1. Introduction

The Cox proportional hazards model is the preeminent regression model when the response variable is survival time and is subject to possible right censoring (Cox 1972). The Cox model is specified through the conditional hazard function

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\exp[\boldsymbol{\gamma}_0^T\mathbf{x}]$$

where $\mathbf{x}$ is the covariate vector, $\boldsymbol{\gamma}_0$ is the true Cox model log relative risk parameter, $\lambda_0(t)$ is the unknown baseline hazard function, and $\exp[\boldsymbol{\gamma}_0^T\mathbf{x}]$ represents the covariate specific relative risk. Due to right censoring, the observed data for this model are (y, δ, x) where y is the minimum of the failure time and the censoring time, and δ is the censoring indicator, with $\delta = 1$ signifying the failure time is smaller.

Estimation and inference of the regression parameter is based on the score function,

$$U_n(\boldsymbol{\gamma}) = \sum_i \delta_i \left\{ \mathbf{x}_i - \frac{\sum_j \mathbf{x}_j I(y_j \geq y_i)\exp[\boldsymbol{\gamma}^T\mathbf{x}_j]}{\sum_j I(y_j \geq y_i)\exp[\boldsymbol{\gamma}^T\mathbf{x}_j]} \right\}$$

derived from the partial likelihood and the proportional hazards assumption. The score function has mean zero when evaluated at $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. It is monotone and continuous in $\boldsymbol{\gamma}$, enabling stable numerical algorithms to be employed and simplifying the asymptotic derivations for the properties of the parameter estimate $\hat{\boldsymbol{\gamma}}$, computed as the zero solution to the score equation $U_n(\boldsymbol{\gamma}) = 0$.

In addition to its numerical stability, the partial likelihood estimate $\hat{\boldsymbol{\gamma}}$ attains a semiparametric efficiency bound when the proportional hazards assumption is satisfied. When this assumption is violated, however, application of the Cox model can produce inconsistent estimates of the relative risk parameter and the asymptotic variance of the relative risk estimate. In this case, the proportional hazards model is likely to lead to an incorrect conclusion regarding the relationship between covariates and survival time. As a result, alternative approaches to modeling survival time are needed.

The most common alternative to the proportional hazards model is the accelerated failure time model

$$\log\, t_i = \boldsymbol{\beta}_0^T \mathbf{x}_i + \epsilon_i \qquad i = 1, \ldots, n$$

where the stochastic errors $\{\epsilon_i\}$ are independent identically distributed with unknown distribution function F and the covariate vector $\mathbf{x}_i$ is independent of ϵ_i. Since F is unknown, an estimating equation is a natural approach for estimation and inference on $\boldsymbol{\beta}_0$. However, the log survival time in the regression residual $\epsilon^{\boldsymbol{\beta}} = \log t - \boldsymbol{\beta}^T \mathbf{x}$ is an indication that estimation and inference is sensitive to small failure times. Rank regression is one approach to regain robustness with respect to the outlying log survival times.

Tsiatis (1990) developed an estimating equation based on a nonparametric family of linear rank tests, with the observed survival times in the logrank statistic replaced by the observed residuals $r^{\boldsymbol{\beta}} = \log y - \boldsymbol{\beta}^T \mathbf{x}$. The regression estimate $\hat{\boldsymbol{\beta}}$ is determined as the zero solution, or due to the discontinuity, the zero crossing of the estimating equation

$$Q_n(\boldsymbol{\beta}) = n^{-1/2} \sum_i \delta_i \left\{ \mathbf{x}_i - \frac{\sum_j \mathbf{x}_j I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})}{\sum_j I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})} \right\},$$

and when $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, this rank estimating function is asymptotically normal with mean zero. There are clear similarities between this rank estimating function and the proportional hazards score function. By replacement of the observed survival times with the observed residuals ($r^{\boldsymbol{\beta}}$), Tsiatis is able to replace the likelihood assumption with less restrictive moment assumptions.

Tsiatis (1990) and Ying (1993) extended this result to a general class of weighted rank estimating functions,

$$Q_n(\boldsymbol{\beta}; w) = n^{-1/2} \sum_i \delta_i w(r_i^{\boldsymbol{\beta}}) \left\{ \mathbf{x}_i - \frac{\sum_j \mathbf{x}_j I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})}{\sum_j I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})} \right\} \tag{1}$$

with weights $w(r)$ chosen to increase the efficiency of the estimate $\hat{\boldsymbol{\beta}}$. The weight function $\lambda'(r)/\lambda(r)$ produces an asymptotically efficient regression estimate. The selection of a weight function based on an efficiency criteria is problematic, since it requires knowledge of the underlying distribution function F, information that is assumed unknown at the outset. Setting $w(r_i) = 1$ is asymptotically efficient when the underlying error distribution is extreme value.

Instead of using the weight function to maximize efficiency, Fygenson and Ritov (1994) selected the weight $w(r_i^{\boldsymbol{\beta}}) = n^{-1} \sum_j I(r_j^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})$ to produce a monotone rank estimating function. The monotonicity with respect to $\boldsymbol{\beta}$ insures a unique solution to the estimating function. Jin et al. (2003) used an approximation to the weighted rank estimating function (1) to create a monotone estimating function, with weight function equal to

$$w(r_i^{\boldsymbol{\beta}}) = \frac{\psi(r_i^{\tilde{\boldsymbol{\beta}}}) \sum_k I(r_k^{\tilde{\boldsymbol{\beta}}} > r_i^{\tilde{\boldsymbol{\beta}}})}{\sum_k I(r_k^{\boldsymbol{\beta}} \geq r_i^{\boldsymbol{\beta}})}$$

where ψ is chosen by the user and $\tilde{\boldsymbol{\beta}}$ is an external consistent estimate of $\boldsymbol{\beta}$.

These modifications do not alter the discontinuity in the rank estimating function with respect to $\boldsymbol{\beta}$. The problem with the discontinuous estimating function is manifested in the Taylor series expansion of $Q_n(\boldsymbol{\beta}; w)$ in a $n^{1/2}$ neighborhood around $\boldsymbol{\beta}_0$

$$Q_n(\boldsymbol{\beta}; w) = Q_n(\boldsymbol{\beta}_0; w) + D_n(\boldsymbol{\beta}_0; w)(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$$

and computation of $D_n(\boldsymbol{\beta}; w)$ requires differentiation of the discontinuous rank estimating function with respect to $\boldsymbol{\beta}$. To enable differentiation, Tsiatis proposed using

$$\tilde{D}_n(\beta; w) = \frac{\partial}{\partial \beta} E[Q_n(\beta; w)]$$

in the Taylor expansion, employing kernel smoothing to estimate the residual density function, a byproduct of the expectation operator. The uncertainty of the accuracy of density estimation in the presence of right censored data has been raised by Hess et al. (1999).

In this paper, a monotone and continuous weighted rank estimating equation is developed. The weights are chosen to produce a regression estimate that is robust against outliers in the covariate space, and since the ranks of the residuals bound the effect of outlying survival times, the resulting regression estimate has bounded influence. In Section 2, a smoothed rank estimating equation is created that is continuous and monotone in $\boldsymbol{\beta}$. The asymptotic properties of the resultant regression estimate are developed. In Section 3, a weight function is introduced to add bounded influence to the properties of the regression estimate. An analysis of lung cancer data is undertaken in Section 4 to illustrate the proposed methodology. In Section 5, simulations are performed to examine the finite sample adequacy of the parameter estimate and coverage based on asymptotic confidence intervals. Section 6 contains concluding remarks.

2. A smooth rank based estimating equation

The rank estimating equation for censored data proposed by Fygenson and Ritov

(1994) is

$$\tilde{S}_n(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i (\mathbf{x}_i - \mathbf{x}_j)[1 - I(r_i^{\boldsymbol{\beta}} > r_j^{\boldsymbol{\beta}})].$$

The estimating function is a U-statistic with degree 2, is asymptotically normal, and when evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, has expectation zero. In addition, this estimating function is monotone in $\boldsymbol{\beta}$. The monotonicity insures that a unique $\boldsymbol{\beta}_0$ exists to produce a zero solution to the population estimating equation. From a data analytic standpoint, the lack of monotonicity may produce multiple zero solutions to the estimating equation.

The Fygenson and Ritov (1994) estimating function is not continuous in $\boldsymbol{\beta}$. The discontinuity, stemming from the indicator function $I(r_i^{\boldsymbol{\beta}} > r_j^{\boldsymbol{\beta}})$, presents a challenge for the derivation of the asymptotic distribution of $\hat{\boldsymbol{\beta}}$. A smooth approximation to the indicator function is the local distribution function $\Phi((r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}})/h)$, where the scale parameter h, termed the bandwidth in the smoothing literature, converges to zero as the sample size increases. Note that if $r_i > r_j$, $\Phi((r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}})/h) \rightarrow 1$ as n gets large, whereas if $r_i < r_j$, $\Phi((r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}})/h) \rightarrow 0$.

Thus, a smooth approximation to the Fygenson and Ritov (1994) monotone rank estimating function is

$$S_n(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i (\mathbf{x}_i - \mathbf{x}_j) \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h}\right)\right]. \qquad (2)$$

It is demonstrated in the theorem below, that by choosing the bandwidth h to converge to zero at a sufficient rate, the local distribution function Φ may be substituted for the indicator function, without changing the asymptotic distribution of the estimating function. As a result, U-statistic theory is used to derive its asymptotic distribution and a Taylor series expansion is applied to derive the asymptotic distribution for the regression estimate $\hat{\boldsymbol{\beta}}$, the zero solution to the estimating equation in

(2). The following conditions are required for the proof of the theorem; the proof is presented in the appendix.

(C1) The parameter vector $\boldsymbol{\beta}$ lies in a p dimensional bounded rectangle $\boldsymbol{\mathcal{B}}$, and for the covariate vector, $E(\mathbf{x}\mathbf{x}^T) < M < \infty$.

(C2) $n^{-1/2}S_n(\boldsymbol{\beta})$ has a bounded first derivative $n^{-1/2}A_n(\boldsymbol{\beta})$ in a compact neighborhood of $\boldsymbol{\beta}_0$, with $n^{-1/2}A_n(\boldsymbol{\beta})$ nonzero in that neighborhood.

(C3) The local distribution function $\Phi(z)$ is continuous and its derivative $\phi(z) = \partial\Phi(z)/\partial z$, is symmetric about zero with $\int z^2\phi(z) < \infty$.

(C4) The bandwidth h is chosen such as $n \to \infty$, $nh \to \infty$ and $nh^4 \to 0$.

Theorem 1: For the accelerated failure time model, under conditions C1-C4, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}VA^{-T})$, where $A = \lim_{n\to\infty} E\{n^{-1/2}\partial S_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ and $V = \lim_{n\to\infty} n^{-1}\ \mathrm{var}\{S_n(\boldsymbol{\beta}_0)\}$.

In practice, the bandwidth h can be set equal to $\hat{\sigma}n^{-0.26}$, where the estimate $\hat{\sigma}$ is the sample standard deviation of the uncensored residuals $r^{\hat{\boldsymbol{\beta}}}$. The exponent -0.26 provides the quickest rate of convergence while satisfying the bandwidth constraint $nh^4 \to 0$. This use of smoothing is distinguished from density estimation needed to approximate the expectation operator applied to the second derivative matrix (Tsiatis 1990). Here, smoothing is used at an earlier level, to approximate the observable, but discontinuous, estimating function. At this level, there are no expectation operators and only the regression residuals ($r^{\boldsymbol{\beta}}$) are required for smoothing, stemming concern for smoothing in a multidimensional space.

Computation of the estimate $\hat{\boldsymbol{\beta}}$ is straightforward and may be accomplished through the standard Newton-Raphson iteration. To incorporate the bandwidth h in the algorithm, an initial estimate $\hat{\boldsymbol{\beta}}_I$, based on the Fygenson and Ritov estimating function, can be used to compute the standard deviation of $r^{\hat{\boldsymbol{\beta}}}$. With h determined, Newton-Raphson is performed until convergence. The (l, m) element of the second derivative matrix is

$$A_{n(l,m)}(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i h^{-1} (x_{il} - x_{jl})(x_{im} - x_{jm}) \phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h}\right)$$

where $\phi(u) = \partial\Phi(u)/\partial u$. A consistent estimate of the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is $A_n^{-1}(\hat{\boldsymbol{\beta}})V_n(\hat{\boldsymbol{\beta}})A_n^{-1}(\hat{\boldsymbol{\beta}})$, where it follows from U-statistics theory that the (l, m) element of V_n is

$$V_{n(l,m)}(\boldsymbol{\beta}) = n^{-3} \sum_i \sum_j \sum_{k \neq j} (x_{il} - x_{jl})(x_{im} - x_{km})(e_{ij}^{\boldsymbol{\beta}} - e_{ji}^{\boldsymbol{\beta}})(e_{ik}^{\boldsymbol{\beta}} - e_{ki}^{\boldsymbol{\beta}})$$

where

$$e_{ij}^{\boldsymbol{\beta}} = \delta_i \left[1 - \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h}\right)\right]$$

3. A bounded influence smooth weighted rank estimating equation

A strong motivation for the rank estimating function is its robustness to outlying survival times. The rank estimating equation, however, remains vulnerable to leverage points in the covariate space. In this section, a weighted rank estimating function is proposed. In contrast to the efficiency weights cited earlier, the weight function is chosen to reduce the influence of outlying covariate values on $\hat{\boldsymbol{\beta}}$ and its asymptotic variance.

Denote the weighted estimating function vector by

$$S_n(\boldsymbol{\beta}; w) = (S_{n(1)}(\boldsymbol{\beta}; w), \ldots, S_{n(p)}(\boldsymbol{\beta}; w))^T$$

where the k^{th} component is

$$S_{n(k)}(\boldsymbol{\beta}; w) = n^{-3/2} \sum_i \sum_j \delta_i (x_{ik} - x_{jk}) w_{ij} \left[1 - \Phi \left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h} \right) \right] \qquad k = 1, \ldots, p. \quad (3)$$

The weight function is defined by

$$w_{ij} = \min \left\{ 1, \frac{1}{\max_k (x_{ik} - x_{jk})^2} \right\} \qquad k = 1, \ldots, p.$$

The weights are constant across covariate values within subject pair, and were constructed to assure that $S_n(\boldsymbol{\beta}; w)$ remains a monotone field (Ritov 1987).

Theorem 2: For the accelerated failure time model, under conditions C1-C4, the weighted rank estimating function vector $S_n(\boldsymbol{\beta}; w)$ is a monotone field, differentiable in $\boldsymbol{\beta}$, and has bounded influence. Let $\hat{\boldsymbol{\beta}}$ denote the regression estimate derived as the zero solution to the estimating function vector $S_n(\boldsymbol{\beta}; w)$. Then $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}(w) V(w) A^{-T}(w))$, where
$A(w) = \lim_{n\to\infty} E\{n^{-1/2} \partial S_n(\boldsymbol{\beta}; w)/\partial \boldsymbol{\beta}\}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ and $V(w) = \lim_{n\to\infty} n^{-1} \operatorname{var}\{S_n(\boldsymbol{\beta}_0; w)\}$.

This theorem provides the asymptotic inferential structure for the estimate $\hat{\boldsymbol{\beta}}$. The monotonicity provides a sufficient condition for a unique solution, which eliminates the concern of multiple solutions. The differentiability of the estimating function with respect to $\boldsymbol{\beta}$, enables the conventional sandwich estimate of the asymptotic variance of $\hat{\boldsymbol{\beta}}$ to be computed. The estimated variance covariance matrix is $A_n^{-1}(\hat{\boldsymbol{\beta}}; w) V_n(\hat{\boldsymbol{\beta}}; w) A_n^{-T}(\hat{\boldsymbol{\beta}}; w)$, where $A_n(\hat{\boldsymbol{\beta}}; w)$ and $V_n(\hat{\boldsymbol{\beta}}; w)$ are computed from (3).

The bounded influence provides stability to the regression estimate $\hat{\boldsymbol{\beta}}$ in the presence of outlying survival times and covariate values.

4. Example - Veterans Administration lung cancer data

The Veterans Administration lung cancer data, found in Kalbfleisch and Prentice (1980), is used to demonstrate the proposed methodology. The data were derived from a clinical trial of 137 men with advanced stage lung cancer. The primary endpoint of the clinical trial was survival time. The maximum follow up time in this data set was 599 days, and only 9 of the 137 survival times were censored. For this example, the association between survival time and the Karnofsky Performance Status (KPS), measured at the time of study entry, was examined. The Karnofsky Performance Status is a clinician rating of the patient's functional impairment. The rating scale is from 10 to 100, with 10 representing a moribund state and 100 indicating no evidence of disease.

To assess the relationship between KPS and survival time, the Cox proportional hazards model was initially employed. The log relative risk estimate from the Cox model $\hat{\boldsymbol{\gamma}}$ and its estimated standard error indicated a strong relationship between KPS and survival time ($\hat{\boldsymbol{\gamma}} = -0.033,\ \mathrm{se}(\hat{\boldsymbol{\gamma}}) = 0.005$). To assess the adequacy of the proportional hazards assumption, a smoothed relationship between the scaled Schoenfeld residuals and the Kaplan-Meier estimate of the survival function was explored (Figure 1). Grambsch and Therneau (1994) demonstrate that if the proportional hazards assumption is satisfied, these variables are asymptotically uncorrelated, and the smoothed function should have an approximate zero slope. Figure 1, however, indicates a nonzero slope, calling into question the assumption of a constant relative risk with respect to time. The rejection of the constant relative risk hypothesis was affirmed using the test statistic developed in Grambsch and Therneau (1994), which produced a p-value less than 0.01.

Since the proportional hazards specification was questionable, the weighted rank estimating function was applied to determine the relationship between KPS and survival. Based on the accelerated failure time model

$$\log t_i = \alpha + \beta x_i + \epsilon_i$$

where the $\{x_i, t_i\}$ are the Karnofsky Performance Status scores and the survival times, the coefficient estimate and standard error were $\hat{\beta} = 0.038$ and $\text{se}(\hat{\beta}) = 0.005$, indicating an association between KPS and survival in this patient population. The bandwidth (h), based on the uncensored residuals, was 0.302. The question, however, of whether the accelerated failure time model is appropriate for this data, must be addressed before acceptance of this result. The adequacy of the loglinear specification was explored graphically. A smoothed Kaplan-Meier estimate of the median survival time was computed conditional on the KPS score. This nonparametric relationship was compared to the conditional median estimates produced from the accelerated failure time model; the intercept $\hat{\alpha} = 1.890$ was derived as the median from the Kaplan-Meier estimate of the residuals $\log t_i - \hat{\beta} x_i$. Figure 2 depicts both curves and demonstrates that the accelerated failure time model provided a good fit to the data.

5. Simulations

Simulations were run to compare the small sample performance of the proposed weighted rank regression estimate to the Fygenson-Ritov estimate, and a censored data version of Kendall's rank regression estimate

$$S_{nk}(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i \; \text{sgn}(x_{ik} - x_{jk}) \; [1 - I(r_i^{\boldsymbol{\beta}} > r_j^{\boldsymbol{\beta}})].$$

A prominent feature of the Kendall estimate is it robustness against covariate outliers. Both the Fygenson-Ritov and Kendall estimating equations are nondifferentiable in $\boldsymbol{\beta}$ and require an alternative method to estimate the asymptotic variance of the regression estimate. For the simulations, the procedure developed by Huang (2002) will be employed; it is briefly described.

For a single covariate, calculation of the asymptotic variance of the Fygenson-Ritov and Kendall estimating equations follow directly from U-statistic theory and may be written generically as

$$\text{var}[n^{-1/2} S_n(\beta_0)] = d^2.$$

Now define $\hat{\beta}$ and $\tilde{\beta}$ through the equations

$$n^{-1/2} S_n(\hat{\beta}) = 0$$

$$n^{-1/2} S_n(\tilde{\beta}) = d.$$

The asymptotic variance of $\hat{\beta}$ is estimated as $(\hat{\beta} - \tilde{\beta})^2$.

The simulations were based on the accelerated failure time model

$$\log t_i = \beta_0 x_i + \epsilon_i,$$

where the regression coefficient used in all simulations was $\beta_0 = 2$. The censoring times were determined by first generating values from a uniform distribution $(0, \tau)$, and then taking the log of these values. The choice of τ determines the percentage of censored observations in each replication. The maximum support (τ) of the uniform was chosen to produce average censoring proportions of $\{0.0, 0.25, 0.50, 0.75\}$. For all simulations the sample size was $n = 100$ and there were 5000 replications for each simulation.

In all simulations the covariate (x_i) and error (ϵ) distributions were generated independently. Three different simulation scenarios are presented. In the first set of simulations, the (x_i, ϵ_i) were generated from a bivariate normal distribution with mean $(0, 1)$ and standard deviation $(1, \sigma)$. The second set of simulations were similarly structured except the error distribution was extreme value (log Weibull). These simulations were carried out to examine the properties of the regression estimator when the error distribution is asymmetric. In the third set of simulations, 95% of the (x_i, ϵ_i) were generated from a bivariate normal distribution with mean $(0, 1)$ and standard deviation $(1, \sigma)$, and 5% of the (x_i, ϵ_i) were generated from a bivariate normal distribution with mean $(-5, 1)$ and standard deviation $(1, 2\sigma)$. These five observations were generated to convey contaminated uncensored data with high leverage, and were used to examine the robustness of the rank based regression estimates. In all simulations, the strength of the relationship between the covariate x and the survival time t was dictated by σ, which ranged from 1 to 4.

The properties of the three rank based estimates: the weighted rank regression estimate $\hat{\boldsymbol{\beta}}_w$, the Fygenson-Ritov regression estimate $\hat{\boldsymbol{\beta}}_{FR}$, and the Kendall regression estimate $\hat{\boldsymbol{\beta}}_K$, and their attendant standard error estimates, are presented for the three sets of simulations in Figure 3. The figure displays the bias of $\hat{\boldsymbol{\beta}}$, the simulation standard error of $\hat{\boldsymbol{\beta}}$, the bias of the estimated asymptotic standard error of $\hat{\boldsymbol{\beta}}$, the

95% empirical coverage based on the confidence interval $\hat{\boldsymbol{\beta}} \pm 1.96 \times \text{se}(\hat{\boldsymbol{\beta}})$, and the asymptotic efficiency of the estimates. The simulation standard error for the coverage probability estimate, based on 5000 replications, is approximately 0.003. A tabulation of all the simulation results may be found on the American Statistical Association website http://www.amstat.org/publications/jasa/supplemental_materials/.

For the normal and log Weibull error simulations, the bias is acceptable for each $\hat{\boldsymbol{\beta}}$, but the weighted regression estimate $\hat{\boldsymbol{\beta}}_w$ has the most pronounced bias. The asymptotic standard error estimate for $\hat{\boldsymbol{\beta}}_w$ is consistently unbiased across simulations. In contrast, the standard error estimate for $\hat{\boldsymbol{\beta}}_{FR}$ and $\hat{\boldsymbol{\beta}}_K$, based on nondifferential estimating equations, is positively biased in the higher censoring simulations. The 95% empirical coverage is adequate for all three estimates.

The robustness of the three rank based approaches is examined using a contaminated normal covariate/error distribution. The bias of the weighted rank estimate $\hat{\boldsymbol{\beta}}_w$ is least affected by contamination. The stability of $\hat{\boldsymbol{\beta}}_w$ stems from the covariate based bounded influence weight function. This weight function, however, produces a small increase in the sampling variability of $\hat{\boldsymbol{\beta}}_w$ relative to $\hat{\boldsymbol{\beta}}_K$. Similar to the uncontaminated simulations, the asymptotic standard error estimate for $\hat{\boldsymbol{\beta}}_{FR}$ and $\hat{\boldsymbol{\beta}}_K$ is positively biased, and the asymptotic standard error estimate for $\hat{\boldsymbol{\beta}}_w$ remains unbiased. Interestingly, the positive bias in the standard error estimate for $\hat{\boldsymbol{\beta}}_K$ somewhat offset the bias of $\hat{\boldsymbol{\beta}}_K$ by providing a larger interval to cover the true $\boldsymbol{\beta}$. The Kendall based coverage probability, however, remains below the nominal 95% rate. The coverage probability using $\hat{\boldsymbol{\beta}}_{FR}$ is poor. The bias in $\hat{\boldsymbol{\beta}}_{FR}$ is too strong to be counterbalanced by the positive bias in its standard error estimate. The coverage probability using $\hat{\boldsymbol{\beta}}_w$ is unaffected by the contaminated observations. In conclusion, the contaminated simulation results demonstrate the stability of the weighted rank estimate $\hat{\boldsymbol{\beta}}_w$ in the presence of a small percentage of contaminated observations; the Kendall estimate $\hat{\boldsymbol{\beta}}_K$ is moderately affected and the Fygenson-Ritov estimate $\hat{\boldsymbol{\beta}}_{FR}$ is strongly influenced by the contaminated observations. The standard error estimate for $\hat{\boldsymbol{\beta}}_w$ is superior to the standard error estimates for $\hat{\boldsymbol{\beta}}_{FR}$ and $\hat{\boldsymbol{\beta}}_K$ over all simulations.

As noted in the contaminated normal simulation results, the robustness of $\hat{\boldsymbol{\beta}}_w$ may be counterbalanced by an increase in variability. To explore the asymptotic efficiency of the three estimators, an additional set of simulations was generated. The simulation structure used for the small sample experiments was used again here except that the sample size was increased to 500, to approximate asymptotic calculations, and only 500 replications were generated for each simulation. To estimate the asymptotic efficiency of the three regression estimators, the asymptotic standard error of these regression estimates were compared to the asymptotic standard of the estimators derived from the asymptotically efficient weighted rank estimated function (1). For the normal and log Weibull error distributions, the asymptotically efficient weights are $w(r) = \lambda(r) - r$ (Ritov 1990) and $w(r) = 1$ (Tsiatis 1990), respectively. For the contaminated normal simulations, additional weights were used to downweight the outlying observations. The estimating function used as the benchmark for the contaminated normal efficiency simulations was

$$\sum_i \delta_i w_1(r_i^\beta) u(x_i) \left\{ x_i - \frac{\sum_j x_j u(x_j) I(r_j^\beta \geq r_i^\beta)}{\sum_j u(x_j) I(r_j^\beta \geq r_i^\beta)} \right\}$$

where

$$w_1(r_i^\beta) = [\lambda(r_i^\beta) - r_i^\beta] \times \min\left\{1, \frac{1}{|\lambda(r_i^\beta) - r_i^\beta|}\right\}, \qquad u(x_i) = \min\{1, 1/|x_i|\}$$

The outcome of the efficiency simulations, depicted in the bottom of Figure 3, were expressed as

$$\text{asymptotic efficiency } = \frac{\text{se}(\hat{\beta}_{eff})}{\text{se}(\hat{\beta})}.$$

For the normal simulations, the Fygenson-Ritov estimator is efficient, whereas the robust estimators $\hat{\boldsymbol{\beta}}_w$ and $\hat{\boldsymbol{\beta}}_K$ obtain relative efficiencies in the 85-100 percent range. The efficiency of the Fygenson-Ritov estimate is attributable to its monotone decreasing weight function, a characteristic it shares with the estimate derived from the asymptotically optimal weighted estimating function for the normal error. In the log Weibull simulations, the relative efficiencies of the three estimators are comparable, and are generally in the range of 75-85 percent. The contaminated normal simulations produce a reduction in efficiency for the Fygenson-Ritov estimator relative to $\hat{\boldsymbol{\beta}}_w$ and $\hat{\boldsymbol{\beta}}_K$.

6. Discussion

For the (log) linear model, with uncensored data, M-estimation is a common choice for the creation of an unbiased estimating equation and subsequent inference on the regression parameters. With right censored data, however, the right support of the underlying survival time distribution may not observable, requiring truncation or weighting devices to the right tail of the survival distribution (Ritov 1990, Lai and Ying 1992). As pioneered by Cox (1972), and adapted to the accelerated failure time model by Tsiatis (1990) and subsequent researchers, the use of ranks provide an unconstrained approach to the construction of an unbiased estimating equation in the presence of right censored data.

One barrier to greater usage of rank based inference in the accelerated failure time model stems from the discontinuity of the estimating equation, and the resulting difficulty in computing the asymptotic variance of the regression estimator. Brown and Wang (2005) explored smoothed approximations to the estimating function for the uncensored case. With censored data, Jin et al. (2003) used a resampling tool to bypass differentiation of the estimating function. This approach, however, is computationally intensive and may be problematic if applied to an iterative process of model building, where each time a set of covariates are selected, the significance of the corresponding regression coefficients require determination. The order of magnitude of these computations could increase further if regression diagnostics are incorporated into the model building approach, requiring a reassessment of the significance of the regression coefficients each time an influential observation is removed.

In this work, a smooth approximation to the monotone rank based estimating function was developed in the censored data case. Smoothing provided an analytic expression for the asymptotic variance of $\hat{\boldsymbol{\beta}}$, with a straightforward plug-in estimate. The differentiability and monotonicity of the estimating function, with respect to $\boldsymbol{\beta}$, insured that the Newton-Raphson algorithm converged quadratically to the proper neighborhood of the true regression parameter. In addition, the proposed estimate has bounded influence, and from the limited simulation studies, provided stable finite sample results in the presence of outlying observations.

Appendix: Proof of theorems

Lemma 1: Under conditions C1 - C4, the Fygenson-Ritov unsmoothed estimating function $\tilde{S}_n(\boldsymbol{\beta})$ and the smoothed estimating function $S_n(\boldsymbol{\beta})$ are asymptotically equivalent

$$\tilde{S}_n(\boldsymbol{\beta}) = S_n(\boldsymbol{\beta}) + o_p(1) \quad \text{uniformly in } \boldsymbol{\beta}$$

Proof:

$$\tilde{S}_n(\boldsymbol{\beta}) - S_n(\boldsymbol{\beta}) = n^{-3/2} \sum_i \sum_j \delta_i(\mathbf{x}_i - \mathbf{x}_j) \left\{ \Phi\left(\frac{r_i^{\boldsymbol{\beta}} - r_j^{\boldsymbol{\beta}}}{h}\right) - I(r_i^{\boldsymbol{\beta}} > r_j^{\boldsymbol{\beta}}) \right\}$$

Letting $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ and $r = r_1 - r_2$,

$$\left|\tilde{S}_n(\boldsymbol{\beta}) - S_n(\boldsymbol{\beta})\right| \leq M n^{1/2} \sup_{\boldsymbol{\beta} \in \mathcal{B}} \left| \int_{\mathbf{x}} \int_r \mathbf{x} \left\{ \Phi\left(\frac{r^{\boldsymbol{\beta}}}{h}\right) - I(r^{\boldsymbol{\beta}} > 0) \right\} d\hat{F}_{n\times n}(r^{\boldsymbol{\beta}}|\mathbf{x}) d\hat{G}_{n\times n}(\mathbf{x}) \right| \tag{A.1}$$

where $\hat{F}_{n\times n}$ and $\hat{G}_{n\times n}$ are the conditional and marginal empirical cumulative distribution functions with jumps at each of the n^2 differences.

The right hand side can be expanded as

$$\int_{\mathbf{x}} \int_r \mathbf{x} \left\{ \Phi\left(\frac{r}{h}\right) - I(r > 0) \right\} \left[d\hat{F}_{n\times n}(r|\mathbf{x}) - dF(r|\mathbf{x}) \right] d\hat{G}_{n\times n}(\mathbf{x})$$
$$+ \int_{\mathbf{x}} \int_r \mathbf{x} \left\{ \Phi\left(\frac{r}{h}\right) - I(r > 0) \right\} dF(r|\mathbf{x}) d\hat{G}_{n\times n}(\mathbf{x})$$

where $F(r|\mathbf{x}) = \lim_{n\to\infty} \hat{F}_{n\times n}(r|\mathbf{x})$ and the superscript $\boldsymbol{\beta}$ is omitted when selecting the supremum over the bounded space $\mathcal{B}$. As a result, (A.1) is bounded by

$$M n^{1/2} |U_1(h) + U_2(0)| + M n^{1/2} |B(h)|$$

where

$$U_1(h) = \int_{\mathbf{x}} \int_r \mathbf{x} \Phi\left(\frac{r}{h}\right) \left[d\hat{F}_{n\times n}(r|\mathbf{x}) - dF(r|\mathbf{x}) \right] d\hat{G}_{n\times n}(\mathbf{x})$$

$$U_2(0) = -\int_{\mathbf{x}}\int_r \mathbf{x}I(r>0)\left[d\hat{F}_{n\times n}(r|\mathbf{x}) - dF(r|\mathbf{x})\right] d\hat{G}_{n\times n}(\mathbf{x})$$

$$B(h) = \int_{\mathbf{x}}\int_r \mathbf{x}\left[\Phi\left(\frac{r}{h}\right) - I(r>0)\right] dF(r|\mathbf{x})d\hat{G}_{n\times n}(\mathbf{x})$$

For $U_1(h)$, a change of variable $z = h^{-1}r$ and integration by parts gives

$$U_1(h) = -\int_{\mathbf{x}}\int_z \mathbf{x}\phi(z)[\hat{F}_{n\times n}(zh|\mathbf{x}) - F(zh|\mathbf{x})]dz\ d\hat{G}_{n\times n}(\mathbf{x}),$$

and hence

$$n^{1/2}|U_1(h)+U_2(0)| = \\ n^{1/2}\left|\int_{\mathbf{x}}\int_z \mathbf{x}\phi(z)\left\{\left[\hat{F}_{n\times n}(zh|\mathbf{x}) - F(zh|\mathbf{x})\right] - \left[\hat{F}_{n\times n}(0|\mathbf{x}) - F(0|\mathbf{x})\right]\right\} dz\ d\hat{G}_{n\times n}(\mathbf{x})\right|.$$

Using the results on oscillations of empirical processes (Shorack and Wellner, 1986; p.531) and assuming $E(\mathbf{x}\mathbf{x}^T) < M < \infty$

$$n^{1/2}|U_1(h)+U_2(0)| = O_p\left(\left[h\log n\ \log\left(\frac{1}{h\log n}\right)\right]^{1/2}\right).$$

For $B(h)$, integration by parts and a two-term Taylor expansion around $h = 0$, produces

$$B(h) = -\frac{h^2}{2}\int_{\mathbf{x}}\int_z \mathbf{x}z^2\phi(z)f'(zh^*|\mathbf{x})dzd\hat{G}_{n\times n}(\mathbf{x})$$

where h^* lies between h and zero, and $f'(u|\mathbf{x}) = \partial^2 F(u|\mathbf{x})/\partial u^2$.

Combining the above arguments,

$$\begin{aligned}\sup_{\boldsymbol{\beta}\in\mathcal{B}}|\tilde{S}_n(\boldsymbol{\beta}) - S_n(\boldsymbol{\beta})| &\leq Mn^{1/2}|U_1(h)+U_2(0)| + Mn^{1/2}|B(h)| \\ &= O_p\left(\left[h\log n\ \log\left(\frac{1}{h\log n}\right)\right]^{1/2} + n^{1/2}h^2\right).\end{aligned}$$

Using the condition $nh^4 \to 0$, it follows that

$$\tilde{S}_n(\boldsymbol{\beta}) - S_n(\boldsymbol{\beta}) = o_p(1) \qquad \text{uniformly in } \ \boldsymbol{\beta}.$$

Theorem 1: Under conditions C1 - C4, $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}VA^{-T})$, where $A = \lim_{n\to\infty} E\{n^{-1/2}\partial S_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ and
$V = \lim_{n\to\infty} n^{-1}\ \mathrm{var}\{S_n(\boldsymbol{\beta}_0)\}$.

Proof: From Lemma 1 and the results in Fygenson and Ritov (1994), $S_n(\boldsymbol{\beta})$ is asymptotically normal with mean zero. Taylor expanding $S_n(\hat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}_0$

$$S_n(\hat{\boldsymbol{\beta}}) = S_n(\boldsymbol{\beta}_0) + A_n(\boldsymbol{\beta}^{\#})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

where $\boldsymbol{\beta}^{\#}$ lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Since $n^{-1/2}A_n(\boldsymbol{\beta}^{\#})$ is bounded and non-zero (condition C2), it follows that $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}VA^{-T})$.

Definition: An estimating function vector is a monotone field if for all $\boldsymbol{\beta}, \boldsymbol{\alpha} \in R^p$ (Ritov 1987)

$$\frac{\partial}{\partial a}\boldsymbol{\alpha}^T S_n(\boldsymbol{\beta} + a\boldsymbol{\alpha}) \geq 0 \qquad a \in R^1.$$

Theorem 2: Under conditions C1-C4, the weighted rank estimating function vector $S_n(\boldsymbol{\beta}; w)$ is a monotone field, differentiable in $\boldsymbol{\beta}$, and has bounded influence. Let $\hat{\boldsymbol{\beta}}$ denote the regression estimate derived as the zero solution to the estimating function vector $S_n(\boldsymbol{\beta}; w)$. Then $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}(w)V(w)A^{-T}(w))$, where $A(w) = \lim_{n\to\infty} E\{n^{-1/2}\partial S_n(\boldsymbol{\beta}; w)/\partial\boldsymbol{\beta}\}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ and
$V(w) = \lim_{n\to\infty} n^{-1}\ \mathrm{var}\{S_n(\boldsymbol{\beta}_0; w)\}$.

PROOF: An argument similar to that found in Narjano and Hettmansperger (1994) is used to show that the influence function is bounded. Since the unsmoothed and smoothed estimating functions are asymptotically equivalent, consider the unsmoothed estimating function

$$\tilde{S}_{n(k)}(\boldsymbol{\beta}_0; w) = n^{-3/2} \sum_i \sum_j \phi(x_{ik} - x_{jk}) \delta_i [1 - I(r_i > r_j)] \quad k = 1, \ldots, p$$

where $\phi(u) = |u|^{-1}$ if $|u| \geq 1$ and $\phi(u) = u$ if $|u| < 1$. The expectation of this estimating function, conditional on $(\epsilon_i, \epsilon_j, x_{ik}, x_{jk})$ is

$$n^{-3/2} \sum_i \sum_j \phi(x_{ik} - x_{jk}) \bar{G}_i(\epsilon_i + \mathbf{x}_i^T \boldsymbol{\beta}_0) \bar{G}_j(\epsilon_i + \mathbf{x}_j^T \boldsymbol{\beta}_0) I(\epsilon_j > \epsilon_i)$$

where $\bar{G}_i(u)$ is the survival function of the censoring random variable conditional on the covariate $\mathbf{x}_i$. It follows that the population estimating function that defines $\boldsymbol{\beta}_0$ is

$$\int_{\mathbf{x}_1, \epsilon_1} \int_{\mathbf{x}_2, \epsilon_2} \phi(x_{1k}, x_{2k}) \bar{G}_1(\epsilon_1 + \mathbf{x}_1^T \boldsymbol{\beta}_0) \bar{G}_2(\epsilon_1 + \mathbf{x}_2^T \boldsymbol{\beta}_0) I(\epsilon_2 > \epsilon_1) dH(\mathbf{x}_1, \epsilon_1) dH(\mathbf{x}_2, \epsilon_2) = 0$$

for $k = 1, 2 \ldots, p$. To compute the influence function for $\boldsymbol{\beta}_0$, define the contaminated cumulative distribution function by $H_v = (1 - v)H + v\psi_0$, where $\psi_0(\mathbf{x}, \epsilon)$ is the distribution function of a point mass at $(\mathbf{x}_0, \epsilon_0)$. The population estimating equation based on the contaminated distribution is written as

$$\int_{\mathbf{x}_1, \epsilon_1} \int_{\mathbf{x}_2, \epsilon_2} \phi(x_{1k} - x_{2k}) \bar{G}_1(\epsilon_1 + \mathbf{x}_1^T \boldsymbol{\beta}(H_v)) \bar{G}_2(\epsilon_1 + \mathbf{x}_2^T \boldsymbol{\beta}(H_v)) I(\epsilon_2 > \epsilon_1) dH_v(\mathbf{x}_1, \epsilon_1) dH_v(\mathbf{x}_2, \epsilon_2) = 0. \tag{A.2}$$

Differentiating both sides of (A.2) with respect to v, setting $v = 0$, and denoting the influence function $(\partial/\partial v)\boldsymbol{\beta}(H_v)|_{v=0}$ by $\dot{\boldsymbol{\beta}}$ produces

$$\int_{\mathbf{x}_1, \epsilon_1} \int_{\mathbf{x}_2, \epsilon_2} \phi(x_{1k} - x_{2k}) \frac{\partial}{\partial \boldsymbol{\beta}} \left\{ \bar{G}_1(\epsilon_1 + \mathbf{x}_1^T \boldsymbol{\beta}) \bar{G}_2(\epsilon_1 + \mathbf{x}_2^T \boldsymbol{\beta}) \right\} \dot{\boldsymbol{\beta}} I(\epsilon_2 > \epsilon_1) dH(\mathbf{x}_1, \epsilon_1) dH(\mathbf{x}_2, \epsilon_2)$$

$$+\int_{\mathbf{x}_1,\epsilon_1} \phi(x_{1k}-x_{0k})\bar{G}_1(\epsilon_1+\mathbf{x}_1^T\boldsymbol{\beta})\bar{G}_0(\epsilon_1+\mathbf{x}_0^T\boldsymbol{\beta})I(\epsilon_0>\epsilon_1)dH(\mathbf{x}_1,\epsilon_1)$$
$$+\int_{\mathbf{x}_2,\epsilon_2} \phi(x_{0k}-x_{2k})\bar{G}_0(\epsilon_0+\mathbf{x}_0^T\boldsymbol{\beta})\bar{G}_2(\epsilon_0+\mathbf{x}_2^T\boldsymbol{\beta})I(\epsilon_2>\epsilon_0)dH(\mathbf{x}_2,\epsilon_2).$$

The influence function evaluated at $(\mathbf{x}_0,\epsilon_0)$ is

$$\dot{\boldsymbol{\beta}} = \left[\int_{\mathbf{x}_1,\epsilon_1}\int_{\mathbf{x}_2,\epsilon_2} \phi(x_{1k}-x_{2k})\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\bar{G}_1(\epsilon_1+\mathbf{x}_1^T\boldsymbol{\beta})\bar{G}_2(\epsilon_1+\mathbf{x}_2^T\boldsymbol{\beta})\right\} I(\epsilon_2>\epsilon_1)dH(\mathbf{x}_1,\epsilon_1)dH(\mathbf{x}_2,\epsilon_2)\right]^{-1}$$
$$\left[\int_{\mathbf{x}_1,\epsilon_1} \phi(x_{1k}-x_{0k})\bar{G}_1(\epsilon_1+\mathbf{x}_1^T\boldsymbol{\beta})\bar{G}_0(\epsilon_1+\mathbf{x}_0^T\boldsymbol{\beta})I(\epsilon_0>\epsilon_1)dH(\mathbf{x}_1,\epsilon_1)\right.$$
$$\left.+\int_{\mathbf{x}_2,\epsilon_2} \phi(x_{0k}-x_{2k})\bar{G}_0(\epsilon_0+\mathbf{x}_0^T\boldsymbol{\beta})\bar{G}_2(\epsilon_0+\mathbf{x}_2^T\boldsymbol{\beta})I(\epsilon_2>\epsilon_0)dH(\mathbf{x}_2,\epsilon_2)\right]$$

which is bounded for any $(\mathbf{x}_0,\epsilon_0)$.

To show that $S_n(\boldsymbol{\beta})$ is a monotone field, consider

$$\begin{aligned}\boldsymbol{\alpha}^T S_n(\boldsymbol{\beta}+a\boldsymbol{\alpha}) &= \alpha_1 n^{-3/2}\sum_i\sum_j \delta_i(x_{i1}-x_{j1})w_{ij1}\left[1-\frac{r_i^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}-r_j^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}}{h}\right]+\ldots\\ &+ \alpha_p n^{-3/2}\sum_i\sum_j \delta_i(x_{ip}-x_{jp})w_{ijp}\left[1-\Phi\left(\frac{r_i^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}-r_j^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}}{h}\right)\right]\end{aligned}$$

where $r_i^{\boldsymbol{\beta}+a\boldsymbol{\alpha}} = y_i-(\beta_1+a\alpha_1)x_{i1}-\ldots-(\beta_p+a\alpha_p)x_{ip}$. Since

$$\frac{\partial}{\partial a}\left[1-\Phi\left(\frac{r_i^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}-r_j^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}}{h}\right)\right] = \frac{\boldsymbol{\alpha}^T(\mathbf{x}_i-\mathbf{x}_j)}{h}\phi\left(\frac{r_i^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}-r_j^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}}{h}\right)$$

it follows that

$$\frac{\partial}{\partial a}\boldsymbol{\alpha}^T S_n(\boldsymbol{\beta}+a\boldsymbol{\alpha}) =$$
$$n^{-3/2}\sum_i\sum_j \phi\left(\frac{r_i^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}-r_j^{\boldsymbol{\beta}+a\boldsymbol{\alpha}}}{h}\right)\left[\frac{\boldsymbol{\alpha}^T(\mathbf{x}_i-\mathbf{x}_j)}{h}\right][\alpha_1(x_{i1}-x_{j1})w_{ij1}+\ldots+\alpha_p(x_{ip}-x_{jp})w_{ijp}]$$

and since $w_{ijk}=w_{ij}>0$

$$\frac{\partial}{\partial a}\boldsymbol{\alpha}^T S_n(\boldsymbol{\beta}+a\boldsymbol{\alpha}) \geq 0 \qquad a\in R^1$$

The proof that $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to $N(0, A^{-1}VA^{-1})$ follows immediately from the proof in Theorem 1.

Acknowledgements

I would like to thank the joint editor Dr. Portnoy, an associate editor, and a reviewer, for comments that improved the content and presentation of this manuscript.

References

Brown, B. M., and Wang, Y. G. (2005), "Standard errors and covariance matrices for smoothed rank estimators," *Biometrika*, 92, 149-158.

Cox, D. R. (1972), "Regression models and life tables (with Discussion)," *Journal of Royal Statistical Society, Series B*, 34, 187-220.

Fygenson, M., and Ritov, Y. (1994), "Monotone estimating equations for censored data," *Annals of Statistics*, 22, 732-746.

Grambsch, P. M., and Therneau, T. M. (1994), "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, 81, 515-526.

Hess, K. R., Serachitopol, D. M., and Brown, B. W. (1999), "Hazard function estimators: a simulation study," *Statistics in Medicine*, 18, 3075-3088.

Huang Y. (2002), "Calibration regression of censored lifetime medical cost," *Journal of the American Statistical Association*, 97, 318-327.

Kalbfleisch, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, Wiley: New York.

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.

Lai, T. L., and Ying, Z. (1991), "Large sample theory of a modified Buckley-James estimator for regression analysis with censored data," *Annals of Statistics*, 19,1370-1402.

Naranjo, J. D., and Hettmansperger, T. P. (1994), "Bounded influence rank regression," *Journal of the Royal Statistical Society, Series B*, 56, 209-220.

Ritov, Y. (1987), "Tightness of monotone random fields," *Journal of the Royal Statistical Society, Series B*, 49, 331-333.

Ritov, Y. (1990), "Estimation in a linear regression model with censored data," *Annals of Statistics*, 18, 303-328.

Shorack, G. R., and Wellner, J. A. (1986), *Empirical Processes with Applications to Statistics*, Wiley: New York.

Tsiatis, A. A. (1990), "Estimating regression parameters using linear rank tests for censored data," *Annals of Statistics*, 18,354-372.

Ying, Z. (1993), "A large sample study of rank estimation for censored regression data," *Annals of Statistics*, 21,76-99.

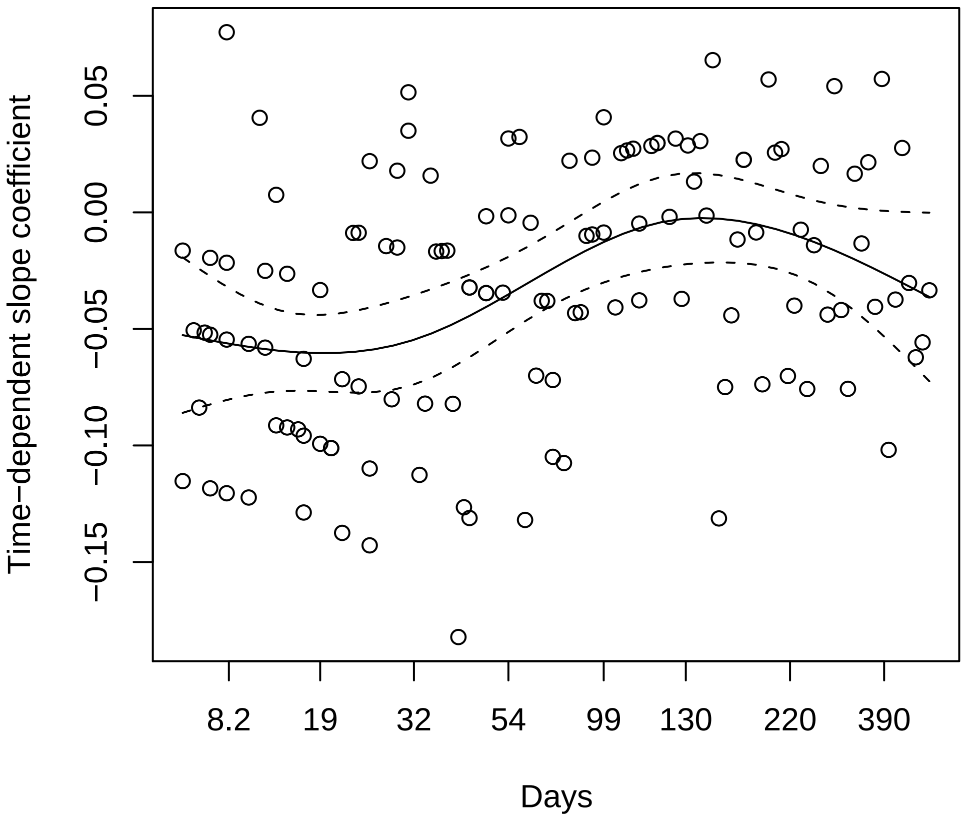

Figure 1: Time-dependent coefficient plot to assess the adequacy of the proportional hazards assumption. The solid line represents the time dependent slope estimate and the dashed lines a 95% confidence interval. A nonzero slope indicates a violation of the proportional hazards assumption.

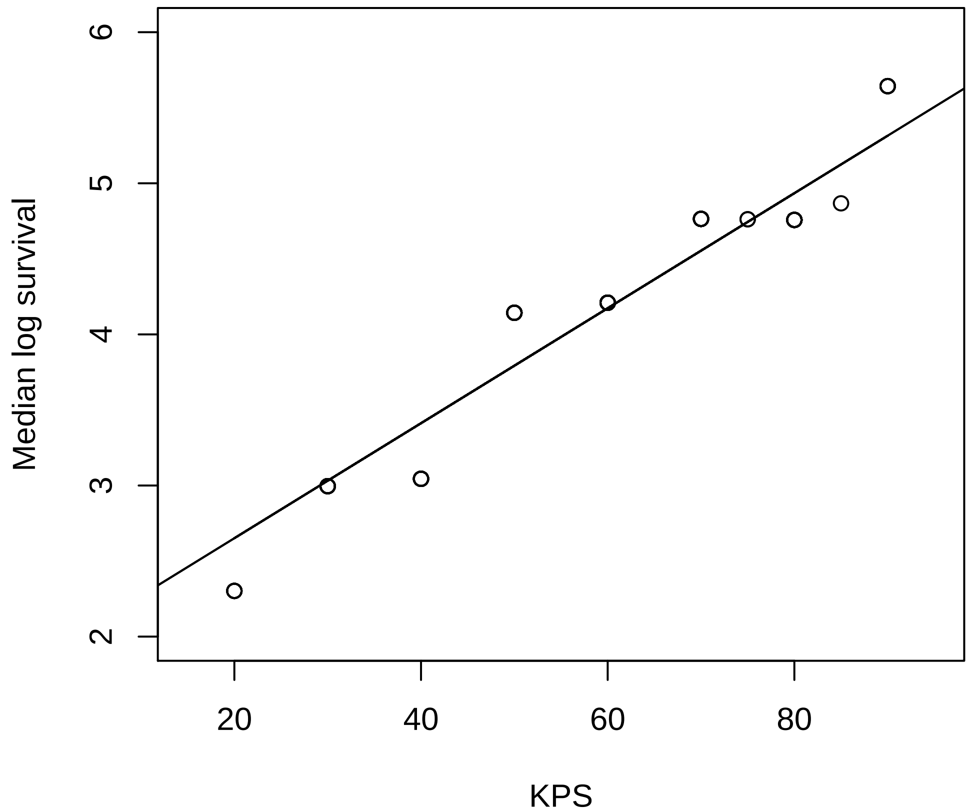

Figure 2: Estimated median log survival times from a smoothed Kaplan-Meier estimator is plotted using circles. Estimated median log survival times from the accelerated failure time model is plotted as a solid line.

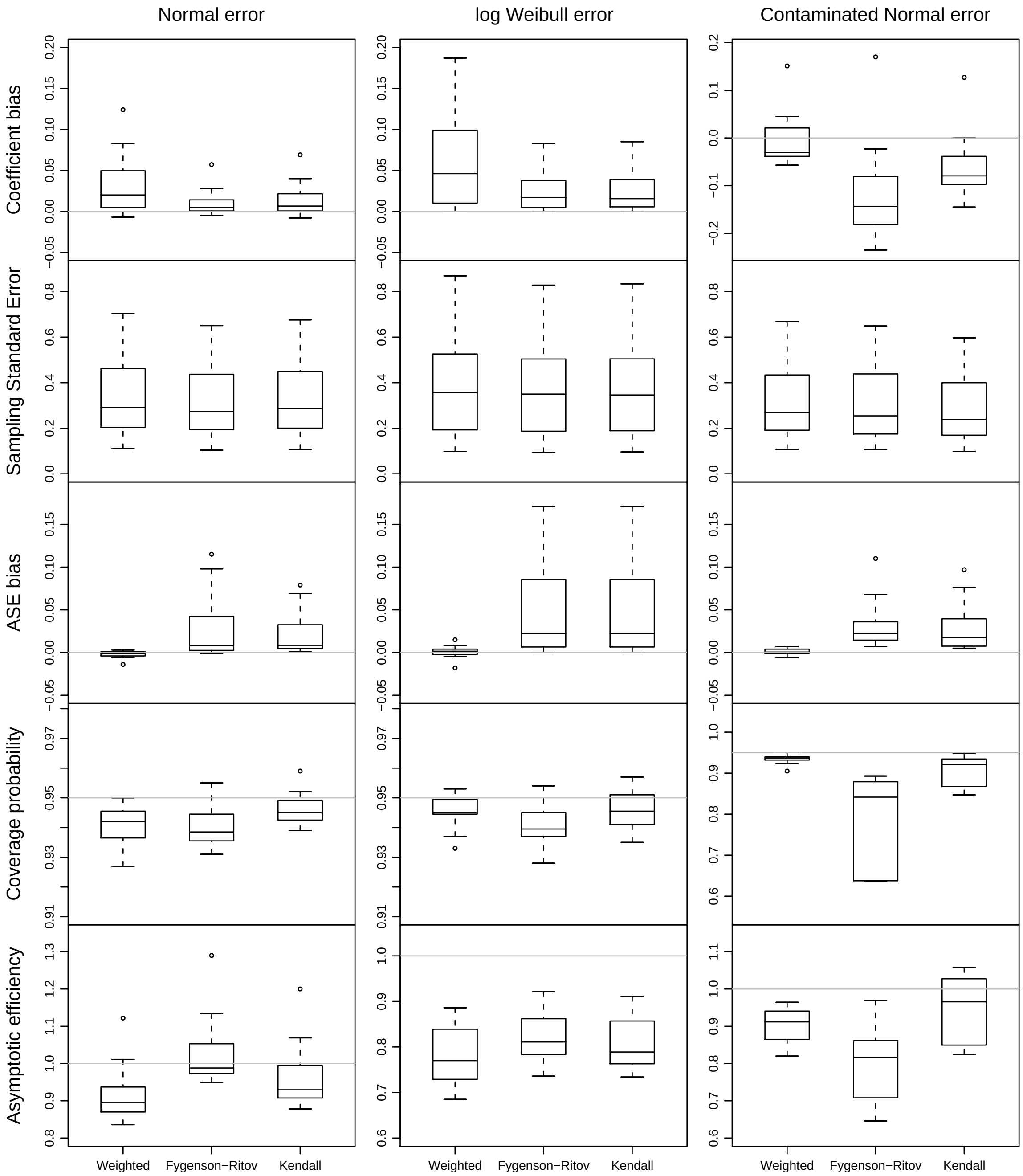

Figure 3: Boxplots comparing the bias and sampling standard error of $\hat{\beta}$, the bias of the asymptotic standard error estimate, the 95% empirical coverage probability, and the asymptotic efficiency of $\hat{\beta}$. The normal, log Weibull, and contaminated normal error distributions were used for the simulations.