

Semiparametric Regression Analysis for Doubly Censored Data

BY T. CAI

*Department of Biostatistics, Harvard University,
Boston, MA 02115, U.S.A.*

tcai@hsph.harvard.edu

and S. CHENG

*Department of Epidemiology and Biostatistics, University of California,
San Francisco, CA 94143-0560, U.S.A.*

scheng@biostat.ucsf.edu

SUMMARY

We analyse doubly censored data using semiparametric transformation models. We provide inference procedures for the regression parameters and derive the asymptotic distributions of the proposed estimators. Procedures for model checking and model selection are also discussed. We illustrate our approach with a viral-load dataset from a recent AIDS clinical trial.

Some key words : Double censoring; Model checking; Semiparametric transformation model.

1 INTRODUCTION

One useful alternative to the popular proportional hazards model (Cox, 1972) for right censored survival data is a class of semiparametric transformation models, which includes the proportional hazards and proportional odds models as special cases. Inference procedures under this class of models for right censored data have been proposed by Clayton & Cuzick (1985), Cheng et al. (1995, 1997), Cai et al. (2000), Xu & Harrington (2001) and Chen et al. (2002).

We consider semiparametric transformation models allowing the response variable T to be subject to double censoring, that is to both left and right censoring. If T represents the failure time, then left censoring occurs when a subject failed prior to the entry of the study, L , and right censoring occurs when a subject has not yet failed by the end of the study, U . Double censoring may also arise when T represents an outcome variable that can only be accurately measured within a certain range, $[L, U]$. For example, a recent AIDS clinical trial was designed to compare the virological responses to three treatments for HIV-infected children. One major endpoint of the study was the plasma HIV-1 RNA level obtained by the NucliSens assay. As a result of the limits of quantification of the assay, the observed RNA copies per millilitre of plasma are highly unreliable if above 750,000 or below 400. Note that we observe $\{\min(U, \max(L, T)), I(T < L), I(T > U)\}$, where $I(\cdot)$ is the indicator function. Double censoring, as we define it, differs from interval censoring, in which case only $\{I(T < L), I(T > U)\}$ is observed, and from doubly interval censored data (Sun, 1995) in which the occurrences of both the originating event and the terminating event are either right or interval censored (Kim et al., 1993; Geskus, 2001). Our censoring variables L and U can be fixed constants, or group-specific or subject-specific random variables.

Previous work on our problem includes single sample estimation, two sample testing and regression analysis. For single sample estimation, Turnbull (1974)'s estimator was studied asymptotically by Tsai & Crowley (1985), Chang (1990) and Gu & Zhang (1993). Two-sample tests with doubly censored data were treated by Gehan (1965), Mantel (1967) and Huges (2000). For the linear regression model, Zhang & Li (1996) and Ren & Gu (1997) proposed M-estimators. These regression methods, requiring that the censoring variables be observed only when T is censored, are somewhat complicated. In this article, we consider an alternative semiparametric model and present a simpler approach to the regression analysis of doubly censored data. To this end, we require the censoring variables to be always observed, as in the aforementioned paediatric AIDS study.

We propose a semiparametric approach to the estimation of both the covariate effect and the underlying distribution function of T at a given covariate level. We consider the transformation model

$$S_Z(t) = \text{pr}(T \geq t \mid Z) = g\{h(t) + \beta^\top Z\}, \quad (1.1)$$

where the continuous, strictly decreasing link function $g(\cdot)$ is given or specified up to a finite dimensional parameter, $h(\cdot)$ is a completely unspecified strictly increasing function, and β is a $p \times 1$ vector of unknown regression coefficients. Note that when $g(\cdot) = \exp\{-\exp(\cdot)\}$ (1.1) gives the Cox proportional hazards model (Cox, 1972), and when $g(\cdot) = 1/\{1 + \exp(\cdot)\}$ it corresponds to the proportional odds model (Bennett, 1983; Clayton & Cuzick, 1985; Murphy et al, 1997; Yang & Prentice, 1999). Note also that the linear model considered by Zhang & Li (1996) and Ren & Gu (1997) is equivalent to (1.1) if g is unspecified and $h(t)$ is known to be t .

We first consider the case in which $g(\cdot)$ is completely specified. Inference procedures for the regression parameters are derived in §2. Pointwise and simultaneous confidence intervals for the distribution function of T at a given covariate level are provided in §3. In §4, we present a graphical method for model checking and a procedure for selecting an optimal link function, $g(\cdot)$, from a class indexed by a single parameter λ . We also provide inference procedures for the estimated parameters accounting for the variation from estimating the additional parameter λ in the link function. Simulation studies in §5 suggest that the proposed methods work well in finite samples. In §6 we apply the methods to the aforementioned AIDS dataset. We close in §7 with some remarks.

2 INFERENCE PROCEDURES FOR REGRESSION PARAMETERS

Let T be the response variable and Z be the corresponding $p \times 1$ vector of bounded covariates. Assume that T and Z are related through the semiparametric transformation model in (1.1). Let L and U denote the left and right censoring variables, respectively. We assume that L and U are always observed. For T , one can only observe the vector $\{X, I(T < L), I(T > U)\}$, where $X = \min\{U, \max(T, L)\}$. Given Z , the censoring variables L and U are assumed to be independent of T . Let $\{(T_i, L_i, U_i, Z_i) : i = 1, \dots, n\}$ be n independent and identically distributed copies of (T, L, U, Z) . Note that $\text{pr}(X_i \geq t > L_i \mid Z_i) = \text{pr}(U_i \geq t > L_i, T_i \geq t \mid Z_i) = \text{pr}(U_i \geq t > L_i \mid Z_i)\text{pr}(T_i \geq t \mid Z_i)$.

Then, under model (1.1),

$$\text{pr}(X_i \geq t > L_i \mid Z_i) = \text{pr}(U_i \geq t > L_i \mid Z_i)g \{h_0(t) + \beta_0^\top Z_i\},$$

where $h_0(\cdot)$ and β_0 are the true values of $h(\cdot)$ and β , respectively. This motivates the following estimating equation for $h_0(t)$ at any given β :

$$\sum_{i=1}^n \left[I(X_i \geq t > L_i) - I(U_i \geq t > L_i)g \{h(t) + \beta^\top Z_i\} \right] = 0, \quad \tau_a \leq t \leq \tau_b, \quad (2.1)$$

where τ_a and τ_b are pre-specified constants such that both $\text{pr}(X_1 < \tau_a)$ and $\text{pr}(X_1 > \tau_b)$ are positive. Let $\hat{h}(t; \beta)$ denote the solution to equation (2.1). Note that $\hat{h}(t; \beta)$ is a step function in t that rises at the distinct jump points of $\{I(X_i \geq t) : i = 1, \dots, n\}$ and $\{I(U_i \geq t > L_i) : i = 1, \dots, n\}$. By mimicking the generalised estimating equation method, we consider the following class of estimating equations for β_0 :

$$\sum_{i=1}^n \int_{\tau_a}^{\tau_b} Z_i \left[I(X_i \geq t > L_i) - I(U_i \geq t > L_i)g \left\{ \hat{h}(t; \beta) + \beta^\top Z_i \right\} \right] d\hat{v}(t) = 0, \quad (2.2)$$

where $\hat{v}(\cdot)$ is a known increasing, but possibly data-dependent, weight function that converges to a deterministic function $v(\cdot)$ uniformly in $t \in [\tau_a, \tau_b]$. For example, we choose the weight function \hat{v} to be the counting process of $\{X_i\}$ in all our analyses. Let $\hat{\beta}$ denote the root of (2.2) and let $\hat{h}(t) = \hat{h}(t; \hat{\beta})$. In Appendix 1, we show that, under mild conditions, $\hat{\beta}$ and $\hat{h}(t)$ are unique for large n and are consistent.

To obtain interval estimators for specific components of β_0 , in Appendix 2 we show that

$$n^{\frac{1}{2}}(\hat{\beta} - \beta_0) \simeq n^{-\frac{1}{2}}A^{-1} \sum_{i=1}^n V_i,$$

where A is defined by (A1.6) in Appendix 1,

$$V_i = \int_{\tau_a}^{\tau_b} \{Z_i - z(t; \beta_0)\} \left[I(X_i \geq t > L_i) - I(U_i \geq t > L_i)g \{h_0(t) + \beta_0^\top Z_i\} \right] dv(t),$$

and $z(t; \beta)$ is the limit of $\bar{Z}(t; \beta)$ defined by (A1.4) in Appendix 1. It follows from the standard central limit theorem that the distribution of $n^{\frac{1}{2}}(\hat{\beta} - \beta_0)$ can be approximated by a zero-mean normal random vector with covariance matrix

$$\Sigma = n^{-1}A^{-1} \left(\sum_{i=1}^n V_i V_i^\top \right) A^{-1}.$$

A reasonable estimator for Σ is

$$n^{-1} \widehat{A}^{-1}(\widehat{\beta}) \left(\sum_{i=1}^n \widehat{V}_i \widehat{V}_i^\top \right) \widehat{A}^{-1}(\widehat{\beta}),$$

where \widehat{V}_i is obtained by replacing all the theoretical quantities in V_i with their empirical counterparts and $\widehat{A}(\beta)$ is defined by (A1.5) in Appendix 1.

3 PREDICTION OF SURVIVAL PROBABILITIES

The regression model (1.1) for the response variable T can also be used to predict the survivorship function of T for future patients with specific covariates, for example, in the AIDS study, for a patient with RNA at a given level z_0 . A natural estimator for the survival probability $S_{z_0}(t)$ is $\widehat{S}_{z_0}(t) = g\{\widehat{h}(t) + \widehat{\beta}^\top z_0\}$. The consistency of $\widehat{\beta}$ and $\widehat{h}(t)$ ensures the consistency of $\widehat{S}_{z_0}(t)$. To obtain the distribution of $\widehat{S}_{z_0}(t)$, in Appendix 2 we show that the process

$$\omega_{z_0}(t) = n^{\frac{1}{2}} \left[g^{-1}\{\widehat{S}_{z_0}(t)\} - g^{-1}\{S_{z_0}(t)\} \right]$$

is asymptotically equivalent to $\check{\omega}_{z_0}(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \check{\omega}_{z_0i}(t)$, where

$$\check{\omega}_{z_0i}(t) = \{z_0 - z(t; \beta_0)\}^\top A^{-1} V_i + a(t)^{-1} \left[I(X_i \geq t > L_i) - I(U_i \geq t > L_i) g\{h_0(t) + \beta_0^\top Z_i\} \right],$$

$a(t)$ is the limit of $n^{-1} \sum_{i=1}^n I(U_i \geq t > L_i) \dot{g}\{h_0(t) + \beta_0^\top Z_i\}$, and $\dot{g}(y) = dg(y)/dy$. In Appendix 2, we also show that $\check{\omega}_{z_0}(t)$ converges weakly to a zero-mean Gaussian process. In practice, to approximate the distribution of $\check{\omega}_{z_0}(t)$, the resampling method proposed by Parzen et al. (1994) can be used. In essence, one can first generate a large number of independent random samples of $\{\mathcal{G}_i, i = 1, \dots, n\}$ from the standard normal distribution. For each sample, construct $\widehat{\omega}_{z_0}(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \widehat{\omega}_{z_0i}(t) \mathcal{G}_i$, where $\widehat{\omega}_{z_0i}(t)$ is obtained by replacing all the theoretical quantities in $\check{\omega}_{z_0i}(t)$ with their empirical counterparts. It can be shown that $\widehat{\omega}_{z_0}(t)$ and $\check{\omega}_{z_0}(t)$ converge to the same Gaussian process. Therefore, the distribution of $\omega_{z_0}(t)$ can be approximated by the realisations of $\widehat{\omega}_{z_0}(t)$. Confidence bands for the $S_{z_0}(t)$ can subsequently be obtained based on these realisations.

4 MODEL CHECKING AND MODEL SELECTION PROCEDURES

The inference procedures we propose in §2 and §3 require the specification of the link function $g(\cdot)$. Here, we present a graphical method for checking if model (1.1) with a given link function $g(\cdot)$ is

appropriate. Note that model (1.1) has an equivalent form:

$$h(T) = -\beta^\top Z + \epsilon,$$

where the distribution function of the error ϵ is $1 - g(\cdot)$. If the model is correctly specified, then the distribution of fitted residuals $\widehat{\epsilon}_i(T_i) = \widehat{h}(T_i) + \widehat{\beta}^\top Z_i$ will be close to the error distribution $1 - g(\cdot)$. As a result of censoring, the $\widehat{\epsilon}_i(T_i)$'s are not always observed. Since $h(\cdot)$ is an increasing function, the censoring pattern of $\widehat{\epsilon}_i(T_i)$ can be approximated by that of T_i . To obtain the empirical distribution for possibly censored $\widehat{\epsilon}_i(T_i)$, we use Turnbull's generalised Kaplan-Meier estimator (Turnbull, 1974) based on the observations $\{\widehat{\epsilon}_i(X_i), I(T_i < L_i), I(T_i > U_i)\}$.

To select an appropriate link function $g(\cdot)$, we consider a family of linear transformation models whose link functions are indexed by a single parameter λ :

$$g_\lambda(x) = \begin{cases} \{1 + \lambda \exp(x)\}^{-\frac{1}{\lambda}}, & \text{if } \lambda > 0; \\ \exp\{-\exp(x)\}, & \text{if } \lambda = 0. \end{cases} \quad (4.1)$$

This type of transformation is closely related to the Box-Cox transformation for the linear regression model. Note that $\lambda = 0$ corresponds to the proportional hazards model and $\lambda = 1$ corresponds to the proportional odds model. Moreover, model (4.1) corresponds to the gamma frailty model integrated over the unknown frailty; see Xu & Harrington (2001) for a discussion on the connection between the time-varying coefficient model, the transformation model and the gamma frailty model.

An optimal link function may be chosen by minimising a measure of discrepancy between the observed and fitted values. To be specific, let λ_0 denote the true value of λ . We can estimate λ_0 by minimising the statistic

$$Q(\lambda) = \sum_{i=1}^n \int_{\tau_a}^{\tau_b} \{e_i(t; \widehat{h}_\lambda, \widehat{\beta}_\lambda, g_\lambda)\}^2 d\widehat{v}(t), \quad (4.2)$$

where, for any given λ , similarly to those of equations (2.1) and (2.2), $\{\widehat{\beta}_\lambda, \widehat{h}_\lambda(\cdot)\}$ are the solutions of

$$\sum_{i=1}^n e_i(t; h, \beta, g_\lambda) = 0, \quad \tau_a \leq t \leq \tau_b, \quad (4.3)$$

$$\sum_{i=1}^n \int_{\tau_a}^{\tau_b} Z_i e_i(t; h, \beta, g_\lambda) d\widehat{v}(t) = 0, \quad (4.4)$$

and $e_i(t; h, \beta, g_\lambda) = I(X_i \geq t > L_i) - I(U_i \geq t > L_i) g_\lambda\{h(t) + \beta^\top Z_i\}$. The statistic $Q(\lambda)$ quantifies the predictive accuracy of the model under link function g_λ . A similar measure of predictive accuracy was used by Graf et al. (1999) for right-censored failure-time data.

Let $\hat{\lambda}$ denote $\operatorname{argmin}_\lambda Q(\lambda)$, and let $\{\tilde{\beta}, \tilde{h}(t)\}$ denote the solution of (4.3) and (4.4) when given $\lambda = \hat{\lambda}$. In Appendix 3, we show the consistency of $\hat{\lambda}$, $\tilde{\beta}$ and $\tilde{h}(t)$. This leads to a reasonable estimator for the survivorship function of T given covariate z_0 :

$$\tilde{S}_{z_0}(t) = g_{\hat{\lambda}}\{\tilde{h}(t) + \tilde{\beta}^\top z_0\}.$$

Asymptotic properties of $\tilde{\beta}$ and $\tilde{S}_{z_0}(t)$ are shown in Appendix 4.

5 SIMULATION STUDIES

Several simulation studies of the proposed methods were conducted for model (1.1). For the case when the link function g is given, in one of the studies, we generated T following the proportional odds model with $h(t) = \log(t/10)$ and $\beta = (\beta_1 = 1, \beta_2 = 2)^\top$. The resulting T has the survivorship function

$$\operatorname{pr}(T \geq t \mid Z_1, Z_2) = \frac{1}{1 + \exp\{\log(\frac{t}{10}) + Z_1 + 2Z_2\}},$$

where Z_1 is a standard normal random variable and Z_2 is a Bernoulli random variable with probability 0.5. Note that the median of T at baseline $Z = (0, 0)^\top$ is 10 in this setting. We generated the left censoring variable L from $\operatorname{Un}(0, 1)$, and the right censoring variable U was generated from $L + \operatorname{Un}(10, 40)$. Sample sizes of 200 and 400 were considered. The choice of h , L and U resulted in, on average, about 17% of the observations being left censored and 20% right censored. For each simulated dataset, we obtained $\hat{\beta}_1, \hat{\beta}_2$, their estimated standard errors and the predicted survival probabilities $\hat{s}_k = \hat{S}_{Z_1=0.5, Z_2=0}(t_k)$, for $k = 1, 2, 3$. Here, t_1, t_2 and t_3 are the 25th, 50th and 75th percentiles of the random variable T . In Table 1, we present empirical biases, sampling standard errors, averages of the standard error estimates and coverage probabilities of the 95% confidence intervals for $\hat{\beta}_j$ and \hat{s}_k . The results suggest that the parameter estimators have negligible biases. The standard error estimates are close to the sampling standard errors. In addition, the 95% confidence intervals have appropriate coverage probabilities. Simulation results, not shown, under settings such that T follows the proportional hazards model, are similar to those in Table 1.

To examine the performance of the estimation procedures accounting for the additional parameter λ in the link function of model (1.1), for example, we let $h(t) = 10t$, $\beta = (1, 2)^\top$ and the link function be (4.1) with true $\lambda = 2$. We simulated data from the following models:

$$\text{pr}(T \geq t \mid Z_1, Z_2) = g_2(10t + Z_1 + 2Z_2), \quad L \sim \text{Un}(-3, -3/4) + 2Z_1, \quad U \sim L + \text{Un}(0, 3) + 2Z_2,$$

where Z_1 and Z_2 are $\text{Un}(0, 1)$ random variables. Note that here censoring variables L and U depend on covariates. Under this set up, on average, about 21% of the observations were left censored and 12% right censored. For each simulated dataset, we obtained estimates for λ along with other parameters in addition to the estimates treating $\lambda = 2$ as given. Summary statistics for $\hat{\beta}_j$, \hat{s}_k , $\tilde{\beta}_j$, $\tilde{s}_k = \tilde{S}_{z_0}(t_k)$ and $\hat{\lambda}$ are presented in Table 2. We find that, like $\hat{\beta}_j$ and \hat{s}_k , the estimators $\tilde{\beta}_j$ and \tilde{s}_k also have negligible biases and their estimated standard errors are close to the empirical standard errors. As expected, the standard errors for $\tilde{\beta}_j$ are larger than those for $\hat{\beta}_j$ due to the additional variation from estimating λ . On the other hand, the standard errors of the predicted survival probabilities appear similar when λ is given and is estimated.

6 EXAMPLE

We now return to the AIDS clinical trial mentioned in §1. This multi-centre, phase II, randomised paediatric clinical trial was conducted in 1997. A total of 298 HIV-infected children from 48 sites were enrolled in the study and were followed for 48 weeks. One primary objective of the study was to compare the virologic responses to the following three treatments: Zidovudine (ZDV) plus lamivudine (3TC) (Arm A), stavudine (d4T) plus ritonavir (RTV) (Arm B), and ZDV plus 3TC plus RTV (Arm C). The main outcome measure is the plasma HIV-1 RNA level assessed using the NucliSens assay. The lower limit of assay quantification for RNA was 400 copies/ml and the upper limit was 750,000 copies/ml. Here we restrict our analysis to subjects who had initial HIV RNA level within the limit of quantification ($n = 195$) and focus on the change in \log_{10} RNA between baseline and at week 24. Let ℓ_k denote the \log_{10} RNA value at week k and let $\Delta = \ell_0 - \ell_{24}$. Then the response variable $T = \Delta$ could be left censored by $L = \ell_0 - 5.88$ and right censored by $U = \ell_0 - 2.60$. For T , there are about 4% of left censoring and 42% of right censoring. Since our preliminary analysis shows no significant treatment difference between Arms A and B, we combine these two treatment groups and compare them to Arm C. Let R denote the treatment group indicator with 1 being Arm C and

0 otherwise. The covariates of interest are ℓ_0 and R , the baseline \log_{10} RNA level and the treatment indicator. Note that the censoring variables depend on the covariates in this setting.

We first fit the model

$$\text{pr}(\Delta \geq t \mid R, \ell_0) = g \{h(t) + \beta_1 R + \beta_2 \ell_0\} ,$$

choosing $g(\cdot)$ appropriately to encapsulate the proportional hazards and odds models. The estimates of the regression parameters are summarised in Table 3. To check whether or not these models are appropriate, in Fig. 1 we compare the empirical distribution of the fitted residuals to the error distribution $1 - g_\lambda(\cdot)$ for different given λ 's. Based on Fig. 1, the model with $\lambda = 3.5$ suggests a better fit to the AIDS data than the commonly used proportional hazards and proportional odds models do as the discrepancy between the solid and dotted curves in (a) and (b) appears more systematic. To select the optimal link function among the class of transformation models described in §4 for this dataset, we fit the more general model,

$$\text{pr}(\Delta \geq t \mid R, \ell_0) = g_\lambda \{h(t) + \beta_1 R + \beta_2 \ell_0\} . \quad (6.1)$$

By minimising $Q(\lambda)$ defined in (4.2), we obtain $\hat{\lambda} = 3.5$. The estimated treatment difference is $\hat{\beta}_1 = -1.87$ with standard error 0.71, indicating that Arm C is more effective than the other treatments; p -value = 0.01. The estimated coefficient for ℓ_0 is -0.64 with standard error 0.49. A negative coefficient for ℓ_0 indicates that the decrease in RNA is more likely for subjects who have higher initial HIV RNA values, but this is not conclusive; p -value = 0.19.

Clinically, an RNA value exceeding 10,000 indicates that the patient is not doing well. To examine the virological response to the treatments for the patients in poor condition, in Fig. 2 we show the estimated survivorship function of Δ for patients with $\ell_0 = 4$ in both treatment groups. These plots also indicate that subjects in Arm C have a larger relative decrease in RNA from baseline. For example, the probability of their RNA level at week 24 having at least a 50% decrease from baseline ($\Delta \geq \log_{10} 2$) is 0.75 (0.065) if they are in treatment Arm C and 0.49 (0.039) if in treatment Arm A or B, where standard errors are given in brackets. Also shown in Fig. 2 are the 95% pointwise and simultaneous confidence intervals for the estimated survival probabilities.

For each treatment group, we also compare the estimated survival functions of Δ based on (6.1) for subjects with ℓ_0 at the sample median, which is 4.15, to Turnbull's generalised Kaplan-Meier estimates of $\text{pr}(\Delta \geq t \mid R)$. Note that, based on Laplace's approximation, $\text{pr}(\Delta \geq t \mid R, \ell_0 = 4.15) \simeq$

$\text{pr}(\Delta \geq t|R)$ since the density function of ℓ_0 peaks roughly at $\ell_0 = 4.15$. As shown in Fig. 3, our estimates are fairly similar to Turnbull’s nonparametric estimates, which supports the choice of model (6.1).

7 DISCUSSION

Practical implementation of our procedures is easy. The Newton Raphson algorithm was used to obtain estimates for the regression parameter β and the baseline function $h(\cdot)$. For the weight function \hat{v} , we recommend using the counting process of observed responses $\{X_i\}$. Empirically we find that the estimators are insensitive to the choice of \hat{v} . For model checking, we used Turnbull’s generalisation of the Kaplan-Meier estimate to obtain the distribution for the fitted residuals through the standard software `Splus 6.0` (function `kaplanMeier`) or `Splus 6.1` (function `survfit`).

When inferences are made about the regression parameter β in the transformed linear model, issues concerning whether the transformation parameter λ should be regarded as specified or as to be estimated were discussed in Bickel & Doksum (1981) and Box & Cox (1982). In the context of doubly censored data, we present inference procedures for both when λ is given and when it is unknown. Bickel & Doksum (1981) indicated that the variances of the estimated β can be inflated by large factors over the conditional variances for fixed λ . Our numerical results in Tables 2 and 3 agree with their findings. As noted in Box & Cox (1982), a relevant question is in what circumstances confidence intervals for β calculated as if $\hat{\lambda}$ were preassigned provide an adequate approximation when λ is unknown. In addition, are inference procedures for prediction more robust to whether λ is specified or estimated? These issues and many aspects of transformations are worthy of further study.

The modifications needed for extending our methods to analyse commonly observed clustered data are straightforward. The semiparametric model (1.1) can be defined in a marginal sense when the data records are clustered. Without considering the within-cluster correlation, our estimating equations provide valid estimates. Then, easy extensions of the proposed inference procedures can be made to account for the correlation between data records within the same cluster. Our method can also be extended to incorporate time-varying covariates and the model selection procedure can be extended to allow for a more general class of link functions.

ACKNOWLEDGEMENT

The authors are grateful to the editor, associate editor and a referee for their helpful comments and suggestions. We also thank L. J. Wei for his insight and supporting discussions. This work was supported in part by grants from the US National Institutes of Health and UCSF General and Pediatric Clinical Research Centre.

APPENDIX 1

Uniqueness and consistency

We assume that g is twice continuously differentiable and the baseline function $h_0(t)$ is continuous. By the strong law of large numbers, for large n , $\delta \geq 0$, $\beta \in D_\delta = \{\beta : \|\beta - \beta_0\| \leq \delta\}$, $t \in [\tau_a, \tau_b]$,

$$\frac{1}{n} \sum_{i=1}^n [I(X_i \geq t > L_i) - I(U_i \geq t > L_i)g\{h_0(t) - \varepsilon + \beta^\top Z_i\}] < 0, \quad (\text{A1.1})$$

$$\frac{1}{n} \sum_{i=1}^n [I(X_i \geq t > L_i) - I(U_i \geq t > L_i)g\{h_0(t) + \varepsilon + \beta^\top Z_i\}] > 0, \quad (\text{A1.2})$$

when ε is sufficiently large, and hence there exists a unique $\hat{h}(t; \beta)$ such that

$$\sum_{i=1}^n [I(X_i \geq t > L_i) - I(U_i \geq t > L_i)g\{\hat{h}(t; \beta) + \beta^\top Z_i\}] = 0. \quad (\text{A1.3})$$

By differentiating both sides of (A1.3) with respect to β , we have the identity

$$-\frac{\partial}{\partial \beta} \hat{h}(t; \beta) = \bar{Z}(t; \beta) \equiv \frac{\sum_{i=1}^n I(U_i \geq t > L_i) \dot{g}\{\hat{h}(t; \beta) + \beta^\top Z_i\} Z_i}{\sum_{i=1}^n I(U_i \geq t > L_i) \dot{g}\{\hat{h}(t; \beta) + \beta^\top Z_i\}}, \quad (\text{A1.4})$$

where $\dot{g}(y) = dg(y)/dy$. To show the existence and uniqueness of $\hat{\beta}$, let $V(\beta)$ be the left-hand side of (2.2). It follows from (A1.4) that $n^{-1} \partial V(\beta) / \partial \beta = -\hat{A}(\beta)$, where

$$\hat{A}(\beta) = \frac{1}{n} \sum_{i=1}^n \int_{\tau_a}^{\tau_b} \{Z_i - \bar{Z}(t; \beta)\}^{\otimes 2} I(U_i \geq t > L_i) \dot{g}\{\hat{h}(t; \beta) + \beta^\top Z_i\} d\hat{v}(t), \quad (\text{A1.5})$$

which is nonpositive definite, and, for any vector x , $x^{\otimes 0} = 1$, $x^{\otimes 1} = x$ and $x^{\otimes 2} = xx^\top$. Furthermore, since (A1.1) and (A1.2) hold for any $\varepsilon > 0$ when and only when $\beta = \beta_0$, we have that $\hat{h}(t; \beta_0) \rightarrow h_0(t)$, uniformly in $t \in [\tau_a, \tau_b]$, and $n^{-1} \partial V(\beta_0) / \partial \beta \rightarrow -A$, where

$$A = E \left[\int_{\tau_a}^{\tau_b} \{Z_1 - z(t; \beta_0)\}^{\otimes 2} \dot{g}\{h_0(t) + \beta_0^\top Z_1\} G_{Z_1}(t) dv(t) \right], \quad (\text{A1.6})$$

$G_Z(t) = \text{pr}(U \geq t > L \mid Z)$, and $z(t; \beta)$ is the limit of $\bar{Z}(t; \beta)$. When Z_i is non-degenerate, A is negative definite.

Since $n^{-1}V(\beta_0) \rightarrow 0$, by the standard inverse theorem, there exists a unique solution $\hat{\beta}$ to the equation $V(\beta) = 0$ in a neighbourhood of β_0 . This, coupled with the nonpositivity of $\hat{A}(\beta)$ for large n , ensures the uniqueness of the root of $V(\beta) = 0$ in the entire domain of β asymptotically. The above proof also implies that $\hat{\beta}$ is strongly consistent with $\hat{h}(t; \hat{\beta}) \rightarrow h_0(t)$, almost surely, uniformly in $t \in [\tau_a, \tau_b]$.

APPENDIX 2

Large sample distribution of $\hat{\beta}$ and $\omega_{z_0}(t)$

The consistency of $\hat{\beta}$ and a Taylor series expansion of $V(\hat{\beta})$ around β_0 give $n^{\frac{1}{2}}(\hat{\beta} - \beta_0) \simeq A^{-1}n^{-\frac{1}{2}}V(\beta_0)$. Taking a Taylor series expansion of $\hat{h}(t; \beta_0)$ around $h_0(t)$, we have

$$n^{-\frac{1}{2}}V(\beta_0) \simeq n^{-\frac{1}{2}} \sum_{i=1}^n \int_{\tau_a}^{\tau_b} \{Z_i - \bar{Z}(t; \beta_0)\} e_i(t; h_0, \beta_0) dv(t),$$

where $e_i(t; h, \beta) = I(X_i \geq t > L_i) - I(U_i \geq t > L_i)g\{h(t) + \beta^\top Z_i\}$. Furthermore, it follows from the uniform law of large numbers (Pollard, 1990, p.41) that $\sup_{t \in [\tau_a, \tau_b]} |\bar{Z}(t; \beta_0) - z(t; \beta_0)| \rightarrow 0$, almost surely as $n \rightarrow \infty$. The functional central limit theorem (Pollard, 1990, p.53) ensures the weak convergence of $n^{-\frac{1}{2}} \sum_{i=1}^n e_i(t; h_0, \beta_0)$. This, coupled with the strong representation theorem and the uniform convergence of $\bar{Z}(t; \beta_0) \rightarrow z(t; \beta_0)$, entails that $n^{-\frac{1}{2}}V(\beta_0)$ is asymptotically equivalent to $n^{-\frac{1}{2}} \sum_{i=1}^n V_i$.

To show the asymptotic distribution of $\omega_{z_0}(t) = n^{\frac{1}{2}}\{\hat{h}(t; \hat{\beta}) - h_0(t) + (\hat{\beta} - \beta_0)^\top z_0\}$, we take Taylor series expansions of $\hat{h}(t; \hat{\beta})$ around β_0 and $\hat{h}(t; \beta_0)$ around $h_0(t)$ and obtain

$$n^{\frac{1}{2}} \left\{ \hat{h}(t; \hat{\beta}) - h_0(t) \right\} \simeq n^{-\frac{1}{2}} \sum_{i=1}^n \left\{ -z(t; \beta_0)^\top A^{-1} V_i + a(t)^{-1} e_i(t; h_0, \beta_0) \right\}.$$

It follows that $\omega_{z_0}(t)$ is asymptotically equivalent to $\check{\omega}_{z_0}(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \check{\omega}_{z_0 i}(t)$. To show the weak convergence of $\check{\omega}_{z_0}(t)$ to a zero-mean Gaussian process, it suffices to show the finite-dimensional convergence and the tightness of $\check{\omega}_{z_0}(t)$. It is straightforward to see that, for any finite number of time points $\{t_1, t_2, \dots, t_m\}$, the joint distribution of $\{\check{\omega}_{z_0}(t_1), \check{\omega}_{z_0}(t_2), \dots, \check{\omega}_{z_0}(t_m)\}$ is asymptotically normal with mean zero. Since $a(t)$, A and $z_0 - z(t; \beta_0)$ are nonrandom, it remains to show that $n^{-\frac{1}{2}} \sum_{i=1}^n V_i$ and $n^{-\frac{1}{2}} \sum_{i=1}^n e_i(t; h_0, \beta_0)$ are tight. Since V_i does not involve t , $n^{-\frac{1}{2}} \sum_{i=1}^n V_i$ is tight.

The tightness of $n^{-\frac{1}{2}} \sum_{i=1}^n e_i(t; h_0, \beta_0)$ follows the basic properties of empirical processes (Shorack & Wellner, 1986).

APPENDIX 3 Consistency of $\hat{\lambda}$

Suppose that λ_0 is an interior point of a compact set $[0, \eta]$. To show that $\hat{\lambda}$, the minimiser of $Q(\lambda)$, is strongly consistent, it suffices to show that $n^{-1}Q(\lambda)$ converges almost surely and uniformly to a deterministic function of λ which has a unique minimiser at λ_0 (Newey & McFadden, 1994). It follows from the uniform law of large numbers that $n^{-1}Q(\lambda) \rightarrow q(\lambda)$ almost surely for any $\lambda \in [0, \eta]$, where

$$q(\lambda) = \int_{\tau_a}^{\tau_b} \int [g_{\lambda_0} \{h_{\lambda_0}(t) + \beta_{\lambda_0}^\top z\} - g_\lambda \{h_\lambda(t) + \beta_\lambda^\top z\}]^2 G_z(t) d\mathcal{L}(z) dv(t),$$

$h_\lambda(t)$ and β_λ are the limits of $\hat{h}_\lambda(t)$ and $\hat{\beta}_\lambda$, respectively, and \mathcal{L} is the distribution function of Z . It is easy to see that $q(\lambda) \geq q(\lambda_0)$ and the minimum of $q(\lambda)$ is achieved if and only if

$$g_{\lambda_0} \{h_{\lambda_0}(t) + \beta_{\lambda_0}^\top z\} = g_\lambda \{h_\lambda(t) + \beta_\lambda^\top z\}, \quad (\text{A3.1})$$

for $t \in [\tau_a, \tau_b]$ and $z \in D_z$ almost surely, where D_z is the domain of Z . For (4.1), following some partial derivatives of (A3.1), it can be shown that (A3.1) holds if and only if $\lambda = \lambda_0$. Thus, λ_0 is the unique minimiser of $q(\lambda)$. It remains to show that the convergence of $n^{-1}Q(\lambda)$ to $q(\lambda)$ is uniform in $\lambda \in [0, \eta]$. To this end, one can obtain $\dot{\beta}_\lambda^* \equiv \partial \hat{\beta}_\lambda / \partial \lambda$ and $\dot{h}_\lambda^*(t) \equiv \partial \hat{h}_\lambda(t) / \partial \lambda$ by first plugging the solutions $\{\hat{\beta}_\lambda, \hat{h}_\lambda(t)\}$ into equations (4.3) and (4.4) and differentiating both sides of the resulting equations with respect to λ . It is easy to show that $\dot{\beta}_\lambda^*$ and \dot{h}_λ^* are bounded for $\lambda \in [0, \eta]$. Therefore $n^{-1}Q(\lambda)$ is equicontinuous in λ , which implies that the convergence of $n^{-1}Q(\lambda)$ to $q(\lambda)$ is uniform in λ . This concludes the consistency of $\hat{\lambda}$ and hence the consistency of $\tilde{\beta}$ and \tilde{h} .

APPENDIX 4 Large sample distribution of $\tilde{\beta}$ and $\tilde{S}_{z_0}(t)$

Let $\dot{Q}(\lambda) = \partial Q(\lambda) / \partial \lambda$, $e_i(t; h, \beta) = e_i(t; h, \beta, g_{\lambda_0})$, $h_0(t) = h_{\lambda_0}(t)$, $\beta_0 = \beta_{\lambda_0}$, $\hat{h}(t) = \hat{h}_{\lambda_0}(t)$ and $\hat{\beta} = \hat{\beta}_{\lambda_0}$. It follows from a Taylor series expansion that

$$n^{\frac{1}{2}}(\hat{\lambda} - \lambda_0) \doteq \left\{ -n^{-1} \frac{\partial \dot{Q}(\lambda_0)}{\partial \lambda} \right\}^{-1} n^{-\frac{1}{2}} \dot{Q}(\lambda_0). \quad (\text{A4.1})$$

Let $\hat{h}_\lambda(t)$ and $\hat{\beta}_\lambda$ denote the limits of $\hat{h}_\lambda^*(t)$ and $\hat{\beta}_\lambda^*$, respectively. The uniform law of large numbers ensures that $\sup_{\lambda \in [0, \eta]} |\hat{\beta}_\lambda^* - \hat{\beta}_\lambda| \rightarrow 0$ and $\sup_{\lambda \in [0, \eta], t \in [\tau_a, \tau_b]} |\hat{h}_\lambda^*(t) - \hat{h}_\lambda(t)| \rightarrow 0$ almost surely. It follows from the uniform convergence of $\hat{h}_\lambda(t)$ and $\hat{\beta}_\lambda$ and the large sample properties of $\hat{\beta}$ and $\hat{h}(t)$ that $n^{-\frac{1}{2}} \hat{Q}(\lambda_0) \simeq n^{-\frac{1}{2}} \sum_{i=1}^n q_i$, where

$$q_i = \int_{\tau_a}^{\tau_b} \left[e_i(t; h_0, \beta_0) \left\{ w_i(t) - \frac{b_x^{(0)}(t)}{a(t)} \right\} - \left\{ b_x^{(1)}(t) - b_x^{(0)}(t) z(t; \beta_0) \right\}^\top A^{-1} V_i \right] dv(t),$$

$w_i(t) = -2I(U_i \geq t > L_i) [\dot{g}_{\lambda_0} \{h_0(t) + \beta_0^\top Z_i\} + \dot{g}_{\lambda_0} \{h_0(t) + \beta_0^\top Z_i\} \{\dot{h}_{\lambda_0}(t) + \dot{\beta}_{\lambda_0}^\top Z_i\}]$, $b_x^{(k)}(t)$ is the limit of $n^{-1} \sum_{i=1}^n w_i(t) \dot{g}_{\lambda_0} \{h_0(t) + Z_i^\top \beta_0\} Z_i^{\otimes k}$ for $k = 0$ and 1 , and for any function $f_\lambda(x)$ we use the notation $\dot{f}_\lambda(x)$ to denote $\partial f_\lambda(x) / \partial \lambda$ and use $f'_\lambda(x)$ to denote $\partial f_\lambda(x) / \partial x$. It is easy to see from the uniform law of large numbers that $-n^{-1} \partial \hat{Q}(\lambda_0) / \partial \lambda \rightarrow \rho \equiv \frac{1}{2} \int_{\tau_a}^{\tau_b} E \{w_1(t)^2\} dv(t)$ almost surely. From (A4-1), we have $n^{\frac{1}{2}}(\hat{\lambda} - \lambda_0) \simeq n^{-\frac{1}{2}} \sum_{i=1}^n \rho^{-1} q_i$. It follows from Taylor series expansions of $\tilde{\beta}$ and $\tilde{h}(t)$ around $h_0(t)$ and β_0 that

$$n^{\frac{1}{2}}(\tilde{\beta} - \beta_0) \simeq n^{-\frac{1}{2}} \sum_{i=1}^n (\dot{\beta}_{\lambda_0} \rho^{-1} q_i + A^{-1} V_i),$$

$$n^{\frac{1}{2}} \{\tilde{h}(t) - h_0(t)\} \simeq n^{-\frac{1}{2}} \sum_{i=1}^n \left\{ \dot{h}_{\lambda_0}(t) \rho^{-1} q_i + \frac{e_i(t; h_0, \beta_0)}{a(t)} - z(t; \beta_0)^\top A^{-1} V_i \right\}.$$

Therefore, the distribution of $\tilde{\beta}$ can be approximated by a normal random vector with mean β_0 and covariance matrix

$$n^{-2} \sum_{i=1}^n \left(\rho^{-2} q_i^2 \dot{\beta}_{\lambda_0} \dot{\beta}_{\lambda_0}^\top + A^{-1} V_i V_i^\top A^{-1} \right).$$

The consistency of $\hat{\lambda}$, $\tilde{h}(t)$ and $\tilde{\beta}$ ensures the consistency of $\tilde{S}_{z_0}(t)$. Arguments similar to those in Appendix 2 can be used to show that $n^{\frac{1}{2}} \{\tilde{S}_{z_0}(t) - S_{z_0}(t)\}$ is asymptotically equivalent to $n^{\frac{1}{2}} \sum_{i=1}^n \mathcal{S}_i(t)$, which converges to a zero-mean Gaussian process, where $S_z(t) = g_{\lambda_0} \{h_0(t) + \beta_0^\top z\}$ and

$$\mathcal{S}_i(t) = \rho^{-1} q_i \left[\dot{g}_{\lambda_0} \{h_0(t) + \beta_0^\top z_0\} + \dot{g}_{\lambda_0} \{h_0(t) + \beta_0^\top z_0\} \left\{ \dot{h}_{\lambda_0}(t) + \dot{\beta}_{\lambda_0}^\top z_0 \right\} \right] \\ + \dot{g}_{\lambda_0} \{h_0(t) + \beta_0^\top z_0\} \left[\frac{e_i(t; h_0, \beta_0)}{a(t)} + \{z_0 - z(t; \beta_0)\}^\top A^{-1} V_i \right].$$

REFERENCES

BENNETT, S. (1983). Analysis of survival data by the proportional odds model. *Statist. Med.* **2**, 273–7.

- BICKEL, P. J. & DOKSUM, K. A. (1981). An analysis of transformations revisited. *J. Am. Statist. Assoc.* **76**, 296–311.
- BOX, G. E. P. & COX, D. R. (1982). An analysis of transformations revisited, rebutted. *J. Am. Statist. Assoc.* **77**, 209–10.
- CAI, T., WEI, L. J. & WILCOX, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika* **87**, 867–878.
- CHANG, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18**, 391–404.
- CHEN, K., JIN, Z. & YING, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89**, 659–68.
- CHENG, S. C., WEI, L. J. & YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–45.
- CHENG, S. C., WEI, L. J. & YING, Z. (1997). Prediction of survival probabilities with semi-parametric transformation models. *J. Am. Statist. Assoc.* **92**, 227–35.
- CLAYTON, D. & CUZICK, J. (1985). Multivariate generalizations of the proportional hazards model (with Discussion). *J. R. Statist. Soc. A* **148**, 82–117.
- COX, D. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc. B* **34**, 187–220.
- GEHAN, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52**, 650–3.
- GESKUS, R. B. (2001). Methods for estimating the AIDS incubation time distribution when date of seroconversion is censored. *Statist. Med.* **20**, 795–812.
- GRAF, E., SCHMOOR, C., SAUERBREI, W. & SCHUMACHER, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statist. Med.* **18**, 2529–45.
- GU, M. G. & ZHANG, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21**, 611–24.
- HUGHES, M. D. (2000). Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in HIV clinical trials. *Statist. Med.* **19**, 3171–91.
- KIM, M. Y., DE GRUTTOLA, V. G. & LAGAKOS, S. W. (1993). Analyzing doubly censored data with covariates, with application to Aids. *Biometrics* **49**, 13–22.

- MANTEL, N. (1967). Ranking procedures for arbitrarily restricted observation. *Biometrics* **23**, 65–78.
- MURPHY, S. A., ROSSINI, A. J. & VAN DER VAART, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *J. Am. Statist. Assoc.* **92**, 968–76.
- NEWBY, W. & MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Ed. D. McFadden and R. Engler, volume IV, pages 2113–245. Amsterdam: North Holland.
- PARZEN, M. I., WEI, L. J. & YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–50.
- POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. Hayward, CA: Institute of Mathematical Statistics.
- REN, J.-J. & GU, M. (1997). Regression M -estimators with doubly censored data. *Ann. Statist.* **25**, 2638–64.
- SHORACK, G. R. & WELLNER, J. A. (1986). *Empirical Processes With Applications to Statistics*. New York: Wiley.
- SUN, J. (1995). Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to Aids studies. *Biometrics* **51**, 1096–104.
- TSAI, W.-Y. & CROWLEY, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency (Corr: V18 p470). *Ann. Statist.* **13**, 1317–34.
- TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Am. Statist. Assoc.* **69**, 169–73.
- XU, R. & HARRINGTON, D. P. (2001). A semiparametric estimate of treatment effects with censored data. *Biometrics* **57**, 875–85.
- YANG, S. & PRENTICE, R. (1999). Semiparametric inference in the proportional odds regression model. *J. Am. Statist. Assoc.* **94**, 125–36.
- ZHANG, C.-H. & LI, X. (1996). Linear regression with doubly censored data. *Ann. Statist.* **24**, 2720–43.

Table 1: Empirical bias (Bias), sampling standard errors (Sse), average of the estimated standard errors (Ase) and 95% empirical coverage probability (Ecp) for $\hat{\beta}_j$ and \hat{s}_k under the proportional odds model ($\lambda = 1$). Results are based on 1000 realisations with $\beta_1 = 1$ and $\beta_2 = 2$.

	$n = 200$					$n = 400$				
	$\hat{\beta}_1$	$\hat{\beta}_2$	\hat{s}_1	\hat{s}_2	\hat{s}_3	$\hat{\beta}_1$	$\hat{\beta}_2$	\hat{s}_1	\hat{s}_2	\hat{s}_3
Bias	0.014	0.026	0.005	0.011	0.003	0.006	0.014	0.002	0.004	0.001
Sse	0.168	0.300	0.043	0.044	0.025	0.114	0.217	0.029	0.030	0.018
Ase	0.156	0.295	0.043	0.043	0.028	0.110	0.208	0.030	0.031	0.020
Ecp	0.932	0.949	0.959	0.944	0.968	0.951	0.933	0.959	0.951	0.968

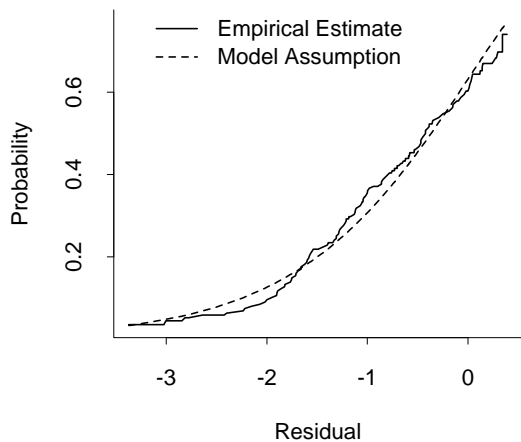
Table 2: Empirical bias (Bias), sampling standard errors (Sse), average of the estimated standard errors (Ase) and 95% empirical coverage probability (Ecp) for $\hat{\beta}_j$ and \hat{s}_k when $\lambda = 2$ is known and $\tilde{\beta}_j$ and \tilde{s}_k when λ is estimated. Results are based on 1000 simulated datasets with $\beta_1 = 1$, $\beta_2 = 2$ and sample size $n = 200$.

		$\hat{\lambda}$	$\hat{\beta}_1$	$\hat{\beta}_2$	\hat{s}_1	\hat{s}_2	\hat{s}_3
Given $\lambda = 2$	Bias	—	0.027	0.038	0.011	0.009	0.006
	Sse	—	0.797	0.735	0.041	0.043	0.034
	Ase	—	0.753	0.727	0.042	0.045	0.035
	Ecp	—	0.937	0.946	0.945	0.954	0.959
		$\hat{\lambda}$	$\tilde{\beta}_1$	$\tilde{\beta}_2$	\tilde{s}_1	\tilde{s}_2	\tilde{s}_3
Estimated λ	Bias	0.104	0.001	0.064	0.053	0.047	0.004
	Sse	1.833	0.942	1.106	0.042	0.042	0.034
	Ase	1.805	0.972	1.219	0.042	0.045	0.036
	Ecp	0.930	0.940	0.951	0.951	0.956	0.962

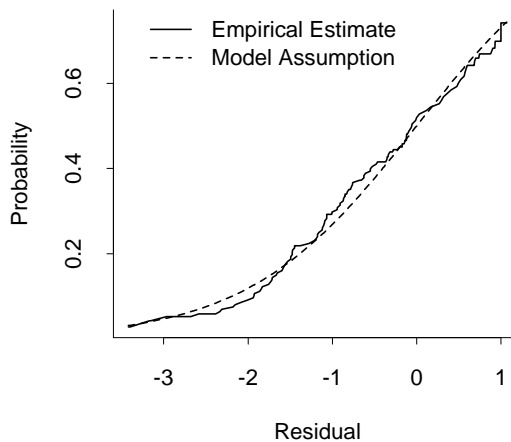
Table 3: Estimated regression parameters, with standard errors in brackets for the RNA data: the proportional hazards model ($\lambda = 0$), the proportional odds model ($\lambda = 1$) and the model with link function $g_\lambda(\cdot)$ at $\lambda = 3.5$.

	λ	ℓ_0	R
Given λ	0	-0.30 (0.17)	-0.90 (0.24)
	1	-0.39 (0.24)	-1.15 (0.30)
	3.5	-0.64 (0.40)	-1.87 (0.49)
Estimated λ	3.5 (3.02)	-0.64 (0.49)	-1.87 (0.71)

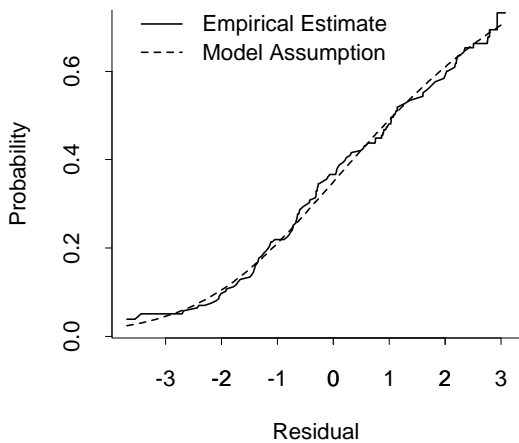
Figure 1: The empirical estimates of the distribution function of $\hat{h}(T_i) + \hat{\beta}^\top Z_i$ compared to the distribution of ϵ_i under the model assumption at various λ 's: (a) $\lambda = 0$, (b) $\lambda = 1$, (c) $\lambda = 3.5$, (d) $\lambda = 5$.



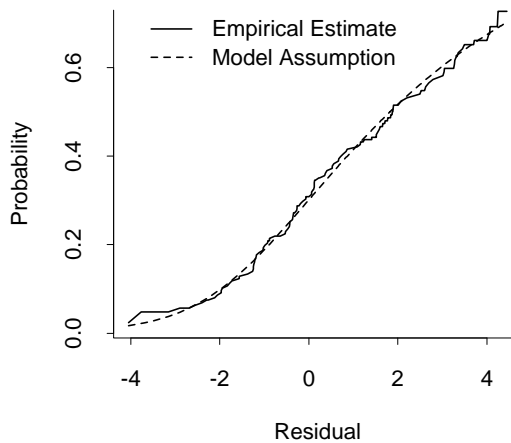
(a) $\lambda = 0$



(b) $\lambda = 1$

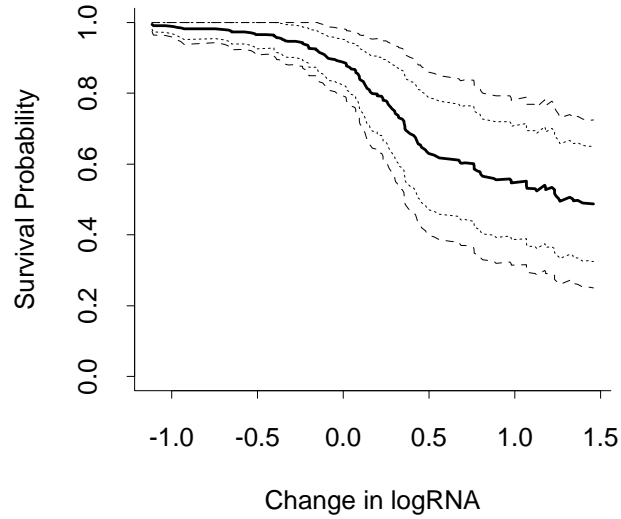


(c) $\lambda = 3.5$

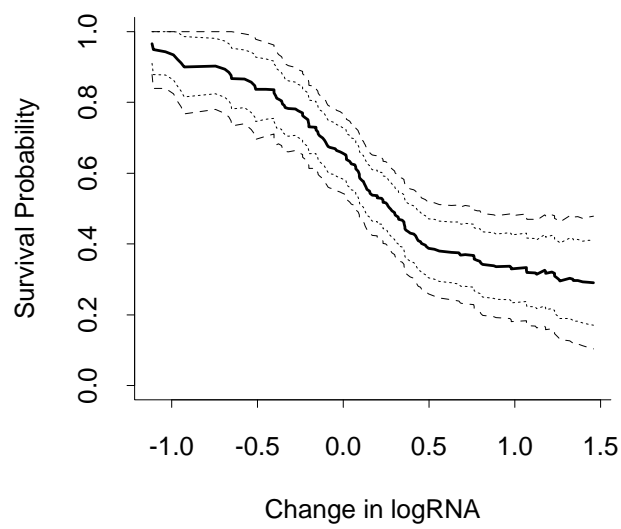


(d) $\lambda = 5$

Figure 2: The estimated survivorship function (solid curve) of Δ for subjects with baseline RNA 10,000 copies/ml in treatment arm (a) C and (b) A or B. Also shown are their 95% pointwise (dotted curve) and simultaneous (dashed curve) confidence intervals.



(a) Treatment Arm C



(b) Treatment Arm A or B

Figure 3: The estimated survivorship function (solid lines) of Δ for subjects with $\ell_0 = 4.15$ in each treatment group. Also shown are Turnbull's generalised Kaplan-Meier estimates (dashed lines) obtained by ignoring the baseline RNA level.

