2011

Adjusted Empirical Likelihood Models with Estimating Equations for Accelerated Life Tests

Ni Wang

Jye-Chyi Lu

Di Chen

Paul H. Kvam
University of Richmond, pkvam@richmond.edu

Adjusted Empirical Likelihood Models with Estimating Equations for Accelerated Life Tests

Ni Wang[+], Jye-Chi Lu[+], Di Chen[*], Paul Kvam[+]

*: Biostatistics, UCB Pharma, Inc. Symrna, GA 30080

+: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205

SUMMARY

This article proposes an Adjusted Empirical Likelihood Estimation (AMELE) method to model and analyze accelerated life testing data. This approach flexibly and rigorously incorporates distribution assumptions and regression structures by estimating equations within a semi-parametric estimation framework. An efficient method is provided to compute the empirical likelihood estimates, and asymptotic properties are studied. Real-life examples and numerical studies demostrate the advantage of the proposed methodology.

KEY WORDS: Asymptotics; Maximum Likelihood Estimation; Percentile Regression; Random Censoring; Reliability.

1 Introduction

In evaluating the reliability of durable products, accelerated life testing (ALT) is commonly applied by stressing specimens at harsher conditions than in normal use, thereby hastening failure time in tests with short duration. Regression models of replicated data at several stress levels are built to provide extrapolated estimates of distribution properties (e.g., 5th or 10th percentile, mean, variance and lifetime distribution function) in the normal-use condition for warranty management, product improvement and risk analysis. For newer products where the physics supporting regression models is not clearly understood for extrapolation, the stress levels are usually set closer to the normal-use condition. Due to high durability of products and limited testing time, this practice results in heavily censored data. For example, in Meeker and LuValle (1995), tests of printed circuit boards revealed that 68.5% of the data in the lowest stress level are censored after 4,078 hours (169.9 days) of testing. This creates challenges in deriving statistical inference procedures for lifetime quantities.

Various parametric approaches have been introduced to solve this inference problem. Typical parametric approaches assume that failure time distributions under various stress levels belong to the same parametric family and there is a (transformed) linear regression structure of the location parameters of these distributions. Most ALT procedures assume a constant variance. There are some exceptions, such as Meeter and Meeker (1994), where it is assumed that the logarithm for each of the scale parameters has a linear regression relationship. In some cases, the traditional ALT models cannot accurately represent the failure time data; the commonly used acceleration function for regression might not be suitable. For example, Meeker and LuValle (1995) used chemical-kinetic knowledge to derive an intricate failure time model which does not fit into the ALT model structure. Because the traditional regression-over-the-mean approach is questionable (especially in the case that the means might not exist), Meeker and LuValle constructed log-linear regression models based on two key chemical-reaction parameters found in differential equations that characterize the failure evolution processes. Although this physics-based approach provides a well-justified ALT model, explicit physical relations are rarely available to aid the data modeling so directly. Thus, there is a need for developing a data exploration approach to entertain potential regression models and to examine the goodness-of-fit of the assumed lifetime distribution.

Other than parametric approaches, the semi-parametric accelerated failure time (AFT) (Kalbfleisch and Prentice, 2002) model regresses the logarithm of the survival time on the stress levels, which is an attractive alternative due to its intuitive physical interpretation. Many approaches have been proposed to estimate the regression parameters, including the nonparametric estimator of Shaked, Zimmer, and Ball (1979), the Buckley-James estimator (Buckley and James, 1979), rank-based estimators (Kalbfleisch and Prentice, 2002), and so on. Lai and Ying (1991) provided theoretical justification and asymptotic properties of rank estimators, which were solved from rank estimating functions. Recently, semiparametric transformation models have been generalized, including the PH model, the AFT model and proportional odds model as special cases (Cheng, Wei, and Ying (1997), Murphy, Rossini and Vaart (1997)), where the parameters are estimated using generalized estimating equations and likelihood based methods. There are also semiparametric inference procedures proposed for median regression models. Ying, Jun and Wei (1995) proposed a method to estimate the parameters by solving from a set of estimating equations, which has similar feature as the least absolute deviation estimator. Yang (1999) approached this problem by specifying estimating equations based on a weighted empirical hazard.

This article considers a semiparametric approach with parameter inference based on empirical likelihood. This has two advantages: first, failure times at different stress levels are not required to have the same underlying distribution; second, the confidence regions are automatically determined using likelihood ratio based methods without estimating the variance of test statistics, which can be difficult in the case of the rank-based regression estimators in censored ALT models.

Empirical likelihood (EL) was developed by Owen (1990) as a general nonparametric inference procedure which combineed the reliability of nonparametric methods with the effectiveness of likelihood methods. It has been extended to more difficult inference problems involving censored or truncated data Owen (2001). (Owen, 2001). Pan and Zhou (2000) studied the EL procedure when the parameter could be written as a function of cumulative hazard functions, with additional constraints that the hazard function are dominated by the Nelson-Aalen estimator. In Li and Wang (2003), the authors considered the EL approach for right censored data, and proposed a new synthetic variable incorporating the failure time and censoring information, then proceeded with model inference using standard EL methods. Chen, Lu and Lin (2005) considered the case of group censored data, where failure time and censored data are observed at pre-specified time intervals. Estimating quations are introduced into empirical likelihood in Qin and Lawless (1994), which demonstrated that estimating equations (EE) can be useful in incorporating distribution knowledge to improve estimation quality. Zhou (2005b) studied the EL inference of rank estimators by using the rank estimating equations in the constraints. In particular, Lu, Chen and Gan (2002) showed that the EL-EE approach is a natural extension of both Generalized Estimating Equations (GEE; Liang and

Zeger, 1986) and Quasi-Likelihood Estimation (QLE) approaches (Wedderburn, 1974) by allowing censored data. The computational issues of EL is also complicated due to the censoring in ALT. Zhou (2005a) proposed an iterative EM algorithm to impute weights of censored data by the survival probability. Here, we also address this issue by proposing an approximate solution which is shown to be much faster.

This paper proposes an Adjusted Empirical Likelihood Estimation (AMELE) method which is easy to implement compared to existing empirical likelihood methods for censored data, and it further studies the asymptotic properties of the parameter estimates as well as the survival functions. Section 2 defines the empirical likelihood with estimating equations for censored data, and proposes AMELE method. Section 3 shows the asymptotic properties for the proposed estimators. Real-life examples and simulation studies are presented in Section 4 and 5 to illustrate and compare the proposed methods with some existing methods. Section 6 provides the conclusion and future work.

2 The Adjusted Empirical Likelihood Estimation Methods

2.1 Empirical Likelihood with Estimating Equations

Let T_j and C_j be the failure time and censoring random variables at stress level j, where $j = 1, \ldots, m$. Let x_j be a $p \times 1$ vector of covariates under stress level j. We assume that C_j and T_j are independent. Denote survival function and the distribution function for the failure time T_j and the censoring time C_j as $S_{T_j}(t)$, $F_{T_j}(t)$, $S_{C_j}(t)$ and $F_{C_j}(t)$.

Many AFT models assume the failure time (or transformed) T_j and x_j are related through regression functions $\mathrm{E}(T_j) = \boldsymbol{\theta}^\top x_j$, where $\boldsymbol{\theta}$ is a $p \times 1$ vector including the intercept term. We generalize the regression functions to $r-$dimensional functionally independent estimating equations $G(T_j, \boldsymbol{\theta}, x_j)$, abbreviated as $G_j(T, \boldsymbol{\theta})$, which satisfies $\mathrm{E}G_j(T, \boldsymbol{\theta}) = 0$. In the following section, we set up an estimation framework using empirical likelihood to solve for $\boldsymbol{\theta}$.

Suppose that there are n_j replicates at stress level j, and k_j distinct failure time $t_{1,j} < t_{2,j} < \cdots < t_{k_j,j} < t_{k_j+1,j} = L_j$, where L_j is an upper level of failure and censoring time. Let c_{ij} be the number of censored data in the interval $(t_{i-1,j}, t_{ij}]$, and P_{ij} be the probability point mass on each observed failure time, we can write the (empirical) likelihood function as follows:

$$L = \prod_{j=1}^{m} \left\{ \prod_{i=1}^{k_j+1} P_{ij} \prod_{i=1}^{k_j+1} \left(\sum_{l=i}^{k_j+1} P_{lj} \right)^{c_{ij}} \right\}, \tag{1}$$

where $P_{ij} = \Pr(t_{i-1,j} < T_j \le t_{ij}) = S_{T_j}(t_{i-1,j}) - S_{T_j}(t_{ij})$.

Under the constraints of the m sets of estimating equations

$$\mathrm{E}[G_j(T_j, \boldsymbol{\theta})] = \sum_{i=1}^{k_j+1} P_{ij} G_j(t_{ij}, \boldsymbol{\theta}) = 0, \tag{2}$$

the optimal parameter estimates maximize the empirical likelihood (1).

2.2 Adjusted Maximum Empirical Likelihood Estimator

For notational simplicity, we drop the subscript j in deriving the parameter estimates since it is straightforward to extend this to multiple levels. We rewrite the original formulation as

$$\begin{aligned} \text{maximize :} \qquad & L = \prod_{i=1}^{k+1} P_i \prod_{i=1}^{k+1} (\sum_{l=i}^{k+1} P_l)^{c_i}, \qquad (3) \\ \text{subject to :} \qquad & \sum_{i=1}^{k+1} P_i = 1, \ \sum_{i=1}^{k+1} P_i G(T_i, \boldsymbol{\theta}) = 0, \quad P_i \in [0,1]. \end{aligned}$$

Using the standard Lagrange multiplier arguments in Owen (1990), Qin and Lawless (1994) and Owen (2001), we have the following implicit intermediate results:

$$\begin{aligned} \hat{P}_i(\boldsymbol{\lambda}) &= \frac{1}{n(1 - a_i(\boldsymbol{\lambda}) + \boldsymbol{\lambda}^\top G(t_i, \boldsymbol{\theta}))}, \qquad (4) \\ a_i(\boldsymbol{\lambda}) &= \frac{1}{n} \sum_{m=1}^{i} \frac{c_m}{\sum_{l=m}^{k+1} \hat{P}_m(\boldsymbol{\lambda})}, \\ 0 &= \sum_{i=1}^{k+1} \frac{G(t_i, \boldsymbol{\theta})}{1 - a_i(\boldsymbol{\lambda}) + \boldsymbol{\lambda}^\top G(t_i, \boldsymbol{\theta})}. \end{aligned}$$

It is well known that when there are no constraints, the optimal $\hat{P}_i$ maximizing the nonparametric likelihood is the Kaplan-Meier estimator (Kaplan and Meier, 1958). We summarize this in the following lemma:

Lemma 1 *When $\boldsymbol{\lambda}$=0, $P_i(0) = [n(1 - a_i(0))]^{-1}$ is the Kaplan-Meier estimator $d\hat{F}_{T,KM}(T_i)$, and $1 - a_i(0)$ is equal to the Kaplan-Meier estimator of the censoring variable $\hat{S}_{C,KM}(t_i)$*

For a given $\boldsymbol{\theta}$, we can find corresponding optimal $\boldsymbol{\lambda}$ and $\hat{P}_i(\boldsymbol{\lambda})$ to maximize the nonparametric likelihood defined in (3). Plugging $\hat{P}_j(\boldsymbol{\lambda})$ back into the likelihood function, we have a profile likelihood of $\boldsymbol{\theta}$ as $L(\boldsymbol{\theta})$. The maximum empirical likelihood estimator of $\boldsymbol{\theta}$ can then be solved by maximizing

$\boldsymbol{\theta}$ over the parameter space. However, the computations for solving optimal values directly is quite complicated since there is no explicit solution available for $\hat{P}_i(\boldsymbol{\lambda})$ from the intermediate results in (4). Zhou (2005a) proposed an EM type algorithm to solve for the mean of failure time under censoring, and studied its asymptotic properties. In order to estimate parameters in more general settings, we propose a simplified computational procedure in this paper, using $a_i(0)$ to approximate $a_i(\boldsymbol{\lambda})$. Therefore, the previous intermediate results in (4) become:

$$\begin{aligned} \tilde{P}_i(\boldsymbol{\lambda}) &= \frac{1}{n(1 - a_i(0) + \boldsymbol{\lambda}^\top G(t_i, \boldsymbol{\theta}))}, \\ a_i(0) &= \frac{1}{n} \sum_{v=1}^{i} \frac{c_v}{\sum_{l=v}^{k+1} \tilde{P}_l(0)}, \\ 0 &= \sum_{i=1}^{k+1} \frac{G(t_i, \boldsymbol{\theta})}{1 - a_i(0) + \boldsymbol{\lambda}^\top G(t_i, \boldsymbol{\theta})}. \end{aligned} \tag{5}$$

Starting from the Kaplan-Meier estimator, we can obtain the estimator of the Lagrangian multiplier $\tilde{\boldsymbol{\lambda}}$, and compute the $\tilde{P}_i(\boldsymbol{\lambda})$ from (5). Plugging the new set of intermediate results into likelihood (3), we can obtain a new estimator $\tilde{\boldsymbol{\theta}}$ by maximizing the profile likelihood. We call this new estimator the *Adjusted Maximum Empirical Likelihood Estimator* (AMELE).

The approximation of $a_i(\boldsymbol{\lambda})$ by $a_i(0)$ is one critical step in simplifying the derivation of asymptotic properties. As shown in Lemma 3, $\boldsymbol{\lambda}$ is tightly bounded when $\boldsymbol{\theta}$ is close to the true value $\boldsymbol{\theta}_0$. Lemma 3 also justifies this approximation by showing the new estimator will be a consistent estimator of true value $\boldsymbol{\theta}_0$.

Remark 1. In the case of no censoring ($a_i = 0$), the equations in (4) reduce to

$$P_i = \frac{1}{n(1 + \boldsymbol{\lambda}^\top G(t_i, \boldsymbol{\theta}))}, 1 \le i \le n, \quad \text{and} \quad \sum_{i=1}^{n} \frac{G(t_i, \boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^\top G(t_i, \boldsymbol{\theta})} = 0,$$

as shown in Qin and Lawless (1994).

Three general assumptions are needed for fully characterizing the properties of the AMELE $\tilde{\boldsymbol{\theta}}$ along with $\tilde{S}_T(t)$.

(**A.1**) The parameter space $\boldsymbol{\Theta} \subset I\!R^p$ is compact, contains a neighborhood of the true parameter $\boldsymbol{\theta}_0$, and $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \{0 < |L(\boldsymbol{\theta})|\} < \infty$.

(**A.2**) Given $\boldsymbol{t} = (t_1, t_2, \ldots, t_k, t_{k+1})$, let $G(\boldsymbol{t}, \boldsymbol{\theta}) = (G(t_i, \boldsymbol{\theta}))_{(k+1)\times r}$. For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, assume that the $r \times r$ matrix $G^\top G$ is nonsingular.

(**A.3**) $\mathrm{E}(||G(T,\boldsymbol{\theta})||^3) < \infty$ and $G(T,\boldsymbol{\theta})$ is second-order differentiable with respect to $\boldsymbol{\theta}$, i.e., $\partial^2 G/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ exists for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The assumption **A.1** is needed to ensure that the maximum of $|L(\boldsymbol{\theta})|$ exists in the interior of the parameter space. Assumptions **A.2** and **A.3** require the non-singularity, continuity and differentiability of the estimating function $G(t,\boldsymbol{\theta})$ to ensure that equation (5) is well defined and the AMELE $\tilde{\boldsymbol{\theta}}$ is in $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < n^{-1/2}$ with probability one, given that n is sufficiently large.

For notational simplicity, define

$$Z_i(\boldsymbol{\theta}) = \frac{G(t_i,\boldsymbol{\theta})}{1 - a_i(0)} = nG(t_i,\boldsymbol{\theta})d\hat{F}_{T,KM}(t_i), \tag{6}$$

where the equivalence of $[n(1 - a_i(0))]^{-1}$ and $d\hat{F}_{T,KM}(t_i)$ is shown in Lemma 1. Then, (5) can be rewritten as

$$\tilde{P}_i(\boldsymbol{\lambda}) \quad = \quad \frac{1}{1 + \boldsymbol{\lambda}^\top Z_j(\boldsymbol{\theta})} d\hat{F}_{T,KM}(t_i), \tag{7}$$

$$\sum_{i=1}^{k+1} \frac{Z_i(\boldsymbol{\theta})}{1 + \boldsymbol{\lambda}^\top Z_i(\boldsymbol{\theta})} = 0. \tag{8}$$

Let $\hat{F}_{T,KM}(t)$ be the Kaplan-Meier estimator of the distribution function and let $\hat{S}_{C,KM}(t)$ be the Kaplan-Meier survival function estimator for the censoring time. We have the follow lemma:

Lemma 2 *Under the assumptions* **A.1-A.3**, *as* $n \to \infty$,

$$(a)\ \frac{1}{n}\sum_{i=1}^{k+1} Z_i(\boldsymbol{\theta}) = \int G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t) = O_p(n^{-1/2}),$$
$$(b)\ \frac{1}{n}\sum_{i=1}^{k+1} Z_i(\boldsymbol{\theta})Z_i(\boldsymbol{\theta})^\top = \boldsymbol{A}(\boldsymbol{\theta}) + o_p(1),$$

uniformly in the ball $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \le n^{-1/2}$, *where*

$$\boldsymbol{A}(\boldsymbol{\theta}) = \int \frac{G(t,\boldsymbol{\theta})G(t,\boldsymbol{\theta})^\top}{\hat{S}_{C,KM}(t)} d\hat{F}_{T,KM}(t). \tag{9}$$

The vector $\boldsymbol{\lambda}$ and Z_i are related through equation (8). For a given $\boldsymbol{\theta}$, a unique $\boldsymbol{\lambda}$ exists, provided that 0 is inside the convex hull of the points $Z_i(\boldsymbol{\theta})$. The following lemma quantifies the magnitude of the $\boldsymbol{\lambda}$ in a small neighborhood of $\boldsymbol{\theta}$.

Lemma 3 *Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ be the true value of the parameter. Under the assumptions* **A.1-A.3***, we have the following results: For $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}\}$ and $\boldsymbol{\lambda}(\boldsymbol{\theta})$ satisfying (8), we have $\boldsymbol{\lambda}(\boldsymbol{\theta}) \overset{w.p.1}{\longrightarrow} 0$, and $\boldsymbol{\lambda}(\boldsymbol{\theta}) = O_p(n^{-1/2})$ uniformly, as $n \to \infty$.*

Given $\boldsymbol{\lambda}(\boldsymbol{\theta})$ and $\tilde{P}_i(\boldsymbol{\lambda})$ through (7), the adjusted log-likelihood is

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{k} \log \tilde{P}_i(\boldsymbol{\theta}) + \sum_{i=1}^{k+1} c_i \log \left(\sum_{l=i}^{k+1} \tilde{P}_l(\boldsymbol{\theta}) \right). \tag{10}$$

Following the same argument in Chen, Lu, and Lin (2003), the score function $\boldsymbol{l}(\boldsymbol{\theta})$ of the adjusted log-likelihood equation can be simplified to

$$\boldsymbol{l}(\boldsymbol{\theta}) = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\boldsymbol{\lambda}^\top(\boldsymbol{\theta}) \sum_{i=1}^{k} \tilde{P}_i(\boldsymbol{\theta}) \frac{\partial G(t_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{11}$$

The solution to the function $\boldsymbol{l}(\boldsymbol{\theta}) = 0$ is the AMELE $\tilde{\boldsymbol{\theta}}$, and the corresponding AMELE for the survival function $S_T(t)$ is then

$$\tilde{S}_T(t) = \sum_{t_i > t} \frac{1}{n\left(1 - a_i(0) + \boldsymbol{\lambda}^\top G(t_i, \tilde{\boldsymbol{\theta}})\right)}. \tag{12}$$

The following lemma justifies the AMELE $\tilde{\boldsymbol{\theta}}$ as a consistent estimator of $\boldsymbol{\theta}_0$, which can be proven using the same arguments in Qin and Lawless (1994). So, we simply state the proposition here without detailed proof.

Lemma 4 *Under the regularity conditions, as $n \to \infty$, likelihood function $L(\boldsymbol{\theta})$ attains its maximum value at some point $\tilde{\boldsymbol{\theta}}$ in the interior of the ball $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}$ with probability one. Thus, the AMELE $\tilde{\boldsymbol{\theta}}$ is a strongly consistent estimate of $\boldsymbol{\theta}_0$.*

3 Asymptotic Properties of the AMELE

To understand of the asymptotic properties of AMELE, we start by investigating the large sample properties of $\boldsymbol{\lambda}(\tilde{\boldsymbol{\theta}})$.

3.1 Asymptotic Distribution of $\boldsymbol{\lambda}$

For $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}\}$, $\boldsymbol{\lambda} = O_p(n^{-1/2})$ according to Lemma 3. After a Taylor expansion of (8) at $\boldsymbol{\lambda} = 0$, we have

$$\begin{aligned} \boldsymbol{\lambda} &= -\left[\frac{1}{n}\sum_{i=1}^{k+1} Z_i(\boldsymbol{\theta}) Z_i^\top(\boldsymbol{\theta})\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{k+1} Z_i(\boldsymbol{\theta})\right] + o_p(n^{-1/2}) \\ &= -\boldsymbol{A}(\boldsymbol{\theta})^{-1} \int G(t,\boldsymbol{\theta}) d\hat{F}_{T,KM}(t) + o_p(n^{-1/2}). \end{aligned} \tag{13}$$

The following Theorem states the asymptotic normality of $\boldsymbol{\lambda}$.

Theorem 1 *For continuous lifetime T and censoring time C, suppose $S_C(L) > 0$, and $S_T(t)$ is continuous at $t = L$. Then, as $n \to \infty$, if $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}\}$,*

$$\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\theta}) \xrightarrow{d} N_r(0, \boldsymbol{\Sigma}_\lambda(\boldsymbol{\theta})),$$

where

$$\boldsymbol{\Sigma}_\lambda(\boldsymbol{\theta}) = \boldsymbol{A}(\boldsymbol{\theta})^{-1} \boldsymbol{\Sigma}_G(\boldsymbol{\theta}) \boldsymbol{A}(\boldsymbol{\theta})^{-1}, \tag{14}$$

$$\boldsymbol{\Sigma}_G(\boldsymbol{\theta}) = \int_0^\infty \left\{\int_x^\infty (G(x,\boldsymbol{\theta}) - G(t,\boldsymbol{\theta})) dS_T(t)\right\} \left\{\int_x^\infty (G(x,\boldsymbol{\theta}) - G(t,\boldsymbol{\theta})) dS_T(t)\right\}^\top \frac{dF_T(t)}{S_T^2(t) S_C(t)}. \tag{15}$$

Because $\hat{F}_{T,KM}(t)$ is uniformly consistent and $\boldsymbol{\lambda} \xrightarrow{w.p.1} 0$, it is easy to see that

$$\sum_{i=1}^{k+1} \tilde{P}_i(\boldsymbol{\theta}) \frac{\partial G(t_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \int \frac{\partial G_{jh}(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{1 + \boldsymbol{\lambda} Z_i(\boldsymbol{\theta})} d\hat{F}_{T,KM}(t) \tag{16}$$

$$\to \int \frac{\partial G(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\hat{F}_{T,KM}(t) \tag{17}$$

$$\to \int \frac{\partial G(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dF_T(t) = \mathrm{E}\frac{\partial G(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{18}$$

as $n \to \infty$, for $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}\}$. The asymptotic normality of $\boldsymbol{l}(\boldsymbol{\theta})$ follows directly from Theorem 1. We state it as the following corollary.

Corollary 1 *Under the conditions of Theorem 1, for given $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}\}$, then $\sqrt{n}\boldsymbol{l}_n(\boldsymbol{\theta})$ is asymptotically normal with mean zero, and covariance matrix*

$$\boldsymbol{\Sigma}_l(\boldsymbol{\theta}) = E\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right) \boldsymbol{\Sigma}_\lambda(\boldsymbol{\theta}) E\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right)^\top, \tag{19}$$

where $\boldsymbol{\Sigma}_\lambda(\boldsymbol{\theta})$ is given by (14), and

$$\mathrm{E}\left(\frac{\partial G(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = \mathrm{E}\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right).$$

3.2 Asymptotic Normality of the AMELE of Model Parameters

The partial derivative of $\boldsymbol{l}(\boldsymbol{\theta})$ with regard to $\boldsymbol{\theta}$ can be expressed as

$$-\frac{\partial \boldsymbol{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \boldsymbol{\lambda}^\top(\boldsymbol{\theta})\frac{\partial}{\partial \boldsymbol{\theta}}\left(\sum_{i=1}^{k+1}\tilde{P}_i(\boldsymbol{\theta})\frac{\partial G(t_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) + \frac{\partial \boldsymbol{\lambda}_j^\top(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left(\sum_{i=1}^{k+1}\tilde{P}_i(\boldsymbol{\theta})\frac{\partial G(t_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right).$$

As $n \to \infty$, the first term of right side in the above equation goes to zero in probability (since $\boldsymbol{\lambda}(\boldsymbol{\theta}) \xrightarrow{w.p.1} 0$), and the convergence of the second part is shown in (16). Thus,

$$\lim_{n\to\infty} -\frac{\partial \boldsymbol{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \lim_{n\to\infty} \mathrm{E}\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right)^\top \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\theta}}.$$

It follows from (13) that

$$\begin{aligned}\lim_{n\to\infty} -\frac{\partial \boldsymbol{\lambda}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial \boldsymbol{A}(\boldsymbol{\theta})^{-1}}{\partial \boldsymbol{\theta}}\mathrm{E}(G(t,\boldsymbol{\theta})) + \boldsymbol{A}(\boldsymbol{\theta})^{-1}\mathrm{E}\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right)\\ &= \boldsymbol{A}(\boldsymbol{\theta})^{-1}\mathrm{E}\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right).\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} -\frac{\partial \boldsymbol{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathrm{E}\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right)^\top \boldsymbol{A}(\boldsymbol{\theta})^{-1}\mathrm{E}\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right).$$

Applying Taylor's expansion to $\boldsymbol{l}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$, we have

$$0 = \boldsymbol{l}(\tilde{\boldsymbol{\theta}}) = \boldsymbol{l}(\boldsymbol{\theta}_0) + \frac{\partial \boldsymbol{l}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p(\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|), \tag{20}$$

which leads to the following theorem.

Theorem 2 *Under the assumptions* **A.1 - A.3**, *as* $n \to \infty$, $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N_p(0, \boldsymbol{\Sigma}_{\theta_0})$, *where*

$$\begin{aligned}\boldsymbol{\Sigma}_\theta &= \boldsymbol{B}(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}_l(\boldsymbol{\theta})\boldsymbol{B}(\boldsymbol{\theta})^{-1},\\ \boldsymbol{B}(\boldsymbol{\theta}) &= E\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right)^\top \boldsymbol{A}(\boldsymbol{\theta})^{-1}E\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right),\end{aligned}$$

and $\boldsymbol{A}(\boldsymbol{\theta})$ *and* $\boldsymbol{\Sigma}_l(\boldsymbol{\theta})$ *are given by (9) and (19).*

Remark 2. Here, we compare our asymptotic results against some well-known benchmark results in the literature. When θ is the population mean and the estimating function is $G(t,\theta) = t - \theta$, $\partial G/\partial\theta = 1$, $\boldsymbol{B}(\theta) = \boldsymbol{A}(\theta)^{-1}$ and $\boldsymbol{\Sigma}_l(\theta) = \boldsymbol{\Sigma}_\lambda(\theta)$, $\boldsymbol{\Sigma}_\theta(\theta)$ reduces to

$$\begin{aligned}\boldsymbol{\Sigma}_G(\theta) &= \int_0^\infty \left(\int_x^\infty (x-t)dS_T(t)\right)^2 \frac{dF_T(x)}{S_T^2(x)S_C(x)}\\ &= \int_0^\infty \left(\int_x^\infty (1-F_T(t))dt\right)^2 \frac{dF_T(x)}{S_T^2(x)S_C(x)},\end{aligned}$$

which is the same result as obtained by Breslow and Crowley (1974).

In the complete-sample case where $S_C(t) = 1$ and $G(\theta)$ is differentiable, it is easy to verify that the asymptotic covariance matrix reduces to

$$\boldsymbol{\Sigma}_\theta = \left[\mathrm{E}\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right) \mathrm{E}[GG^\top]^{-1} \mathrm{E}\left(\frac{\partial G}{\partial \boldsymbol{\theta}}\right)^\top \right]^{-1},$$

which is the same result as obtained by Qin and Lawless (1994).

The asymptotic properties for the AMELE of the survival function $S_T(t)$ can be proved in the similar manner with the detailed proof in the Appendix.

Theorem 3 *Under the conditions of Theorem 1, as $n \to \infty$, $\sqrt{n}(\tilde{S}_{T,}(t) - S_T(t)) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{S(t)})$, where*

$$\begin{aligned} \boldsymbol{\Sigma}_{S(t)} &= S_T^2(t) \int_0^t \frac{dF_T(x)}{S_T(x)^2 S_C(x)} + \\ & \int_0^\infty \left[\int_x^\infty \boldsymbol{\gamma}^\top(t, \boldsymbol{\theta})(G(x, \boldsymbol{\theta}) - G(t, \boldsymbol{\theta})) dS_T(t) \right]^2 \frac{dF_T(x)}{S_T^2(x) S_C(x)} + \\ & 2S_T(t) \int_0^t \left(\int_s^\infty \boldsymbol{\gamma}^\top(t, \boldsymbol{\theta})(G(s, \boldsymbol{\theta}) - G(x, \boldsymbol{\theta}) dS_T(x) \right) \frac{dF_T(s)}{S_C(s)}, \textit{and} \end{aligned}$$

$$\boldsymbol{\gamma}(t, \boldsymbol{\theta}) = A(\boldsymbol{\theta})^{-1} \int_t^\infty \frac{G(x, \boldsymbol{\theta}) dF_T(x)}{S_C(x)}.$$

4 Examples

In the following example, a set of real-life data from accelerated life tests is analyzed, and we compare the results of our AMELE estimates with results from a more traditional Weibull regression. Meeker and LuValle (1995) reported on this experiment for testing printed-circuit-boards (PCB) at four high relative humidity (RH) conditions: 49.5% RH, 62.8% RH, 75.4% RH and 82.4% RH. The normal-use condition in this case has RH at 10%. Figure 1 shows the Weibull probability plot of the data from three higher stress levels. The curvature in the plot indicates that the Weibull lifetime distribution does not adequately fit these data. Note that there are only 22 (out of 70) failures in the lowest stress level with 68.6% of data censored after 169.9 days of testing.

Since a lower percentile (such as 5%) of lifetime is observed for all stress levels and lower percentiles are important in reliability applications, we explore the regression structures based on them. Figure 2 shows that a linear relationship between the logarithm of failure times and the logit transformation of RH is plausible.

Because 25th percentiles are also available for all stress levels, we explore the trend of the difference of the logarithm of the 25th and 5th percentiles over changing stress levels. Figure 2 shows that this percentile-difference is not constant, but rather a linear function with much larger difference in the normal-use condition. For estimating the lifetime distribution, one approach is to assume that after a proper "re-scaling" of the data using the percentile and percentile-difference, the lifetime distributions at all stress levels would be approximately the same. Then, the AMELE gives the estimate and its point-wise confidence intervals.

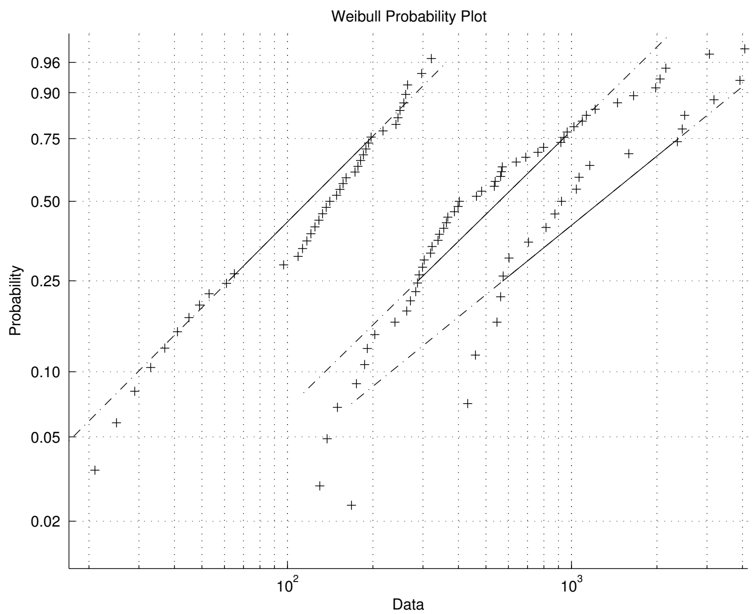

Figure 1: Weibull probability plot for the failure time data under different stress levels with RH = 49.5, 62.8, 75.4% (from right to left)

Remark 3. In previous sections, we consider a generic smooth function G in order to have desirable properties. In situations when many observations are being censored, especially at lower stress levels, estimating equations can be constructed on lower percentiles. Consider the qth percentile of T as $\boldsymbol{\theta}^\top x$, and then specify the structural relationship in estimating functions:

$$G(T, \boldsymbol{\theta}) = I(T < \boldsymbol{\theta}^\top x) - q, \tag{21}$$

where q is the percentile of the lifetime, and $I(\cdot)$ denotes the indicator function.

In order to smooth the non-differentialable constraint fuctions G, we introduce a kernel function to smooth the estimating equations. Similar to the techniques proposed in Chen and Hall (1993),

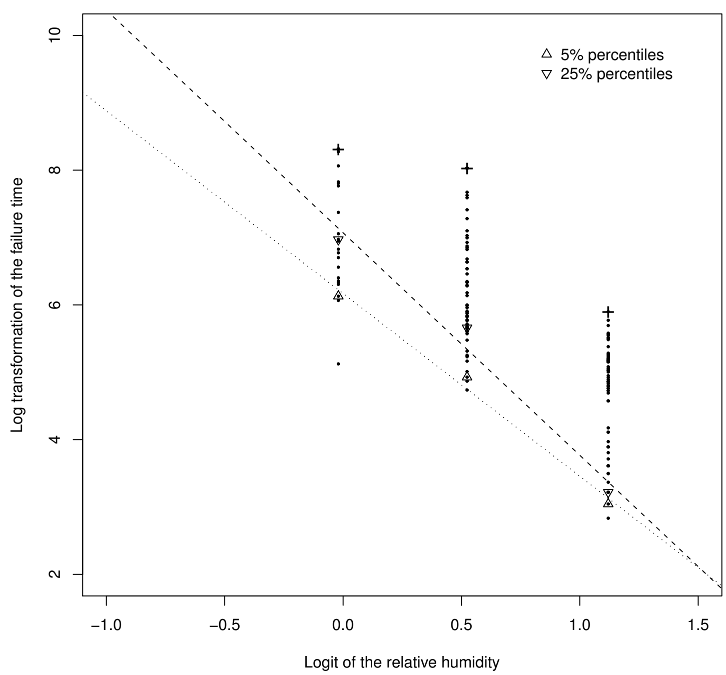

Figure 2: Empirical sample quantiles at different levels and the regression lines

and Whang (2006), we use the kernel K (rth-order) that is bounded and compactly supported on [-1,1], satisfying

$$\int u^j K(u)du = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } 1 \leq j \leq r-1, \\ \kappa & \text{if } j = r, \end{cases} \tag{22}$$

where $r \geq 2$ and constant $\kappa \neq 0$. Define $\mathcal{K}(x) = \int_{y<x} K(y)dy$, and $\mathcal{K}_h(x) = \mathcal{K}(x/h)$. Then, we can have a smoothed version of constraint function G, given by

$$G_h = \mathcal{K}_h(\boldsymbol{\theta}^\top x - T) - q. \tag{23}$$

In our experiment, we use a fourth-order kernel given by

$$\mathcal{K}(u) = \begin{cases} 0, & \text{if } u < -1 \\ 0.5 + \frac{105}{64}[u - \frac{5}{3}u^3 + \frac{7}{5}u^3 - \frac{3}{7}u^7], & \text{if } |u| \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{24}$$

The smoothing parameter h can be chosen using cross-validations, it is fixed at 0.2 for simplicity based on our preliminary studies. □

Remark 4. Let T be the survival time from location-scale families with parameters μ and σ, and $Z = (T - \mu)/\sigma$ is the standardized survival time. In general, the pth quantile of any location-scale family η_p is $\mu + c_p\sigma$, where c_p is the pth quantile for the standardized variable Z. For example, the extreme-value distribution has $c_p = \log(-\log(1-p))$. For the location-scale family, the difference of two percentiles is $\eta_{p2} - \eta_{p1} = (\mu + c_{p2}\sigma) - (\mu + c_{p1}\sigma) = (c_{p2} - c_{p1})\sigma$, which is a linear function of the scale parameter σ. Thus, the percentile-difference is a simple and proper replacement of the scale parameter σ for the heavy censoring case. □

Now, we compare our procedure with the commonly used Weibull regression model. Following Meeker and LuValle's formulation (1995),

$$F_{T_j}(t; \beta_0, \beta_1, \sigma) = \Phi_{EV}(T_j), \quad Z = (T_j - \mu(x_j))/\sigma,$$

where

$$\mu(x_j) = \beta_0 + \beta_1 \text{logit}(x_j), \quad x_j = \text{RH}/100, \quad \text{logit}(p) = \log[p/(1-p)], \tag{25}$$

and Φ_{EV} is the cdf of the standard extreme-value distribution. In this model, σ is the same at all levels, and the logit-transformation is justified (Meeker and LuValle, 1995). The parametric MLEs for model parameters are calculated as $\hat{\beta}_0 = 9.10$, $\hat{\beta}_1 = -3.78$, $\hat{\sigma} = 0.93$, respectively.

Figure 3 shows the profile likelihood plots for each of the three Weibull regression parameters. The horizontal lines on Figure 3 are drawn such that the their intersection with the profile likelihood provide approximate 95% confidence intervals (CIs) based on inverting the likelihood-ratio (LR) test. Using these plots, one can obtain the 95% LR-based CIs for β_0, β_1 and σ as $(8.82, 9.43)$, $(-4.17, -3.43)$ and $(0.83, 1.05)$, respectively.

The estimate of the pth quantile η_p of T is $\hat{\eta}_p(x) = \hat{\beta}_0 + \hat{\beta}_1 logit(x) + w_p\hat{\sigma}$, where $w_p = \log[-\log(1-p)]$. Confidence intervals for $\eta_p(x)$ can be obtained by using the large-sample normal approximation with the asymptotic variance calculated from the Fisher information matrix (Lawless 1982). Under the normal-use condition (RH=10%), the point estimate and CIs for the 5th percentile $\eta_{0.05}$ are calculated as 14.64 and $(13.54, 15.70)$, where the scales are in hours after the log-transformation.

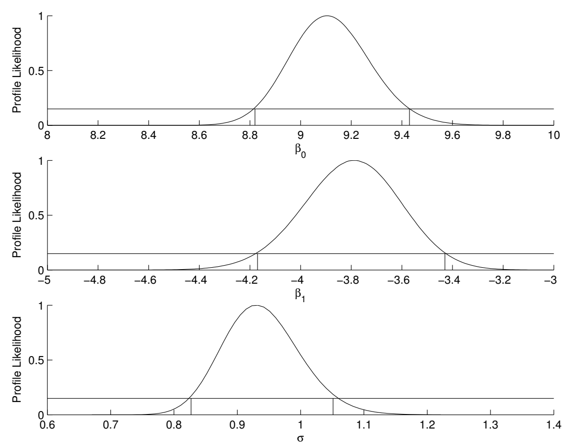

Figure 3: Profile likelihoods of β_0, β_1 and σ using the Weibull regression model

Figure 4 compares the confidence intervals using the AMELE method, where the percentile regression coefficients γ_0 and γ_1 are specified directly through regression functions $\eta_p(x) = \gamma_0 + \gamma_1 logit(x)$. Using the delta method, the corresponding point estimate and CI of the 5th percentile lifetime at the normal-use condition are 12.30 and $(11.83, 12.78)$, where the units are in hours after log-transformation. Note that the width of this CI is only about 44% of the width for the CI calculated using the Weibull regression model. After back-transforming the estimate to the original time scale, the 5th percentile lifetime is predicted as 25 years. The result that relies on the physics-based kinetics model given in Meeker and LuValle (1995) cannot produce a proper prediction for the 5th percentile, since the proportion of product failing is less than 1% under the normal-use condition.

Next, we explore the difference in predicting the survival functions. Specifically, we examine

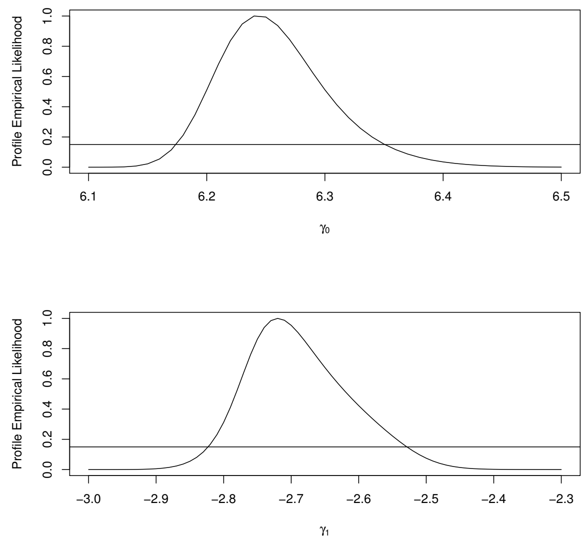

Figure 4: Profile empirical likelihoods for γ_0 and γ_1 using the AMELE method

the survival function of the failure time at the normal-use condition under different distribution assumptions. The data exploration analysis in Figure 2 shows that the 5th percentile regression and the percentile-difference regression provide possible adjustments for location and scale of the lifetime distributions at three stress levels. Consider the following two cases for this comparison.

- Case (i) − After adjusting the 5th percentiles, lifetime distributions are the same.
- Case (ii) − After adjusting the 5th percentiles and re-scaling with the percentile-difference (25th - 5th percentile), lifetime distributions are the same.

Both cases can be justified by applying the nonparametric two-sample Wilcoxon-test to the adjusted-data at the higher stress levels. In Case (i), we have one regresson function on the 5th percentile with values computed in the previous paragraph. In Case (ii), we have two regression functions on both the 5th and the 25th percentiles. The AMELE in Case (i) estimates the 5th percentile regression parameters (γ_0, γ_1) as (6.25, -2.71). Correspondingly, the AMELE in Case (ii) leads to (6.30, -2.68). The lifetime prediction of the 5th percentile lifetime are 20 and 22 years for Case (i) and (ii), respectively. Note that with the adjustment from the scale, the lifetime distribution in Case (ii) should be much more spread out than the one in Case (i). This shows in the estimates of the survival function plotted in Figure 5. Figure 6 provides the point-wise confidence intervals for the survival function in Case (ii). Because there are more censored observations in the right tail, those intervals are larger than the ones in the left tail.

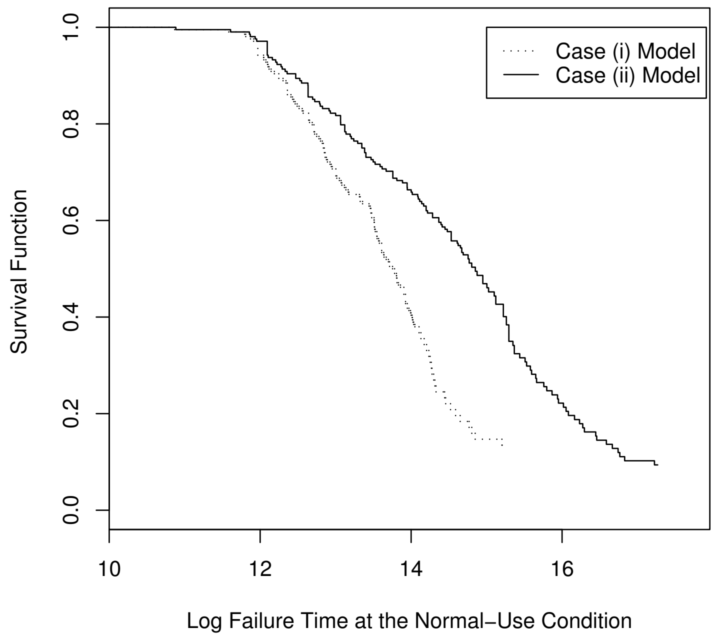

Figure 5: AMELE of survival functions under different assumptions.

5 Simulation Study

For the first example, we present more extensive simulation study on the finite-sample properties of the proposed methods by varying the sample sizes. Then, we compare our methods with two well-known median regression methods when failure times are less frequently censored. Finally, to examine the computational efficiency, we compute median failure time to compare the AMELE method with an iterative EM-type algorithm.

Our simulation studies will focus on the location-scale family of failure time distributions. Let $\eta_{p,k}$ be observed pth quantiles of survival times T_k at the stress level k. Since we do not observe the location and scale parameters directly, it is more sensible to apply regression functions on lower percentiles, $\eta_{p1,k} = \beta_{01} + \beta_{11}x_k$ and $\eta_{p2,k} = \beta_{02} + \beta_{12}x_k$, where x_k is the stress at the level k. Thus, after the following transformation:

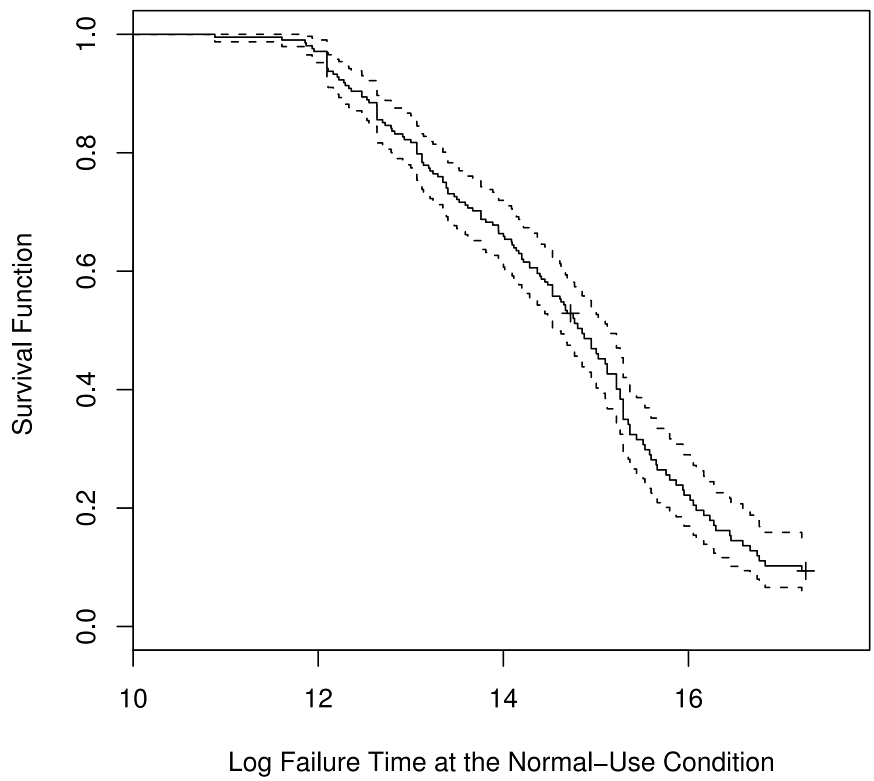

Figure 6: AMELE of survival function of the failure time

$$
\begin{aligned}
(T_k - \eta_{p2,k})/(\eta_{p2,k} - \eta_{p1,k}) &= (T_k - \mu_k - c_{p2}\sigma_k)/[(c_{p2} - c_{p1})\sigma_k] \\
&= Z/(c_{p2} - c_{p1}) - c_{p2}/(c_{p2} - c_{p1}),
\end{aligned}
$$

where $Z = (T_k - \mu_k)/\sigma_k$ is the standardized survival time. Note that the pth percentile of any location-scale family η_p is $\mu + c_p\sigma$, where c_p is the pth quantile for the standardized variable Z. With this transformation we are able to normalize the survival data at different levels, thus improve the estimation quality for the survival function. In all these simulation examples, we use the same smoothing function G. The first two examples simulate data from extreme-value distribution, while the third example uses log-normal distribution.

5.1 Simulation of Accelerated Life Test

In this simulation study, we consider three levels with $x_k = k$, and assume the failure times are governed by the extreme-value distribution with 10th and 25th percentiles following regression function $\eta_{0.1,k} = -5 + x_k$, and $\eta_{0.25,k} = -2 + 0.5x_k$. The censoring variable is simulated from an exponential distribution such that failure times are censored at 50% under different stress levels. An iterative algorithm is used to search for the optimal level of a particular β, while fixing the values of others. The following table shows the performance of parameter estimates with increasing sample size n. Bias and MSE are computed from bootstrap sampling with 1000 samples.

Sample size	$\beta_{01} = -5$	$\beta_{11} = 1$	$\beta_{02} = -2$	$\beta_{12} = 0.5$
$n = 20$	-0.122(0.940)	-0.051(0.309)	-0.142(0.548)	-0.069(0.23)
$n = 50$	-0.015(0.547)	-0.012(0.203)	-0.018(0.344)	-0.007(0.140)
$n = 100$	-0.053(0.368)	-0.035(0.155)	-0.018(0.260)	-0.010(0.109)

Table 1: Bias(MSE) of AMELE parameter estimates for simulated accelerated life test data with different sample sizes

By transforming observations using $(T - \eta_{p2})/(\eta_{p2} - \eta_{p1})$, we compute the AMELE of the survival curves using all the observation at different levels. In Figure 7, we show the 95% confidence bands from the survival curves of 200 bootstrap samples with sample size $n = 50$. For comparison, we also plot the confidence bands from the Kaplan-Meier estimates at one level. We see that the AMELE survival bands are narrower, especially in the lower percentiles. The AMELE incorporates more information than the one-level estimate, so it is not surprising that it produces narrower confidence bands. The regression functions are setup for 10th and 25th percentiles, so the estimation quality is better in that area than high percentile area.

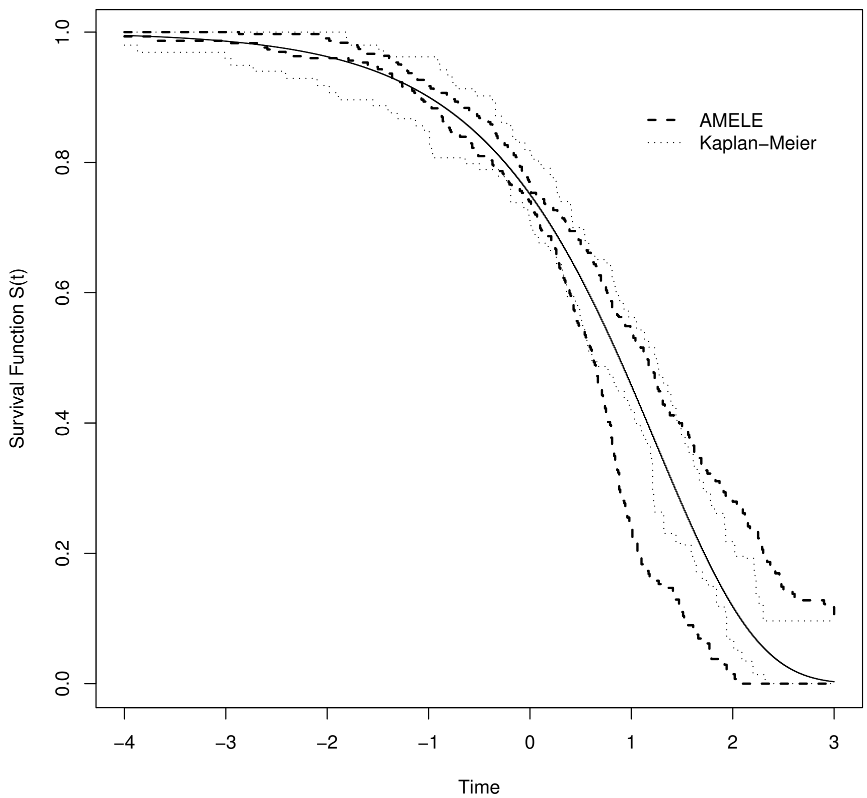

Figure 7: Comparison of confidence bands of survival functions using AMELE and Kaplan-Meier methods

5.2 Comparison of Several other Semiparametric Methods

In the previous section, we demonstrated that the AMELE method can estimate the regression parameters and the distribution function simultaneously. In the case of infrequent censoring when we observe median failure times at different levels, we can compare our methods with some benchmark results of semi-parametric median regressions from Ying (1995) and Yang (1999).

Consider four stress levels: $x =$ -1, 1, 2 and 3, which correspond to normal, low, medium and high stress levels. The failure time data are simulated from the extreme-value distribution with shape parameter $u = 4 - x$, and scale parameter $b = 3 - x/2$. Then the regression function on the median η is $\eta = u + b \times \log(-\log(0.5))$. The censoring time is chosen such that the censoring proportions are: 10%, 20% and 30% at different stress levels. After calculating the regression parameters, we compute the estimated median at the normal-use condition ($x = -1$) through extrapolation. In Table 2, results based on 1000 iterations of different sample sizes show that the finite-sample properties of median estimates using different methods. The performance of different methods are quite close, which shows that AMELE achieves comparable results to those specialized median regression models.

Sample Size	AMELE Bias(MSE)	Ying95 Bias(MSE)	Yang99 Bias(MSE)
20	0.055(0.149)	0.071(0.120)	0.01(0.119)
50	0.007(0.059)	0.024(0.049)	0.004(0.053)
100	-0.006(0.029)	0.003(0.025)	-0.001(0.027)

Table 2: Performance comparison of median regression using $AMELE$ and several nonparametric methods with different sample sizes

5.3 Comparison of AMELE and MELE

In Zhou (2005a), an EM-based method is proposed to compute the estimator directly for censored data. Here, we compare the median estimates using our AMELE estimator with the MELE estimator which has been implemented in *emplik* package in R (2009).

We consider the case of median estimation for right-censored failure times, and compare the estimates properties such as coverage, length of confidence intervals and mean square error (MSE) from simulation of 1000 iterations. In each iteration, we first simulate n random samples of failure-times from a normal distribution, in which the mean equals the median. We then generate censoring times from the exponential distributions, where the rates are chosen to ensure a certain proportion of censoring occurs. For each iteration, a grid search is used to locate the optimal estimate, and we also compute the 95% profile likelihood-ratio based confidence intervals.

Table 5.3 shows the performance comparison of the two methods. We see that overall, the MSE and CI length of MELE is about 82% of the MSE and CI length of AMELE. Although AMELE does not have the estimation efficiency as the MELE, it does offers more efficient computational steps, which does not increase as we have higher censoring. We test the speed of two approaches based on 10000 iterations of function calls. In each function call, it computes the empirical likelihood estimates of survival probability from simulated samples with size 50. In the case of no censoring, the computation time for the MELE is 60% greater than that of the AMELE, which is a significant increase. With 20% censored data, however, it takes over 15 times as long due to the slow convergence of the EM algorithm. This demonstrates that AMELE is an efficient alternative to compute the empirical likelihood estimator, especially during cases of heavy censoring.

Censoring proportion	Sample size	AMELE			MELE		
		Coverage	CI Length	MSE	Coverage	CI Length	MSE
0%	20	0.98	1.729	0.248	0.94	1.4380	0.185
	50	0.91	1.261	0.147	0.97	1.071	0.077
	100	0.97	0.917	0.056	0.93	0.762	0.041
10%	20	0.97	1.701	0.278	0.89	1.428	0.210
	50	0.89	1.284	0.165	0.95	1.109	0.081
	100	0.94	1.017	0.068	0.95	0.774	0.044
20%	20	0.975	1.679	0.277	0.92	1.466	0.230
	50	0.940	1.380	0.134	0.96	1.158	0.079
	100	0.960	1.070	0.063	0.97	0.833	0.041

Table 3: Performance comparison of $MELE$ versus $AMELE$ for median estimates under different censoring rate

6 Concluding Remarks

In the ALT experiment for printed ciruit boards, the AMELE method provides a reasonable estimator for PCB lifetime and has important advantages over previous estimators. The proposed data-exploration based percentile and percentile-difference regressions are effective in overcoming the difficulty of observing mean lifetimes in the heavily censored data case for constructing commonly used mean and variance regression models in ALT studies. Numerical studies show that the AMELE is reasonably competitive against other semi-parametric MLE methods, and also compares favorably to the MELE method. Based on the properties derived in this article, the AMELE method should be a strong candidate for handling challenging data modeling and statistical inference problems.

Appendix: Proof of Lemmas and Theorems

Proof of Lemma 1: If we set $\boldsymbol{\lambda} = 0$ such that constraint is imposed on the empirical likelihood, we have:

$$\begin{aligned} a_i(0) &= \frac{1}{n}\sum_{j=1}^{i} \frac{c_j}{\sum_{l=j}^{k+1} P_{l,0}}, \\ P_{i,0} &= \frac{1}{n(1-a_i(0))}, \quad \sum_{i=1}^{k+1} P_{i,0} = 1, \quad i = 1, 2, \ldots, k. \end{aligned} \tag{26}$$

Let $S_i = 1 - \sum_{j=1}^{i} P_j$ be the survival probability, and $h_i = (S_{i-1} - S_i)/S_{i-1}$ be the hazard rate, where $i = 1, 2, \ldots, k+1$, and $S_0 = 1$. We have $P_i = h_i S_{i-1} = h_i \prod_{j=1}^{i-1}(1-h_j)$. Thus, (26) can be rewritten as

$$h_i \prod_{j=1}^{i-1}(1-h_j) = \frac{1}{n - \sum_{j=1}^{i} c_j/S_{j-1}}. \tag{27}$$

Denote by $n_i = n - \sum_{j=1}^{i} c_j - i + 1$ the number of subjects at risk at time t_i. By further simplification of (27), it is easy to see $H_1 = P_1/S_0 = 1/(n - c_1)$, and $H_i = 1/n_i$. So, the survival function is expressed as

$$S_{T,0}(t) = \prod_{t_i \le t}(1-h_i) = \hat{S}_{T,KM}(t). \tag{28}$$

Following similar arguments, we can show that

$$1 - a_i(0) = \hat{S}_{C,KM}(t_i), \tag{29}$$

which is the Kaplan-Meier estimate of the survival function $S_C(t)$ at time t_i. □

Proof of Lemma 2: Because $\mathrm{E}[G(T,\boldsymbol{\theta})] = \int_0^\infty G(t,\boldsymbol{\theta})dF_T(t) = 0$, we know

$$\int_0^\infty G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t) = \int_0^\infty G(t,\boldsymbol{\theta})d(\hat{F}_{T,KM}(t) - F_T(t)) = -\int_0^\infty G(t,\boldsymbol{\theta})d(\hat{S}_{T,KM}(t) - S_T(t)).$$

Using integration by parts, it follows that

$$\int_0^\infty G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t) = \int_0^\infty (\hat{S}_{T,KM}(t) - S_T(t))dG(t,\boldsymbol{\theta}). \tag{30}$$

According to Breslow and Crowley (1974, Theorem 5), $\sqrt{n}(\hat{S}_{T,KM}(t) - S_T(t))$ converges to a Gaussian process $W(t)$, with $\mathrm{E}(W(t)) = 0$ and

$$\mathrm{Cov}(W(s), W(t)) = S_T(s)S_T(t)\int_0^{\min(t,s)} \frac{dF_T(x)}{S_T(x)^2 S_C(x)}. \tag{31}$$

In the neighborhood of $\boldsymbol{\theta}_0$,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{k+1} Z_i(\boldsymbol{\theta}) &= \int G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t) \\ &= \int G(t,\boldsymbol{\theta}_0)d\hat{F}_{T,KM}(t) + \int \frac{\partial G(t,\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)d\hat{F}_{T,KM}(t) + O(n^{-1/2}) \\ &= \int G(t,\boldsymbol{\theta}_0)d\hat{F}_{T,KM}(t) - \int G(t,\boldsymbol{\theta}_0)dF_T(t) + O(n^{-1/2}) \\ &= \int (\hat{S}_{T,KM}(t) - S_T(t))dG(t,\boldsymbol{\theta}_0) + O(n^{-1/2}),\end{aligned}$$

which proves the part (a) of the lemma.

As shown in Lemma 1, $1 - a_i(0) = \hat{S}_{C,KM}(t_i)$, and we can write

$$\frac{1}{n}\sum_{i=1}^{k+1} Z_i(\boldsymbol{\theta})Z_i(\boldsymbol{\theta})^\top = \int \frac{G(t,\boldsymbol{\theta})G(t,\boldsymbol{\theta})^\top}{\hat{S}_{C,KM}(t)}d\hat{F}_{T,KM}(t) \tag{32}$$

$$\rightarrow \int \frac{G(t,\boldsymbol{\theta})G(t,\boldsymbol{\theta})^\top}{S_C(t)}dF_T(t). \tag{33}$$

Following the uniform consistency of Kaplan-Meier estimate and the bounded derivatives of $G(t,\boldsymbol{\theta})$, part (b) is proved. □

Proof of Lemma 3: $\boldsymbol{\lambda}$ is solved implicitly from $\sum_{i=1}^{k+1} Z_i(\boldsymbol{\theta})/(1+\boldsymbol{\lambda}^\top Z_i(\boldsymbol{\theta})) = 0$. Since we have $\frac{1}{n}\sum_{i=1}^{k+1} Z_i(\boldsymbol{\theta})Z_i(\boldsymbol{\theta})\top < \infty$, it is easy to verify that $\max_i||Z_i(\boldsymbol{\theta})|| = o(n^{-1/2})$. Following the steps used in Owen (1990), we can establish that

$$\frac{||\boldsymbol{\lambda}||}{1+||\boldsymbol{\lambda}||\max_i||Z_i(\boldsymbol{\theta})||} = O_p(n^{-1/2}),$$

which implies that $||\boldsymbol{\lambda}|| = O_p(n^{-1/2})$.

Proof of Theorem 1:

According to (13), $\boldsymbol{\lambda} = -A(\boldsymbol{\theta})^{-1}\left[\int_0^\infty G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t)\right] + o_p(n^{-1/2})$. It follows from (30) that (under condition **A3**) as $n \to \infty$,

$$\sqrt{n}\int_0^\infty G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t) \xrightarrow{p} \int_0^\infty W(t)dG(t,\boldsymbol{\theta}).$$

Using Gaussian process properties, we know that $\int_0^\infty W(t)\ dG(t,\boldsymbol{\theta})$ is distributed normal with mean zero and covariance matrix $\boldsymbol{\Sigma}_G(\boldsymbol{\theta})$ defined as

$$\boldsymbol{\Sigma}_G(\boldsymbol{\theta}) = \int_0^\infty \left\{\int_x^\infty (G(x,\boldsymbol{\theta}) - G(t,\boldsymbol{\theta}))dS_T(t)\right\}\left\{\int_x^\infty (G(x,\boldsymbol{\theta}) - G(t,\boldsymbol{\theta}))dS_T(t)\right\}^\top \frac{dF_T(t)}{S_T^2(t)S_C(t)}.$$

Thus, $\sqrt{n}\boldsymbol{\lambda}$ is asymptotically normal with mean zero and covariance matrix

$$\boldsymbol{\Sigma}_\lambda(\boldsymbol{\theta}) = \boldsymbol{A}(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}_G(\boldsymbol{\theta})\boldsymbol{A}(\boldsymbol{\theta})^{-1},$$

where $\boldsymbol{A}(\boldsymbol{\theta})$ is given by (9). □

Proof of Theorem 3:

Again, consider $\boldsymbol{\theta}$ in a ball $\{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n^{-1/2}\}$. Because $\boldsymbol{\lambda}(\boldsymbol{\theta}) = O_p(n^{-1/2})$, we can construct a Taylor expansion of $\tilde{S}_T(t)$ at $\boldsymbol{\lambda} = 0$ following a similar procedure in Qin and Lawless (1994), which results in

$$\begin{aligned}
\tilde{S}_T(t) &= \sum_{t_i > t}\left(\frac{1}{n(1-a_i(0))} + \frac{G^\top(t_i,\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta})}{n(1-a_i(0))^2} + o_p(n^{-1/2})\right) \\
&= \hat{S}_{T,KM}(t) + \int_t^\infty \frac{G^\top(x,\boldsymbol{\theta})d\hat{F}_{T,KM}(x)}{\hat{S}_{C,KM}(x))}\boldsymbol{\lambda}(\boldsymbol{\theta}) + o_p(n^{-1/2}),
\end{aligned}$$

where $\hat{F}_{T,KM}(t) = 1 - \hat{S}_{T,KM}(t)$, and $\hat{S}_{T,KM}(t)$ and $\hat{S}_{C,KM}(t)$ are the Kaplan-Meier estimates of $S_T(t)$ and $S_C(t)$. By replacing $\boldsymbol{\lambda}(\boldsymbol{\theta})$ with $-\boldsymbol{A}^{-1}(\boldsymbol{\theta})\int_0^\infty G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t)$,

$$\tilde{S}_T(t) = \hat{S}_{T,KM}(t) + \left(\int_t^\infty \frac{G^\top(x,\boldsymbol{\theta})d\hat{F}_{T,KM}(x)}{\hat{S}_{C,KM}(x))}\right)\boldsymbol{A}^{-1}(\boldsymbol{\theta})\int_0^\infty G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t) + o_p(n^{-1/2}).$$

It follows that

$$\begin{aligned}
\sqrt{n}(\tilde{S}_T(t) - S_T(t)) &= \sqrt{n}(\hat{S}_{T,KM}(t) - S_T(t)) + \\
&\quad \sqrt{n}\left(\int_t^\infty \frac{G^\top(x,\boldsymbol{\theta})d\hat{F}_{T,KM}(x)}{\hat{S}_{C,KM}(x)}\right)\boldsymbol{A}^{-1}(\boldsymbol{\theta})\int_0^\infty G(t,\boldsymbol{\theta})d\hat{F}_{T,KM}(t).
\end{aligned}$$

Denote

$$\boldsymbol{\gamma}(t,\boldsymbol{\theta}) = \lim_{n\to\infty}\boldsymbol{A}^{-1}(\boldsymbol{\theta})\int_t^\infty \frac{G(x,\boldsymbol{\theta})d\hat{F}_{T,KM}(x)}{\hat{S}_{C,KM}(x)} = \boldsymbol{A}^{-1}(\boldsymbol{\theta})\int_t^\infty \frac{G_(x,\boldsymbol{\theta})dF_T(x)}{S_C(x)},$$

then we have

$$\sqrt{n}(\tilde{S}_T(t) - S_T(t)) = \sqrt{n}(\hat{S}_{T,KM}(t) - S_T(t)) + \sqrt{n}\boldsymbol{\gamma}^\top(t,\boldsymbol{\theta})\int_0^\infty G(x,\boldsymbol{\theta})d\hat{F}_{T,KM}(x) = W_{n1}(t) + W_{n2}(t),$$

which are decomposed into summation of two Gaussian processes, with

$$\begin{aligned}
\lim_{n\to\infty} W_{n2}(t) &= \lim_{n\to\infty}\sqrt{n}\boldsymbol{\gamma}^\top(t,\boldsymbol{\theta})\int_0^\infty G(x,\boldsymbol{\theta})d\hat{F}_{T,KM}(x) \\
&= \lim_{n\to\infty}\boldsymbol{\gamma}^\top(t,\boldsymbol{\theta})\int_0^\infty \sqrt{n}(\hat{S}_{nT,KM}(x) - S_T(x))dG(x,\boldsymbol{\theta}) \\
&= \boldsymbol{\gamma}^\top(t,\boldsymbol{\theta})\int_0^\infty W_1(x)dG(x,\boldsymbol{\theta}).
\end{aligned}$$

Note that $\mathrm{E}(W_1(t) + W_2(t)) = 0$, so that the asymptotic variance of $\sqrt{n}(\tilde{S}_T(t) - S_T(t))$ reduces to

$$\sigma_{\tilde{S}(t)} = \mathrm{Var}(W_1(t)) + \mathrm{Var}(W_2(t)) + 2\mathrm{Cov}(W_1(t), W_2(t)),$$

which could be proven following derivation of standard Gaussian process properties. □

References

[1] Breslow, N., and Crowley, J. (1974), "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship," *The Annals of Statistics*, 2(3), 437-453.

[2] Chen, D., Lu, J.-C., and Lin, S. C. (2005), "Asymptotic Distribution of Semiparametric Maximum Likelihood Estimations With Estimating Equations for Group-Censored Data," *Aust.N.Z.J.Stat.*, 47(2), 173-192.

[3] Chen, S.X., Hall, P. (1993), "Smoothed Empirical Likelihood Confidence Intervals for Quantiles," *Ann. Statist.*, 21, 1166-1181.

[4] Cheng, S.C., Wei, L.J., and Ying, Z. (1997), "Predicting Survival Probabilities With Semiparametric Transformation Models," *J. Amer. Statist. Assoc.*, 92, 227- 235.

[5] Kaplan, E. L., and Meier, P. (1958), "Nonparametric estimation from incomplete observation," *J. Amer. Statist. Assoc.*, 58, 457-481.

[6] Kalbfleish, J. D., and Prentice, R. L, (2002). *The Statistical Analysis of Failure Time Data*, Wiley.

[7] Lai, T. L., Ying, Z. (1991), "Rank Regression Methods for Left-Truncated and Right-Censored Data," *Ann. Statist.*, 19, 531-566.

[8] Lawless, J. F. (1982). *Statistical Models and Methods for lifetime Data*. John Wiley and Sons: New York.

[9] Li, G., Wang, Q. H. (2003), "Empirical Likelihood Regression Analysis for Right Censored Data," *Statistica Sinica*, 13, 51-68.

[10] Liang, K. Y., and Zeger, S. L. (1986), "Logitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 12-22.

[11] Lu, J.-C., Chen, D. and Gan, N. (2002), "Semi-Parametric Modeling And Likelihood Estimation with Estimating Equations," *Aust. N.Z. J. Stat.*, 44(2), 193-212.

[12] Meeter, C.A., Meeker, W.Q. (1994), "Optimum Accelerated Life Tests With a Nonconstant Scale Parameter," *Technometrics*, 36, 71-83.

[13] Meeker, W. Q., Luvalle, M.J. (1995), "An Accelerated Life Test Model Based on Reliability Kinetics," *Technometrics*, 37, 133-146.

[14] Murphy, S.A., Rossini, A.J., van der Vaart, A.W. (1997), "Maximum Likelihood Estimation in the Proportion Odds Model," *J. Amer. Statist. Assoc.*, 25, 1471-1509.

[15] Murphy, S.A., van der Vaart, A.W. (1997), "Semparametric Likelihood Ratio Inference," *Ann. Statist.*, 25, 1471-1509.

[16] Owen, A. B. (2001), *Empirical Likelihood*, Chapman & Hall/CRC.

[17] Owen, A. B. (1990), "Empirical Likelihood Ratio Confidence Regions," *Ann. Statist.*, 18, 90-120.

[18] Pan, X.R., Zhou, M. (2000), "Empirical Likelihood Ratio in terms of Cumulative Hazard Function for Censored Data," Technical report, University of Kentucky, Department of Statistics.

[19] Qin, J., and Lawless, J. F. (1994), "Empirical Likelihood and General Estimating Equations," *Ann. Statist.*, 22, 300-325.

[20] R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, http://www.R-project.org.

[21] Shaked, Zimmer, and Ball (1979), "A Nonparametric Approach to Accelerated Life Testing," *Journal of the American Statistical Association*, 74, 694-699.

[22] Wedderburn, R. W. M. (1974), "Quasi-Likelihood Functions, Generalized Linear Models, and Gause-Newton Method," *Biometrika*, 61, 439-447.

[23] Whang, Y.J. (2006), "Smoothed Empirical Likelihood Methods for Quantile Regression Models," *Econometric Theory*, 22, 173-205.

[24] Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Survival and Hazard Functions," *Journal of the American Statistical Association*, 94, 137-145.

[25] Ying, Z., Jung, S.H., and Wei, L.J., (1995) "Survival Analysis with Median Regression Models," *Journal of the American Statistical Association*, 90, 178-184.

[26] 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.

[27] Zhou, M. (2005b), "Empirical Likelihood Analysis of the Rank Estimator for the Censored Accelerated Failure Time Model," *Biometrika*, 92, 492-498.