genetics. In *Convergent Issues in Genetics and Demography* (J. Adams, D. A. Lam, A. I. Herman and P. E. Smouse, eds.) 156–170. Oxford Univ. Press.

VAUPEL, J. W., MANTON, K. G. and STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16** 439–454.

VAN DER VAART, A. (1992). Efficiency of infinite dimensional m-estimators. Unpublished manuscript.

VAN DER VAART, A. and WELLNER, J. A. (1995). *Weak Convergence and Empirical Processes*. Springer, Berlin. To appear.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802-2111