

Empirical Likelihood Analysis for the Heteroscedastic Accelerated Failure Time Model

BY MAI ZHOU

Department of Statistics, University of Kentucky, Lexington, KY 40506 USA
mai@ms.uky.edu

MI-OK KIM

Center for Epidemiology and Biostatistics, Cincinnati Children's Medical Center
MLC 5041, 3333 Burnet Avenue, Cincinnati, OH 45229-3039
miok.kim@cchmc.org

AND ARNE BATHKE

Department of Statistics, University of Kentucky, Lexington, KY 40506 USA
arne@ms.uky.edu

SUMMARY

In this manuscript, we discuss the distinction of two types of data generating scheme for the accelerated failure time model. We identify two different empirical likelihood formulations under random right censoring, *case-wise* and *residual-wise*, each reflecting the relevant features of the stochastic model assumed to have generated the data. We specifically propose the *case-wise* empirical likelihood as a computationally simple inference method for the accelerated failure time model with heteroscedastic errors. A nonparametric version of Wilks' theorem is shown to hold for the resulting likelihood ratio. The results are also applicable to censored quantile regression.

Some key words: Accelerated failure time model; Empirical likelihood; Wilks' theorem; Censored quantile regression

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 ,

$$\log(T_i) = Y_i = X_i\beta + e_i, \quad (1)$$

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 $\delta_i = I_{[Y_i \leq C_i]}$ 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 analyzing 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. In this paper, we consider empirical likelihood as a computationally simple inference method for the AFT model, particularly for the model that admits heteroscedastic errors.

Before we can formulate and study the empirical likelihood, we need to distinguish two different types of data generating schemes for the accelerated failure time model under random right censoring. In the accelerated failure time *regression* model, X is fixed and the errors are “homoscedastic”. In the accelerated failure time *correlation* model, X is random and the vectors (X_i, Y_i) are “homoscedastic”, while the errors are in general “heteroscedastic”. 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 not only different estimators unlike the uncensored case of Freedman (1981), but also different assumptions on the censoring times C_i . We observe a trade-off 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) .

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 the empirical likelihood method 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 empirical likelihood construction be based on the homoscedasticity of the errors 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.

We show that a non-parametric version of Wilks’ theorem holds for the empirical likelihood ratio of the *case-wise* likelihood. We also show that the results are applicable to censored

quantile regression. It may be emphasized that the computation involved in this likelihood is simple, not requiring resampling. Hence we do not include other empirical likelihood approaches in the discussion that do not admit chi-squared limiting distributions and require calibration via resampling.

2. Correlation and Regression Model

In this section, we summarize the main characteristics of the two data generating models and corresponding estimators for the accelerated failure time model. Our main focus will be on the correlation AFT model. For better comparison, we also 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 (see Freedman, 1981). The true value of the p -dimensional parameter vector β solves $\int (y - x\beta)x dF_e = 0$, where F_e denotes the error distribution. The main assumptions can be summarized as follows.

Condition 1 *The covariates, x_1, \dots, x_n , are row vectors of p -dimensional fixed constants, forming a matrix of full rank.*

Condition 2 *The errors, e_1, \dots, e_n , are independent, with common distribution F_e having mean 0 and finite variance σ^2 : both F_e and σ^2 are unknown.*

Condition 3 *The censoring time variables C_1, \dots, 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 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 $\delta_i = I_{[Y_i \leq C_i]}$. Let b be a vector, and define the residuals with respect to b

as $r_i(b) = z_i - x_i b$. 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

$$L_e(F) = \prod_{\delta_i=1} p_i \prod_{\delta_i=0} (1 - \sum_{r_j(b) \leq r_i(b)} p_j), \quad (2)$$

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

$$R_e(b) = \frac{\sup\{L_e(F) | F \in \mathbb{F}^b\}}{\sup\{L_e(F) | F \in \mathfrak{F}^b\}}, \quad (3)$$

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 (3) is maximized by the Kaplan-Meier estimator (Kaplan and Meier, 1958) 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 (3) are ‘centered’ at the Buckley-James estimate. By formulating similar constraints with respect to rank estimators, Zhou (2005) proposed *residual-wise* likelihood for a log-rank or Gehan-type of estimators. In each case, the resulting likelihood ratio admits chi-squared limiting distributions (Zhou & Li, 2008; Zhou, 2005).

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 $\iint (y - x\beta)x dF_{xy} = 0$, where F_{xy} denote the joint distribution of x and y . The main assumptions are as follows.

Condition 4 *The vectors (X_i, Y_i) are assumed independent, with a common unknown $(p + 1)$ variate distribution F_{xy} .*

Condition 5 *The $p \times p$ covariance matrix of the rows of X , $EX^\top X$ is positive definite and $E\|(X, Y)\|^3$ exist.*

Condition 6 *The censoring time variables C_1, \dots, C_n , are independent with common unknown distribution G and independent of (Y_i, X_i) .*

The estimation method we shall consider for this model is defined by the (casewise weighted) estimating equation below. Weighted least squares and M-estimation methods have been proposed by Koul *et. al.* (1982), Zhou (1992b), Stute (1993, 1996), and Gross & Lai (1996). The estimator b can be expressed as the solution of the estimating equation

$$\sum_{i=1}^n w_i (Z_i - X_i b) X_i = 0. \quad (4)$$

In contrast to the ‘synthetic data’ approach of Koul et al. (1981), Leurgans (1987), and various generalizations, the ‘casewise 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 (1993, 1996) ordered the Z_i such that $\delta_{(i)}$ is the censoring indicator δ corresponding to the i th order statistic $Z_{(i)}$, and rewrote the jumps of the Kaplan-Meier estimator of the marginal distribution of Y as

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

He used the Δ_i as weights in (4).

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 $(Z_i, 1 - \delta_i)$.

We note that 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 ,

$$[1 - \hat{F}(t)][1 - \hat{G}(t)] = 1 - \hat{H}(t) \quad (5)$$

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 (5), 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 (4) with $w_i = \Delta_i$ *case-wise* weighted estimator.

Next we define the *case-wise* empirical likelihood.

3. Main Results

In the AFT correlation model, the vectors (X_i, Y_i) are independent and identically distributed, although the observations are censored. We propose formulating the empirical likelihood *case-wise*.

Consider the following estimating equation. $\int \int (y - x\beta) x dF_{xy} = 0$. Note that for any integrable function $\phi(x, y)$, the integration $\int \phi(x, y) dF_{xy} = \int \int \phi(x, y) dF_{x|y} dF_y$ where $F_{x|y}$ denotes the conditional distribution of X given Y .

Based on the data (X_i, Z_i, δ_i) $i = 1, \dots, n$, 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 estimate of F_y , the Kaplan-Meier estimator, one obtains an estimator that is identical to Stute's \hat{F}_{xy} mentioned above.

Using this relationship, the *casewise* empirical likelihood is given by

$$L_{xy}(F_y, F_{x|y}) = \prod_{\delta_i=1} 1 \prod_{\delta_i=0} p_i \prod (1 - \sum_{Z_j \leq Z_i} p_j), \quad (6)$$

where $p_i = dF_y[Z_i]$ is the probability that F_y places on the i th *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 as follows:

$$L_{xy}(F_y) = \prod_{\delta_i=1} p_i \prod_{\delta_i=0} (1 - \sum_{Z_j \leq Z_i} p_j). \quad (7)$$

The likelihood ratio is given by

$$R_{xy}(b) = \frac{\sup\{L_{xy}(F) | F \in \tilde{\mathbb{F}}^b\}}{\sup\{L_{xy}(F) | F \in \mathfrak{F}\}}, \quad (8)$$

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

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

We comment that the constraint above can also be interpreted as

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

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

The denominator of (8) is provided when F is the Kaplan-Meier estimator and the maximization is required for the numerator only, which can be obtained using (Zhou 2005a) among others. When b is the *case-wise* weighted estimator (the solution to estimating equation (4)), then $R_{xy}(b) = 1$ and confidence regions based on (8) are ‘centered’ at this estimator.

We are now ready to state our main result. Consider the AFT correlation model equation (1). We are interested in testing the null hypothesis $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_0$. Then, we have the following theorem for the proposed *case-wise* empirical likelihood ratio statistic for least squares regression. Proof will be given in the appendix.

Theorem 1 *Consider the AFT correlation model as specified in Section 2.2. Under $H_0 : \beta = \beta_0$ and assuming regularity conditions (A1)-(A3) from the appendix as well as the assumptions for Theorem 3.1 in Zhou (1992a), we have $-2 \log R_{xy}(\beta_0) \rightarrow \chi_p^2$ in distribution as $n \rightarrow \infty$.*

The proof is given in the appendix. Similar results hold for censored quantile models when the τ th conditional quantile of Y_i is modeled by

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

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 $Y_i = X_i\beta_\tau + e_i$, but the error e_i are just independent random variables with zero τ th quantiles. When $\tau = 0.5$, this is the censored median regression and Huang *et al* (2005) 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

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

and $\psi_\tau(u)$ is the derivative of the so called check function $\rho_\tau(u) = u(\tau - I_{[u < 0]})$ of Koenker & Bassett (1978). Similarly to the censored accelerated failure time model, the denominator is maximized by the Kaplan-Meier estimator and the maximization is only needed for the numerator.

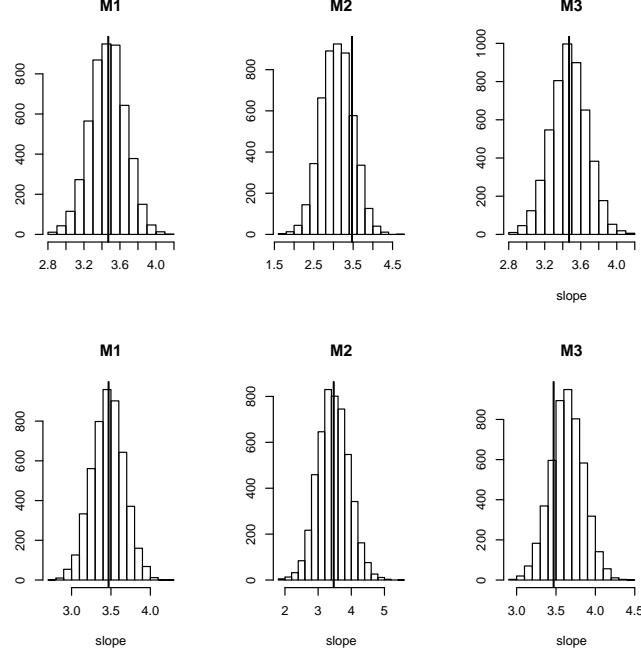
Theorem 2 *Under $H_0 : \beta = \beta_0$ and assuming regularity conditions A1, A3-A6 listed in the appendix, for given τ , $-2 \log \mathbb{R}_{xy}(\beta_\tau) \rightarrow \chi_p^2$ in distribution as $n \rightarrow \infty$.*

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

Figure 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.



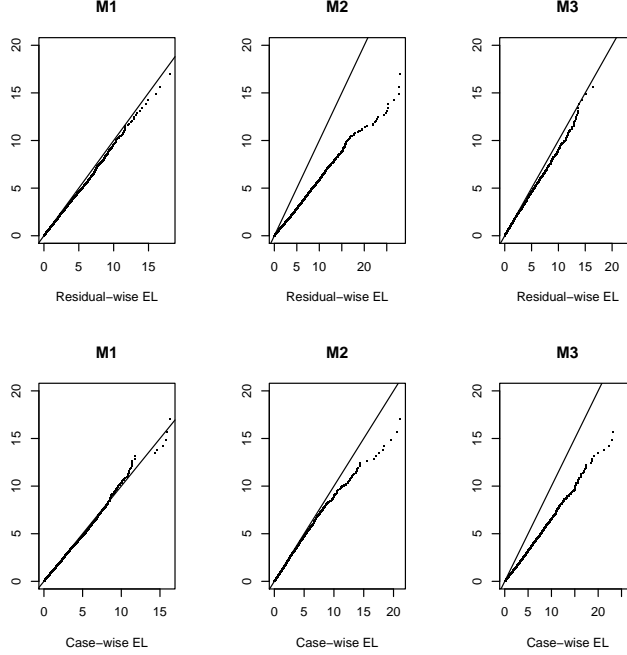
censoring (M2), and homoscedastic errors with censoring dependent on x (M3).

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

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 (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 1 shows that the Buckley-James estimator is biased for the heteroscedastic

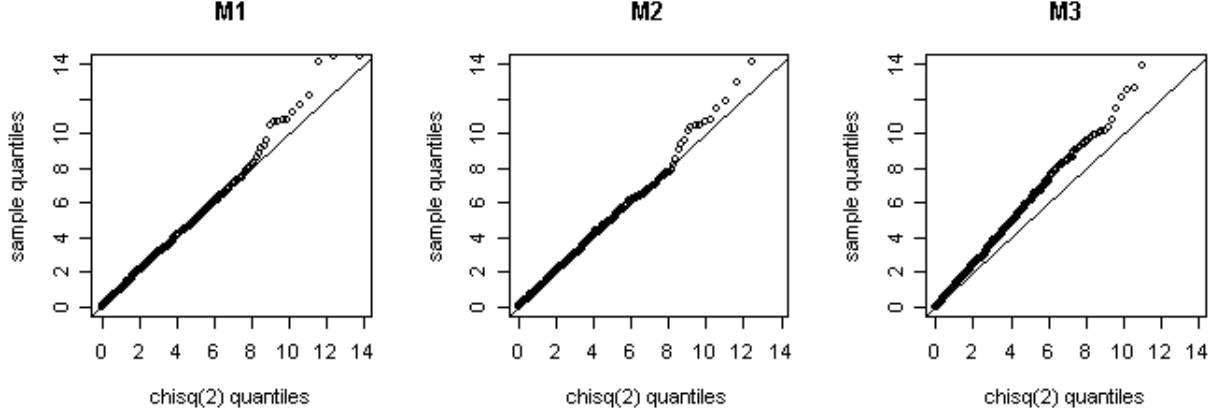
Figure 2: Q-Q plots of quantiles of χ^2_2 versus $-2 \log R_e(\beta)$ (*residual-wise*) and $-2 \log R_{xy}(\beta)$ (*case-wise*) respectively based on 5000 simulations with $n = 400$ with 28.5% censoring in each case



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 the empirical likelihood to be formulated differently. Q-Q plots in Figure 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 X_i 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.

Figure 3: Q-Q plots of the quantiles χ_2^2 versus $-2\log R_{xy}(\beta_0)$ (*casewise*) based on 1000 simulations with sample size $n = 200$ and about 30% censoring in each case.



4.2 Simulation Study 2

We examine the performance of the *casewise* empirical likelihood for censored quantile regression using the following three models that are similar to those used in Simulation Study 1.

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

where $X_i = (1, X_{1i})$ with $X_{1i} \sim U(0, 1)$. The error $e_i \sim N(0, 0.75^2)$ in M1 and M3, $e_i \sim N(0, 0.5^2)$ in M2. The parameter $\beta = (0.5, 1.5)$, and for the censoring $\epsilon_i \sim 0.5 + \exp(0.5)$ 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 $\tau = 0.25$. The quantile regression Q-Q plots in Figure 3 exhibit similar properties as the mean regression in Simulation Study 1.

5. Small-Cell Lung Cancer Data

We consider a lung cancer data set (Maksymiuk et al., 1994) that has been analyzed by Ying et al. (1995) using median regression, and by Huang et al. (2005) 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 X_{1i} the treatment indicator variable, and X_{2i} the entry age for the i th patient, where $X_{1i} = 1$ if the patient is in group B . Let Y_i be the base 10 logarithm of the i th patient's failure time. We assume the AFT model

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

Note that the estimated parameter values obtained using the approach described in this paper need to be equal to the ones from Huang et al. (2005), 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 4.

The following median regression estimates were obtained by Ying et al. (1995) and Huang et al. (2005).

$$\hat{\beta}_0 = 3.028, \hat{\beta}_1 = -0.163, \text{ and } \hat{\beta}_2 = -0.004 \quad (\text{Ying et al., 1995})$$

$$\hat{\beta}_0 = 2.693, \hat{\beta}_1 = -0.146, \text{ and } \hat{\beta}_2 = 0.001 \quad (\text{Huang et al., 2005})$$

Huang et al. (2005) did not always treat the largest Y observation as uncensored. This resulted in weights that sum to less than one. We recommend to always treat the largest Y observation as uncensored so that the weights always sum to one. The median regression estimates then become

$$\hat{\beta}_0 = 2.603, \hat{\beta}_1 = -0.263, \text{ and } \hat{\beta}_2 = 0.004 \quad (\text{with last weight}).$$

6. Concluding Remarks

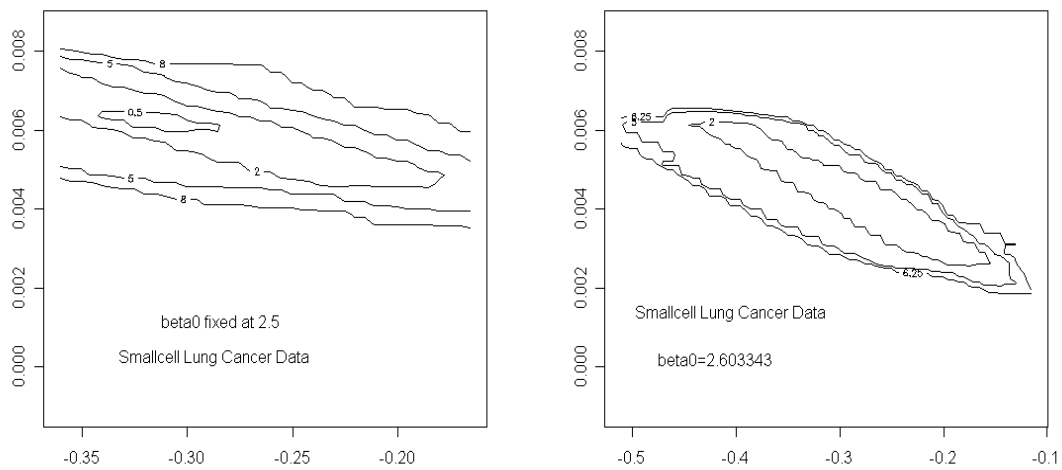
We note some differences and similarities of the *case-wise* and *residual-wise* empirical likelihood. The computation of (8) does not involve X_i with $\delta_i = 0$. Therefore, the *case-wise* empirical

likelihood allows missingness in X_i with $\delta_i = 0$, while the *residual-wise* empirical likelihood does not. As shown 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.

Nevertheless both methods provide a computationally simple inference method. The computations involved in either empirical likelihood are simple, and software is readily available in *R*. 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. This is the case for both, correlation and regression model. The resampling method proposed by Jin *et. al.* (2003) is one of few viable options but its computation is much more involved.

When censored quantile regression is concerned, Portnoy (2003) investigated the censored quantile regression process considering all τ that are estimable with the given data. He proposed estimators and a resampling based inference method under a very liberal condition only requiring

Figure 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 $\beta_0 = 2.5$ and $\beta_0 = 2.603$ are displayed. The contour lines correspond to different critical values.



linearity of the quantile regression process, accommodating both heteroscedastic errors and conditionally independent censoring. His method may not be computationally simple, though. Even in a situation when we are interested in a single quantile regression, we need to compute all the quantiles up to the one of our interest for the estimation and repeat the successive estimation in all bootstrapped resamples for the inference. In such occasions, the proposed method provides a simple alternative with slightly more restrictive conditions.

A. Assumptions

Some of these assumptions have been formulated in related work by Zhou (1992a), Stute (1993, 1996), and Huang *et al* (2005).

(A1) The (transformed) survival times Y_i and the censoring times C_i are independent. Furthermore, $P(Y_i \leq C_i | X_i, Y_i) = P(T_i \leq C_i | Y_i)$.

(A2) The survival functions $P(Y_i \geq t)$ and $P(C_i \geq t)$ are continuous and $\forall t < \infty : P(Z_i \geq t) > 0$.

(A3) The X_i are independent random variables, identically distributed according to some distribution with finite, nonzero variance, and they are independent of Y_i and of C_i .

(A4) Let $F_e(\cdot|x)$ be the conditional distribution of the residuals e_i given $X = x$, and $f_e(\cdot|x)$ the corresponding conditional density function. For a given τ , $F_e(0|x) = \tau$, and $f_e(u|x)$ is continuous in u in a neighborhood of $F_e^{-1}(\tau|x)$.

(A5) For given τ and x , let $\xi(\tau, x) = F_e^{-1}(\tau|x)$. Then, $E(XX'f_e(\xi(\tau, X)|X))$ is finite and nonsingular.

(A6) The covariate X is bounded and the right end point of the support of $X'\beta_0$ is strictly less than τ_Z . The expected value of $\|X\|^2\gamma_0^2(Y)\delta$ is finite, and for $j = 0, 1, \dots, d$, $\int |x_j|D^{1/2}(w)\tilde{F}^0(dx, dw) < \infty$, where $D(y) = \int_0^{y^-} [(1-H(w))(1-G(w))]^{-1}G(dw)$, and H and G are the distribution functions of the observations after censoring, Z_i , and the censoring times C_i , respectively.

Let $(x_1, y_1), \dots, (x_n, y_n)$ be a random sample of vectors from a joint distribution $F(x, y)$, where the y_i are subject to right censoring. Suppose t_i are the censored values of y_i : ($t_i = \min(y_i, c_i)$), along with the censoring indicator δ_i . The independent censoring variables c_i are assumed to have distribution $G(t)$.

A reasonable estimate for the marginal distribution of y is the Kaplan-Meier estimate \hat{F}

based on t_i and δ_i . A reasonable estimate of the conditional distribution $F(x|y)$ is

$$\hat{F}(x|y = t_i) = \text{point mass at } x_i, \text{ if } \delta_i = 1; \quad (9)$$

otherwise $\hat{F}(x|y)$ is undefined. This gives rise to an estimator of the joint distribution, identical to the one proposed by Stute (1993).

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 as in (9):

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

Notice that this also defines an estimator of the joint distribution $F(x, y)$ if we take the function ϕ to be the indicator function. See Stute (1993) for the law of large numbers and central limit theorem for this estimator. The ϕ function for the correlation AFT model is $\phi(x, y) = (y - x\beta)x$.

We now outline the proof of Theorem 1. To compute the supremum of the numerator in the likelihood ratio (7), we first fix a direction/path that F changes along and search the supremum on this path. The directional derivative of the path will be h . Later we shall take the supremum over h .

We define a class of functions $h(x, y)$;

$$\mathcal{H}_\phi^{F_0} = \left\{ h \mid h \text{ is left continuous, } \int h^2(x, y) dF(x, y) < \infty, \phi(x, y)h(x, y) \geq 0 \right\}.$$

Given a function h from the above class, we define a one-parameter family of distribution functions

$$\mathcal{A}_h^{F_0} = \left\{ F_\lambda(t) \mid F_\lambda(t) = \sum_{i: T_i \leq t} \Delta F_\lambda(T_i) \right\},$$

where

$$\Delta F_\lambda(t_i) = \Delta \hat{F}_{KM}(t_i) \times \frac{1}{1 + \lambda h(x_i, t_i)} \times \frac{1}{C(\lambda)}, \quad i = 1, 2, \dots, n, \quad (10)$$

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

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

This family of distributions has the same support as the Kaplan-Meier estimator and is well defined when λ is in a neighborhood of zero. For $\lambda = 0$, we obtain the Kaplan-Meier estimator:

$$F_{\lambda=0} = \hat{F}_{KM}.$$

Coupled with the conditional distribution $F_{x|y}$ defined in Section 3 above, this family of distributions can also give rise to a family of joint distributions $F_\lambda(x, y)$. Within this family of joint distributions (parameterized by λ), we seek for one that satisfies the constraint equation

$$\int \phi(x, t) dF_\lambda(x, t) = \frac{1}{C(\lambda)} \sum_{i=1}^n \Delta \hat{F}_{KM}(t_i) \frac{\phi(x_i, t_i)}{1 + \lambda h(x_i, t_i)} = 0. \quad (11)$$

There is only one distribution within the family satisfying the constraint equation (see Lemma A). 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 this family of distributions, we define a (profile) empirical likelihood ratio functions as follows:

$$\mathcal{R}_h = \left\{ \frac{L_{xy}(F_{\lambda_0})}{L_{xy}(\hat{F}_{KM})} \mid F \in \mathcal{A}_h^{F_0} \right\},$$

where L_{xy} is defined in Section 3.

Lemma A Assume all the conditions in Lemma 1. Then, as $n \rightarrow \infty$, (1) $\lambda_0 = O_p(n^{-1/2})$, (2) λ_0 has the representation given in (11) below.

PROOF: (outline) Expanding (11), we have

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

from there we have

$$\lambda_0 = \frac{\sum_{i=1}^n \phi(x_i, t_i) \Delta \hat{F}_{KM}(t_i)}{\sum_{i=1}^n \phi(x_i, t_i) h(x_i, t_i) \Delta \hat{F}_{KM}(t_i)} + o_p(n^{-1/2}). \quad (12)$$

◇

Notice that by the LLN and CLT of Stute, we have ([assume $\int \phi(x, t) dF(x, t) = 0$] see Stute 1996), as $n \rightarrow \infty$,

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

in distribution, and

$$\sum_{i=1}^n \phi(x_i, t_i) h(x_i, t_i) \Delta \hat{F}_{KM}(t_i) \rightarrow \int \phi(x, y) h(x, y) dF(x, y)$$

in probability.

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

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

Furthermore,

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

PROOF: Assume $p = 1$. Define a function of λ (the marginal empirical likelihood as a function of λ)

$$f(\lambda) = \log \prod_{i=1}^n (\Delta F_\lambda(T_i))^{\delta_i} (1 - F_\lambda(T_i))^{1-\delta_i}, \quad (13)$$

where $|\lambda| \leq |\lambda_0|$ and $F \in \mathcal{A}_h^{F_0}$. 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 $f(\lambda_0)$:

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

Substituting (10) in (13),

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

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

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

Similar calculations show that the third derivative of f evaluated at ξ is

$$f'''(\xi) = o_p(n^{2/3}). \quad (14)$$

Now

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

Rewriting the above, we obtain

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

where

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

Finally

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

Calculations show that the infimum of r_h over h is one, thus by Stute's CLT we have

$$\frac{[\sqrt{n} \sum \phi(x_i, t_i) \Delta F_{KM}(t_i)]^2}{\sigma_{KM}^2(\phi)}$$

converge to chi square in distribution as $n \rightarrow \infty$.

Acknowledgement: Research supported in part by NSF grant DMS-0604920.

References

- Buchinsky, M., and Hahn (1998), "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, **66**, **3**, 653–672.
- Buckley, J.J., and 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., and 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. and Xie, H. (2005), "Least Absolute Deviations Estimation for the Accelerated Failure Time Model," Technical Report 350, University of Iowa, Dept. of Statistics and Actuarial Science.
- Ihaka, R., and Gentleman, R. (1996), "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, **5**, **3**, 299–314.
- Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003), "Rank-Based Inference for the Accelerated Failure Time Model," *Biometrika*, **90**, **2**, 341–353.
- Jin, Z., Lin, D. Y. & Ying, Z. (2006). "On least-squares regression with censored data," *Biometrika*, **93**, **1**, 147–161.
- Kalbfleisch, J. D., and Prentice, R.L. (2002), *The Statistical Analysis of Failure Time Data, Second Edition*, New York: Wiley.
- Kaplan, E. L., and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, **53**, **282**, 457–481.
- Koul, H., Susarla, V., and Van Ryzin, J. (1981), "Regression Analysis With Randomly Right-Censored Data," *Annals of Statistics*, **9**, 1276–1288.
- Koul, H., Susarla, V, and 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.
- van der Laan, M., and Robins, J. M. (2003), *Unified Methods for Censored Longitudinal Data and Causality*, New York: Springer.
- Lai, T. L., and 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., and Wang, Qi-Hua (2003), "Empirical Likelihood Regression Analysis for Right Censored Data," *Statistica Sinica*, **13**, 51–68.
- 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., and 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.
- Murphy, S., and van der Vaart, A. (1997), "Semi-Parametric Likelihood Ratio Inference," *Annals of Statistics* **25**, 1471–1509.
- Owen, A. (1991), "Empirical Likelihood for Linear Models," *Annals of Statistics*, **19**, **4**, 1725–1747.

- Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, **98**, 1001–1012.
- Qin, G. S., and Jing, B. Y. (2001), "Empirical Likelihood for Censored Linear Regression," *Scandinavian Journal of Statistics*, **28**, 661–673.
- Rotnitzky A, Robins J. M. (2005), "Inverse Probability Weighted Estimation in Survival Analysis," in: *The Encyclopedia of Biostatistics, 2nd edition*, Ed. P. Armitage and T. Coulton, New York: Wiley.
- Stute, W. (1993), "Consistent Estimation Under Random Censorship When Covariables are Present," *Journal of Multivariate Analysis*, **45**, 89–103.
- Stute, W. (1996), "Distributional Convergence Under Random Censorship When Covariables are Present," *Scandinavian Journal of Statistics* **23**, 461–471.
- Stute, W. (1999), "Nonlinear Censored Regression," *Statistica Sinica*, **9**, 4, 1089–1102.
- Wang, Q.H., and Li, G. (2002), "Empirical Likelihood Semiparametric Regression Analysis Under Random Censorship," *Journal of Multivariate Analysis*, **83**, 469–486.
- Wei, L. J. (1992), "The Accelerated Failure Time Model: A Useful Alternative to the Cox Regression Model in Survival Analysis," *Statistics in Medicine*, **11**, 1871–1879.
- Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Hazard and Survival Functions," *Journal of the American Statistical Association*, **94**, 137–145.
- Ying, Z., Jung, S. H., and Wei, L. J. (1995), "Survival Analysis with Median Regression Models," *Journal of the American Statistical Association* **90**, **429**, 178–184.
- Zheng, Z. (1984), "Regression with Randomly Censored Data," unpublished Ph.D. dissertation, Columbia University, Dept. of Statistics.
- Zhou, M. (1992a), "Asymptotic Normality of the 'Synthetic Data' Regression Estimator for Censored Survival Data," *Annals of Statistics*, **20**, **2**, 1002–1021.
- Zhou, M. (1992b), "M-Estimation in Censored Linear Models," *Biometrika*, **79**, **4**, 837–841.
- Zhou, M. (2005a), "Empirical Likelihood Ratio With Arbitrarily Censored/Truncated Data by a Modified EM algorithm," *Journal of Computational and Graphical Statistics*, **14**, 643–656.
- Zhou, M. (2005b), "Empirical Likelihood Analysis of the Rank Estimator for the Censored Accelerated Failure Time Model," *Biometrika*, **92**, 492–498.
- Zhou, M., and Li. G. (2008), "Empirical Likelihood Analysis of the Buckley-James Estimator," *J. Multivariate Analysis*, **99**, 649–664.