EMPIRICAL LIKELIHOOD ANALYSIS FOR THE HETEROSCEDASTIC ACCELERATED FAILURE TIME MODEL

Mai Zhou, Mi-Ok Kim, and Arne C. Bathke

Department of Statistics, University of Kentucky

Division of Biostatistics and Epidemiology, Cincinnati Children's Medical Center

Abstract: In this manuscript, we provide a general framework for the accelerated failure time (AFT) model that ties different data generating mechanisms, namely correlation and regression models, with different estimators that have been proposed for the AFT model. Furthermore, we develop empirical likelihood methods for inference that yield asymptotic chi-squared results by adequately reflecting the underlying data generating mechanisms in the construction of the likelihood. The results are also applicable to censored quantile regression, as illustrated in an example.

Key words and phrases: Censored quantile regression, correlation model, regression model, Wilks' theorem.

1. Introduction

The semi-parametric accelerated failure time (AFT) model is an extension of linear regression to the analysis of survival data such that for some survival times T_i,

$$\log(T_i) = Y_i = X_i^\top \beta + e_i \ , \tag{1.1}$$

where the distribution of the error term e_i is unspecified. Due to censoring on the responses, we observe $Z_i = \min(Y_i, C_i)$ and $\delta_i = I_{[Y_i \le C_i]}$ for some censoring time variables C_i instead of Y_i.

This model has emerged as a useful alternative to the popular Cox proportional hazards model for analysing censored data, as it provides a direct interpretation of the results in terms of quantification of survival times instead of the more abstract hazard rates. However, it is less utilized than it should, among other reasons, because inference for the model has been difficult. Under random right censoring, inference methods that require a direct estimation of the covariance matrices of the estimators are difficult to implement because the covariance matrices involve nonparametric estimation of the underlying distribution. When least squares estimating equation (3) is concerned, a Wald type test and associated confidence interval, although less desirable for the reason given above, can be constructed as in Stute (1999). However, for median (or quantile) estimating equations used in Theorem 3, the asymptotic variance of $\hat{\beta}$ involves the density at the median (or the quantile of interest), of which a reliable estimation is problematic even without censoring. Resampling methods, as proposed by Jin *et al.* (2003), constitute an alternative, but they are computationally rather involved. Here we propose a new empirical likelihood ratio approach that is computationally simple and can apply to both cases equally well.

Qin & Jing (2001) and Li & Wang (2003) considered empirical likelihood, a nonparametric inference method based on the likelihood principle. For linear models, Owen (1991) showed that a nonparametric equivalent of Wilks' theorem holds for the $-2\log$ empirical likelihood ratio. Qin & Jing (2001) and Li & Wang (2003) kept the likelihood of Owen (1991) unchanged and replaced the constraint equations by $\sum_{i=1}^{n} p_i X_i (Y_i^* - X_i^\top b) = 0$, where Y^* is the 'synthetic data' of Koul *et al.* (1981). The approach by Qin & Jing (2001) and Li & Wang (2003) has two major drawbacks. First, the inference method is based on an estimator that does not perform well (see Simulation Study 3 below), which implies that any inference based on this estimator cannot perform well either. Second, their methods did not yield the non-parametric version of Wilks' theorem since the likelihood construction in their case did not appropriately reflect the data generating mechanism. Indeed, the limiting distribution of the $-2\log$ empirical likelihood ratio is that of a linear combination of weighted independent χ_1^2 random variables with the weights to be estimated. In practice, the limiting distribution needs to be calibrated via resampling. These drawbacks have motivated our work.

We first provide a general framework for the AFT model that ties together different data generating mechanisms with different estimators that have been proposed for the AFT model. We distinguish two different types of data generating schemes under random right censoring, the accelerated failure time *regression* model and the accelerated failure time *correlation* model. This differentiation is similar to that made by Freedman (1981) regarding linear models. However, we find that in the accelerated failure time model, the different data generation models require different estimators, unlike the uncensored case of Freedman (1981), and different assumptions on the censoring times C_i as well. We provide details in Section 2.

Furthermore, we develop empirical likelihood methods that yield asymptotic χ^2 results by adequately reflecting the underlying data generating mechanisms in the construction of the likelihood. With regard to the bootstrap, Freedman (1981) noted that the resampling scheme must reflect the relevant features of the stochastic model assumed to have generated the data. Owen (1991) recognized this with empirical likelihood for linear models. He suggested that the empirical likelihood be constructed based on the homoscedasticity of (X_i, Y_i) for the correlation model, while the likelihood be based on the homoscedasticity of the errors e_i for the regression model. We extend this distinction to the accelerated failure time model under random right-censoring and call the first type of empirical likelihood formulation *case-wise* and the latter *residual-wise*. Without censoring, the two different likelihood formulations yield the same likelihood function and hence produce identical p-values and confidence regions, although interpreted differently (see Owen (1991)). With censoring however, the *case-wise* empirical likelihood is no longer identical to the *residual-wise* one, and estimates, p-values, and confidence regions differ.

The proposed likelihood has a complex form and is not explicit under the constraints. We show that nevertheless the resulting inference methods are computationally simple, particularly for the AFT model that admits heteroscedastic errors, and they yield the standard asymptotic χ^2 results. We show that the results are also applicable to censored quantile regression.

2. Correlation and Regression Model

In this section, we summarize the main characteristics of the two data generating models and corresponding estimators. Our main focus will be on the correlation AFT model. For comparison, we include a brief discussion of the empirical likelihood results regarding the AFT regression model in Subsection 2.1.

2.1 Regression Model

The regression model is appropriate if, for example, the measurement error of the response is the main source of uncertainty (Freedman 1981). The true value of the p-dimensional parameter vector β solves $\int(y - x^\top\beta)x\,dF_e = 0$, where F_e denotes the error distribution. The main assumptions in the regression model are that the covariates, $x_1, \cdots, x_n$, are row vectors of $p-$dimensional fixed constants, forming a matrix of full rank, the errors, $e_1, \cdots, e_n$, are independent, with common distribution F_e having mean 0 and finite variance σ^2 (both F_e and σ^2 unknown), and the censoring time variables, $C_1, \cdots, C_n$, are independent with common unknown distribution G and independent of Y_i conditionally on x_i.

Popular estimators of the parameter vector β in this model with censored data include rank based estimators (see Chapter 7 of Kalbfleisch & Prentice (2002) and references therein; Jin *et al.* 2003) and the Buckley–James estimator (Buckley & James 1979; Lai & Ying 1991).

The following empirical likelihood approach was proposed by Zhou & Li (2008). Let $Z_i = \min(Y_i, C_i)$ and $\delta_i = I_{[Y_i \le C_i]}$. Let b be a vector, and define the residuals with respect to b as $r_i(b) = z_i - x_i^\top b$. Zhou & Li (2008) proposed empirical likelihood be formulated with respect to $(r_i(b), \delta_i)$ as follows: given b, the *residual-wise* empirical likelihood for some univariate distribution F is defined as

$$L_e(F) = \prod_{\delta_i=1} p_i \prod_{\delta_i=0} (1 - \sum_{r_j(b) \le r_i(b)} p_j)\,,$$

where $p_i = dF[r_i(b)]$ is the probability placed by F on the ith residual. The likelihood ratio is

$$R_e(b) = \frac{\sup\{L_e(F) | F \in \mathbb{F}^b\}}{\sup\{L_e(F) | F \in \mathfrak{F}^b\}}\,, \tag{2.1}$$

where $\mathfrak{F}^b$ denotes the set of all univariate distributions that place positive probabilities on each uncensored $r_i(b)$, as $L_e(F) = 0$ for any F that places zero probability on any uncensored $r_i(b)$, and $\mathbb{F}^b$ denotes a subset of $\mathfrak{F}^b$ that satisfies the constraints $\sum_{i=1}^n p_i \delta_i r_i(b) \tilde{x}_i$ for $\tilde{x}_i = x_i + \sum_{\delta_j=0,\, j:j<i} m[j,i] x_j$. Here, $m[j,i]$ denotes the weights derived from the Buckley–James estimating equation. We refer to Zhou & Li (2008) for more details. As the term in the denominator of (2.1) is maximized by the Kaplan–Meier estimator (Kaplan & Meier 1958) of the residuals $r_i(b)$ whose calculation is straightforward, maximization is only required for the numerator, analogously to the uncensored case. When $\hat{b}$ is the Buckley–

James estimator, $R_e(\hat{b}) = 1$ and confidence regions based on (2.1) are 'centered' at the Buckley–James estimate. By formulating similar constraints with respect to rank estimators, Zhou (2005a) proposed a *residual-wise* likelihood for log-rank or Gehan-type estimators. In each case, the resulting likelihood ratio admits chi-squared limiting distributions (Zhou 2005a; Zhou & Li 2008).

2.2 Correlation Model

The correlation model is appropriate if, for example, the goal is to estimate the regression plane for a certain population on the basis of a simple random sample (Freedman 1981). The true value of the parameter β solves $\iint (y - x^\top \beta) x \, dF_{xy} = 0$, where F_{xy} denotes the joint $(p+1)$-variate distribution of x and y. Here, we assume that the vectors (X_i, Y_i) are assumed independent, the $p \times p$ covariance matrix of the rows of X, $EX^\top X$ is positive definite, and $E||(X,Y)||^3$ exist. Some more technical conditions are listed in the appendix.

The estimation method we shall consider for this model is defined by the (case-wise weighted) estimating equation below. Weighted least squares and M-estimation methods have been proposed by Koul *et al.* (1982), Zhou (1992a), Stute (1996), and Gross & Lai (1996). The estimator b can be expressed as the solution of the estimating equation

$$\sum_{i=1}^{n} w_i (Z_i - X_i^\top b) X_i = 0 \,. \tag{2.2}$$

In contrast to the 'synthetic data' approach of Koul *et al.* (1981), Leurgans (1987), and various generalizations, the 'case-wise weighted' approach never creates any new response values (i.e. synthetic data). Instead, it tries to recoup the effect of censored responses by properly weighting the uncensored responses. On the other hand, it does not require iteration in the calculation of the estimator, as opposed to the Buckley–James estimator.

Two different weighting schemes are known in the literature to determine the weights w_i. Stute (1996) ordered the Z_i such that $\delta_{(i)}$ is the censoring indicator δ corresponding to the ith order statistic $Z_{(i)}$, and rewrote the jumps of the Kaplan–Meier estimator of the marginal distribution of Y as

$$\Delta_1 = \delta_{(1)}/n \quad \text{and} \quad \Delta_i = \frac{\delta_{(1)}}{n-i+1} \prod_{j=1}^{i-1} \left(\frac{n-j}{n-j+1} \right)^{\delta_{(j)}} , \; i = 2, \cdots, n \,.$$

He used the Δ_i as weights in (2.2).

On the other hand, inverse censoring probability weights have been used in many different places, for example in van der Laan & Robins (2003) and Rotnitzky & Robins (2005). The weights there are given by

$$w_i^* = \frac{\delta_i}{1 - \hat{G}(Z_i)}$$

with $\hat{G}(\cdot)$ being the Kaplan–Meier estimator of the censoring distribution G based on $(Z_i, 1-\delta_i)$.

These two weighting schemes are in fact identical. Inverse censoring probability weighting is equivalent to weighting by the jumps of the Kaplan–Meier. Indeed, for all t,

$$[1 - \hat{F}(t)][1 - \hat{G}(t)] = 1 - \hat{H}(t) \tag{2.3}$$

where $\hat{F}(t)$ and $\hat{G}(t)$ are the Kaplan–Meier estimators for Y_i based on (Z_i, δ_i) and the censoring variable C_i based on $(Z_i, 1-\delta_i)$ respectively, and $\hat{H}(t)$ is the empirical distribution based on Z_i. From (2.3), we observe that when $t = Z_i$ with $\delta_i = 1$, then $\Delta_i[1 - \hat{G}(t)] = 1/n$, from which it follows that

$$\Delta_i = \frac{\delta_i}{n[1 - \hat{G}(Z_i)]} = \frac{w_i^*}{n} .$$

Also, Stute (1996) proposed

$$\hat{F}_{xy}(A) = \sum_{i=1}^{n} \Delta_i I_{\{(Z_i, X_i) \in A\}} \quad \text{for some set } A \text{ in } R^{(p+1)}$$

as a multivariate extension of the univariate Kaplan–Meier estimator. Based on these two observations, we call a solution to (2.2) with $w_i = \Delta_i$ *case-wise* weighted estimator.

3. Main Results

We define the *case-wise* empirical likelihood for the AFT correlation model. Reflecting the independent and identical distributions of the vectors (X_i, Y_i), although the observations are censored, we propose formulating the empirical likelihood *case-wise* as follows:

Consider the estimating equation $\int\int (y - x^\top \beta) x dF_{xy} = 0$. For any integrable function $\phi(x, y)$, equality holds for the two integrals $\int \phi(x, y) dF_{xy}$ and

$\int\int \phi(x,y)dF_{x|y}dF_y$, where $F_{x|y}$ denotes the conditional distribution of X given Y. Based on the data (X_i, Z_i, δ_i) $i = 1, \cdots n$, a reasonable estimator of $F_{x|y}$ when $y = Z_i$ and $\delta_i = 1$ is a point mass at X_i. In fact, using this conditional distribution coupled with the marginal distribution estimator of F_y, namely the Kaplan–Meier estimator, one obtains an estimator that is identical to Stute's $\hat{F}_{xy}$ mentioned above (see the appendix for more details of this equivalency).

Using this relationship, the *case-wise* empirical likelihood is

$$L_{xy}(F_y, F_{x|y}) = \prod 1 \prod_{\delta_i=1} p_i \prod_{\delta_i=0} (1 - \sum_{Z_j \le Z_i} p_j)\,,$$

where $p_i = dF_y[Z_i]$ is the probability that F_y places on the ith *case*. Since the conditional distribution $F_{x|y}$ will remain as a point mass throughout (as discussed above) and not change, we will from now on drop $F_{x|y}$ from L_{xy} and denote F_y simply as F. Similarly we drop the constant point mass from the likelihood:

$$L_{xy}(F_y) = \prod_{\delta_i=1} p_i \prod_{\delta_i=0} (1 - \sum_{Z_j \le Z_i} p_j)\,.$$

The likelihood ratio is

$$R_{xy}(b) = \frac{\sup\{L_{xy}(F) | F \in \tilde{\mathbb{F}}^b\}}{\sup\{L_{xy}(F) | F \in \mathfrak{F}\}}\,, \tag{3.1}$$

where $\mathfrak{F}$ denotes the set of univariate distributions that place positive probabilities on each uncensored case (as $L_{xy}(F) = 0$ for any F that places zero probability on some uncensored (Z_i, δ_i)), and $\tilde{\mathbb{F}}^b$ denotes a subset of $\mathfrak{F}$ that satisfies the constraints

$$\sum_{i=1}^n p_i \delta_i (Z_i - X_i^\top b) X_i = 0\,.$$

The constraint above can also be interpreted as

$$\sum_{Z_i} \sum_{X_j} p_i \delta_i (Z_i - X_j^\top b) X_j d_{ij} = 0 \quad \text{or} \quad \int\int (y - x^\top b) x d\hat{F}_{xy} = 0$$

where $d_{ij} = 1$ if and only if $i = j$ and zero otherwise, and $\hat{F}_{xy}$ is similar to Stute's estimator except that we identify $p_i = \Delta_i$.

The denominator of (3.1) is provided when F is the Kaplan–Meier estimator and the maximization is required for the numerator only, which can be obtained using the method of Zhou (2005b), among others. When b is the *case-wise* weighted estimator (the solution to estimating equation (2.2)), then $R_{xy}(b) = 1$ and confidence regions based on (3.1) are 'centered' at this estimator.

Consider the AFT correlation model equation (1.1) and testing the null hypothesis $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_0$. The following theorem contains the main result for the proposed *case-wise* empirical likelihood ratio statistic for least squares regression. A proof can be found in the appendix.

Theorem 1 *Consider the accelerated failure time correlation model as specified in Section 2.2. Under $H_0 : \beta = \beta_0$ and assuming regularity conditions (C1)-(C3) and (C6) from the appendix as well as the assumptions for Theorem 3.1 in Zhou (1992b), we have $-2 \log R_{xy}(\beta_0) \to \chi^2_p$ in distribution as $n \to \infty$.*

In many applications, inference is only sought for a part of the β vector. A usual approach is to "profile" out the components that are not under consideration. Profiling has been proposed for empirical likelihood in uncensored cases (e.g. Qin & Lawless, 1994). Under censoring, Lin & Wei (1992) proposed profiling with Buckley-James estimators. We similarly consider a profile empirical likelihood ratio: let $\beta = (\beta_1, \beta_2)$ where $\beta_1 \in R^q$ with $q < p$, is the part of the parameter under testing. Similarly we let $\beta_0 = (\beta_{10}, \beta_{20})$. A profile empirical likelihood ratio for β_1 is given by $\sup_{\beta_2} R_{xy}(\beta = (\beta_1, \beta_2)$. The below theorem shows that Wilks' theorem holds for the profile empirical likelihood.

Theorem 2 *Consider the accelerated failure time correlation model as specified in Section 2.2 and assume conditions of Theorem 1 above. Under $H_0 : \beta_1 = \beta_{10}$, we have $-2 \log \sup_{\beta_2} R_{xy}(\beta = (\beta_{10}, \beta_2)) \to \chi^2_q$ in distribution as $n \to \infty$.*

Similar results hold for censored quantile models when the τth conditional quantile of Y_i is modelled by

$$Q_\tau(\log T_i | X_i) = Q_\tau(Y_i | X_i) = X_i^\top \beta_\tau \,,$$

and, instead of Y_i, we observe $Z_i = \min(Y_i, C_i)$ and $\delta_i = I_{[Y_i \leq C_i]}$ for some censoring time variables C_i. This model may also be written as $Y_i = X_i^\top \beta_\tau + e_i$, but the error terms e_i are just independent random variables with zero τth quantiles. When $\tau = 0 \cdot 5$, this is the censored median regression, and Huang *et al.*

(2007) proposed a *case-wise* weighted estimator, which is a special case of our *case-wise* weighted estimator. Here, we propose the following *case-wise* empirical likelihood inference for the general censored quantile regression using

$$\mathbb{R}_{xy}(b) = \frac{\sup\{L_{xy}(F)|F \in \tilde{\mathbb{F}}^b\}}{\sup\{L_{xy}(F)|F \in \mathfrak{F}\}}\,,$$

where $\tilde{\mathbb{F}}^b$ denotes a subset of $\mathfrak{F}$ that satisfies the constraints

$$\sum_{i=1}^{n} p_i \delta_i \psi_\tau(Z_i - X_i^\top b) X_i = 0\,,$$

and $\psi_\tau(u)$ is the derivative of the so called check function $\rho_\tau(u) = u(\tau - I_{[u<0]})$ of Koenker & Basset (1978). Similar to the censored accelerated failure time model, the denominator of $\mathbb{R}_{xy}(b)$ is maximized by the Kaplan–Meier estimator, and thus when calculating $\mathbb{R}_{xy}(b)$ the maximization is only needed for the numerator.

Theorem 3 *Under $H_0 : \beta_\tau = \beta_0$ and assuming regularity conditions (C1)-(C6) listed in the appendix, for given τ, $-2\log \mathbb{R}_{xy}(\beta_0) \to \chi^2_p$ in distribution as $n \to \infty$. If $\beta_\tau = (\beta_1, \beta_2)$ with $\beta_1 \in R^q$ with $q < p$, under $H_0 : \beta_1 = \beta_{10}$, we have $-2\log \sup_{\beta_2} R_{xy}(\beta_\tau = (\beta_{10}, \beta_2)) \to \chi^2_q$ in distribution as $n \to \infty$.*

4. Simulation Studies

4.1 Simulation Study 1

We compared *case-wise* and *residual-wise* empirical likelihoods using the following three models: Homoscedastic errors with independent censoring (M1), heteroscedastic errors with independent censoring (M2), and homoscedastic errors with censoring dependent on x (M3).

$$\begin{aligned}
&\mathrm{M1}: Y_i = X_i^\top \beta + e_i\,, && C_i = \epsilon_i\,,\\
&\mathrm{M2}: Y_i = X_i^\top \beta + e_i \exp(X_i^\top \gamma)\,, && C_i = \epsilon_i\,,\\
&\mathrm{M3}: Y_i = X_i^\top \beta + e_i\,, && C_i = X_i^\top \eta + \epsilon_i\,,
\end{aligned}$$

where $X_i = (1, X_{1i})$ with $X_{1i} \sim U(0,1)$, $e_i \sim N(0,1)$ and ϵ_i is from a mixture of $N(3, 3^2)$ and $U(-2, 18)$. β, γ and η were chosen such that the censoring in each model amounts to 28·5% and the error heteroscedasticity in (M2) and the conditional dependency of C_i on X_i in (M3) is non-negligible. Due to the heteroscedastic errors, the R^2 of the least squares regression analysis of (M2)

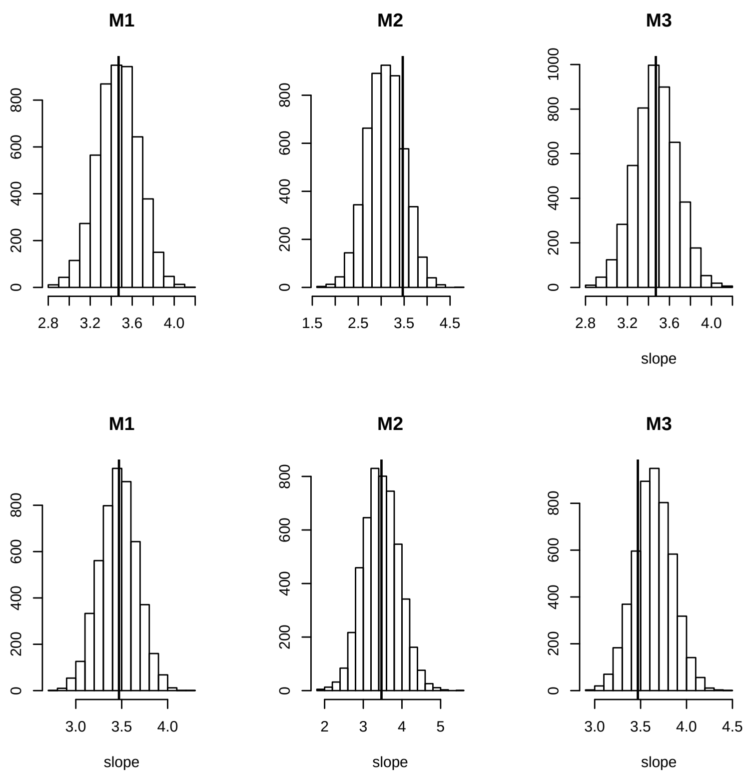

Figure 3.1: Histogram plots of slope estimates by the Buckley–James estimator and *case-wise* weighted estimator based on 5000 simulations with $n = 400$ with 28·5% censoring in each case. Vertical lines indicate the true value of the slope parameter. First row is Buckley–James estimator.

(without censoring) was on average reduced to 0·25 from 0·5 of an equivalent analysis of (M1), and an average R^2 of 0·28 was yielded for the least squares regression analysis of C_i on X_i for (M3).

We first examined the slope estimate by the *case-wise* weighted and the Buckley–James estimator. Figure 3.1 shows that the Buckley–James estimator is biased for the heteroscedastic errors model (M2), while the *case-wise* weighted estimator is not. When the censoring variables are dependent on X_i (M3), however, the *case-wise* weighted estimator is biased.

We confirmed that the differences in the data generating models require

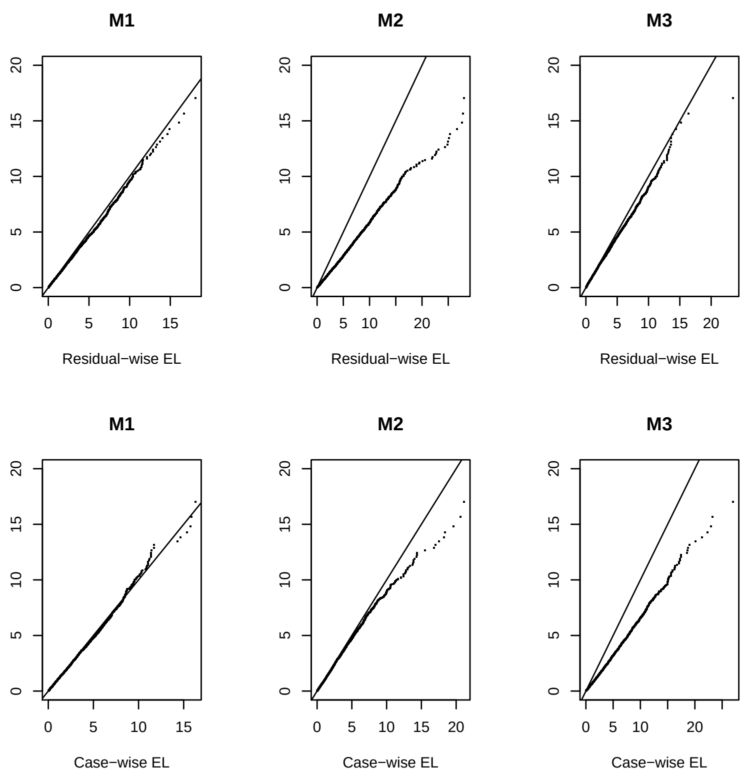

Figure 4.2: Q-Q plots of quantiles of χ^2_2 versus $-2\log R_e(\beta)$ (*residual-wise*) and $-2\log R_{xy}(\beta)$ (*case-wise*) respectively based on 5000 simulations with $n = 400$ with 28·5% censoring in each case.

the empirical likelihood to be formulated differently. Q-Q plots in Figure 4.2 show that only the *case-wise* empirical likelihood is valid for the heteroscedastic errors model (M2) and only the *residual-wise* likelihood is valid for covariate-dependent censoring (M3). Deviation from the limiting chi-squared distribution in the likelihood ratio statistic corresponds to bias in the estimates. Both empirical likelihoods were appropriate for (M1), as they are equivalent to the first order with the same limiting chi-squared distribution.

4.2 Simulation Study 2

We examined the performance of the *case-wise* empirical likelihood for censored quantile regression using the following three models that are similar to

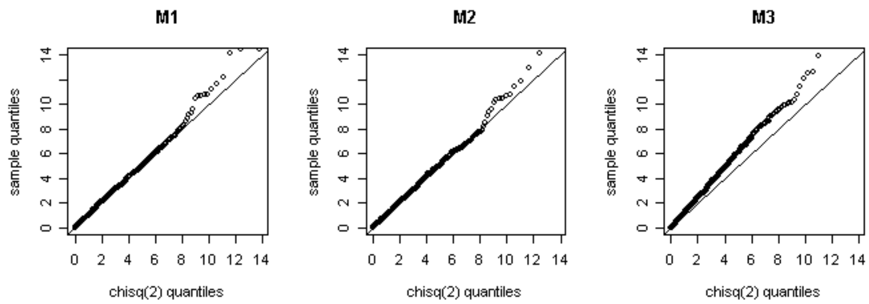

Figure 4.3: Q-Q plots of the quantiles χ^2_2 versus $-2\log R_{xy}(\beta_0)$ (*case-wise*) based on 1000 simulations with sample size $n = 200$ and about 30% censoring in each case.

those used in Simulation Study 1.

$$\begin{aligned} &\mathrm{M1} : Y_i = X_i^\top \beta + e_i\,, && C_i = \epsilon_i\,,\\ &\mathrm{M2} : Y_i = X_i^\top \beta + e_i(X_i + 1)\,, && C_i = \epsilon_i\,,\\ &\mathrm{M3} : Y_i = X_i^\top \beta + e_i\,, && C_i = X_{1i} + \epsilon_i\,, \end{aligned}$$

where $X_i = (1, X_{1i})$ with $X_{1i} \sim U(0, 1)$. The error $e_i \sim N(0, 0\cdot 75^2)$ in (M1) and (M3), $e_i \sim N(0, 0\cdot 5^2)$ in (M2). The parameter $\beta = (0\cdot 5, 1\cdot 5)$, and for the censoring $\epsilon_i \sim 0\cdot 5 + \exp(0\cdot 5)$ in (M1) and (M2), and $\epsilon_i \sim \exp(0.5)$ in (M3). The censoring percentage is about 30%. We fit $Q_\tau(Y_i|X_i)$ at $\tau = 0\cdot 25$. The quantile regression Q-Q plots in Figure 4.3 exhibit similar properties as the mean regression in Simulation Study 1. Specifically, the *case-wise* empirical likelihood approach for censored quantile regression is not adversely affected by heteroscedastic errors, but it is biased in the presence of covariate-dependent censoring.

4.3 Simulation Study 3

In order to compare the proposed method with the empirical likelihood method proposed by Qin & Jing (2001) and Li & Wang (2003), it is instructive to compare the estimators that form the basis of the proposed inferential methods. In the just mentioned articles, the estimator is based on the 'synthetic data' approach of Koul *et al.* (1981), we will refer to it as $\hat\beta_{KSV}$. The estimator considered in the present manuscript is based on Zhou (1992a) and Stute (1996), denoted $\hat\beta_{ZS}$.

Table 4.1: Simulation results comparing $\hat{\beta}_{KSV}$ with $\hat{\beta}_{ZS}$

	With intercept				Without intercept	
	$\hat{\beta}_{KSV}$		$\hat{\beta}_{ZS}$		$\hat{\beta}_{KSV}$	$\hat{\beta}_{ZS}$
	intercept	slope	intercept	slope	slope	slope
Mean	-0·00088	0·99964	-0·00122	1·00002	0·99899	0·99907
Variance	0·02061	0·02002	0·01395	0·01133	0·00312	0·00230

For the direct comparison, we assumed the simplest model from the previous section, namely (M1), and used the setting $X_i = (1, X_{1i})$, $X_{1i} \sim U(1, 0 \cdot 5^2)$, $e_i \sim N(0, 0 \cdot 5^2)$, $\beta = (0, 1)$, and $\epsilon_i \sim N(6 \cdot 1, 4^2)$. Table 4.1 provides results based on 10,000 simulations for sample size $n = 100$. Clearly, both estimators appeared to be unbiased, but the variance of the components of $\hat{\beta}_{ZS}$ was at least 30% and 40% smaller for the model including the intercept and at least 25% smaller for the without intercept model. The simulation indicates that the estimator $\hat{\beta}_{KSV}$ is far from efficient. Therefore, any inference method based on this estimator can not be expected to perform as well as inferential procedures based on $\hat{\beta}_{ZS}$.

5. Small-Cell Lung Cancer Data

We consider a lung cancer data set (Maksymiuk *et al.* 1994) that has been analysed by Ying *et al.* (1995) using median regression, and by Huang *et al.* (2007) using a least absolute deviations method in the accelerated failure time (AFT) model. In this study, 121 patients with limited-stage small-cell lung cancer were randomly assigned to one of two different treatment sequences A and B, with 62 patients assigned to A and 59 patients to B. Each death time was either observed or administratively censored, and the censoring variable did not depend on the covariates *treatment* and *age*.

Denote the treatment indicator variable by X_{1i}, and the entry age for the ith patient by X_{2i}, where $X_{1i} = 1$ if the patient is in group B. Let Y_i be the base 10 logarithm of the ith patient's failure time. We assume the AFT model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \sigma(X_{1i}, X_{2i})\epsilon_i \ .$$

The estimated parameter values obtained using the approach described in this paper need to be equal to the ones from Huang *et al.* (2007), provided weighting has been done in the same way. The major difference is in inference about the parameters, where empirical likelihood has the advantage that it is not necessary to estimate the asymptotic variance of the estimator in order to perform hypothesis tests and to construct confidence regions. An empirical likelihood confidence region for this data set is displayed in Figure 5.4.

The following median regression estimates were obtained by Ying *et al.* (1995) and Huang *et al.* (2007).

$$\hat{\beta}_0 = 3{\cdot}028,\ \hat{\beta}_1 = -0{\cdot}163, \ \text{ and }\ \hat{\beta}_2 = -0{\cdot}004 \qquad \text{(Ying } et\ al. \text{ 1995)}$$
$$\hat{\beta}_0 = 2{\cdot}693,\ \hat{\beta}_1 = -0{\cdot}146, \ \text{ and }\ \hat{\beta}_2 = 0{\cdot}001 \qquad \text{(Huang } et\ al. \text{ 2007)}$$

Huang *et al.* (2007) did not always treat the largest Y observation as uncensored. This resulted in weights that sum to less than one (the sum of the weights without the last observation is 0·85). We recommend to always treat the largest Y observation as uncensored so that the weights always sum to one. Otherwise, the estimation may be biased since the information from the largest Y observation is ignored. Treating the largest Y as uncensored, the median regression estimates become

$$\hat{\beta}_0 = 2{\cdot}603,\ \hat{\beta}_1 = -0{\cdot}263, \ \text{ and }\ \hat{\beta}_2 = 0{\cdot}0038 \qquad \text{(with last weight).}$$

While the different approaches lead to the same conclusions with regard to the treatment (predictor X_1), the data analysis results suggest possibly conflicting interpretations regarding the role of the predictor X_2 (entry age). Ying *et al.* (1995) calculate a negative coefficient (-0·004), while Huang *et al.* (2007) and the approach proposed in this manuscript obtain a positive estimate (0·001 and 0·003837, respectively, depending on whether or not the largest observation is treated as uncensored). However, the confidence interval for β_2 that is provided by Ying *et al.* (1995) includes zero (-0·0162, 0·003), and so does the empirical likelihood confidence interval based on the newly proposed inference method (-0·0024, 0·0151). Thus, the different methods do agree that the predictor "entry age" is not significant. The empirical likelihood based confidence interval is about 10% shorter than the interval provided by Ying *et al.* (1995). This is not surprising when considering that (a) the latter is not based on the likelihood and may therefore not be efficient, and (b) the former does not have to rely on inverting the estimated variance-covariance matrix of an estimating function, which could lead to unstable estimates. Perhaps this explains the differences in the estimated intercept that are leading to rather different estimated median survival times between $10^{2 \cdot 6} = 398$ and $10^3 = 1000$ days.

6. Concluding Remarks

We note some differences and similarities of the *case-wise* and *residual-wise*

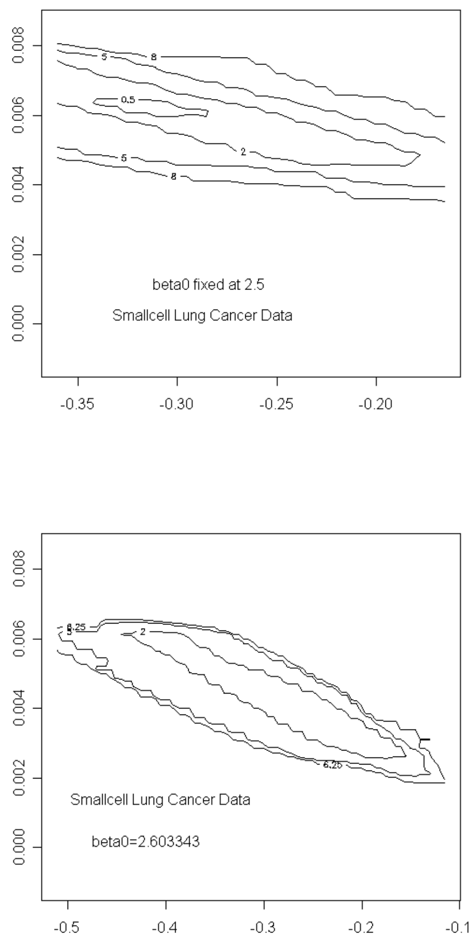

Figure 5.4: Confidence Region for (β_1, β_2) in the small-cell lung cancer data. The original parameter space is 3-dimensional, therefore only two two-dimensional cuts at $\beta_0 = 2 \cdot 5$ and $\beta_0 = 2 \cdot 603$ are displayed. The contour lines correspond to different critical values.

empirical likelihood. A trade-off exists between assumptions on the error terms and the censoring time variables: The homoscedastic errors assumption of the regression model is relaxed for the correlation model, while the conditional independence assumption on C_i is strengthened such that the C_i need to be independent of the random vector (X_i, Y_i) (or at least satisfy assumption (C1) in the appendix). This is confirmed in the simulation study: The *case-wise* empirical likelihood is biased when C_i are only conditionally independent of Y_i, while the *residual-wise* one is biased in the presence of heteroscedastic errors. Hence, the *case-wise* empirical likelihood is more appropriate when error heteroscedasticity is more a concern than independent censoring. The computation of (3.1) does not involve X_i with $\delta_i = 0$. Therefore, the *case-wise* empirical likelihood allows missingness in X_i with $\delta_i = 0$, while the *residual-wise* empirical likelihood does not.

Nevertheless both methods provide a computationally simple inference method, and software is readily available in *R*. This contrasts with other methods that require a direct estimation of the covariance matrices of the estimators. For example, the median regression approach proposed by Ying *et al.* (1995) requires for inference the inversion of an estimated variance-covariance matrix which may be unstable, in particular at the tails of the survival functions. Also, this approach is expected to lack efficiency because it is not based on the likelihood.

Portnoy (2003) investigated the censored quantile regression process using a recursive algorithm that fits the entire quantile regression process successively from below. The model accommodates both heteroscedastic errors and conditionally independent censoring. His method, however, requires a strong assumption that the entire qunatile process is linear in x_i. This implies that the validity of quantile estimates at, for example the median depends on the linearity of the entire conditional functionals at all lower quantiles and non-linear relations at any of the lower quantiles will bias the median estimates. Peng & Huang (2008)'s method is similarly restricted by the global linearity assumption. Wang & Wang relaxed the global linearity assumption by a local Kaplan-Meier method. However, their method is subject to the so-called "curse-of-dimensionality". For inference, all these methods rely on resampling. The proposed *case-weighted* estimator and accompanying *case-wise* empirical likelihood method provides a simple and practical alternative with only slightly more restrictive conditions.

Acknowledgment The research was supported in part by National Science Foundation grant DMS-0604920.

Appendix

We let $(X_1, Y_1), \cdots (X_n, Y_n)$ be a random sample of vectors from a joint distribution $F(x, y)$, where the Y_i are subject to right censoring, and define a function $\phi(x, y)$ such that $\phi(x, y) = (y - x^\top \beta)x$ for the AFT model, and $\phi(x, y) = \psi_\tau(y - x^\top \beta)x$ for the quantile regression model below.

A.1 Assumptions

Some of these assumptions have been formulated in related works (Zhou 1992b; Stute 1996; Huang *et al.* 2007).

(C1) The (transformed) survival times Y_i and the censoring times C_i are independent. Furthermore, $\mathrm{pr}(Y_i \le C_i | X_i, Y_i) = \mathrm{pr}(Y_i \le C_i | Y_i)$.

(C2) The survival functions $\mathrm{pr}(Y_i \ge t)$ and $\mathrm{pr}(C_i \ge t)$ are continuous and either $\xi_Y < \xi_C$ or $\forall t < \infty, \mathrm{pr}(Z_i \ge t) > 0$. Here, for any random variable U, ξ_U denotes the right end point of the support of U.

(C3) The X_i are independent, identically distributed according to some distribution with finite, nonzero variance, and they are independent of Y_i and of C_i.

(C4) Let $F_e(\cdot|x)$ be the conditional distribution of the e_i given $X = x$, and $f_e(\cdot|x)$ the corresponding conditional density function. For the given τ, $F_e(0|x) = \tau$, and $f_e(u|x)$ is continuous in u in a neighbourhood of 0 for almost all x.

(C5) $E(XX^\top f_e(0|X))$ is finite and nonsingular.

(C6) $\sigma^2_{KM}(\phi) < \infty$ where $\sigma^2_{KM}(\phi)$ is the asymptotic variance of $\sqrt{n}\sum_i \phi(x_i, t_i)\Delta \hat{F}_{KM}(t_i)$.

A.2 Equivalency Between Two Kaplan–Meier Estimators

We show that in the correlation model, the Kaplan–Meier estimator of the marginal distribution F_y is identical to Stute's $\hat{F}_{xy}$, multivariate extension of the univariate Kaplan–Meier estimator. Recall that Z_i are the censored values of Y_i: $Z_i = \min(Y_i, C_i)$, along with the censoring indicator δ_i.

Under the assumption of (C1), a reasonable estimate for the marginal distribution of Y is the Kaplan–Meier estimate $\hat{F}_{KM}$ based on Z_i and δ_i. A reasonable estimate of the conditional distribution $F(x|y)$ is

$$\hat{F}(x|y = z_i) = \text{point mass at } x_i, \text{ if } \delta_i = 1; \tag{3.1}$$

otherwise $\hat{F}(x|y = z_i)$ is left undefined. This gives rise to an estimator of the joint distribution, identical to the one proposed by Stute (1996).

For the function of the random vector $\phi(x, y)$, the expectation

$$E\phi(x, y) = \int \phi(x, y)dF(x, y) = \int\int \phi(x, y)dF(x|y)dF(y)$$

can be estimated using the above joint distribution estimator, that is, the Kaplan–Meier estimator for the marginal distribution, and the conditional distribution defined in (3.1):

$$\sum_i [\sum_j \phi(x_j, t_i)I_{[j=i]}\delta_i]\Delta\hat{F}_{KM}(t_i) = \sum_i \phi(x_i, t_i)\Delta\hat{F}_{KM}(t_i).$$

If we take the function ϕ to be the indicator function $I[s \leq x; t \leq y]$, this also defines an estimator of the joint distribution $F(x, y)$.

Furthermore, by the law of large numbers and central limit theorem for this estimator (see Stute (1996)) and the assumptions $\int \phi(x, t)dF(x, t) = 0$, we have, as $n \to \infty$,

$$\sqrt{n}\left(\sum_{i=1}^{n} \phi(x_i, t_i)\Delta\hat{F}_{KM}(t_i)\right) \to N(0, \sigma^2_{KM}(\phi))$$

in distribution, where $\sigma^2_{KM}(\phi) < \infty$ by (C6). The variance $\sigma^2_{KM}(\phi)$ can be written in several different but equivalent ways. See Akritas (2000) for a form that the following estimate is based on. The variance $\sigma^2_{KM}(\phi)$ can be consistently estimated by

$$\hat{\sigma^2} = \sum_i [\phi(x_i, t_i) - \bar{\phi}(t_i)]^2 \frac{\Delta\hat{F}_{KM}(t_i)}{1 - \hat{G}_{KM}(t_i)} \tag{3.2}$$

where $\bar{\phi}(s) = \sum_{j:t_j>s} \phi(x_j, t_j)\Delta\hat{F}_{KM}(t_j)/[1 - \hat{F}_{KM}(s)]$, the so called advanced-time transformation. See Akritas (2000) for details.

A.3 Outline of the Proofs for Theorems 1 and 3.

Recall that when computing the numerator in the likelihood ratio (3.1), we need to compute the supremum of the empirical likelihood over all F that is dominated by the Kaplan–Meier estimator and satisfy the constrain equations.

We first construct those distributions indexed by an $h(\,)$ function. Later we shall take the supremum over h.

Notice that any distribution that are dominated by the Kaplan–Meier estimator can be written as (by its jumps)

$$\Delta F(t_i) = \Delta \hat{F}_{KM}(t_i) \frac{1}{1+h(t_i)}$$

for some h. On the other hand this distribution may not satisfy the constrain equations. Motivated by the above we define F that are dominated by the Kaplan–Meier *and* satisfy the constraint equation (3.4) by

$$\Delta F_\lambda(t_i) = \Delta \hat{F}_{KM}(t_i) \times \frac{1}{1+\lambda h(x_i, t_i)} \times \frac{1}{C(\lambda)}, \quad i = 1, 2, ..., n, \tag{3.3}$$

and $C(\lambda)$ is just a normalizing constant

$$C(\lambda) = \sum_{i=1}^{n} \frac{\Delta \hat{F}_{KM}(t_i)}{1+\lambda h(x_i, t_i)}.$$

The parameter λ in the above is chosen so that the resulting F satisfies the constraint equations, i.e. λ is the solution of

$$\int \phi(x,t) dF_\lambda(x,t) = \frac{1}{C(\lambda)} \sum_{i=1}^{n} \Delta \hat{F}_{KM}(t_i) \frac{\phi(x_i, t_i)}{1+\lambda h(x_i, t_i)} = 0. \tag{3.4}$$

We suppose the function $h = h(x, y)$ satisfies the following requirements:

(i) $\|h(x,y)\| = 1$

(ii) $|\int \phi(x,y) h(x,y) dF(x,y)| \geq \epsilon > 0$.

Lemma A below gives an asymptotic representation of this λ value that makes F satisfy the constraint equation. We denote the parameter value for this unique distribution as λ_0. With this λ_0, the distribution F_{λ_0} satisfies the conditions $F_{\lambda_0} \ll \hat{F}_{KM}$ *and* $\int \phi(x,t) dF_{\lambda_0}(x,t) = 0$.

When restricted to these distributions (indexed by h), we define a (profile) empirical likelihood ratio function as follows:

$$\mathcal{R}_h = \left\{ \frac{L_{xy}(F_{\lambda_0})}{L_{xy}(\hat{F}_{KM})} \mid h \text{ satisfies } (i), (ii) \text{ above} \right\},$$

where L_{xy} is defined in Section 3. By Theorem A below we have

$$-2\log R_{xy} = -2\log(\sup_h \mathcal{R}_h) = \chi^2_{(p)} + o_p(1) \ ,$$

and the proof is complete.

Lemma A *Assume all the conditions in Theorem 3.1. and (i), (ii) above. Then,*
(1) With probability tending to one as $n \to \infty$, λ_0 is well defined;
(2) $\lambda_0 = O_p(n^{-1/2})$ as $n \to \infty$;
(3) λ_0 has the representation given in (3.5) below for $n \to \infty$.

Proof. Expanding (3.4), we have

$$\begin{aligned} 0 &= \sum_{i=1}^n \Delta\hat{F}_{KM}(t_i)\frac{\phi(x_i,t_i)}{1+\lambda_0 h(x_i,t_i)} \\ &= \sum_{i=1}^n \phi(x_i,t_i)\Delta\hat{F}_{KM}(T_i) - \lambda_0\sum_{i=1}^n \phi(x_i,t_i)h(x_i,t_i)\Delta\hat{F}_{KM}(t_i) \\ &\quad +\lambda_0^2\sum_{i=1}^n \frac{\phi(x_i,t_i)h^2(x_i,t_i)}{1+\lambda_0 h(x_i,t_i)}\Delta\hat{F}_{KM}(t_i). \end{aligned}$$

It follows from this that

$$\lambda_0 = \frac{\sum_{i=1}^n \phi(x_i,t_i)\Delta\hat{F}_{KM}(t_i)}{\sum_{i=1}^n \phi(x_i,t_i)h(x_i,t_i)\Delta\hat{F}_{KM}(t_i)} + o_p(n^{-1/2}). \tag{3.5}$$

The uniqueness of λ_0 follows from observing that in display (3.4), $C(\lambda)$ is positive, and the remaining term is strictly monotone in λ. Existence requires feasibility of the parameter under hypothesis (for a detailed discussion of feasibility in censored empirical likelihood see Pan & Zhou 2002). However, for $n \to \infty$, the probability that the true parameter is feasible approaches one.

Theorem A *If the conditions in Lemma A hold, then, as $n \to \infty$*

$$-2\log\mathcal{R}_h = -\lambda_0 f''(0)\lambda_0 + o_p(1)$$

where $f(\cdot)$ is defined in (3.6) below.

Furthermore,

$$-2\log R_{xy} = -2\log(\sup_h \mathcal{R}_h) = \chi^2_{(p)} + o_p(1) \ .$$

Proof. For simplicity assume $p = 1$. Define a function of λ (the marginal empirical likelihood as a function of λ)

$$f(\lambda) = \log \prod_{i=1}^{n} (\Delta F_\lambda(T_i))^{\delta_i} (1 - F_\lambda(T_i))^{1-\delta_i}, \tag{3.6}$$

where $|\lambda| \le C|\lambda_0|$. From the definition, we can see that

$$f(0) = \log \prod_{i=1}^{n} (\Delta \hat{F}_{KM}(T_i))^{\delta_i} (1 - \hat{F}_{KM}(T_i))^{1-\delta_i} = L_{xy}(\hat{F}_{KM}).$$

By Lemma A, $\lambda_0 = O_p(n^{-1/2})$. Hence we can apply Taylor's expansion for $f(\lambda_0)$:

$$f(\lambda_0) = f(0) + \lambda_0 f'(0) + \frac{\lambda_0^2}{2} f''(0) + \frac{\lambda_0^3}{3!} f'''(\xi), \qquad |\xi| \le |\lambda_0|.$$

Substituting (3.3) in (3.6),

$$\begin{aligned} f(\lambda) \; &= \; \sum_{i=1}^{n} \delta_i \log \Delta \hat{F}_{KM}(T_i) - \sum_{i=1}^{n} \delta_i \log(1 + \lambda h(x_i, T_i)) - n \log \left(\sum_{i=1}^{n} \frac{\Delta \hat{F}_{KM}(T_i)}{1 + \lambda h(x_i, T_i)} \right) \\ &\quad + \sum_{i=1}^{n} (1 - \delta_i) \log \left(\sum_{j:T_j > T_i} \frac{\Delta \hat{F}_{KM}(T_j)}{1 + \lambda h(x_j, T_j)} \right). \end{aligned}$$

Some tedious but straightforward calculation shows that $f'(0) = 0$ and that the second derivative of f with respect to λ, evaluated at $\lambda = 0$ is

$$\begin{aligned} f''(0) \; &= \; n\Big(\sum_{i=1}^{n} h(x_i, T_i) \Delta \hat{F}_{KM}(T_i)\Big)^2 - n \sum_{i=1}^{n} h^2(x_i, T_i) \Delta \hat{F}_{KM}(T_i) \\ &+ \; \sum_{i=1}^{n} (1 - \delta_i) \frac{\sum_{j:T_j > T_i} h^2(x_j, T_j) \Delta \hat{F}_{KM}(T_j)}{1 - \hat{F}_{KM}(T_i)} - \sum_{i=1}^{n} (1 - \delta_i) \frac{\Big(\sum_{j:T_j > T_i} h(x_j, T_j) \Delta \hat{F}_{KM}(T_j)\Big)^2}{(1 - \hat{F}_{KM}(T_i))^2}. \end{aligned}$$

Similar calculations show that the third derivative of f evaluated at ξ is $f'''(\xi) = o_p(n^{2/3})$. Now

$$\begin{aligned} -2 \log \mathcal{R}_h \; &= \; 2[f(0) - f(\lambda_0)] \\ &= \; 2 \left(f(0) - f(0) - \lambda_0 f'(0) - \frac{\lambda_0^2}{2} f''(0) - \frac{\lambda_0^3}{3!} f'''(\xi) \right) \\ &= \; -\lambda_0^2 f''(0) - \frac{\lambda_0^3}{3} f'''(\xi) \\ &= \; n \lambda_0^2 \frac{-f''(0)}{n} + o_p(1). \end{aligned}$$

This proves the first assertion. Rewriting the above using Lemma A, we obtain

$$-2\log\mathcal{R}_h = \frac{[\sqrt{n}\sum\phi(x_i,t_i)\Delta F_{KM}(t_i)]^2}{\hat{\sigma^2}} \times r_h + o_p(1)$$

where

$$r_h = \frac{\hat{\sigma^2}}{[\sum\phi(x_i,t_i)h(x_i,t_i)\Delta F(t_i)]^2}\frac{-f''(0)}{n} \ ,$$

and $\hat{\sigma^2}$ is given in (3.2). Finally

$$\begin{aligned} -2\log\mathcal{R}_{xy} &= -2\log(\sup_h \mathcal{R}_h) = \inf_h -2\log\mathcal{R}_h \\ &= \frac{[\sqrt{n}\sum\phi(x_i,t_i)\Delta F_{KM}(t_i)]^2}{\hat{\sigma^2}} \ \inf_h r_h + o_p(1). \end{aligned}$$

It can be shown by a Cauchy Schwarz inequality and the definition of $\hat{\sigma^2}$ in (3.2) that for any sample size n the infimum of r_h over h is 1. Thus by Stute's central limit theorem and the Slutsky Lemma we have

$$-2\log\mathcal{R}_{xy} = \frac{[\sqrt{n}\sum\phi(x_i,t_i)\Delta F_{KM}(t_i)]^2}{\hat{\sigma^2}} + o_p(1) = \chi^2_{(1)} + o_p(1) \ . \tag{3.7}$$

Finally we show $r_h \geq 1$. Recall that we already calculated $f''(0)$ before. After several tedious simplifications using the self-consistency property of the Kaplan–Meier estimator among other things, we have

$$-f''(0)/n = \sum_{i=1}^{n}(h(x_i,t_i) - \bar{h}(t_i))^2[1-\hat{G}_{KM}(t_i)]\Delta\hat{F}_{KM}(t_i) \ .$$

An application of the Cauchy Schwarz inequality shows that $r_h \geq 1$, and the lower bound 1 is achieved when

$$h(x_i,t_i) - \bar{h}(t_i) = \frac{\phi(x_i,t_i) - \bar{\phi}(t_i)}{1-\hat{G}_{KM}(t_i)} \ .$$

Proof of Theorem 2 Note that the first equality in the equation (3.7) in a multivariate version is simply

$$-2\log\mathcal{R}_{xy} = \hat{\top}\theta\hat{\Sigma}^{-1}\hat{\theta} + o_p(1) \,,$$

where $\hat{\Sigma}$ denote a multivariate version of $\hat{\sigma}$ defined in (3.2) and $\hat{\theta} = \sqrt{n}(\sum\phi(x_i,t_i)\Delta\hat{F}_{KM}(t_i))$. Standard result on the profiling of the quadratic form then lead to the conclusion of Theorem 2 (see e.g.) Basawa & Koul (1980)).

References

Akritas, M. (2000). The central limit theorem under censoring, *Bernoulli* **6**, 1109–1120.

Basawa, I. V. & Koul, H. L. (1980). Large-sample statistics based on quadratic dispersion, *Int Statist Rev.* **56**, 199–219.

Buckley, J. J. & James, I. R. (1979), Linear Regression With Censored Data, *Biometrika*, **66**, 429–36.

Freedman, D. A. (1981), Bootstrapping Regression Models, *Annals of Statistics*, **9**, 1218–1228.

Gross, S. & Lai, T. L. (1996), Nonparametric Estimation and Regression Analysis With Left Truncated and Right Censored Data, *Journal of the American Statistical Association*, **91**, 1166–1180.

Huang, J., Ma, S. & Xie, H. (2007), Least Absolute Deviations Estimation for the Accelerated Failure Time Model, *Statistica Sinica*, **17**, 1533–1548.

Jin, Z., Lin, D. Y., Wei, L. J. & Ying, Z. (2003), Rank-Based Inference for the Accelerated Failure Time Model, *Biometrika*, **90, 2**, 341–353.

Kalbfleisch, J. D. & Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data, Second Edition*, New York: Wiley.

Kaplan, E. L. & Meier, P. (1958), Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, **53, 282**, 457–481.

Koenker, R. & Bassett, G. (1978), Regression Quantiles, *Econometrica*, **46, 1**, 33–50.

Koul, H., Susarla, V. & Van Ryzin, J. (1981), Regression Analysis With Randomly Right-Censored Data, *Annals of Statistics*, **9**, 1276–1288.

Koul, H., Susarla, V. & Van Ryzin, J. (1982), Least Squares Regression Analysis With Censored Survival Data, In *Topics in Applied Statistics*, Ed. Y. P. Chaubey and T. D. Dwivedi, New York: Marcel Dekker, pp. 151–165.

Lai, T. L. & Ying, Z. L. (1991), Large Sample Theory of a Modified Buckley-James Estimator for Regression Analysis With Censored Data, *Annals of Statistics*, **19, 3**, 1370–1402.

Leurgans, S. (1987), Linear Models, Random Censoring and Synthetic Data, *Biometrika*, **74, 2**, 301–309.

Li, G. & Wang, Q. H.(2003), Empirical Likelihood Regression Analysis for Right Censored Data, *Statistica Sinica*, **13**, 51–68.

Lin, J. S. & Wei, L. J. (1992), Linear Regression Analysis Based on Buckley-James Estimating Equation, *Biometrics*, **48**, 679–681.

Maksymiuk, A. W., Jett, J. R., Earle, J. D., Su, J. Q., Diegert, F. A., Mailliard, J. A., Kardinal, C. G., Krook, J. E., Veeder, M. H., Wiesenfeld, M., Tschetter, L. K. & Levitt, R. (1994), Sequencing and Schedule Effects of Cisplatin Plus Etoposide in Small Cell Lung Cancer, Results of a North Central Cancer Treatment Group Randomized Clinical Trial, *Journal of Clinical Oncology*, **12**, 70–76.

Owen, A. (1991), Empirical Likelihood for Linear Models, *Annals of Statistics*, **19, 4**, 1725-1747.

Pan, X.-R. & Zhou, M. (2002), Empirical Likelihood Ratio in Terms of Cumulative Hazard Function for Censored Data, *Journal of Multivariate Analysis* **80**, 166–188.

Peng, L. and Huang, Y. (2008). Survival Analysis with Quantile Regression Models. *J. Amer. Statist. Assoc.*, **103**, 637-649.

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

Qin, G. S. & Jing, B. Y. (2001), Empirical Likelihood for Censored Linear Regression, *Scandinavian Journal of Statistics*, **28**, 661–673.

Qin, J. & Lawless, J. (2001), Empirical likelihood and general estimating equations, *Ann. Statist.*, **22**, 300–325.

Rotnitzky A. & Robins J. M. (2005), Inverse Probability Weighted Estimation in Survival Analysis, *Encyclopedia of Biostatistics.*

Stute, W. (1996), Distributional Convergence Under Random Censorship When Covariables are Present, *Scandinavian Journal of Statistics* **23**, 461–471.

van der Laan, M. & Robins, J. M. (2003), *Unified Methods for Censored Longitudinal Data and Causality*, New York: Springer.

Wang, H., and Wang, L. (2009), Locally Weighted Censored Quantile Regression. *J. Amer. Statist. Assoc.*, **104**, 1117-1128.

Ying, Z., Jung, S. H. & Wei, L. J. (1995), Survival Analysis with Median Regression Models, *Journal of the American Statistical Association* **90, 429**, 178–184.

Zhou, M. (1992a), M-Estimation in Censored Linear Models, *Biometrika*, **79, 4**, 837–841.

Zhou, M. (1992b), Asymptotic Normality of the 'Synthetic Data' Regression Estimator for Censored Survival Data, *Annals of Statistics*, **20, 2**, 1002–1021.

Zhou, M. (2005a), Empirical Likelihood Analysis of the Rank Estimator for the Censored Accelerated Failure Time Model, *Biometrika*, **92**, 492–498.

Zhou, M. (2005b), Empirical Likelihood Ratio With Arbitrarily Censored/Truncated Data by a Modified EM algorithm, *Journal of Computational and Graphical Statistics*, **14**, 643–656.

Zhou, M. & Li. G. (2008), Empirical Likelihood Analysis of the Buckley-James Estimator, *J. Multivariate Analysis*, **99**, 649-664. in: *The Encyclopedia of Biostatistics, 2nd edition*, Ed. P. Armitage and T. Coulton, New York: Wiley.

Mai Zhou, Department of Statistics, University of Kentucky, Lexington, Kentucky 40506-0027, USA

E-mail: mai@ms.uky.edu

Mi-Ok Kim, Division of Biostatistics and Epidemiology, Cincinnati Children's Medical Center MLC 5041, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039

E-mail: MiOk.Kim@cchmc.org

Arne C. Bathke, Department of Statistics, University of Kentucky, Lexington, Kentucky 40506-0027, USA

E-mail: arne@uky.edu