Statistica Sinica **22** (2012), 509-530
doi:http://dx.doi.org/10.5705/ss.2010.109

MODEL CHECKING TECHNIQUES FOR ASSESSING FUNCTIONAL FORM SPECIFICATIONS IN CENSORED LINEAR REGRESSION MODELS

Larry F. León and Tianxi Cai

Genentech and Harvard University

Abstract: In this paper we develop model checking techniques for assessing functional form specifications of covariates in censored linear regression models. These procedures are based on a censored data analog to taking cumulative sums of "robust" residuals over the space of the covariate under investigation. These cumulative sums are formed by integrating certain Kaplan-Meier estimators and may be viewed as "robust" censored data analogs to the processes considered by Lin, Wei, and Ying (2002). The null distributions of the stochastic processes can be approximated by the distributions of certain zero-mean Gaussian processes whose realizations can be generated by computer simulation. Each observed process can then be graphically compared with a few realizations from the Gaussian process. We also develop formal test statistics for numerical comparison. Such comparisons enable one to assess objectively whether an apparent trend seen in a residual plot reflects model misspecification or natural variation. We illustrate the methods with a well-known dataset. In addition, we examine the finite sample performance of the proposed test statistics in simulation experiments. In them, the proposed test statistics have good power for detecting misspecification while at the same time controlling the size of the test.

Key words and phrases: Censored linear regression, goodness-of-fit, partial linear model, partial residual, quantile regression, resampling method, rank estimation.

1. Introduction

Regression models are widespread statistical tools applied in the analysis of experimental and observational data. When the model is misspecified, this can seriously affect the validity and efficiency of inference procedures. Unfortunately, investigators do not routinely check the adequacy of the specified model for their particular data analysis. Model checking techniques have been developed for some commonly used regression models, including the linear regression model (Stute (1997); Stute et al. (1998)), and generalized linear models (Su and Wei (1991); Lin, Wei, and Ying (2002); Stute and Zhu (2002)). When analyzing failure time outcomes subject to censoring, a commonly used model is the Cox proportional hazards model (Cox (1972)). For the Cox model, the effects of misspecification have been well studied (Gail, Wieand, and Piantadosi (1984), Lagakos and Schoenfeld (1984), Morgan (1986), Struthers and Kalbfleisch (1986), DiRienzo and Lagakos (2001)) and model checking procedures have also been developed (Lin, Wei, and Ying (1993); León and Tsai (2004)). However, the proportional hazards assumption may not be appropriate for certain applications. A useful alternative is the Accelerated Failure Time (AFT) model (Wei (1992)) which has been studied extensively in recent years for the standard regression setting (Buckley and James (1979), Tsiatis (1990), Ritov (1990), Jin et al. (2003)). Model checking techniques for the AFT model are not well developed. Recently Lopez and Patilea (2009) developed omnibus test procedures. Their tests however can be sensitive to the choice of the bandwidth and an approach for selecting a bandwidth is not provided. Moreover, their omnibus test can be sensitive to the degree of censoring, with poor performance under moderate censoring. In this paper, we propose an alternative procedure for assessing the adequacy of the AFT model with respect to the functional form specification of covariates. The AFT model assumes that a transformation (e.g., log-transformation) of the survival time is linearly related to covariates via

$$T = \alpha + \beta' X + \gamma Z + e, \tag{1.1}$$

where X is a $p \times 1$ vector, Z is a scalar random variable, α , β, and γ are unknown parameters, and e is an independent error term with unspecified distribution constrained so that α is identifiable. Here and in the sequel, we use T to denote the transformed survival time. Under correct specification β and γ may be estimated by various procedures, including the Buckley and James estimator (Buckley and James (1979)), the inverse probability weighted estimator (Koul, Susarla, and Van Ryzin (1981)), and the rank based estimators (Tsiatis (1990)). However, when the linear functional form of Z is misspecified, these procedures may result in biased estimators for β and thus lead to invalid inference. To examine the appropriateness of the linearity assumption, we consider the alternative class of partial linear model,

$$T = \alpha_0 + \beta_0' X + g(Z) + \epsilon, \tag{1.2}$$

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. Model (1.2) reduces to (1.1) when $g(z) = z$. Here we view (1.2) as the "true" model and obtain a consistent estimate for β in (1.1), regardless of whether or not the functional form for Z is correctly specified therein. We apply the estimator of León, Cai, and Wei (2009) who have extended model (1.2) to the censored linear regression model and developed valid inference procedures for β_0 without involving estimation of $g(\cdot)$ in (1.2).

To check the functional form of Z in model (1.1) for non-censored data, Lin, Wei, and Ying (2002) proposed residual plots and Kolmogorov-Smirnov (KS) type lack of fit tests based on the cumulative residual process

$$\hat{Q}_0(v) = n^{-1}\sum_{i=1}^{n}[T_i - \hat{\alpha} - \hat{\beta}' X_i - \hat{\gamma} Z_i] I(Z_i \leq v),$$

where $(\hat{\alpha}, \hat{\beta}', \hat{\gamma})'$ is the least squares estimate of the parameter vector $(\alpha, \beta', \gamma)'$ in model (1.1) and $I(\cdot)$ is the indicator function. Under correct specification, the process $\hat{Q}_0(\cdot)$ fluctuates randomly around zero. It is important to note that, under functional form misspecification of Z, the parameter estimate $\hat{\beta}$ corresponding to the correctly specified covariate vector X can be seriously biased, especially when X and Z are highly correlated (Berk and Booth (1995); Grambsch, Therneau, and Fleming (1995)). The ability of the cumulative residual process to identify the form of misspecification can therefore be jeopardized. An extreme situation would be $(\beta_0 - \bar{\beta})' E[X|Z = z] \approx g(z) - \bar{\gamma} z$ for each z in the support of Z where $g(\cdot)$ is non-linear; the process $\hat{Q}$ would have little power for detecting the non-linearity in such a case. In general, the bias of $\hat{\beta}$ can lead to misleading residual plots and/or consistency of the KS-type test W_0 can fail. To overcome such difficulties, we propose an analog of $\hat{Q}_0$ for the censored linear regression model but based on a robust estimator of β_0. That is, we employ the León, Cai, and Wei (2009) estimator of β_0 that is consistent under misspecification of the functional form of Z. This robustness enables us to develop residual plots and goodness-of-fit statistics that have the ability to identify misspecification and suggest the particular form.

The paper is outlined as follows. In Section 2 we state the censored linear regression model under consideration and outline the parameter estimates used in forming our "robust" residuals. In addition, we develop a censored data analog to the cumulative residual process $\hat{Q}_0$ and propose functional form goodness-of-fit test statistics. In Section 3 we show how to approximate the null distribution of our cumulative residual process and provide a monte carlo technique for estimating the p-values of the goodness-of-fit statistics. Because our cumulative residual process requires the selection of a smoothing bandwidth, in Section 4 we develop a cross-validation procedure. In Section 5 we apply our methods to the Mayo Clinic's primary biliary cirrhosis (PBC) data set and we examine the finite sample performance of our test statistics in simulation experiments. Section 6 concludes with some remarks. Technical details are relegated to the Appendix.

2. Censored Linear Regression Models

2.1. Model and estimators

Let C be the censoring variable. For T, we only observe $Y = \min(T, C)$ and $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. The censoring times C are assumed to be independent of the survival times T and covariates (X, Z), with a common distribution function G. The covariates (X, Z) are assumed to be bounded. Without loss of generality, we assume that the support of Z is $[\mathfrak{z}_l, \mathfrak{z}_r]$ with $-\infty < \mathfrak{z}_l < \mathfrak{z}_r < \infty$. Let $(T_i, C_i, X_i, Z_i), i = 1, \ldots, n$, be independent realizations of (T, C, X, Z).

To construct a residual process for assessing the covariate functional form, we first estimate the regression parameters $\psi = (\beta', \gamma)'$. For β, we consider the León, Cai, and Wei (2009) estimator based on a quantile regression method under the *working* model

$$T = a + b'X + c'\ \hat{\kappa}(Z) + e, \tag{2.1}$$

where a is the intercept, b and c are $p \times 1$ vectors of parameters, and e is the error term. Here $\hat{\kappa}(Z) = \sum_{i=1}^{n} K_\zeta(Z_i - z)X_i / \sum_{i=1}^{n} K_\zeta(Z_i - z)$ is a nonparametric kernel estimate of $\kappa(Z) = E(X \mid Z)$, where $K_\zeta(\cdot) = \zeta^{-1}K(\cdot/\zeta)$, $K(\cdot)$ is a symmetric density function with $\int x^2 K(x)dx < \infty, \zeta \to 0$, and $(\log n)^{-1} n\zeta \to \infty$ as $n \to \infty$. Their estimator, $\hat{\theta}_\rho = (\hat{a}_\rho, \hat{b}'_\rho, \hat{c}'_\rho)'$, say, is defined as the solution to the ρ^{th} quantile estimating function

$$\Xi_\rho(\theta) = \sum_{i=1}^{n} \hat{U}_i \Big\{ \frac{I(Y_i \geq \theta'\hat{U}_i)}{\hat{G}(\theta'\hat{U}_i)} - (1 - \rho) \Big\}, \tag{2.2}$$

where $\hat{U}_i' = (1, X_i', \hat{\kappa}(Z_i)')$ and $\hat{G}(\cdot)$ is the K-M estimator of $G(\cdot)$. Under mild regularity conditions, $\hat{b} = \hat{b}_\rho$ is consistent for β_0 in the partial linear model (1.2). For the regression parameter γ_0 in the null model (1.1), we consider the Gehan estimator (Ritov (1990), Tsiatis (1990), Wei, Ying, and Lin (1990)) defined as follows. Let $\tilde{\psi}$ be the minimizer of

$$L(\psi) = \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_i \{e_i(\psi) - e_j(\psi)\}^-, \tag{2.3}$$

where $e_i(\psi) = Y_i - \beta' X_i - \gamma Z_i$ and $\{a\}^- = |a| I(a \leq 0)$. Note that $\tilde{\psi}$ can be obtained by standard linear programming techniques. Our estimator for γ, denoted by $\hat{\gamma}$, is the corresponding component of $\tilde{\psi}$ corresponding to γ. Let $\hat{\psi} = (\hat{b}', \hat{\gamma})'$ and denote the limit of $\hat{\gamma}$ (in probability) by $\bar{\gamma}$, assumed to be finite. By León, Cai, and Wei (2009) and U-process theory, it is straightforward to show that $\hat{\psi}$ converges to its limit, $\bar{\psi} = (\beta_0', \bar{\gamma})'$, in probability.

2.2. Functional form checking techniques

We propose to construct a residual process based on the the "robust" residuals $e(\hat{\psi}) = Y - \hat{b}'X - \hat{\gamma}Z$. The residuals $e(\hat{\psi})$ are robust in the sense that, under correct specification, they do not depend on the covariate vector (X, Z), whereas under misspecification of the functional form of Z, they are asymptotically free of X (by the consistency of $\hat{b}$ for β_0) and only depend on Z (through $g(Z) - \hat{\gamma}Z$). Without censoring, the cumulative residual process

$$n^{-1} \sum_{i=1}^{n} \left\{ \epsilon_i(\hat{\psi}) - \bar{\epsilon}(\hat{\psi}) \right\} I(Z_i \leq v)$$

is a convenient choice for assessing the functional form $g(\cdot)$, where $\epsilon_i(\psi) = T_i - \beta' X_i - \gamma Z_i$ and $\bar{\epsilon}(\psi) = n^{-1} \sum_{i=1}^{n} \epsilon_i(\psi)$. However, $\epsilon(\psi)$ is not observable in the presence of censoring. To incorporate censoring, we note that $n^{-1} \sum_{i=1}^{n} \{\epsilon_i(\hat{\psi}) - \bar{\epsilon}(\hat{\psi})\} I(Z_i \leq v)$ converges in probability to

$$E\left\{\epsilon_i(\bar{\psi}) I(Z_i \leq v)\right\} - E\left\{\epsilon_i(\bar{\psi})\right\} H(z) = -\int_{\mathfrak{z}_l}^{v} \left[\int t d\{S_z(t, \bar{\psi}) - S(t, \bar{\psi})\}\right] dH(z), \tag{2.4}$$

where $S_z(t, \psi) = \Pr(\epsilon(\psi) \geq t \mid Z = z)$, $S(t, \psi) = \Pr(\epsilon(\psi) \geq t)$, and $H(z) = \text{pr}(Z \leq z)$. This motivates us to consider a residual process based on an empirical version of (2.4) with $S_z(t, \psi)$ estimated using a local Kaplan Meier estimator, $S(t, \psi)$ estimated using the unconditional Kaplan Meier estimator, and H estimated using the empirical distribution function of $\{Z_i, i = 1, \ldots, n\}$.

To be specific, we estimate $S(t, \psi)$ as the product-integral functional

$$\hat{S}(t, \psi) = \Pi_{s \leq t}[1 - d\hat{\Lambda}(s, \psi)],$$

(cf., Andersen et al. (1993)), where $\hat{\Lambda}(t; \psi) = \int_{-\infty}^{t} d\bar{N}(s, \psi) / \hat{\pi}(s, \psi)$,

$$\bar{N}(t, \psi) = n^{-1} \sum_{i=1}^{n} N_i(t, \psi), \quad \hat{\pi}(t, \psi) = n^{-1} \sum_{i=1}^{n} I(e_i(\psi) \geq t),$$

and $N_i(t, \psi) = I(e_i(\psi) \leq t)\Delta_i$. By the empirical process results of Lai and Ying (1988), $\hat{S}(t, \psi)$ is a uniformly consistent estimator of $S(t, \psi)$ for $t \leq \tau$ and $\psi \in \Omega$, where Ω is a compact set containing $\bar{\psi}$ as an interior point and τ is a pre-selected endpoint such that $\Pr(e(\psi) > \tau) > 0$ for all $\psi \in \Omega$. Here and in the sequel we take $\psi \in \Omega$ and $t \leq \tau$ without mention. The conditional survival function $S_z(t, \psi)$ can be consistently estimated by the Beran estimator (Beran (1981)),

$$\hat{S}_z(t, \psi) = \Pi_{s \leq t}[1 - d\hat{\Lambda}_z(s, \psi)],$$

where $\hat{\Lambda}_z(t;\psi) = \int_{-\infty}^{t} d\bar{N}_z(s,\psi)/\hat{\pi}_z(s,\psi)$, $\bar{N}_z(t,\psi) = n^{-1}\sum_{i=1}^{n} N_i(t,\psi)K_h(Z_i - z)$, $\hat{\pi}_z(t,\psi) = n^{-1}\sum_{i=1}^{n} I(e_i(\psi) \geq t)K_h(Z_i - z)$, $K_h(\cdot) = h^{-1}K(\cdot/h)$ is the standardized kernel function, and $K(\cdot)$ is a symmetric density function. The bandwidth $h = h_n$ satisfies $h \to 0$ and $n^{1/4}h = o_p(1)$. The bandwidth h of order $n^{-1/4}$ represents the undersmoothing that is common in semiparametric estimation. In Section 5 we present a cross-validation procedure for choosing h.

Now, under the linear model when $g(Z) = \gamma_0 Z$, $\hat{S}(t,\hat{\psi}) \approx \hat{S}_z(t,\hat{\psi}) \approx S(t,\psi_0)$ for $t \leq \tau$, where $\psi_0 = (\beta_0', \gamma_0)'$. Thus the residual process

$$\int_{\mathfrak{z}_l}^{v} \left\{ \int_{-\infty}^{\tau} td[\hat{S}_z(t,\hat{\psi}) - \hat{S}(t,\hat{\psi})] \right\} d\hat{H}(z) \tag{2.5}$$

fluctuates randomly about zero. Without censoring, for $\tau = \infty$, $-\int_{-\infty}^{\tau} td\hat{S}_z(t,\hat{\psi})$ is identical to the Nadaraya-Watson estimate of the conditional mean $E(\epsilon(\hat{\psi})|Z = z)$ and $-\int_{-\infty}^{\tau} td\hat{S}(t,\hat{\psi})$ is the unconditional mean of the residuals, $n^{-1}\sum_{i=1}^{n}\epsilon_i(\hat{\psi})$. Therefore (2.5) is a natural analog of $\hat{Q}_0$ for the censored linear regression model. In practice, however, S_z may only be estimable on the support of the weighted observations which may be less than τ. For this reason we let $\hat{S}_z(\tau,\hat{\psi}) = \hat{S}(\tau,\hat{\psi})$. This is analogous to Efron's way (Efron (1981)) of setting the K-M estimator to zero at the last observation. Incorporating this modification into (2.5), we have our final residual process for assessing the functional form $g(\cdot)$: $\hat{Q}(v) =$

$$\int_{\mathfrak{z}_l}^{v} \Big\{ \int_{-\infty}^{\tau^-} td[\hat{S}_z(t,\hat{\psi}) - \hat{S}(t,\hat{\psi})] \Big\} d\hat{H}(z) - \tau \int_{\mathfrak{z}_l}^{v} [\hat{S}_z(\tau^-,\hat{\psi}) - \hat{S}(\tau^-,\hat{\psi})] d\hat{H}(z), \tag{2.6}$$

where τ^- is the point just prior to τ.

In applications we propose plotting $\hat{Q}$ against v and supplementing this plot with realizations of $\hat{Q}$ generated under correct specification of the linear model. In addition, as numerical measures of lack-of-fit, we consider the KS and Cramer-VonMises (CvM)-type tests, defined respectively, by

$$W = n^{1/2} \sup_{v \in [\mathfrak{z}_l, \mathfrak{z}_r]} |\hat{Q}(v)|, \tag{2.7}$$

$$D = n \int_{\mathfrak{z}_l}^{\infty} \hat{Q}(v)^2 d\hat{H}(v). \tag{2.8}$$

In Appendix A, we establish consistency of tests based on W and D. That is, when g is non-linear, the probability of rejecting the null based on W or D goes to 1 as $n \to \infty$.

3. Approximation to the Null Distribution of $\hat{Q}(\cdot)$

In Appendix B, we show that, under correct specification, $n^{1/2}\hat{Q}(v)$ is asymptotically equivalent to a sum of i.i.d random variables, $n^{1/2}\bar{Q}(v) =$

$$\int_{-\infty}^{\tau^-} n^{1/2}\{t[\hat{\eta}_0(t,v,\psi_0)-\bar{\eta}(t,v,\psi_0)]dS(t,\psi_0)+tS(t,\psi_0)[\hat{\eta}_0(dt,v,\psi_0)-\bar{\eta}(dt,v,\psi_0)]\}$$
$$+n^{1/2}\tau S(\tau^-,\psi_0)[\hat{\eta}_0(\tau^-,v,\psi_0)-\bar{\eta}(\tau^-,v,\psi_0)],$$

where $n^{1/2}\hat{\eta}_0(t,v,\psi_0) =$

$$n^{1/2}(\hat{\psi}-\psi_0)\int_{\mathfrak{z}_l}^{v}\left\{\frac{\partial(\Lambda_z(t,\psi)-\Lambda(t,\psi))}{\partial\psi}\bigg|_{\psi=\psi_0}\right\}dH(z) \tag{3.1}$$

and $n^{1/2}\bar{\eta}(t,v,\psi_0) =$

$$n^{-1/2}H(v)\sum_{i=1}^{n}\int_{-\infty}^{t}\frac{dN_i(u,\psi_0)-I(e_i(\psi_0)\geq u)\Lambda(du,\psi_0)}{\pi(u,\psi_0)} \tag{3.2}$$
$$-n^{-1/2}\sum_{i=1}^{n}I(Z_i\leq v)\dot{H}(Z_i)\int_{-\infty}^{t}\frac{dN_i(u,\psi_0)-I(e_i(\psi_0)\geq u)\Lambda_{Z_i}(du,\psi_0)}{\pi_{Z_i}(u,\psi_0)}. \tag{3.3}$$

We approximate the null distribution of $n^{1/2}\hat{Q}(v)$ by properly perturbing each term in (3.1)$-$(3.3) to generate $\hat{\eta}_0^*(t,v)$ and $\bar{\eta}^*(t,v)$ such that conditional on the data, the distribution of $n^{1/2}Q^*(v) =$

$$\int_{-\infty}^{\tau^-} n^{1/2}\left\{t[\hat{\eta}_0^*(t,v)-\bar{\eta}^*(t,v)]d\hat{S}(t,\hat{\psi})+t\hat{S}(t,\hat{\psi})[\hat{\eta}_0^*(dt,v)-\bar{\eta}^*(dt,v)]\right\} \tag{3.4}$$
$$+n^{1/2}\tau\hat{S}(\tau^-,\hat{\psi})[\hat{\eta}_0^*(\tau^-,v)-\bar{\eta}^*(\tau^-,v)]$$

is the same as that of $\bar{Q}(v)$ in the limit. A similar technique for approximating the distributions of complex empirical processes in survival analysis has been utilizied by Lin, Wei, and Ying (1993), Lin, Fleming, and Wei (1994), Goldwasser, Tian, and Wei (2004), and León, Cai, and Wei (2009). In Appendix C, we provide the exact forms of $\hat{\eta}_0^*(t,v)$ and $\bar{\eta}^*(t,v)$. To approximate the p-value of the KS-type test W, (2.7), let w denote its observed value. The p-value, $\Pr(W\geq w)$, can be approximated by $\Pr(W^*\geq w \mid \text{Data})$, where $W^*=\sup_{v\in[\mathfrak{z}_l,\mathfrak{z}_r]}|n^{1/2}Q^*(v)|$. We estimate $\Pr(W^*\geq w)$ by generating a large number of realizations from $Q^*(\cdot)$. The p-value of the CvM-type test (2.8) can be approximated analogously.

4. Delete-k Cross-validation Bandwidth Selection

In this section we develop a cross-validation bandwidth selection procedure. Without censoring the kernel regression smooth of the *partial* residuals $T - \hat{b}'X$ estimates, up to a constant, $g(z)$ in model (1.2) (Cook (1993)). In developing our cumulative residual plots, we are concerned with choosing the bandwidth h that estimates g as closely as possible such that the integrated error is minimal. Here we propose a cross-validation procedure based on the censored *partial* residuals $\hat{r}_i = Y_i - \hat{b}'X_i$, $i = 1, \ldots, n$.

First consider the non-censored case and let $\{(\hat{r}_1, Z_1), \ldots, (\hat{r}_k, Z_k)\}$ denote k randomly selected $(\hat{r}, Z)$ pairs. For bandwidth h, let $\hat{g}_h(Z_j)$ denote the non-parametric kernel estimate of $E[\hat{r}|Z_j]$ based on the remaining $n-k$ pairs for $j = 1, \ldots, k$. Since $\hat{g}_h(z) \approx E[\hat{r}|Z = z] = g_0(z) = \alpha_0 + g(z)$ (Cook (1993)), we have $k^{-1}\sum_{j=1}^{k} I(Z_j \leq v)[\hat{r}_j - \hat{g}_h(Z_j)] \approx EI(Z \leq v)\{E[\hat{r}|Z] - \hat{g}_h(Z)\} = \int_{\mathfrak{z}l}^{v}[g_0(z) - \hat{g}_h(z)]dH(z)$. Thus, one may choose h that mimimizes an empirical version of $E[\Gamma(h)]$, where

$$\Gamma(h) = \int_{\mathfrak{z}l}^{\infty}\left\{\int_{\mathfrak{z}l}^{v} g_0(z)dH(z) - \int_{\mathfrak{z}l}^{v}\hat{g}_h(z)dH(z)\right\}^2 dw(v)$$

and $w(v)$ can be a trimming function, or $H(v)$, or a combination thereof. Here $\Gamma(h)$ compares the error in $\hat{g}_h$ in terms of its integrated error. In practice, we propose to obtain h as the minimizer of $M^{-1}\sum_{m=1}^{M}\tilde{\gamma}_m(h)$, where

$$\tilde{\gamma}_m(h) = \int_{\mathfrak{z}l}^{\infty}\left\{k^{-1}\sum_{j\in\mathcal{C}_m} I(Z_j \leq v)\left[\hat{r}_j - \hat{g}_h(Z_j)\right]\right\}^2 dw(v), \quad m = 1, \ldots, M, \quad (4.1)$$

and $\mathcal{C}_m$ denotes a collection of k randomly drawn $(\hat{r}, Z)$ pairs.

To incorporate censoring, we replace the term $k^{-1}\sum_{j\in\mathcal{C}_m} I(Z_j \leq v)[\hat{r}_j - \hat{g}_h(Z_j)]$ in (4.1) with the integrated Kaplan-Meier process

$$\hat{p}_m(v)\int_{-\infty}^{\tau_0} td[\tilde{S}_m(t) - \tilde{S}_{v,m}(t)], \quad (4.2)$$

where $\hat{p}_m(v) = k^{-1}\sum_{j\in\mathcal{C}_m} I(Z_j \leq v)$, $\tilde{S}_m(\cdot)$ is the Kaplan-Meier estimator based on all pairs $\{(\hat{r}_j - \hat{g}_h(Z_j), Z_j) : j \in \mathcal{C}_m\}$, $\tilde{S}_{v,m}(\cdot)$ is the Kaplan-Meier estimator based on pairs $\{(\hat{r}_j - \hat{g}_h(Z_j), Z_j) : Z_j \leq v, j \in \mathcal{C}_m\}$, and τ_0 denotes the largest non-censored partial residual. Note that $\hat{g}_h(Z_j) = -\int td\hat{S}_{Z_j}(t, \hat{\psi})$ where $\hat{S}_{Z_j}(\cdot, \hat{\psi})$, defined in Section 2.2, is based on the $n-k$ *partial* residual pairs not contained within the $\mathcal{C}_m$ sample (In addition, $\hat{S}_{Z_j}(\tau_m, \hat{\psi}) = 0$ where τ_m denotes the largest non-censored *partial* residual not contained within the $\mathcal{C}_m$ sample.). Subsequently, we propose to select h by minimizing $M^{-1}\sum_{m=1}^{M}\hat{\gamma}_m(h)$, where

$$\hat{\gamma}_m(h) = \int_{\mathfrak{z}l}^{\infty}\left\{\hat{p}_m(v)\int_{-\infty}^{\tau_0} td[\tilde{S}_m(t) - \tilde{S}_{v,m}(t)]\right\}^2 dw(v), \quad m = 1, \ldots, M, \quad (4.3)$$

with $\tilde{S}_m(\tau_0) - \tilde{S}_{v,m}(\tau_0) = 0$. For $w(v) = \hat{H}(v)$, the resulting process $\hat{\gamma}_m(h)$ can be viewed as a CvM goodness-of-fit process.

In applications the partial residuals can be more heavily censored in the left tail of the Z-covariate space. Consequently, the Kaplan-Meier estimator restricted to observations with $Z \leq v$, $\tilde{S}_{v,m}(\cdot)$, may be unreliable. Alternatively, let $\tilde{S}^{+}_{v,m}(\cdot)$ denote the Kaplan-Meier estimator based on pairs $(\hat{r}_j - \hat{g}_h(Z_j), Z_j)$, restricted to $Z_j \geq v$, $j \in \mathcal{C}_m$. Then, estimate the expectation of $\Gamma^{+}(h) = \int_{\mathfrak{z}l}^{\infty}\{\int_{v}^{\infty} g_0(z)dH(z) - \int_{v}^{\infty} \hat{g}_h(z)dH(z)\}^2 dw(v)$ by $M^{-1}\sum_{m=1}^{M} \hat{\gamma}^{+}_m(h)$ where

$$\hat{\gamma}^{+}_m(h) = \int_{\mathfrak{z}l}^{\infty} \left\{ \hat{p}^{+}_m(v) \int_{-\infty}^{\tau_0} td[\tilde{S}_m(t) - \tilde{S}^{+}_{v,m}(t)] \right\}^2 dw(v), \tag{4.4}$$

$\hat{p}^{+}_m(v) = k^{-1}\sum_{j\in\mathcal{C}_m} I(Z_j \geq v)$, and $\tilde{S}_m(\tau_0) - \tilde{S}^{+}_{v,m}(\tau_0) = 0$, $m = 1, \ldots, M$.

In the context of model selection for the non-censored linear regression model, Shao (1993) has shown that k (the "validaton" dataset) should be much larger than $n - k$ (the "construction" dataset). We therefore recommend that k be some integer larger that $n/2$.

As is common with goodness-of-fit tests based on smoothing procedures, an undersmoothed bandwidth is generally required to maintain the size. Based on our simulation experiments we recommend choosing the slightly undersmoothed bandwidth $h/n^{0.1}$.

5. Example and Simulation Study

In this section we apply our methods to the well-known PBC data set developed by the Mayo Clinic and listed in Appendix D of Fleming and Harrington (1991). For comparison, we also apply the omnibus test procedure proposed by Lopez and Patilea (2009). The data contains $n = 418$ patients with primary biliary cirrhosis (PBC), a fatal chronic liver disease. At the time of data collection, 161 patients had died. Approximately 61% of the survival times were censored. Recently, Cai, Tian, and Wei (2003) analyzed this data set using the Box-Cox transformation model

$$T = h_\lambda(T^0) = \beta' U + \epsilon, \tag{5.1}$$

where $\lambda = 0.102$, $h_\lambda(t) = (t^\lambda - 1)/\lambda$ if $\lambda \neq 0$, and $\log(t)$ if $\lambda = 0$, and U consists of age, log(albumin), log(protime), log(bilirubin), and edema.

For illustration we consider this model except with log(bilirubin) in (5.1) replaced by bilirubin. We standardize the covariates (subtracting the mean and dividing by sample standard deviation) and because of the heavy censoring, we apply the León, Cai, and Wei (2009) estimator using the 25^{th} quantile. The endpoint τ is chosen as the 90^{th} percentile of the non-censored residuals. In addition, the p-values of the test statistics and cross-validation curves are estimated from 1,000 monte carlo realizations.

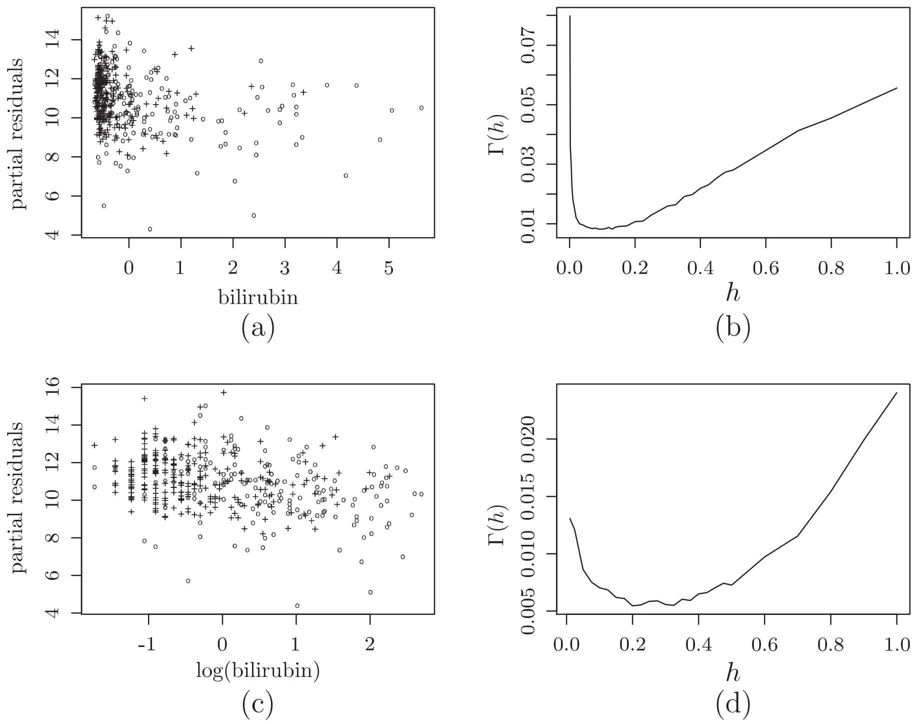

Figure 1. PBC data. (a) Partial residuals, $Y - \hat{\beta}'X$, for linear bilirubin model ((5.1) with log(bilirubin) replaced by bilirubin). Crosses denote censored points. (b) Cross-validation curve, via (4.4), for linear bilirubin model. (c) Partial residuals for log(bilirubin) model. Crosses denote censored points. (d) Cross-validation curve, via (4.4), for log(bilirubin) model.

Our first step is to choose a bandwidth according to the delete-k cross-validation procedure described in Section 5. For k we chose the integer value of $0.65\times n$. In Figure 1(a) we plot the partial residuals, $Y - \hat{\beta}'X$, that we note are heavily censored in the left tail of the covariate space and thus estimate the cross-validation curve via (4.4). Figure 1(b) displays the cross-validation curve where, in (4.4), $dw(v) = j(v)d\hat{H}(v)$ with $j(v)$ denoting a 2.5%-trimming function. The minimizer of the cross-validation curve, Figure 1(b), is approximately 0.13. We therefore chose the undersmoothed bandwidth $h = 0.13/n^{0.1}$. Figure 2(a) displays the residuals and Figure 2(b) displays the observed $\hat{Q}(\cdot)$ process along with 50 realizations of $Q^*(\cdot)$, (3.4). The observed process is extreme relative to the 50 simulated realizations and follows a logarithmic shape. In addition, the estimated p-values of both tests, (2.7) and (2.8), are < 0.001. The residual plots and functional form tests therefore indicate that the linear functional form for bilirubin is misspecified. In contrast, the omnibus test procedure

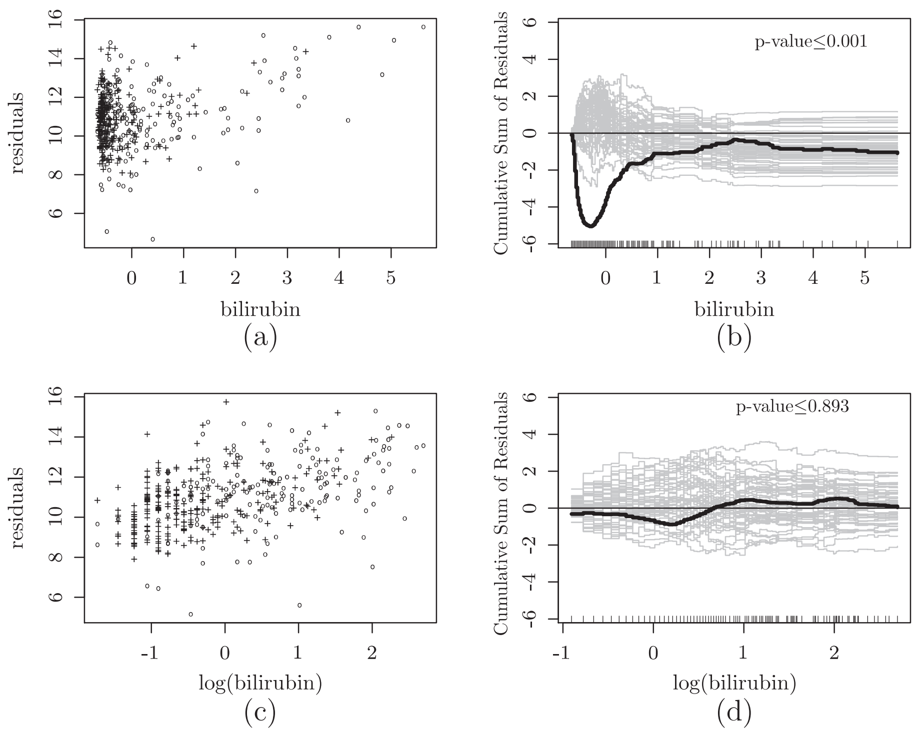

Figure 2. PBC data. (a) Residuals for linear bilirubin model ((5.1) with log(bilirubin) replaced by bilirubin), with crosses denoting censored observations. (b) Observed cumulative residual process, $\hat{q}(\cdot)$, denoted by solid curve, and 50 realizations of $Q^*(\cdot)$, (3.4). (c) Residuals for logarithmic bilirubin model, (5.1). (d) Observed cumulative residual process, $\hat{q}(\cdot)$, and 50 realizations of $Q^*(\cdot)$.

of Lopez and Patilea (2009) does not indicate lack-of-fit; the p-values of their "WLS" test range from 0.81 (for bandwidth=0.10) to 0.14 (for bandwidth=0.50).

We therefore consider the model with log(bilirubin), model (5.1), and choose $h = 0.20/n^{0.1}$ (Figure 1(c) displays the partial residuals and Figure 1(d) the cross-validation curve, (4.4), with k and dw chosen as above.). Note that Figure 2(c) displays the residuals which are heavily censored for log(bilirubin) values less than -1. We therefore estimate $\hat{Q}(v)$ for $v \geq -1$ (see Figure 2(d)). Here the observed process fluctuates around a constant and is not extreme relative to the 50 simulated realizations. The estimated p-values of both tests are > 0.80 indicating that the logarithmic functional form of bilirubin appears to be adequate. In addition the p-values for checking the other continuous covariates, age (bandwidth=0.425), log(albumin) (bandwidth=0.50), and log(protime) (bandwidth=0.70) are all > 0.90 indicating that the functional forms for these covariates appear to be adequate.

Note that this conclusion is consistent with the omnibus test of Lopez and Patilea (2009) where the p-values of their "WLS" test range from 0.91 (for bandwidth=0.10) to 0.08 (for bandwidth=0.5).

We have illustrated the methods for checking the functional form for the single covariate bilirubin. In general applications we recommend checking covariates in a one-at-a-time manner by first focusing on the covariate (if any) with a residual pattern that suggests a particular functional form. As a guideline, one can use the prototype residual plots provided in Figure 2 of Lin, Wei, and Ying (2002) which displays typical residual patterns under various forms of misspecification in the non-censored generalized linear model setting (we have found these to also be representative for the AFT model with censoring).

To investigate the performance of our procedures in moderate sample sizes we conducted a series of simulation experiments. Our key experiment was a design to mimic the PBC data set by simulating survival times from the model

$$T_i = 12.5 + 0.74{\times}X_i - 1.44{\times}\log(Z_i) + \epsilon_i, \quad i = 1, \ldots, 418, \tag{5.2}$$

where X and Z were fixed at the observed (standardized) values of log(albumin) and bilirubin, respectively, from the PBC data, and $\epsilon \sim N(0, 2)$. The parameters of (5.2) are obtained by fitting the corresponding Gaussian model to the PBC data (transformed via (5.1)). To mimic the observed censoring pattern of the PBC data we simulate censoring times from the observed K-M estimate, $1-\hat{G}(\cdot)$. We generate 1000 simulations and estimated the p-values of the KS and CvM-type tests, W and D, from 100 monte carlo realizations. The censoring proportion for the simulated datasets was about 64%. The powers of W and D in detecting misspecification of Z in the model $T = \beta' X + \gamma Z$ were 0.84 and 0.88, respectively. On the other hand, when the model is correctly specified the empirical sizes of W and D (based on $\hat{Q}(v)$, $v \geq -1$) were 0.023 and 0.034, respectively. Note that the bandwidths in the above simulations were fixed at $h = 0.10/n^{0.1}$ for the misspecified model and $h = 0.18/n^{0.1}$ for the correctly specified model. These bandwidth values were chosen based on applying the cross-validation procedure to 25 randomly selected datasets and then taking the mean of the resulting estimated bandwidths.

In sharp contrast to the performance of the W and D functional form tests, the power of the "WLS" omnibus test procedure of Lopez and Patilea (2009) was only 0.038, 0.053, and 0.052 for the bandwidths 0.01, 0.10, and 0.25, respectively. However when the censoring times were U(0,25), resulting in an average censoring of 50%, the power of the "WLS" omnibus test improved to 0.61 and the size was 0.04 (bandwidth=0.10).

We also investigated the performance of our methods using kernels other than the normal. In particular, we evaluated the powers for the tests based on the popular Epanechnikov kernel (Härdle and Marron (1985)) and the "twicing" kernel $K(u) = 2\exp(-u^2/2)/\sqrt{2\pi} - \exp(-u^2/4)/\sqrt{4\pi}$ (Newey, Hsieh, and Robins (2004)). The "twicing" kernels are higher-order kernels that have a small bias property in the context of semiparametric estimation (Newey, Hsieh, and Robins (2004)). For the Epanechnikov kernel (bandwidth=0.35) the powers of the W and D tests were approximately 0.78 and 0.84, whereas for the "twicing" kernel (bandwidth=0.25), the powers were approximately 0.70 and 0.80. Although "twicing" kernels and other higher-order kernels have theoretical advantages with respect to bias, it is not clear that such properties translate into advantages in the model checking setting. Though the simulations are limited in scope, we recommend that the normal kernel be used in applications.

In addition, we investigated the robustness of our methods to violation of the assumption that the censoring distribution is independent of covariates. Specifically, we generated covariate-dependent censoring times, C_i, via

$$C_i = 11.8 + 0.33 \times X_i + 0.32 \times L_i + \epsilon_i, \quad i = 1, \ldots, 418, \tag{5.3}$$

where X and L were fixed at the observed (standardized) values of log(albumin) and log(protime) from the PBC data, and $\epsilon \sim N(0, 0.97)$. The survival times T_i were generated according to (5.2). The censoring proportion for the simulated datasets was about 64%. The powers of W and D in detecting misspecification of Z were 0.80 and 0.85. When the model was correctly specified the empirical sizes of W and D (based on $\hat{Q}(v)$, $v \geq -1$) were 0.047 and 0.085. Though the D test was slightly anti-conservative, the methods performed well in this simulation setting under violation to the covariate-independent censoring assumption.

6. Remarks

For the censored linear regression model practically useful model checking techniques have not yet been well developed. The omnibus tests proposed by Lopez and Patilea (2009) can be sensitive to the choice of the bandwidth and an approach for selecting a bandwidth is not provided. Moreover, in the simulation experiments considered here we have found their "WLS" omnibus test to be sensitive to the degree of censoring, with poor performance under moderate censoring (e.g. $\geq 50\%$). In this paper we developed residual plots and goodness-of-fit statistics for assessing the functional form specifications. Our procedure is based on robust estimators of the parameters corresponding to correctly specified covariates. This robustness is crucial when covariates are correlated as procedures based on non robust estimators can lead to misleading residual plots and/or test statistics with low power.

Our model checking procedure depends on the quantile regression estimator of León, Cai, and Wei (2009) that requires the assumption that the censoring times are independent of the covariates (X, Z) with a common distribution G. The León, Cai, and Wei (2009) estimator is a generalization of the Ying, Jung, and Wei (1995) median estimator which relies on the same censoring assumptions. As noted by Ying, Jung, and Wei (1995), the assumption that the censoring times are independent of the covariates may be strong for some observational studies but is often satisfied in randomized clinical trials. The PBC data are from an observational study, and applying a Cox model to the censoring times suggests that the censoring times are associated with log(albumin) and log(protime) (p-values < 0.001). Ying, Jung, and Wei (1995) provide numerical evidence that their procedure can be robust to violations of the aforementioned assumptions on the censoring distribution. Since the estimator of León, Cai, and Wei (2009) is a generalization of the Ying, Jung, and Wei (1995) estimator it appears reasonable that such robustness properties may apply to theirs as well. This is indeed confirmed by the simulation results given in Section 5. Future work is required to develop alternative estimators based on replacing the unconditional K-M estimator of the censoring distribution $\hat{G}$ utilized in the León, Cai, and Wei (2009) estimator by its conditional K-M analog. For example, one may impose an AFT or Cox model relating the censoring distribution to the covariates and then replace $\hat{G}(\cdot)$ by an estimate of $P(C > \cdot|U_i)$ in the estimation of θ_ρ.

Our procedure requires the selection of a smoothing parameter that can be difficult to implement for small sample sizes. Another approach is to develop a cumulative residual process based on ρth quantile residuals. In addition, model checking techniques based on robust estimators for assessing the link function would also be useful to practitioners. We plan to investigate both of these topics in future work.

The computational load for implementing the cross-validation and resampling procedures is extensive compared to the computational requirements of Lopez and Patilea (2009). Computing code (R Development Core Team (2010)) is available from the first author, as well as strategies for reducing the computational load.

Appendix A. Consistency of the Test Statistics

Throughout, we assume that model (1.2) holds, where ϵ has a cumulative distribution function $F(x) = \text{pr}(\epsilon < x)$. The derivative functions of F and G are assumed to be continuous and uniformly bounded. The covariates X and

Z are assumed to be bounded and, without loss of generality, we assume that the support of Z is $[\mathfrak{z}_l, \mathfrak{z}_r]$ with $-\infty < \mathfrak{z}_l < \mathfrak{z}_r < \infty$. To ensure the validity of the kernel-smoothed estimators, we assume that the density function of Z, $\dot{H}(\cdot)$, has bounded continuous first and second derivatives that are also bounded away from 0. Here and in the sequel, for any function $\xi(\cdot)$ we write $\dot{\xi}(\cdot)$ to denote the derivative. The bandwidth $h \to 0$ with $n^{1/4}h = o_p(1)$. This corresponds to Assumption A of Dabrowska (1989). The kernel function K is a symmetric density vanishing outside $(-1, 1)$ and the total variation of K is bounded as in Assumption B of Dabrowska (1989). We note that, for theoretical derivations, K is typically assumed to have finite support, although in practical settings the Gaussian kernel appears to work well. Furthermore, we assume that there is a constant $\tilde{t}$ such that $P(Y > \tilde{t}) > 0$ and $P(\theta'_{\rho 0}U_i < \tilde{t}) = 1$, where $U_i = (1, X'_i, \kappa(Z_i)')'$ and $\theta_{\rho 0}$ is the solution to $E[U_i\{I(T_i \geq \theta' U_i) - (1-\rho)\}] = 0$. This assumption ensures the consistency of $\hat{b}$, as discussed in Ying, Jung, and Wei (1995) and León, Cai, and Wei (2009). We assume that $\bar{\psi}$ is an interior point of a compact set Ω. Throughout, unless specified otherwise, the supremum is taken over $(-\infty, \tau]$ for t, $[\mathfrak{z}_l, \mathfrak{z}_r]$ for z, and Ω for ψ.

We establish consistency of the KS-type test (2.7), based on $\hat{Q}(\cdot)$, under model (1.2) where $g(\cdot)$ is a non-linear function. Note that under model (1.2), $\hat{b}$ is consistent for β_0 while $\hat{c}$ converges in probability to some $\bar{c}$. Moreover, for $t \leq \tau$, $\hat{S}_z(t, \hat{\psi})$ is uniformly consistent for $\text{pr}\{\epsilon(z) \geq t \mid z\} = 1 - F\{t - \mathcal{R}(z)\}$, where $\epsilon(z) = \mathcal{R}(z) + \epsilon$ and $\mathcal{R}(z) = \alpha_0 + g(z) - \bar{c}z$. Therefore, $\hat{S}_z(\tau^-, \hat{\psi})$ converges to $1 - F\{\tau - \mathcal{R}(z)\}$ by the continuity of $F(\cdot)$ and $-\int_{-\infty}^{\tau^-} td\hat{S}_z(t, \hat{\psi})$ is consistent for the restricted mean, $m(z) = E[\epsilon(z)I(\epsilon(z) < \tau)]$. In addition, $-\int_{-\infty}^{\tau^-} td\hat{S}(t, \hat{\psi}) + \tau\hat{S}(\tau^-, \hat{\psi})$ converges to some constant, m_0, say.

Now, the consistency of the K-S test,

$$\sup_{v \in [\mathfrak{z}_l, \mathfrak{z}_r]} \left| \int_{\mathfrak{z}_l}^{v} \left[\int_{-\infty}^{\tau^-} td\left\{\hat{S}_z(t, \hat{\psi}) - \hat{S}(t, \hat{\psi})\right\} - \tau\left\{\hat{S}_z(\tau^-, \hat{\psi}) - \hat{S}(\tau^-, \hat{\psi})\right\} \right] d\hat{H}(z) \right|,$$

fails if

$$m(z) + \tau\{1 - F(\tau - \mathcal{R}(z))\} = m_0 \quad \text{for every } z \in [\mathfrak{z}_l, \mathfrak{z}_r].$$

That is, consistency fails if $\dot{m}(z) + \tau\dot{\mathcal{R}}(z)\dot{F}\{\tau - \mathcal{R}(z)\} = 0$ for every $z \in [\mathfrak{z}_l, \mathfrak{z}_r]$. Note that $m(z) = E[(\epsilon + \mathcal{R}(z))I(\epsilon \leq \tau - \mathcal{R}(z))] = \int_{-\infty}^{\tau - \mathcal{R}(z)} tdF(t) + \mathcal{R}(z)F(\tau - \mathcal{R}(z)) = \tau F(\tau - \mathcal{R}(z)) - \int_{-\infty}^{\tau} F(t - \mathcal{R}(z))dt$, and

$$\dot{m}(z) + \tau\dot{\mathcal{R}}(z)\dot{F}\{\tau - \mathcal{R}(z)\} = \dot{\mathcal{R}}(z)F(\tau - \mathcal{R}(z)).$$

Consequently, if $\dot{\mathcal{R}}(z) > 0$, then consistency fails if and only if $F(\tau - \mathcal{R}(z)) = 0$ for every $z \in [\mathfrak{z}_l, \mathfrak{z}_r]$. Therefore, consistency follows if we assume that ϵ has infinite support. The above arguments also establish consistency of the CvM-type test

$$\int_{\mathfrak{z}l}^{\infty}\left\{\int_{\mathfrak{z}l}^{v}\int_{-\infty}^{\tau} td[\hat{S}_z(t)-\hat{S}(t,\hat{\psi})]d\hat{H}(z)\right\}^2 d\hat{H}(v).$$

Appendix B. Derivation of Asymptotic Null Distribution of $\hat{Q}(\cdot)$

We now establish the asymptotic null distribution of

$$n^{1/2}\hat{Q}(v) = n^{1/2}\int_{\mathfrak{z}l}^{v}\Big\{\int_{-\infty}^{\tau^-} td[\hat{S}_z(t,\hat{\psi})-\hat{S}(t,\hat{\psi})]\Big\}d\hat{H}(z) \tag{B.1}$$

$$- n^{1/2}\tau\int_{\mathfrak{z}l}^{v}[\hat{S}_z(\tau^-,\hat{\psi})-\hat{S}(\tau^-,\hat{\psi})]d\hat{H}(z), \tag{B.2}$$

where τ^- is the point just prior to τ. We write $\Lambda(t)$ and $S(t)$ for $\Lambda(t,\psi_0)$ and $S(t,\psi_0)$, respectively. By the empirical process results of Lai and Ying (1988), the processes $\bar{N}(t,\psi)$ and $\hat{\pi}(t,\psi)$ converge in probability at rate $n^{1/2}$, uniformly in (t,ψ), to $A(t,\psi) = E\{N(t,\psi)\}$ and $\pi(t,\psi) = \mathrm{pr}\{e_i(\psi) \geq t\}$, respectively. Moreover, by arguments as in Horowitz (1996) it can be shown that $\sup_{t,z,\psi}|\bar{N}_z(t,\psi)-A_z(t,\psi)| = o_p(n^{-1/4})$ and $\sup_{t,z,\psi}|\hat{\pi}_z(t,\psi)-\pi_z(t,\psi)| = o_p(n^{-1/4})$, where $A_z(t,\psi) = E\{N_i(t,\psi) \mid Z_i = z\}$ and $\pi_z(t,\psi) = \mathrm{pr}\{e_i(\psi) \geq t \mid Z_i = z\}$. It then follows from Lemma (A.3) of Bilias, Gu, and Ying (1997) that

$$\sup_{t,z,\psi}\left|\hat{\Lambda}_z(t,\psi)-A_z(t,\psi)\right| = o_p(n^{-1/4}), \quad \sup_{t,\psi}\left|\hat{\Lambda}(t,\psi)-A(t,\psi)\right| = o_p(n^{-1/4}).$$

Now, the process $\sqrt{n}(\hat{H}(z)-H(z))$ converges weakly to a mean zero Gaussian process. Applying these results and Lemma (A.3) of Bilias, Gu, and Ying (1997) we can show that, under correct specification, (B.1) is asymptotically equivalent to

$$\begin{aligned}
&n^{1/2}\int_{\mathfrak{z}l}^{v}\int_{-\infty}^{\tau^-} t\left\{de^{-\hat{\Lambda}_z(t,\hat{\psi})}-de^{-\hat{\Lambda}(t,\hat{\psi})}\right\}dH(z)\\
&= n^{1/2}\int_{\mathfrak{z}l}^{v}\int_{-\infty}^{\tau^-} tS(t)\Big[\{e^{-\hat{\Lambda}(t,\hat{\psi})+\Lambda(t)}-1\}d\hat{\Lambda}(t,\hat{\psi})-\{e^{-\hat{\Lambda}_z(t,\hat{\psi})+\Lambda(t)}-1\}d\hat{\Lambda}_z(t,\hat{\psi})\\
&\qquad\qquad - d\{\hat{\Lambda}_z(t,\hat{\psi})-\hat{\Lambda}(t,\hat{\psi})\}\Big]dH(z)\\
&\approx -n^{1/2}\int_{\mathfrak{z}l}^{v}\int_{-\infty}^{\tau^-}\left\{t\hat{\eta}(t,v,\hat{\psi})dS(t)+tS(t)\hat{\eta}(dt,v,\hat{\psi})\right\}dH(z),
\end{aligned} \tag{B.3}$$

where $\hat{\eta}(t,v,\psi) = \int_{\mathfrak{z}l}^{v}\{\hat{\Lambda}_z(t,\psi)-\hat{\Lambda}(t,\psi)\}dH(z)$. In addition, using similar arguments we have

$$\text{(B.2)} \approx n^{1/2}\tau S(\tau^-)\hat{\eta}(\tau-,v,\hat{\psi}), \tag{B.4}$$

and thus $n^{1/2}\hat{Q}(v) \approx$ (B.3) + (B.4). Now, we write $\hat{\eta}(t,v,\psi) = \eta_0(t,v,\psi) + \hat{\eta}_1(t,v,\psi) - \hat{\eta}_2(t,v,\psi)$, where $\eta_0(t,v,\psi) = \int_{\mathfrak{z}l}^{v}\{\Lambda_z(t,\psi) - \Lambda(t,\psi)\}dH(z)$ and

$$\hat{\eta}_1(t,v,\psi) = \int_{\mathfrak{z}l}^{v} \left\{\hat{\Lambda}_z(t,\psi) - \Lambda_z(t,\psi)\right\} dH(z),$$
$$\hat{\eta}_2(t,v,\psi) = \int_{\mathfrak{z}l}^{v} \left\{\hat{\Lambda}(t,\psi) - \Lambda(t,\psi)\right\} dH(z).$$

It follows from (B.3) and (B.4) that $n^{1/2}\hat{Q}(v)$ is asymptotically equivalent to

$$n^{1/2}\int_{-\infty}^{\tau^-} t\{\hat{\eta}_2(t,v,\psi)-\hat{\eta}_1(t,v,\psi)\}dS(t)+n^{1/2}\int_{-\infty}^{\tau^-} tS(t)\{\hat{\eta}_2(dt,v,\psi)-\hat{\eta}_1(dt,v,\psi)\} \tag{B.5}$$

$$-n^{1/2}\int_{-\infty}^{\tau^-}\left\{t\eta_0(t,v,\hat{\psi})dS(t) + tS(t)\eta_0(dt,v,\hat{\psi})\right\} \tag{B.6}$$

$$+n^{1/2}\tau S(\tau^-)\hat{\eta}(\tau^-,v,\hat{\psi}). \tag{B.7}$$

We show that each of these terms can be approximated by a sum of independent and identically distributed terms. To this end, we first obtain expansions for $n^{1/2}\hat{\eta}_1(t,v,\psi)$ and $n^{1/2}\hat{\eta}_2(t,v,\psi)$. By the uniform convergence of $\bar{N}_z(u,\psi) \to A_z(u,\psi)$ and $\hat{\pi}_z(u,\psi) \to \pi_z(u,\psi)$, we have

$$n^{1/2}\hat{\eta}_1(t,v,\psi) \approx n^{1/2}\int_{\mathfrak{z}l}^{v}\int_{-\infty}^{t}\left\{\frac{\bar{N}_z(du,\psi)}{\pi_z(u,\psi)} - \frac{\hat{\pi}_z(u,\psi)}{\pi_z(u,\psi)^2}A_z(du,\psi)\right\}dH(z).$$

By Taylor expansion and a change in the order of integration, the first term in the last display can be written as

$$n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} I(z\le v)K_h(Z_i - z)\left\{\int_{-\infty}^{t}\frac{N_i(du,\psi)}{\pi_z(u,\psi)}\right\}\dot{H}(z)dz$$
$$= n^{-1/2}\sum_{i=1}^{n} N_i(t,\psi)\frac{I(Z_i\le v)\dot{H}(Z_i)}{\pi_{Z_i}(e_i(\psi),\psi)} + O_p(n^{1/2}h^2), \quad \text{as} \quad h\to 0.$$

Similarly, the second term in the aforementioned display can be written as

$$n^{-1/2}\sum_{i=1}^{n} I(e_i(\psi)\ge t)I(Z_i\le v)\dot{H}(Z_i)\frac{\dot{A}_{Z_i}(e_i(\psi),\psi)}{\pi_{Z_i}(e_i(\psi),\psi)^2} + O_p(n^{1/2}h^2), \quad \text{as} \quad h\to 0,$$

where $\dot{A}_z(t,\psi) = \partial A_z(t,\psi)/\partial t$. Therefore, under the assumption that $n^{1/4}h = o_p(1)$,

$$n^{1/2}\hat{\eta}_1(t,v,\psi)\approx n^{-1/2}\sum_{i=1}^{n} I(Z_i\le v)\dot{H}(Z_i)\int_{-\infty}^{t}\frac{N_i(du,\psi)-I(e_i(\psi)\ge u)\Lambda_{Z_i}(du,\psi)}{\pi_{Z_i}(u,\psi)}. \tag{B.8}$$

By analogous arguments, $n^{1/2}\hat{\eta}_2(t,v,\psi) = n^{1/2}H(v)\{\hat{\Lambda}(t,\psi)-\Lambda(t,\psi)\}$ is asymptotically equivalent to

$$n^{-1/2}H(v)\sum_{i=1}^{n}\int_{-\infty}^{t}\frac{N_i(du,\psi)-I(e_i(\psi)\geq u)\Lambda(du,\psi)}{\pi(u,\psi)}. \tag{B.9}$$

From (B.8) and (B.9), applying the Functional Central Limit Theorem (Pollard (1990)), it follows that the processes $n^{1/2}\hat{\eta}_1$ and $n^{1/2}\hat{\eta}_2$ converge weakly to Gaussian processes in (t,v,ψ), and are thus equicontinuous in ψ. Therefore, (B.5) is asymptotically equivalent to

$$\int_{-\infty}^{\tau^-} n^{1/2}\left\{t\bar{\eta}(t,v,\psi_0)dS(t)+tS(t)\bar{\eta}(dt,v,\psi_0)\right\}, \tag{B.10}$$

where $n^{1/2}\bar{\eta}(t,v,\psi_0) =$

$$n^{-1/2}\sum_{i=1}^{n}\left\{H(v)\int_{-\infty}^{t}\frac{N_i(du,\psi_0)-I(e_i(\psi_0)\geq u)\Lambda(du,\psi_0)}{\pi(u,\psi_0)}\right.$$
$$\left. -I(Z_i\leq v)\dot{H}(Z_i)\int_{-\infty}^{t}\frac{N_i(du,\psi_0)-I(e_i(\psi_0)\geq u)\Lambda_{Z_i}(du,\psi_0)}{\pi_{Z_i}(u,\psi_0)}\right\}.$$

The expansion for (B.6) follows directly from a Taylor series expansion:

$$\text{(B.6)} \approx -n^{1/2}(\hat{\psi}-\psi_0)\int_{-\infty}^{\tau^-}\left\{ta(t,v,\psi_0)dS(t)+tS(t)a(dt,v,\psi_0)\right\}, \tag{B.11}$$

where $a(t,v,\psi) = \partial\eta_0(t,v,\psi)/\partial\psi$. Lastly, by similar arguments,

$$\text{(B.7)} \approx \tau S(\tau^-)\left\{n^{1/2}(\hat{\psi}-\psi_0)a(\tau^-,v,\psi_0) - n^{1/2}\bar{\eta}(\tau^-,v,\psi_0)\right\}. \tag{B.12}$$

Combining (B.10), (B.11) and (B.12), we obtain $\hat{Q}(v)\approx$

$$n^{1/2}\left[\int_{-\infty}^{\tau^-}\left\{t\bar{\eta}(t,v,\psi_0)dS(t)+tS(t)\bar{\eta}(dt,v,\psi_0)\right\}-\tau S(\tau^-)n^{1/2}\bar{\eta}(\tau^-,v,\psi_0)\right]$$
$$-n^{1/2}(\hat{\psi}-\psi_0)\left[\int_{-\infty}^{\tau^-}\{ta(t,v,\psi_0)dS(t)+tS(t)a(dt,v,\psi_0)\}-\tau S(\tau^-)a(\tau^-,v,\psi_0)\right].$$

This, combined with a Functional Central Limit Theorem and the asymptotic linear expansion for $n^{1/2}(\hat{\psi}-\psi_0)$ given in León, Cai, and Wei (2009), implies that $\hat{Q}(v)$ converges weakly to a zero-mean Gaussian process.

Appendix C. Resampling Procedures

To find expressions for $\hat{\eta}_0^*$ and $\bar{\eta}^*$, let $\{V_i, i = 1, \ldots, n\}$ be a random sample from a population with mean 0 and variance one that is independent of the data $\{(Y_i, \Delta_i, X_i, Z_i), i = 1, \ldots, n\}$. First, applying the resampling technique of León, Cai, and Wei (2009) and Jin et al. (2003), it can be shown that (3.1) is approximated by $n^{1/2}\hat{\eta}_0^*(t, v) =$

$$n^{1/2}\int_{\mathfrak{z}l}^{v}\Big\{[\hat{\Lambda}_z(t,\psi^*) - \hat{\Lambda}_z(t,\hat{\psi})] - [\hat{\Lambda}(t,\psi^*) - \hat{\Lambda}(t,\hat{\psi})]\Big\}\, d\hat{H}(z), \tag{C.1}$$

where, conditional on the data, $n^{1/2}(\psi^* - \hat{\psi})$ has the same asymptotic distribution as $n^{1/2}(\hat{\psi} - \psi_0)$, $\psi^* = (b^{*\prime}, \gamma^*)'$ generated as follows. For b^*, recall that $\hat{b} = \hat{b}_\rho$ denotes the estimate corresponding to the parameter b in θ. We therefore wish to approximate the distribution of $n^{1/2}(\hat{\theta} - \theta_0)$, where $\theta_0 = (\bar{a}, \beta_0', \bar{c})$ (the limiting value of $\hat{\theta}_\rho$ under model (1.2)). León, Cai, and Wei (2009) showed that the distribution of $n^{1/2}(\hat{\theta} - \theta_0)$ can be approximated by the conditional distribution of $n^{-1/2}(\theta^* - \hat{\theta})$ given the data, where $\theta^* = \theta_\rho^*$ is the solution to

$$n^{-1/2}\Xi_\rho(\theta^*) = B_\rho^*,$$

$$\begin{aligned}
B_\rho^* = n^{-1/2}&\sum_{i=1}^{n}\hat{U}_i\Big\{\frac{I(Y_i - \hat{\theta}_\rho'\hat{U}_i \ge 0)}{G^*(\hat{\theta}_\rho'\hat{U}_i)} - \frac{I(Y_i - \hat{\theta}_\rho'\hat{U}_i \ge 0)}{\hat{G}(\hat{\theta}_\rho'\hat{U}_i)}\Big\}\\
&+n^{-1/2}\sum_{i=1}^{n}\hat{U}_i\Big\{\frac{I(Y_i - \hat{\theta}_\rho'\hat{U}_i \ge 0)}{\hat{G}(\hat{\theta}_\rho'\hat{U}_i)} - (1-\rho)\Big\}V_i\\
&+n^{-1/2}\sum_{i=1}^{n}\hat{U}_i\Big\{\frac{I(Y_i - \hat{\theta}_\rho'U_i^* \ge 0)}{\hat{G}(\hat{\theta}_\rho'U_i^*)} - \frac{I(Y_i - \hat{\theta}_\rho'\hat{U}_i \ge 0)}{\hat{G}(\hat{\theta}_\rho'\hat{U}_i)}\Big\}\\
&+n^{-1/2}\sum_{i=1}^{n}(U_i^* - \hat{U}_i)\Big\{\frac{I(Y_i - \hat{\theta}_\rho'\hat{U}_i \ge 0)}{\hat{G}(\hat{\theta}_\rho'\hat{U}_i)} - (1-\rho)\Big\},\\
U_i^* = (1, X_i'&, \kappa^*(Z_i))', \quad G^*(t) = \hat{G}(t)\Big[1 - \sum_{i=1}^{n}\Big\{\int_{-\infty}^{t}\frac{dM_i^C(s)}{\sum_{j=1}^{n}I(Y_j \ge s)}\Big\}V_i\Big],
\end{aligned}$$

$$\kappa^*(z) = \hat{\kappa}(z) + n^{-1}\sum_{i=1}^{n}\frac{K_\zeta(Z_i - z)(X_i - \hat{\kappa}(Z_i))V_i}{n^{-1}\sum_{j=1}^{n}K_\zeta(Z_j - z)},$$

$\hat{M}_i^C(s) = I(Y_i \le s)(1 - \Delta_i) - \int_{-\infty}^{s} I(Y_i \ge u)d\hat{\Lambda}^C(u)$, and $\hat{\Lambda}^C(\cdot)$ is the Nelson-Aalen estimate for the cumulative hazard function of the censoring variable C. The realized b^* is then defined as the component of θ^* corresponding to b. To obtain γ^*, we apply the techniques of Jin et al. (2003), and define

$$L^*(\psi) = \sum_{i=1}^{n}\sum_{j=1}^{n} \Delta_i\{e_i(\psi) - e_j(\psi)\}^-(V_i + 1).$$

Let $\tilde{\psi}^*$ be a minimizer of $L^*(\psi)$ and take γ^* as the component of $\tilde{\psi}^*$ corresponding to γ in ψ. It follows from Parzen, Wei, and Ying (1994) that the distribution of $n^{1/2}\hat{\eta}_0(t, v, \psi_0)$ can be approximated by the conditional distribution of $n^{1/2}\hat{\eta}_0^*(t, v)$.

Next, we can show that the term $n^{1/2}\bar{\eta}(t, v, \psi_0)$ defined by (3.2) and (3.3) can be approximated by

$$\begin{aligned} n^{1/2}\bar{\eta}^*(t, v) = n^{-1/2}\hat{H}(v)&\sum_{i=1}^{n}\Big\{\int_{-\infty}^{t} \frac{d\hat{M}_i(u, \hat{\psi})}{\hat{\pi}(u, \hat{\psi})}\Big\}V_i \\ &- n^{-1/2}\sum_{i=1}^{n}\Big\{\int_{-\infty}^{t} \frac{d\hat{M}_{Z_i}(u, \hat{\psi})}{\hat{\pi}_{Z_i}(u, \hat{\psi})}\Big\}V_i I(Z_i \le z)\hat{f}_Z(Z_i), \end{aligned} \qquad \text{(C.2)}$$

where $\hat{M}_i(t, \hat{\psi}) = N_i(t, \hat{\psi}) - \int_{-\infty}^{t} I(e_i(\hat{\psi}) \ge u)\hat{\Lambda}(du, \hat{\psi})$, $\hat{M}_{Z_i}(t, \hat{\psi}) = N_i(t, \hat{\psi}) - \int_{-\infty}^{t} I(e_i(\hat{\psi}) \ge u)\hat{\Lambda}_{Z_i}(du, \hat{\psi})$ and $\hat{f}_Z(\cdot) = \sum_{j=1}^{n} K_h(Z_j - \cdot)$.

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer Verlag.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Department of Statistics, University of California, Berkeley.

Berk, K. N. and Booth, D. E. (1995). Seeing a curve in multiple regression. *Technometrics* **37**, 385-398.

Bilias, Y.,Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.* **25**, 662-682.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**(3), 429.

Cai, T., Tian, L. and Wei, L. (2003). Semi-parametric box-cox power tranformation models for censored survival observations. Technical report, Harvard University Biostatistics Working Paper Series.

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

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

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.* **17**, 1157-1167.

DiRienzo, A. G. and Lagakos, S. W. (2001). Effects of model misspecification on tests of no randomized treatment effect arising from Cox's proportional hazards model. *J. Roy. Statist. Soc.* **63**, 745-757.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **76**, 312-319.

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

Gail, M. H., Wieand, S. and 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. and Wei, L. (2004). Statistical inference for infinite-dimensional parameters via asymptotically pivotal estimating functions. *Biometrika* **91**, 81-94.

Grambsch, P. M., Therneau, T. M. and Fleming, T. R. (1995). Diagnostic plots to reveal functional form for covariates in multiplicative intensity models. *Biometrics* **51**, 1469-1482.

Härdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13**, 1465-1481.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable (STMA V37 2604). *Econometrica* **64**, 103-137.

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

Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276-1288.

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

Lai, T. L. and Ying, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *J. Multivariate Anal.* **27**, 334-358.

León, L., Cai, T. and Wei, L. (2009). Robust inferences for covariate effects on survival time with censored linear regression models. *Statist. Biosciences*, 1-15.

León, L. F. and Tsai, C.-L. (2004). Functional form diagnostics for Cox's proportional hazards model. *Biometrics* **60**, 75-84.

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

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

Lin, D. Y., Wei, L. J. and Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58**, 1-12.

Lopez, O. and Patilea, V. (2009). Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. *J. Multivariate Anal.* **100**, 210-230.

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

Newey, W. K., Hsieh, F. and Robins, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica* **72**, 947-962.

Parzen, M. I., Wei, L. J. and 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.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303-328.

Shao, J. (1993). Linear model selection by cross-validation. *J. Amer. Statist. Assoc.* **88**, 486-494.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363-369.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613-641.

Stute, W., Gonzalez Manteiga, W., González Manteiga, W. and Presedo Quindimil, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93**, 141-149.

Stute, W. and Zhu, L.-X. (2002). Model checks for generalized linear models. *Scand. J. Statist.* **29**, 535-545.

Su, J. Q. and Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *J. Amer. Statist. Assoc.* **86**, 420-426.

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

Wei, L. (1992). The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statist. Medicine* **11**, 1871-1871.

Wei, L. J., Ying, Z. and Lin, D. Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845-851.

Ying, Z., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90**, 178-184.

Department of Biostatistics / Health Outcomes and Payer Support, Genentech, South San Francisco, CA 94080, U.S.A.

E-mail: larry.leon.05@post.harvard.edu

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

E-mail: tcai@hsph.harvard.edu

(Received May 2010; accepted May 2011)