PARTIAL LIKELIHOOD IN TRANSFORMATION MODELS WITH CENSORED DATA*

by

Dorota M. Dabrowska

and

Kjell A. Doksum

Technical Report No. 47
July 1985
(revised March 1987)

*Research Partially Supported by
National Science Foundation Grants DMS-83-01716
and DMS-85-41787, and the
National Institute of General Medical Sciences Grant SSS-Y1RO1 GM35416.

Department of Statistics
University of California
Berkeley, California

PARTIAL LIKELIHOOD IN TRANSFORMATION MODELS WITH CENSORED DATA*

by

Dorota M. Dabrowska
Carnegie-Mellon University

and

Kjell A. Doksum
University of California, Berkeley

Abstract

In regression analysis, including generalized linear models, the scale at which the response is to be measured to make the regression linear, is known. We consider linear transformation models where this scale is unknown and the transformed response variable and the covariates satisfy a linear regression model. Such models provide a unified approach to semiparametric models where some rate function given the covariates can be expressed as a product of a baseline rate function and a parametric function involving the covariates. When regarding the transformation as a nuisance parameter, a partial likelihood is useful to obtain estimates of the regression parameters. A resampling scheme (likelihood sampler) is used to approximate the partial likelihood and the maximum partial likelihood estimate. Counting processes techniques are used to establish conditions under which the partial likelihood is an approximation to the likelihood. Asymptotic and Monte Carlo methods are used to compare the approximate maximum partial likelihood estimate with an estimate obtained from a local approximation to the likelihood. Our limited Monte Carlo results indicate that our methods work very well when the regression effect is small or moderate compared to the residual variation, in fact, in this case, they have mean squared errors equal to or close to that of the parametric estimate appropriate when the transformation is known. On the other hand, our estimates are badly biased when the regression parameter is large relative to residual variation. The effect of misspecification of the model on the performance of the estimates is also investigated and found to be very small.

Key words: partial likelihood, semiparametric transformation models, counting processes, proportional hazards, odds and γ-odds.

1. Introduction. We consider models where a response variable Y and p covariates $x_1,\ldots,x_p$ are related through an equation of the form

$$h(Y) = \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon. \tag{1.1}$$

Here h is an increasing, unknown function, $\beta^T = (\beta_1,\ldots,\beta_p)$ are parameters and ε has a known continuous distribution function Ψ. Since h is unknown, it is assumed that any multiplicative or additive constants are absorbed into h. If Y, $x_1,\ldots,x_p$ satisfy (1.1) we say that they follow a *linear transformation model* with parameters β and h. The function h is thought of as the scale where the transformed response variable h(Y) satisfies a linear model relationship with the covariates $x_1, \cdots, x_p$. The usefulness of such models for data analysis has been discussed in detail by Anscombe & Tukey (1954) and Box & Cox (1964). These authors considered parametric functions h, while in our case h is unspecified.

When h is unknown, there has been a considerable discussion about the interpretation of β in the model (1.1). See Bickel & Doksum (1981), Hinkley & Runger (1984), Bickel (1984), Rubin (1984), and Doksum (1984, 1987). Here we note that β/σ_e where σ_e denotes the standard deviation of the error ε, measures the amount of systematic variation relative to residual variation on the linear model scale. Thus when testing whether certain covariates contribute to systematic variation, β/σ_E is the appropriate parameter. Moreover, the ratio β_i/β_j is a measure of the relative importance of the covariates x_i and x_j in the systematic variation. Although the controversy involving the meaning of β in (1.1) is interesting, there is, for a variety of cases, a way of avoiding it. This involves reparametrizing the problem in such a way that β has a meaning independent of the scale h. The next four examples illustrate this point.

EXAMPLE 1.1. (Proportional Hazard Model). Let Y be a nonnegative survival or failure time with hazard function λ_Y. The Cox (1972, 1975) *proportional hazard model* is given by $\lambda_Y(y) = \lambda(y)\exp(-\Sigma\, \beta_j x_j)$ where λ is an unknown nonnegative function on $(0,\infty)$. It was observed by Kalbfleisch (1978) and Prentice (1978) that this is a special case of the linear transformation model. Here (1.1) is satisfied with $h(y) = \log \int_0^y \lambda(t)dt$ and Ψ equal to the extreme value distribution $1-\exp(-e^t)$. The parameter β_j is the increase in the log hazard as x_j is increased by one unit. Special cases of this model were considered earlier by Lehmann (1953), Savage (1956, 1957), and Rao, Savage & Sobel (1960), among others.

EXAMPLE 1.2. (Proportional Odds Rate Model). For a nonnegative random variable Y define the *odds on death* or *odds function* at time t as

$$r_Y(t) \ = \ P(Y \le t)/[1-P(Y \le t)].$$

The *proportional odds rate model* (Bennett (1983), Pettitt (1984)) is given by

$$r_Y(y) \ = \ r(y)\exp\{-\Sigma\, \beta_j x_j\}$$

where r is an unknown increasing function on $(0,\infty)$. The term *odds rate* refers to the derivative of the function r_Y. Thus the ratio of two odds rates corresponding to different values of the covariates does not depend on r(y). Here (1.1) is satisfied with $h(y) = \log r(y)$ and Ψ equal to the logistic distribution $[1+\exp(-t)]^{-1}$. The parameter β_j is the increase in the log odds function as x_j is increased by one unit. This model was considered in the two-sample uncensored case by Ferguson (1967, p. 257). In the case of dose response studies, it is the natural extension of the logit model for binary data (Berkson (1944)) to continuous data.

EXAMPLE 1.3. (Proportional γ odds Model). For $Y \ge 0$, define the γ *odds function* at time $t>0$ as

$$\Gamma_Y(t) \ = \ \frac{1}{\gamma}\left[\frac{1-P^\gamma(Y>t)}{P^\gamma(Y>t)}\right], \quad \gamma>0$$

$$= \ -\log P(Y>t), \quad \gamma \ = \ 0.$$

For $\gamma = 0$, Γ_Y is the integrated hazard rate, and for $\gamma = 1$, Γ_Y is the odds function. If γ is a positive integer, say k, then $k\Gamma_Y(t)$ is the odds that k independent individuals with the same covariate values $x_1, \cdots, x_p$ will not all survive to time t. To see this, note that if $\tilde{Y}_1, \cdots, \tilde{Y}_k$ are the survival times of the k individuals, then $P^k(Y>t) = P(\min(\tilde{Y}_1, \cdots, \tilde{Y}_k)>t)$ and

$1-P^k(Y>t) = P(\min(\tilde{Y}_1, \cdots, \tilde{Y}_k) \le t)$. For γ an arbitrary rational number, a similar interpretation of $\gamma\Gamma_Y(t)$ in terms of series systems can be given. $\gamma\Gamma_Y(t)$ is called the *γ-odds on death by time* t. The *proportional γ-odds model* is

$$\Gamma_Y(t) = \Gamma(t)\exp\{-\sum \beta_j x_j\}$$

where Γ is an unknown increasing baseline γ-odds function on $(0,\infty)$. Here (1.1) is satisfied with $h(y) = \log\Gamma(y)$ and Ψ equal to the log Burr (1942) distribution $1-(1+\gamma e^t)^{-1/\gamma}$, $\gamma>0$; $1-\exp(-e^t)$, $\gamma = 0$. The parameter β_j is the increase in the log γ-odds as x_j is increased by one unit. This model is equivalent to models considered by Harrington & Fleming (1982), Clayton & Cuzick (1986), and Bickel (1986), among others.

EXAMPLE 1.4. (Generalized probit model). In the case of dose response studies, the natural extension of the probit model for binary data (Bliss (1935)) to continuous data is given by

$$\Phi^{-1}(P(Y \le t)) = \Phi^{-1}(F_0(t)) - \sum \beta_j x_j,$$

where F_0 is some unknown distribution function and Φ is the standard normal distribution function. Here (1.1) is satisfied with $\Psi = \Phi$ and $h(t) = \Phi^{-1}(F_0(t))$.

Next we give the classical parametric transformation model and one of its most important special cases.

EXAMPLE 1.5. (Power Transformation Model). Here h(y) is specified as

$$h(y) = y^\lambda \quad \text{or} \quad h(y) = [y^\lambda - 1]/\lambda,\ \lambda \ne 0,\ = \log y,\ \lambda = 0,\ y>0$$

and Ψ is taken to be the standard normal distribution Φ. Such transformations have been considered by Anscombe & Tukey (1954), Tukey (1957), Box & Cox (1964), Bickel & Doksum (1981), and Johnson (1982), among others. When Y is a survival or failure time, the above specifications are not compatible since h(Y) will be bounded below which is not possible when $\Psi = \Phi$. Thus, in this case, h(Y) should be regarded as the transform of the log of the failure time. Moreover, to avoid problems with roots of negative numbers, the transformation $\mathrm{sign}(z)|z|^{\lambda}$ or $h(z) = [\mathrm{sign}(z)(|z|^{\lambda}) - 1]/\lambda$, where $z = \log y$, should be used.

EXAMPLE 1.6. (The Accelerated Failure Time Model). This is a special case of Example 1.5 with $h(z) = z$. It has been considered by Miller (1976), Buckley & James (1979) and Koul, Susarla & van Ryzin (1981), among others.

Note that in examples 1-3, semiparametric models are constructed by giving an equation of the form: (rate function | **x**) = (baseline rate function) (parametric function of **x**). Model (1.1) unifies the various rate function approaches by defining a semiparametric model which equates a transformation of the response variable with a parametric function of **x** plus an error term.

In addition to the references listed in Examples 1-4, the case of nonparametric h has also been considered by Fisher (1946, Example 46.2), Kruskal (1965), and Breiman & Friedman (1985), among others. The alternating conditional expectation (ACE) procedure of Breiman & Friedman is more flexible in that it allows arbitrary (not necessarily monotone) transformations of both the response variable and the covariates. However, even though the procedures are hard to compare because of their somewhat different purposes, we expect our procedure to perform well in comparison with ACE since it is likelihood based.

Next consider n independent responses $Y_1, \ldots, Y_n$ satisfying (1.1). In this case we use the notation

$$h(Y_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i,\ i = 1, \ldots, n$$

where $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{ip})$ are the covariates corresponding to Y_i, and $\varepsilon_1, \ldots, \varepsilon_n$ are independent identically distributed with distribution Ψ. Since all additive constants can be incorporated in h, we assume $\sum_{i=1}^{n} x_{ij} = 0$, $j = 1, \ldots, p$.

For the estimation of $\boldsymbol{\beta}$, we use a partial (marginal or rank) likelihood which can be computed without knowing h. In Section 2, we show that the partial likelihood is the projection of the h - known likelihood onto the space of rank statistics, and we obtain a Hoeffding (1951) type formula where the partial likelihood is expressed as an expected value of relative likelihoods. In Section 3, we use this expression to construct a resampling scheme, *the likelihood sampler* , to compute approximations to the partial likelihood and the maximum partial likelihood estimate. We refer to this estimate as the *approximate maximum partial likelihood estimate* (*AMPLE*). The likelihood sampler is illustrated on the Pike (1966) cancer data and the Stanford Heart Transplant Data (Miller & Halpern (1982)).

In Section 4, we use Aalen's (1978) counting processes approach to obtain a Le Cam type locally asymptotically normal (LAN) representation of the h - known likelihood, and then show that in the case of i.i.d. censoring, a likelihood defined from the censored rank vector approximates this likelihood in a neighborhood of $\beta = 0$. This generalizes to censored data the Le Cam - Hájek - Šidák result (see Hájek & Šidák (1967), pp 245 and 275) that the ranks are asymptotically sufficient. Next we use this rank approximation to the likelihood to construct estimates of β called *local maximum partial likelihood estimates* (*LMPLE*).

One important question relates to the choice of model: If we use the estimate derived assuming the proportional hazard model, how well will this estimate behave if the model generating the data is actually the proportional odds model? More generally, in our linear transformation model framework, this question is: If we use an estimate $\hat{\beta}$ assuming error distribution Ψ when in fact the true error distribution is Ψ_0, how well will $\hat{\beta}$ behave? It turns out that asymptotically, as $n \rightarrow \infty$, the squared bias dominates the variance in the usual breakup of the MSE (Mean Squared Error) into the sum of squared bias and variance. In fact, $n(\text{bias})^2 \rightarrow \infty$, while n(variance) stays bounded. On the other hand, Monte Carlo simulation shows that for moderate sample sizes and for a range of parameter values, squared bias does not dominate the variance. In order to have an asymptotic theory that reflects the actual behaviour of MSE for moderate sample sizes we resort to the trick of letting the parameter set shrink as sample size increases. More precisely, we consider the parameter set

$$\Omega_n = \{\beta : \sum(\mu_i - \bar{\mu})^2 \le K^2,\ \max|\mu_i - \bar{\mu}| = o_n(1)\}$$

where $\mu_i = \sum x_{ij}\beta_j$ and K^2 is an arbitrary positive constant not depending on n. This is the usual contiguous parameter set used in asymptotic testing theory. In estimation, a similar set has been used by Pettitt (1983) to derive approximate partial likelihood estimates and by Solomon (1984) to study the performance of estimates derived from the proportional hazard model in the accelerated life model and vice versa. It also appears to be an implicit assumption in part of the statistical literature, e.g. Box & Tsiao (1973). Thus, if $\beta \in \Omega_n$, the apparently contradictory results of Box & Tsiao (1973, Section 10.4) and Bickel & Doksum (1981) can be reconciled. See also Doksum & Wong (1983) and Carroll (1982).

In Section 5 we find that Ω_n does the trick: It gives (1) good approximations to the MSE and (2) formulae that yield insight into the properties of the estimates. It may be argued that the model we use to achieve (1) and (2) is unrealistic since in order to keep the squared bias and variance of the same order we must let the parameter set shrink as n tends to infinity. But note that for the moderate sample sizes used in medical trials, our parameter set is quite substantial, e.g. for $n=40$, it is $[-2,2]$ in the two sample case. Moreover, the purpose of asymptotic theory has always been to achieve good approximations and insight into properties of statistical procedures since, inherently, $n = \infty$ is unrealistic. Using the Ω_n asymptotics and Monte Carlo simulation, we find that for $p = 1$, using the wrong Ψ in the estimation procedure has a very small effect in the performance of the estimates. Moreover, for β in a substantial neighborhood of 0, the variance dominates the squared bias and the MSE's of the rank estimates are equal to or exceed by a small to moderate amount the MSE's of the optimal h-known estimates.

We conclude this section with some assumptions and notation. We assume that the design matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is of full rank and that Y_i is nonnegative. Thus our model assumes that for some increasing function h, (A.1) $h(Y_i) = x_i^T\beta + \varepsilon_i,\ i = 1,\ldots,n,\ \sum_{i=1}^{n} x_i = 0,\ \varepsilon_i \sim \Psi$ are i.i.d., $Y_i \geq 0$.

(A.2) $$\text{Rank}\,(\mathbf{X}) = p.$$

The $Y\,'$ s are censored on the right by $\tilde{Y}_1,\ldots,\tilde{Y}_n$ and we observe

$$T_i = \min\{Y_i, \tilde{Y}_i\} \text{ and } D_i = I[Y_i \leq \tilde{Y}_i],\ i = 1,\ldots,n$$

where I denotes the indicator function. In many instances, we need to assume

(C.1) $\tilde{Y}_1,\ldots,\tilde{Y}_n$ are independent random variables independent of $Y_1,\ldots,Y_n$.

(C.2) The support of $\tilde{Y}_i$ intersects the support of Y_i, $i = 1, \ldots, n$.

Finally, we assume

(A.3) The distribution Ψ has probability density ψ such that $\psi(x) > 0$ for all $x \in R$ and h is a transformation mapping the support of Y_i onto the real line.

To simplify notation we shall write $\mu_i = x_i^T\beta$, $f_{\mu_i}(t) = \psi(h(t) - \mu_i)$, $F_{\mu_i}(t) = \Psi(h(t) - \mu_i)$, $\bar{F}_{\mu_i} = 1 - F_{\mu_i}$, $\bar{\Psi} = 1 - \Psi$. Further, let $\lambda(t) = \psi(t)/(1 - \Psi(t))$ be the hazard rate and $\Lambda(t) = -\log(1 - \Psi(t))$ be the integrated hazard. Set $\lambda_{\mu_i}(t) = \lambda(h(t) - \mu_i)$ and $\Lambda_{\mu_i}(t) = \Lambda(h(t) - \mu_i)$.

2. Rank, marginal and partial likelihood. Suppose that h is regarded as a nuisance parameter in H = { increasing transformations mapping the support of Y onto R}. Our model is invariant under the group G = {increasing transformations mapping the support of Y onto the support of Y}. For the model (A.1) with h in H, the rank likelihood, which we describe below, is equivalent to the more general concepts of partial likelihood and marginal likelihood. See Cox (1972, 1975), Kalbfleisch & Prentice (1973, 1980, p. 72), Kalbfleisch (1978), Prentice (1978), Pettitt (1982, 1983, 1984), and Wong (1986).

Let $S = (S_1,\ldots,S_n)$ be the vector of ranks of $Y_1,\ldots,Y_n$ in the *uncensored version of the experiment*. Owing to censoring, S is only partially observable, however information carried by the data can be used to construct a set $\mathcal{S}$ of possible values of S. $\mathcal{S}$ can be described in terms of $D_1,\ldots,D_n$ and the *censored rank vector* $\mathbf{R} = (R_1,\ldots,R_n)$ defined by

$$R_i = \sum_{j=1}^{n} D_j I[T_j \le T_i], \quad i = 1,\ldots,n.$$

Note that here we rank the uncensored observations among themselves and then assign to each censored observation the same rank as to the nearest uncensored observation on the left. The set $\mathcal{S}$ is now taken to be the collection of all rank vectors $(s_1,\ldots,s_n)$ of $y_1,\ldots,y_n$ compatible with the observed censored rank vector $r_1,\ldots,r_n$ and the indicators $d_1,\ldots,d_n$ of censoring.

Let

$$L(\beta) = P(S \in \mathcal{S})$$

be the *censored rank likelihood,* where $P = \tilde{P}_\beta$ refers to the (unconditional) distribution of $Y_1,\ldots,Y_n$ in the underlying *uncensored* version of the experiment. The following proposition establishes the connection between $L(\beta)$ and the likelihood

$$\tilde{L}(\beta) = d\tilde{P}_\beta/d\tilde{P}_o = \prod_{i=1}^{n} \frac{\psi(h(y_i)-\mu_i)}{\psi(h(y_i))}$$

appropriate when h is known and there is no censoring.

PROPOSITION 2.1. *For the model* (A.1), (A.2), (A.3)

$$E_{\tilde{P}_o}(\tilde{L}(\beta) \mid S \in \mathcal{S}) = L(\beta)/L(0).$$

PROOF. We have $L(0)E_{\tilde{P}_0}(\tilde{L}(\beta) \mid S \in \mathcal{S}) = \int_{\mathcal{S}} \tilde{L}(\beta)d\tilde{P}_0 = L(\beta)$.

REMARK 2.1. The above result shows that the censored rank likelihood $L(\beta)$ is proportional to the projection of the likelihood $\tilde{L}(\beta)$ onto the space of functions of S which are constant on $\mathcal{S}$. Thus, in the class of functions of $\underset{\sim}{R}$ and $\underset{\sim}{D}$, $L(\beta)/L(0)$ is the closest (in the $L_2(\tilde{P}_o)$ sense) to $\tilde{L}(\beta)$.

The following result gives a Hoeffding (1951) type representation of the partial likelihood. Let $k = \sum_{i=1}^{n} d_i$, be the number of uncensored observations in the sample.

PROPOSITION 2.2. *For the model defined by* (A.1), (A.2) *and* (A.3)

$$L(\beta) = \frac{1}{k!} E \left\{ \prod_{i=1}^{n} \left[\frac{\psi(V^{(r_i)} - \mu_i)}{\psi(V^{(r_i)})} \right]^{d_i} \left[\overline{\Psi}(V^{(r_i)} - \mu_i) \right]^{(1-d_i)} \right\} \tag{2.1}$$

where $V^{(1)} < \ldots < V^{(k)}$ *are the order statistics in a sample of size* k *from* Ψ *and* $V^{(0)} = -\infty$.

PROOF. The proof is very similar to Kalbfleisch & Prentice's (1973) marginal likelihood derivation and is omitted.

REMARK 2.2. Since the ranks are invariant under increasing transformations, $L(\beta)$ does not depend on h.

3. A resampling scheme for estimating the partial likelihood. Results of Section 2 imply that in the transformation model (A.1), the regression parameters β can be estimated without first estimating the transformation h. We shall consider now a resampling scheme, called the likelihood sampler, for estimating the partial likelihood and the parameter β. A somewhat different estimation procedure will be considered in Section 4.

The likelihood sampler can be described as follows. On the computer, generate B independent ordered samples $V_j^{(1)} < \ldots < V_j^{(k)}$, $j = 1, \ldots, B$, from Ψ. Then approximate $k!L(\beta)$ by $L_B(\beta)$, where

$$L_B(\beta) = \frac{1}{B} \sum_{j=1}^{B} g_j(\beta), \text{ and} \tag{3.1}$$

$$g_j(\beta) = \prod_{i=1}^{n} \left[\frac{\psi(V_j^{(r_i)} - \mu_i)}{\psi(V_j^{(r_i)})} \right]^{d_i} \left[\overline{\Psi}(V_j^{(r_i)} - \mu_i) \right]^{(1-d_i)}$$

$$= \prod_{i=1}^{n} \left[\frac{\lambda(V_j^{(r_i)} - \mu_i)}{\psi(V_j^{(r_i)})} \right]^{d_i} \overline{\Psi}(V_j^{(r_i)} - \mu_i)$$

Now, by Proposition 2.2 and the strong law of large numbers.

PROPOSITION 3.1. *Under the conditions of Proposition* 2.2, $L_B(\beta)$ *converges almost surely to* $k!L(\beta)$ *as* $B \to \infty$.

We estimate β by $\hat{\beta}_B$ maximizing $L_B(\beta)$ for $B = 200$ or $B = 400$. Experience shows that not much accuracy is gained by going beyond $B = 400$ when $p = 1$. For $p > 1$ larger B will be required but we do not have enough experience to give exact recommendations.

REMARK 3.1 *Ties*. Ties among the censored observations are automatically handled by the above approach. We can easily modify our procedure to handle data sets with tied uncensored observations. We view ties as the result of rounding off in the underlying continuous model. Moreover, we take the point of view that we are trying to compute the same $L(\beta)$ as before. Suppose there are γ_i tied uncensored observations at t_i^0, $i = 1, \ldots, e$. At the same time we select $V_j^{(1)} < \ldots < V_j^{(k)}$, we independently select $\sum_{i=1}^{e} \gamma_i = s$ uniform $(-\delta, 0]$ variables $W_j^1, \ldots, W_j^s$ and add one to each of the tied observations. δ is chosen to equal 1/2 times the minimum absolute difference between uncensored, untied observations. Now compute the ranks of the new uncensored observations, the censored observations, $g_j(\beta)$, and proceed as before.

EXAMPLE 3.1. We illustrate the likelihood sampler on the Pike cancer data (Pike (1966), Kalbfleisch & Prentice (1980), pp. 2, 82). This data gives days to cancer mortality for two groups of rats distinguished by a pretreatment regime prior to exposure to a carcinogen.

Group 1: 143, 164, 188, 188, 190, 192, 206, 209, 213, 216, 220, 227, 230, 234, 246, 265, 304, 216*, 244*

Group 2: 142, 156, 163, 198, 205, 232, 232, 233, 233, 233, 233, 239, 240, 261, 280, 280, 296, 296, 323, 204*, 344*

Censoring times are indicated with a star. There are no ties between groups. The ties within groups can be broken in any way at all without changing the analysis. Let $\hat{\beta}_B$ denote the value of β that maximizes $L_B(\beta)$.

For the normal likelihood sampler where $\Psi = \Phi$, the estimate $\hat{\beta}_B$ was equal to .454, .467, .451 and .463 for $B=100$, 200, 400 and 800, respectively. The corresponding $L_B(\hat{\beta}_B)$ values were .549, .579, .595 and .621. The random variables used in the likelihood sampler with $B = 100$ were independent of those used in the $B = 200$ sampler, and so on. For the logistic likelihood sampler, the estimate $\hat{\beta}_B/\sigma_L$ was equal to .473, .479, .486 and .501 for $B=100$, 200, 400 and 800. Here σ_L is the standard deviation of the logistic distribution. The corresponding $L_B(\hat{\beta}_B)$ values were .801, .901, .877 and .971. Thus the likelihood principle suggests the logistic transformation (proportional odds) model as the better one for this data set. For the fitted logistic model the interpretation is that the odds rate for group 2 rats is about 40% ($e^{-.9} = .4$) that of group 1 rats. Note that both the logistic and the normal model yield values of the estimates of group difference to residual variability close to .5, but remember that the two models refer to different transformations, h. It would have been of interest to estimate the two transformations for comparison and interpretation, but we have chosen not to study methods for such estimation in the present paper.

To further study the effect of the random numbers used in the resampling scheme on the value of $\hat{\beta}_B$, we ran 50 independent trials where in each trial we used the normal likelihood sampler with $B = 100$ to compute a value for $\hat{\beta}_B$. The mean and standard deviation of the resulting 50 values of $\hat{\beta}_B$ were 0.460 and 0.020, respectively. We repeated this experiment using $B = 400$. This time the mean and standard deviation of the 50 $\hat{\beta}_B$ values were 0.457 and 0.0085, respectively. The standard deviation 0.0085 should be compared to the standard error 0.333 of the asymptotically optimal estimate in the censored normal model (from Example 4.3). In other words, when $B = 400$ the extra randomness introduced by the likelihood sampler is very small when compared to the variability already present in the data.

EXAMPLE 3.2. We consider the Stanford Heart Transplant Data as reported by Miller & Halpern (1982). The response variable Y is survival time after transplant, the covariate x is age at time of transplant and n = 184. We find that for the normal sampler with B = 400, $\hat{\beta}_B = -.0093$ with standard error (from Section 4) $(\hat{I}\,\Sigma x_i^2)^{-\frac{1}{2}} = .0082$. Thus we cannot reject the hypothesis that β is zero.

REMARK 3.2. Monte Carlo sampling techniques have also been used by Diggle & Gratton (1984) and Beran & Millar (1986) to estimate likelihood functions. The likelihood sampler is different in that it takes advantage of Hoeffding's formula.

4. A local approximation to the likelihood. Suppose that $Y_1, \ldots, Y_n$ satisfy the linear transformation model (A.1). In this section we shall consider conditions under which the regression parameters β can be estimated adaptively when the transformation h is unknown. A convenient tool is provided by the theory of multivariate counting processes and stochastic integrals. We refer to Aalen (1978), Gill (1980), Andersen et al. (1982), Andersen & Gill (1982) and Andersen & Borgan (1985) for a detailed discussion of relevant results.

Recall that Ψ has an absolutely continuous density ψ and that $\lambda = \psi/(1-\Psi)$. Define

$$l = \log\lambda,$$

and assume

(A.4) $\quad l' = (\psi'/\psi) + \lambda$ has bounded variation and $0 < I(\Psi) = \int (l')^2 d\Psi < \infty$.

Let $G_{i(n)}$ denote the distribution of the censoring variable $\tilde{Y}_i$, $i = 1, \ldots, n$. We assume conditions (C.1) and (C.2), and

(C.3) $\quad$ There exist distribution functions $G_1, \ldots, G_n$ and G such that $\lim_{n\to\infty} G_{i(n)}(t) = G_i(t)$ and

$$\lim_{n\to\infty} n^{-1}\sum_i G_{i(n)}(t) = G(t),\ t \in R.$$

Define processes

$$N_i(t) = I(T_i \le t, D_i = 1), \quad N(t) = \sum_{i=1}^{n} N_i(t)$$

$$Z_i(t) = I(T_i \ge t), \quad Z(t) = \sum_{i=1}^{n} Z_i(t)$$

N_i and N are counting processes with intensity processes $Z_i\lambda_{\mu_i}$ and $\sum Z_i\lambda_{\mu_i}$ respectively, where λ_{μ_i} is the hazard rate of Y_i and $\mu_i = x_i^T\beta$. Then, using results of Aalen (1978),

$$M_i(t) = N_i(t) - \int_0^t Z_i d\Lambda_{\mu_i}, \quad M(t) = N(t) - \int_0^t \sum Z_i d\Lambda_{\mu_i}$$

are mean zero square integrable martingales with predictable variation processes

$$\begin{aligned} \langle M_i, M_j \rangle(t) &= 0 \quad \text{if } i \ne j \\ &= \int_0^t Z_i d\Lambda_{\mu_i} \quad i = j \end{aligned}$$

$$\langle M, M \rangle(t) = \int_0^t \sum_{i=1}^{n} Z_i d\Lambda_{\mu_i}.$$

Here Λ_{μ_i} is the integrated hazard of Y_i.

For the moment we shall assume that the transformation h is known. For any $\boldsymbol{\beta} \in R^P$, let P_{β} denote the probability measure induced by $T_1, \ldots, T_n$ and $D_1, \ldots, D_n$. Further, let $\mathbf{B}_n(\beta)$ be the p×p symmetric matrix

$$\mathbf{B}_n(\beta) = \sum_{i=1}^{n} I(\mu_i, \Psi, G_i)\mathbf{x}_i\mathbf{x}_i^T$$

where $I(\mu_i, \Psi, G_i) = \int_0^{\infty} (l'_{\mu_i})^2 z_i d\Lambda_{\mu_i}$, $z_i = \overline{G}_i\overline{F}_{\mu_i}$, and $l'_{\mu_i}(t) = l'(\dot{h}(t) - \mu_i)$.

We shall assume

(A.5) $\max_i x_i^T(X^TX)^{-1}x_i \to 0$ as $n \to \infty$

(A.6) $\lim_n \inf I(\mu_i, \Psi, G_i) = \eta > 0$ for some η.

REMARK 4.1. Assumption (A.6) ensures that the distributions G_i of the censoring variables do not have supports disjoint with the support of F_{μ_i}.

REMARK 4.2. By (A.2), (A.4), (A.6), for any $a \neq 0$, $a \in R^p$ we have

$$I(\Psi)a^TX^TXa \geq a^TB_n(\beta)a \geq \eta a^TX^TXa > 0$$

so that the matrix $B_n(\beta)$ is positive definite. By (A.5) and (A.6),

$$\max_i x_i^TB_n^{-1}(\beta)x_i \leq \eta^{-1}\max_i x_i^T(X^TX)^{-1}x_i \to 0.$$

Fix $\beta_0 \in R^p$ and define a $p\times 1$ vector $T(t)$ by

$$T(t) = -B_n^{-1/2}(\beta_0)\sum_{i=1}^n x_i\int_0^t l'_{\mu_i}(dN_i - Z_i d\Lambda_{\mu_i})$$

$$= -B_n^{-1/2}(\beta_0)\sum_{i=1}^n x_i\int_0^t l'_{\mu_i} dM_i$$

THEOREM 4.1. *Suppose that the conditions* (A.1) - (A.5), (C.1) *and* (C.2) *are satisfied and let* $\beta_0 \in R^P$ *be such that* (A.6) *holds*

(i) *Let* $\beta = \beta_0 + B_n^{-1/2}(\beta_0)a_n$ *where* $\{a_n\}$ *is a bounded sequence in* R^p. *Then, under* P_{β_0}

$$\log(dP_\beta/dP_{\beta_0}) = a_n^TT(\infty) - \frac{1}{2}a_n^Ta_n + o_{P_{\beta_0}}(1)$$

which converges weakly to a normal distribution $N(-a_n^Ta_n/2,\ a_n^Ta_n)$.

(ii) *Let* $u: R^P \to R$ *be a subconvex loss function. Then*

$$\lim_{k\to\infty}\liminf_n \inf_{\hat\beta} \sup_{a_n^Ta_n \leq K^2} E_\beta u[B_n^{-1/2}(\beta_0)(\hat\beta - \beta)] \geq (2\pi)^{-p/2}\int_{R^p} u(s)\exp(-s^Ts/2)ds$$

where K *is a constant independent of* n, $\beta = \beta_0 + B_n^{-1/2}(\beta_0)a_n$ *and the infimum is taken over sequences of estimates* $\hat{\beta}$ *that are regular at* β_0 *is the sense of Hájek (1972).*

Part (i) of the theorem establishes that when the transformation h is known, the family of underlying distributions is locally asymptotically normal (LAN) in the sense of Le Cam (1969). By the Hájek (1970) convolution theorem, asymptotically any sequence of estimates regular at β_0 is at least as dispersed as $N(0, B_n^{-1}(\beta_0))$. Part (ii) provides the asymptotic minimax bound for the risk associated with the loss function u. To prove (i), we sketch, in Section 6, an argument based on Rebolledo's martingale central limit theorem (Theorem I.2 in Andersen & Gill (1982)). The technical details are similar to those of the two-sample case considered in Gill (1980, pp. 118-122). An alternative proof is given by Koul & Wang (1984).

REMARK 4.3. Note that if $\beta = \beta_0 + B_n^{-1/2}(\beta_0)a_n$, $\mu_{i0} = x_i^T\beta_0$, $\mu_i = x_i^T B_n^{-1/2}(\beta_0)a_n + \mu_{i0}$ and $a_n^T a_n \le K^2$, then by Remark 4.3 and the Cauchy-Schwarz inequality we have

$$(\mu - \mu_0)^T(\mu - \mu_0) = a_n^T B_n^{-1/2}(\beta_0) X^T X B_n^{-1/2}(\beta_0) a_n \le \eta^{-1} K^2 \quad \text{and}$$

$$\max_i (\mu_i - \mu_{i0})^2 = \max_i (x_i^T B_n^{-1/2}(\beta_0) a_n)^2 \le K^2 \max_i x_i^T B_n^{-1}(\beta_0) x_i \to 0$$

REMARK 4.4. Assuming that the transformation h is known, under conditions of Theorem 4.1, one can construct asymptotically efficient estimates of β using Le Cam's (1969) one-step maximum likelihood procedure. Under additional conditions, namely that $\log\psi$ and $\log\overline{\Psi}$ are continuously twice differentiable and have negative second derivatives, Barndorff-Nielsen & Blaesild (1980) show that the maximum likelihood estimate exists and is unique. Theorem 4.1 can be applied to establish its asymptotic efficiency.

We now drop the assumption that the transformation h is known. In general, the parameter β cannot be estimated adaptively, i.e. with the same accuracy as when h is known.

Consider for instance the Cox proportional hazard model. In the two sample case with both uncensored and censored data, Begun (1982) and Begun & Wellner (1983) showed that if $\beta_0 \neq 0$ then every rank estimator of β that is regular at β_0, has a limiting distribution more dispersed than $N(0, \mathbf{B}_n^{-1}(\beta_0))$. Moreover, the optimal rank estimator (the Cox partial likelihood estimate) is asymptotically best among all regular estimates of β. See Begun et al. (1983) for the treatment of the general Cox proportional hazards regression model.

It is conceivable that this situation carries over to other transformation models. The maximum rank likelihood estimate and its likelihood sampler version is expected to be optimal among rank estimates and asymptotically best among all regular estimates of β. The information contained in the data cannot be used, however, to "estimate away" the unknown transformation h and thereby estimate β as if h were known. To show this one needs to extend Theorem 4.1 to a LAN and a Hájek-Le Cam asymptotic minimax result so that both components of the model -- the parametric β and nonparametric h -- are taken into account. This a difficult and open problem.

We shall consider now local neighborhoods of $\beta_0 = 0$ and show that equal limiting censoring is a sufficient condition for adaptability of β. Define the local parameter set

$$\Omega_n = \{\beta : \beta = \mathbf{B}_n^{-1/2}\mathbf{a}_n,\ \mathbf{a}_n^T\mathbf{a}_n \leq K^2\}$$

where $\mathbf{a}_n$ is a $p \times 1$ vector, K is a constant independent of n and $\mathbf{B}_n = \mathbf{B}_n(0)$. Note that $\mathbf{B}_n$ can be estimated consistently by

$$\hat{\mathbf{B}}_n = \sum \mathbf{x}_i \mathbf{x}_i^T \int_0^\infty l'(\Psi^{-1}(\hat{F}))^2 \, dN_i$$

where $\hat{F}$ is the left continuous version of the Kaplan-Meier (1958) estimator based on the combined sample.

Define a $p \times 1$ vector statistic

$$(4.1) \qquad S(t) = -\hat{B}_n^{-1/2}\sum_{i=1}^{n} x_i \int_0^t l'(\Psi^{-1}(\hat{F}))[dN_i - (V/Z)Z_i dN]$$

where $V(t) = I(Z(t) > 0)$ and $0/0 = 0$. Note that the statistic S depends entirely on the processes N_i and Z_i, in particular it does not depend on the unknown transformation h or the parameters β.

Let $F(t) = F_0(t) = \Psi(h(t))$. We assume

(A.7) l' either has a limit at $x = -\infty$ and is bounded on $(-\infty,\infty)$, or alternatively l' is bounded on $[x,\infty)$ for each $x > -\infty$.

(A.8) For all $\varepsilon > 0$

$$\lim_{t \downarrow 0} \overline{\lim_{n\to\infty}} P\left(\int_0^t [l'(\Psi^{-1}(\hat{F}))]^2 dF > \varepsilon\right) = 0.$$

The following theorem is an extension of results obtained by Gill (1980, Chapter 5) and Andersen et al. (1982) in the k-sample set-up. The proof is given in Section 6.

THEOREM 4.2. *Suppose that the conditions* (A.1) - (A.5), (A.7), (A.8), (C.1) - (C.3) *are satisfied and that* (A.6) *holds for* $\beta_0 = 0$. *Let* $\beta \in \Omega_n$ *and let* $\mathbf{b}_n$ *be a bounded sequence in* R^p. *Assuming that* h *is known, then, under* P_0

$$(\log(dP_\beta / dP_0), \mathbf{b}_n^T S(\infty))^T$$

converges weakly to a bivariate normal distribution with mean $(-1/2\mathbf{a}_n^T\mathbf{a}_n, 0)^T$ *and covariance matrix*

$$\Sigma_{ab} = \begin{bmatrix} \mathbf{a}_n^T\mathbf{a}_n & \mathbf{b}_n^T(I-C_n)\mathbf{a}_n \\ \mathbf{b}_n^T(I-C_n)\mathbf{a}_n & \mathbf{b}_n^T(I-C_n)\mathbf{b}_n \end{bmatrix}$$

where I *is a* p×p *identity matrix,*

$$C_n = \frac{1}{n} B_n^{-1/2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j^T \int_0^\infty [l'(\Psi^{-1}(F))]^2 (v/z) z_i z_j d\Lambda_0\right) B_n^{-1/2}$$

and $z_i = (1-G_i)(1-F)$, $z = (1-G)(1-F)$, $v(t) = I(z(t)>0)$.

We shall assume now that

(C.4) the limiting censoring distributions G_i are equal.

Under this assumption $z_i = z$ and

$$B_n = (X^TX)\int_0^\infty l'(\Psi^{-1}(F))^2 z\,d\Lambda_0$$

which can be estimated consistently by

$$\tilde{B}_n = (X^TX)[n^{-1}\int_0^\infty l'(\Psi^{-1}(\hat{F}))^2 dN].$$

Replacing $\hat{B}_n$ by $\tilde{B}_n$ in the definition of $S(\infty)$ we conclude that for $\beta \in \Omega_n$, $a_n^T S(\infty) - \frac{1}{2} a_n^T a_n$ has the same asymptotic distribution as $\log(dP_\beta/dP_0)$ with the transformation h being known. More precisely, let

$$L_R(\beta) = \exp\{a_n^T S(\infty) - \frac{1}{2} a_n^T a_n\}$$

with a_n as in the definition of Ω_n.

THEOREM 4.3 *Under* (C.4) *and the conditions of theorem* 4.2, *there exists a sequence* c_n *depending on the data only through* $S(\infty)$ *such that for* $\beta \in \Omega_n$, $c_n L_R(\beta)\,dP_0$ *is a density,* $c_n \to 1$ *in probability as* $n \to \infty$, *and*

$$\lim_{n\to\infty} \sup_{\beta \in \Omega_n} \int |(dP_\beta/dP_0) - c_n L_R(\beta)|\,dP_0 = 0.$$

COROLLARY 4.1. *Under the conditions of Theorem 4.3, for sequences of alternatives in* Ω_n, *the maxmin most powerful level* α *test based on the censored data ranks* $R_1, \ldots, R_n$ *and the indicator variables* $D_1, \ldots, D_n$ *is asymptotically maxmin most powerful in the class of all level* α *tests based on the censored data* $\min\{h(Y_i), h(\tilde{Y}_i)\}$, $I[h(Y_i) \le h(\tilde{Y}_i)]\}$, $i = 1,\ldots,n$, *with* h *assumed known.*

These results follow from Theorem 4.2 and the arguments in Hájek & Šidák (1967, pp. 242-249).

REMARK 4.5. Theorem 4.3 extends to censored data the Le Cam-Hájek-Šidák (Hájek & Šidák, pp. 245 and 275) result that the ranks are asymptotically sufficient. This follows since $S(\infty)$ is easily seen to be a function of the censored ranks $R_1, \ldots, R_n$ and the indicator variables $D_1, \ldots, D_n$ only.

Since $L_R(\boldsymbol{\beta})$ is close to the likelihood, we will use the value of $\boldsymbol{\beta}$ that maximizes $L_R(\boldsymbol{\beta})$ as an estimate of $\boldsymbol{\beta}$. $L_R(\boldsymbol{\beta})$ is maximized by $\mathbf{a}_n^* = S(\infty)$ so that for $\boldsymbol{\beta} \in \Omega_n$ the maximum is achieved at $\boldsymbol{\beta} = \mathbf{B}_n^{-1/2} S(\infty)$. Substituting $\tilde{\mathbf{B}}_n$ for $\mathbf{B}_n$ we obtain the estimate

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{A}/\hat{\mathbf{I}} \tag{4.2}$$

where $\mathbf{A}^T = (A_1, \ldots, A_n)$ and A_i is the censored rank score

$$A_i = \int_0^\infty l'(\Psi^{-1}(\hat{F}))[dN_i - (V/Z)Z_i dN] \tag{4.3}$$

and

$$\hat{\mathbf{I}} = n^{-1}\int_0^\infty l'(\Psi^{-1}(\hat{F}))^2 dN. \tag{4.4}$$

REMARK 4.6. Note that $\tilde{\boldsymbol{\beta}}$ is the least squares estimate computed from $(\tilde{Y}_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, where $\tilde{Y}_i = A_i/\hat{\mathbf{I}}$.

COROLLARY 4.2. *Under the conditions of Theorem* 4.3, *for* $\boldsymbol{\beta} \in \Omega_n$, $(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ *has an asymptotic* p*-variate normal distribution* $N(0, \mathbf{B}_n^{-1})$.

COROLLARY 4.3. *Under the conditions of Theorem* 4.3, *for* $\boldsymbol{\beta} \in \Omega_n$, $\tilde{\boldsymbol{\beta}}$ *is an adaptive estimator in the sense of having the same limiting distribution as an efficient estimator of* $\boldsymbol{\beta}$ *with* h *assumed to be known.*

Here are some examples. Let $C = (c_{ji})$ be the p×n matrix given by $C = (X^TX)^{-1}X^T$.

EXAMPLE 4.1. (Proportional hazards). In the case of $\Psi(t) = 1 - \exp\{-e^t\}$ we have $l' \equiv 1$, (4.3) reduces to the familiar log-rank scores and

$$\tilde{\beta}_j = \frac{-n}{\sum_{i=1}^n D_i} \sum_{i=1}^n c_{ji} \int_0^\infty [dN_i - (V/Z)Z_i dN].$$

For the Pike data of Example 3.1, $\tilde{\beta} = .511$. Moreover $\tilde{\beta}/\sigma_E = .398$ with estimated standard error $\sigma_E^{-1}(\Sigma x_i^2(36/40))^{-1/2} = 0.260$ and standardized score $\tilde{\beta}/$ (standard error $\tilde{\beta}$) equal to 1.531. Here σ_E is the standard deviation of the extreme value distribution.

EXAMPLE 4.2. (Proportional γ-odds). The choice $\Psi(t) = 1 - (1 + \gamma e^t)^{-1/\gamma}$, gives $l'(t) = (1 - \Psi(t))^\gamma$ and leads to scores considered by Harrington & Fleming (1982) in the two-sample problem. The estimator $\tilde{\beta}_j$ takes the form

$$\tilde{\beta}_j = \frac{-n}{\int_0^\infty (1 - \hat{F})^{2\gamma} dN} \sum_{i=1}^n c_{ji} \int_0^\infty (1 - \hat{F})^\gamma [dN_i - (V/Z)Z_i dN].$$

Recall that $\gamma = 1$ yields the proportional odds rate model. In this case, for the Pike data, $\tilde{\beta} = .903$. Moreover, $\tilde{\beta}/\sigma_L = .498$ with estimated standard error $\sigma_L^{-1}(\hat{I}\,\Sigma x_i^2)^{-1/2} = .308$ and standardized score 1.617. Here σ_L is the standard deviation of the logistic distribution. For the Stanford Heart Transplant Data of Example 3.2, $\tilde{\beta}/\sigma_L = -.015$ with estimated standard error .0076.

EXAMPLE 4.3. In the case of $\Psi = \Phi$, the standard normal distribution, $l'(t) = -t + [1 - \Phi(t)]^{-1}\phi(t)$. If we replace the Kaplan-Meier estimator $\hat{F}$ by $\tilde{F} = (n\hat{F} + 1)/(n+1)$ as in Gill (1980, p.. 127), then the asymptotically optimal estimate takes the form

$$\tilde{\beta}_j = -n \sum_{i=1}^n c_{ji} \int_0^\infty J(\tilde{F})[dN_i - (V/Z)Z_i dN] \Big/ \int_0^\infty J^2(\tilde{F}) dN$$

where $J(u) = -\Phi^{-1}(u) + (1-u)^{-1}\phi(\Phi^{-1}(u))$. For the Pike data, $\tilde{\beta} = .459$ with estimated standard error $(\hat{I}\,\Sigma x_i^2)^{-1/2} = .333$ and standardized score 1.378. For the Stanford Heart Transplant Data $\tilde{\beta} = -.014$ with estimated standard error .0082.

REMARK 4.7. The scores A_i in (4.3) are consistent with those proposed by Andersen et al. (1982) and Harrington & Fleming (1982) in the context of the k-sample testing problem. Alternative scores can be derived from Prentice (1978), Prentice & Marek (1979) or Kalbfleisch & Prentice (1980, Chapter 6). In general, the scores can be written as $A_i = \int a_u dN_i + \int a_c dN_i'$ where $N_i'(t) = I(Y_i \le t, D_i = 0)$. The processes a_u and a_c depend on N and Z only and satisfy $Z(t)da_c(t) = (a_c(t) - a_u(t))dN(t)$. The scores which yield asymptotically efficient estimates are derived through processes a_u and a_c that estimate $-\psi'(\Psi^{-1}(F))/\psi(\Psi^{-1}(F))$ and $\lambda(\Psi^{-1}(F))$, respectively. In particular, the processes a_u and a_c corresponding to estimates $\tilde{\beta}$ of Examples 4.1 - 4.3, are given by

$a_u = \hat{\Lambda} - 1$	$a_c = \hat{\Lambda}$	$\Psi =$ extreme value
$a_u = (1 + 1/\gamma)(1 - \hat{F})^{\gamma} - 1/\gamma$	$a_c = [(1 - \hat{F})^{\gamma} - 1]/\gamma$	$\Psi =$ log Burr
$a_u = \Phi^{-1}(\tilde{F})$	$a_c = -(1 - \tilde{F})^{-1}\ \phi(\Phi^{-1}(\tilde{F}))$	$\Psi =$ normal

where $\hat{\Lambda}(t) = \int_0^t (V/Z)dN$ is the Nelson-Aalen estimate of the integrated hazard function Λ. The exact scores for $\Psi =$ logistic and approximate scores for $\Psi =$ normal proposed by Prentice (1978) have a similar form with $\hat{F}$ replaced by $\tilde{\tilde{F}}^{+}$ and $\hat{F}$ by $\tilde{\tilde{F}}$ where

$$\tilde{\tilde{F}}(t) = 1 - \prod_{s<t}(1 - dN/(Z+1))$$

is an estimate close to the Kaplan-Meier estimate. We refer to Andersen et al. (1982) and Cuzick (1985) for further discussion.

REMARK 4.8. The problem of extending Corollaries 4.2 and 4.3 to the approximate partial likelihood estimate $\hat{\beta}_B$ can be attacked as follows: Let $\hat{\beta}_j$ be the value of β that maximizes $g_j(\beta)$. Extending arguments of Bell & Doksum (1965) we can find conditions under which $\sqrt{n}(\hat{\beta}_j - \beta)$ converges in probability to zero, $j = 1, \ldots, B$. Using convexity arguments (Barndorff - Nielsen & Blaesild (1980)), $\hat{\beta}_B$ can be shown to be closer in Euclidean norm to β than each $\hat{\beta}_j$. Thus, if B is fixed as $n \to \infty$, $\sqrt{n}(\hat{\beta}_B - \beta)$ converges to zero in probability and Corollaries 4.2 and 4.3 hold if $\hat{\beta}$ is replaced by $\hat{\beta}_B$.

5. Bias, mean squared error and Monte Carlo results. In this section, we first use the results of Section 4 to find approximate expressions for the bias and mean squared error of the local maximum partial likelihood estimate (LMPLE) $\tilde{\beta}$ of Section 4, then we compute Monte Carlo bias and mean squared error of $\tilde{\beta}$ and the approximate maximum partial likelihood estimate (AMPLE) $\hat{\beta}_B$ of Section 3.

We consider the following questions:

(i) Suppose both Ψ and h are known. Then, presumably, we can do much better by using the MLE obtained by maximizing the ordinary likelihood of the censored data. We investigate how close the mean squared error of the AMPLE and the LMPLE is to the mean squared error of the MLE, i.e. how much is lost by not knowing h.

(ii) Suppose we use the estimates $\hat{\beta}$ and $\tilde{\beta}$ corresponding to a certain specified Ψ. What are the biases and mean squared errors of these estimates when the true error distribution is Ψ_0? The cases $\Psi = \Psi_0$ and $\Psi \neq \Psi_0$, are both considered.

For any distribution function Ψ let $J(u) = -l'(\Psi^{-1}(u))$ where the function l is defined as in Section 4. Further, for distribution functions Ψ, Ψ_0 and G set

$$I_h(\Psi,\Psi_0,G) = \int J(u)J_0(u)\bar{G}(F^{-1}(u))du, \quad I_h(\Psi,G) = I_h(\Psi,\Psi,G)$$

$$b_h(\Psi,\Psi_0,G) = 1 - I_h(\Psi,\Psi_0,G) / I_h(\Psi_0,G)$$

where $F(t) = \Psi_0(h(t))$ and $B_n = X^T X I_h(\Psi,G)$, $B_{0n} = X^T X I_h(\Psi_0,G)$. In Section 6, we show

PROPOSITION 5.1. *Under the conditions of Corollary* 4.1, *for the model* A.1 *with error distribution* Ψ_0 , $(\tilde{\beta}-\beta)+b_h(\Psi,\Psi_0,G)\beta$ *has an asymptotic p-variate normal distribution* $N(0, \mathbf{B}_n^{-1})$.

In what follows, we shall assume that the covariates are univariate, i.e. $p = 1$. In this case, by Proposition 5.1, the bias and mean squared error of $\tilde{\beta}$ are approximately

$$\text{bias}(\tilde{\beta}) \cong b_h(\Psi,\Psi_0,G)\beta \tag{5.1}$$

$$\text{MSE}(\tilde{\beta}) \cong b_h^2(\Psi,\Psi_0,G)\beta^2+[I_h(\Psi,G)\Sigma x_i^2]^{-1}.$$

Note that these expressions are related to test efficiency. In fact, if the approximately standard normal test statistic $\tilde{\beta}(\hat{I}\,\Sigma x_i^2)^{\frac{1}{2}}$ is used to test $H_0 : \beta = 0$ versus $H_1: \beta>0$, the efficacy for the model with error distribution Ψ_0 is $[1-b_h(\Psi,\Psi_0,G)]^2 I_h(\Psi,G)\Sigma x_i^2]$. In the two - sample case, this reduces to the efficiency given by Harrington & Fleming (1982) and Cuzick (1985).

To compare the estimates $\hat{\beta}_B$ and $\tilde{\beta}$, in terms of bias and mean squared error, we shall examine some special choices of distributions Ψ, Ψ_0 and G.

5(a) *The normal likelihood sampler and the normal* LMPLE. Here $\hat{\beta}_B$ denotes the estimate derived from the normal likelihood sampler of Section 3 and $\tilde{\beta}$ is the normal LMPLE of Example 4.3. In our parametrization, when $\Psi = N(0,1)$, the estimates $\tilde{\beta}$ and $\hat{\beta}$ are estimates of the slope β in a transformed regression model with $Var(\varepsilon_i) = 1$. Thus, the error distribution Ψ_0 should have variance one. The formulas (5.1) have been evaluated for Ψ_0 = logistic and no censoring, in which case the $(\text{bias})^2$ in (5.1) is negligible compared to the variance for $|\beta|\leq 2$ and sample sizes such that $\Sigma x_i^2 \leq 100$ (see Doksum (1987)).

Our first table deals with the transformed linear regression model with $x_i = (i-13)/12$, $i = 1,\ldots,25$ and $h(Y_i) = \beta x_i+\varepsilon_i$, $\varepsilon_i \sim \Psi_0$. The data are either uncensored or the censoring is extreme value. BIAS and MSE refers to Monte Carlo results over $M = 500$ repetitions. The standard error (st. error) of the MSE over the $M = 500$ Monte Carlo trials is reported in parenthesis below the MSE. The AMPLE is based on a likelihood sampler with B = 100 replications.

Table 5.1 about here.

To check how much is lost in efficiency by not knowing h, we compare the MSE entries with the asymptotically optimal MSE when h is known (ideal MSE). We see that no efficiency is lost by using the AMPLE $\hat{\beta}$ for β in the range $(-1.5, 1.5)$. However, for $\beta = +2.5$, $\hat{\beta}$ is quite biased in favor of 1.8, and the mean squared error is 5 times larger than the optimal h - known mean squared error. The LMPLE $\tilde{\beta}$ is more biased than $\hat{\beta}$ and its performance is poor for $|\beta| > 1.5$.

This table also shows that the performance of the estimates is not much affected by the error distribution nor by the censoring here considered. Although $\hat{\beta}$ and $\tilde{\beta}$ of this subsection were designed for normal errors, they perform very well for logistic and extreme value errors for β in the range $(-1.5, 1.5)$.

Our next table deals with the two sample case with $x_i = -1/2$, $i = 1,\ldots,20$; $x_i = 1/2$, $i = 21,\ldots,40$. We consider the same error and censoring distributions as before.

Table 5.2 about here.

From Table 5.2 we learn that the AMPLE $\hat{\beta}$ has efficiency close to the optimal h - known efficiency for β in the interval $(-2,2)$. Thus $\hat{\beta}$ now has a larger ***range of adaptability*** than in the regression model of Table 5.1. Again, the LMPLE also performs well but it is generally less efficient than $\hat{\beta}$. Moreover, the normal AMPLE $\hat{\beta}$ and LMPLE $\tilde{\beta}$ are not sensitive to change in the error distributions.

5(b) *The logistic likelihood sampler and the logistic* LMPLE. Here we shall assume that the estimate $\hat{\beta}_B$ is derived from the logistic likelihood sampler and $\tilde{\beta}$ is the logistic LMPLE of Example 4.2.

In Table 5.3, the performance of the logistic AMPLE and LMPLE is investigated. The AMPLE $\hat{\beta}_B$ is based on $B = 50$ terms in the likelihood sampler, and the number of Monte Carlo trials in the simulation is $M = 300$.

Similarly to the normal AMPLE, the logistic AMPLE $\hat{\beta}_B$ performs very well for β values in the interval $(-1.5, 1.5)$, however for values $|\beta| > 2.0$, the performance of the logistic AMPLE is poorer than the normal AMPLE. The performance of the logistic LMPLE is remarkably close to that of the normal LMPLE for all values of β and for the three error distributions considered.

Again we note the robustness with respect to the error distribution Ψ of both the AMPLE and the LMPLE: The mean squared error of these estimates hardly changes as the error distribution changes from the logistic to the normal and extreme value distribution.

Table 5.3 about here.

We next give the local approximation to the bias and mean squared error of our estimates based on Ψ = logistic when the true error distribution Ψ_0 is extreme value distribution with the same variance as the logistic, i.e. $\overline{\Psi}_0(t) = \exp[-\exp(t/\sqrt{2})]$. We have

$$I_h(\Psi;\, \Psi_0, G) = \int \overline{F}\overline{G}dF, \;\; I_h(\Psi;\, G) = \int (\overline{F})^2\overline{G}dF.$$

We will assume that the censoring distribution satisfies $\overline{G} = \overline{F}^{\theta}$ for some $\theta > 0$. This case includes the interesting model where the failure times Y_i have the Weibull distribution $F(w_i y)$, where $\overline{F}(t) = \exp(y^{a/\sqrt{2}})$ and $w_i = \exp(-\beta x_i)$. For this h, the censoring distribution is Weibull with $\overline{G}(t) = \exp(y^{\theta a/\sqrt{2}})$.

Formula (5.1) yields

$$\mathrm{BIAS}(\tilde{\beta}) \cong [1-(\theta+3)/\sqrt{2}(\theta+2)]\beta, \quad \mathrm{MSE}(\tilde{\beta}) \cong (\mathrm{BIAS})^2 + (\sum_{i=1}^{n} x_i^2)^{-1}(\theta+3).$$

The case $\theta \to 0$ corresponds to no censoring and in this case the approximate bias is $(1-\sqrt{9/8})\beta$. As censoring increases so that $\theta \to \infty$, the bias tends to $[1-(1/\sqrt{2})]\beta$. The bias is negative for little censoring, positive for much censoring and zero at $\theta = \sqrt{2}-1 \cong .41$.

5(c) *Estimates generated by the extreme value distribution.* When Ψ is the extreme value distribution, the MPLE $\hat{\beta}$ is the usual Cox estimate and there is no need to use the likelihood sampler. The LMPLE $\tilde{\beta}$ defined in Example 4.1, has a simple and intuitive expression. It is the least squares estimate based on $(\tilde{Y}_i, x_i)$, $i = 1, \ldots, n$, where $\tilde{Y}_i = (n/k)[\hat{\Lambda}(Y_i) - d_i]$, $\hat{\Lambda}(t) = \int_0^t (V/Z)dN$ is the Nelson-Aalen estimate of the integrated hazard. Thus when $p = 1$, $\tilde{\beta}$ is just the slope of the least squares line to the standardized Nelson-Aalen plot $(\tilde{Y}_i, x_i)$, $i = 1, \ldots, n$.

Suppose the error distribution Ψ_0 is the logistic with the same variance as the extreme value distribution, i.e. $\Psi(t) = [1+\exp(-\sqrt{2}t)]^{-1}$, then $I_h(\Psi,\Psi_0,G) = \int \overline{F}\overline{G}dF$, $I_h(\Psi,G) = \int \overline{G}dF$. In the case $\overline{G} = \overline{F}^{\theta}$, (5.1) yields

$$\mathrm{BIAS}(\tilde{\beta}) = \left[1 - \frac{\sqrt{2}(\theta+1)}{(\theta+2)}\right]\beta, \quad \mathrm{MSE}(\tilde{\beta}) = (\mathrm{BIAS})^2 + (\sum_{i=1}^{n} x_i^2)^{-1}(\theta+1).$$

Thus when there is no censoring, there is a positive bias of $(1-\sqrt{2}/2)\beta$. The bias decreases with θ, crosses zero at $\theta = \sqrt{2}$ and tends to $(1-\sqrt{2})$ as $\theta \to \infty$.

5(d) *Summary.* To summarize the tables, we note that the mean squared error of the partial likelihood estimates are remarkably close to the ideal h-known mean squared error when the systematic variation is small and moderate compared to the residual variation. However, the bias of these estimates is large when the systematic variation dominates the residual variation. This bias effect is more severe for the local estimate LMPLE than for the AMPLE. Further, the variances of both estimates are fairly stable and decrease slowly as the systematic variation becomes large relative to the residual variation. These observations apply to both uncensored and censored data and the error distributions considered. The AMPLE performs overall better than LMPLE.

Figure 5.1, which is based on Table 5.2, presents a picture of the typical behavior of the approximate partial likelihood estimate. Note that the curves can be extended to the left by symmetry and that the variance can be found as the gap between the MSE and $(\text{BIAS})^2$ curves.

Figure 5.1 about here

6. Proofs.

PROOF of Theorem 4.1. Set $W_i = 2[(\lambda_{\mu_i}/\lambda_{\mu_{i0}})^{1/2} - 1]$ and $W(t) = \sum_{i=1}^{n} \int_0^t W_i dM_i$. Then

$$\log(dP_\beta/dP_{\beta_0}) = W(\infty) - \frac{1}{2}\sum \int_0^\infty W_i^2 Z_i d\Lambda_{\mu_{i0}} + r_n \tag{6.1}$$

where

$$r_n = \sum \int_0^\infty \{2\log(\lambda_{\mu_i}/\lambda_{\mu_{i0}})^{1/2} - W_i + \frac{1}{4}W_i^2\} dN_i - \frac{1}{4}\sum \int_0^\infty W_i^2 dM_i.$$

Using techniques similar to those in Gill (1980), it can be shown that r_n converges in P_{β_0} - probability to 0. Further, W and $a_n^T T$ are mean zero square integrable martingales with predictable variation processes given by

$$\langle W, W\rangle(t) = \sum \int_0^t W_i^2 Z_i d\Lambda_{\mu_{i0}}$$

$$\langle \mathbf{a}_n^T\mathbf{T},\mathbf{a}_n^T\mathbf{T}\rangle(t) = \sum(\mu_i-\mu_{i0})^2\int_0^t (l'_{\mu_{i0}})^2 Z_i d\Lambda_{\mu_{i0}}$$

$$\langle \mathbf{a}_n^T\mathbf{T},\mathbf{W}\rangle(t) = -\sum(\mu_i-\mu_{i0})\int_0^t W_i l'_{\mu_{i0}} Z_i d\Lambda_{\mu_{i0}}.$$

Using Remark 4.4, Theorem 1.1 in van Zuijlen (1978), and Theorem V.1.3 in Hâjek & Šidâk (1967), we can show (after some algebra) that

$$\langle \mathbf{a}_n^T\mathbf{T}-\mathrm{W},\mathbf{a}_n^T\mathbf{T}-\mathrm{W}\rangle(t) = o_{P_{\beta_0}}(1)$$

$$\langle \mathbf{a}_n^T\mathbf{T},\mathbf{a}_n^T\mathbf{T}\rangle(t) = \sum(\mu_i-\mu_{i0})^2\int_0^t (l'_{\mu_{i0}})^2 Z_i d\Lambda_{\mu_{i0}} + o_{P_{\beta_0}}(1)$$

and

$$\langle \mathbf{a}_n^T\mathbf{T}-\mathrm{W},\mathbf{a}_n^T\mathbf{T}-\mathrm{W}\rangle(\infty) = o_{P_{\beta_0}}(1),$$

$$\langle \mathbf{a}_n^T\mathbf{T},\mathbf{a}_n^T\mathbf{T}\rangle(\infty) = \mathbf{a}_n^T\mathbf{a}_n + o_{P_{\beta_0}}(1).$$

It follows that the second term in (4.1) is asymptotically equivalent to $-\mathbf{a}_n^T\mathbf{a}_n/2$.

Further, Remark 4.4 and A.4 imply that the statistic $\mathbf{a}_n^T\mathbf{T}$ satisfies the Lindeberg condition

$$\sum\frac{(\mu_i-\mu_{i0})^2}{\mathbf{a}_n^T\mathbf{a}_n}\int_0^\infty (l'_{\mu_{i0}})^2 I[(\mu_i-\mu_{i0})^2(l'_{\mu_{i0}})^2 > \varepsilon\mathbf{a}_n^T\mathbf{a}_n] Z_i d\Lambda_{\mu_{i0}} = o_{P_{\beta_0}}(1)$$

for all $\varepsilon>0$. Rebolledo's Central Limit Theorem completes the proof of (i). Part (ii) follows directly from the Hâjek-Le Cam asymptotic minimax theorem for LAN families (e.g. Le Cam (1969)).

PROOF of Theorem 4.2. We can write $\mathbf{b}_n^T S(t)$ as $\mathbf{b}_n^T S(t) = \sum\int_0^t U_i dN_i$ where $U_i = \nu_i l'(\Psi^{-1}(\hat{F})) - \sum \nu_k l'(\Psi^{-1}(\hat{F}))(V/Z)Z_k$ and $\nu_i = \mathbf{b}_n^T B^{-1/2} x_i$.

By Theorem 4.1, under P_0, $\log(dP_\beta/dP_0)$ has the same asymptotic distribution as $a_n^T T - \frac{1}{2} a_n^T a_n$ and $a_n^T T$ is asymptotically normal $N(0, a_n^T a_n)$. Under P_0, $b_n^T S$ is a mean zero square integrable martingale. Using Gill (1980, pp. 129-130) it can be verified that the predictable variation processes of $b_n^T S$ and $a_n^T T$ satisfy

$$\langle b_n^T S, b_n^T S\rangle(t) = \sum_{i=1}^{n} v_i^2 \int_0^t [l'(\Psi^{-1}(F))]^2 z_i d\Lambda_0$$

$$- \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{v_i v_j}{n} \int_0^t [l'(\Psi^{-1}(F))]^2 (v/z) z_i z_j d\Lambda_0 + o_{P_{\beta_0}}(1)$$

$$\langle b_n^T S, a_n^T T\rangle(t) = \sum_{i=1}^{n} \mu_i v_i \int_0^t [l'(\Psi^{-1}(F))]^2 z_i d\Lambda_0$$

$$- \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mu_i v_j}{n} \int_0^t [l'(\Psi^{-1}(F))]^2 (v/z) z_i z_j d\Lambda_0 + o_{P_{\beta_0}}(1)$$

where $\mu_i = a_n^T B_n^{-1/2} x_i$. Furthermore,

$$\langle b_n^T S, b_n^T S\rangle(\infty) = b_n^T (I - C_n) b_n + o_{P_0}(1)$$

$$\langle b_n^T S, a_n^T T\rangle(\infty) = b_n^T (I - C_n) a_n + o_{P_0}(1).$$

We shall verify now the Lindeberg condition

(6.2) $$\frac{1}{b_n^T (I - C_n) b_n} \int_0^t \sum U_i^2 Z_i I[U_i^2 > \varepsilon b_n^T (I - C_n) b_n] d\Lambda_0 \xrightarrow{P_0} 0$$

for all $\varepsilon > 0$ and t such that $v(t) = 1$. For each i, the set $[U_i^2 > \varepsilon b_n^T (I - C_n) b_n]$ is contained in $[(l'(\Psi^{-1}(\hat{F})))^2 > \varepsilon b_n^T (I - C_n) b_n / 4 \max v_j^2]$. Let $I(\varepsilon)$ be its indicator function. By the elementary inequality $|a-b|^2 \le 2a^2 + 2b^2$,

$$U_i^2 \le 2\{v_i^2 + (\Sigma v_j Z_j V/Z)^2\}(l'(\Psi^{-1}(\hat{F})))^2 \le 2\{v_i^2 + \Sigma v_j^2 Z_j V/Z\}(l'(\Psi^{-1}(\hat{F})))^2.$$

It follows that the integral in (4.2) is bounded above by

$$4\int \sum v_i^2 Z_i (l'(\Psi^{-1}(\hat{F})))^2 I(\varepsilon) d\Lambda_0 \le 4\int \sum (v_i^2 Z_i / z_i)(l'(\Psi^{-1}(\hat{F})))^2 I(\varepsilon) dF.$$

Remark 4.4. Assumption A.4 and Theorem 1.1 in van Zuijlen (1978) imply (4.2). The Rebolledo's Central Limit Theorem completes the proof of the asymptotic normality of $\mathbf{b}_n^T S(t)$, for t such that $v(t) = 1$. Assumption A.8 and a similar argument as in Gill (1980, p. 130) implies asymptotic normality of $\mathbf{b}_n^T S(\infty)$. The joint asymptotic normality of $\mathbf{a}_n^T \mathbf{T}$ and $\mathbf{b}_n^T \mathbf{S}$ follows from the Cramér-Wold device as in Hájek & Šidák (1967, pp. 216-218) and Gill (1980, pp. 131-133).

PROOF of Corollary 4.1. Let $\boldsymbol{\beta} = \mathbf{B}_n^{-1/2}\mathbf{a}_n \in \boldsymbol{\Omega}_n$ and let $\mathbf{b}_n$ be a bounded sequence in R^p. By Theorem 4.2, $(\log(dP_\beta/dP_0), \mathbf{b}_n^T S(\infty))$ have an asymptotic bivariate normal distribution with mean $(-\mathbf{a}_n^T\mathbf{a}_n/2, 0)$ and covariance

$$\boldsymbol{\Sigma}_{ab} = \begin{bmatrix} \mathbf{a}_n^T\mathbf{a}_n & \mathbf{b}_n^T\mathbf{a}_n \\ \mathbf{b}_n^T\mathbf{a}_n & \mathbf{b}_n^T\mathbf{b}_n \end{bmatrix}$$

By Le Cam's Third Lemma (Hájek & Šidák (1967), p. 208), under P_β, $\mathbf{b}_n^T S_n(\infty)$ has an asymptotic normal distribution $N(\mathbf{b}_n^T\mathbf{a}_n, \mathbf{b}_n^T\mathbf{b}_n)$. Therefore, by the Cramér-Wold device, under P_β, $S(\infty) - \mathbf{a}_n$ has a p - variate standard normal distribution. To complete the proof it is enough to note that $\mathbf{a}_n = \mathbf{B}_n^{1/2}\boldsymbol{\beta}_n$, $S(\infty) = \mathbf{B}_n^{1/2}\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{B}}_n$ is a consistent estimate of $\mathbf{B}_n$.

PROOF of Proposition 5.1. Let $\boldsymbol{\beta} = \mathbf{B}_{0n}^{-1/2}\mathbf{a}_n \in \boldsymbol{\Omega}_n$ and let S(t) be given by (4.2). Proceeding as in the proof of Theorem 4.1, it can be verified that for any bounded sequence $\mathbf{b}_n$ in R^p, under P_0, $(\log(dP_\beta/dP_0), \mathbf{b}_n^T S)$ are jointly asymptotically normal with mean $(-\frac{1}{2}\mathbf{a}_n^T\mathbf{a}_n, 0)$ and covariance

$$\boldsymbol{\Sigma}_{ab} = \begin{bmatrix} \mathbf{a}_n^T\mathbf{a}_n & \mathbf{b}_n^T\mathbf{a}_n\rho \\ \mathbf{b}_n^T\mathbf{a}_n\rho & \mathbf{b}_n^T\mathbf{b}_n \end{bmatrix}$$

where

$$\rho = I_h(\Psi,\Psi_0,G)I_n(\Psi,G)^{-1/2}I_h(\Psi_0,G)^{-1/2}.$$

By Le Cam's Third Lemma (Hájek & Šidák (1967), p. 208) and the Cramér-Wold device, $S(\infty) - a_n\rho$ has a p-variate standard normal distribution. The conclusion follows from the relations $S(\infty) = B_n^{1/2}\tilde{\beta}$ and $a_n\rho = B_n^{-1/2}\beta - B_n^{-1/2}b_h(\Psi,\Psi_0,G)\beta$.

Acknowledgements. We are grateful to P. G. Neville for programming the likelihood sampler and the local estimates. This research was supported in part by the National Science Foundation Grants DMS-83-01716 and DMS-85-41787, and the National Institute of General Medical Sciences Grant SSS-Y1RO1 GM35416.

TABLE 5.1. Monte Carlo results for the normal approximate maximum partial likelihood estimate $\hat{\beta}$ and the local maximum partial likelihood estimate $\tilde{\beta}$ in the transformed linear regression model. $\mathrm{Var}(\varepsilon_i) = 1$, $\Sigma x_i^2 = 9$. The data are uncensored in (a), (b) and (c). In (d), (e) and (f) the censoring is extreme value with standard deviation equal to 1 and shift parameter equal to 1. The number of trials in the resampling scheme is B = 100 and the number of Monte Carlo trials is M = 500.

	β	0	0.5	1.0	1.5	2.0	2.5	3.0
(a)	BIAS($\hat{\beta}$)	-.029	-.005	-.016	-.136	-.390	-.737	-1.14
normal	BIAS($\tilde{\beta}$)	-.027	-.055	-.172	-.414	-.752	-1.15	-1.58
error,	9 MSE($\hat{\beta}$)	1.24	1.22	1.02	.904	1.79	5.18	11.9
ideal	(st. error)	(.079)	(.072)	(.059)	(.059)	(.078)	(.106)	(.139)
9MSE=1	9 MSE($\tilde{\beta}$)	1.04	.896	.795	1.83	5.24	11.9	22.6
	(st. error)	(.058)	(.053)	(.055)	(.069)	(.079)	(.081)	(.084)
(b)	BIAS($\hat{\beta}$)	-.286	.188	.020	-.111	-.378	-.736	-1.13
logistic	BIAS($\tilde{\beta}$)	-.272	-.350	-.148	-.399	-.743	-1.15	-1.58
error,	9MSE($\hat{\beta}$)	1.24	1.26	1.08	.893	1.76	5.16	11.9
ideal	(st. error)	(.080)	(.076)	(.060)	(.061)	(.086)	(.108)	(.140)
9MSE=.912	9MSE($\tilde{\beta}$)	1.05	.885	.748	1.74	5.16	11.9	22.7
	(st. error)	(.058)	(.052)	(.054)	(.071)	(.088)	(.091)	(.093)
(c)	BIAS($\hat{\beta}$)	-.014	.078	.064	-.074	-.353	-.714	-1.14
extreme	BIAS($\tilde{\beta}$)	-.013	.012	-.119	-.381	-.733	-1.14	-1.58
value	9MSE($\hat{\beta}$)	1.17	1.26	1.06	.823	1.62	4.94	11.9
error,	(st. error)	(.076)	(.084)	(.063)	(.059)	(.085)	(.119)	(.151)
ideal	9MSE($\tilde{\beta}$)	1.01	.814	.628	1.62	5.01	11.8	22.6
9MSE=.608	(st. error)	(.060)	(.052)	(.050)	(.071)	(.088)	(.099)	(.103)
(d)	BIAS($\hat{\beta}$)	-.012	-.003	-.047	-.150	-.368	-.736	-1.14
normal	BIAS($\tilde{\beta}$)	-.019	-.047	-.159	-.378	-.696	-1.15	-1.58
error	9MSE($\hat{\beta}$)	1.63	1.73	1.59	1.36	2.09	5.69	13.1
	(st. error)	(.112)	(.111)	(.098)	(.098)	(.112)	(.118)	(.155)
	9MSE($\tilde{\beta}$)	1.11	.996	.852	1.63	4.56	13.2	25.0
	(st. error)	(.067)	(.062)	(.060)	(.073)	(.088)	(.100)	(.102)
(e)	BIAS($\hat{\beta}$)	-.005	.057	.052	-.050	-.290	-.632	-1.01
logistic	BIAS($\tilde{\beta}$)	-.003	.002	-.098	-.334	-.669	-1.06	-1.48
error	9MSE($\hat{\beta}$)	1.49	1.47	1.43	1.07	1.55	4.23	9.8
	(st. error)	(.099)	(.092)	(.089)	(.071)	(.088)	(.136)	(.204)
	9 MSE($\tilde{\beta}$)	1.02	.838	.616	1.30	4.21	10.2	19.8
	(st. error)	(.057)	(.051)	(.044)	(.059)	(.082)	(.096)	(.105)

TABLE 5.1 (continued)

	β	0	0.5	1.0	1.5	2.0	2.5	3.0
(f)	BIAS($\hat{\beta}$)	.018	.178	.209	.080	-.198	-.557	-.982
extreme	BIAS($\tilde{\beta}$)	.008	.057	-.041	-.285	-.624	-1.02	-1.45
value	9MSE($\hat{\beta}$)	1.91	2.05	1.81	1.04	1.04	3.47	9.33
error	(st. error)	(.141)	(.130)	(.115)	(.069)	(.061)	(.120)	(.204)
	9MSE($\tilde{\beta}$)	1.15	.954	.594	1.08	3.69	9.47	19.0
	(st. error)	(.065)	(.053)	(.046)	(.064)	(.080)	(.098)	(.111)

TABLE 5.2. Monte Carlo results for the normal approximate maximum partial likelihood estimate $\hat{\beta}$ and the local maximum partial likelihood estimate $\tilde{\beta}$ in the transformed two-sampled shift model. $\mathrm{Var}(\varepsilon_i) = 1$, $\Sigma x_i^2 = 10$, $n = 40$. The data are uncensored in (a), (b) and (c). In (d), (e) and (f) the censoring is extreme value with standard deviation equal to and shift parameter equal to 1. The number of trials in the resampling scheme is B = 100 and the number of Monte Carlo trials is M = 500.

	β	0	0.5	1.0	1.5	2.0	2.5	3.0
(a)	BIAS($\hat{\beta}$)	.005	.014	.012	-.078	-.290	-.626	-1.03
normal	BIAS($\tilde{\beta}$)	.005	-.020	-.109	-.306	-.611	-1.00	-1.44
error,	10MSE($\hat{\beta}$)	.991	.991	.980	.740	1.24	4.13	10.8
ideal	(st. error)	(.060)	(.063)	(.064)	(.048)	(.060)	(.084)	(.111)
10 MSE = 1	10MSE($\tilde{\beta}$)	.912	.792	.668	1.24	3.87	9.96	20.6
	(st. error)	(.053)	(.050)	(.047)	(.057)	(.069)	(.067)	(.054)
(b)	BIAS($\hat{\beta}$)	.005	.033	.037	-.062	-.292	-.647	-1.05
logistic	BIAS($\tilde{\beta}$)	.005	-.001	-.092	-.297	-.613	-1.01	-1.45
error,	10 MSE($\hat{\beta}$)	1.15	1.17	1.18	.840	1.34	4.47	11.3
ideal	(st. error)	(.072)	(.072)	(.067)	(.055)	(.073)	(.102)	(.131)
10MSE = .912	10 MSE($\tilde{\beta}$)	1.04	.901	.724	1.27	3.96	10.2	21.0
	(st. error)	(.062)	(.053)	(.048)	(.065)	(.084)	(.092)	(.088)
(c)	BIAS($\hat{\beta}$)	-.005	.076	.096	-.027	-.279	-.643	-1.05
extreme	BIAS($\tilde{\beta}$)	-.005	.033	-.051	-.276	-.606	-1.00	-1.46
value	10 MSE($\hat{\beta}$)	.992	1.12	1.14	.752	1.29	4.44	11.3
error,	(st. error)	(.060)	(.066)	(.067)	(.046)	(.073)	(.109)	(.149)
ideal	10 MSE($\tilde{\beta}$)	.912	.822	.590	1.11	3.88	10.3	21.2
10MSE = .608	(st. error)	(.053)	(.045)	(0.36)	(.058)	(.085)	(.100)	(.103)
(d)	BIAS($\hat{\beta}$)	-.001	.008	-.005	-.088	-.280	-.585	-1.01
normal	BIAS($\tilde{\beta}$)	.012	-.002	-.077	-.252	-.523	-.875	-1.28
error	10 MSE($\hat{\beta}$)	1.55	1.53	1.36	1.13	1.44	3.93	10.5
	(st. error)	(.107)	(.107)	(.086)	(.075)	(.080)	(.115)	(.170)
	10 MSE($\tilde{\beta}$)	1.16	.986	.739	1.05	2.95	7.78	16.4
	(st. error)	(.073)	(.060)	(.050)	(.061)	(.077)	(.089)	(.095)
(e)	BIAS($\hat{\beta}$)	.021	.061	.077	-.014	-.225	-.578	-1.01
logistic	BIAS($\tilde{\beta}$)	.027	.033	-.358	-.222	-.511	-.873	-1.28
error	10 MSE($\hat{\beta}$)	1.41	1.43	1.44	1.06	1.17	3.95	10.5
	(st. error)	(.098)	(.090)	(.088)	(.072)	(.072)	(.125)	(.178)
	10 MSE($\tilde{\beta}$)	1.05	.965	.683	.894	2.84	7.75	16.5
	(st. error)	(.065)	(.059)	(.048)	(.056)	(.076)	(.092)	(.102)

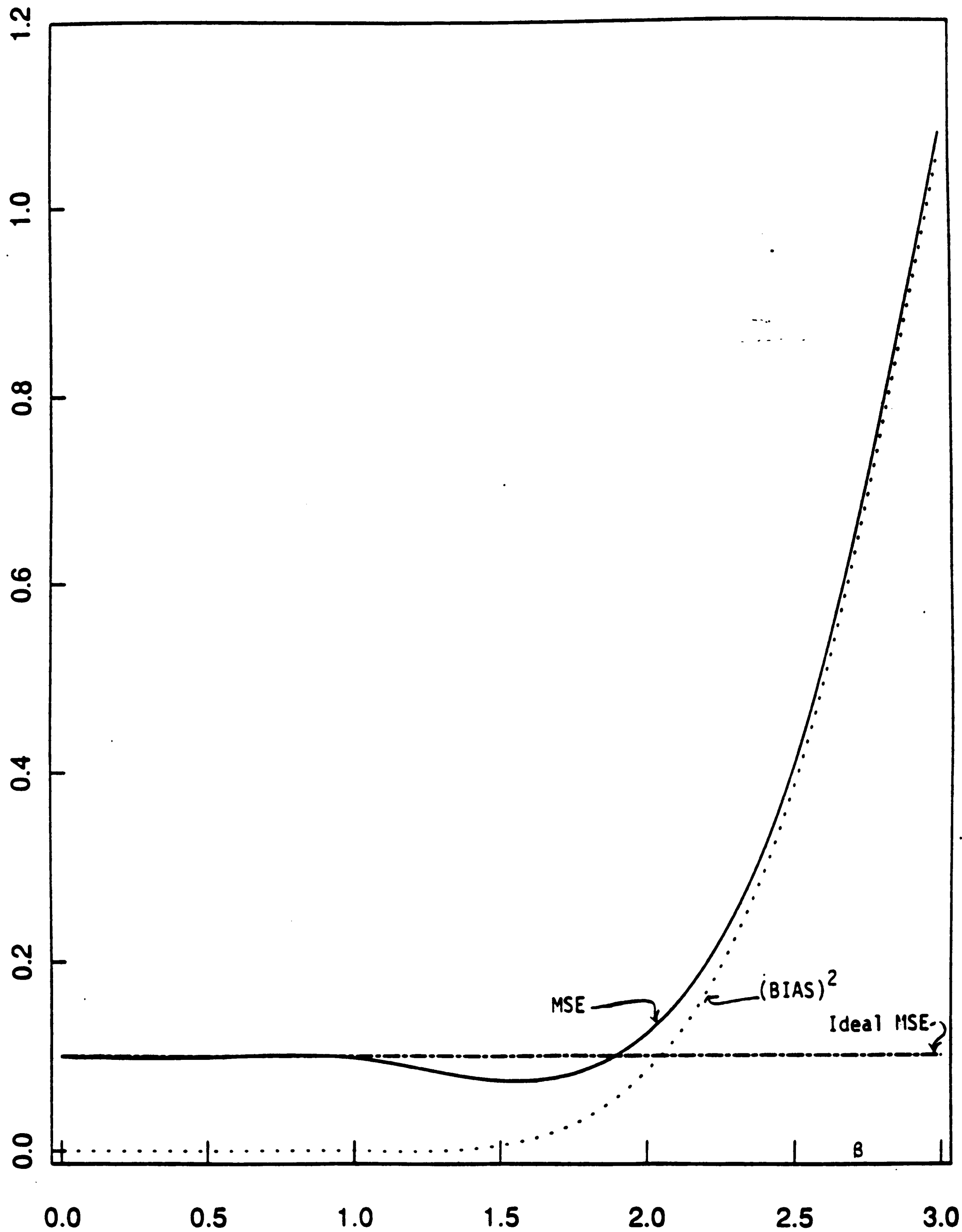
TABLE 5.2 (continued)

	β	0	0.5	1.0	1.5	2.0	2.5	3.0
(f)	BIAS($\hat{\beta}$)	.002	.178	.240	.141	-.132	-.561	-1.03
extreme	BIAS($\tilde{\beta}$)	-.008	.064	.018	-.164	-.465	-.829	-1.24
value	10MSE($\hat{\beta}$)	1.68	1.92	1.85	1.09	.762	3.62	11.1
error	(st. error)	(.106)	(.111)	(.112)	(.077)	(.051)	(.106)	(.196)
	10MSE($\tilde{\beta}$)	1.11	.951	.612	.634	2.36	7.00	15.5
	(st. error)	(.065)	(.057)	(.039)	(.043)	(.065)	(.088)	(.104)

TABLE 5.3. Monte Carlo results for the logistic approximate maximum partial likelihood estimate $\hat{\beta}$ and the local maximum partial likelihood estimate $\tilde{\beta}$ in the transformed two sample shift model. $Var(\varepsilon_i) = 1$, $\Sigma x_i^2 = 10$, $n = 40$. The data are uncensored in all cases. The number of trials in the resampling scheme is $B = 50$ and the number of Monte Carlo trials is $M = 300$.

	β / σ_E	0	0.5	1.0	1.5	2.0	2.5	3.0
(a)	BIAS ($\hat{\beta}$)	.012	.020	.015	-.121	-.405	-.799	-1.22
logistic	BIAS ($\tilde{\beta}$)	.012	-.001	-.094	-.297	-.606	-.992	-1.43
error,	10 MSE ($\hat{\beta}$)	.903	.943	.928	.777	2.03	6.65	15.2
ideal	(st. error)	(.149)	(.153)	(.141)	(.137)	(.194)	(.268)	(.348)
10 MSE = .912	10 MSE ($\tilde{\beta}$)	.847	.785	.671	1.24	3.86	9.93	20.5
	(st. error)	(.135)	(.127)	(.119)	(.155)	(.205)	(.228)	(.219)
(b)	BIAS ($\hat{\beta}$)	.012	-.013	-.043	-.164	-.421	-.799	-1.22
normal	BIAS ($\tilde{\beta}$)	.012	-.036	-.136	-.329	-.673	-.992	-1.42
error,	10 MSE ($\hat{\beta}$)	.903	.930	.931	.925	2.16	6.62	15.0
ideal	(st. error)	(.149)	(.152)	(.145)	(.151)	(.196)	(.255)	(.325)
10 MSE = 1	10 MSE ($\tilde{\beta}$)	.847	.803	.774	1.44	4.02	9.88	20.2
	(st. error)	(.135)	(.131)	(.132)	(.165)	(.204)	(.204)	(.168)
(c)	BIAS ($\hat{\beta}$)	-.012	.023	.024	-.110	-.400	-.799	-1.23
extreme	BIAS ($\tilde{\beta}$)	-.012	-.000	-.088	-.287	-.601	-.992	-1.43
value	10 MSE ($\hat{\beta}$)	.907	.906	.937	.710	1.99	6.66	15.3
error,	(st. error)	(.150)	(.130)	(.128)	(.117)	(.179)	(.258)	(.349)
ideal	10 MSE ($\tilde{\beta}$)	.847	.756	.636	1.18	3.84	9.96	20.8
10 MSE = .608	(st. error)	(.135)	(.115)	(.100)	(.145)	(.201)	(.233)	(.230)

Figure 5.1. Monte Carlo results for the normal approximate maximum partial likehood estimate $\hat{\beta}$ in the two-sample shift model.



REFERENCES

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701-726.

ANDERSEN, P. K. & BORGAN, O. (1985). Counting process models for life history data: a review. *Scand. J. Statist.* **12**, 97-158.

ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100-1120.

ANDERSEN, P. K., BORGAN, O., GILL, R.D., KEIDING, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data. *Internat. Stat. Review.* **50** 219-258.

ANSCOMBE, F. J. & TUKEY, J. W. (1954). The criticism of transformation. Unpublished manuscript.

BARNDORFF-NIELSEN, O. & BLAESILD, P. (1980). Global maxima and likelihood in linear models. Aarhus Research Report.

BEGUN, J. M. (1981). A class of rank estimates of relative risk. Tech. Report, University of North Carolina, Chapel Hill.

BEGUN, J. M. & WELLNER, J. A. (1983). Asymptotic efficiency of relative risk estimates. In *Contributions to statistics: essays in honour of N. L. Johnson* (P. K. Sen, Ed.). North Holland Pub. Co., 47-62.

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

BELL, C. B. & DOKSUM, K.A. (1965). Some new distribution-free statistics. *Ann. Math. Statist.* **36**, 203-214.

BENNETT, S. (1983). Log-logistic regression models for survival data. *Appl. Statist.* **32** 165-171.

BERAN, R. & MILLAR, P. W. (1985). (Preprint). Stochastic estimation and testing.

BERKSON, J. (1944). Application of the logistic function to bio-assay. *J. Amer. Statist. Assoc.* **39**, 357-365.

BICKEL, P.J. (1984). Discussion of "The analysis of transformed data" by Hinkley and Runger. *J. Amer. Statist. Assoc.* **76**, 293-311.

BICKEL, P. J. (1986). (Preprint). Efficient testing in a class of transformation models.

BICKEL, P. J. & DOKSUM, K. A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.* **76** 293-311.

BLISS, C.I. (1935). The calculation of the dosage mortality curve. *Ann. Applied Biology* **22**, 134.

BOX, G. E. P. & COX, D. R. (1964). An analysis of transformation. *J. Roy. Statist. Soc. Ser. B* **26** 211-252.

BOX, G. E. P. & TIAO, G. E. (1973). *Bayesian inference in statistical analysis.* Addison-Wesley, Menlo Park, California.

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

BUCKLEY, J. & JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 89-99.

BURR, I. W. (1942). Cumulative frequency distributions. *Ann. Math. Statist.* **13** 215-232.

CARROLL, R. J. (1982). Tests for regression parameters in power transformation models. *Scand. J. Statist.* **9** 217-222.

CLAYTON, D. & CUZICK, J. (1986). The semiparametric Pareto model for regression analysis of survival times. *Proc. ISI*, Amsterdam.

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B.* **34** 187-202.

COX, D.R. (1975). Partial likelihood. *Biometrika* **62** 269-276.

CUZICK, J. (1985). Asymptotic properties of censored linear rank tests. *Ann. Statist.* **13** 133-141.

DIGGLE, P. J. & GRATTON, R. J. (1984) Monte Carlo methods of inference for implicit statistical models. *J. R. Statist. Soc., Ser. B* **46** 193-227.

DOKSUM, K. A. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.* **15** (March).

DOKSUM, K. A. & WONG, C-W. (1983). Statistical tests based on transformed data. *J. Amer. Statist. Assoc.* **78** 411-417.

FERGUSON, T. S. (1967). *Mathematical statistics.* Academic Press, New York.

FISHER, R. A. (1946). *Statistical methods for research workers, 10th ed.* Oliver and Boyd, Edinburgh-London.

GILL, R. D. (1980). *Censoring and stochastic integrals.* Mathematical Centre Tracts 124, Amsterdam.

HAJEK, J. A. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14** 401-414.

HAJEK, J. A. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175-194. University of California Press, Berkeley.

HAJEK, J. A. & ŠIDAK, Z. (1967). *Theory of rank tests.* Academic Press, New York.

HARRINGTON, D. P. & FLEMING, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69** 533-546.

HINKLEY, D. & RUNGER, G. (1984). The analysis of transformed data. *J. Amer. Statist. Assoc.* **79**, 302-309.

HOEFFDING, W. (1951). "Optimum" nonparametric tests. *Proc. Second Berkeley Symposium Math. Statist. Probab* 83-92. University of California Press, Berkeley.

JOHNSON, R. A. (1982). Transformation of survival data. In *Survival analysis* (Crowley, J. J. and Johnson, R. A., Eds.). IMS Lecture Notes, Monograph Series **2** 118-136.

KALBFLEISCH, J. D. (1978). Likelihood methods and nonparametric tests. *J. Amer. Statist. Assoc.* **73** 167-170.

KALBFLEISCH, J. D. & PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression model. *Biometrika* **60** 267-278.

KALBFLEISCH, J. D. & PRENTICE, R. S. (1980). *The statistical analysis of failure time data.* Wiley, New York.

KAPLAN, E. L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457-481.

KOUL, H., SUSARLA, V. & VAN RYZIN, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9** 1276-1288.

KOUL, H. L. & WANG, W. H. (1984). Local asymptotic normality of randomly censored linear regression model. *Statistics and Decisions* **1** 17-30.

KRUSKAL, J. B. (1965). Analysis of factorial experiments by estimating monotone transformation of the data. *J. Roy. Statist. Soc. B* **27** 251-263.

LE CAM, L. M. (1969). *Theorie asymptotique de la decision statistique*. Les Presses de l'Universitie de Montreal.

LE CAM, L. M. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41** 802-828.

LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23-43.

MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63** 449-464.

MILLER, R. & HALPERN, J. (1982). Regression with censored data. *Biometrika* **60** 521-531.

PETTITT, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *J. Roy. Statist. Soc. Ser. B.* **44** 234-243.

PETTITT, A. N. (1983). Approximate methods using ranks for regression with censored data. *Biometrika* **70** 121-132.

PETTITT, A. N. (1984). Proportional odds model for survival data and estimates using ranks. *J. Roy. Statist. Soc. Ser. C* **33** 169-175.

PIKE, M. C. (1966). A method of analysis of certain class of experiments in carcinogenesis. *Biometrics* **22** 142-161.

PRENTICE, R. S. (1978). Linear rank tests with censored data. *Biometrika* **65** 167-179.

PRENTICE, R. S. & MAREK, P. (1979). A qualitative discrepancy between censored data rank tests. *Biometrics* **35** 861-867.

RAO, U. V. R., SAVAGE, I. R. & SOBEL, M. (1960). Contributions to the theory of rank order statistics: two sample censored case. *Ann. Math. Statist.* **31** 415-426.

RUBIN, D.B. (1984). Discussion of "The analysis of transformed data" by Hinkley and Runger. *J. Amer. Statist. Assoc.* **79**, 309-312.

SAVAGE, I. R. (1956). Contributions to the theory of rank order statistics: the two sample case. *Ann. Math. Statist.* **27** 590-615.

SAVAGE, I. R. (1957). Contributions to the theory of rank order statistics: the trend case. *Ann. Math. Statist.* **28** 968-977.

SOLOMON, P. J. (1984). Effect of misspecification of regression models in the analysis of survival data. *Biometrika* **71** 291-298.

VAN ZUIJLEN, M. C. A. (1978). Properties of the empirical distribution function for independent nonidentically distributed random variables. *Ann. Probab.* **6** 250-266.

WONG, W. H. (1986). Theory of partial likelihood. *Ann. Statist.* **14** 88-123.