The Annals of Statistics
1995, Vol. 23, No. 2, 376–395

ESTIMATING THE REAL PARAMETER IN A TWO-SAMPLE PROPORTIONAL ODDS MODEL

BY COLIN O. WU

Johns Hopkins University

This paper considers efficient estimation of the Euclidean parameter θ in the proportional odds model $G(1-G)^{-1} = \theta F(1-F)^{-1}$ when two independent i.i.d. samples with distributions F and G, respectively, are observed. The Fisher information $I(\theta)$ is calculated based on the solution of a pair of integral equations which are derived from a class of more general semiparametric models. A one-step estimate is constructed using an initial $\sqrt{N}$-consistent estimate and shown to be asymptotically efficient in the sense that its asymptotic risk achieves the corresponding minimax lower bound.

1. Introduction. Transformation models, including the well-known *proportional hazard model*, the *proportional odds model* and so forth, have been widely used in survival analysis, reliability theory and medical and epidemiological studies. Examples and applications of these models can be found in Cox (1972, 1975) and McCullagh and Nelder (1989). In the semiparametric framework, Wellner (1985) and Bickel (1986) considered a class of *transformation models* which adopted the following form: Let X and Y be two random variables on $(0,\infty)$ with cumulative distribution functions F and G, respectively, and let f and g be the corresponding densities with respect to Lebesgue measure. Let $X_1,\dots,X_m$ and $Y_1,\dots,Y_n$ be two independent i.i.d. samples of X and Y, respectively. Then F and G satisfy such a model if there is a nonnegative transformation $B(\theta,u)$ on $(0,\infty)\times[0,1]$ such that f and g satisfy

$$g(t) = B(\theta, F(t))f(t) \quad \text{for all } t>0 \text{ and some } \theta>0, \tag{1}$$

where $\int_0^1 B(\theta,u)\,du = 1$ for all $\theta\in(0,\infty)$.

An important subclass of (1) is the *generalized proportional odds-rate model* considered by Clayton and Cuzick (1986), Dabrowska and Doksum (1988) and Dabrowska, Doksum and Miura (1989), among others. Define $\Lambda_T(t|c)$ to be the *generalized odds ratio* for any random variable T such that

$$\Lambda_T(t|c) = \begin{cases} c^{-1}\left[(1-Q(t))^{-c}-1\right], & t>0, c>0,\\ -\log(1-Q(t)), & t>0, c=0,\end{cases}$$

where Q is the c.d.f. of T. Then, F and G satisfy the *generalized proportional*

Received March 1992; revised January 1994.
AMS 1991 *subject classifications*. 62F10, 62F12, 62G20, 62G05.
Key words and phrases. Semiparametric models, proportional odds model, generalized proportional odds-rate model, asymptotic efficiency, one-step estimate.

odds-rate model if there is a known constant $c \geq 0$ such that

$$\Lambda_Y(t|c) = \theta\Lambda_X(t|c) \quad \text{for all } t > 0 \text{ and some } \theta > 0. \tag{2}$$

It is straightforward to verify that (1) reduces to (2) when the transformation B is defined by

$$B_c(\theta, u) = \begin{cases} \theta\left[\theta - (\theta - 1)(1 - u)^c\right]^{-(c+1)/c}, & \text{if } c > 0, \\ \theta(1 - u)^{\theta - 1}, & \text{if } c = 0. \end{cases} \tag{3}$$

In particular, when $c = 0$, (2) yields the proportional hazard model

$$1 - G(t) = [1 - F(t)]^{\theta} \quad \text{for all } t > 0 \text{ and some } \theta > 0. \tag{4}$$

When $c = 1$, (2) gives the proportional odds model

$$\frac{G(t)}{1 - G(t)} = \theta\frac{F(t)}{1 - F(t)} \quad \text{for all } t > 0 \text{ and some } \theta > 0. \tag{5}$$

By (1) and (3), if $g_{\theta, f}$ denotes the density of Y based on the parameter (θ, f), (5) is equivalent to

$$g_{\theta, f}(t) = \theta[1 + (\theta - 1)F(t)]^{-2}f(t) \quad \text{for all } t > 0 \text{ and some } \theta > 0. \tag{6}$$

For c other than 0 or 1, the interpretations of $\Lambda_T(t|c)$ and model (2) were also discussed by Dabrowska and Doksum (1988).

In most applications of model (2), the relative risk θ is the principal parameter of interest. When F belongs to a regular parametric family, the estimation of θ follows the framework of parametric estimation where the usual procedures, such as the maximum likelihood estimate, Le Cam's one-step MLE, give asymptotically efficient estimators of θ, in the sense of asymptotic efficiency corresponding to either the Hájek–Le Cam convolution theorem or the local asymptotic minimax theorem (LAM) [see Le Cam (1972, 1986) or Millar (1979, 1983)]. When the structure of F is unknown, one reasonable nonparametric model for F is the family of all probability distribution functions on the real line. The estimation of θ follows the semiparametric framework with F as an infinite dimensional nuisance parameter. Theory and methods of constructing efficient estimators of θ under semiparametric models were developed by a number of authors, such as Koshevnik and Levit (1976), Levit (1975, 1978) and Begun, Hall, Huang and Wellner (1983), where the asymptotic efficiency of estimating θ was determined by the convolution theorem or the LAM theorem of the corresponding least favorable parametric submodels. For the special case of the proportional hazard model (4), θ can be efficiently estimated by the Cox partial maximum likelihood estimate or the two-step procedure of Begun and Wellner (1983). For model (2), Dabrowska and Doksum (1988) considered a class of estimators of θ defined by

$$\hat{\theta} = \frac{\int_0^\infty W(\hat{F}_m(t))[1 - \hat{G}_n(t)]^{-(c+1)} \, d\hat{G}_n(t)}{\int_0^\infty W(\hat{F}_m(t))[1 - \hat{F}_m(t)]^{-(c+1)} \, d\hat{F}_m(t)},$$

where $\hat{F}_m$ and $\hat{G}_n$ are left-continuous empirical distributions based on $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$, respectively, and $W(\cdot)$ is some score function. Under some mild regularity conditions, Dabrowska and Doksum showed that $(m + n)^{1/2}(\hat{\theta} - \theta)$ has an asymptotically normal distribution and that for $\theta = 1$,

$$W(u) = W_0(u) = (c + 1)(1 - u)^{2c+1}$$

corresponds to a fully efficient estimate. However, for $\theta \neq 1$, the efficient score function $W(\cdot)$ remains unknown.

With a slightly different sampling scheme, Klaassen (1989) considered the problem of estimating θ under a particular type of the transformation models (1), and derived a second order differential equation whose solution determines the influence function of an efficient estimator. His result shows the existence of a solution for this equation, and that efficient estimation of θ is possible in the Clayton and Cuzick semiparametric Pareto model. However, there are two obstacles to extend Klaassen's approach directly to the problem of estimating θ of (2) from samples $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$. First, since (2) is different from the models he considered, a new differential equation for estimating θ under the two-sample model (2) needs to be developed. Second, the estimation procedure he suggested relies on solving the differential equation numerically, and does not lead to an explicit form of an efficient estimator except for the proportional hazard model. Therefore, it is worthwhile to consider other approaches so that more practical and explicit efficient estimators can be obtained.

We consider in this paper the special case of the two-sample proportional odds model (5) and show that an efficient estimator of θ can be constructed based on the solution of a pair of integral equations. Because of the special structure of the proportional odds model, the solution of our equations, which determines the Fisher information of θ, can be obtained in a closed form. Thus an asymptotically efficient estimate of θ is constructed explicitly. It is also important to note that our integral equations can be extended to the transformation models (1), and their solutions should lead to the Fisher information of θ under many different models of this class. In particular, some routine computation shows that the solution of the integral equations under the proportional hazard model leads to the results of Efron (1977) and Begun and Wellner (1983). However, since we do not have explicit solutions for models (1) or (2), efficient estimates of θ for these general cases are not constructed in this paper.

The main results are given in Section 2, where we present a LAM theorem for the proportional odds model and propose an estimate of θ based on a one-step procedure for which the locally asymptotic maximum risk achieves the corresponding minimax lower bound. In Section 3, we solve the integral equations for the proportional odds model (5) explicitly and derive the closed form of the Fisher information of θ. The proofs of the main technical results, the LAM theorem and asymptotic efficiency of the one-step estimate are presented in Section 4.

2. The main results. This section summarizes our main technical results on efficient estimation of θ for the proportional odds model (5). Based on a preliminary result on Hellinger differentiability, we first derive a minimax lower bound for the asymptotic risk of any estimator of θ, and then show that this minimax bound is attainable by a one-step estimator. The proofs of Theorems 2.1 and 2.2 are deferred to later sections.

2.1. *Hellinger differentiability.* Let $\|\cdot\|_\mu$ be the usual L_2-norm with respect to any sigma finite measure μ on the real line and let $\|\cdot\|$ be the L_2-norm with respect to Lebesgue measure. Given any fixed distribution F_0 with density f_0 on $(0,\infty)$ and $\theta_0 > 0$, let $g_0(x) = g_{\theta_0, f_0}(x)$ satisfy (6) and let G_0 be the corresponding c.d.f. of g_0. Denote by $\mathbf{H}$ the collection of all mean zero functions of $L_2(F_0)$, that is,

$$\mathbf{H} = \left\{h(x): x \in (0,\infty), \int_0^\infty h(x)\, dF_0(x) = 0, \int_0^\infty h^2(x)\, dF_0(x) < \infty\right\},$$

$\mathscr{F}(h)$ the family of all density sequences $\{f_m\}$ on $(0,\infty)$ such that

$$\lim_{m\to\infty} \left\|\sqrt{m}\left(f_m^{1/2} - f_0^{1/2}\right) - \tfrac{1}{2} h f_o^{1/2}\right\| = 0 \quad \text{for some } h \in \mathbf{H} \tag{7}$$

and $\mathscr{F}$ the union of all $\{\mathscr{F}(h): h \in \mathbf{H}\}$. It is easily seen that $\mathbf{H}$ is a subspace of $L_2(F_0)$. Throughout this paper, we assume that $N = m + n$,

$$\lim_{n\to\infty} (m/N) = \lambda \quad \text{and} \quad \lim_{N\to\infty} (n/N) = 1 - \lambda,$$

where $0 < \lambda < 1$ is a constant. Given any $\{f_m\} \in \mathscr{F}$ with corresponding c.d.f.'s F_m and $\{\theta_N \in (0,\infty): \sqrt{N}|\theta_N - \theta_0| \le \alpha\}$, where $0 < \alpha < \infty$, define $\{g_{\theta_N, f_m}\}$ to be a sequence of densities on $(0,\infty)$ satisfying (6).

The next lemma shows the Hellinger differentiability of g_{θ_N, f_m} at (θ_0, f_0). For the general definition of Hellinger differentiability, see Begun, Hall, Huang and Wellner (1983).

LEMMA 2.1. *Given any $\theta_0 > 0$, f_0 and g_0, suppose that $\{f_m\} \in \mathscr{F}(h)$ for some $h \in \mathbf{H}$, $\{\theta_N\}$ satisfies $\sqrt{N}|\theta_N - \theta_0| < \infty$ and $\theta_N > 0$, $N \ge 1$, and $g_N = g_{\theta_N, f_m}$ satisfies* (6). *Then*

$$\lim_{N\to\infty} \frac{\left\|g_N^{1/2} - g_0^{1/2} - (1/2)\left((\theta_N - \theta_0)\xi_{\theta_0, F_0} g_0^{1/2} + m^{-1/2} A h\right)\right\|}{|\theta_N - \theta_0| + m^{-1/2}} = 0,$$

where A is a bounded linear operator given by

$$(8)\qquad Ah(t) = h(t)g_0^{1/2}(t) - \frac{2(\theta_0 - 1)g_0^{1/2}(t)}{1 + (\theta_0 - 1)F_0(t)} \int_0^t h(s)\, dF_0(s)$$

and

$$(9)\qquad \xi_{\theta, F}(t) = \frac{1}{\theta} - \frac{2F(t)}{1 + (\theta - 1)F(t)}.$$

REMARK. 2.1. It necessarily follows from the above differentiability that $\xi_{\theta_0, F_0} g_0^{1/2} \perp g_0^{1/2}$ and $Ah \perp g_0^{1/2}$ in $L_2(\mu)$, that is, $\int_0^\infty \xi_{\theta_0, F_0}(t)\, dG_0(t) = 0$ and $\int_0^\infty Ah(t) g_0^{1/2}(t)\, dt = 0$ for any $h \in \mathbf{H}$ and for which there exists $\{f_m\} \in \mathscr{F}(h)$.

PROOF. Using $F_m(t) - F_0(t) = \int_0^t (f_m^{1/2}(s) - f_0^{1/2}(s))(f_m^{1/2}(s) + f_0^{1/2}(s))\, ds$ and the Cauchy–Schwarz inequality, we can show that

$$(10)\qquad \lim_{N\to\infty} \sup_{t\in R} \left| \sqrt{m}\,(F_m(t) - F_0(t)) - \int_0^t h(s)\, dF_0(s) \right| = 0.$$

Next we notice the following inequality:

$$(11)\qquad \begin{aligned} &\sqrt{N} \left\| g_{\theta_N, f_m}^{1/2} - g_0^{1/2} - \tfrac{1}{2}(\theta_N - \theta_0)\xi_{\theta_0, F_0} g_0^{1/2} - \tfrac{1}{2} m^{-1/2} Ah \right\| \\ &\quad \le \sqrt{N} \left\| g_{\theta_N, f_m}^{1/2} - g_{\theta_0, f_m}^{1/2} - \tfrac{1}{2}(\theta_N - \theta_0)\xi_{\theta_0, F_m} g_{\theta_0, f_m}^{1/2} \right\| \\ &\quad\quad + \tfrac{1}{2}\sqrt{N} |\theta_N - \theta_0| \left\| \xi_{\theta_0, F_m} g_{\theta_0, f_m}^{1/2} - \xi_{\theta_0, F_0} g_0^{1/2} \right\| \\ &\quad\quad + \sqrt{N} \left\| g_{\theta_0, f_m}^{1/2} - g_0^{1/2} - \tfrac{1}{2} m^{-1/2} Ah \right\|. \end{aligned}$$

Notice further that both $B_0^{1/2}(\theta, F_m) = \theta^{1/2}[1 + (\theta - 1)F_m]^{-1}$ and its partial derivative with respect to θ are uniformly bounded in F_m and in θ within a neighborhood of θ_0. Then, for the first term on the right-hand side of (11), we have that [see Le Cam and Yang (1990), Lemma 6.3.1]

$$\begin{aligned} &\sqrt{N} \left\| g_{\theta_N, f_m}^{1/2} - g_{\theta_0, f_m}^{1/2} - \frac{1}{2}(\theta_N - \theta_0)\xi_{\theta_0, F_m} g_{\theta_0, f_m}^{1/2} \right\| \\ &\quad = \sqrt{N} \left\{ \int \left[B_0^{1/2}(\theta_N, F_m(t)) - B_0^{1/2}(\theta_0, F_m(t)) \right.\right. \\ &\quad\quad \left.\left. - (\theta_N - \theta_0) \frac{\partial}{\partial\theta} B_0^{1/2}(\theta, F_m(t))|_{\theta=\theta_0} \right]^2 dF_m(t) \right\}^{1/2} \\ &\quad = \sqrt{N} \left\{ \int \left[B_0^{1/2}(\theta_N, u) - B_0^{1/2}(\theta_0, u) \right.\right. \\ &\quad\quad \left.\left. - (\theta_N - \theta_0) \frac{\partial}{\partial\theta} B_0^{1/2}(\theta, u)|_{\theta=\theta_0} \right]^2 du \right\}^{1/2} \\ &\quad \to 0 \quad \text{as } N \to \infty. \end{aligned}$$

For the second term, since $(\partial/\partial\theta) B_0^{1/2}(\theta, F_m)$ is also continuous in F_m, (10) implies that

$$\| \xi_{\theta_0, F_m} g_{\theta_0, f_m}^{1/2} - \xi_{\theta_0, F_0} g_0^{1/2} \| \to 0 \quad \text{as } N \to \infty.$$

For the third term, we first observe that

$$\frac{\partial}{\partial u} B_0^{1/2}(\theta_0, u) = -\frac{\theta_0^{1/2}(\theta_0 - 1)}{(1 + (\theta_0 - 1)u)^2}$$

and

$$\begin{aligned}
&\sqrt{N} \left\| g_{\theta_0, f_m}^{1/2} - g_0^{1/2} - \frac{1}{2} m^{-1/2} A h \right\| \\
&\quad \le \sqrt{N} \left\| \left(B_0^{1/2}(\theta_0, F_m) - B_0^{1/2}(\theta_0, F_0) \right) \left(f_m^{1/2} - f_0^{1/2} \right) \right\| \\
&\qquad + \sqrt{N} \left\| B_0^{1/2}(\theta_0, F_0) \left(f_m^{1/2} - f_0^{1/2} - \frac{1}{2} m^{-1/2} h f_0^{1/2} \right) \right\| \\
&\qquad + \sqrt{N} \left\| \left[B_0^{1/2}(\theta_0, F_m) - B_0^{1/2}(\theta_0, F_0) \right. \right. \\
&\qquad\qquad \left. \left. - \frac{1}{2} m^{-1/2} \frac{\partial}{\partial u} B_0^{1/2}(\theta_0, u)|_{u = F_0} \left(\int_0^t h \, dF_0 \right) \right] f_0^{1/2} \right\|.
\end{aligned}$$

Now since $\lim_{m \to \infty} \sup_{t > 0} |B_0^{1/2}(\theta_0, F_m(t)) - B_0^{1/2}(\theta_0, F_0(t))| = 0$, $f_m \in \mathscr{F}$ and that $B_0(\theta, u)$ is bounded for all $0 \le u \le 1$ and θ within a neighborhood of θ_0, we have that, as $N \to \infty$,

$$\sqrt{N} \left\| \left(B_0^{1/2}(\theta_0, F_m) - B_0^{1/2}(\theta_0, F_0) \right) \left(f_m^{1/2} - f_0^{1/2} \right) \right\| \to 0$$

and

$$\sqrt{N} \left\| B_0^{1/2}(\theta_0, F_0) \left(f_m^{1/2} - f_0^{1/2} - \frac{1}{2} m^{-1/2} h f_0^{1/2} \right) \right\| \to 0.$$

Finally by (10) we have that

$$\begin{aligned}
&\sqrt{m} \left\| \left[B_0^{1/2}(\theta_0, F_m) - B_0^{1/2}(\theta_0, F_0) \right. \right. \\
&\qquad \left. \left. - \frac{1}{2} m^{-1/2} \frac{\partial}{\partial u} B_0^{1/2}(\theta_0, u)|_{u = F_0} \left(\int_0^t h \, dF_0 \right) \right] f_0^{1/2} \right\| \\
&\quad \le \sqrt{m} \left\| \left(B_0^{1/2}(\theta_0, F_m) - B_0^{1/2}(\theta_0, F_0) \right. \right. \\
&\qquad\qquad \left. \left. + \frac{m^{-1/2} \theta_0^{1/2}(\theta_0 - 1)}{(1 + (\theta_0 - 1)F_0)^2} (F_m - F_0) \right) f_0^{1/2} \right\| \\
&\qquad + \sqrt{m} \left\| \frac{\theta_0^{1/2}(\theta_0 - 1)}{(1 + (\theta_0 - 1)F_0)^2} \left(\int_0^t h \, dF_0 - (F_m - F_0) \right) f_0^{1/2} \right\| \\
&\quad \to 0
\end{aligned}$$

as $m \to \infty$. Hence the conclusion of the lemma follows. □

2.2. *Minimax lower bound.* Let $f_m^m g_N^n$ be the joint density of the observations $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$. If the density sequence $\{f_m\}$ were known to belong to a regular parametric family, that is, h of (7) were replaced by $(\theta_N - \theta_0)h_*$ with h_* known, then $\{f_m^m g_N^n\}$ belongs to a regular parametric model and the results of the locally asymptotical minimax theorems [e.g., Le Cam and Yang (1990), pages 83 and 84] imply that the asymptotic maximum risk of any estimator T_N of θ_0 is greater than or equal to the risk of a $N(0, 1/I_*(\theta_0))$ random variable, where

$$I_*(\theta_0) = \lambda \|h_* f_0^{1/2}\|^2 + (1-\lambda)\|\xi_{\theta_0, F_0} g_0^{1/2} + Ah_*\|^2$$

is a two sample analogue of the usual parametric Fisher information of θ_0. However, under the semiparametric framework where h_* is unknown, the asymptotic minimax risk of any estimator T_N may not reach the lower bound given by the parametric models. Then it is adequate to describe the efficiency of T_N through the corresponding least favorable parametric submodel of $\{f_m^m g_N^n\}$. Let Ψ be a score function defined by

$$\Psi(t) = \xi_{\theta_0, F_0}(t) + Ah(t) g_0^{-1/2}(t). \tag{12}$$

The Fisher information $I(\theta_0)$ of θ_0 for the full model of $\{f_m^m g_N^n\}$ is now given by

$$\begin{aligned} I(\theta_0) &= \inf_{h \in \mathbf{H}} \left\{\lambda \|h f_0^{1/2}\|^2 + (1-\lambda)\|\Psi g_0^{1/2}\|^2\right\} \\ &= \inf_{h \in \mathbf{H}} \left\{\lambda \int_0^\infty h^2 \, dF_0 + (1-\lambda)\int_0^\infty \Psi^2 \, dG_0\right\}. \end{aligned} \tag{13}$$

This corresponds to the two-sample Fisher information of a least favorable parametric subfamily of $\{f_m^m g_N^n\}$. For the geometric interpretation of $I(\theta_0)$ in the general semiparametric models, see Begun, Hall, Huang and Wellner (1983) or Bickel, Klaassen, Ritov and Wellner [(1993), Chapter 3].

A crucial step in evaluating the efficiency of T_N under such a semiparametric model is to find a least favorable parametric subfamily, hence a h which attains the infimum in (13).

For this purpose, it is more convenient to write (13) in the following equivalent form:

$$\begin{aligned} I(\theta_0) &= \inf_{h \in \mathbf{H}} \left\{ \sum_{\delta = 0, 1} \int_0^\infty [\delta h(t) + (1-\delta)\Psi(t)]^2 Q_0(dt, \delta) \right\} \\ &= \inf_{h \in \mathbf{H}} \|\delta h + (1-\delta)\Psi\|_{Q_0}^2, \end{aligned} \tag{14}$$

where $Q_0(t, \delta) = \lambda \delta F_0(t) + (1-\lambda)(1-\delta)G_0(t)$ is a probability measure on $(0, \infty) \times \{0, 1\}$ and $Q_0(dt, \delta) = \lambda \delta \, dF_0(t) + (1-\lambda)(1-\delta)\, dG_0(t)$. This formula will be used again in Section 3 in calculating the specific form of the least favorable h.

Now let $\mathbf{M}_N(\alpha)$ be a Hellinger neighborhood of $f_0^m g_0^n$, so that

$$\mathbf{M}_N(\alpha) = \Big\{ f_m^m g_N^n : f_m \in \mathscr{F} \text{ and } \big\| (f_m^m g_N^n)^{1/2} - (f_0^m g_0^n)^{1/2} \big\| \le 2 - \alpha, 0 < \alpha < 2 \Big\}.$$

The next main result, which is a special case of Theorem 3.2 of Begun, Hall, Huang and Wellner (1983), gives a lower bound for the local asymptotic minimax risk of estimating θ.

THEOREM 2.1. *Suppose that $\{f_m\} \in \mathscr{F}$ and Lemma 2.1 holds for the family $\{g_N, N \ge 1\}$. Let $l(\cdot): R \to [0, \infty)$ be any subconvex loss function, that is, $\{x: l(x) \le y\}$ is closed, convex and symmetric for every $y \ge 0$, such that $El(\gamma Z_0) < \infty$ for all $\gamma > 0$, where Z_0 has $N(0,1)$ distribution. Then:*

(i) *The infimum in* (13) [*or in* (14)] *is attained by*

$$h_0(t) = \begin{cases} \Gamma(\theta_0, F_0; t) - \int_0^\infty \Gamma(\theta_0, F_0 : t)\, dF_0(t), & \text{if } \theta_0 \ne 1, \\ -(1-\lambda)(1 - 2F_0(t)) & \text{if } \theta_0 = 1, \end{cases} \tag{15}$$

and $\Psi_o(t)$ is defined in (12) *with h replaced by h_0; hence,*

$$I(\theta_0) = \lambda \int_0^\infty h_0^2(t)\, dF_0(t) + (1-\lambda) \int_0^\infty \Psi_0^2(t)\, dG_0(t),$$

where $\Gamma(\theta, F; t) = \Gamma_1(\theta, F; t) + \Gamma_2(\theta, F; t)$,

$$\Gamma_1(\theta, F; t) = (\lambda(1-\lambda)\theta)^{-1/2} \tau(\theta) \times \arctan\left(\left(\frac{\lambda}{\theta(1-\lambda)} \right)^{1/2} (1 + (\theta - 1)F(t)) \right),$$

$$\Gamma_2(\theta, F; t) = \frac{\tau(\theta)(1 + (\theta-1)F(t))}{\lambda(1 + (\theta-1)F(t))^2 + \theta(1-\lambda)}$$

and

$$\tau(\theta) = \frac{\sqrt{\lambda(1-\lambda)\theta}}{\theta\left(\arctan\sqrt{\lambda\theta/(1-\lambda)} - \arctan\sqrt{\lambda/(\theta(1-\lambda))}\right)}.$$

(ii) *For any estimator T_N of θ,*

$$\lim_{\alpha \to 0} \liminf_{N \to \infty} \inf_{T_N} \sup_{f_m^m g_N^n \in \mathbf{M}_N(\alpha)} E_{f_m^m g_N^n} l\{\sqrt{N}(T_N - \theta_N)\} \ge El(Z),$$

where $E_{f_m^m g_N^n}(\cdot)$ is the expectation under the joint density $f_m^m g_N^n$ and Z has a normal distribution with mean zero and variance $I^{-1}(\theta_0)$.

REMARK. 2.2. Since $h_0(t)$ and $\Psi_0(t)$ are bounded for all $t > 0$, we have that $0 < I(\theta_0) < \infty$. Furthermore, by (8), (12) and (15), straightforward computation shows that

$$\text{(16)}\quad \Psi_0(t) = \begin{cases} \Gamma^*(\theta_0, F_0; t) - \int_0^\infty \Gamma^*(\theta_0, F_0; t)\, dG_0(t), & \text{if } \theta_0 \neq 1, \\ \lambda(1 - 2F_0(t)), & \text{if } \theta_0 = 1, \end{cases}$$

where $\Gamma^*(\theta, F; t) = \Gamma_2(\theta, F; t) - \Gamma_1(\theta, F; t)$. Some further straightforward but tedious computation based on (8), (12) and (15), shows that h_0 and Ψ_0 have the following relationship:

$$\text{(17)}\quad \begin{aligned} h_0(t) &= \frac{(1-\lambda)\lambda^{-1}\theta_0\Psi_0(t)}{[1 + (\theta_0 - 1)F_0(t)]^2} \\ &\quad - \int_0^t \frac{2(1-\lambda)\lambda^{-1}\theta_0(\theta_0 - 1)\Psi_0(s)}{[1 + (\theta_0 - 1)F_0(s)]^3}\, dF_0(s) + a, \end{aligned}$$

where a, which only depends on θ_0 and F_0, is determined by $\int_0^\infty h_0(t)\, dF_0(t) = 0$. Ψ_0 and the first term on the right-hand side of (17) will be used next as score functions in constructing an efficient one-step estimate of θ.

2.3. *Efficient estimates.* We now discuss the attainability of the lower bound $El(Z)$. An estimator T_N is called *asymptotically efficient* or simply *efficient*, if its local asymptotic maximum risk attains the lower bound, that is,

$$\text{(18)}\quad \lim_{\alpha \to 0} \lim_{N \to \infty} \sup_{f_m^m g_N^n \in \mathbf{M}_N(\alpha)} E_{f_m^m g_N^n} l\{\sqrt{N}(T_N - \theta_N)\} = El(Z).$$

The construction of $\sqrt{N}$-consistent estimates for the two-sample proportional odds model is relatively intuitive and straightforward. In fact, by selecting any suitable and convenient score functions, the estimators $\hat{\theta}$ constructed by Dabrowska and Doksum (1988) are not only $\sqrt{N}$-consistent, but also asymptotically normal in the sense that $\sqrt{N}(\hat{\theta} - \theta_N)$ has asymptotically a normal distribution uniformly for all $\theta_N \in \{\theta : \sqrt{N}|\theta - \theta_0| \le \alpha\}, 0 < \alpha < \infty$. In particular, if we take the score function to be $W(u) = (1 - u)^2$, a Dabrowska–Doksum estimate is given by

$$\hat{\theta}_p = \int_0^\infty \left(1 - \hat{F}_m(t)\right)^2 \left(1 - \hat{G}_n(t)\right)^{-2} d\hat{G}_n(t),$$

where $\hat{F}_m$ and $\hat{G}_n$ are left-continuous empirical distribution functions based on $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$, respectively. An easy application of Theorem 1 of Dabrowska and Doksum (1988) shows that the asymptotic distribution of $\sqrt{mn/(m-n)}\,(\hat{\theta}_p - \theta_0)$ is normal with mean zero and variance

$$\sigma^2 = \lambda_1 \left\{ \int_0^\infty \gamma_1^2\, dF_0 - \left(\int_0^\infty \gamma_1\, dF_0 \right)^2 \right\} + \lambda_0 \left\{ \int_0^\infty \gamma_2^2\, dG_0 - \left(\int_0^\infty \gamma_2\, dG_0 \right)^2 \right\},$$

where $\gamma_1(t) = -2\int_0^t (1 - F_0)(1 - G_0)^{-2}\, dG_0$ and $\gamma_2(t) = -2\int_0^t (1 - F_0)(1 - G_0)^{-2}\, dF_0$. Note that our notation here does not totally agree with that of Dabrowska and Doksum (1988). Moreover, under $f_m^m g_N^n \in \mathbf{M}_N(\alpha)$ for any

$\alpha < \infty$, $\sqrt{mn/(m+n)}\,(\hat{\theta}_p - \theta_N)$ is also asymptotically normal with mean zero and variance σ^2. Comparing with Theorem 2.1, we see easily that $\hat{\theta}_p$ is $\sqrt{N}$-consistent, but inefficient.

Our aim here is to apply Le Cam's one-step procedure to our semiparametric model in order to construct efficient estimates based on some $\sqrt{N}$-consistent, but possible inefficient, initial estimates, for example, the Dabrowska–Doksum estimates, such as $\hat{\theta}_p$.

Let $\hat{\theta}_N$ be any $\sqrt{N}$-consistent estimator of θ. We now define the one-step estimate of θ by

$$T_N = \hat{\theta}_N + I^{-1}\big(\hat{\theta}_N;\hat{F}_m,\hat{G}_n\big) \times\left[\frac{1}{N}\sum_{j=1}^{n}\Psi\big(\hat{\theta}_N,\hat{F}_m;Y_j\big) - \frac{1}{N}\sum_{i=1}^{m}\frac{(1-\lambda)\lambda^{-1}\hat{\theta}_N\Psi\big(\hat{\theta}_N,\hat{F}_m;X_i\big)}{\big[1+\big(\hat{\theta}_N-1\big)\hat{F}_m(X_i)\big]^2}\right], \tag{19}$$

where $I(\theta;F_m,G_n) = \lambda\int_0^\infty h^2(\theta,F;t)\,dF(t) + (1-\lambda)\int_0^\infty \Psi^2(\theta,F;t)\,dG(t)$, and $h(\theta,F;t)$ and $\Psi(\theta,F;t)$ are as defined in (15) and (16) with θ_0 and F_0 being replaced by θ and F, respectively.

The next result shows that the local asymptotic minimax lower bound of Theorem 2.1 is attained by T_N.

THEOREM 2.2. *Let $\hat{\theta}_N$ be any $\sqrt{N}$-consistent preliminary estimate of θ, for example, $\hat{\theta}_N$ may be chosen as $\hat{\theta}_p$ and T_N may be the one-step estimate defined in* (19). *Then, for all $\alpha > 0$ and $\theta_N \in \{\theta_N: \sqrt{N}|\theta_N - \theta_0| \le \alpha\}$, the asymptotic distribution of $\sqrt{N}(T_N - \theta_N)$ under the joint density $f_m^m g_N^n$ is normal with mean zero and variance $I^{-1}(\theta_0)$. Thus* (18) *with l bounded holds for T_N.*

REMARK 2.3. In general the construction of efficient one-step estimators requires specially *discretized* $\sqrt{N}$-consistent initial estimators as described in Le Cam and Yang [(1990), Section 5.3], Millar [(1983), Chapter VII], and Bickel, Klaassen, Ritov and Wellner [(1993), Chapter 7], among others. The motivation for such discretization is to eliminate those possible unreasonable estimators obtained through the likelihood functions [e.g., Le Cam and Yang (1990), page 59]. Here, because $h(\theta,F;t)$ and $\Psi(\theta,F;t)$ are bounded and continuous differentiable in θ within a neighborhood of θ_0 for all $t > 0$, the discretization requirement can be removed for $\hat{\theta}_N$ in (19).

3. Fisher information and integral equations. Finding the Fisher information is a crucial step in proving Theorems 2.1 and 2.2. In this section, we develop a pair of integral equations for the proportional odds model (5) and compute the explicit form of $I(\theta_0)$ based on the solution of these equations.

Following the definition of $I(\theta_0)$ in (14), our aim here is to find a particular h such that the function $\delta h + (1-\delta)\Psi$ has the smallest norm in the $L_2(Q_0)$

space for all $h \in \mathbf{H}$. Let

$$\mathbf{S} = \left\{ S_0(z)(1-\delta) + S_1(z)\delta : \int_0^\infty S_0(t)\,dG_0(t) = \int_0^\infty S_1(t)\,dF_0(t) = 0, \right.$$
$$\left. S_0 \in L_2(G_0), S_1 \in L_2(F_0) \text{ and } \delta \in \{0,1\} \right\}$$

be a L_2-space with respect to Q_0 and let $\|\cdot\|_{Q_0}$ and $\langle \cdot, \cdot \rangle_{Q_0}$ be its norm and inner product, respectively. Let

$$\mathbf{T} = \left\{ h(z)\delta + (Ah(z))g_0^{-1/2}(z)(1-\delta) : h \in \mathbf{H}, \delta \in \{0,1\} \right\}.$$

Since $\mathbf{T}$ itself is a vector space and a subset of $\mathbf{S}$, it is a subspace of $\mathbf{S}$. Denote by $\Pi(\xi_{\theta_0, F_0}(z)(1-\delta))$ the projection of $\xi_{\theta_0, F_0}(z)(1-\delta)$ into $\mathbf{T}$ and by $\mathbf{T}^\perp$ the orthogonal complement of $\mathbf{T}$ in $\mathbf{S}$ such that $\mathbf{S} = \mathbf{T} \oplus \mathbf{T}^\perp$. Thus $\mathbf{T}^\perp$ is also a subspace of $\mathbf{S}$ and

$$\left\{ \xi_{\theta_0, F_0}(z)(1-\delta) - \Pi\left(\xi_{\theta_0, F_0}(z)(1-\delta) \right) \right\} \in \mathbf{T}^\perp.$$

Basic properties of Hilbert spaces imply that there are unique $h \in \mathbf{H}$ and $\{M_0(z)(1-\delta) + M_1(z)\delta\} \in \mathbf{T}^\perp$ such that

$$\begin{aligned} \xi_{\theta_0, F_0}(z)(1-\delta) &= h(z)\delta + (Ah(z))g_0^{-1/2}(z)(1-\delta) \\ &\quad + M_0(z)(1-\delta) + M_1(z)\delta \end{aligned} \tag{20}$$

and by (14),

$$\begin{aligned} I(\theta_0) &= \left\| \xi_{\theta_0, F_0}(z)(1-\delta) - \Pi\left(\xi_{\theta_0, F_0}(z)(1-\delta) \right) \right\|_{Q_0}^2 \\ &= \| M_0(z)(1-\delta) + M_1(z)\delta \|_{Q_0}^2. \end{aligned} \tag{21}$$

LEMMA 3.1. *If $M_0(z)(1-\delta) + M_1(z)\delta$ is any element of $\mathbf{T}^\perp$, then*

$$\begin{aligned} \lambda M_1(t) &= (1-\lambda) \int_t^\infty \frac{2\theta_0(\theta_0 - 1)}{[1 + (\theta_0 - 1)F_0(s)]^3} M_0(s)\,dF_0(s) \\ &\quad - \frac{(1-\lambda)\theta_0 M_0(t)}{[1 + (\theta_0 - 1)F_0(t)]^2} + (1-\lambda)K_0, \end{aligned}$$

where

$$K_0 = \int_0^\infty \frac{-2\theta_0(\theta_0 - 1)F_0(t)}{[1 + (\theta_0 - 1)F_0(t)]^3} M_0(t)\,dF_0(t).$$

PROOF. Let $M_0(z)(1-\delta) + M_1(z)\delta$ be any element of $\mathbf{S}$. If it is orthogonal to $\mathbf{T}$, then the inner product between $M_0(z)(1-\delta) + M_1(z)\delta$ and any element in **T** must be zero, that is, for any $h \in \mathbf{H}$,

$$\begin{aligned}\langle h(z)\delta + (Ah(z))g_0^{-1/2}(z)(1-\delta), M_0(z)(1-\delta) + M_1(z)\delta\rangle_{Q_0} \\ = \sum_{\delta=0,1}\Big\{\int \big(h(z)\delta + (Ah(z))g_0^{-1/2}(z)(1-\delta)\big) \\ \times (M_0(z)(1-\delta) + M_1(z)\delta)Q_0(dz,\delta)\Big\} \\ = 0.\end{aligned}$$

Equivalently, we have that, for any $h \in \mathbf{H}$,

$$\int_0^\infty\int_0^\infty (\lambda h(x) + (1-\lambda)Ah(y)g_0^{-1/2}(y)) \times (M_1(x) + M_0(y))\,dF_0(x)\,dG_0(y) = 0.$$

Thus, by (6) and the definition of A in (8), the following equalities hold for all $h \in \mathbf{H}$:

$$\begin{aligned}&\lambda\int_0^\infty h(x)M_1(x)\,dF_0(x) + (1-\lambda)\int_0^\infty Ah(y)g_0^{-1/2}(y)M_0(y)\,dG_0(y) \\ &= \lambda\int_0^\infty h(x)M_1(x)\,dF_0(x) + (1-\lambda)\int_0^\infty h(y)M_0(y)\,dG_0(y) \\ &\quad + (1-\lambda)\int_0^\infty\left(\frac{-2(\theta_0-1)}{1+(\theta_0-1)F_0(y)}\int_0^y h(t)\,dF_0(t)\right)M_0(y)\,dG_0(y) \\ &= \int_0^\infty\Bigg[\lambda M_1(x) + \frac{(1-\lambda)\theta_0 M_0(x)}{[1+(\theta_0-1)F_0(x)]^2} \\ &\qquad - (1-\lambda)\int_x^\infty \frac{2\theta_0(\theta_0-1)M_0(t)}{[1-(\theta_0-1)F_0(t)]^3}\,dF_0(t)\Bigg]h(x)\,dF_0(x) \\ &= 0.\end{aligned}$$

The last equality holds if and only if, for some constant K_0,

$$\begin{aligned}\lambda M_1(t) = (1-\lambda)\int_t^\infty \frac{2\theta_0(\theta_0-1)}{[1+(\theta_0-1)F_0(s)]^3}M_0(s)\,dF_0(s) \\ - \frac{(1-\lambda)\theta_0 M_0(t)}{[1+(\theta_0-1)F_0(t)]^2} + (1-\lambda)K_0.\end{aligned}$$

Then K_0 is determined by (6), $\int_0^\infty M_1(x)\,dF_0(x) = 0$ and

$$\int_0^\infty M_0(x)\,dG_0(x) = \int_0^\infty \frac{\theta_0 M_0(x)}{[1+(\theta_0-1)F_0(x)]^2}\,dF_0(x) = 0.$$

The conclusion of the lemma holds. □

Lemma 3.1 implies that (20) can be specified as

$$\begin{aligned}\xi_{\theta_0, F_0}(z)(1-\delta) &= h(z)\delta + Ah(z)g_0^{-1/2}(z)(1-\delta) \\ &\quad + M_0(z)(1-\delta) - \frac{(1-\lambda)\lambda^{-1}\theta_0\delta M_0(z)}{[1+(\theta_0-1)F_0(z)]^2} \\ &\quad + (1-\lambda)\lambda^{-1}\delta\int_z^\infty \frac{2\theta_0(\theta_0-1)F_0(s)}{[1+(\theta_0-1)F_0(s)]^3}\,dF_0(s) \\ &\quad + (1-\lambda)\lambda^{-1}\delta K_0\end{aligned} \tag{22}$$

for all $z \in (0,\infty)$ and $\delta = 0, 1$. Furthermore, by (8), (9), $\int h\,dF_0 = \int M_0\,dG_0 = 0$ and routine integration, (22) is equivalent to the following pair of integral equations in $h(x)$ and $M_0(y)$:

$$\begin{aligned}&\frac{1}{\theta_0} - \frac{2F_0(y)}{1+(\theta_0-1)F_0(y)} \\ &\qquad = -\frac{2(\theta_0-1)}{1+(\theta_0-1)F_0(y)}\int_0^y h(t)\,dF_0(t) + h(y) + M_0(y), \\ &h(x) = (1-\lambda)\lambda^{-1}\theta_0 M_0(x)\big(1+(\theta_0-1)F_0(x)\big)^{-2} \\ &\qquad + (1-\lambda)\lambda^{-1}\int_0^x \frac{2\theta_0(\theta_0-1)M_0(t)}{[1+(\theta_0-1)F_0(t)]^3}\,dF_0(t) \\ &\qquad - (1-\lambda)\lambda^{-1}b,\end{aligned} \tag{23}$$

where x and y are arbitrary points in $(0,\infty)$ and

$$b = \int_0^\infty \frac{2\theta_0(\theta_0-1)(1-F_0(s))}{[1+(\theta_0-1)F_0(s)]^3} M_0(s)\,dF_0(s).$$

It is interesting to note that, when $h(t) = -h_0(t)$ and $M_0(t) = \Psi_0(t)$ as defined in (15) and (16), the second equation of (23) is reduced to the special case (17).

LEMMA 3.2. *The solution of* (23) *is given by* $h(t) = -h_0(t)$ *and* $M_0(t) = \Psi_0(t)$ *as defined in* (15) *and* (16), *respectively. Consequently, when* (θ_0, F_0, G_0) *satisfies* (5),

$$I(\theta_0) = \|h_0(z)\delta + M_0(z)(1-\delta)\|_{Q_0}^2.$$

PROOF. Suppose that $h(t) = -h_0(t)$ and $M_0(t)$ are the solution of (23). It is easily seen from (20) that $h(z) = -M_1(z)$. Thus the last assertion of the lemma follows from (21) and $h_0(t) = M_1(t)$.

We now consider the solution of (23). If $\theta_0 = 1$, (23) reduces to the following special case:

$$1 - 2F_0(t) = h(t) + M_0(t) \quad \text{and} \quad h(t) = (1-\lambda)\lambda^{-1}M_0(t).$$

Thus the solution of the above equations is

$$M_0(t) = \lambda(1 - 2F_0(t)) \quad \text{and} \quad h(t) = -h_0(t) = (1-\lambda)(1-2F_0(t)).$$

Now we consider the case of $\theta_0 \neq 1$. First notice that, by direct integration, (15) and (16) are the results of the following two integrals:

$$h_0(t) = \int_0^t \frac{(1-\lambda)\theta_0 c(\theta_0)}{\left[\lambda U^2(\theta_0, F_0) + (1-\lambda)\theta_0\right]^2}\, dF_0 - b_1, \tag{24}$$

$$\Psi_0(t) = -\int_0^t \frac{\lambda c(\theta_0) U^2(\theta_0, F_0)}{\left[\lambda U^2(\theta_0, F_0) + (1-\lambda)\theta_0\right]^2}\, dF_0(s) + b_2, \tag{25}$$

where $U(\theta, F; t) = 1 + (\theta - 1)F(t)$,

$$c(\theta) = 2\left(\int_0^1 \frac{\theta}{\lambda[1 + (\theta-1)u]^2 + (1-\lambda)\theta}\, du\right)^{-1} \quad \text{for } \theta > 0$$

and b_1 and b_2 are determined by $\int h_0\, dF_0 = 0$ and $\int \Psi_0\, dG_0 = 0$, respectively. Then, by (24), (25) and direct computation, we have that

$$\Psi_0(y) = \frac{c(\theta_0)U(\theta, F_0;y)}{(\theta_0 - 1)\left[\lambda U^2(\theta_0, F_0;y) + (1-\lambda)\theta_0\right]} - \frac{c(\theta_0)}{(\theta_0 - 1)(\lambda + (1-\lambda)\theta_0)}$$
$$- \int_0^y \frac{(1-\lambda)\theta_0 c(\theta_0)}{\left[\lambda U^2(\theta_0, F_0;t) + (1-\lambda)\theta_0\right]^2}\, dF_0(t) + b_2,$$

$$\frac{2(\theta_0 - 1)}{U(\theta_0, F_0; y)} \int_0^y h_0(t)\, dF_0(t)$$
$$= 2\int_0^y \frac{(1-\lambda)\theta_0 c(\theta_0)}{\left[\lambda U^2(\theta_0, F_0;t) + (1-\lambda)\theta_0\right]^2}\, dF_0(t) - \frac{2(\theta_0 - 1)b_1 F_0(y)}{U(\theta_0, F_0; y)}$$
$$+ \frac{(1-\lambda)\theta_0 c(\theta_0)}{\lambda(\theta_0 - 1)U(\theta_0, F_0; y)\left[\lambda U^2(\theta_0, F_0; y) + (1-\lambda)\theta_0\right]}$$
$$- \frac{(1-\lambda)\theta_0 c(\theta_0)}{\lambda(\theta_0 - 1)(\lambda + (1-\lambda)\theta_0)U(\theta_0, F_0; y)}$$

and

$$\frac{(1-\lambda)\theta_0 \Psi_0(y)}{\lambda U^2(\theta_0, F_0; y)} + (1-\lambda)\lambda^{-1} \int_0^y \frac{2\theta_0(\theta_0 - 1)\Psi_0(t)}{U^3(\theta_0, F_0;t)}\, dF_0(t)$$
$$= -\int_0^y \frac{(1-\lambda)\theta_0 c(\theta_0)}{\left[\lambda U^2(\theta_0, F_0;t) + (1-\lambda)\theta_0\right]^2}\, dF_0(t) + (1-\lambda)\lambda^{-1} b_2 \theta_0.$$

Now it is straightforward to verify that both equations of (23) are satisfied when $M_0(t) = \Psi_0(t)$ and $h(t) = -h_0(t)$. □

REMARK. 3.1. Extending the method of this section to the general two-sample transformation model (1), it is also possible to develop a pair of general integral equations such that their solution gives a least favorable parametric subfamily for estimating θ. In particular, applying these general equations to the two-sample proportional hazard model (4), we can derive an explicit form of the Fisher information of θ which coincides with that given in Efron (1977) and Begun and Wellner (1983). Unfortunately the solution for the general case is unknown, hence, we do not have explicit forms of the Fisher information of θ for the general models (1) or (2). Efficient estimation of θ under the general two-sample transformation models certainly deserves further study.

4. Proofs. We provide in this section the proofs of Theorems 2.1 and 2.2.

PROOF OF THEOREM 2.1. The first assertion is essentially the conclusion of Lemma 3.2. Thus we only consider the statement in (ii). For $\{f_m\}$ and $\{g_N\}$ as in Lemma 2.1, define L_N to be the local log likelihood ratio such that

$$L_N = 2\log\left\{\prod_{i=1}^{m}\frac{f_m^{1/2}(X_i)}{f_0^{1/2}(X_i)} \times \prod_{j=1}^{n}\frac{g_N^{1/2}(Y_j)}{g_0^{1/2}(Y_j)}\right\}$$
$$= \sum_{i=1}^{m}\log\frac{f_m(X_i)}{f_0(X_i)} + \sum_{j=1}^{n}\log\frac{g_N(Y_j)}{g_0(Y_j)}.$$

Then Lemma 2.1 implies that L_N is locally asymptotically normal, that is, Lemma 2.1. of Begun, Hall, Huang and Wellner (1983) holds for L_N and the conditions of Theorem 3.2 of Begun, Hall, Huang and Wellner (1983) are satisfied. It can be deduced from the proof of Theorem 3.2 of Begun, Hall, Huang and Wellner (1983) that

$$\lim_{c_*\to\infty}\lim_{N\to\infty}\inf_{T_N}\sup_{\substack{f_m\in\mathbf{B}_F(c_*)\\ g_N\in\mathbf{B}_G(c_*)}} E_{f_m^m g_N^n} l\{\sqrt{N}(T_N-\theta_N)\} \ge El(Z), \tag{26}$$

where

$$\mathbf{B}_F(c_*) = \{f_m : f_m \in \mathscr{F} \text{ and } \sqrt{N}\|f_m^{1/2} - f_0^{1/2}\| \le c_*\}$$

and

$$\mathbf{B}_G(c_*) = \{g_N : g_N \text{ satisfies Lemma 2.1 and } \sqrt{N}\|g_N^{1/2} - g_0^{1/2}\| \le c_*\}.$$

Now, by the independence of $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$, for sufficiently large N,

$$\begin{aligned}&\left\|(f_m^m g_N^n)^{1/2} - (f_0^m g_0^n)^{1/2}\right\|^2 \\ &\quad = 2 - 2\left(1 - \tfrac{1}{2}\|f_m^{1/2} - f_0^{1/2}\|^2\right)^m \left(1 - \tfrac{1}{2}\|g_N^{1/2} - g_0^{1/2}\|^2\right)^n \\ &\quad = 2 - 2\exp\left\{\left(-\tfrac{1}{2}N\right)\left(\lambda_1\|f_m^{1/2} - f_0^{1/2}\|^2 + \lambda_0\|g_N^{1/2} - g_0^{1/2}\|^2\right)\right\} + o(1).\end{aligned}$$

Consequently, for any estimator T_N and sufficiently large N,

$$\begin{aligned}&\sup_{f_m^m g_N^n \in \mathbf{M}_N(2 - 2\exp(-c_*^2/2) + o(1))} E_{f_m^m g_N^n} l\left(\sqrt{N}(T_N - \theta_N)\right) \\ &\quad \geq \sup_{\substack{f_m \in \mathbf{B}_F(c_*) \\ g_N \in B_G(c_*)}} E_{f_m^m g_N^n} l\left(\sqrt{N}(T_N - \theta_N)\right).\end{aligned}$$

Therefore, (26) completes the proof of the theorem. □

PROOF OF THEOREM 2.2. We first notice the following inequalities:

$$\begin{aligned}&\left|\int h^2\left(\hat{\theta}_N, \hat{F}_m\right) d\hat{F}_m - \int h_0^2\, dF_0\right| \\ &\quad \leq \sup_{x \in (0,\infty)} \left|h^2\left(\hat{\theta}_N, \hat{F}_m; x\right) - h_0^2(x)\right| + \left|\int h_0^2 d\left(\hat{F}_m - F_0\right)\right|\end{aligned} \tag{27}$$

and

$$\begin{aligned}&\left|\int \Psi^2\left(\hat{\theta}_N, \hat{F}_m\right) d\hat{G}_n - \int \Psi_0^2\, dG_0\right| \\ &\quad \leq \sup_{x \in (0,\infty)} \left|\Psi^2\left(\hat{\theta}_N, \hat{F}_m; x\right) - \Psi_0^2(x)\right| + \left|\int \Psi_0^2 d\left(\hat{G}_n - G_0\right)\right|.\end{aligned} \tag{28}$$

Since $\hat{\theta}_N$ and $\hat{F}_m$ are $\sqrt{N}$-consistent and both $h_0(t)$ and $\Psi_0(t)$ are bounded in $t \in (0, \infty)$, it is easy to verify from the definition of $h(\theta, F; t)$ and $\Psi(\theta, F; t)$ given in (19) that

$$\begin{aligned}&\sup_{t \in (0,\infty)} \left|h^2\left(\hat{\theta}_N, \hat{F}_m; t\right) - h_0^2(t)\right| \to 0, \\ &\sup_{t \in (0,\infty)} \left|\Psi^2\left(\hat{\theta}_N, F_m; t\right) - \Psi_0^2(t)\right| \to 0,\end{aligned} \tag{29}$$

in probability as $N \to \infty$ and, by the law of large numbers,

$$\begin{aligned}&\left|\int h^2(\theta_0, F_0) d\left(\hat{F}_m - F_0\right)\right| \to 0, \\ &\left|\int \Psi^2(\theta_0, F_0) d\left(\hat{G}_n - G_0\right)\right| \to 0,\end{aligned} \tag{30}$$

in probability as $N \to \infty$. Relations (27) through (30) imply that as $N \to \infty$,

$$|I(\hat{\theta}_N; \hat{F}_m, \hat{G}_n) - I(\theta_0)| \to 0 \quad \text{in probability.} \tag{31}$$

Since $I(\theta_0) > 0$, (31) implies that

$$I^{-1}(\hat{\theta}_N; \hat{F}_m, \hat{G}_n) \to I^{-1}(\theta_0) \quad \text{in probability as } N \to \infty. \tag{32}$$

Next the error sequence $\sqrt{N}(T_N - \theta_N)$ of (19) can be written as

$$\begin{aligned} \sqrt{N}(T_N - \theta_N) &= \sqrt{N}(\hat{\theta}_N - \theta_N) + I^{-1}(\hat{\theta}_N; \hat{F}_m, \hat{G}_n) \\ &\quad \times \sqrt{N}\left\{\frac{1}{N}\sum_{j=1}^{n} \Psi(\hat{\theta}_N, \hat{F}_M; Y_j)\right. \\ &\quad \left. - \frac{1}{N}\sum_{i=1}^{m} \frac{(1-\lambda)\lambda^{-1}\hat{\theta}_N \Psi(\hat{\theta}_N, \hat{F}_m; X_i)}{\left[1 + (\hat{\theta}_N - 1)\hat{F}_m(X_i)\right]^2}\right\}. \end{aligned} \tag{33}$$

Let $\{F_{\hat{\theta}_N}\}$ be a sequence of distribution functions with densities $\{f_{\hat{\theta}_N}\}$ such that

$$f_{\hat{\theta}_N}(t) = f_m(t) + (\hat{\theta}_N - \theta_N)h(\theta_N, F_m; t)f_m(t).$$

Note that $f_{\hat{\theta}_N}$ is indeed a density function when N is sufficiently large. Let $\{G_{\hat{\theta}_N}\} \equiv \{G_{\hat{\theta}_N, f_{\hat{\theta}_N}}\}$ be a sequence of distributions as defined in (5). The $\sqrt{N}$-consistency of $\hat{\theta}_N$, $\sqrt{N}|\theta_N - \theta_0| \le \alpha$, $\alpha > 0$, and $\lim_{m\to\infty}\sqrt{m}\,\|f_m^{1/2} - f_0^{1/2}\| < \infty$ imply that both

$$\sup_{t\in(0,\infty)} \sqrt{N}|F_{\hat{\theta}_N}(t) - F_0(t)| \quad \text{and} \quad \sup_{t\in(0,\infty)} \sqrt{N}|G_{\hat{\theta}_N}(t) - G_0(t)|$$

are bounded in probability. Thus the second term of the right side of (33) can be written as

$$\begin{aligned} &\sqrt{N}\left(\frac{1}{N}\sum_{j=1}^{n} \Psi(\hat{\theta}_N, \hat{F}_m; Y_j) - \frac{1}{N}\sum_{i=1}^{m} \frac{(1-\lambda)\lambda^{-1}\hat{\theta}_N \Psi(\hat{\theta}_N, \hat{F}_m; X_i)}{\left[1 + (\hat{\theta}_N - 1)\hat{F}_m(X_i)\right]^2}\right) \\ &= (1-\lambda)\sqrt{N}\left\{\int_0^\infty \Psi(\hat{\theta}_N, \hat{F}_m; t)\, d\hat{G}_n(t)\right. \\ &\quad \left. - \int_0^\infty \frac{\hat{\theta}_N \Psi(\hat{\theta}_N, \hat{F}_m; t)}{\left[1 + (\hat{\theta}_N - 1)\hat{F}_m(t)\right]^2}\, d\hat{F}_m(t)\right\} + o_p(1) \end{aligned}$$

$$= (1-\lambda)\sqrt{N}\Bigg\{\int_0^\infty \Psi\big(\hat\theta_N, \hat F_m\big) d\big(\hat G_n - G_{\hat\theta_N}\big) - \int_0^\infty \frac{\hat\theta_N \Psi\big(\hat\theta_N, \hat F_m\big)}{\big[1+\big(\hat\theta_N - 1\big)\hat F_m\big]^2} d\big(\hat F_m - F_{\hat\theta_N}\big) + \int_0^\infty \Psi\big(\hat\theta_N, \hat F_m\big)\, dG_{\hat\theta_N} - \int_0^\infty \frac{\hat\theta_N \Psi\big(\hat\theta_N, \hat F_m\big)}{\big[1+\big(\hat\theta_N - 1\big)\hat F_m\big]^2}\, dF_{\hat\theta_N}\Bigg\} + o_p(1)$$

$$= (1-\lambda)\sqrt{N}\Bigg\{\int_0^\infty \Psi(\theta_N, F_m) d\big(\hat G_n - G_{\hat\theta_N}\big) - \int_0^\infty \frac{\theta_N \Psi(\theta_N, F_m)}{[1+(\theta_N - 1)F_m]^2} d\big(\hat F_m - F_{\hat\theta_N}\big) + \int_0^\infty \left(\int_0^x \frac{-2(\theta_N - 1)\theta_N \Psi(\theta_N, F_m)}{[1+(\theta_N-1)F_m]^3}\, dF_{\hat\theta_N}\right) \times d\big(\hat F_m(x) - F_{\hat\theta_N}(x)\big)\Bigg\} + o_p(1)$$

$$= (1-\lambda)\sqrt{N}\Bigg\{\int_0^\infty \Psi(\theta_N, F_m) d\big(\hat G_n - G_{\hat\theta_N}\big) - \int_0^\infty \frac{\theta_N \Psi(\theta_N, F_m)}{[1+(\theta_N - 1)F_m]^2} d\big(\hat F_m - F_{\hat\theta_N}\big) + \int_0^\infty \left(\int_0^x \frac{-2(\theta_N - 1)\theta_N \Psi(\theta_N, F_m)}{[1+(\theta_N-1)F_m]^3}\, dF_m\right) \times d\big(\hat F_m(x) - F_{\hat\theta_N}(x)\big)\Bigg\} + o_p(1),$$

where the second and third equality signs hold because $g_{\hat\theta_N} = \hat\theta_N(1 + (\theta_N - 1)F_{\hat\theta_N})^{-2} f_{\hat\theta_N}$. Direct computation, (22) and Lemma 3.2 show that

$$\partial g_{\hat\theta_N}(t)/\partial\hat\theta_N|_{\hat\theta_N = \theta_N} = \Psi(\theta_N, F_m, t) g_N(t)$$

and that the next equality follows from the usual Taylor expansions

$$\sqrt{N}\left(\frac{1}{N}\sum_{j=1}^{n}\Psi\left(\hat{\theta}_N,\hat{F}_m;Y_j\right)-\frac{1}{N}\sum_{i=1}^{m}\frac{(1-\lambda)\lambda^{-1}\hat{\theta}_N\Psi\left(\hat{\theta}_N,\hat{F}_m;X_i\right)}{\left[1+\left(\hat{\theta}_N-1\right)\hat{F}_m(X_i)\right]^2}\right) \tag{34}$$
$$=\sqrt{N}\Delta_N-\sqrt{N}I(\theta_0)\left(\hat{\theta}_N-\theta_N\right)+o_p(1),$$

where

$$\sqrt{N}\Delta_N+(1-\lambda)\sqrt{N}\left\{\int_0^\infty\Psi(\theta_N,F_m)\,d\left(\hat{G}_n-G_N\right)\right.$$
$$-\int_0^\infty\frac{\theta_N\Psi(\theta_N,F_m)}{[1+(\theta_N-1)F_m]^2}\,d\left(\hat{F}_m-F_m\right)$$
$$+\int_0^\infty\left(\int_0^x\frac{-2(\theta_N-1)\theta_N\Psi(\theta_N,F_m)}{[1+(\theta_N-1)F_m]^3}\,dF_m\right)$$
$$\left.\times\, d\left(\hat{F}_m(x)-F_m(x)\right)\right\}.$$

The central limit theorem then implies that the asymptotic distribution of $\sqrt{N}\Delta_N$ is normal with mean zero and variance $I(\theta_0)$. Thus (31) through (34) imply that

$$\sqrt{N}(T_N-\theta_N)=\sqrt{N}\left(\hat{\theta}_N-\theta_N\right)+\sqrt{N}I^{-1}\left(\hat{\theta}_N;\hat{F}_m,\hat{G}_n\right)\Delta_N$$
$$-\sqrt{N}I^{-1}\left(\hat{\theta}_N;\hat{F}_m,\hat{G}_n\right)I(\theta_0)\left(\hat{\theta}_N-\theta_N\right)+o_p(1)$$
$$\to Z\quad\text{in distribution as } N\to\infty,$$

where the distribution of Z is normal with mean zero and variance $I^{-1}(\theta_0)$. It now follows that (18) holds for T_N with all bounded and subconvex loss functions. The proof of the theorem is complete. □

Acknowledgments. The author would like to thank Professor Peter J. Bickel, Professor P. W. Millar, Professor David Oakes and Professor Robert J. Serfling for many fruitful discussions, comments and encouragement in the preparation of this paper. The author would also like to thank Professor Alan J. Goldman for his help in polishing the final version of this paper. The author is also grateful to an Associate Editor and the referee for their careful reading of the manuscript and insightful suggestions. Their suggestions have greatly improved the presentation of this paper.

REFERENCES

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.

BEGUN, J. M. and WELLNER, J. A. (1983). Asymptotic efficiency of relative risk estimates. In *Contribution to Statistics: Essays in Honor of Norman Lloyd Johnson* (P. K. Sen, ed.) 47–62. North-Holland, Amsterdam.

BICKEL, P. J. (1986). Efficient testing in a class of transformation models. Technical report, Dept. Statistics, Univ. California, Berkeley.

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.

CLAYTON, D. and CUZICK, J. (1986). The semiparametric Pareto model for regression analysis of survival times. In *Proceedings of the 45th Session of the International Statistical Institute* 23.3-1–23.3-18. Internat. Statist. Inst., Amsterdam.

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

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–279.

DABROWSKA, D. M. and DOKSUM, J. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *J. Amer. Statist. Assoc.* **83** 774–749.

DABROWSKA, D. M., DOKSUM, J. A. and MIURA, R. (1989). Rank estimates in a class of semi-parametric two-sample models. *Ann. Inst. Statist. Math.* **41** 63–79.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.

KLAASSEN, C. A. J. (1989). Efficient estimation in the Clayton–Cuzick model for survival data. Technical Report 88-06, Dept. Mathematics, Univ. Leiden.

KOSHEVNICK, YU. A. and LEVIT, B. YA. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738–753.

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. California Press, Berkeley.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, Berlin.

LE CAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics: Some Basic Concepts.* Springer, Berlin.

LEVIT, B. YA. (1975). On the efficiency of a class of non-parametric estimates. *Theory Probab. Appl.* **20** 723–740.

LEVIT, B. YA. (1978). Infinite-dimensional informational inequalities. *Theory Probab. Appl.* **23** 371–377.

MCCULLAGH, D. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

MILLAR, P. W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **48** 233–252.

MILLAR, P. W. (1983). *The Minimax Principle in Asymptotic Theory. Lecture Notes in Math.* **976** 76–276. Springer, Berlin.

WELLNER, J. A. (1985). Semiparametric models: progress and problems. In *Proceedings of the 45th Session of the International Statistics Institute* 23.1-1–23.1-20. Internat. Statist. Inst., Amsterdam.

DEPARTMENT OF MATHEMATICAL SCIENCES
JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218-2689