Harvard University
Harvard University Biostatistics Working Paper Series

Year 2003 *Paper* 6

Semi-parametric Box-Cox Power Transformation Models for Censored Survival Observations

Tianxi Cai* Lu Tian† L. J. Wei‡

*Harvard University, tcai@hsph.harvard.edu
†Harvard University, ltian@hsph.harvard.edu
‡Harvard University, wei@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

http://biostats.bepress.com/harvardbiostat/paper6

Copyright ©2003 by the authors.

SEMI-PARAMETRIC BOX-COX POWER TRANSFORMATION MODELS FOR CENSORED SURVIVAL OBSERVATIONS

Tianxi Cai, Lu Tian and L.J. Wei

Department of Biostatistics, Harvard University

655 Huntington Ave, Boston, MA 02115

SUMMARY

The accelerated failure time model specifies that the logarithm of the failure time is linearly related to the covariate vector without assuming a parametric error distribution. In this article, we consider the semi-parametric Box-Cox transformation model, which includes the above regression model as a special case, to analyse possibly censored failure time observations. Inference procedures for the transformation and regression parameters are proposed via a resampling technique. Prediction of the survival function of future subjects with a specific covariate vector is also provided via point-wise and simultaneous interval estimates. All the proposals are illustrated with the data sets from two clinical studies.

Some key words: Accelerated failure time model; Prediction; Resampling method; Simultaneous confidence interval.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

1 INTRODUCTION

The most popular semi-parametric regression model, which *directly* links the failure time to its covariates, for analysing censored survival data is the accelerated failure time model. This model simply relates the logarithm of the failure time linearly to its covariates without specifying a parametric error distribution (Kalbfleisch & Prentice, 2002, Ch.7). Recently, Jin et al. (2003) proposed an iterative estimating procedure for the regression coefficients of this log-linear model based on a class of monotone estimating functions, which can be implemented efficiently via the standard linear programming technique. With this new proposal, such a log-linear regression model can be a useful, practical alternative to the Cox proportional hazards model (Cox, 1972) for analysing survival observations.

The accelerated failure time model may not fit the data well due to, for example, the mis-specification of the log-link function. One possible remedy is to consider a class of flexible Box-Cox transformations for the response variable. For the case that there are no censored observations, the parametric Box-Cox power transformation model has been extensively studied (Box & Cox, 1964, 1982; Bickel & Doksum, 1981; Carroll & Rupert, 1985, 1987; Taylor, 1987). The transformed response yields a linear regression model with normal error and constant variance. Without assuming a parametric error distribution, inference procedures for the semi-parametric Box-Cox transformation model have been proposed, for example, by Han (1987), Newey (1990), Robinson (1991), Wang and Ruppert (1995) and Foster et al. (2001).

In this article, we study the semi-parametric power transformation models for analysing possibly censored survival data, where the censoring variable may depend on the covariates.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper6

This class of models includes the accelerated failure time model as a special case. Inference procedures for the transformation and regression parameters are proposed via a simple re-sampling scheme. Although the physical interpretation of the regression coefficients under a transformation model may not be obvious, this type of model is quite flexible and useful, for example, for predicting the survival function of future patients with a specific covariate vector. In this paper, we also show how to construct point-wise and simultaneous confidence intervals for such a survival function. The new proposals are illustrated extensively with the data sets from two clinical studies.

2 ESTIMATING TRANSFORMATION AND REGRESSION PARAMETERS

Let T be the time to the event of interest and Z be the corresponding $p \times 1$ vector of bounded covariates. Assume that the support of Z is not contained in a $(p-1)$ dimensional hyper-plane. A semi-parametric Box-Cox transformation model specifies that

$$g_{\lambda_0}(T) = \beta_0' Z + \epsilon, \tag{2·1}$$

where

$$g_\lambda(t) = \begin{cases} \frac{t^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \log(t) & \text{if } \lambda = 0, \end{cases}$$

β_0 is a $p \times 1$ vector of unknown regression parameters, λ_0 is an unknown transformation parameter and ϵ has a completely unspecified, continuous density function, which is free of Z. We assume that $\theta_0 = \binom{\lambda_0}{\beta_0}$ is an interior point of a compact set. Note that β_0 does not include the intercept term. The failure time T may be censored by a variable C. Let $X = \min(T, C)$ and

COBRA
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

$\Delta = 1$, if T is observed, 0, otherwise. The distribution of C may depend on Z, but conditional on Z, T and C are independent of each other. Also, assume that $(T_i, Z_i, C_i, \epsilon_i), i = 1, \cdots, n$, are independent copies generated from (T, Z, C, ϵ). In Appendix 1, we show that in the presence of censoring, β_0, λ_0 are identifiable and the error distribution function $F(\cdot)$ is identifiable in the support of $g_{\lambda_0}(X) - \beta_0' Z$. We are interested in making inferences about λ_0 and β_0 based on $\{(X_i, Z_i, \Delta_i)\}$, n independent copies of (X, Z, Δ).

Suppose that λ_0 is known, the regression parameter β_0 can be estimated based on the weighted logrank estimating functions studied by Tsiatis (1990), Ritov (1990) and Wei et al. (1990). A special rank estimator $\widehat{\beta}_{\lambda_0}$ with the Gehan-type weight function can be obtained by minimising a simple U-process of β. Numerically this minimisation can be implemented easily via the standard linear programming technique (Jin et al., 2003). Specifically, for a fixed λ, $\widehat{\beta}_\lambda$ is obtained by minimising the function

$$L(\theta) \equiv n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i \left\{ \varepsilon_i(\theta) - \varepsilon_j(\theta) \right\} I \left\{ \varepsilon_i(\theta) \geq \varepsilon_j(\theta) \right\}, \tag{2·2}$$

with respect to β, where $\theta = \binom{\lambda}{\beta}$, $I(\cdot)$ is the indicator function, and $\varepsilon_i(\theta) = g_\lambda(X_i) - \beta' Z_i$. Then $\widehat{\beta}_{\lambda_0}$ is a consistent estimator for β_0 and its distribution can be approximated by a normal, whose covariance matrix can be obtained via a simple resampling method (Jin et al., 2003).

Since (2·2) is a U-process (Honore and Powell, 1994), it follows that $\widehat{\beta}_\lambda$ converges to β_λ, uniformly in λ, where β_λ is the minimiser of the limit of (2·2), and $\beta_{\lambda_0} = \beta_0$. Therefore, if one can obtain a "good" estimator $\widehat{\lambda}$ for λ_0, $\widehat{\beta}_{\widehat{\lambda}}$ is expected to be a reasonably good estimator for β_0. Now, for a fixed λ, consider the estimated martingale residual $\{\widehat{M}_i(t, \widehat{\theta}_\lambda), t \geq 0\}$ for the

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper6

counting process $\{N_i(t) = I(X_i \leq t)\Delta_i, t \geq 0\}, i = 1, \cdots, n$, where

$$\widehat{M}_i(t, \theta) = N_i(t) - \widehat{A}_i(t, \theta), \tag{2·3}$$

$\widehat{\theta}_\lambda = \binom{\lambda}{\widehat{\beta}_\lambda}$, $\widehat{A}_i(t, \theta) = \int_0^t I(X_i \geq u) d\widehat{\Lambda}(g_\lambda(u) - \beta' Z_i, \theta)$, and

$$\widehat{\Lambda}(y, \theta) = \sum_{i=1}^{n} \frac{\Delta_i I(\varepsilon_i(\theta) \leq y)}{\sum_{j=1}^{n} I\{\varepsilon_j(\theta) \geq \varepsilon_i(\theta)\}}. \tag{2·4}$$

Note that $\widehat{\Lambda}(\cdot, \widehat{\theta}_{\lambda_0})$ is a consistent estimator for the cumulative hazard function of the error term ϵ in (2·1). When $\lambda = \lambda_0$, the expected value of the process $\{\widehat{M}_i(t, \widehat{\theta}_\lambda), t \geq 0\}$ is approximately 0. To examine the adequacy of the link function g_λ for Model (2·1), consider the following process in s and t, which consists of partial sums of the estimated martingale residuals:

$$\widehat{Q}_\lambda(s, t) = n^{-1} \sum_{i=1}^{n} I(\widehat{\beta}_\lambda' Z_i \leq s) \widehat{M}_i(t, \widehat{\theta}_\lambda).$$

The above process is expected to be around 0 when $\lambda = \lambda_0$. A "large deviation" of the observed value $\{\widehat{Q}_\lambda(s, t)\}$ from 0 suggests that the assumed link function g_λ is not correct. This type of model checking techniques was proposed by Lin et al. (1993, 2002) for examining the adequacy of the link function for various regression models and for the Cox model in the presence of censoring.

A reasonable lack-of-fit measure for the assumed link function g_λ of Model (2·1) is

$$\widehat{Q}(\lambda) = \int_{-\infty}^{\infty} \int_0^{\tau} \widehat{Q}_\lambda^2(s, t) dw(t) dv(s), \tag{2·5}$$

where $w(\cdot)$ and $v(\cdot)$ are positive, bounded, differentiable, strictly increasing, possibly data dependent functions that converge to deterministic functions uniformly. This motivates us

Hosted by The Berkeley Electronic Press

to estimate λ_0 by a minimiser $\widehat{\lambda}$ for (2·5). Since $\widehat{Q}(\lambda)$ is a function of a single parameter λ, its global minimiser can be readily obtained within a reasonable interval, for example, $\{\lambda : |\lambda| \leq 2\}$ (Carroll, 1982). To estimate β_0, we let $\widehat{\beta} = \widehat{\beta}_{\widehat{\lambda}}$. We show in Appendix 2 that $\widehat{\lambda}$ and $\widehat{\beta}$ are strongly consistent for λ_0 and β_0, respectively. Let $\widehat{\theta} = \binom{\widehat{\lambda}}{\widehat{\beta}}$. We also show in Appendix 3 that $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$ converges in distribution to a zero-mean multivariate normal, whose covariance matrix, however, depends on the unknown density function of the error term in (2·1) and is difficult to estimate well directly under the nonparametric setting.

Here, we present a relatively simple resampling method to estimate the distribution of $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$. Let (x_i, δ_i, z_i) be the observed value of (X_i, Δ_i, Z_i). Also, let $\{V_i, i = 1, ..., n\}$ be a set of independent copies generated from a positive random variable V with a known distribution whose mean and variance are one. Then, for a fixed λ, we consider a stochastically perturbed $L(\theta)$ in (2·2):

$$L^*(\theta) \equiv n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} V_i \delta_i \left\{\tilde{\varepsilon}_i(\theta) - \tilde{\varepsilon}_j(\theta)\right\} I(\tilde{\varepsilon}_i(\theta) \geq \tilde{\varepsilon}_j(\theta)), \tag{2·6}$$

where $\tilde{\varepsilon}(\theta)$ is the observed value of $\varepsilon(\theta)$. Let $\beta^*_\lambda = \text{argmin}_\beta L^*(\theta)$. To obtain a corresponding perturbed $\widehat{Q}(\lambda)$ in (2·5), first we perturb the estimated martingale residuals (2·3):

$$M^*_i(t, \theta) = I(x_i \leq t)\delta_i - \int_0^t \frac{I(x_i \geq u) d\left[\sum_{j=1}^n I\{\tilde{\varepsilon}_j(\theta) \leq g_\lambda(u) - \beta' z_i\}\delta_j V_j\right]}{\sum_{l=1}^n I\{\tilde{\varepsilon}_l(\theta) \geq g_\lambda(u) - \beta' z_i\} V_l}.$$

This results in a perturbed $\widehat{Q}(\lambda)$, which is

$$Q^*(\lambda) = \int_{-\infty}^{\infty} \int_0^{\tau} \left\{ n^{-1} \sum_{i=1}^{n} I(z_i' \beta^*_\lambda \leq s) M^*_i(t, \theta^*_\lambda) V_i \right\}^2 dw(t) dv(s), \tag{2·7}$$

where $\theta^*_\lambda = \binom{\lambda}{\beta^*_\lambda}$. Let λ^* be a minimiser of (2·7), $\beta^* = \beta^*_{\lambda^*}$, $\theta^* = \binom{\lambda^*}{\beta^*_{\lambda^*}}$. In Appendix 4 we show that for large n, the unconditional distribution of $n^{1/2}(\widehat{\theta} - \theta_0)$ can be approximated by that of

http://biostats.bepress.com/harvardbiostat/paper6

$n^{1/2}(\theta^* - \tilde{\theta})$, where $\tilde{\theta}$ be the observed value of $\hat{\theta}$. In practice, to obtain the above approximation, one may generate a large number, $\mathcal{M}$, of random samples $\{V_i, i = 1, \cdots, n\}$, and for each realized sample, obtain λ^* and β^* by minimising (2·6) and then (2·7). The covariance matrix of $(\theta^* - \tilde{\theta})$ or $(\hat{\theta} - \theta_0)$ can then be approximated via the sample covariance matrix based on those $\mathcal{M}$ realizations of θ^*.

Possible choice for $w(\cdot)$ and $v(\cdot)$ in (2.5) are the empirical distribution function based on $\{X_i, i = 1, \ldots, n\}$ and $n^{-1}\sum_{i=1}^{n} I(\hat{\beta}_\lambda' Z_i \leq s)$, respectively. We find via numerical examples in Section 4 that the resulting point and interval estimates are quite stable with respect to these two weight functions.

3 PREDICTING CUMULATIVE HAZARD AND SURVIVAL FUNCTIONS FOR FUTURE SUBJECTS

For a given covariate vector z_0, the cumulative hazard function $\Lambda_{z_0}(t)$ is $\Lambda(t_{\theta_0})$, where $\Lambda(\cdot)$ is the cumulative hazard function of ϵ in model (2·1) and $t_\theta = g_\lambda(t) - \beta' z_0$. It follows that a consistent estimator $\widehat{\Lambda}_{z_0}(t)$ for $\Lambda_{z_0}(t)$ is $\widehat{\Lambda}(t_{\widehat{\boldsymbol{\theta}}}, \widehat{\theta})$ for all t such that $t_{\boldsymbol{\theta}_0}$ is within the support of $g_{\lambda_0}(X) - \beta_0' Z$, where $\widehat{\Lambda}(\cdot, \cdot)$ is defined in (2·4). Now, consider the process $\mathcal{V}(t) = n^{\frac{1}{2}}\left\{\widehat{\Lambda}_{z_0}(t) - \Lambda_{z_0}(t)\right\}$, which can be written as

$$n^{\frac{1}{2}}\left\{\widehat{\Lambda}(t_{\widehat{\boldsymbol{\theta}}}, \widehat{\theta}) - \widehat{\Lambda}(t_{\boldsymbol{\theta}_0}, \theta_0)\right\} + n^{\frac{1}{2}}\left\{\widehat{\Lambda}(t_{\boldsymbol{\theta}_0}, \theta_0) - \Lambda(t_{\boldsymbol{\theta}_0})\right\}. \tag{3·1}$$

It follows from the arguments given in Appendix 3 that the first term of (3·1) can be approximated by $n^{\frac{1}{2}}\left\{\Lambda(t_{\widehat{\boldsymbol{\theta}}}, \widehat{\theta}) - \Lambda(t_{\boldsymbol{\theta}_0})\right\}$, which is asymptotically equivalent to

$$A(t, \theta_0) n^{\frac{1}{2}}(\widehat{\theta} - \theta_0), \tag{3·2}$$

Hosted by The Berkeley Electronic Press

where $A(t, \theta_0)$ is a deterministic matrix. The second term in (3·1) is

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_{-\infty}^{t_{\theta_0}} \frac{d\left[\Delta_i I\{\varepsilon_i(\theta_0) \leq y\} - \int_{-\infty}^{y} I\{\varepsilon_i(\theta_0) \geq x\} d\Lambda(x, \theta_0)\right]}{\sum_{j=1}^{n} I\{\varepsilon_j(\theta_0) \geq y\}}. \tag{3·3}$$

In Appendix 3, we showed that $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$ is asymptotically equivalent to a sum of independent, identically distributed random vectors. This, coupled with the martingale central limit theorem, justifies the weak convergence of $\mathcal{V}(t) = (3{\cdot}2) + (3{\cdot}3)$ to a zero mean Gaussian process in t.

The covariance function of the process $\mathcal{V}(t)$, however, is prohibitively complex and the large sample properties of the above limiting Gaussian process are rather difficult, if not impossible, to obtain analytically. Here, we use a resampling technique similar to that presented in Section 2 to approximate the distribution of $\mathcal{V}(t)$. To this end, let θ^* be the random vector generated by the resampling method in the previous section. Then, it follows from the same arguments given in Appendix 4 that the distribution of (3.2)+(3.3) can be approximated by the conditional distribution of

$$\mathcal{V}^*(t) = n^{\frac{1}{2}}\{\tilde{\Lambda}(t_{\theta^*}, \theta^*) - \tilde{\Lambda}(t_{\tilde{\theta}}, \tilde{\theta})\} + n^{\frac{1}{2}} \int_{-\infty}^{t_{\tilde{\theta}}} \frac{\sum_{i=1}^{n} \left[\delta_i dI\{\tilde{\varepsilon}_i(\widetilde{\theta}) \leq u\} - I\{\tilde{\varepsilon}_i(\widetilde{\theta}) \geq u\} d\widetilde{\Lambda}(u, \widetilde{\theta})\right] (V_i - 1)}{\sum_{i=1}^{n} I\{e_i(\widetilde{\theta}) \geq u\}},$$

where $\tilde{\Lambda}(\cdot, \cdot)$ is the observed value of $\hat{\Lambda}(\cdot, \cdot)$.

By the functional δ-method, for any given differentiable function $\mathcal{G}(\cdot)$, the distribution of $\mathcal{V}_{\mathcal{G}}(t) = n^{\frac{1}{2}}\left[\mathcal{G}\left\{\widehat{\Lambda}_{z_0}(t)\right\} - \mathcal{G}\left\{\Lambda_{z_0}(t)\right\}\right]$ and $\mathcal{V}^*_{\mathcal{G}}(t) = \dot{\mathcal{G}}\left\{\widetilde{\Lambda}(t_{\widetilde{\theta}}, \widetilde{\theta})\right\} \mathcal{V}^*(t)$ converges weakly to the same Gaussian process, where $\dot{\mathcal{G}}$ is the derivative of $\mathcal{G}$. To obtain an approximation to the distribution of the random process $\mathcal{V}_{\mathcal{G}}(\cdot)$, we generate $\mathcal{M}$ realizations of θ^*, and obtain the corresponding realizations of $\mathcal{V}^*_{\mathcal{G}}(\cdot)$.

COBRA
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper6

Now, suppose that we are interested in constructing $(1-\alpha)$ point-wise and simultaneous confidence intervals for $\Lambda_{z_0}(t)$, where $0 < \alpha < 1$. To this end, let $\mathcal{G}(y) = \log(y)$, and let $\tilde{\sigma}_{z_0}(t)$ be the observed estimated standard error for $\log(\hat{\Lambda}_{z_0}(t))$, which may be obtained via the above $\mathcal{M}$ realizations of $\mathcal{V}^*_{\mathcal{G}}(t)$. Then, the $(1-\alpha)$ interval is

$$\hat{\Lambda}_{z_0}(t) \exp\left\{\pm c_\alpha \hat{\sigma}_{z_0}(t)\right\}. \tag{3·4}$$

For the point-wise interval, c_α is the $100(1-\alpha/2)$th percentile of the standard normal. For the simultaneous interval for $\Lambda_{z_0}(t)$, $0 < t < \tau$, the cutoff point c_α is chosen such that

$$\text{pr}\left\{\sup_{0\le t\le\tau}\left|\frac{\mathcal{V}^*_{\mathcal{G}}(t)}{n^{1/2}\tilde{\sigma}_{z_0}(t)}\right| < c_\alpha\right\} \approx 1-\alpha. \tag{3·5}$$

For estimating the corresponding survival function $S_{z_0}(t)$, since $S_{z_0}(t) = \exp(-\Lambda_{z_0}(t))$, the $(1-\alpha)$ interval of such a function is

$$\exp\left[\hat{\Lambda}_{z_0}(t) \exp\left\{\mp c_\alpha \hat{\sigma}_{z_0}(t)\right\}\right], \tag{3·6}$$

where c_α is either the upper $\alpha/2$ cutoff point from the standard normal or obtained via (3·5).

4 EXAMPLE

We use two examples to illustrate our estimation procedure. The first one is from the well-known Mayo primary biliary cirrhosis study (Fleming & Harrington, 1991, Appendix D). The data set in this example consists of information from 418 patients on the survival time and prognostic factors. Recently, Jin et al. (2003) and Park & Wei (2003) used the accelerated failure time model to analyse this set of data with five covariates: age, oedema, log(bilirubin), log(albumin) and log(protime). Based on the semi-parametric Box-Cox model (2·1) with the

Hosted by The Berkeley Electronic Press

above set of covariates, $\widehat{\lambda} = 0.102$ and its estimated standard error is 0.097. In Table 1, we present the point estimates with the corresponding estimated standard errors for these five covariates, which are practically identical to those reported in Jin et al. (2003). Here, for $\hat{Q}(\lambda)$ in (2.5) and $Q^*(\lambda)$ in (2.7), we let $\tau = 12$ years, $v(s) = n^{-1}\sum_{i=1}^{n} I(\widehat{\beta}_\lambda' Z_i \leq s)$, $w(t) = 200^{-1}\sum_{k=1}^{200} I(\eta_k \leq t)$, where η_k is the $(k/2)$th percentile of the empirical distribution function based on $\{X_i, i = 1, \ldots, 418\}$, V is the unit exponential and $\mathcal{M} = 1000$. Note that although we let $w(\cdot)$ be a crude approximation to the empirical distribution of $\{X_i\}$ to reduce the amount of computing, empirically we find that $\hat{\theta}$ and θ^* are almost identical to their counterparts with $w(\cdot)$ being the empirical distribution function of $\{X_i\}$. Moreover, we find that these estimators are not sensitive to the choice of V when the sample size is moderate and the censoring is not heavy.

For the second example, we consider a recent trial for treating the advanced AIDS patients, which was sponsored by the AIDS Clinical Trials Group (Henry et al., 1998). This multi-centre randomised, double-blind, placebo-controlled study was conducted from June 1993 to June 1996. Thirteen hundred and thirteen HIV-infected patients with CD4 counts ≤ 50 cells/mm^3 were randomise to one of four treatment groups. One of the major goals of the study is to examine the effect from the three-drug combination, AZT+ddI+Nevirapine, with respect to survival. For illustrating our methods, we let the first component of the covariate vector Z be the treatment indicator, which is one if the patient was treated by the triple therapy, otherwise, is 0. There were 330 patients assigned to the three-drug group and 983 patients assigned to either two-drug groups or the alternating drug groups. The censoring rate for the entire study cohort is about 60%. In Model (2·1), we also include the patient's baseline CD4

http://biostats.bepress.com/harvardbiostat/paper6

count and age in the covariate vector Z. Here, we let $\tau = 2.8$ years. Under the same setting for $w(t)$, $v(s)$, V and $\mathcal{M}$ as that in the previous example, the point estimates for λ_0 is 0.62 with the estimated standard error of 0.30. The point estimates for the regression coefficients of the treatment indicator, age and baseline CD4 count are 0.24, -0.02, and 0.52 with the estimated standard errors of 0.08, 0.004 and 0.08, respectively. In Figure 1, we present two sets of point and interval estimates of the survival function for a 37- year old patient with baseline CD4 count of 20. The plots on the left panel predict the survival function for this patient with the triple therapy, and those on the right panel predict the survival function with the two-drug or alternating drug therapy. For this type of patients, the three-drug combination treatment is a much better choice than the two-drug alternatives.

To compare the results based on our model (2·1) with those from the standard accelerated failure time model, that is, by setting $\lambda_0 = 0$, we consider a 25-year old patient with the baseline CD4 count of 45 and the triple therapy. In Figure 2, we present the estimated survival curves for this particular patient. The solid curve is based on the Box-Cox model and the dotted curve is from the accelerated failure time model. These two curves appear to be markedly different, for example, the latter one overly estimates the survival probabilities by more than 10% around Year 2.5.

5. REMARKS

Recently Foster et al. (2001) proposed an estimation procedure for the semi-parametric Box-Cox transformation model with completely observed data. Unfortunately their method cannot be generalised to the case when the censoring variable for the failure time may depend

Hosted by The Berkeley Electronic Press

on the covariate vector under a non-parametric setting.

Although the first stage of our estimation procedure is based on a Gehan-type estimator $\hat{\beta}_\lambda$ by minimising a U-process (2.2), one may replace this estimator by that obtained from any of the monotone estimating functions considered by Jin et al. (2003). Moreover, the new proposal is still valid even with a general weighted log-rank estimator for β_λ (Tsiatis, 1990; Wei et al., 1990), which can be approximated well by the iterative procedure studied by Jin et al. (2003) within a finite number of iterations.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper6

APPENDIX 1

Identifiability of Model (2·1) in the Presence of Censoring

Let $F_z(t)$ be the distribution function of T given $Z = z$, which is identifiable for $t \leq \tau_z$, where $P(X > \tau_z \mid Z = z) > 0$. Let e_0 be the $p \times 1$ vector of zeros and e_k be the $p \times 1$ vector of zeros except for the kth element being 1, $k = 1, ..., p$. Without loss of generality, we assume that e_0, e_1, ..., e_p are possible values of Z. Then, $F_{e_k}(t) = F\left\{g_{\lambda_0}(t) - \beta_0' e_k\right\}$, $k = 0, ..., p$ and $\frac{\partial}{\partial t}\left\{g_{\lambda_0}(t)\right\} = \frac{\partial}{\partial t} g_{\lambda_0}\left[F_{e_0}^{-1}\{F_{e_1}(t)\}\right]$. This implies that

$$\lambda_0 = \frac{\log \frac{d}{dt} F_{e_0}^{-1}\{F_{e_1}(t)\}}{\log t - \log F_{e_0}^{-1}\{F_{e_1}(t)\}} + 1, \quad \text{and} \quad g_{\lambda_0}(t) - \beta_0' e_k = g_{\lambda_0}\left[F_{e_0}^{-1}\{F_{e_k}(t)\}\right], \quad 1 \leq k \leq p.$$

Therefore, λ_0 and β_0 are identifiable. Furthermore, $F(y) = F_z\left\{g_{\lambda_0}^{-1}(y + \beta_0' z)\right\}$. It follows that $F(\cdot)$ is identifiable in the support of $g_{\lambda_0}(X) - \beta_0' Z$.

APPENDIX 2

Consistency of $\widehat{\lambda}$ and $\widehat{\beta}$

Suppose θ_0 lies in a compact set $\Omega_{\lambda_0} \times \Omega_{\beta_0}$. To show that $\widehat{\lambda}$ is a consistent estimate of λ_0, it suffices to show that $\widehat{Q}(\lambda)$ converges to a deterministic function $q(\lambda)$, uniformly in λ, and $q(\lambda)$ has a unique minimiser λ_0 (Newey & McFadden, 1994).

First, it follows from the empirical process theory (Pollard, 1990, Chapter 8) that $\widehat{\Lambda}(y, \theta)$ converges to a deterministic function $\Lambda(y, \theta)$ almost surely, uniformly in y and θ. This, coupled with the uniform convergence of $\widehat{\theta}_\lambda$, implies that $\widehat{\Lambda}(y, \widehat{\theta}_\lambda)$ converges, uniformly in λ, to $\Lambda(y, \theta_\lambda)$. Next, let $Y_i(t) = I(X_i \geq t)$, $A_i(t, \theta) = \int_0^t Y_i(u) d\Lambda\{g_\lambda(u) - \beta' Z_i, \theta\}$, $Q_\lambda(s, t) = n^{-1} \sum_{i=1}^n I(\beta_\lambda' Z_i \leq s)\{N_i(t) - A_i(t, \theta_\lambda)\}$, $Q(\lambda) = \int_{-\infty}^{\infty} \int_0^\tau Q_\lambda^2(s, t) dw(t) dv(s)$, and

Hosted by The Berkeley Electronic Press

$q(\lambda) = \int_{-\infty}^{\infty}\int_0^{\tau}[E\{Q_\lambda(s,t)\}]^2 dw(t)dv(s)$. Since $Q(\lambda)$ is a U-process in λ, it follows from Theorem 1 of Honore and Powell (1994) that $Q(\lambda) \to q(\lambda)$, uniformly in $\lambda \in \Omega_{\lambda_0}$. Furthermore, the convergence of $\widehat{\Lambda}(y, \widehat{\theta}_\lambda) \to \Lambda(y, \theta_\lambda)$ implies that

$$C_n \equiv \max_{1\le i\le n} \sup_{\lambda\in\Omega_{\lambda_0}, t\le\tau} \left|\widehat{A}_i(t,\widehat{\theta}_\lambda) - A_i(t,\theta_\lambda)\right| \to 0. \tag{A2·1}$$

With the uniform convergence of $n^{-1}\sum_{i=1}^n I(\beta' Z_i \le z)A_i(t,\theta_\lambda)$,

$$\left|\widehat{Q}(\lambda) - Q(\lambda)\right| \le n^{-1}\sum_{i=1}^n \{2 + 2\Lambda(g_\lambda(\tau) - \beta_\lambda' Z_i, \theta_\lambda)\}\, C_n \int_0^\tau dw(t)\int_{-\infty}^{\infty} dv(s) \; + o(1).$$

It follows from (A2·1) and the continuity of β_λ in λ that $\sup_{\lambda\in\Omega_{\lambda_0}}\left|\widehat{Q}(\lambda) - Q(\lambda)\right| \to 0$ almost surely. Therefore, $Q(\lambda) \to q(\lambda)$ almost surely, uniformly in λ.

It remains to show that $q(\lambda)$ has a unique minimiser at λ_0. Without loss of generality, we assume that $\lambda_0 = 1$, $\epsilon \ge 0$, $\beta_0' Z \ge 0$ and $f_\epsilon(x) > 0$ for any $x \ge 0$, where $f_e(\cdot)$ denotes the density function of the random variable e. Assume that there is another minimiser λ for $q(\lambda)$. Let ϵ_λ be the random variable with cumulative hazard function $\Lambda(\cdot, \theta_\lambda)$ and let $T_\lambda = g_\lambda^{-1}(\epsilon_\lambda + \beta_\lambda' Z)$. Since $q(\lambda) \ge 0$ and $q(\lambda_0) = 0$, then $q(\lambda) = 0$. This implies that for any $t \in [0,\tau]$ and s, we have

$$0 = E\left[\int_0^t I(X \ge u, \beta_\lambda' Z \le s)\left\{d\Lambda(g_{\lambda_0}(u) - \beta_0' Z, \theta_{\lambda_0}) - d\Lambda(g_\lambda(u) - \beta_\lambda' Z, \theta_\lambda)\right\}\right],$$

and hence

$$0 = E\left[I(\beta_\lambda' Z \le s)\int_0^t \left\{P(C \ge u, T \ge u \mid Z) f_{T_\lambda|Z}(u) - P(C \ge u, T_\lambda \ge u \mid Z) f_{T|Z}(u)\right\} du\right],$$

where $f_{T_\lambda|Z}$ and $f_{T|Z}$ are density functions of T_λ and T given Z, respectively. Let

$$R_\lambda(s,t) = \frac{\mathrm{pr}(T \ge t, C \ge t \mid \beta_\lambda' Z = s)}{P(T_\lambda \ge t \mid \beta_\lambda' Z = s)\mathrm{pr}(C \ge t \mid \beta_\lambda' Z = s)}. \tag{A2·2}$$

http://biostats.bepress.com/harvardbiostat/paper6

Then, $R_\lambda(s,t) = t^{1-\lambda}\,\frac{\partial}{\partial x}\mathrm{pr}(T \le x \mid C \ge t, \beta_\lambda' Z = s)\big|_{x=t} / f_{\epsilon_\lambda}(g_\lambda(t) - s)$. Note that the right-hand side of (A2·2) is positive and bounded for any s when $t = 0$. Moreover, when $\lambda = 1$, there exists s_0 such that $f_{\epsilon_\lambda}(g_\lambda(0) - s_0) > 0$. Hence, if $\lambda < 1,\ R_\lambda(s_0, 0^+) = 0$. On the other hand, if $\lambda > 1$, $R_\lambda(s, 0^+) \to \infty$, when s approaches to the upper bound of the support of $\beta_\lambda' Z$. This implies that $\lambda = \lambda_0 = 1$. The consistency of $\widehat{\beta}$ follows from the uniform consistency of $\widehat{\beta}_\lambda$ and the consistency of $\widehat{\lambda}$.

APPENDIX 3

Asymptotic Normality of $\widehat{\theta}$

To justify asymptotic normality of $n^{\frac{1}{2}}(\widehat{\theta} - \theta_0)$, we need to show that the minimand $\widehat{Q}(\lambda)$ has a "good" quadratic expansion around λ_0. To this end, let $\widehat{q}_\lambda(s,t) = n^{-1}\sum_{i=1}^n I(\widehat{\beta}_\lambda' Z_i \le s)\{A_i(t,\theta_0) - A_i(t,\theta_\lambda)\}$. We first show that $n^{\frac{1}{2}}\{\widehat{Q}_\lambda(s,t) - \widehat{q}_\lambda(s,t)\}$ is asymptotically equivalent to a sum of independent and identically distributed (iid) terms and converges weakly to a mean zero Gaussian process. Note that

$$
\begin{aligned}
n^{\frac{1}{2}}\left\{\widehat{Q}_\lambda(s,t) - \widehat{q}_\lambda(s,t)\right\} &= n^{-\frac{1}{2}}\sum_{i=1}^n I(\widehat{\beta}_\lambda' Z_i \le s) M_i(t) + o_p(1) \\
&- \int_{-\infty}^{\infty} n^{-\frac{1}{2}} \sum_{i=1}^n \Bigg[I\{\widehat{\beta}_\lambda' Z_i \le s, g_\lambda(0) \le y + \widehat{\beta}_\lambda' Z_i \le g_\lambda(t)\} Y_i \{g_\lambda^{-1}(y + \widehat{\beta}_\lambda' Z_i)\} \\
&\qquad - I\{\widehat{\beta}_\lambda' Z_i \le s, g_\lambda(0) \le y + \beta_\lambda' Z_i \le g_\lambda(t)\} Y_i \{g_\lambda^{-1}(y + \beta_\lambda' Z_i)\} \Bigg] d\widehat{\Lambda}(y, \widehat{\theta}_\lambda) \qquad \text{(A3·1)} \\
&- \int_{-\infty}^{\infty} n^{-1} \sum_{i=1}^n I\{\widehat{\beta}_\lambda' Z_i \le s, g_\lambda(0) \le y + \beta_\lambda' Z_i \le g_\lambda(t)\} Y_i \left\{g_\lambda^{-1}(y + \beta_\lambda' Z_i)\right\} d\mathcal{L}(y,\lambda), \qquad \text{(A3·2)}
\end{aligned}
$$

where $\mathcal{L}(y,\lambda) = n^{\frac{1}{2}}\left\{\widehat{\Lambda}(y,\widehat{\theta}_\lambda) - \Lambda(y,\theta_\lambda)\right\}$. Now, the process $\mathcal{L}(y,\lambda)$ can be written as

$$
n^{\frac{1}{2}}\{\widehat{\Lambda}(y,\widehat{\theta}_\lambda) - \widehat{\Lambda}(y,\theta_\lambda)\} + n^{\frac{1}{2}}\{\widehat{\Lambda}(y,\theta_\lambda) - \Lambda(y,\theta_\lambda)\}. \qquad \text{(A3·3)}
$$

Hosted by The Berkeley Electronic Press

It follows from the standard empirical process theory (Pollard, 1990, Chapter 10) that the first term of (A3·3) $\approx n^{\frac{1}{2}}\left\{\Lambda(y,\widehat{\theta}_\lambda)-\Lambda(y,\theta_\lambda)\right\}$, which is asymptotically equivalent to

$$n^{\frac{1}{2}}(\widehat{\beta}_\lambda-\beta_\lambda)'\frac{\partial\Lambda(y,\theta_\lambda)}{\partial\beta_\lambda}. \tag{A3·4}$$

Since $\widehat{\beta}_\lambda$ is obtained by minimising a U-process with respect to β, using the argument similar to that in Jin et al (2003), one can show that

$$n^{\frac{1}{2}}(\widehat{\beta}_\lambda-\beta_\lambda)\approx n^{-\frac{1}{2}}\sum_{i=1}^{n}\eta_i(\lambda), \tag{A3·5}$$

where $\eta_i(\lambda)$, $i=1,...,n$, are iid terms. By the martingale central limit theorem, the second term in (A3·3) is asymptotically equivalent to a sum of iid terms. This implies that (A3·2) can be approximated by a sum of iid terms. Furthermore, the integrand of (A3.1) can be approximated by $n^{1/2}(\hat{\beta}_\lambda-\beta_\lambda)'$ multiplying by a deterministic vector, and hence asymptotically (A3·1) is equivalent to a sum of iid elements. As a consequence, $n^{\frac{1}{2}}\{\widehat{Q}_\lambda(s,t)-\widehat{q}_\lambda(s,t)\}$ converges weakly to a mean zero Gaussian process. Moreover for $|\lambda-\lambda_0|=o(1)$,

$$n^{\frac{1}{2}}\left\{\widehat{Q}_\lambda(s,t)-\widehat{q}_\lambda(s,t)\right\}-n^{\frac{1}{2}}\left\{\widehat{Q}_{\lambda_0}(s,t)-\widehat{q}_{\lambda_0}(s,t)\right\}=o_p(1)$$

and $n^{\frac{1}{2}}(\hat{q}_\lambda(s,t)-q_\lambda(s,t))$ converges weakly to a Gaussian process indexed by (λ,t,s), where $q_\lambda(s,t)=E[I(\beta_\lambda'Z_i\le s)\{A_i(t,\theta_0)-A_i(t,\theta_\lambda)\}]$. It follows that for $|\lambda-\lambda_0|=o(1)$,

$$\begin{aligned}\widehat{Q}(\lambda)-\widehat{Q}(\lambda_0) \;=\; & (\lambda-\lambda_0)\int_{-\infty}^{\infty}\int_0^t 2\widehat{Q}_{\lambda_0}(s,t)\dot{q}_{\lambda_0}(s,t)dw(t)dv(s)+\\ & (\lambda-\lambda_0)^2\int_{-\infty}^{\infty}\int_0^t \dot{q}_{\lambda_0}(s,t)^2dw(t)dv(s)+o_p(n^{-1}+|\lambda-\lambda_0|^2),\end{aligned} \tag{A3·6}$$

where $\dot{q}_\lambda(s,t)=\partial q_\lambda(s,t)/\partial\lambda$. Similar to the arguments used for the expansion of $n^{\frac{1}{2}}\{\widehat{Q}_\lambda(s,t)-\widehat{q}_\lambda(s,t)\}$, one can show that $n^{\frac{1}{2}}\widehat{Q}_{\lambda_0}(s,t)\approx n^{-\frac{1}{2}}\sum_{i=1}^{n}\mathcal{Q}_i(s,t)$, where $\mathcal{Q}_i(s,t)$, $i=1,...,n$, are iid

http://biostats.bepress.com/harvardbiostat/paper6

random processes. Now, let $\mathfrak{a} = \int_{-\infty}^{\infty}\int_0^t \dot{q}_{\lambda_0}(s,t)^2 dw(t)dv(s)$, since $\widehat{Q}(\widehat{\lambda}) \leq \widehat{Q}(\lambda_0)$, $\widehat{Q}_{\lambda_0}(s,t) = O_p(n^{-\frac{1}{2}})$, it follows from (A3·6) that $\mathfrak{a}(\widehat{\lambda}-\lambda_0+O_p(n^{-\frac{1}{2}}))^2+O_p(n^{-1}) \leq o_p(|\widehat{\lambda}-\lambda_0|^2)$. This implies that $|\widehat{\lambda}-\lambda_0| = O_p(n^{-\frac{1}{2}})$ and

$$n^{\frac{1}{2}}(\widehat{\lambda}-\lambda_0) = \mathfrak{a}^{-1}\int_{-\infty}^{\infty}\int_0^t n^{\frac{1}{2}}\widehat{Q}_{\lambda_0}(s,t)\dot{q}_{\lambda_0}(s,t)dw(t)dv(s) + o_p(1). \tag{A3·7}$$

It follows from the equicontinuity of $n^{\frac{1}{2}}(\widehat{\beta}_\lambda - \beta_\lambda)$ that

$$n^{\frac{1}{2}}(\widehat{\beta}-\beta_0) \approx n^{\frac{1}{2}}(\widehat{\beta}_{\lambda_0}-\beta_{\lambda_0}) + n^{\frac{1}{2}}(\beta_{\widehat{\lambda}}-\beta_{\lambda_0}) \approx n^{\frac{1}{2}}(\widehat{\beta}_{\lambda_0}-\beta_{\lambda_0}) + n^{\frac{1}{2}}(\widehat{\lambda}-\lambda_0)\dot{\beta}_{\lambda_0}, \tag{A3·8}$$

where $\dot{\beta}_\lambda = d\beta_\lambda/d\lambda$. This, coupled with (A3·5), (A3·7) and a multivariate central limit theorem, implies that $n^{\frac{1}{2}}(\widehat{\theta}-\theta_0)$ can be expressed as a sum of iid random vectors, which converges in distribution to a zero-mean multivariate normal.

APPENDIX 4

Justification for the Resampling Method

First, consider the following unconditional version of $L^*(\theta)$, and $Q^*(\theta)$:

$$\widehat{L}^*(\theta) \equiv n^{-2}\sum_{i=1}^n\sum_{j=1}^n \Delta_i V_i \left\{\varepsilon_i(\theta)-\varepsilon_j(\theta)\right\} I\left\{\varepsilon_i(\theta)\geq\varepsilon_j(\theta)\right\},$$

$$\widehat{Q}^*(\lambda) = \int_{-\infty}^{\infty}\int_0^{\tau}\left\{n^{-1}\sum_{i=1}^n I(Z_i'\widehat{\beta}_\lambda^* \leq s)\widehat{M}_i^*(t,\widehat{\theta}_\lambda^*)V_i\right\}^2 dw(t)dv(s)$$

where $\widehat{\beta}_\lambda^* = \text{argmin}_\beta\, \hat{L}_G^*(\beta,\lambda)$, $\widehat{\theta}_\lambda^* = (\lambda, \widehat{\beta}_\lambda^*)$ and

$$\widehat{M}_i^*(t,\theta) = N_i(t) - \int_0^t \frac{Y_i(u)d\sum_{j=1}^n I\{\varepsilon_j(\theta)\leq g_\lambda(u)-\beta' Z_i\}\Delta_j V_j}{\sum_{j=1}^n I\{\varepsilon_j(\theta)\geq g_\lambda(u)-\beta' Z_i\}V_j}.$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Let $\widehat{\lambda}^* = \text{argmin}_\lambda \widehat{Q}^*(\lambda)$. It follows from the arguments in Appendix 3, that $\widehat{\lambda}^*$ and $\widehat{\beta}^* = \widehat{\beta}^*_{\widehat{\lambda}^*}$ are consistent and asymptotically normal. More specifically,

$$n^{\frac{1}{2}} \begin{pmatrix} \widehat{\lambda}^* - \lambda_0 \\ \widehat{\beta}^* - \beta_0 \end{pmatrix} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \begin{pmatrix} P_i^* V_i \\ \eta_i(\lambda_0) V_i + \dot{\beta}_{\lambda_0} P_i^* V_i \end{pmatrix} + o_{p^*}(1)$$

where

$$P_i^* = \mathfrak{a}^{-1} \int_{-\infty}^{\infty} \int_0^{\tau} \dot{q}_{\lambda_0}(s,t) \mathcal{Q}_i(s,t) dw(t) dv(s),$$

and $o_{p^*}(1)$ is with respect to the product probability measure generated by $\mathcal{D} = \{(X_i, \Delta_i, Z_i), i = 1, ..., n\}$ and $\{V_i, i = 1, ..., n\}$. Therefore,

$$n^{\frac{1}{2}} \begin{pmatrix} \widehat{\lambda}^* - \widehat{\lambda} \\ \widehat{\beta}^* - \widehat{\beta} \end{pmatrix} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \begin{pmatrix} P_i^* \\ \eta_i(\lambda_0) + \dot{\beta}_{\lambda_0} P_i^* \end{pmatrix} (V_i - 1) + o_{p^*}(1) \qquad \text{(A4·1)}$$

Conditional on the data, it follows from Lindeberg-Feller Central Limit Theorem that the conditional distribution of $n^{\frac{1}{2}}(\widehat{\lambda}^* - \widehat{\lambda}, \widehat{\beta}^* - \widehat{\beta})$ converges to a multivariate normal with mean $\mathbf{0}$ and covariance Σ_θ. This implies that, for any $\varepsilon > 0$, there exists a N_0 such that when $n > N_0$, the probability, with respect to $\mathcal{D}$, of the event

$$\sup_{\mathbf{u} \in \mathbb{R}^{p+1}} \left| \text{pr} \left(n^{\frac{1}{2}} \begin{pmatrix} \widehat{\lambda}^* - \widehat{\lambda} \\ \widehat{\beta}^* - \widehat{\beta} \end{pmatrix} \leq \mathbf{u} \,\middle|\, \mathcal{D} \right) - \text{pr} \left(n^{\frac{1}{2}} \begin{pmatrix} \widehat{\lambda} - \lambda_0 \\ \widehat{\beta} - \beta_0 \end{pmatrix} \leq \mathbf{u} \right) \right| < \varepsilon,$$

is at least $1 - \varepsilon$.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper6

References

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. (1964). An analysis of transformations. *J. R. Statist. Soc.* B **26**, 211–52.

BOX, G. E. P. & COX, D. R. (1982). An analysis of transformations revisited, rebutted. *J. Am. Statist. Assoc.* **77**, 209–10.

CARROLL, R. J. (1982). Tests for regression parameters in power transformation models. *Scandinavian Journal of Statistics* **9**, 217–222.

CARROLL, R. J. & RUPPERT, D. (1985). Transformations in regression: A robust analysis. *Technometrics* **27**, 1–12.

CARROLL, R. J. & RUPPERT, D. (1987). Diagnostics and robust estimation when transforming the regression model and the response. *Technometrics* **29**, 287–299.

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

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting processes and survival analysis*. John Wiley and Sons.

FOSTER, A. M., TIAN, L. & WEI, L. J. (2001). Estimation for the Box-Cox transformation model without assuming parametric error distribution. *J. Am. Statist. Assoc.* **96**, 1097–1101.

HAN, A. K. (1987). A non-parametric analysis of transformations. *J. of Econ.* **35**, 191–209.

HENRY, K., ERICE, A., TIERNEY, C., BALFOUR, H., FISCHL, M., KMACK, A., LIOU, S., KENTON, A., HIRSCH, M., PHAIR, J., MARTINEZ, A. & JO, K. (1998). A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced aids. *Journal of Acquired Immune Deficiency Syndromes* **19**, 339–49.

HONERE, B. E. & POWELL, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *J. of Econ.* **64**, 241–78.

JIN, Z. Z., LIN, D. Y., WEI, L. J. & YING, Z. L. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–53.

KALBFLEISCH, J. D. & PRENTICE, R. L. (2002). *The statistical analysis of failure time data*. John

COBRA
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Wiley and Sons.

LIN, D. Y., WEI, L. J. & YING, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572.

LIN, D. Y., WEI, L. J. & YING, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58**, 1–12.

NEWEY, 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.

NEWEY, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica* **58**, 809–837.

PARK, Y. & WEI, L. J. (2003). Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika* **90**, 717–23.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications.* Hayward, CA: Institute of Mathematical Statistics.

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303–28.

ROBINSON, P. M. (1991). Best nonlinear three-stage least squares estimation of certain econometric models (STMA V33 0996). *Econometrica* **59**, 755–786.

TAYLOR, J. M. G. (1987). Using a generalized mean as a measure of location. *Biometrical Journal. Journal of Mathematical Methods in Biosciences. [Continues: Biometrische Zeitschrift. Zeitschrift fr mathematische Methoden in den Biowissenschaften]* **29**, 731–738.

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–372.

WANG, N. & RUPPERT, D. (1995). Nonparametric estimation of the transformation in the transform-both-sides regression model. *J. Am. Statist. Assoc.* **90**, 522–534.

WEI, L. J., YING, Z. & LIN, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper6

Table 1: Estimated Covariate Effects on Failure Time with Mayo Biliary Cirrhosis Data.

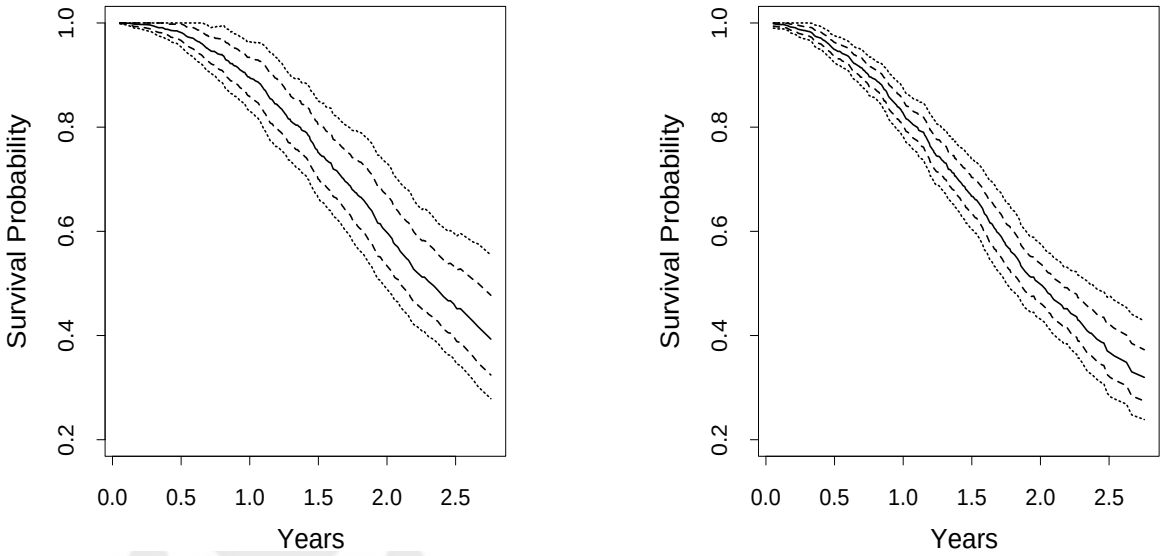
	β_{Age}	β_{Oedema}	$\beta_{\text{Bilirubin}}$	β_{Albumin}	β_{Protime}
Box-Cox Model	-0.028	-0.926	-0.618	1.630	-2.877
	$(0.007)^*$	(0.277)	(0.106)	(0.618)	(0.920)
Accelerated Failure Time Model	-0.026	-0.878	-0.589	1.591	-2.768
	(0.006)	(0.277)	(0.070)	(0.535)	(0.908)

* Estimated standard error

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Figure 1: Survival probabilities for a 37-year-old patient with baseline CD4 count of 20 treated by the triple therapy or two-drug alternatives. (—— : point estimates; − − −: pointwise 95% confidence intervals; · · · · · · : simultaneous 95% confidence intervals).



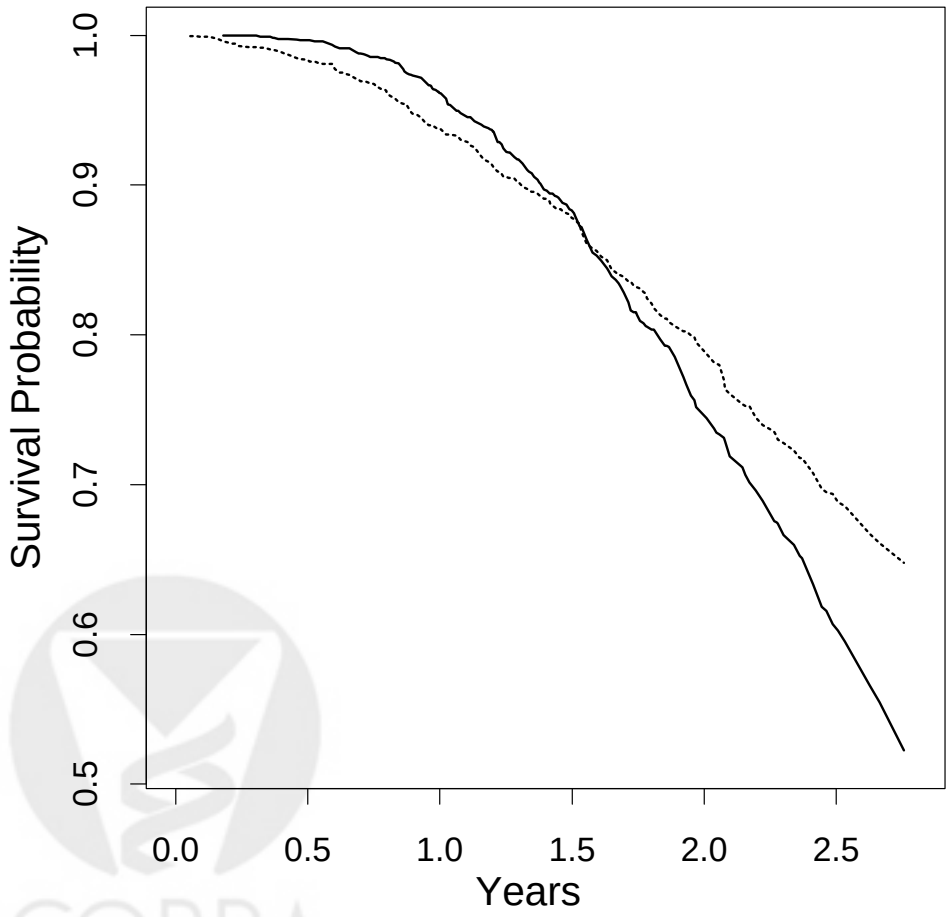
(a) Triple Therapy

(b) Two drug alternatives

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

http://biostats.bepress.com/harvardbiostat/paper6

Figure 2: Predicted survival probabilities for a 25-year-old patient with baseline CD4 count of 45 on three drug combination based on the Box-Cox model (solid curve) and on the accelerated failure time model (dotted curve).



COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press