Harvard University
Harvard University Biostatistics Working Paper Series

Year 2005 *Paper* 20

Robust Inferences For Covariate Effects On Survival Time With Censored Linear Regression Models

Larry Leon* Tianxi Cai†
L. J. Wei‡

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

Robust Inferences For Covariate Effects On Survival Time With Censored Linear Regression Models

Larry Leon, Tianxi Cai, and L. J. Wei

Abstract

Various inference procedures for linear regression models with censored failure times have been studied extensively. Recent developments on efficient algorithms to implement these procedures enhance the practical usage of such models in survival analysis. In this article, we present robust inferences for certain covariate effects on the failure time in the presence of "nuisance" confounders under a semiparametric, partial linear regression setting. Specifically, the estimation procedures for the regression coefficients of interest are derived from a working linear model and are valid even when the function of the confounders in the model is not correctly specified. The new proposals are illustrated with two examples and their validity for cases with practical sample sizes is demonstrated via a simulation study.

ROBUST INFERENCES FOR COVARIATE EFFECTS ON SURVIVAL TIME WITH CENSORED LINEAR REGRESSION MODELS

Larry F. León, Tianxi Cai and L.J. Wei
Department of Biostatistics
Harvard University
677 Huntington Ave.
Boston, MA 02115
lleon@hsph.harvard.edu, tcai@hsph.harvard.edu, wei@hsph.harvard.edu

Summary

Various inference procedures for linear regression models with censored failure times have been studied extensively. Recent developments on efficient algorithms to implement these procedures enhance the practical usage of such models in survival analysis. In this article, we present robust inferences for certain covariate effects on the failure time in the presence of "nuisance" confounders under a semiparametric, partial linear regression setting. Specifically, the estimation procedures for the regression coefficients of interest are derived from a working linear model and are valid even when the function of the confounders in the model is not correctly specified. The new proposals are illustrated with two examples and their validity for cases with practical sample sizes is demonstrated via a simulation study.

Some key words: Censored linear regression; Partial linear model; Resampling method; Rank estimation.

1. INTRODUCTION

Suppose that we are interested in making inferences about the covariate effect from a $p \times 1$ vector X on the failure time in the presence of a continuous confounder Z. When the failure time is subject to censoring, the Cox model (Cox, 1972) is commonly used for estimating the covariate effect. Alternatively, one may assume that a transformation T of the failure time is linearly related to X and Z. Rank-based inferences for such censored linear regression models have been studied extensively, for example, by Tsiatis (1990), Ritov (1990), Wei, Ying & Lin (1990), and Kalbfleisch & Prentice (2002, Ch. 7). Recent developments on efficient algorithms for implementing these methods by Jin et al. (2003) and Tian et al. (2004) enhance the practical usage of linear regression models in survival analysis.

When the working Cox model is misspecified, the standard inference procedures for the covariate effects generally are not valid (Gail, Wieand & Piantadosi, 1984; Lagakos & Schoenfeld, 1984; Morgan, 1986; Struthers & Kalbfleisch, 1986; DiRienzo & Lagakos, 2001b). Lin & Wei (1989) proposed robust variance estimates of the maximum partial likelihood estimators for the regression parameters. Recently, DiRienzo & Lagakos (2001a) provided bias correction for score tests derived from misspecified proportional hazards models.

In this paper, we are interested in exploring if valid inferences for the covariate effects can be made when the censored linear regression model may not be correctly specified. Specifically, let the true model for T and its covariate vector X and confounder Z be

$$T = \alpha_0 + \beta_0' X + g(Z) + \epsilon, \tag{1.1}$$

where α_0 and β_0 are unknown parameters, $g(\cdot)$ is a completely unspecified function, and the unknown distribution function of the error term ϵ has zero mean and is free of X and Z. We are mainly interested in estimating the vector of the regression parameters β_0. Without censoring, the partial linear model (1.1) has been studied extensively (Wahba, 1984; Green et al., 1985; Engle et al., 1986; Heckman, 1986; Rice, 1986; Shiau et al., 1986; Robinson, 1988; Speckman, 1988; Chen, 1988; Chen & Shiau, 1991; Cuzick, 1992; Bhattacharya & Zhao, 1997; Hong & Cheng, 1999). Recently, Qin & Jing (2001) used (1.1) to fit the so-called synthetic data created from censored observations (Koul et al., 1981; Leurgans, 1987). Their procedure, however, is only valid for a rather rare case that the support of the censoring variable is at least as large as that of the failure time. Moreover, in the presence of censoring, the nonparametric function $g(\cdot)$ may not be estimated well in practice, which may affect the performance of the estimate for β_0.

In this paper, we introduce a *working* model for T and its covariates and present valid inference procedures for β_0 without involving estimation of the unknown function $g(\cdot)$ in (1.1). To be specific, we consider the following working linear model

$$T = a + b'X + c' \ \hat{v}(Z) + e, \tag{1.2}$$

where $\hat{v}(Z)$ is a nonparametric or parametric estimate of $v(Z)$, the expected value of X given Z, a is the intercept, b and c are $p \times 1$ vectors of parameters and e is the error term. When there are no censored observations, the least squares and most rank-based estimates for b are consistent with respect to β_0 in the true model (1.1) (Cook, 1993). Unfortunately, in the presence of censoring, it is not clear whether the estimators based on the weighted logrank test statistics proposed by Tsiatis (1990), Ritov (1990) and Wei et al. (1990) are consistent.

Under mild regularity conditions, we show that a class of estimators $\hat{b}$ for b in (1.2) derived from the censored quantile regression (Ying et al., 1995) is consistent with respect to β_0. We then demonstrate that the distribution of $\hat{b}$ can be approximated by a normal whose variance can be estimated well via a novel resampling method. The new proposal is valid even when the support of the censoring variable is shorter than that of the failure time. All the procedures are illustrated with two examples and their validity for cases with practical sample sizes is demonstrated via a simulation study.

2. CONSISTENT ESTIMATORS FOR REGRESSION COEFFICIENTS IN A PARTIAL LINEAR MODEL

Suppose that T may be censored by C. Let $Y = \min(T, C)$ and $\Delta = I(Y = T)$, where $I(\cdot)$ is the indicator function. Also, let $(T_i, C_i, X_i, Z_i), i = 1, \ldots, n$, be n independent copies of (T, C, X, Z), where X and Z are bounded. Assume that censoring is independent of T and Z. Let the common survival function of C be denoted by $G(\cdot)$ and $\hat{G}(\cdot)$ be its corresponding Kaplan-Meier estimate. The *working* model (1.2) can be expressed as a τ^{th} quantile regression of T, $0 < \tau < 1$. That is, conditioning on X and $\hat{v}(Z)$, the $100\tau^{\text{th}}$ percentile of T is $\{a_\tau + b'_\tau X + c'_\tau\ \hat{v}(Z)\}$. To estimate $\eta_\tau = (a_\tau, b'_\tau, c'_\tau)'$, we consider the following estimating function

$$Q_\tau(\eta) = \sum_{i=1}^{n} \hat{W}_i \left\{ \frac{I(Y_i \geq \eta' \hat{W}_i)}{\hat{G}(\eta' \hat{W}_i)} - (1 - \tau) \right\}, \tag{2.1}$$

where $\hat{W}'_i = (1, X'_i, \hat{v}(Z_i)')$. Note that (2.1) is a generalization of the estimating function for the censored median regression proposed by Ying et al. (1995). Let $\hat{\eta}'_\tau = (\hat{a}_\tau, \hat{b}'_\tau, \hat{c}'_\tau)$ be a solution to the equation $Q_\tau(\eta) = 0$. Here, we restrict our search for $\hat{\eta}_\tau$ in the set of η such that $\hat{G}(\eta' \hat{W}_i) > 0, i = 1, \ldots, n$. It is important to note that due to censoring, $\hat{\eta}_\tau$ may not exist for relatively large τ. For (2.1), one may use a parametric estimate $\hat{v}(\cdot)$, but in this article, we use a nonparametric, kernel estimate $\hat{v}(\cdot)$ for $v(\cdot)$. That is,

$$\hat{v}(z) = \frac{\sum_{i=1}^{n} K_h(Z_i - z) X_i}{\sum_{i=1}^{n} K_h(Z_i - z)},$$

where $K_h(\cdot) = h^{-1} K(\cdot/h)$, $K(\cdot)$ is a symmetric density function with $\int x^2 K(x) dx < \infty, h \to 0$, and $(\log n)^{-1} nh \to \infty$, as $n \to \infty$.

Let t_0 be a pre-specified time point such that $\text{pr}(Y > t_0) > 0$. The Kaplan-Meier estimate $\hat{G}(t)$ converges uniformly to $G(t)$ for $t \leq t_0$ (Csörgő & Horváth, 1983, p. 418). Moreover, under certain regularity conditions $\hat{v}(z)$ is uniformly convergent to

$v(z)$ for z in any compact set of the support of Z (Härdle et al., 1988). It follows that uniformly in η, $n^{-1}Q_\tau(\eta) =$

$$n^{-1}\sum_{i=1}^{n} W_i \left\{ \frac{I(Y_i - \eta' W_i \geq 0)}{G(\eta' W_i)} - (1-\tau) \right\} + o(1)$$

which converges uniformly to $E[W\{\text{pr}(T - \eta' W \geq 0) - (1-\tau)\}]$, for η such that $\text{pr}(G(\eta' W) > 0) = 1$, where $W' = (1, X', v(Z)')$. Now, let $\eta_{\tau 0} = (a_{\tau 0}, b'_{\tau 0}, c'_{\tau 0})'$ be the unique solution to the equation

$$E[W\{\text{pr}(T - \eta' W \geq 0 \mid W) - (1-\tau)\}] = 0. \tag{2.2}$$

Suppose that $\text{pr}(\eta'_{\tau 0} W \leq t_0) = 1$. This condition is required for establishing the consistency of the quantile regression parameters even when (1.2) is the true model (Ying et al., 1995).

It follows that $\hat{\eta}_\tau$ converges to $\eta_{\tau 0}$, almost surely. Since the solution to (2.2) is unique, if we can show that there exists a solution such that $b_{\tau 0} = \beta_0$, then $\hat{b}_\tau$ is a consistent estimator for β_0. To this end, we rewrite the left hand side of (2.2) as

$$E\left[\begin{pmatrix} 1 \\ X \\ v(Z) \end{pmatrix} \{F(a_{\tau 0} - \alpha_0 + (b_{\tau 0} - \beta_0)' X + c_{\tau 0}\, v(Z) - g(Z)) - (1-\tau)\}\right], \tag{2.3}$$

where $F(\cdot)$ is the distribution function of the error term ϵ in (1.1). Now, if we let $b_{\tau 0} = \beta_0$ in (2.3), the second and third components of (2.3) are identical. Furthermore, since

$$E\left[\begin{pmatrix} 1 \\ v(Z) \end{pmatrix} \{F(a - \alpha_0 + c' v(Z) - g(Z)) - (1-\tau)\}\right] \tag{2.4}$$

is a monotone function of a and c, there is a unique solution to the equation $(2.4) = 0$. This implies that $(a_{\tau 0}, \beta'_0, c'_{\tau 0})'$ is indeed a unique solution to (2.2).

Let $0 < \tau_1 < \cdots < \tau_K < 1$, such that $\text{pr}(\eta'_{\tau_k 0} W < t_0) = 1, k = 1, \ldots, K$. It follows that any linear combination $\sum_{k=1}^{K} e_k \hat{b}_{\tau_k}$, where $\sum_k e_k = 1$, is consistent with respect to β_0. In the next section, we derive the optimal linear combination $\hat{b}$ of $\hat{b}_{\tau_k}, k = 1, \ldots, K$, which has the smallest asymptotic variance among all the aforementioned linear combinations.

3. APPROXIMATION TO THE DISTRIBUTION OF $\hat{b}$

To obtain the distribution of a linear combination of $\hat{b}_\tau, \tau \in I$, where $I = \{\tau_k, k = 1, \ldots, K\}$, we need to show that for large n, the joint distribution of $\{\hat{\eta}_\tau, \tau \in I\}$ can be approximated by a normal distribution. To this end, first we demonstrate that the estimating function $Q_\tau(\eta)$ is approximately linear in a small neighborhood of $\eta_{\tau 0}$ and $n^{-1/2} Q_\tau(\eta_{\tau 0})$ is asymptotically normal. Now, let $S_W(t) = \text{pr}(T \geq t | W)$. We assume that the density functions of the error ϵ in (1.1) and the covariate Z

are continuously differentiable with bounded derivatives. Furthermore, we let the smoothing parameter $h = o_p(n^{-1/4})$ for $\hat{v}(\cdot)$. Consider the process

$$\psi(\eta) = \sum_{i=1}^{n} \hat{W}_i \left\{ \frac{I(Y_i - \eta'\hat{W}_i \geq 0)}{\hat{G}(\eta'\hat{W}_i)} - S_{W_i}(\eta'W_i) \right\}. \quad (3.1)$$

Note that

$$Q_\tau(\eta) = \psi(\eta) + \sum_{i=1}^{n} \hat{W}_i (S_{W_i}(\eta'W_i) - (1-\tau)). \quad (3.2)$$

In the Appendix, we show that the process $n^{-1/2}\psi(\eta)$ is asymptotically equivalent to a standardized sum of independent and identically distributed random processes of η, and converges weakly to a zero mean Gaussian process in η, for all η such that $\text{pr}(G(\eta'W) > 0) = 1$. Since the second term on the right hand side of (3.2) is a smooth function of η, it follows that $Q_\tau(\eta)$ is locally linear around $\eta_{\tau 0}$. Therefore, there exists a deterministic matrix A_τ such that $n^{1/2}(\hat{\eta}_\tau - \eta_{\tau 0})$ is asymptotically equivalent to $A_\tau n^{-1/2} Q_\tau(\eta_{\tau 0})$.

Furthermore, from (6.11)-(6.14) in the Appendix, $\{n^{-1/2}Q_\tau(\eta_{\tau 0}), \tau \in I\}$ converges weakly to a mean-zero normal. It follows that $\{n^{1/2}(\hat{\eta}_\tau - \eta_{\tau 0}), \tau \in I\}$ converges weakly to a mean-zero normal. This implies that the limiting distribution of $\{\hat{b}_\tau, \tau \in I\}$ is normal with mean 0 and a $K \times K$ covariance matrix Ω. If one can obtain a consistent estimate $\hat{\Omega}$ for Ω, then the optimal linear combination of $\{\hat{b}_\tau\}$ for estimating β_0 is

$$\hat{b} = \frac{d'\hat{\Omega}^{-1}(\hat{b}_{\tau_1}, \ldots, \hat{b}_{\tau_K})'}{d'\hat{\Omega}^{-1}d}, \quad (3.3)$$

where $d = (1, \ldots, 1)'$, a $K \times 1$ vector. An estimate for the variance of $\hat{b}$ is $(d'\hat{\Omega}^{-1}d)^{-1}$. The covariance matrix Ω, however, cannot be estimated well directly. Here, we show how to use a perturbation method to approximate the joint distribution of $\{\hat{b}_\tau, \tau \in I\}$.

Based on the asymptotic expansion for $Q_\tau(\eta)$ given in (6.11)-(6.14), it is not difficult to show that $n^{-1/2}Q_\tau(\eta_{\tau 0})$ is asymptotically equivalent to $B_\tau =$

$$n^{-1/2} \sum_{i=1}^{n} W_i \left\{ \frac{I(Y_i - \eta'_{\tau 0}W_i \geq 0)}{\hat{G}(\eta'_{\tau 0}W_i)} - \frac{I(Y_i - \eta'_{\tau 0}W_i \geq 0)}{G(\eta'_{\tau 0}W_i)} \right\} \quad (3.4)$$

$$+ n^{-1/2} \sum_{i=1}^{n} W_i \left\{ \frac{I(Y_i - \eta'_{\tau 0}W_i \geq 0)}{G(\eta'_{\tau 0}W_i)} - (1-\tau) \right\} \quad (3.5)$$

$$+ n^{-1/2} \sum_{i=1}^{n} W_i \left\{ \frac{I(Y_i - \eta'_{\tau 0}\hat{W}_i \geq 0)}{G(\eta'_{\tau 0}\hat{W}_i)} - \frac{I(Y_i - \eta'_{\tau 0}W_i \geq 0)}{G(\eta'_{\tau 0}W_i)} \right\} \quad (3.6)$$

$$+ n^{-1/2} \sum_{i=1}^{n} (\hat{W}_i - W_i) \left\{ \frac{I(Y_i - \eta'_{\tau 0}W_i \geq 0)}{G(\eta'_{\tau 0}W_i)} - (1-\tau) \right\}. \quad (3.7)$$

Conditional on the data, we perturb each term of the summations in (3.4)-(3.7) properly to generate a random vector $\{B^*_\tau, \tau \in I\}$, which has the same asymptotic distribution as that of $\{B_\tau, \tau \in I\}$. This technique has been utilized successfully, for example, to approximate the distributions of complex empirical processes in survival analysis (Lin et al., 1993; Lin et al., 1994; Goldwasser et al., 2004). Specifically, let $\{V_i, i = 1, \ldots, n\}$ be a random sample from a population with mean 0 and variance one, which is independent of the data $\{(Y_i, \Delta_i, X_i, Z_i)\}$. First, in (3.4) we replace $\hat{G}(\cdot)$ by its perturbed version $G^*(\cdot)$ and $G(\cdot)$ by its observed Kaplan-Meier estimate $\hat{G}(\cdot)$, where

$$G^*(t) = \hat{G}(t)\left[1 - \sum_{i=1}^{n}\left\{\int_{-\infty}^{t} \frac{dM_i(s)}{\sum_j I(Y_j \geq s)}\right\} V_i\right], \tag{3.8}$$

$M_i(s) = I(Y_i \leq s, \Delta_i = 0) - \int_{-\infty}^{s} I(Y_i \geq u) d\hat{\Lambda}(u)$, and $\hat{\Lambda}(\cdot)$ is the standard Nelson-Aalen estimate for the cumulative hazard function of the censoring variable C. We then replace W_i by $\hat{W}_i$. For (3.5), we multiply the i^{th} term in the summation by $V_i, i = 1, \ldots, n$ and then replace W_i by $\hat{W}_i$. For (3.6) and (3.7), we replace $\hat{W}_i$ by W_i^* and W_i by the observed $\hat{W}_i$, where $W_i^* = (1, X_i', v^*(Z_i))'$ and

$$v^*(z) = \hat{v}(z) + n^{-1}\sum_i \frac{K_h(Z_i - z)(X_i - \hat{v}(Z_i))V_i}{n^{-1}\sum_j K_h(Z_j - z)}. \tag{3.9}$$

We then replace all the theoretical quantities G and $\eta_{\tau 0}$ in (3.4)-(3.7) by their observed empirical counterparts $\hat{G}$ and $\hat{\eta}_\tau$. This results in $B_\tau^* =$

$$n^{-1/2}\sum_{i=1}^{n} \hat{W}_i \left\{\frac{I(Y_i - \hat{\eta}_\tau'\hat{W}_i \geq 0)}{G^*(\hat{\eta}_\tau'\hat{W}_i)} - \frac{I(Y_i - \hat{\eta}_\tau'\hat{W}_i \geq 0)}{\hat{G}(\hat{\eta}_\tau'\hat{W}_i)}\right\} \tag{3.10}$$

$$+n^{-1/2}\sum_{i=1}^{n} \hat{W}_i \left\{\frac{I(Y_i - \hat{\eta}_\tau'\hat{W}_i \geq 0)}{\hat{G}(\hat{\eta}_\tau'\hat{W}_i)} - (1-\tau)\right\} V_i \tag{3.11}$$

$$+n^{-1/2}\sum_{i=1}^{n} \hat{W}_i \left\{\frac{I(Y_i - \hat{\eta}_\tau' W_i^* \geq 0)}{\hat{G}(\hat{\eta}_\tau' W_i^*)} - \frac{I(Y_i - \hat{\eta}_\tau'\hat{W}_i \geq 0)}{\hat{G}(\hat{\eta}_\tau'\hat{W}_i)}\right\} \tag{3.12}$$

$$+n^{-1/2}\sum_{i=1}^{n} (W_i^* - \hat{W}_i) \left\{\frac{I(Y_i - \hat{\eta}_\tau'\hat{W}_i \geq 0)}{\hat{G}(\hat{\eta}_\tau'\hat{W}_i)} - (1-\tau)\right\}. \tag{3.13}$$

To obtain an approximation to the distribution of $\{\hat{\eta}_\tau, \tau \in I\}$, one may use the resampling method proposed by Parzen et al. (1994). To be specific, let $\tilde{Q}_\tau(\eta)$ be the observed value of $Q_\tau(\eta)$. Define a random vector η_τ^* such that

$$n^{-1/2}\tilde{Q}_\tau(\eta_\tau^*) = B_\tau^*, \ \tau \in I, \tag{3.14}$$

where $\hat{G}(\eta_\tau^{*\prime}\hat{W}_i) > 0$, $i = 1, \ldots, n$.

It follows from Parzen et al. (1994) that the distribution of $\{(\hat{\eta}_\tau - \eta_{\tau 0}), \tau \in I\}$ can be approximated by the conditional distribution of $\{(\eta_\tau^* - \hat{\eta}_\tau), \tau \in I\}$ given the data. Let b_τ^* in η_τ^* be the counterpart of $\hat{b}_\tau$ in $\hat{\eta}_\tau$. Then, the distribution of $\{n^{1/2}(\hat{b}_\tau - \beta_0), \tau \in I\}$ can be approximated by the conditional distribution of $\{n^{1/2}(b_\tau^* - \hat{b}_\tau), \tau \in I\}$.

To obtain the optimal linear combination $\hat{b}$ in (3.3) to make inferences about β_0, first one may generate $\mathcal{M}_1$ independent realizations from $\{b^*_\tau, \tau \in I\}$ based on which we construct the standard sample covariance matrix $\hat{\Omega}$ to estimate the weights of $\hat{b}$. We then generate additional $\mathcal{M}_2$ realizations from $\{b^*_\tau, \tau \in I\}$ to obtain $\mathcal{M}_2$ corresponding realized linear combinations

$$b^* = \frac{d'\hat{\Omega}^{-1}(b^*_{\tau_1}, \ldots, b^*_{\tau_K})'}{d'\hat{\Omega}^{-1}d}. \tag{3.15}$$

Inferences about β_0 can be made based on these $\mathcal{M}_2$ independent realized b^*, for example, via a robust variance estimate $\hat{\sigma}^2$ for $\hat{b}$ and the normal approximation to the distribution of $\hat{b}$. This two-stage procedure works well in practice and is illustrated with examples in the next section.

It is important to note that in theory, the variance for $\hat{b}$ decreases as K, the number of quantiles involved in the linear combination, increases. However, in practice the covariance matrix estimate $\hat{\Omega}$ can be rather unstable when K is large and the resulting variance estimate for $\hat{b}$ can substantially under-estimate the true variance of $\hat{b}$. Empirically we find that $\hat{b}$ performs well with three properly chosen τ's. More discussions on these issues are given in the next section.

4. EXAMPLE AND NUMERICAL STUDIES

In this section, we use two examples to illustrate the new proposal and a simulation study to show the validity of the resulting confidence interval estimates derived from $\hat{b}$ for β_0. For the nonparametric function estimate $\hat{v}(Z)$ we use the Epanechnikov kernel (Härdle & Marron, 1985). For each covariate, all the observed values are standardized by their sample mean and standard deviation in the analysis. For the present case, one needs to choose a slightly under-smoothed parameter $h = o_p(n^{-1/4})$. To this end, we first obtain the optimal bandwidth h_c via the cross-validation procedure, which is of order $n^{-1/5}$ (Härdle & Marron, 1985, p. 1467-68). We then let $h = h_c n^{-r+1/5}$, where $r = 1/4+0.01$. Moreover, we use 5-10% trimming as the weighting in the above cross-validation depending on the sparseness of the observed Z values.

To obtain $\hat{\eta}_\tau$ and η^*_τ from (3.14) we use the Nelder & Mead (1965) procedure implemented via R software (R Development Core Team, 2004) to minimize, respectively, $||n^{-1/2}\tilde{Q}_\tau(\eta)||$ and $||n^{-1/2}\tilde{Q}_\tau(\eta) - B^*_\tau||$, for each realized B^*_τ. In the above minimizations we restrict our search to the set of η such that $\hat{G}(\eta'\hat{W}_i) > 0$, $i = 1, \ldots, n$. Since we locate $\hat{\eta}_\tau$ and η^*_τ via a minimization process, we need to check if $\hat{\eta}_\tau$ and η^*_τ indeed give reasonably small values of $n^{-1/2}\tilde{Q}(\hat{\eta}_\tau)$ and $n^{-1/2}\tilde{Q}_\tau(\eta^*_\tau) - B^*_\tau$, respectively. For the two examples, we let $\mathcal{M}_1 = 250$ for constructing $\hat{\Omega}$ and let $\mathcal{M}_2 = 750$ for constructing a sample variance estimate $\hat{\sigma}^2$, but with 2.5% trimming on both tails.

The first example is from a Veterans Administration lung cancer trial with 137 patients (Kalbfleisch & Prentice, 2002, Appendix A). We are interested in estimating the tumor cell type differences on survival by adjusting for the patient's performance status, a measure of general fitness on a scale from 0 to 100. There are four cell types, squamous, small, adeno and large. To fit the data with Model (1.2), we let T be the natural logarithm of the survival time, X be a 3×1 vector consisting of three binary covariates by taking large cell type as the reference, and Z be the performance score. The covariate X and confounder Z are expected to be correlated. There are nine censored survival times. Here, $h = 0.48$ based on the above cross-validation procedure. In Table 1, we report two sets of point estimates $\hat{b}$ and their estimated standard errors. One is the optimal combination $\hat{b}_{(3)}$ with three quantiles: $\tau = 0.25, 0.5, 0.75$, and the other, $\hat{b}_{(5)}$, is with five quantiles: $\tau = 0.25, 0.35, 0.5, 0.65, 0.75$. The results based on $\hat{b}_{(5)}$ are practically identical to those based on $\hat{b}_{(3)}$. We also report $\hat{b}$ for $\tau = 0.5$ only. The standard errors of the median regression estimators are larger than those based on $\hat{b}_{(3)}$ and $\hat{b}_{(5)}$. Furthermore, for comparisons, we report the results based on the Gehan estimate for b by treating (1.2) as the accelerated failure time model with $\hat{v}(Z)$ replaced by Z (Jin et al., 2003). Except for the non-significant cell type difference, squamous vs. large, the Gehan estimates and their estimated standard errors are similar to ours, largely due to the fact that $\hat{v}(Z)$ is approximately linear in Z.

The second example is from a recent clinical study on treating HIV infected patients sponsored by the AIDS Clinical Trials Group (Henry et al., 1998). This is a multi-center, randomized, double-blind trial conducted from June 1993 to June 1996 to evaluate if a three-drug combination (AZT+ddI+Nevirapine) is better than various two-drug combinations with respect to survival. Thirteen hundred and thirteen patients were randomized to the study and 330 patients were assigned to the three drug combination group. Study patients with the three drug combination tend to survive significantly longer than those in the two drug combination group. Here, we address another important question with this data set, that is, whether the patient's short term CD4 count change is a good surrogate marker for survival (Prentice, 1989; Fleming et al., 1994). To this end, we let T be the natural logarithm of the survival time, $X = 1$, if the patient is in the three drug combination group, zero, otherwise, and Z be the patient's CD4 count change from the randomization date to week 8. There are 893 patients who had CD4 counts at Week 8. Among these patients approximately 52% had censored survival times. We assume that T and X, Z are related via Model (1.1). Note that for the present case, we are not interested in estimating the function $g(\cdot)$ in the model. If the change of CD4 count is a potential surrogate marker, a necessary condition is that the regression coefficient $\beta_0 \approx 0$. For the present case, X is highly correlated with Z. With the working model (1.2) and $h = 0.205$ in $\hat{v}(\cdot)$, we first constructed $\hat{b}$ with three quantiles: $\tau = 0.4, 0.5, 0.7$ under the same setting as that for the analysis of the above lung cancer data. The estimate is 0.014 with estimated standard error of 0.031, indicating β_0 is very likely in a tight interval which contains 0. We also obtained $\hat{b}$ with five quantiles: $\tau = 0.4, 0.5, 0.6, 0.7, 0.8$. The resulting estimate is 0.022 with estimated standard error of 0.027. On the other hand, if we use the Gehan estimate with the working model (1.2), but by replacing $\hat{v}(Z)$ with Z, the estimate is 0.05 with estimated standard error of 0.029.

To examine if the new proposal is valid for cases with practical sample sizes, we conducted a simulation study. To create a model (1.1) for generating the underlying survival times, we mimicked the lung cancer study. First, we fitted the lung cancer data with (1.1) by letting $X = 1$, if the patient has the large cell type, zero, otherwise,

$g(Z) = Z$, the performance score, and the error ϵ is normal with mean 0 and an unknown variance. We then used the maximum likelihood procedure to estimate the unknown parameters. This results in the following model

$$T = 4.16 - 0.31 \times X + 0.75 \times Z + \epsilon, \tag{4.1}$$

where ϵ is normal with mean 0 and variance 1.1. We then used the observed $n = 137$ covariate vectors in the lung cancer data set repeatedly to generate 500 sets of $\{T_i, i = 1, \ldots, 137\}$ via Model (4.1). Moreover, we let the censoring variable C be a uniform variable on the interval $(3, c_u)$, where c_u is chosen to obtain a pre-specified censoring proportion. Here, $h = 1.12$. For each generated data set, we let $M_1 = 100$, $M_2 = 200$, and $\tau = 0.35, 0.5, 0.75$. We find that our interval estimation procedure behaves well. For example, for the case without censoring or with 20% censoring, the empirical coverage probability of the 0.95 confidence interval for β_0 is about 0.96.

In the second part of the simulation study, we considered a case where the function $g(Z)$ is not proportional to $E(X|Z) = v(Z)$ to examine if our confidence interval estimates still have correct coverage probabilities in practice. Here, the true model (1.1) for T and the binary variable X and continuous confounder Z is

$$T = 6 - 0.75 \times X + 1.25 \times Z - 1.5 \times Z^2 + \epsilon, \tag{4.2}$$

where ϵ is normal with mean 0 and variance 0.25. The sample size is $n = 100$. The corresponding Z's were generated from a uniform $(0, 10)$, and X's were generated via a logistic regression: $\text{logit}(\text{pr}(X = 1|Z)) = -0.25 - 0.75 \times Z + 0.5 \times Z^2$. For this specific set of 100 covariate vectors, 500 samples were generated via (4.2). Here, $h = 0.91$. Under the same setting as that in the first part of the simulation study, for the case without censoring the empirical coverage probability of our 0.95 interval based on the optimal combination of three quantiles, $\tau = 0.35, 0.5, 0.75$, is 0.944. For the case with 20% censoring, the empirical level is 0.924. On the other hand, if we use (1.2) with $\hat{v}(Z)$ replaced by Z to analyze these simulated data, the empirical level of the 0.95 interval derived from the Gehan score (Jin et al., 2003) is about 0.5 for the non-censored or censored case.

5. REMARKS

For the censored linear regression, practically useful model checking and selection procedures are not available. Thus, a robust inference procedure such as the one we proposed here, appears to be a valuable tool for censored data analysis.

In the presence of censoring, $\hat{b}_\tau$ exists for τ in a subset of the interval $(0, 1)$, for example, $0 < \tau \le \psi < 1$ such that $\hat{\eta}'_\psi \hat{W}_i < t_0, i = 1, \ldots, n$. Theoretically one may utilize all the $\hat{b}_\tau, 0 < \tau < \psi$ to obtain an optimal linear combination $\int_0^\psi \hat{b}_\tau d\omega(\tau)$, where the weight function $\omega(\cdot)$ satisfies the condition $\int_0^\psi d\omega(\tau) = 1$, to estimate β_0 (Carrasco & Florens, 2000). In practice, unlike the case when the working model is correctly specified and there are no censored observations (Portnoy & Koenker, 1989), the weight function $\omega(\cdot)$ is rather difficult to estimate well. Although through a limited simulation study we find that an optimal combination $\hat{b}$ of three quantile estimates $\hat{b}_\tau$, whose τ's are approximately equally spaced over the interval $(0, \psi)$, works well, it is not clear how to justify such a choice in general. A data-dependent, adaptive way to choose an appropriate set of quantile regression estimates for $\hat{b}$ may be a feasible solution to this challenging problem.

If the confounder Z is multivariate, the analysis with a nonparametric regression estimate $\hat{v}(\cdot)$ can be quite complicated if not impossible. An alternative way to handle this case is to use a rich parametric model for the mean of the covariate X given Z.

In this article, we assume that the censoring distribution $G(\cdot)$ is free of X and Z. Naturally one may generalize this case by modeling the censoring distribution with the covariates and confounders semiparametrically. Research on developing robust procedures without modeling the censoring distribution warrants further investigation, for example, via a novel approach recently taken by Portnoy (2003).

Recently Huang (1999) proposed a novel estimation procedure for the Cox model with the partial linear structure similar to model (1.1). However, there is no valid variance estimate available for his estimate of the regression coefficients of interest.

6. APPENDIX
DERIVATION OF LARGE SAMPLE PROPERTIES OF $\hat{b}$

First, we show that $\psi(\eta)$ is locally linear in a neighborhood of $\eta_{\tau 0}$. To this end, note that

$$n^{-1/2}\psi(\eta) = n^{-1/2}\sum_{i=1}^{n} W_i \left\{ \frac{I(Y_i - \eta'\hat{W}_i \geq 0)}{\hat{G}(\eta'\hat{W}_i)} - S_{W_i}(\eta' W_i) \right\} \tag{6.1}$$

$$+n^{-1/2}\sum_{i=1}^{n}(\hat{W}_i - W_i) \left\{ \frac{I(Y_i - \eta'\hat{W}_i \geq 0)}{\hat{G}(\eta'\hat{W}_i)} - S_{W_i}(\eta' W_i) \right\}. \tag{6.2}$$

Furthermore, (6.2) =

$$\int n^{1/2} \left\{ \frac{1}{\hat{G}(u)} - \frac{1}{G(u)} \right\} d\hat{C}_\eta(u) \tag{6.3}$$

$$+n^{-1/2}\sum_{i=1}^{n}(\hat{W}_i - W_i) \left\{ \frac{I(Y_i - \eta'\hat{W}_i \geq 0)}{G(\eta'\hat{W}_i)} - S_{W_i}(\eta' W_i) \right\}, \tag{6.4}$$

where $\hat{C}_\eta(u) = n^{-1}\sum_{i=1}^{n}(\hat{W}_i - W_i) I(Y_i - \eta'\hat{W}_i \geq 0, \eta'\hat{W}_i \leq u)$.

It follows from the uniform consistency of $\hat{v}(\cdot)$ and the uniform law of large numbers (Pollard, 1990, p. 53) that $\hat{C}_\eta(u)$ converges weakly to 0, uniformly in η and u. Moreover, $n^{1/2}(\hat{G}(t) - G(t))$ is asymptotically equivalent to a standardized sum of independent and identically distributed random quantities (Fleming & Harrington, 1991). It follows from Lemma A.3 of Bilias et al. (1997) that (6.3) is $o_p(1)$ uniformly in η. Furthermore, it is straightforward to show that the variance of (6.4) goes to 0, uniformly in η, as $n \to \infty$. It follows that $n^{-1/2}\psi(\eta)$ is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n} W_i \left\{ \frac{I(Y_i - \eta'\hat{W}_i \geq 0)}{\hat{G}(\eta'\hat{W}_i)} - \frac{I(Y_i - \eta'\hat{W}_i \geq 0)}{G(\eta'\hat{W}_i)} \right\} \tag{6.5}$$

$$+n^{-1/2}\sum_{i=1}^{n} W_i \left\{ \frac{I(Y_i - \eta'\hat{W}_i \geq 0)}{G(\eta'\hat{W}_i)} - S_{W_i}(\eta'\hat{W}_i) \right\} \tag{6.6}$$

$$+n^{-1/2}\sum_{i=1}^{n} W_i \left\{ S_{W_i}(\eta'\hat{W}_i) - S_{W_i}(\eta' W_i) \right\}. \tag{6.7}$$

Now,

$$(6.5) = \int n^{1/2} \left\{ \hat{G}(u) - G(u) \right\} d\hat{D}_\eta(u),$$

where $\hat{D}_\eta(u) = n^{-1}\sum_{i=1}^{n} \frac{W_i I(Y_i - \eta'\hat{W}_i \geq 0, \eta'\hat{W}_i \geq u)}{\hat{G}(\eta'\hat{W}_i) G(\eta'\hat{W}_i)}$. It follows from the uniform convergence of $\hat{G}(\cdot)$, $\hat{v}(\cdot)$ and a uniform law of large numbers that $\hat{D}_\eta(u)$ converges to a deterministic function $D_\eta(u)$, in probability, uniformly in η and u. This implies that (6.5) is asymptotically equivalent to $\int n^{1/2}\left\{ \hat{G}(u) - G(u) \right\} dD_\eta(u)$, a standardized sum of independent and identically distributed random quantities.

For (6.6), consider the following quantity

$$n^{-1/2}\sum_{i=1}^{n} W_i\left\{\frac{I(Y_i \geq \eta'\hat{W}_i)}{G(\eta'\hat{W}_i)} - S_{W_i}(\eta'\hat{W}_i) - \frac{I(Y_i \geq \eta' W_i)}{G(\eta' W_i)} + S_{W_i}(\eta' W_i)\right\}. \qquad (6.8)$$

It follows from the uniform convergence of $\hat{v}(\cdot)$, the variance of (6.8) goes to 0, uniformly in η. Therefore, (6.6) is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n} W_i\left\{\frac{I(Y_i - \eta' W_i \geq 0)}{G(\eta' W_i)} - S_{W_i}(\eta' W_i)\right\}.$$

Next, by the Taylor series expansion, (6.7) is asymptotically equivalent to

$$n^{-1/2}\sum_{i} W_i \dot{S}_{W_i}(\eta' W_i)\eta'(\hat{W}_i - W_i) = n^{1/2}\int b_2'(\hat{v}(u) - v(u))d\hat{E}_\eta(u), \qquad (6.9)$$

where $\dot{S}_{W_i}(u)$ is the derivative of $S_{W_i}(u)$, and $\hat{E}_\eta(u) = n^{-1}\sum_i W_i I(Z_i \leq u)\dot{S}_{W_i}(\eta' W_i)$. It follows from the weak convergence of the process $n^{-1/2}\{\hat{E}_\eta(u) - E_\eta(u)\}$, the uniform consistency of $\hat{v}(\cdot)$ and Lemma A.3 of Bilias et al. (1997) that (6.9) =

$$n^{1/2}\int b_2'(\hat{v}(u) - v(u))dE_\eta(u) + o_p(1), \qquad (6.10)$$

where $E_\eta(u) = E\{\hat{E}_\eta(u)\}$. With some elementary algebraic operations and using $h = o(n^{-1/4})$ and $nh(\log n)^{-1} \to \infty$, in probability, (6.10) =

$$n^{-1/2}\sum_i b_2'(X_i - v(Z_i))\{f_z(Z_i)\}^{-1}\dot{E}_\eta(Z_i) + o_p(1),$$

where $\dot{E}_\eta(u)$ is the derivative of $E_\eta(u)$, and $f_z(\cdot)$ is the density function of Z. This implies that $n^{-1/2}\psi(\eta)$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^n \Psi(\eta; Y_i, \Delta_i, X_i, Z_i)$, where Ψ is a deterministic function. Since the density functions of ϵ and Z are assumed to be continuously differentiable with bounded derivatives, it follows from (5.2) of Pollard (1990) that $\Psi(\eta; Y_i, \Delta_i, X_i, Z_i)$ has finite pseudo dimension and thus the process $n^{-1/2}\sum_{i=1}^n \Psi(\eta; Y_i, \Delta_i, X_i, Z_i)$ is manageable. This, coupled with a functional central limit theorem (Pollard, 1990, p. 53), implies that the process $n^{-1/2}\psi(\eta)$ converges weakly to a zero mean Gaussian process in the neighborhood of $\eta_{\tau 0}$. This establishes the local linearity property for $\psi(\eta)$ around $\eta_{\tau 0}$.

We now show that $n^{-1/2}Q_\tau(\eta_{\tau 0})$ is asymptotically normal. From (3.2) and the above derivation of the local linearity property of $\psi(\eta)$, $n^{-1/2}Q_\tau(\eta_{\tau 0})$ is asymptotically equivalent to

$$\int n^{1/2}\left\{\hat{G}(u) - G(u)\right\} dD_{\eta_{\tau 0}}(u) \qquad (6.11)$$

$$+n^{-1/2}\sum_{i=1}^{n} W_i\left\{\frac{I(Y_i - \eta_{\tau 0}' W_i \geq 0)}{G(\eta_{\tau 0}' W_i)} - (1-\tau)\right\} \qquad (6.12)$$

$$+n^{-1/2}\sum_i \beta_0'(X_i - v(Z_i))\{f_z(Z_i)\}^{-1}\dot{E}_{\eta_{\tau 0}}(Z_i) \tag{6.13}$$

$$+n^{-1/2}\sum_i \begin{pmatrix} 0_{(p+1)\times 1} \\ X_i - v(Z_i) \end{pmatrix}\{f_z(Z_i)\}^{-1}\dot{H}_{\eta_{\tau 0}}(Z_i), \tag{6.14}$$

where $\dot{H}_\eta(u)$ is the derivative of $H_\eta(u)$, and $H_\eta(u) = E(I(Z_i \leq u)(S_{W_i}(\eta' W_i) - (1 - \tau)))$. This establishes the weak convergence of $n^{-1/2}Q_\tau(\eta_{\tau 0})$.

Lastly, we briefly show how the perturbation method presented in Section 3 works for approximating the distribution of $Q_\tau(\eta_{\tau 0})$. Note that for B_τ^*, (3.10) $\approx$

$$n^{1/2}\int (G^*(u) - \hat{G}(u))dD_{\eta_{\tau 0}}(u). \tag{6.15}$$

The quantity (3.11) $\approx$

$$n^{-1/2}\sum_{i=1}^n W_i\left\{\frac{I(Y_i - \eta_{\tau 0}' W_i \geq 0)}{G(\eta_{\tau 0}' W_i)} - (1-\tau)\right\}V_i. \tag{6.16}$$

Also, (3.12) $\approx$

$$n^{-1/2}\sum_i W_i\{S_{W_i}(\eta_{\tau 0}' W_i^*) - S_{W_i}(\eta_{\tau 0}' \hat{W}_i)\} \approx$$

$$n^{1/2}\int \beta_0'(v^*(u) - \hat{v}(u))dE_{\eta_{\tau 0}}(u). \tag{6.17}$$

Moreover, (3.13) $\approx$

$$n^{1/2}\int \begin{pmatrix} 0_{(p+1)\times 1} \\ v^*(u) - \hat{v}(u) \end{pmatrix} dH_{\eta_{\tau 0}}(u). \tag{6.18}$$

Then, using the definitions (3.7) and (3.8) for $G^*(\cdot)$ and $v^*(\cdot)$, it can be shown that conditional on the data, the limiting distribution of $\{B_\tau^*, \tau \in I\}$ is the same as the unconditional limiting distribution of $\{n^{-1/2}Q_\tau(\eta_{\tau 0}), \tau \in I\}$.

Table 1. Robust estimates (standard errors) of tumor type differences for lung cancer data adjusting for performance score

Tumor type versus large	$\hat{b}_{(3)}$	$\hat{b}_{(5)}$	Median*	Gehan†
adeno	-0.247 (0.073)	-0.237 (0.070)	-0.276 (0.092)	-0.294 (0.093)
small	-0.296 (0.074)	-0.289 (0.073)	-0.429 (0.098)	-0.317 (0.107)
squamous	0.112 (0.083)	0.109 (0.081)	-0.139 (0.115)	0.001 (0.119)

* Median: Estimate with $\tau = 0.50$ only
† Gehan: Estimate based on Gehan score for the accelerated failure time model

References

Bhattacharya, P. K. & Zhao, P.-L. (1997). Semiparametric inference in a partial linear model. *The Annals of Statistics* **25**, 244–262.

Bilias, Y., Gu, M. & Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *The Annals of Statistics* **25**, 662–682.

Carrasco, M. & Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* **16**, 797–834.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *The Annals of Statistics* **16**, 136–146.

Chen, H. & Shiau, J.-J. H. (1991). A two-stage spline smoothing method for partially linear models. *Journal of Statistical Planning and Inference* **27**, 187–201.

Cook, R. D. (1993). Exploring partial residual plots. *Technometrics* **35**, 351–362.

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

Csörgő, S. & Horváth, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete [Became: @J(ProbTher)]* **62**, 411–426.

Cuzick, J. (1992). Efficient estimates in semiparametric additive regression models with unknown error distribution. *The Annals of Statistics* **20**, 1129–1136.

DiRienzo, A. G. & Lagakos, S. W. (2001a). Bias correction for score tests arising from misspecified proportional hazards regression models. *Biometrika* **88**, 421–434.

DiRienzo, A. G. & Lagakos, S. W. (2001b). Effects of model misspecification on tests of no randomized treatment effect arising from Cox's proportional hazards model. *Journal of the Royal Statistical Society, Series B, Methodological* **63**, 745–757.

Engle, R. F., Granger, C. W. J., Rice, J. & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association* **81**, 310–320.

Fleming, T. R. & Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley.

Fleming, T. R., Prentice, R. L., Pepe, M. S. & Glidden, D. (1994). Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. *Statistics in Medicine* **13**, 955–968.

Gail, M. H., Wieand, S. & Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71**, 431–444.

Goldwasser, M. A., Tian, L. & Wei, L. (2004). Statistical inference for infinite-dimensional parameters via asymptotically pivotal estimating functions. *Biometrika* **91**, 81–94.

Green, P., Jennison, C. & Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B, Methodological* **47**, 299–315.

Härdle, W., Janssen, P. & Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *The Annals of Statistics* **16**, 1428–1449.

Härdle, W. & Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics* **13**, 1465–1481.

Heckman, N. E. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society, Series B, Methodological* **48**, 244–248.

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**, 339–349.

Hong, S.-Y. & Cheng, P. (1999). Normal approximation rate of the kernel smoothing estimator in a partial linear model. *Journal of Multivariate Analysis* **70**, 207–220.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *The Annals of Statistics* **27**, 1536–1563.

Jin, Z., Lin, D. Y., Wei, L. J. & Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Koul, H., Susarla, V. & Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *The Annals of Statistics* **9**, 1276–1288.

Lagakos, S. W. & Schoenfeld, D. A. (1984). Properties of proportional-hazards score tests under misspecified regression models. *Biometrics* **40**, 1037–1048.

Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309.

Lin, D. Y., Fleming, T. R. & Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73–81.

Lin, D. Y. & Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84**, 1074–1078.

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

Morgan, T. M. (1986). Omitting covariates from the proportional hazards model. *Biometrics* **42**, 993–995.

Nelder, J. A. & Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal* **7**, 308–313.

Parzen, M. I., Wei, L. J. & Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

Pollard, D. (1990). *Empirical processes: theory and applications.* Institute of Mathematical Statistics.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association* **98**, 1001–1012.

Portnoy, S. & Koenker, R. (1989). Adaptive L-estimation for linear models.

The Annals of Statistics **17**, 362–381.

PRENTICE, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* **8**, 431–440.

QIN, G. & JING, B.-Y. (2001). Censored partial linear models and empirical likelihood. *Journal of Multivariate Analysis* **78**, 37–61.

R DEVELOPMENT CORE TEAM (2004). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

RICE, J. (1986). Convergence rates for partially splined models. *Statistics & Probability Letters* **4**, 203–208.

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *The Annals of Statistics* **18**, 303–328.

ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56**, 931–954.

SHIAU, J., WAHBA, G. & JOHNSON, D. (1986). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two and three dimensional objective analysis. *J. Atmos. Ocean. Technol.* **3**, 714–725.

SPECKMAN, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B, Methodological* **50**, 413–436.

STRUTHERS, C. A. & KALBFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363–369.

TIAN, L., LIU, J., ZHAO, M. & WEI, L. (2004). Statistical inferences based on non-smooth estimating functions. *Biometrika, in press* .

TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *The Annals of Statistics* **18**, 354–372.

WAHBA, G. (1984). Cross-validated spline methods for the estimation of multivariate functions from data on functionals. In David, H. A. & David, H. T., editors, *Statistics: An Appraisal*, pages 205–235. Iowa State University.

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

YING, Z., JUNG, S. H. & WEI, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90**, 178–184.