*The Annals of Statistics*
1993, Vol. 21, No. 3, 1648–1662

# SEMIPARAMETRIC ESTIMATION OF ASSOCIATION IN A BIVARIATE SURVIVAL FUNCTION

BY GANGAJI MAGULURI

*Rutgers University*

Clayton's model for association in bivariate survival data is both of intrinsic importance and an interesting example of a semiparametric estimation problem, that is, a problem where inference about a parameter is required in the presence of nuisance functions. The joint distribution of the two survival times in this model is absolutely continuous and a single parameter governs the association between the two survival times. In this paper we describe an algorithm to derive the asymptotic lower bound for the information of the parameter governing the association. We discuss the construction of one-step estimators and compare their performance to that of other estimators in a Monte Carlo study.

**1. Introduction.** Many models are possible for association between two nonnegative random variables $S$ and $T$ with an arbitrary continuous joint distribution. One such model is the Clayton's (1978) model. While studying familial tendency in disease incidence, Clayton (1978) proposed a bivariate survival function as a way of representing association between two survival times. Let $(S, T)$ be a pair of survival times with joint survival function

$$(1.1)\quad F(s,t;\theta) = \left[\left\{\frac{1}{G(s)}\right\}^{\theta-1} + \left\{\frac{1}{H(t)}\right\}^{\theta-1} - 1\right]^{-1/(\theta-1)}, \qquad \theta > 1,$$

where $G(s) = \Pr(S > s)$ and $H(t) = \Pr(T > t)$ are the marginal survivor functions of $S$ and $T$. Also the joint distribution is (right) continuous in $\theta$ at $\theta = 1$, which corresponds to independence of the two survival times.

The preceding model has gained wide acceptance in analyzing bivariate survival data, partly due to a random effect interpretation. The association between $S$ and $T$ is explained by their common dependence on an unobservable gamma random variable (frailty) through a proportional hazard structure. For details, see Clayton (1978) or Oakes (1982).

Estimation of $\theta$ in this semiparametric model is an interesting problem and has been considered by several authors. Given a fully parametric specification of the marginals, inference about $\theta$ is in principle straightforward. Oakes (1982) discussed such parametric inference for the case when both $G$ and $H$

Received February 1989; revised February 1992.
*AMS* 1991 *subject classifications*. 62G05, 62G20.
*Key words and phrases*. Alternating projections, efficient score, information bounds, operators.



Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

are exponential and derived explicitly the Fisher information matrix in the absence of censoring.

An alternate approach, which has been termed semiparametric [Oakes (1982) and Begun, Hall, Huang and Wellner (1983)], is to make inferences about $\theta$ in a way that does not depend on the form of the marginal distributions, that is, to treat $G$ and $H$ as nuisance functions.

Clayton (1978) proposed a method for estimating $\theta$ in the absence of knowledge of the marginal distributions. However, his pseudo-likelihood is not completely clear and its asymptotic properties were later questioned by Oakes (1982). As an alternative, Oakes (1982) proposed a simpler estimator called the concordance estimator, which is based on Kendall's coefficient of concordance, and evaluated its asymptotic variance. Clayton and Cuzick (1985) have also suggested various approximations to the likelihood of the ranks. These involve replacing the observations by expected values of order statistics and maximizing a parametric likelihood. They also gave a representation of Clayton's (1978) estimator as a weighted form of Oakes' concordance estimator, whose limiting variance was derived explicitly by Oakes (1986a), in the absence of censoring. All these estimators are consistent and can be used as preliminary estimates in finding an efficient estimator.

The main goal of the present paper is to calculate the information bounds for estimation of $\theta$ in the presence of the unknown nuisance parameters $G$ and $H$. If $G$ and $H$ are known, then any regular estimator of $\theta$ would have a limiting distribution at least as dispersed as $N(0, 1/I_\theta)$, where $I_\theta$ is the usual parametric Fisher information for $\theta$. On the other hand, if $G$ and $H$ are unknown the information for $\theta$ will, in general, be smaller and it will be harder to estimate $\theta$. The natural questions that arise in such a case are, how much smaller the information is likely to be and how much larger the asymptotic variance of a "best" estimate would be.

In the present paper, we study these questions. We show that the asymptotic lower bounds for estimation of $\theta$, when $G$ and $H$ are unknown, are determined by the geometry of the scores. In fact, the bivariate model which we treat here is just one example of a large class of semiparametric models in which the efficient score or the information involves a projection on a subspace of a Hilbert space with a sum space structure. When the subspaces involved in forming the sum spaces are orthogonal, explicit formulas are usually possible since the projection on the sum space is then the sum of the individual projections. However, when orthogonality fails (as it does in the present model), explicit formulas are often not available, and we have to use the alternating projections algorithm for calculating the projection on this sum space. The alternating projections algorithm for calculating a projection on a sum subspace of a Hilbert space is originally due to von Neumann (1949, 1950). Appendix A.4 of Bickel, Klaassen, Ritov and Wellner (1993) gives a treatment suited to semiparametric models.

Alternating projection methods have received considerable interest and attention in statistics within the past few years in connection with nonparametric (additive) regression and correlation; see, for example, Breiman and Friedman (1985), Buja (1985) and Buja, Hastie and Tibshirani (1989).

In Section 2, we describe this algorithm in detail. In Section 3, we propose an estimator which intuitively seems natural and which might be efficient. Efficiency is not proved however. One possible explanation for the lack of a satisfactory solution to this problem is that the influence function of any efficient estimator cannot be calculated in closed form. We compare the performance of this estimator with that of Oakes' concordance estimator and Clayton's weighted concordance estimator in a Monte Carlo study.

**2. Information calculations for θ.** In this section we will give a detailed description of the algorithm to compute the efficient score and efficient information for the association parameter $\theta$ in the model (1.1) in the absence of censoring. To do this, we begin with the calculation of the scores for both the parametric and the nonparametric components of the model.

Suppose $(S, T)$ has joint density $f$ given by

$$f(s,t) = \frac{\partial^2}{\partial s\,\partial t} F(s,t) = \frac{\theta g(s) h(t)}{\{G(s)H(t)\}^{\theta}\{D(s,t)\}^{2+1/(\theta-1)}}, \tag{2.1}$$

where

$$D(s,t) = \left[\left\{\frac{1}{G(s)}\right\}^{\theta-1} + \left\{\frac{1}{H(t)}\right\}^{\theta-1} - 1\right]. \tag{2.2}$$

Then

$$\begin{aligned} \log f(s,t) &= \log\theta - \theta\log[G(s)H(t)] + \log[g(s)h(t)] \\ &\quad - \left(2 + \frac{1}{\theta-1}\right)\log D(s,t). \end{aligned} \tag{2.3}$$

A straightforward calculation yields the score for $\theta$ to be

$$\begin{aligned} \dot{l}_{\theta}(s,t) &\equiv \frac{\partial}{\partial\theta}\log f(s,t) \\ &= \frac{1}{\theta} - \log[G(s)H(t)] + \frac{1}{(\theta-1)^2}\log D(s,t) \\ &\quad + \left(2 + \frac{1}{\theta-1}\right)\frac{1}{D(s,t)}\left\{\frac{\log G(s)}{G(s)^{\theta-1}} + \frac{\log H(t)}{H(t)^{\theta-1}}\right\}. \end{aligned} \tag{2.4}$$

Similarly, let $\{g_\eta \colon \eta \in \mathbb{R}^1\}$ and $\{h_\gamma \colon \gamma \in \mathbb{R}^1\}$ be regular parametric families with $g_0 = g_\eta$ and $h_0 = h_\gamma$ and let $f_{\eta,\gamma}$ be the corresponding regular parametric subfamily of the original model. Then we have

$$\frac{\partial}{\partial\eta}\log f_{\eta,\gamma}(s,t) = a(s) + \frac{1}{G(s)}\left[\theta - (2\theta-1)\frac{G(s)^{-\theta+1}}{D(s,t)}\right]\int_s^\infty a\,dG,$$

where $a \equiv (\partial/\partial\eta)\log g_\eta$. A similar relation holds for the $\gamma$ score. This leads us to the so-called score operators

$$\dot{l}_g a(s,t) = a(s) + \frac{1}{G(s)}\left[\theta - (2\theta - 1)\frac{G(s)^{-\theta+1}}{D(s,t)}\right]\int_s^\infty a\,dG, \tag{2.5}$$
$$\dot{l}_h b(s,t) = b(t) + \frac{1}{H(t)}\left[\theta - (2\theta - 1)\frac{H(t)^{-\theta+1}}{D(s,t)}\right]\int_t^\infty b\,dH,$$

where $b \equiv (\partial/\partial\gamma)\log h_\gamma$. Note from (2.5) that $\dot{l}_g$ defines a liner transformation from functions of $s$ to functions of $(s,t)$ and $\dot{l}_h$ from functions of $t$ to functions of $(s,t)$. Let $L_2(G)$, $L_2(H)$ and $L_2(F)$ be the usual $L_2$ spaces of square integrable functions. Let $L_2^0(G) = \{a \in L_2(G)\colon \int a\,dG = 0\}$ and $L_2^0(H) = \{b \in L_2(H)\colon \int b\,dH = 0\}$. Define the linear operators $R_g\colon L_2(G) \to L_2^0(G)$ and $R_h\colon L_2(H) \to L_2^0(H)$ by

$$R_g a(s) = a(s) + \frac{\int_s^\infty a\,dG}{G(s)}, \tag{2.6}$$
$$R_h b(t) = b(t) + \frac{\int_t^\infty b\,dH}{H(t)}.$$

Then $\dot{l}_g$ and $\dot{l}_h$ can be expressed as

$$\dot{l}_g a(s,t) = a(s) - \left[\theta - (2\theta - 1)\frac{G(s)^{-\theta+1}}{D(s,t)}\right]\big(a(s) - R_g a(s)\big), \tag{2.7}$$
$$\dot{l}_h b(s,t) = b(t) - \left[\theta - (2\theta - 1)\frac{H(t)^{-\theta+1}}{D(s,t)}\right]\big(b(t) - R_h b(t)\big).$$

Now it is easy to see that $\dot{l}_g$ and $\dot{l}_h$ are linear transformations from $L_2^0(G)$ to $L_2(F)$ and $L_2^0(H)$ to $L_2(F)$, respectively.

LEMMA . *$\dot{l}_g$ and $\dot{l}_h$ are bounded operators.*

PROOF. The boundedness of $\dot{l}_g$ follows from the boundedness of $R_g$ [for a proof see Bickel, Klaassen, Ritov and Wellner (1993)], and the inequality $G(s)^{-\theta+1} \le D(s,t)$. The same argument holds for $\dot{l}_h$. □

Let us define

$$\mathscr{H} = \overline{\left\{\dot{l}_g a + \dot{l}_h b\colon a \in L_2^0(G),\, b \in L_2^0(H)\right\}},$$
$$\mathscr{H}_1 = \operatorname{range}\big(\dot{l}_g\big) \quad\text{and}\quad \mathscr{H}_2 = \operatorname{range}\big(\dot{l}_h\big).$$

To compute the efficient score function and information for $\theta$, we need to compute the projection of $\dot{l}_\theta$ onto the orthocomplement of $\mathscr{H}$ or onto $\mathscr{H}$ itself. But $\mathscr{H}$ is the subspace of the Hilbert space $L_2(F)$ with a sum space structure and the subspaces $\mathscr{H}_1$ and $\mathscr{H}_2$ involved in forming this sum space structure are not orthogonal. As a result, the projection of $\dot{l}_\theta$ onto $\mathscr{H}$ cannot be computed in closed form and we have to use the alternating projections algorithm. To do this, we have to first calculate the projections of $\dot{l}_\theta$ onto $\mathscr{H}_1$ and $\mathscr{H}_2$.

We first show that $\mathscr{H}_1$ is closed. This implies that the projection on $\mathscr{H}_1$ exists and can be computed in terms of the projection operator

$$\dot{l}_g\left(\dot{l}_g^T\dot{l}_g\right)^{-1}\dot{l}_g^T,$$

where $\dot{l}_g^T$ is the adjoint of $\dot{l}_g$. Since $\dot{l}_g$ is a bounded linear operator from $L_2^0(G)$ to $L_2(F)$ and $L_2^0(G)$ is complete, it follows from the following proposition that $\mathscr{H}_1$ is closed [cf. Corollary A.1.2 of Bickel, Klaassen, Ritov and Wellner (1993)]. (Similar results hold for $\mathscr{H}_2$ and the projection operator on $\mathscr{H}_2$.)

PROPOSITION 2.1. *$\dot{l}_g^{-1}$ exists on $\mathscr{H}_1$ and is bounded.*

PROOF. An easy calculation shows that for $Q(s,t) = \dot{l}_g a(s,t)$,

$$\begin{aligned}\dot{l}_g^{-1}Q(s) &= E[Q(S,T)|S=s]\\ &= a(s) - \left[a(s) - R_g a(s)\right]\int_0^\infty\left[\theta - (2\theta-1)\frac{G(s)^{-\theta+1}}{D(s,t)}\right]\frac{f(s,t)}{g(s)}\,dt\\ &= a(s),\end{aligned}$$

since the integral on the right-hand side of the above expression is zero. It now follows that $\dot{l}_g^{-1}$ exists and is bounded. □

We next show that the sum space $\mathscr{H}_1 + \mathscr{H}_2$ is closed. We will need this to compute the projection of $\dot{l}_\theta$ onto $\mathscr{H} = \overline{\mathscr{H}_1 + \mathscr{H}_2} = \mathscr{H}_1 + \mathscr{H}_2$. We define two functions $\alpha$ and $\gamma$ as follows:

$$\begin{aligned}\alpha(s) &= E\left[\frac{1}{G^2(S)}\left\{\theta - (2\theta-1)\frac{G(S)^{-\theta+1}}{D(S,T)}\right\}^2\middle|S=s\right],\\ \gamma(t) &= E\left[\frac{1}{H^2(T)}\left\{\theta - (2\theta-1)\frac{H(T)^{-\theta+1}}{D(S,T)}\right\}^2\middle|T=t\right].\end{aligned}\tag{2.8}$$

An easy calculation shows that

$$\begin{aligned}\alpha(s) &= cG(s)^{-2},\\ \gamma(t) &= cH(t)^{-2},\end{aligned}\tag{2.9}$$

where $c \equiv \theta(\theta-1)^2/(3\theta-2)$.

PROPOSITION 2.2.

$$\rho(\mathscr{H}_1, \mathscr{H}_2) \equiv \sup\left\{\langle \dot{l}_g, \dot{l}_h\rangle : \|\dot{l}_g a\| = \|\dot{l}_h b\| = 1\right\} \leq \left(\frac{4c}{1+4c}\right)^{1/2} < 1,$$

*and hence* $\mathscr{H}_1 + \mathscr{H}_2$ *is closed.*

PROOF. Note that, by definition of $\dot{l}_g a$,

$$\begin{aligned}
\langle \dot{l}_g a, \dot{l}_h b\rangle &= E\left[\dot{l}_g a \dot{l}_h b\right] \\
&= E\left[a(S)\dot{l}_h b(S,T)\right] \\
&\quad + E\left[\frac{1}{G(S)}\left\{\theta - (2\theta-1)\frac{G(S)^{-\theta+1}}{D(S,T)}\right\}\int_S^\infty a\,dG\,\dot{l}_h b(S,T)\right] \qquad \text{(A1)}\\
&= E\left[\frac{1}{G(s)}\left\{\theta - (2\theta-1)\frac{G(s)^{-\theta+1}}{D(s,t)}\right\}\int_S^\infty a\,dG\,\dot{l}_h b(S,T)\right],
\end{aligned}$$

since

$$E\left[a(S)\dot{l}_h b(S,T)\right] = E\left[a(S)E\left[\dot{l}_h b(S,T)|S=s\right]\right] = E[a(S)0] = 0.$$

From (A1) it follows that, for $a, b$ with $\|\dot{l}_g a\| = 1 = \|\dot{l}_h b\|$,

$$\begin{aligned}
\left|\langle \dot{l}_g a, \dot{l}_h b\rangle\right| &\leq E\left[\frac{1}{G(S)^2}\left\{\theta - (2\theta-1)\frac{G(S)^{-\theta+1}}{D(S,T)}\right\}^2\right. \qquad \text{(A2)}\\
&\qquad\qquad \left.\times\left(\int_S^\infty a\,dG\right)^2\right]^{1/2} \equiv \sqrt{M^2},
\end{aligned}$$

where

$$1 = \|\dot{l}_g a\|^2 = E\left[\dot{l}_g a\right]^2 = \int_0^\infty a^2\,dG + M^2, \qquad \text{(A3)}$$

since

$$\begin{aligned}
E\left[\dot{l}_g a\right]^2 &= \operatorname{Var}\left[\dot{l}_g a\right] \\
&= \operatorname{Var}\left[E\left(\dot{l}_g a|S\right)\right] + E\left[\operatorname{Var}\left(\dot{l}_g a|S\right)\right] \\
&= \operatorname{Var}[a(S)] \qquad \text{(A4)}\\
&\quad + E\left\{E\left[\left(\frac{1}{G(S)}\left\{\theta - (2\theta-1)\frac{G(S)^{-\theta+1}}{D(S,T)}\right\} \times \int_S^\infty a\,dG\right)^2\right]\middle| S = s\right\},
\end{aligned}$$

in view of

$$E\left[\dot{l}_g a(S,T)\middle|S\right] = a(S) \quad \text{(see Proposition 2.1)}.$$

Thus

$$\text{(A5)} \qquad \begin{aligned} E\left[\dot{l}_g a\right]^2 &= \int_0^\infty a^2\,dG + E\left\{\left(\frac{1}{G(S)}\int_S^\infty a\,dG\right)^2 G(S)^2 \alpha(S)\right\} \\ &\le (1+4c)\int_0^\infty a^2\,dG \end{aligned}$$

by (2.9) and from Hardy's inequality. From (A3) and (A5) it follows that

$$\text{(A6)} \qquad \int_0^\infty a^2\,dG \ge \frac{1}{1+4c}.$$

Thus from (A3), we obtain

$$\text{(A7)} \qquad M^2 = 1 - \int_0^\infty a^2\,dG \le 1 - \frac{1}{1+4c} = \frac{4c}{1+4c}$$

and the inequality of the proposition follows from (A2) and (A7). Closedness of $\mathscr{H}_1 + \mathscr{H}_2$ is implied by Aronszajn [(1950), pages 375–380]; compare Theorem A.4.2 of Bickel, Klaassen, Ritov and Wellner (1993). □

Now we turn to the study of the projection operator

$$P_1 = \Pi(\cdot|\mathscr{H}_1) = \dot{l}_g\left(\dot{l}_g^T\dot{l}_g\right)^{-1}\dot{l}_g^T.$$

From (2.7) we have (dropping the subscript $g$ of $R_g$)

$$\dot{l}_g a(s,t) = a(s) - \left[\theta - (2\theta-1)\frac{G(s)^{-\theta+1}}{D(s,t)}\right](a(s) - Ra(s)).$$

A straightforward calculation shows that

$$\text{(2.10)} \qquad \int_0^s Ra\,\frac{dG}{G} = -\frac{\int_s^\infty a\,dG}{G(s)} = a(s) - Ra(s),$$

so that we can write

$$\text{(2.11)} \quad \dot{l}_g a(s,t) = Ra(s) - \left[(\theta-1) - (2\theta-1)\frac{G(s)^{-\theta+1}}{D(s,t)}\right]\int_0^s Ra\,\frac{dG}{G}.$$

This equation shows that $\dot{l}_g$ is the composition of two bounded linear operators $L$ and $R$, where $L$: range($R$) $\to L_2(F)$ is given by

$$\text{(2.12)} \quad La(s,t) = a(s) - \left[(\theta-1) - (2\theta-1)\frac{G(s)^{-\theta+1}}{D(s,t)}\right]\int_0^s a\,\frac{dG}{G},$$

for any function $a$ in the range space of $R$. Therefore, we have

$$P_1 = \dot{l}_g\left(\dot{l}_g^T \dot{l}_g\right)^{-1}\dot{l}_g^T = LR(R^T L^T L R)^{-1} R^T L^T.$$

One may verify $R^T = R^{-1}$; compare Proposition A.1.8 of Bickel, Klaassen, Ritov and Wellner (1993). Thus $P_1$ reduces to $L(L^T L)^{-1} L^T$. So in order to determine $P_1$, we need to find the forms of $L^T$, $L^T L$ and $(L^T L)^{-1}$ if it exists.

PROPOSITION 2.3. *The adjoint* $L^T$: $L_2(F) \to \text{range}(R)$ *of* $L$ *and* $L^T L$ *are given by*

$$(2.13)\qquad \begin{aligned} L^T a(s) &= \int_0^\infty a(s,t)\frac{f(s,t)}{g(s)}\,dt \\ &\quad + G(s)\int_s^\infty \int_0^\infty a(u,t)k(u,t) f(u,t)\,dt\,du, \end{aligned}$$

*where*

$$k(s,t) = (\theta - 1) - (2\theta - 1)\frac{G(s)^{-\theta+1}}{D(s,t)},$$

*and*

$$(2.14)\qquad L^T L a(s) = a(s) - c\int_0^s a\,\frac{dG}{G} - c\frac{1}{G(s)}\int_s^\infty a\,dG,$$

*with*

$$c \equiv \frac{\theta(\theta - 1)^2}{3\theta - 2} > 0.$$

PROOF. Follows from straightforward calculations. □

We next show that $(L^T L)^{-1}$ exists and compute the form of $(L^T L)^{-1}$.

PROPOSITION 2.4. $(L^T L)^{-1}$ *exists and is bounded.*

PROOF. From (2.14) we have

$$(2.15)\qquad Ja(s) \equiv L^T L a(s) = a(s) - c\int_0^s a\,\frac{dG}{G} - c\frac{1}{G(s)}\int_s^\infty a\,dG.$$

Also from the extended version of (2.10) and the fact that $R^T = R^{-1}$, it follows that

$$(2.16)\qquad R^T a(s) = a(s) + \int_0^s a\,\frac{dG}{G},$$

so that

$$\begin{aligned} Ja(s) &= a(s) - cR^T a(s) - cRa(s) + 2ca(s) \\ &= (1+2c)a(s) - c(R^T a(s) + Ra(s)) \\ &= (1+2c)\left[I - d(R^T + R)\right]a(s), \end{aligned}$$

where

$$d = \frac{c}{1+2c}.$$

Since $\|R^T\| = \|R\| = 1$, $\|R^T + R\| \le 2$ and hence $d\|R^T + R\| \le 2c/(1+2c) < 1$. Therefore, it follows that $J^{-1}$ exists and we can write

$$J^{-1}a(s) = \frac{1}{1+2c}\sum_{j=0}^{\infty} d^j (R^T + R)^j a(s),$$

which shows that

$$\|J^{-1}\| \le \frac{1}{1+2c}\sum_{j=0}^{\infty}(2d)^j = 1,$$

and hence $J^{-1}$ is bounded. □

THEOREM 2.1. *The inverse operator* $(L^T L)^{-1}$ *is given by*

$$\begin{aligned} (L^T L)^{-1}\beta(s) &= \beta(s) + \Delta G(s)^{-1/2-\delta}\int_s^{\infty} G(u)^{-1/2+\delta}\beta(u)g(u)\,du \\ &\quad + \Delta G(s)^{-1/2+\delta}\int_0^{s} G(u)^{-1/2-\delta}\beta(u)g(u)\,du \qquad (2.17) \\ &\quad - \frac{(1/2+\delta)^2}{2\delta}G(s)^{-1/2+\delta}\int_0^{\infty} G(u)^{-1/2+\delta}\beta(u)g(u)\,du, \end{aligned}$$

*where*

$$\Delta = \frac{1/4 - \delta^2}{2\delta} = -\frac{c}{2\delta}.$$

PROOF. To compute $(L^T L)^{-1}$ we proceed as follows: Let $x \equiv G(s)$, $\tilde{a}(x) \equiv a(G^{-1}(x))$, $\tilde{g}(x) \equiv g(G^{-1}(x))$ and $\beta(x) \equiv L^T L\tilde{a}(x)$. With this transformation, we obtain from (2.14)

$$\beta(x) = \tilde{a}(x) + c\int_x^1 \frac{\tilde{a}(w)}{w}\,dw + c\frac{1}{x}\int_0^x \tilde{a}(w)\,dw. \qquad (2.18)$$

Writing $A(u) \equiv \int_0^u \tilde{a}\,dx$, $B(u) \equiv \int_0^u \beta\,dx$, this becomes by differentiation [by writing $\int_0^x \tilde{a}(u)\,du = -\int_x^1 \tilde{a}(u)\,du$]

$$A''(x) - cx^{-2}A(x) = B''(x). \qquad (2.19)$$

We want to solve this equation subject to $A(0) = A(1) = 0$ [since $A(u) \equiv \int_0^u \tilde{a}(x)\,dx$]. The general solution of the homogeneous part of (2.19) has the form

$$A(x) = c_1 A_1(x) + c_2 A_2(x),$$

where

$$A_1(x) = x^{1/2+\delta}, \qquad A_2(x) = x^{1/2-\delta}$$

with

$$\delta = \sqrt{\tfrac{1}{4} + c} > \tfrac{1}{2}.$$

The Wronskian

$$A_1(x)A_2'(x) - A_1'(x)A_2(x) = -2\delta$$

is nontrivial and, consequently, we can form the Green's function

$$G(x,t) \equiv \begin{cases} -\dfrac{1}{2\delta} t^{1/2+\delta} x^{1/2-\delta} = -\dfrac{1}{2\delta} A_1(t) A_2(x), & 0 \le t \le x, \\ -\dfrac{1}{2\delta} t^{1/2-\delta} x^{1/2+\delta} = -\dfrac{1}{2\delta} A_2(t) A_1(x), & x \le t \le 1. \end{cases} \tag{2.20}$$

A particular solution of (2.19) is then given by

$$\begin{aligned} A_P(x) &= \int_0^1 G(x,t) B''(t)\,dt \\ &= -\frac{1}{2\delta}\left\{ x^{1/2-\delta} \int_0^x t^{1/2+\delta} B''(t)\,dt + x^{1/2+\delta} \int_x^1 t^{1/2-\delta} B''(t)\,dt \right\}. \end{aligned} \tag{2.21}$$

Now

$$A(x) = A_P(x) + c_1 A_1(x) + c_2 A_2(x) \tag{2.22}$$

is a general solution of (2.19); we want to choose the constants $c_1, c_2$ such that $A(0) = A(1) = 0$.

Note from (2.21) and (2.20) that $A_P(0) = 0$. Consequently, $A(0) = c_2 A_2(0)$ implies $c_2 = 0$ and $A(1) = A_P(1) + c_1 A_1(1) = 0$ gives

$$c_1 = \frac{1}{2\delta} \int_0^1 t^{1/2+\delta} B''(t)\,dt.$$

Hence

$$\begin{aligned} A(x) = {} & \frac{(1/2 + \delta)}{2\delta} x^{1/2-\delta} \int_0^x t^{-1/2+\delta} B'(t)\,dt \\ & + \frac{(1/2 - \delta)}{2\delta} x^{1/2+\delta} \int_x^1 t^{-1/2-\delta} B'(t)\,dt \\ & - \frac{(1/2 + \delta)}{2\delta} x^{1/2+\delta} \int_0^1 t^{-1/2+\delta} B'(t)\,dt \end{aligned} \tag{2.23}$$

satisfies the differential equation (2.19) subject to the boundary conditions

$A(0) = A(1) = 0$. Thus, we have

$$(L^TL)^{-1}\beta(x) = \beta(x) + \Delta\left\{x^{-1/2-\delta}\int_0^x t^{-1/2+\delta}\beta(t)\,dt\right.$$

$$\left.+x^{-1/2+\delta}\int_x^1 t^{-1/2-\delta}\beta(t)\,dt\right\} \tag{2.24}$$

$$-\frac{(1/2+\delta)^2}{2\delta}x^{-1/2+\delta}\int_0^1 t^{-1/2+\delta}\beta(t)\,dt.$$

Now transforming back to the original variable $s$, we get the required form (2.17). □

Thus once we have the forms for $L^T$ and $(L^TL)^{-1}$, we can compute the projection of $\dot{l}_\theta$ onto $\mathscr{H}_1$ by using the formula $P_1 = L(L^TL)^{-1}L^T$.

Since we now know how to compute the projections of $\dot{l}_\theta$ onto $\mathscr{H}_1$ and $\mathscr{H}_2$, we can obtain the projection of $\dot{l}_\theta$ onto the space $\mathscr{H} = \overline{\mathscr{H}_1 + \mathscr{H}_2} = \mathscr{H}_1 + \mathscr{H}_2$ by the method of alternating projections which we next describe.

Let $a_i \equiv P_i(\dot{l}_\theta) = \Pi(\dot{l}_\theta|\mathscr{H}_i)$, $i = 1, 2$, where $\Pi$ is the projection operator. Set

$$a_2^{(1)} \equiv P_2\dot{l}_\theta = a_2$$

and proceed inductively: for $m \ge 1$, set

$$a_1^{(m)} \equiv P_1\big(\dot{l}_\theta - a_2^{(m)}\big) = a_1 - P_1a_2^{(m)},$$

$$a_2^{(m+1)} \equiv P_2\big(\dot{l}_\theta - a_1^{(m)}\big) = a_2 - P_2a_1^{(m)}.$$

This is known as "back-fitting" in regression. When the projection operators $P_1$ are conditional expectation operators, this is just the "inner loop" of the "alternating conditional expectation" or ACE algorithm of Breiman and Friedman (1985). The fact that $\|a_1^{(m)} + a_2^{(m)} - \Pi(\dot{l}_\theta|\mathscr{H})\| \to 0$ as $m \to \infty$ is due to von Neumann. Compare Theorem A.4.2 of Bickel, Klaassen, Ritov and Wellner (1993).

The efficient score $\dot{l}_\theta^*$ and the efficient information $I_\theta^*$ for $\theta$ are then given by

$$\dot{l}_\theta^* = \dot{l}_\theta - \Pi\big(\dot{l}_\theta|\mathscr{H}\big) \tag{2.25}$$

and

$$I_\theta^* = E\big[\dot{l}_\theta^*\big]^2. \tag{2.26}$$

We therefore have an algorithm to compute the efficient score and information for $\theta$ in the presence of nuisance functions $G$ and $H$.

**3. Construction of efficient estimates.** In this section, we will discuss the construction of one-step estimates for $\theta$ and present some numerical results. The investigation in this section is a preliminary attempt at the problem of constructing optimal estimates; asymptotic normality of these estimators still remains to be proved.

Let $(S_1, T_1), (S_2, T_2), \dots, (S_n, T_n)$ be iid with density function $f(\cdot, \cdot; \theta, G, H)$ given by (2.1). Let $\mathbb{F}_n$ be the empirical measure of $(S_i, T_i)$'s. Let $\hat{G}_n$ and $\hat{H}_n$ be some consistent estimators of $G$ and $H$. Then defining $\hat{\theta}_n$ as a root of the efficient score equation

$$(3.1)\quad \int\int \dot{l}_\theta^*\big(s, t; \theta, \hat{G}_n, \hat{H}_n\big)\, d\mathbb{F}_n(s, t) = \frac{1}{n} \sum_{i=1}^{n} \dot{l}_\theta^*\big(s_i, t_i; \theta, \hat{G}_n, \hat{H}_n\big) = 0,$$

we should have

$$(3.2)\qquad \hat{\theta}_n = \theta + [I_\theta^*]^{-1} \frac{1}{n} \sum_{i=1}^{n} \dot{l}_\theta^*(s_i, t_i; \theta, G, H) + o_p(n^{-1/2}).$$

We will now describe the steps involved in the construction of a one-step approximation to the above estimator.

STEP 1. Consider a preliminary consistent estimator $\tilde{\theta}_n$ of $\theta$, for example, Oakes' concordance estimator.

STEP 2. Compute the marginal empirical survivor functions $G_n$ and $H_n$ from the sample. Clearly $G_n$ and $H_n$ are consistent estimators of $G$ and $H$, respectively.

STEP 3. Using $\tilde{\theta}_n$, $G_n$ and $H_n$ obtain estimates $\hat{P}_1$ and $\hat{P}_2$ for the projections of $\dot{l}_\theta$ onto $\mathscr{H}_1$ and $\mathscr{H}_2$.

STEP 4. Estimate the score $\dot{l}_\theta$ for $\theta$ using $\tilde{\theta}_n$, $G_n$ and $H_n$.

STEP 5. Using $\hat{P}_1$ and $\hat{P}_2$ in the iterative algorithm of the alternating projections obtain an estimate of the projection of $\dot{l}_\theta$ onto $\mathscr{H}$.

STEP 6. Compute the estimated efficient score function $\hat{\dot{l}}_\theta^*$ and estimated information $\hat{I}_\theta^*$ with the estimated score $\hat{\dot{l}}_\theta$ and the estimated projection of the score onto $\mathscr{H}$.

STEP 7. Replacing $\theta$, $\dot{l}_\theta^*$ and $I_\theta^*$ by their estimated values, calculate the one-step estimate of $\theta$ as

$$\hat{\theta}_n = \tilde{\theta}_n + \Big[\hat{I}^*\big(\tilde{\theta}_n, G_n, H_n\big)\Big]^{-1} \frac{1}{n} \sum_{i=1}^{n} \hat{\dot{l}}_\theta^*\big(s_i, t_i; \tilde{\theta}_n, G_n, H_n\big).$$

*Monte Carlo numerical results.* To illustrate and compare our estimator with others, we present a summary of the Monte Carlo results obtained by

TABLE 1

| $\theta$ | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ |
|---|---|---|---|---|
| 1.3 | 1.131 (0.178) | 1.312 (0.162) | 1.310 (0.161) | 1.313 (0.162) |
| 1.5 | 1.527 (0.223) | 1.522 (0.202) | 1.519 (0.202) | 1.523 (0.202) |
| 1.6 | 1.628 (0.236) | 1.620 (0.218) | 1.617 (0.218) | 1.621 (0.218) |
| 2.0 | 2.027 (0.291) | 2.029 (0.276) | 2.025 (0.273) | 2.030 (0.276) |
| 2.5 | 2.521 (0.379) | 2.516 (0.360) | 2.513 (0.360) | 2.518 (0.359) |
| 3.0 | 3.012 (0.443) | 3.004 (0.399) | 3.002 (0.399) | 3.004 (0.396) |
| 3.5 | 3.575 (0.526) | 3.559 (0.484) | 3.557 (0.484) | 3.556 (0.479) |
| 4.0 | 4.075 (0.603) | 4.071 (0.586) | 4.069 (0.587) | 4.064 (0.579) |
| 4.5 | 4.623 (0.695) | 4.624 (0.661) | 4.623 (0.662) | 4.610 (0.649) |
| 5.0 | 5.103 (0.766) | 5.122 (0.704) | 5.122 (0.705) | 5.103 (0.690) |

simulating bivariate data with unit exponential marginals. We briefly describe the simulation procedure as follows:

For different values of $\theta$ we simulated gamma random numbers $W_i$ from a gamma distribution with parameters $(1/(\theta - 1), 1)$. Next generating two independent sets of uniform random numbers $U_1$ and $U_2$ and using the relations

$$S = \frac{1}{\lambda(\theta - 1)} \log\left[1 - \frac{1}{W} \log U_1\right],$$

$$T = \frac{1}{\mu(\theta - 1)} \log\left[1 - \frac{1}{W} \log U_2\right],$$

we generated bivariate data $(S_i, T_i)$.

At each value of $\theta$ we generated 500 samples each of sample size $n = 100$. Then following the steps described above we obtained one-step estimates for $\theta$. We also computed estimates using Clayton's (weighted concordance) and Oakes' (concordance) methods. The means and standard deviations of the 500 estimates for each value of $\theta$ are presented in Table 1. We used the following notation for different estimators of $\theta$:

$\hat{\theta}_1$ = concordance estimator,

$\hat{\theta}_2$ = Clayton's (weighted concordance) estimator,

$\hat{\theta}_3$ = one-step estimator,

$\hat{\theta}_4$ = MLE.

The quantities in the parentheses are the standard deviations. Based on this simulation study, it appears that both the weighted concordance (Clayton's) estimator and the one-step estimator have the same level of performance. They both seem to be reasonable compared to the concordance estimator. Since the asymptotic variance of the one-step estimator cannot be computed in closed form, it is hard to say about its asymptotic relative efficiency with respect to the other estimators.

**Acknowledgments.** This paper is based on my doctoral dissertation. I am grateful to my advisor, Professor Jon A. Wellner, for his generous guidance and encouragement. I also would like to thank an Associate Editor and the referees for their helpful suggestions.

## REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1965). *Handbook of Mathematical Functions*. Dover, New York.

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

BEGUN, J. M. (1980). Estimation of relative risk in proportional hazards model. Ph.D. dissertation, Univ. Rochester.

BEGUN, J. M., HALL, W. J., HUANG, W. M. and WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.

BERAN, R. (1977). Estimating a distribution function. *Ann. Statist.* **5** 400–404.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press.

BOYCE, W. E. and DIPRIMA, R. C. (1977). *Elementary Differential Equations*. Wiley, New York.

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* **80** 580–619.

BUJA, A. (1985). Theory of bivariate ACE. Technical Report, 74, Dept. Statistics, Univ. Washington.

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear smoothers and additive models (with discussion). *Ann. Statist.* **17** 453–555.

CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** 141–151.

CLAYTON, D. G. and CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *J. Roy. Statist. Soc. Ser. A* **148** 82–108.

COX, D. R. and OAKES, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.

DAVIS, H. F. (1963). *Fourier Series and Orthogonal Functions*. Allyn and Bacon, Boston.

GRADSHTEYN, I. S. and RYZHIK, I. M. (1965). *Table of Integrals, Series and Products*, 4th ed. Academic, New York.

HARDY, G. H., LITTLEWOOD, J. E. and POLYA, G. (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press.

HECKMAN, J. and SINGER, B. (1984). The identificability of the proportional hazards model. *Rev. Econom. Stud.* **51** 231–241.

HUANG, W. M. (1982). Parameter estimation in the presence of nuisance functions. Ph.D. dissertation, Univ. Rochester.

LAWLESS, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

MILLER, R. G. (1981). *Survival Analysis*. Wiley, New York.

OAKES, D. (1982). A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B* **44** 414–422.

OAKES, D. (1986a). Semi-parametric inference in a model for association in bivariate survival data. *Biometrika* **73** 353–361.

OAKES, D. (1986b). A model for bivariate survival data. In *Modern Statistical Methods in Chronic Disease Epidemiology* (R. L. Prentice and S. H. Moolgavkar, eds.) 151–166. Wiley, New York.

POLLARD, D. (1980). The minimum distance method of testing. *Metrika* **27** 43–70.

PYKE, R. (1970). Asymptotic results for rank statistics. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 21–37. Cambridge Univ. Press.

REED, M. and SIMON, B. (1980). *Functional Analysis*. Academic, New York.

RUDIN, W. (1974). *Real and Complex Analysis*. McGraw-Hill, New York.

TRICOMI, F. G. (1955). *Integral Equations*. Dover, New York.

VON NEUMANN, J. (1949). On rings of operators. Reduction theory. *Ann. of Math.* **50** 401–485.

VON NEUMANN, J. (1950). *Functional Operators* **2**. *The Geometry of Orthogonal Spaces*. Princeton Univ. Press.

WELLNER, J. A. (1982). Asymptotic optimality of the product limit estimator. *Ann. Statist.* **10** 595–602.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903