An Empirical Likelihood With Estimating Equation Approach for Modeling Heavy Censored Accelerated Life-Test Data

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 uses empirical likelihood with estimating equations to model and analyze heavy censored accelerated life testing data. This approach flexibly and rigorously incorporates distribution assumptions and regression structures into estimating equations in a nonparametric estimation framework. Real-life examples of using available data to explore the regression functional relationship and distribution assumption are provided. Derivation of asymptotic properties of the proposed method provides an opportunity to compare its estimation quality to commonly used parametric MLE methods in the situation of misspecified regression models. These real-life examples and asymptotic studies show a significant potential of the proposed methodology.

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

+ Corresponding Author : jclu@isye.gatech.edu

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 lifetime quantities (e.g., 5th or 10th percentile, mean, variance and lifetime distribution) 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. Because high durability of products and limited testing time, this practice results in heavy data censoring. 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 of scale parameters has a linear regression relationship.

According to the research in Hutton and Monaghan (2002) and Pascual and Montepiedra (2005), selecting an inappropriate lifetime distribution could have significant impact (in terms of estimation bias and variance). However, in data exploration, it is often that several lifetime distributions (e.g., lognormal and Weibull) and are not rejected from goodness-of-fit tests. Several recent studies are semiparametric based (e.g., Yang (1999) and Yu and Wong (2005)).

This article will focus on semiparametric approaches. Semiparametric approaches drop the assumption of distribution forms, but retain the functional relationship of distributions at different stress levels. The widely used Cox Proportional Hazards (PH) model (Kalbfleisch and Prentice, 1980) assumes that hazard rates under different stresses are proportional to a baseline hazard rate. The Accelerated Failure Time (AFT) model assumes that the logarithm of the survival functions under different stresses differ only in location parameters. For instance, Yang (1999) assumed that distributions pertaining to various stress levels differ only by a median shift.

In some cases, the traditional AFT 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 AFT model structure. Because the traditional regression-over-the-mean approach is questionable (especially in this 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 AFT model, explicit physical relations are rarely available to aid the data modeling so directly. Thus, there is a need of developing a *data exploration* approach to entertain potential regression models and to examine the goodness-of-fit of the assumed lifetime distribution. For example, the regression relationships between percentiles are used in Section 2 for exploring models.

In this article, the regression relationships in the semiparametric models are treated as estimating equations (EE) serving as constraints in maximizing the Empirical Likelihood (EL). Compared to the recent semiparametric modeling methods, our $EL - EE$ approach is more flexible, rigorous and easy to understand. For example, Yu and Wong (2005) formulated a profile likelihood function for the mean regression parameters $\mathbf{b}$. The survival function $\hat{S}_{\mathbf{b}}$, used in the likelihood to describe the censored data, was estimated by the product-limit method. The probability density function needed to describe the complete samples is estimated from a kernel method based on $\hat{S}_{\mathbf{b}}$. Section 3 shows that the profile likelihood can be formulated rigorously through the proposed semiparametric maximum likelihood estimation (SMLE) methods without having to apply a kernel method. Another commonly used method in survival analysis is to estimate the hazard and/or survival function $\hat{S}_{\mathbf{b}}$ with weighted empirical function, where the weights are usually constructed based on intuition. Because $\hat{S}_{\mathbf{b}}$ is implicitly a function of the mean regression parameters $\mathbf{b}$ defined in the mean-adjusted data, it forms an estimating equation for finding the solution for $\mathbf{b}$. Yang (1999) provides several references for examples using this approach. Lu, Chen and Gan (2002) showed that the EL-EE method provided the weights determined by the SMLE method. Moreover, the EL-EE approach has the flexibility to include other estimating equations such as regression function of range (or variance) parameters (see Section 5.2 for details).

Empirical likelihood was developed by Owen (1990) as a general nonparametric inference procedure. Qin and Lawless (1994) demonstrated that the empirical likelihood method with additional estimating equations can be useful in incorporating distribution knowledge to improve estimation quality. Recently, the empirical likelihood method has been shown to work well in difficult inference problems involving censored or truncated data (Pan and Zhou 2000). 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. Furthermore, the confidence regions are automatically determined without estimating the variance of test statistics, which can be difficult in the case of the rank-based regression estimators in censored ALT models.

The current literature in empirical likelihood does not provide the technical derivation of the asymptotic properties of the SMLE for *randomly censored data*. Chen, Lu and Lin (2005) only considered interval-censored data and derivations based on empirical-process theory was not involved. The main challenge in the asymptotic derivation is that the SMLE does not have a closed form expression and is defined only through several implicit functions. Thus, derivation of its asymptotic distribution is much more difficult than in the complete sample situation (see Qin and Lawless (1994) for the results in this case). The traditional large-sample studies in the survival analysis (e.g., Breslow and Crowley (1974) and Tsiatis (1981)) of the product-limit estimate and the Cox regression do not include estimating equations, which make the SMLE messy. Because the log-profile-likelihood cannot be expressed as a sum of independent and identically distributed (iid) random variables, it is challenging to derive the asymptotic properties of the SMLE when random censoring occurs.

Section 2 shows some data exploration studies based on two reliability data sets from industrial tests. Section 3 defines the empirical likelihood and formulates the ALT regression model in the estimating equations. Then, the semiparametric maximum likelihood estimation (SMLE) method is proposed. Section 4 shows the asymptotic properties for the proposed estimators. Real life examples of ALT data and asymptotic-efficiency studies are presented in Section 5 to illustrate and compare the proposed methods with parametric MLEs. Section 6 provides a few concluding remarks.

2 Examples from Industry

We employ two different examples in this section based on circuit board testing and reliability for switches. Our main focus is on a conventionally ALT experiment (Example 1). The second example illustrates how life testing can be an important in reliability improvement studies. Besides the limited failure time data, other source of information such as degradation data (and physical knowledge) could be helpful in setting up estimating equations in the proposed EL-EE method.

EXAMPLE 1. Meeker and LuValle (1995) reported on an accelerated life test experiment for 72 or

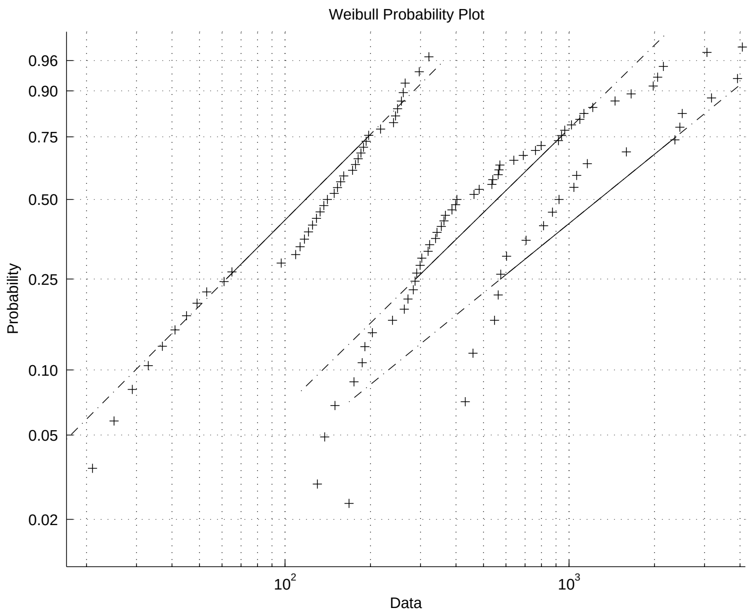

Figure 1: Weibull Probability Plot for the Failure Data Under RH = 49.5, 62.8, 75.4% (from right to left)

so printed-circuit-boards (PCB) at four high relative humidity (RH) conditions: 49.5% RH, 62.8% RH, 75.4% RH and 82.4% RH. They integrated chemical kinetics into a probability model to derive the failure time distribution of PCBs at the normal-use conditions (10% and 20% RH). Their data exploration showed that the failure time distribution for data collected at the highest stress level (82.4% RH) seemed to be different from those at other lower stress levels. As stated in their paper, the physics of PCB failure at the 82.4% RH level is not well understood. Since it is the furthest from the normal-use condition, the 72 failure data collected at 82.4% RH level were discarded. Figure 1 shows the Weibull probability plot of the data from three other 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.

The chemical-kinetics-probability distribution derived in their paper is

$$\begin{aligned} Pr(T < t) &= F_T(t; \beta_0^{[k_1]}, \beta_1^{[k_1]}, \beta_0^{[k_2]}, \beta_1^{[k_2]}, \sigma) \\ &= \Phi \left\{ \left[-\log[[(k_1 + k_2)/k_1]\{1 - \exp[-(k_1 + k_2)t]\}^{-1} - 1] + 6 \right] / \sigma \right\} \end{aligned} \tag{1}$$

and includes a tiny proportion (less than 1%) of population failing at the normal-use condition based on estimated model parameters. In fact, the mean (and variance) of the distribution are not

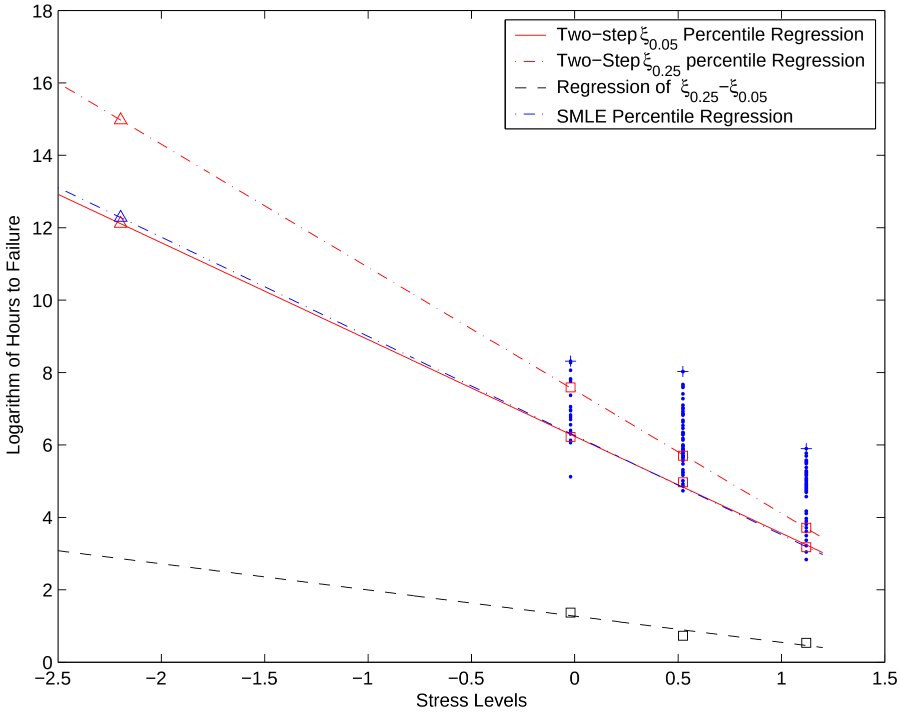

Figure 2: Percentile Regression and Prediction

finite, which makes the mean-variance based warranty studies impossible. The following humidity relationships for (1) are specified as:

$$k_1 = \exp[\beta_0^{[k_1]} + \beta_1^{[k_1]} g(H)], \quad k_2 = \exp[\beta_0^{[k_2]} + \beta_1^{[k_2]} g(H)],$$

where $\Phi(.)$ is the standard normal cumulative distribution function (cdf) and $g(H) = logit(RH/100)$.

Since lower percentile lifetime such as 5% is observable for all stress levels and important in reliability applications, we explore the regression structures based on them. Figure 2 shows that a simple log-linear regression model is suitable for the EL-EE approach. The proposed SMLE estimation elaborated in Section 5 shows that the estimated 5th percentile lifetime of PCBs at 10% RH is 25 years, where the regression prediction based on sample 5th percentiles is 21 years (see the dashed line with shot dots in Figure 2 for the regression). Because Meeker and LuValle's model will predict infinity for the 5th percentile lifetime, it cannot be used as a comparison measure for reliability improvement studies. The percentiles obtained from this data exploration give a more reasonable prediction for comparison measures. The examples presented below reinforce this finding. The width of the confidence interval constructed for the SMLE is 44% shorter than the one deduced from Weibull regression. Section 5 contains a more detailed comparison of the confidence intervals.

Since 25th percentiles are also available for all stress levels, we explore the trend of the "quantile-ranges" (QR) of the logarithm of the 25th and 5th percentiles over three stress levels. Figure 2 shows that the QR is not constant, but rather a linear function with much larger QR 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 QR, the lifetime distributions at all stress levels would be the same. Then, the SMLE gives the estimate and its point-wise confidence intervals. See Section 5 for an example.

EXAMPLE 2. Joseph and Yu (2005) developed robust parameter design methodology for improving product reliability based on degradation data. Their example of a window wiper switch experiment (Wu and Hamada, 2000, page 560) consists of one four-level factor and four two-level factors in an eight-run $OA(4^1 2^4)$ design with four products tested in each run. For each switch, the initial voltage drop (degradation measure) across multiple contacts is recorded every 20,000 cycles up to 180,000. Consider the "soft-failure" definition (Su, Lu, Chen and Hughes-Oliver, 1999) as a drop of voltage over 120. Table 1 presents the failure time. Note that Run #6 has only one failure. Table 2 presents degradation records for the censored cases, which are useful in predicting failure time. Focus on factor effects defined as the difference between the 12.5th percentile life at the high (+1) and low (-1) levels of each factor. Following procedures similar to those used in Joseph and Yu (2005), we identify D and E as significant effects with large differences and we obtained the following percentile regression model based on these differences.

$$y_{12.5\%-tile} = 5.33 - 0.46D - 1.36E.$$

Table 3 reports the failure time data organized in the D and E factor levels useful in the proposed semiparametric method for updating the percentile regression model and estimating the failure time distribution. This table shows that the combination of (D, E) at $(-1, -1)$ gives the largest 12.5th percentile lifetime (the first observation in the table).

EXAMPLE 3. From the empirical likelihood method explained in Section 3, we will find that when the data are heavily censored, there are insufficient failure time points for estimating probability mass. This makes the estimation of the failure time distribution especially troublesome. This example extends Example 2 by exploring ways to handle problems caused by heavy censoring.

Similar to the two effects identified in Joseph and Yu's (2005) mean regression model, B and E are the significant effects for median lifetimes. Run #4 and #6 in Table 1 show there is only one failure

Run	Factor					Failure Time* of Replicates			
	A	B	C	D	E	1	2	3	4
1	0	−	−	−	−	7.54	8.44	8.73	+(10.52)
2	0	+	+	+	+	4.10	4.69	5.31	8.37
3	1	−	−	+	+	3.89	4.45	4.45	6.99
4	1	+	+	−	−	8.70	+(10.56)	+(12.50)	+(36.75)
5	2	−	+	−	+	4.05	6.46	8.59	9.01
6	2	+	−	+	−	8.75	+(21.56)	+(22.81)	+(26.74)
7	3	−	+	+	−	5.85	6.46	7.07	7.35
8	3	+	−	−	+	7.43	8.65	9.82	+(12.64)

∗: the failure time is given in the unit of (time*−1)*20,000 cycles.

+: censored data (and projected failure time).

Table 1: Lifetime of Wiper Switches

Run	Replicate No.	Inspection Time									
		1	2	3	4	5	6	7	8	9	10
1	4	24	30	38	46	57	71	73	91	98	104
4	2	54	51	64	66	78	84	90	93	106	109
4	3	47	54	63	68	70	77	88	86	91	102
4	4	47	45	50	53	58	57	61	55	61	66
6	2	44	50	48	46	55	63	65	71	68	76
6	3	43	44	55	56	58	62	66	66	72	72
6	4	40	46	45	49	55	62	61	61	64	66
8	4	65	68	69	75	79	84	95	96	101	100

Table 2: Degradation Data for Censored Cases

Factor		Failure Time*							
D	E	1	2	3	4	5	6	7	8
−	−	7.54	8.44	8.70	8.73	+(10.52)	+(10.56)	+(12.50)	+(36.75)
−	+	4.05	6.46	7.43	8.59	8.65	9.01	9.82	+(12.64)
+	−	5.85	6.46	7.07	7.35	8.75	+(21.56)	+(22.82)	+(26.74)
+	+	3.89	4.10	4.45	4.45	4.69	5.31	6.99	8.37

Table 3: Reorganized Wiper Switch Testing Data

time and all of the other seven cases are censored. Since these two runs are from the best combination of B and E (at +1 and −1 levels, respectively) leading to the largest median lifetime, it is important to develop a method to estimate the lifetime distribution there; we present two approaches below.

First, one could assume that the lifetime distributions at different factor levels are the same (as done in most of ALT experiments) and use either mean or percentile to adjust the data into the same "distribution-scale" such that all the observed failure time points can be used to construct the empirical likelihood. This approach utilizes only the failure time data and not the degradation data. See Section 5.1 for an example.

Alternatively, when degradation data are available, one can either use the degradation path to "project" the failure time or impute the censored data based on the failure time distribution derived from the degradation model. An example for projecting the failure time is to extend the linear regression line of the degradation path for reaching the threshold defining the failure. Sometimes, the long-range extrapolation in ALT (e.g., see Figure 2) causes concern in projecting the failure time. However, if there exists physical knowledge supporting the derivation of failure time distribution, it undoubtedly improves the semiparametric estimation. This example will illustrate how to derive the failure time distribution in the case of heavy censoring. Then, we can combine this knowledge with the observed failure time data at other factor levels for semiparametric estimation.

Consider the following simplification from Joseph and Yu's (2005) model.

$$dY_t/dt = \beta + \sigma W,$$

where Y_t is the voltage degradation drop, t represents the testing-cycle time, and β is the (positive) intercept. The parameter σ is the standard deviation of the white noise W. By integrating with respect to t, we obtain

$$Y_t = Y_0 + \beta t + \sigma W t,$$

where Y_0 is the initial amount of degradation at $t = 0$.

For deriving the failure time distribution, following the traditional random-coefficient model (e.g., Lu and Meeker, 1993; Lu, Park and Yang, 1997), consider β as a $N(\mu_\beta, \sigma_\beta)$ random variable associated with material (or manufacturing) variations. If W has the standard normal distribution, Y_0 is distributed $N(\mu_{Y_0}, \sigma_{Y_0})$. Denote T the time when degradation reaches the threshold y_f and (soft) failure occurs. Thus,

$$P(T \le t) = P[(y_f - Y_0)/(\beta + \sigma W)] \approx P[y_f \le Y_0 + (\beta + \sigma W)t],$$

where the approximation came from the assumption that $P[(\beta + \sigma W) \le 0] \approx 0$. From aggregating the correlated normal random variables, the cdf of T can be expressed as

$$F_T(t) = 1 - \Phi\left\{(\mu_{Y_0} + \mu_\beta t - y_f)/[{\sigma_{Y_0}}^2 + {\sigma_\beta}^2 t^2 + 2t\rho_{Y_0,\beta}\sigma_{Y_0}\sigma_\beta + \sigma^2]^{1/2}\right\},$$

where we assume the independence between the noise W and β, Y_0, but there is a possible correlation between β and Y_0 due to the material property of products.

This model is similar to the failure time cdf derived in Lu, Park and Yang (1997) as a generalization of Berstein's model (Gertsbakh and Kordonskiy, 1969 page 88), where the moments do not exist. However, the percentile lifetime is available for data exploration analysis. For example, the estimates of the parameters in $F_T(t)$ for products tested in Run #6 are given as $\mu_{Y_0} = 46.52$, $\sigma_{Y_0} = 7,37$, $\mu_\beta = 4.25$, $\sigma_\beta = 1.95$ and $\rho_{Y_0,\beta} = 0.98$. Thus, the median lifetime for Run #6 is "derived" as 17.29 in the failure time unit reported in Table 1 and 3. This failure time is equivalent to 325,788 cycles in the experimental setup.

Following the hybrid of parametric and nonparametric maximum likelihood estimation studied in Qin (2000) and Wang, Lu and Kvam (2005), the physics-based failure time distribution for the low-stress-level degradation data can be combined with the empirical likelihood for the high-stress-level failure time data to estimate lifetime in normal-use conditions.

3 Semiparametric MLE (SMLE)

3.1 The Empirical Likelihood and ALT Regression Model

Suppose that the accelerated life test is conducted under m different stress levels $\{x_1, ..., x_m\}$, and there are n_j replicates at stress level x_j. Denote the normal-use stress level by x_0. For the jth sample (with stress level x_j), let T_j and C_j be the general notation for the failure time and censoring random variables, respectively. Assume that C_j and T_j are independent.

Denote the probability density function (pdf), the survival function (SF) and the cdf for the failure time T_j as $f_{T_j}(t)$, $S_{T_j}(t) \equiv \Pr(T_j > t) = \int_t^\infty f_{T_j}(x)dx$, $F_{T_j}(t) = 1 - S_{T_j}(t)$, respectively. Similarly, let $f_{C_j}(t)$, $S_{C_j}(t)$ and $F_{C_j}(t)$ be the pdf, SF and cdf of the censoring time C_j.

The observed data are of the form $(\tilde{T}_j, \delta_j, X_j)$, where $\tilde{T}_j = \min(T_j, C_j)$ is the censored or failure data and $\delta_j = I\{T_j \le C_j\}$ is the failure indicator. Similarly, $\tilde{T}_{ij}$ represents the ith replication in sample j. Suppose there are $k_j \le n_j$ *distinct* failure times $t_{1j} < t_{2j} < \cdots < t_{k_j j} < t_{k_j+1,j} \equiv L_j$, where L_j is the largest censored observation. Suppose that in the interval $[t_{i-1,j}, t_{i,j})$ there are c_{ij} observed censoring times C_{ij}^v $(v = 1, 2, \ldots, c_{ij})$. Then, assuming the m samples that correspond to the different stress levels are independent and can include censoring times before the first or after the last observed failure time, the observed likelihood function can be written as

$$L_0 = \prod_{j=1}^{m} \left\{ \prod_{i=1}^{k_j} \left[\left(\prod_{r=1}^{c_{ij}} S_{T_j}(C_{ij}^v) \right) \left[S_{T_j}(t_{ij}) - S_{T_j}(t_{ij}+0) \right] \right] \left[\prod_{v=1}^{c_{i,k_j+1}} S_{T_j}(C_{i,k_j+1}^v) \right] \right\}.$$

Using the same argument as in Lawless (1982, page 75), to maximize L_0 with respect to $S_{T_j}(t)$, $j = 1, 2, \ldots, m$, it is only necessary to maximize

$$L = \prod_{j=1}^{m} \left\{ \prod_{i=1}^{k_j} P_{ij} \prod_{i=1}^{k_j+1} \left(\sum_{l=i}^{k_j+1} P_{lj} \right)^{c_{ij}} \right\}, \qquad (2)$$

where, for $i = 1, 2, \ldots, k_j, \quad j = 1, 2, \ldots, m,$

$$\begin{aligned} P_{ij} &= \Pr(t_{i-1,j} < T_j \le t_{ij}) = S_{T_j}(t_{i-1,j}) - S_{T_j}(t_{ij}) \\ P_{k_j+1,j} &= \Pr(T > t_{k_j,j}) = S_T(t_{k_j,j}) = 1 - \sum_{i=1}^{k_j} P_{ij}. \end{aligned}$$

Note that, for stress level x_j, the P_{ij} $(i = 1, 2, \ldots, k_j + 1)$ define an empirical distribution with a point mass at each failure time t_{ij}, $i = 1, 2, \ldots, k_j$, $j = 1, 2, \ldots, m$.

The standard nonparametric MLE (NPMLE) of P_{ij}, $i = 1, 2, \ldots, k_j+1$ can be obtained separately for each sample by finding P_{ij}'s to maximize (2) subject to the constraints $\sum_{i=1}^{k_j+1} P_{ij} = 1$ and $P_{ij} \ge 0$, $i = 1, 2, \ldots, k_j + 1$. With the estimates of P_{ij}'s, we can obtain the estimate of $F_{T_j}(t)$ and $S_{T_j}(t)$.

The SMLE approach will maximize the empirical likelihood under the constraints of the m sets of estimating equations,

$$\mathrm{E}[\boldsymbol{G}_j(T_j, \boldsymbol{\beta})] = \sum_{i=1}^{k_j+1} P_{ij} \boldsymbol{G}_j(t_{ij}, \boldsymbol{\beta}) = 0, \quad j = 1, 2, \ldots, m, \qquad (3)$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of the regression parameters. Each set of estimating equations could include r functions,

$$\boldsymbol{G}_j(T, \boldsymbol{\beta}) = (g_{1j}(T_j, \boldsymbol{\beta}), \ldots, g_{rj}(T_j, \boldsymbol{\beta}))^\top, \quad j = 1, 2, \ldots, m.$$

Thus, there are $s = r \times m (\geq p)$ independent functions to estimate p parameters. As a simple example, a function of means $\mathrm{E}(T_j) = \psi(x_j, \boldsymbol{\beta})$ is commonly used in ALT regression, where ψ is known (usually, a linear or log-linear function). In this case, $r = 1$ and $m > p$, and

$$\boldsymbol{G}(T, \boldsymbol{\beta}) = (T_1 - \psi(x_1, \boldsymbol{\beta}), T_2 - \psi(x_2, \boldsymbol{\beta}), \ldots, T_m - \psi(x_m, \boldsymbol{\beta}))^\top.$$

Stronger assumptions on the estimating equations can be used to improve the estimation quality. For example, higher moment assumptions can be incorporated via $\boldsymbol{G}$ as

$$\boldsymbol{G}_v(T_j, T_{j'}, x_j, x_{j'}, \boldsymbol{\beta}) = (g(T_j, x_j, \boldsymbol{\beta})^v - g(T_{j'}, x_{j'}, \boldsymbol{\beta})^v)^\top, \ (j, j') \in \{1, 2, \ldots, m\},$$

where v could be any real number and $g(T_j, x_j, \boldsymbol{\beta})$ can be $T_j - \psi(x_j, \boldsymbol{\beta})$ in the traditional ALT or

$$g(t, x, \boldsymbol{\beta}) = -\log\left(\beta_1 x^{\beta_2} + \frac{\beta_3 x^{\beta_4}}{1 - \exp(-\beta_1 x^{\beta_2}(1 + \beta_3 x^{\beta_4})t)}\right) \tag{4}$$

as derived in Meeker and LuValle (1995). Note that in the traditional ALT model, time T in $T_j - \psi(x_j, \boldsymbol{\beta})$ is separated (in an additive form) from the regression part, which is not the case in (4). See Chen, Lu and Lin (2005) for more discussion of these models.

As discussed in Section 2, percentile regression is more suitable than the mean regression when heavy censoring occurs. The estimating function can then be set as

$$\boldsymbol{G}(t_1, t_2, \ldots, t_m, \boldsymbol{\beta}) = (I(t_1 < f(x_1, \boldsymbol{\beta})) - q, \ldots, I(t_m < f(x_m, \boldsymbol{\beta})) - q))^\top,$$

where $100q$ is the percentile of the lifetime.

3.2 SMLE Estimation Procedure

Let $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset I\!R^p$ be the p-dimensional parameter vector in the estimating equations, where $\boldsymbol{\Theta}$ is the parameter space containing a neighborhood of the true parameter $\boldsymbol{\theta}_0$. Given $\boldsymbol{\theta}$, maximizing the empirical likelihood L subject to the constraint (3), we can obtain the SMLE of P_{ij} in terms of $\boldsymbol{\theta}$, i.e., $\hat{P}_{ij} = \hat{P}_{ij}(\boldsymbol{\theta})$. Plugging $\hat{P}_{ij}(\boldsymbol{\theta})$ in (2), we have the profile likelihood of $\boldsymbol{\theta}$ as

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{m}\left[\prod_{i=1}^{k_j}\hat{P}_{ij}(\boldsymbol{\theta})\prod_{i=1}^{k_j+1}\left(\sum_{l=i+1}^{k_j+1}\hat{P}_{lj}(\boldsymbol{\theta})\right)^{c_{ij}}\right].$$

Maximizing $L(\boldsymbol{\theta})$ over parameter space $\boldsymbol{\Theta}$, we will obtain the SMLE $\hat{\boldsymbol{\theta}}$.

The SMLE can be found via the following Lagrange multiplier method. For given $\boldsymbol{\theta}$, let

$$H^* = \sum_{j=1}^{m} \left[\sum_{i=1}^{k_j} \log P_{ij} + \sum_{i=1}^{k_j+1} c_{ij} \log(\sum_{l=i}^{k_j+1} P_{lj}) - n_j \lambda_{0j} (\sum_{i=1}^{k_j+1} P_{ij} - 1) - n_j \boldsymbol{\lambda}_j^\top \sum_{i=1}^{k_j+1} P_{ij} G_j(t_{ij}, \boldsymbol{\theta}) \right],$$

where λ_{0j}, $\boldsymbol{\lambda}_j^\top = (\lambda_{1j}, \ldots, \lambda_{rj})$ are Lagrange multipliers.

Taking derivatives with respect to P_{ij}, λ_{0j}, and $\boldsymbol{\lambda}_j$, we have

$$\frac{\partial H^*}{\partial P_{ij}} = \frac{1}{P_{ij}} + \sum_{v=1}^{i} \frac{c_{vj}}{(\sum_{l=v}^{k_j+1} P_{lj})} - n_j \lambda_{0j} - n_j \boldsymbol{\lambda}_j^\top G_j(t_{ij}, \boldsymbol{\theta}),\ i = 1, 2, \ldots, k_j,\ j = 1, 2, \ldots, m,$$

and

$$\frac{\partial H^*}{\partial \lambda_{0j}} = n_j (1 - \sum_{i=1}^{k_j+1} P_{ij}), \quad \frac{\partial H^*}{\partial \boldsymbol{\lambda}_j} = -n_j \sum_{i=1}^{k_j+1} P_{ij} G_j(t_{ij}, \boldsymbol{\theta}), \quad j = 1, \ldots, m.$$

Equating these derivatives to zeros, we find $\lambda_{0j} = 1$ due to $\sum_{i=1}^{k_j+1} P_{ij} \partial H^* / \partial P_{ij} = 0$. Given $\boldsymbol{\theta}$, for $i = 1, 2, \ldots, k_j$, $j = 1, 2, \ldots, m$,

$$\begin{aligned} P_{ij}(\boldsymbol{\theta}) &= \frac{1}{n_j (1 - a_{ij} + \boldsymbol{\lambda}_j^\top G_j(t_{ij}, \boldsymbol{\theta}))}, \qquad (5) \\ P_{k_j+1,j}(\boldsymbol{\theta}) &= 1 - \sum_{i=1}^{k_j} P_{i,j}(\boldsymbol{\theta}), \end{aligned}$$

where

$$a_{ij} = \frac{1}{n_j} \sum_{v=1}^{i} \frac{c_{vj}}{\sum_{l=v}^{k_j+1} P_{lj}(\boldsymbol{\theta})}, \qquad (6)$$

and $\boldsymbol{\lambda}_j$ is the solution to

$$\sum_{i=1}^{k_j+1} \frac{G_j(t_{ij}, \boldsymbol{\theta})}{1 - a_{ij} + \boldsymbol{\lambda}_j^\top G_j(t_{ij}, \boldsymbol{\theta})} = 0, \quad j = 1, 2, \ldots, m. \qquad (7)$$

In the case that there is only one sample and no censoring is observed, i.e., $a_{ij} = 0$, and $m = 1$, equations (5) and (7) reduce to

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

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

Under the same moderate conditions specified in Lemma 3.1 of Chen, Lu and Lin (2005), it is easy to show that, given $\boldsymbol{\theta}$, the implicit function $\boldsymbol{\lambda}_j(\boldsymbol{\theta})$ as the solution of (7) exists uniquely. We state the result in the following lemma, and omit the proof here.

Lemma 3.1 For a given $\boldsymbol{t}_j = (t_{1j}, t_{2j}, \ldots, t_{k_j,j}, t_{k_j+1,j})$, let $\boldsymbol{G}_j(\boldsymbol{t}, \boldsymbol{\theta}) = (G_j(t_{ij}, \boldsymbol{\theta}))_{(k_j+1)\times r}$. For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, assume that the $r \times r$ matrix $\boldsymbol{G}_j(\boldsymbol{t}, \boldsymbol{\theta})^\top \boldsymbol{G}_j(\boldsymbol{t}, \boldsymbol{\theta})$ is nonsingular. Then, the solution of the implicit functions $\boldsymbol{\lambda}_j(\boldsymbol{\theta})$ from equations (5) - (7) exists uniquely.

It follows from Lemma 3.1 that, for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the log-profile-likelihood

$$H(\boldsymbol{\theta}) = \sum_{j=1}^{m} \left\{ \sum_{i=1}^{k_j} \log P_{ij}(\boldsymbol{\theta}) + \sum_{i=1}^{k_j+1} c_{ij} \log \left(\sum_{l=i}^{k_j+1} P_{lj}(\boldsymbol{\theta}) \right) \right\}. \tag{8}$$

is well defined. The likelihood equation $\partial H(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = 0$ can be simplified to

$$\frac{\partial H(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\sum_{j=1}^{m} \left[n_j \boldsymbol{\lambda}_j^\top(\boldsymbol{\theta}) \sum_{i=1}^{k_j} \frac{\partial G_j(t_{ij}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} P_{ij}(\boldsymbol{\theta}) \right] = 0. \tag{9}$$

Denote $\hat{\boldsymbol{\theta}}$ the solution of equation (9) as the SMLE for the parameter $\boldsymbol{\theta}$, and the corresponding SMLE for the survival function $S_{T_j}(t)$ can be written as

$$\hat{S}_{T_j}(t) = \sum_{t_{ij}>t} \frac{1}{n_j(1 - a_{ij} + \boldsymbol{\lambda}_j^\top G_j(t_{ij}, \hat{\boldsymbol{\theta}}))}, \tag{10}$$

where a_{ij} and $\boldsymbol{\lambda}_j$ are defined by (5) - (7) with respect to $\hat{\boldsymbol{\theta}}$.

3.3 SMLE and Kaplan-Meier Estimate

From equations (5) - (7), it is clear that $\boldsymbol{\lambda}_j, \ldots, \boldsymbol{\lambda}_m$ play an important role in the SMLE. If we set $\boldsymbol{\lambda}_j = 0$ $(j = 1, 2, \ldots, m)$, i.e., no constraint is imposed on the empirical likelihood, the SMLE reduces to the NPMLE. Then, the estimate of the survival function (10) reduces to

$$\hat{S}_{T_j,0}(t) = \sum_{t_{ij}>t} \frac{1}{n_j(1 - a_{ij,0})} \tag{11}$$

where "0" is placed in the subscript to emphasize $\boldsymbol{\lambda}_j = 0$, and

$$a_{ij,0} = \frac{1}{n} \sum_{v=1}^{i} \frac{c_{vj}}{\sum_{l=v}^{k_j+1} P_{lj,0}}, \;\; i = 2, \cdots, k_j + 1, \quad a_{n1,0} = 0, \tag{12}$$

$$P_{ij,0} = \frac{1}{n_j(1 - a_{ij,0})}, \quad i = 1, 2, \ldots, k_j, \;\; P_{(k_j+1)j,0} = 1 - \sum_{i=1}^{k_j} P_{ij,0}.$$

When additional information about the lifetime distribution regarding to parameter $\boldsymbol{\theta}$ can be provided by the estimating equations, the SMLE will incorporate it through $\boldsymbol{\theta}$ via $\boldsymbol{\lambda}_j$ and $\boldsymbol{G}(t, \boldsymbol{\theta})$.

The following show that $\hat{S}_{T_j,0}(t)$ is the well-known Kaplan-Meier estimate (Kaplan and Meier, 1958). Denote by $W_{ij} = 1 - \sum_{v=1}^{i} P_{vj}$, $\beta_{ij} = W_{ij}/W_{(i-1)j}$, $i = 1, 2, \ldots, k_j + 1$, $j = 1, 2 \ldots, m$, so that $P_{ij} = W_{ij} - W_{(i-1)j} = (1-\beta_{ij})\prod_{v=1}^{i-1}\beta_{vj}$, and $1-\sum_{v=1}^{i} P_{vj} = \prod_{v=1}^{i}\beta_{vj}$. With this notation, it follows from (11) and (12) that for $i = 2, 3\ldots, k_j,\quad j = 1, 2\ldots, m$,

$$\begin{aligned} 1-\beta_{ij} &= \frac{1}{\prod_{v=1}^{i-1}\beta_{vj}(n_j - \sum_{u=1}^{i} c_{uj}/\prod_{s=1}^{v}\beta_{sj})}, \\ 1-\beta_{1j} &= P_{1j} = \frac{1}{n-c_{1j}}. \end{aligned} \tag{13}$$

Note that, from (13), β_{ij} can be expressed in terms of $\beta_{1j}, \beta_{2j}, \ldots, \beta_{(i-1)j}$, and $\beta_{1j} = (n-1-c_{1j})/(n-c_{1j})$. After further simplification,

$$\beta_{ij} = \frac{n_j - \sum_{v=1}^{i}(c_{vj}+1)}{n_j - \sum_{v=1}^{i-1}(c_{vj}+1) - c_{ij}}, \quad 2, 3, \ldots, k_j, \quad j = 1, 2\ldots, m.$$

Denote by $n_{ij} = n_j - \sum_{v=1}^{i-1}(c_{vj}+1) - c_{ij}$ the number of subjects at risk at time t_{ij} and let $n_{1j} = n_j$. Then β_{ij} can be written as $(n_{ij}-1)/n_{ij}$, and $\hat{S}_{T_j,0}(t)$ can be expressed as a product limit estimator:

$$\hat{S}_{T_j,0}(t) = \sum_{t_{ij}>t} P_{ij} = \prod_{t_{ij}\le t}\beta_{ij} = \prod_{t_{ij}\le t}\frac{(n_{ij}-1)}{n_{ij}} = \hat{S}_{T_j,KM}(t). \tag{14}$$

Consider $a_{ij,0}$ defined in (12). It follows from (14) that

$$a_{ij,0} = \sum_{v=1}^{i-1}\frac{c_{vj}}{n_j\hat{S}_{T_j,KM}(t_{vj})} = \sum_{v=1}^{i-1}\frac{c_{vj}}{n_j\prod_{l=1}^{v}\frac{(n_{lj}-1)}{n_{lj}}}, \quad i = 2, 3, \ldots, k+1.$$

With the same notation used before, let $W_{vj} = 1 - a_{vj,0}$, $\beta_{vj} = W_{vj}/W_{(v-1)j}$, $i = 2, 3, \ldots, k+1$. Then,

$$Q_{vj} = W_{vj} - W_{v-1,j} = \frac{c_{v-1,j}}{n_j\prod_{l=1}^{v-1}\frac{(n_{lj}-1)}{n_{lj}}} = (1-\beta_{vj})\prod_{l=1}^{(v-1)j}\beta_{lj},$$

and $1 - a_{ij,0} = 1 - \sum_{v=1}^{i-1} Q_{vj} = \prod_{v=1}^{i-1}\beta_{vj}$. It follows that

$$1-\beta_{ij} = \frac{c_{i-1,j}}{n_j\prod_{v=1}^{i-1}\frac{(n_{vj}-1)}{n_{vj}}}\prod_{v=1}^{i-1}\beta_{vj}, \quad i = 3, 4, \ldots, k+1, \quad \text{and} \quad 1-\beta_{2j} = \frac{c_{1j}}{n_{1j}-c_{1j}},$$

where $n_{ij} = n_j - \sum_{v=1}^{i-1}(c_{vj}+1)$, for $i > 1$, the number of subjects at risk at time point t_{ij} and $n_{1j} = n_j$, then the β_{ij} can be written as $(n_{ij}-c_{ij})/n_{ij}$. Therefore,

$$1 - a_{ij,0} = \sum_{v=i}^{k+1} Q_{vj} = \prod_{v=1}^{i-1}\beta_{vj} = \prod_{v=1}^{i-1}\frac{(n_{ij}-c_{vj})}{n_{ij}} = \hat{S}_{nC_j,KM}(t_{ij}), \tag{15}$$

which is the Kaplan-Meier estimate of the survival function $S_{C_j}(t)$ at time t_{ij}. Eqs(14) and (15) will play critical roles in deriving the asymptotic distribution of SMLE.

4 Asymptotic Properties of the SMLE

4.1 Regularity Condition and Consistency of the SMLE

Let $n_{max} = max_j\{n_j\}$, $n_{min} = min_j\{n_j\}$. The following regularity conditions are necessary in discussing the properties of the SMLE $\hat{\boldsymbol{\theta}}_n$ and $\hat{S}_{nT_j}(t)$, where the subscript $n = \sum_{j=1}^m n_j$ represents that the estimators are based on empirical data.

(**R.1**) Parameter space $\boldsymbol{\Theta} \subset R^p$ is compact and contains a neighborhood of true parameter $\boldsymbol{\theta}_0$, and, for $H(\boldsymbol{\theta})$ given in (8), $|R| < \infty$, where

$$|R| = \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \{0 < |H(\boldsymbol{\theta})|\}.$$

(**R.2**) Given $\boldsymbol{t}_j = (t_{1j}, t_{2j}, \ldots, t_{k_j,j}, t_{k_j+1,j})$, let $\boldsymbol{G}_j(\boldsymbol{t}_j, \boldsymbol{\theta}) = (G_j(t_{ij}, \boldsymbol{\theta}))_{(k_j+1)\times r}$. For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, assume that $r \times r$ matrix $\boldsymbol{G}_j^\top \boldsymbol{G}_j$, $j = 1, 2, \ldots, m$, is nonsingular.

(**R.3**) For $j = 1, 2, \ldots, m$, $\mathrm{E}(||G_j(T, \boldsymbol{\theta})||^3) < \infty$ and $G_j(T, \boldsymbol{\theta})$ is second-order differentiable with respect to $\boldsymbol{\theta}$, i.e., $\partial^2 G_j(T, \boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ exists for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The regularity condition **R1** is to ensure that the maximum of $|H(\boldsymbol{\theta})|$ exists in the interior of $\boldsymbol{\Theta}$. **R2** and **R3** require the non-singularity, continuity and differentiability of estimating function $G_j(t, \boldsymbol{\theta})$ to ensure that equation (9) is well defined and the SMLE $\hat{\boldsymbol{\theta}}$ is in $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < n_{max}^{-1/3}$ with probability one, given that n_{min} is sufficiently large.

In the investigation of the asymptotic properties of SMLE, we start with understanding the large sample properties of $\boldsymbol{\lambda}_j(\hat{\boldsymbol{\theta}}_n)$. In fact, similar to the results of Qin and Lawless (1994, Lemma 1) in the complete sample case, we have the following conclusions for the censored data.

Theorem 4.1 *Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ be the true value of the parameter. Under the regularity conditions* **R1-R3**, *we have the following results.*

(1) *There exist $\hat{\boldsymbol{\theta}}_n \in \boldsymbol{\Theta}$ and $\boldsymbol{\lambda}_j = \boldsymbol{\lambda}_j(\hat{\boldsymbol{\theta}}_n)$ satisfying equations (9) and (7).*

(2) *For $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \le n_{max}^{-1/3} \le n_j^{-1/3}\}$ and $\boldsymbol{\lambda}_j(\boldsymbol{\theta})$ satisfying (7), we have $\boldsymbol{\lambda}_j(\boldsymbol{\theta}) \stackrel{w.p.1}{\longrightarrow} 0$, $j = 1, 2, \ldots m$, and $\boldsymbol{\lambda}_j(\boldsymbol{\theta}) = O_p(n_j^{-1/2})$, as $n_{min} \to \infty$.*

(3) *When n_{min} is large, $H(\boldsymbol{\theta})$ attains its maximum value at some point $\hat{\boldsymbol{\theta}}_n$ in the interior of the ball $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \le n_{max}^{-1/3}$ with probability one. Thus, the SMLE $\hat{\boldsymbol{\theta}}$ is a strongly consistent estimate of $\boldsymbol{\theta}$.* □

Theorem 4.1 can be proved by following similar procedures used by Qin and Lawless (1994) and Owen (1990) and thus is omitted here. According to Theorem 4.1, it follows form (14) that, for $t > 0$,

$$\lim_{n_j\to\infty} \hat{S}_{nT_j}(t) = \lim_{n_j\to\infty} \sum_{t_{ij}\geq t} \frac{1}{n_j(1 - a_{ij} + \boldsymbol{\lambda}_j^\top(\hat{\boldsymbol{\theta}}_n)G_j(t_{ij}, \hat{\boldsymbol{\theta}}_n))} = \lim_{n_j\to\infty} \hat{S}_{nT_j,0}(t),$$

where $\hat{S}_{nT_j,0}(t) = \hat{S}_{T_j,KM}(t)$ is the Kaplan-Meier estimate of survival function for T_j. This leads to the following result for $\hat{S}_{nT_j}$.

Theorem 4.2 *For continuous lifetime T_j and censoring time C_j, suppose $S_{C_j}(L_j) > 0$, and $S_{T_j}(t)$ is continuous at $t = L_j$. Then, as $n_j \to \infty$,*

$$\sup_{0\leq t\leq L_j} |\hat{S}_{nT_j}(t) - S_{T_j}(t)| \xrightarrow{p} 0, \quad j = 1, 2, \ldots, m.$$

The proof of the Theorem follows by noting that

$$|\hat{S}_{nT_j}(t) - S_{T_j}(t)| \leq |\hat{S}_{nT_j}(t) - \hat{S}_{nT_j,KM}(t)| + |\hat{S}_{nT_j,KM}(t) - S_{T_j}(t)|.$$

It follows from Wang (1987, Corollary 1) that the second part converges to zero uniformly in probability. Using the fact that both $\hat{S}_{nT_j}(t)$ and $\hat{S}_{nT_j,KM}(t)$ are bounded by one, and that $\hat{S}_{nT_j}(t)$ is continuous in λ, (i.e., for any $\epsilon > 0$, there exists $\delta > 0$ such that $\sup_{0\leq t\leq L_j} |\hat{S}_{nT_j}(t) - \hat{S}_{nT_j,KM}(t)| < \epsilon$ as long as $\|\boldsymbol{\lambda}\| \leq \delta$), the first part goes to zero uniformly for $t \in [0, L_j]$, as $n_j \to \infty$. □

4.2 Asymptotic Distribution of $\boldsymbol{\lambda}_j$

Denote

$$h(\boldsymbol{\lambda}) = \sum_{i=1}^{k_j+1} \frac{G_j(t_{ij}, \boldsymbol{\theta})}{n_j(1 - a_{ij} + \boldsymbol{\lambda}^\top G_j(t_{ij}, \boldsymbol{\theta}))},$$

and suppose $\boldsymbol{\lambda}_j$ satisfies equation $h(\boldsymbol{\lambda}_j) = 0$. For $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, because $\boldsymbol{\lambda}_j = O_p(n_j^{-1/2})$, according to Theorem 4.1, the Taylor expansion of $h(\boldsymbol{\lambda}_j)$ at $\boldsymbol{\lambda} = 0$ yields

$$\begin{aligned} 0 &= h(\boldsymbol{\lambda}_j) = h(0) + \left[\frac{\partial h(0)}{\partial \boldsymbol{\lambda}_j}\right]^\top \boldsymbol{\lambda}_j + o_p(n_j^{-1/2}) \\ &= \sum_{i=1}^{k_j+1} \frac{G_j(t_{ij}, \boldsymbol{\theta})}{n_j(1 - a_{ij})} + \sum_{i=1}^{k_j+1} \frac{G_j(t_{ij}, \boldsymbol{\theta})G^\top(t_{ij}, \boldsymbol{\theta})}{n_j(1 - a_{ij})^2}\boldsymbol{\lambda}_j + o_p(n_j^{-1/2}). \end{aligned}$$

It follows that

$$\boldsymbol{\lambda}_j = -\left[\sum_{i=1}^{k_j+1} \frac{G_j(t_{ij},\boldsymbol{\theta})G_j^\top(t_{ij},\boldsymbol{\theta})}{n_j(1-a_{ij})^2}\right]^{-1} \left[\sum_{i=1}^{k_j+1} \frac{G_j(t_{ij},\boldsymbol{\theta})}{n_j(1-a_{ij})}\right] + o_p(n_j^{-1/2}). \tag{16}$$

According to (15) and Theorem 4.1, we have, with probability one,

$$\lim_{n_j\to\infty}(1-a_{ij}) = \lim_{n_j\to\infty}(1-a_{ij,0}) = \lim_{n_j\to\infty}\hat{S}_{nC_j,KM}(t_{ij}).$$

Note that, under the condition that T_j and C_j will never go beyond L_j, it follows from the uniform consistency of the Kaplan-Meier estimate that, on the interval $[0, L_j]$

$$\lim_{n_j\to\infty}\sum_{i=1}^{k_j+1} \frac{G_j(t_{ij},\boldsymbol{\theta})G_j^\top(t_{ij},\boldsymbol{\theta})}{n_j(1-a_{ij})^2} = \lim_{n_j\to\infty}\int_0^\infty \frac{G_j(t,\boldsymbol{\theta})G_j^\top(t,\boldsymbol{\theta})}{\hat{S}_{C_j,KM}(t)}d\hat{F}_{T_j}(t)$$

$$= \int_0^\infty \frac{G_j(t,\boldsymbol{\theta})G_j^\top(t,\boldsymbol{\theta})}{S_{C_j}(t)}dF_{T_j}(t) \equiv \boldsymbol{A}_j(\boldsymbol{\theta}), \tag{17}$$

where the distribution $F_{T_j}(t)$ considered has been truncated at L_j. It follows that

$$\boldsymbol{\lambda}_{j,0} = -A_j(\boldsymbol{\theta})^{-1}\left[\int_0^\infty G_j(t,\boldsymbol{\theta})d\hat{F}_{T_j}(t)\right] \tag{18}$$

and $\boldsymbol{\lambda}_j$ have the same asymptotic distribution, where $\hat{F}_{T_j}(t) = 1 - \hat{S}_{T_j,KM}(t)$. We use the following Theorem to state the asymptotic normality of $\boldsymbol{\lambda}_j$.

Theorem 4.3 *For continuous lifetime T_j and censoring time C_j, suppose $S_{C_j}(L) > 0$, and $S_{T_j}(t)$ is continuous at $t = L_j$. Then, as $n_{min} \to \infty$, if $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \leq n_{max}^{-1/3}\}$,*

$$\sqrt{n_j}\boldsymbol{\lambda}_j(\boldsymbol{\theta}) \xrightarrow{d} N_r(0, \boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta})), \quad j = 1, 2, \ldots m,$$

where

$$\boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta}) = \boldsymbol{A}_j(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}_{G_j}(\boldsymbol{\theta})\boldsymbol{A}_j(\boldsymbol{\theta})^{-1}, \tag{19}$$

$$\boldsymbol{\Sigma}_{G_j}(\boldsymbol{\theta}) = \int_0^\infty \left\{G_j(t,\boldsymbol{\theta})[1-F_{T_j}(t)] - \int_0^t G_j(s,\boldsymbol{\theta})dF_{T_j}(t)\right\}^2 \frac{dF_{T_j}(t)}{[1-F_{T_j}(t)]^2[1-F_{C_j}(t)]}, \tag{20}$$

and $\boldsymbol{A}_j(\boldsymbol{\theta})$ is given in (17).

Proof: It is sufficient to derive the asymptotic normality of $\boldsymbol{\lambda}_{j,0}$ defined in (18). Because $\mathrm{E}[G_j(T_j,\boldsymbol{\theta})] = \int_0^\infty G_j(t,\boldsymbol{\theta})dF_{T_j}(t) = 0$, we know

$$\int_0^\infty G_j(t,\boldsymbol{\theta})d\hat{F}_{nT_j}(t) = \int_0^\infty G_j(t,\boldsymbol{\theta})d(\hat{F}_{nT_j}(t) - F_{T_j}(t)) = -\int_0^\infty G_j(t,\boldsymbol{\theta})d(\hat{S}_{nT_j}(t) - S_{T_j}(t)).$$

Using integration by parts, it follows that

$$\int_0^\infty G_j(t,\boldsymbol{\theta})d\hat{F}_{nT_j}(t) = \int_0^\infty (\hat{S}_{nT_j}(t) - S_{T_j}(t))dG_j(t,\boldsymbol{\theta}). \tag{21}$$

According to Breslow and Crowley (1974, Theorem 5), $\sqrt{n}(\hat{S}_{nT_j}(t) - S_{T_j}(t))$ converges to a Gaussian process $Z_j(t)$, with $\mathrm{E}(Z_j(t)) = 0$ and

$$\mathrm{Cov}(Z_j(s), Z_j(t)) = S_{T_j}(s)S_{T_j}(t)\int_0^{\min(t,s)} \frac{dF_{T_j}(x)}{(S_{T_j}(x)^2 S_{C_j}(x))}. \tag{22}$$

It follows from (21) that (under condition **R3**) as $n_j \to \infty$,

$$\sqrt{n_j}\int_0^\infty G_j(t,\boldsymbol{\theta})d\hat{F}_{nT_j}(t) \xrightarrow{p} \int_0^\infty Z_j(t)dG_j(t,\boldsymbol{\theta}).$$

Using Gaussian process properties (see Appendix Lemma A.1 for details), we can obtain that $\int_0^\infty Z_j(t)\ dG_j(t,\boldsymbol{\theta})$. is normal with mean zero and covariance matrix $\boldsymbol{\Sigma}_{G_j}(\boldsymbol{\theta})$ defined in (20) (see Appendix Corollary A.1). Thus, $\sqrt{n_j}\boldsymbol{\lambda}_j$ is asymptotic normal with mean zero and covariance matrix

$$\boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta}) = \boldsymbol{A}_j(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}_{G_j}(\boldsymbol{\theta})\boldsymbol{A}_j(\boldsymbol{\theta})^{-1},$$

where $\boldsymbol{A}_j(\boldsymbol{\theta})$ is given by (17). This completes the proof. □

Considering (9), denote

$$\boldsymbol{l}(\boldsymbol{\theta}) = \frac{1}{n}\frac{\partial H(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^m \left\{ \frac{n_j}{n}\boldsymbol{\lambda}_j^\top(\boldsymbol{\theta}) \sum_{i=1}^{k_j+1} P_{ij}(\boldsymbol{\theta})\frac{\partial G_j(t_{ij},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\}$$

the partial derivative of the profile likelihood, where $n = \sum_{j=1}^m n_j$. Note that, as $n_{min} \to \infty$,

$$\sum_{i=1}^{k_j+1} P_{ij}(\boldsymbol{\theta})\frac{\partial G_j(t_{ij},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \xrightarrow{p} \int_0^\infty \frac{\partial G_j(t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dF_{T_j}(t) = \mathrm{E}\left[\frac{\partial G_j(t_{ij},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right].$$

From the independence of the m samples, the asymptotic normality of $\boldsymbol{l}(\boldsymbol{\theta})$ follows directly from Theorem 4.3. We state it as the following corollary.

Corollary 4.1 *Let $n = \sum_{j=1}^m n_j$, and assume $n_j/n \to \gamma_j$ as $n_{min} \to \infty$, where $0 \le \gamma_j \le 1$. Under the conditions of Theorem 1, for given $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \le n_{max}^{1/3}\}$, then $\sqrt{n}\boldsymbol{l}_n(\boldsymbol{\theta})$ is asymptotically normal with mean zero, and covariance matrix*

$$\boldsymbol{\Sigma}_l(\boldsymbol{\theta}) = \sum_{j=1}^m \sqrt{\gamma_j} E\left[\frac{\partial G_j(T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] \boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta}) E\left[\frac{\partial G_j(T,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^\top, \tag{23}$$

where $\boldsymbol{\Sigma}_{\lambda_j}(\boldsymbol{\theta})$ is given by (19).

4.3 Asymptotic Normality of the SMLE of Model Parameters

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

$$0 = \boldsymbol{l}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{l}(\boldsymbol{\theta}) - \boldsymbol{l}(\hat{\boldsymbol{\theta}}_n)) - \frac{\partial \boldsymbol{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) + o_p(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|). \tag{24}$$

Because

$$\frac{\partial \boldsymbol{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{m} \left\{ \frac{n_j}{n} \boldsymbol{\lambda}_j^\top(\boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\theta}} \left(\sum_{i=1}^{k_j+1} P_{ij}(\boldsymbol{\theta}) \frac{\partial G(t_{ij}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) + \frac{n_j}{n} \frac{\partial \boldsymbol{\lambda}_j^\top(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left(\sum_{i=1}^{k_j+1} P_{ij}(\boldsymbol{\theta}) \frac{\partial G_j(t_{ij}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \right\},$$

and we know that, as $n \to \infty$, the first term of right side in the above equation goes to zero in probability (since $\boldsymbol{\lambda}_j(\boldsymbol{\theta}) \xrightarrow{w.p.1} 0$), and then

$$\lim_{n\to\infty} \frac{\partial \boldsymbol{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \lim_{n\to\infty} \sum_{j=1}^{m} \frac{n_j}{n} \frac{\partial \boldsymbol{\lambda}_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathrm{E}\left[\frac{\partial G_j(T_j, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right].$$

It follows from (16) that

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

Therefore,

$$\lim_{n\to\infty} \frac{\partial \boldsymbol{l}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{m} \left\{ \gamma_j \mathrm{E}\left[\frac{\partial G_j(T, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^\top \boldsymbol{A}_j(\boldsymbol{\theta})^{-1} \mathrm{E}\left[\frac{\partial G_j(T, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \right\},$$

where $n_j/n \to \gamma_j$. Applying Corollary 4.1 and (24), we have the following theorem.

Theorem 4.4 *Under the conditions of Theorem 4.1, suppose $n_j/n \to \gamma_j$ as $n \to \infty$, where $0 \le \gamma_j \le 1$. Then, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_p(0, \boldsymbol{\Sigma}_\theta)$, where*

$$\boldsymbol{\Sigma}_\theta = \boldsymbol{B}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{\Sigma}_l(\boldsymbol{\theta}_0) \boldsymbol{B}(\boldsymbol{\theta}_0)^{-1},$$

$$\boldsymbol{B}(\boldsymbol{\theta}_0) = \sum_{j=1}^{m} \left\{ \gamma_j E\left[\frac{\partial G_j(T, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right]^\top \boldsymbol{A}_j(\boldsymbol{\theta}_0)^{-1} E\left[\frac{\partial G_j(T, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right] \right\},$$

and $\boldsymbol{A}_j(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}_l(\boldsymbol{\theta}_0)$ are given by (17) and (23), respectively.

Next, we will benchmark these results against some well-known results in the literature. Consider the one-sample case where $m = 1$. If θ is the population mean, the estimating function is $G(t, \theta) =$

$t-\theta$. Because $\partial G/\partial\theta = 1$, $\boldsymbol{B}(\theta) = \boldsymbol{A}(\theta)^{-1}$ 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, i.e., $S_C(t) = 1$, (22) reduces to $\mathrm{Cov}(Z(x), Z(t)) = S_T(t)(1-S_T(x))$. Then,

$$\begin{aligned}&\mathrm{Var}\left(\int_0^\infty Z(t)dG(t,\boldsymbol{\theta})\right) \\ =& \int_0^\infty \left[\int_0^\infty \mathrm{E}(Z(t)Z(x))dG(x,\boldsymbol{\theta})\right] dG^\top(t,\boldsymbol{\theta}) \\ =& \int_0^\infty \left[\int_0^t S_T(t)(1-S_T(x))dG(x,\boldsymbol{\theta}) + \int_t^\infty S_T(x)(1-S_T(t))dG(x,\boldsymbol{\theta})\right] dG^\top(t,\boldsymbol{\theta}) \\ =& \int_0^\infty \left[\int_0^t S_T(t)dG(x,\boldsymbol{\theta}) + \int_t^\infty S_T(x)dG(x,\boldsymbol{\theta}) - \int_0^\infty S_T(x)S_T(t)dG(x,\boldsymbol{\theta})\right] dG^\top(t,\boldsymbol{\theta}) \\ =& \int_0^\infty \left[\int_0^t (S_T(t)-S_T(x))dG(x,\boldsymbol{\theta}) + (1-S_T(x))\int_0^\infty S_T(x)dG(x,\boldsymbol{\theta})\right] dG^\top(t,\boldsymbol{\theta}). \qquad (25)\end{aligned}$$

Using integration by parts and the fact that

$$\int_0^\infty G(x,\boldsymbol{\theta}))dS_T(x) = -\int_0^\infty G(x,\boldsymbol{\theta}))dF_T(x) = \mathrm{E}[G(T,\boldsymbol{\theta})] = 0,$$

we have

$$\int_0^\infty S_T(x)dG(x\boldsymbol{\theta}) = -G(0,\boldsymbol{\theta}). \qquad (26)$$

Again, using integration by parts, we know that

$$\int_0^t (S_T(t)-S_T(x))dG(x,\boldsymbol{\theta}) = (1-S_T(t))G(0,\boldsymbol{\theta}) - \int_0^t G(x,\boldsymbol{\theta})dF_T(x). \qquad (27)$$

Plugging (26) and (27) into (25) and using integration by parts repeatedly, one can obtain that

$$\begin{aligned}\boldsymbol{\Sigma}_G(\boldsymbol{\theta}) &= \mathrm{Var}\left(\int_0^\infty Z(t)dG(t,\boldsymbol{\theta})\right) = \int_0^\infty \left[\int_0^t G(x,\boldsymbol{\theta})dF_T(x)\right] dG^\top(t,\boldsymbol{\theta}) \\ &= \int_0^\infty G(t,\boldsymbol{\theta})G^\top(t,\boldsymbol{\theta})dF_T(t) = \mathrm{E}[G(T,\boldsymbol{\theta})G^\top(T,\boldsymbol{\theta})].\end{aligned}$$

Note that $\boldsymbol{A}(\boldsymbol{\theta}) = \mathrm{E}[G(T,\boldsymbol{\theta})G^\top(T,\boldsymbol{\theta})] = \boldsymbol{\Sigma}_G(\boldsymbol{\theta})$, and it follows that the asymptotic covariance matrix reduces to

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

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

The following theorem shows that in terms of the likelihood ratio, the SMLE and the parametric MLE are similar.

Theorem 4.5 *Under the regularity conditions* **R.1 - R.3**,

(1) *When* $H_{10}: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *is true,* $\Lambda_{1n}(\boldsymbol{\theta}_0) = 2(H(\hat{\boldsymbol{\theta}}_n) - H(\boldsymbol{\theta}_0)) \xrightarrow{d} \chi^2(p)$*, where* p *is the number of the parameters in* $\boldsymbol{\theta}$*, and* H *is the log-likelihood defined in (8).*

(2) *Let* $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$*, where* $\boldsymbol{\theta}_1$ *and* $\boldsymbol{\theta}_2$ *are* $q \times 1$ *and* $(p-q) \times 1$ *vectors, respectively. For* H_{20} : $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$*, the profile empirical likelihood ratio test (PELRT) statistic is*

$$\Lambda_{2n}(\boldsymbol{\theta}_{10}) = 2(H(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) - H(\boldsymbol{\theta}_{10}, \tilde{\boldsymbol{\theta}}_2)),$$

where $\tilde{\boldsymbol{\theta}}_2$ *maximizes* $H(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_2)$ *with respect to* $\boldsymbol{\theta}_2$*. Under* H_{20}*,* $\Lambda_{2n} \xrightarrow{d} \chi^2(q)$*, as* $n \to \infty$*.*

The above Theorem can be proved following similar procedures used in Qin and Lawless (1994).

Part (2) of Theorem 4.5 can be used to test whether the assumptions $\mathrm{E}(G(T, \boldsymbol{\theta})) = 0$ are adequate, where $G = \{G_1, G_2, \ldots, G_m\}$. For testing this hypothesis, the PELRT statistic is $\Lambda_{3n}(\boldsymbol{\theta}) = 2(H^* - H(\hat{\boldsymbol{\theta}}_n))$, where H^* is the maximum of the empirical log-likelihood in the NPMLE approach without any constraint (given by the Kaplan-Meier estimate), and $H(\hat{\boldsymbol{\theta}}_n)$ is the log-likelihood function with $\hat{\boldsymbol{\theta}}_n$ as the SMLE estimates.

To derive the asymptotic distribution of $\Lambda_{3n}(\boldsymbol{\theta})$, we first recall that $G(t, \boldsymbol{\theta}) = (g_1(t, \boldsymbol{\theta}), g_2(t, \boldsymbol{\theta}), \ldots, g_s(t, \boldsymbol{\theta}))$, where $s = r \times m$; and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p)$. Note that any p of s $(s > p)$ equations $\mathrm{E}(g_i(T, \boldsymbol{\theta})) = 0$, $i = 1, 2, \ldots, s$, can define parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p)$. For convenience, let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p)$ be defined by the first p equations, i.e., $\mathrm{E}(g_i(T, \boldsymbol{\theta})) = 0$, $i = 1, 2, \ldots, p$. Denote $\beta_j = \mathrm{E}(g_j(T, \boldsymbol{\theta}))$, $j = p+1, p+2, \ldots, s$, and $\tau = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p, \beta_1, \beta_2, \ldots, \beta_{s-p})$. Then, the r-dimensional parameter vector τ can be determined by the estimating equations $\mathrm{E}(G^*(T, \delta)) = 0$, where $G^*(t, \delta) = (g_1^*(t, \delta), g_2^*(t, \delta), \ldots, g_s^*(t, \delta))$, and $g_i^*(t, \delta) = g_i(t, \boldsymbol{\theta})$, $i = 1, 2, \ldots, p$, $g_i^*(t, \delta) = g_i(t, \boldsymbol{\theta}) - \beta_{i-p}$, $i = p+1, p+2, \ldots, s$.

With this re-parameterization, testing the model $H_{20}: \mathrm{E}(G(T, \boldsymbol{\theta})) = 0$ is equivalent to testing $H_{20}^*: \boldsymbol{\beta} = 0$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_{s-p})$. The PELRT statistic in this case becomes

$$\tilde{\Lambda}_{2n} = 2(H(\hat{\tau}) - H(\hat{\theta}, 0)).$$

In the case of $p = s$, the estimating equations $\mathrm{E}(G^*(T, \tau)) = 0$ make no constraint on τ, and $H(\hat{\tau})$ is equal to H^*, the maximum of the empirical likelihood. With $H_n(\hat{\theta}_n, 0) = H(\hat{\theta}_n)$, the maximum likelihood in the SMLE approach, we have

$$\tilde{\Lambda}_{3n} = 2(H(\hat{\tau}) - H(\hat{\theta}_n, 0)) = 2(H^* - H(\hat{\theta}_n)) = \Lambda_{2n}.$$

This leads to the following corollary.

Corollary 4.2 *If* $\mathrm{E}(G_j(T, \boldsymbol{\theta})) = 0$, *then under the regularity conditions* **R.1 - R.3**, *we have* $\Lambda_{3n}(\boldsymbol{\theta}) = 2(H^* - H(\hat{\boldsymbol{\theta}}_n)) \xrightarrow{d} \chi^2(s - p)$, *where* s *is the number of total independent functions specified in* $G_j(T, \boldsymbol{\theta})$, $j = 1, 2, \ldots, m$, *and* p *is dimension of vector* $\boldsymbol{\theta}$.

4.4 Asymptotics of the SMLE of the Survival Function

Because $\boldsymbol{\lambda}_j(\boldsymbol{\theta}) = O_p(n_j^{-1/2})$, the Taylor expansion of $\hat{S}_{nT_j}(t)$ at $\boldsymbol{\lambda}_j = 0$ results in

$$\begin{aligned} \hat{S}_{nT_j}(t) &= \sum_{t_{ij} > t} \left(\frac{1}{n(1 - a_{ij})} + \frac{G_j^\top(t_{ij}, \hat{\boldsymbol{\theta}}_n) \boldsymbol{\lambda}_j(\hat{\boldsymbol{\theta}}_n)}{n(1 - a_{ij})^2} + o_p(n_j^{-1/2}) \right) \\ &= \hat{S}_{nT_j,0}(t) + \int_t^\infty \frac{G_j^\top(x, \hat{\boldsymbol{\theta}}_n) d\hat{F}_{nT_j,0}(x)}{\hat{S}_{nC_j,0}(x))} \boldsymbol{\lambda}_n(\hat{\boldsymbol{\theta}}_n) + o_p(n_j^{-1/2}), \end{aligned}$$

where $\hat{F}_{nT_j,0}(t) = 1 - \hat{S}_{nT_j,0}(t)$, and $\hat{S}_{nT_j,0}(t)$ and $\hat{S}_{nC_j,0}(t)$ are the Kaplan-Meier estimates of $S_{T_j}(t)$ and $S_{C_j}(t)$.

According to (16), we know that

$$\hat{S}_{nT_j}(t) = \hat{S}_{nT_j,0}(t) + \left(\int_t^\infty \frac{G_j^\top(x, \hat{\boldsymbol{\theta}}_n) d\hat{F}_{nT_j,0}(x)}{\hat{S}_{nC_j,0}(x))} \right) \boldsymbol{A}_{nj}^{-1}(\hat{\boldsymbol{\theta}}_n) \int_0^\infty G_j(t, \hat{\boldsymbol{\theta}}_n) d\hat{F}_{nT_j,0}(t) + o_p(n_j^{-1/2}),$$

where

$$\boldsymbol{A}_{nj}(\hat{\boldsymbol{\theta}}_n) = \int_0^\infty \frac{G_j(t, \hat{\boldsymbol{\theta}}_n) G_j^\top(t, \hat{\boldsymbol{\theta}}_n)}{\hat{S}_{nC_j,0}(t)} d\hat{F}_{nT_j,0}(t).$$

It follows that

$$\begin{aligned} \sqrt{n_j}(\hat{S}_{nT_j}(t) - S_{T_j}(t)) &= \sqrt{n_j}(\hat{S}_{nT_j,0}(t) - S_{T_j}(t)) + \\ &\quad \sqrt{n_j} \left(\int_t^\infty \frac{G_j^\top(x, \hat{\boldsymbol{\theta}}_n) d\hat{F}_{nT_j,0}(x)}{\hat{S}_{nC_j,0}(x)} \right) \boldsymbol{A}_{nj}^{-1}(\hat{\boldsymbol{\theta}}_n) \int_0^\infty G_j(t, \hat{\boldsymbol{\theta}}_n) d\hat{F}_{nT_j,0}(t). \end{aligned}$$

Denote

$$\boldsymbol{\beta}_{nj}(t, \hat{\boldsymbol{\theta}}) = \boldsymbol{A}_{nj}^{-1}(\hat{\boldsymbol{\theta}}_n) \int_t^\infty \frac{G_j(x, \hat{\boldsymbol{\theta}}_n) d\hat{F}_{nT_j}(x)}{\hat{S}_{nC_j}(x)},$$

then, we have

$$\sqrt{n_j}(\hat{S}_{nT_j}(t) - S_{T_j}(t)) = \sqrt{n_j}(\hat{S}_{nT_j,0}(t) - S_{T_j}(t)) + \sqrt{n_j}\boldsymbol{\beta}_{nj}^\top(t, \hat{\boldsymbol{\theta}}_n) \int_0^\infty G_j(x, \hat{\boldsymbol{\theta}}_n) d\hat{F}_{nT_j}(x).$$

It is easy to show that, as $n_{min} \to \infty$,

$$\boldsymbol{\beta}_{nj}(t, \hat{\boldsymbol{\theta}}_n) \xrightarrow{p} \boldsymbol{\beta}_j(\boldsymbol{\theta}) = \boldsymbol{A}_j^{-1}(\boldsymbol{\theta}) \int_t^\infty \frac{G_j(x, \boldsymbol{\theta}) dF_{T_j}(x)}{S_{C_j}(x)}.$$

The asymptotic normality follows from the fact that both $Z_{n1j}(t) = \sqrt{n_j}(\hat{S}_{nT_j,0}(t) - S_{T_j}(t))$ and $Z_{n2j}(t) = \sqrt{n_j}\boldsymbol{\beta}_{nj}^\top(t, \hat{\boldsymbol{\theta}}_n) \int_0^\infty G_j(x, \hat{\boldsymbol{\theta}}_n) d\hat{F}_{nT_j}(x)$ are asymptotic normal, as $n_{min} \to \infty$.

Note that, according to Breslow and Crowley (1974, Theorem 5), $Z_{n1j}(t)$ converges to a Gaussian process $Z_{1j}(t)$, with $\mathrm{E}(Z_{1j}(t)) = 0$,

$$\mathrm{Cov}(Z_{1j}(s), Z_{1j}(t)) = S_{T_j}(s) S_{T_j}(t) \int_0^{\min(t,s)} \frac{dF_T(x)}{S_{T_j}(x)^2 S_{C_j}(x)},$$

and

$$\begin{aligned} \lim_{n_j \to \infty} Z_{n2j}(t) &= \lim_{n_j \to \infty} \sqrt{n_j}\boldsymbol{\beta}_{nj}^\top(t, \hat{\boldsymbol{\theta}}_n) \int_0^\infty G_j(x, \boldsymbol{\theta}) d\hat{F}_{nT_j}(x) \\ &= \lim_{n_j \to \infty} \boldsymbol{\beta}_{nj}^\top(t, \hat{\boldsymbol{\theta}}_n) \int_0^\infty \sqrt{n_j}(\hat{S}_{nT_j,0}(x) - S_{T_j}(x)) dG_j(x, \boldsymbol{\theta}) \end{aligned}$$

converges to a Gaussian process $Z_{2j}(t)$, where

$$Z_{2j}(t) = \boldsymbol{\beta}_j^\top(t, \boldsymbol{\theta}) \int_0^\infty Z_{1j}(x) dG_j(x, \boldsymbol{\theta}).$$

Therefore,

$$\sqrt{n_j}(\hat{S}_{nT_j}(t) - S_{T_j}(t)) = Z_{n1j}(t) + Z_{n2j}(t) \xrightarrow{p} Z_{1j}(t) + Z_{2j}(t).$$

Note that $\mathrm{E}(Z_{1j}(t) + Z_{2j}(t)) = 0$, so that the asymptotic variance of $\sqrt{n_j}(\hat{S}_{nT_j}(t) - S_{T_j}(t))$ reduces to

$$\sigma_{\hat{S}_j(t)} = \mathrm{Var}(Z_{1j}(t)) + \mathrm{Var}(Z_{2j}(t)) + 2\mathrm{Cov}(Z_{1j}(t), Z_{2j}(t)),$$

where

$$\begin{aligned} &\mathrm{Cov}(Z_{1j}(t), Z_{2j}(t)) = \mathrm{E}(Z_{1j}(t) Z_{2j}(t)) = \boldsymbol{\beta}_j^\top(t, \boldsymbol{\theta}) \int_0^\infty \mathrm{E}(Z_{1j}(t) Z_{1j}(x)) dG_j(x, \boldsymbol{\theta}) \\ &= \boldsymbol{\beta}_j^\top(t, \boldsymbol{\theta}) \left(\int_0^t \mathrm{E}(Z_{1j}(t) Z_{1j}(x)) dG_j(x, \boldsymbol{\theta}) + \int_t^\infty \mathrm{E}(Z_{1j}(t) Z_{1j}(x)) dG_j(x, \boldsymbol{\theta}) \right) \\ &= \boldsymbol{\beta}_j^\top(t, \boldsymbol{\theta}) \left\{ \int_0^t \left(S_{T_j}(t) S_{T_j}(x) \right) \int_0^x \frac{dF_{T_j}(s)}{S_{T_j}(s)^2 S_{C_j}(s)} \right) dG_j(x, \boldsymbol{\theta}) + \\ &\int_t^\infty \left(S_{T_j}(t) S_{T_j}(x) \right) \int_0^t \frac{dF_{T_j}(s)}{S_{T_j}(s)^2 S_{C_j}(s)} \right) dG_j(x, \boldsymbol{\theta}) \right\} \end{aligned}$$

$$= S_{T_j}(t)\boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})\left\{\int_0^t \frac{dF_{T_j}(s)}{S_{T_j}(s)^2 S_{C_j}(s)}\int_s^t S_{T_j}(x))dG_j(x,\boldsymbol{\theta}) + \int_0^t \frac{dF_{T_j}(s)}{S_{T_j}(s)^2 S_{C_j}(s)}\int_t^\infty S_{T_j}(x))dG_j(x,\boldsymbol{\theta})\right\}$$

$$= S_{T_j}(t)\int_0^t \left(\int_s^\infty \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})(G_j(s,\boldsymbol{\theta}) - G_j(x,\boldsymbol{\theta})dS_{T_j}(x)\right)\frac{dF_T(s)}{S_{T_j}(s)^2 S_{C_j}(s)}.$$

Note that

$$\mathrm{Var}(Z_{1j}(t)) = S_{T_j}^2(t)\int_0^t \frac{dF_{T_j}(x)}{S_{T_j}(x)^2 S_{C_j}(x)},$$

and it follows from Lemma A.1 that

$$\mathrm{Var}(Z_{2j}(t)) = \int_0^\infty \left[\int_x^\infty \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})(G_j(x,\boldsymbol{\theta}) - G_j(t,\boldsymbol{\theta}))dS_{T_j}(t)\right]^2 \frac{dF_{T_j}(x)}{S_{T_j}^2(x)S_{C_j}(x)}.$$

The results for the asymptotic normality of SMLE $\hat{S}_{nT}(t)$ are stated formally below.

Theorem 4.6 *Under the conditions of Theorem 4.1, as $n_{min}\to\infty$, $\sqrt{n_j}(\hat{S}_{nT_j},(t) - S_{T_j}(t)) \xrightarrow{d} N_m(0,\boldsymbol{\Sigma}_{S_j(t)})$, where*

$$\begin{aligned}\sigma_{S_j(t)}^2 \quad = \quad & S_{T_j}^2(t)\int_0^t \frac{dF_{T_j}(x)}{S_{T_j}(x)^2 S_{C_j}(x)} + \\ & \int_0^\infty \left[\int_x^\infty \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})(G_j(x,\boldsymbol{\theta}) - G_j(t,\boldsymbol{\theta}))dS_{T_j}(t)\right]^2 \frac{dF_{T_j}(x)}{S_{T_j}^2(x)S_{C_j}(x)} + \\ & 2S_{T_j}(t)\int_0^t \left(\int_s^\infty \boldsymbol{\beta}_j^\top(t,\boldsymbol{\theta})(G_j(s,\boldsymbol{\theta}) - G_j(x,\boldsymbol{\theta})dS_{T_j}(x)\right)\frac{dF_{T_j}(s)}{S_{C_j}(s)},\end{aligned}$$

$$\boldsymbol{\beta}_j(t,\boldsymbol{\theta}) = A_j(\boldsymbol{\theta})^{-1}\int_t^\infty \frac{G_j(x,\boldsymbol{\theta})dF_{T_j}(x)}{S_{C_j}(x)},$$

and $A_j(\boldsymbol{\theta})$ is given in (17).

4.5 Iterating Algorithm for SMLE Computation

Because of the complexity of equations (5)- (7), it is difficult to obtain the SMLE by solving these equations in practical applications. Motivated by Theorem 4.1 and (16), the following iterating algorithm is proposed for the SMLE computation to obtain $\hat{\boldsymbol{\theta}}$.

Step 1. Let

$$\boldsymbol{\lambda}_j^{(0)}(\boldsymbol{\theta}) = -\left[\int_0^\infty \frac{G_j(t,\boldsymbol{\theta})G_j^\top(t,\boldsymbol{\theta})}{\hat{S}_{C_j}(t)}d\hat{F}_{T_j}(t)\right]^{-1}\left[\int_0^\infty G_j(t,\boldsymbol{\theta})d\hat{F}_{T_j}(t)\right],$$

where $\hat{F}_{T_j}(t)$ and $\hat{S}_{C_j}(t)$ are the K-M estimates of $F_{T_j}(t)$ and $S_{C_j}(t)$, respectively. Solve the equation

$$\sum_{j=1}^{m} n_j \boldsymbol{\lambda}_j^{(0)}(\boldsymbol{\theta})^\top \sum_{i=1}^{k_j+1} P_{ij}^{(0)}(\boldsymbol{\theta}) \frac{\partial G_j(t_{ij}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$$

to obtain an initial $\hat{\boldsymbol{\theta}}^{(0)}$, where

$$P_{ij}^{(0)}(\boldsymbol{\theta}) = \frac{1}{n_j(1 - a_{ij}^{(0)} + \boldsymbol{\lambda}_j^{(0)}(\boldsymbol{\theta})^\top G_j(t_{ij}, \boldsymbol{\theta}))}, \quad i = 1, 2, \ldots, k_j, \quad \text{and} \quad P_{(k_j+1)j}^{(0)} = 1 - \sum_{i=1}^{k_j} P_{ij}^{(0)}.$$

Step 2. For given $\hat{\boldsymbol{\theta}}^{(b)}$, solve the equation (7) to get $\boldsymbol{\lambda}_j^{(b)} = \boldsymbol{\lambda}_j(\hat{\boldsymbol{\theta}}^{(b)})$.

Step 3. Update $\boldsymbol{l}^{(b)} = \boldsymbol{l}(\hat{\boldsymbol{\theta}}^{(b)})$ and $[\partial \boldsymbol{l}^{(b)}/\partial \boldsymbol{\theta}]$ according to (23) and (25), and let

$$\hat{\boldsymbol{\theta}}^{(b+1)} = \hat{\boldsymbol{\theta}}^{(b)} - \left[\frac{\partial \boldsymbol{l}^{(b)}}{\partial \boldsymbol{\theta}}\right]^{-1} \boldsymbol{l}(\hat{\boldsymbol{\theta}}^{(b)}),$$

where

$$\frac{\partial \boldsymbol{l}^{(b)}}{\partial \boldsymbol{\theta}} = \sum_{j=1}^{m} \frac{n_j}{n} E_j(\hat{\boldsymbol{\theta}}^{(b)})^\top A_j(\hat{\boldsymbol{\theta}}^{(b)})^{-1} E_j(\hat{\boldsymbol{\theta}}^{(b)}),$$

and

$$E_j(\hat{\boldsymbol{\theta}}^{(b)}) = \sum_{i=1}^{k_j+1} P_{ij}^{(b)} \frac{\partial G(t_{ij}, \hat{\boldsymbol{\theta}}^{(b)})}{\partial \boldsymbol{\theta}},$$

$$A_j(\hat{\boldsymbol{\theta}}^{(b)}) = \sum_{i=1}^{k_j+1} \frac{G_j(t_{ij}, \hat{\boldsymbol{\theta}}^{(b)}) G^\top(t_{ij}, \hat{\boldsymbol{\theta}}^{(b)})}{n_j(1 - a_{ij}^{(b)} + \boldsymbol{\lambda}_j^{(b)\top} G_j(t_{ij}, \hat{\boldsymbol{\theta}}^{(b)}))^2}.$$

Step 4. Repeat Steps 2 and 3 until $\|\hat{\boldsymbol{\theta}}^{(b+1)} - \hat{\boldsymbol{\theta}}^{(b)}\| < \epsilon$, where ϵ is some prespecified level of precision.

The proposed algorithm plugs in the Kaplan-Meier estimate as an initial estimator of the SMLE. It is updated with $\hat{P}_{ij} = P_{ij}(\hat{\boldsymbol{\theta}})$ from (5) - (7) after obtaining $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(0)}$ and $\boldsymbol{\lambda}_j^{(0)} = \boldsymbol{\lambda}_j(\hat{\boldsymbol{\theta}}^{(0)})$. According to Theorem 4.1, if the sample size n_{min} is large enough, the iterative solution will converge with probability one. One can also use the result from Step 1 (without iterating) as a simple estimate. Although it is not the exact SMLE, it is also strongly consistent and has the same asymptotic normal distribution as the SMLE.

In the next section, we perform some numerical evaluation for comparing the asymptotic efficiency of the SMLE and the parametric MLE under mis-specified models and illustrate the SMLE with a real-life application.

5 Numerical Evaluation

5.1 Asymptotic Bias and Variance Studies

Consider interval-censored data. Chen, Lu and Lin (2005) investigated the asymptotic efficiency of the SMLE versus the parametric MLE for mis-specified lifetime distributions. They concluded that if the distribution assumption is correct for the parametric MLE, the performance of the SMLE is still close to the MLE with more than 80% asymptotic efficiency. When the distribution assumption is incorrect, the asymptotic Mean Squared Error (MSE) of SMLE is less than 10% of asymptotic MSE of misspecified MLE (misMLE) in the censored sample cases. Since the ALT studies involve extrapolations, the mis-specification of the regression model could be more important than the mis-specification of the lifetime distribution. For this reason, we focus our study on the mis-specified regression model.

Suppose that the main interest is to estimate the 10% lifetime at the normal-use condition. Our goal is to compare the percentile regression against the mis-specified model, mean (and variance) regression. When the lifetime distribution is in a location-scale family, the percentile is a linear function of location and scale parameters. In the case of a constant scale, the percentile is a linear function of the location parameter, and the mean and percentile regressions will lead to the same result. When the lifetime distribution is not in the location-scale family, the separation between the mean and percentile regression can be significant. For example, consider the $Gamma(\theta, \kappa)$ distribution, where $\theta > 0$ is a scale parameter and $\kappa > 0$ is a shape parameter. The q-th percentile is $\xi_q = \theta \Gamma_I^{-1}(q; \kappa)$, where Γ_I is the incomplete Gamma function (Meeker & Escobar, 1998, page 99).

When the parametric model is mis-specified, the estimator obtained by maximizing the log-likelihood is no longer asymptotically optimal. Under proper regularity conditions, White (1982) showed that the misMLE will converge to a well defined limit of θ^*, which minimizes the Kullback-Leibler Information Criterion, $\mathrm{E}_g(\log[g(\theta, \boldsymbol{t})/f(\theta; \boldsymbol{t})])$, where f is the mis-specified pdf, g is the true pdf, and the expectations are taken with respect to the true distribution with the parameter θ. The asymptotic variance for the misMLE can be evaluated as

$$\Sigma(\theta^*) = A(\theta^*)^{-1} B(\theta^*) A(\theta^*)^{-1}, \tag{28}$$

where

$$\begin{aligned} A(\theta) &= \mathrm{E}_g(\partial^2 \log f(\boldsymbol{t}, \theta)/\partial\theta_i \partial\theta_j), \\ B(\theta) &= \mathrm{E}_g[\partial \log f(\boldsymbol{t}, \theta)/\partial\theta_i \cdot \partial \log f(\boldsymbol{t}, \theta)\partial\theta_j]. \end{aligned}$$

Consider an ALT experiment with three levels, where low, middle, and high stress levels are re-scaled as 0.5, 0.75 and 1, respectively, such that the normal-use stress level $x_D = 0$. Replicated samples are allocated at each stress levels according to a $4 : 2 : 1$ proportion (Meeker and Hahn, 1985) of their sample sizes. We will consider the log-normal and Gamma distributions, representing location-scale and non-location-scale families, respectively.

Experiment #1 - Log-normal Distribution: Assume that the location and scale parameters in $LN(\mu, \sigma)$ can be modeled using linear function of stress levels x as $\mu(x) = 1 - x$ and $\sigma(x) = 2 - x$. Then, the $100q\%$ lifetime in log-scale is $\xi_q(x) = \mu(x) + z_q\sigma(x) = (1 + 2z_q) + (z_q - 1)x = x'\beta$, where z_q is the q-quantile of the standard normal distribution. Four regression cases and two censoring are explored.

1. $\mu(x) = x'\alpha$, $\sigma(x) =$ constant
2. $\mu(x) = x'\alpha$, $\log \sigma(x) = x'\gamma$
3. $\mu(x) = x'\alpha$, $\sigma(x) = x'\gamma$ (true parametric model)
4. $\xi_q(x) = x'\beta$, using the SMLE method to estimate β

The first three models are for the parametric MLE and the last one is for the SMLE. Since the percentile here is a function of both location and scale parameters, a mis-specification of the scale (or mean) model does not destroy the percentile's linear relationship with the stress variable. The first case studies the impact of a mis-specified variance model. The second case is similar to the first. However, when the variance model is nonlinear, the percentile regression becomes nonlinear.

Table 4, summarizing the proportions of failure data for the simulated data, shows that the MSE in the misMLE Cases 1(a-b) is more than 74% larger than the MSE in SMLE's in all cases studied. The bias in the MisMLE makes the MSE significantly larger than the SMLE's. Since the log-linear variance model used in Case 1(b) can approximate the true linear model better than the constant variance model, Case 1(b) has smaller bias. Compared with the correct parametric MLE, the SMLE's asymptotic variance is about 28% larger. When the model is mis-specified, the asymptotic variance could be seriously under- or -over estimated. For example, the variance in Case 1(b) is about 34% larger than the variance in the SMLE (Case 1(d)).

Experiment #2 - Gamma Distribution: Consider a non-location-scale $Gamma(\theta, \kappa)$ family of distributions, where $\theta > 0$ is a scale parameter and $\kappa > 0$ is a shape parameter. Assume a linear

Proportion Failing	n=301	Case 1(a)	Case 1(b)	Case 1(c)	Case 1(d)
(0.70, 0.80, 0.90)	bias^2	0.557	0.102	0.00	0.00
	Var($\hat{\xi}_q$)	0.089	0.335	0.158	0.221
(0.44, 0.50, 0.60)	bias^2	0.334	0.143	0.00	0.00
	Var($\hat{\xi}_q$)	0.078	0.369	0.163	0.221

Table 4: Simulation Results of Experiment #1.

Proportion Failing	n=301	Case 2(a)	Case 2(b)	Case 2(c)
(0.73, 0.82, 0.90)	bias^2	1.22	8.30	0.000
	Var($\hat{\xi}_q$)	0.38	3.77	0.25

Table 5: Simulation Results of Experiment #2.

percentile regression, $\xi_q(x_j) = \beta_0 + \beta_1 x_j = \theta_j \Gamma_I^{-1}(q, \kappa_j)$. For simplicity, we fix the $\kappa_j = \kappa$ for all stress levels, and choose $\kappa = 2$ indicating an increasing hazard rate function. Let $\beta_0 = 2$, $\beta_1 = -1$.

Although the Gamma distribution doesn't belong to the location-scale family, the following parametrization is often recommended for numerical stability: $\mu = \log(\theta)$, $\sigma = 1/\sqrt{\kappa}$. Thus, the scale parameters $\theta_L, \theta_M, \theta_H$ are $2.8205, 2.3505, 1.8804$, respectively. Since $\theta = exp(\mu)$ and the true model has a linear regression structure on percentile $\xi_q(x_j) = exp(\mu)\Gamma_I^{-1}(q, \kappa_j)$, $\theta = exp(\mu)$ is a linear function of stress when $\kappa_j = \kappa$. Three models are investigated below; the first two give simple, alternative structures of the regression function of μ.

1. $\mu = x'\alpha$, $\sigma =$ constant.

2. $\log(\mu) = x'\alpha$ (or $\mu = exp(x'\alpha)$), $\sigma =$ constant.

3. $\xi_q(x) = x'\beta$, using SMLE method to estimate β.

Table 5 shows that the misspecification leads to an even larger asymptotic bias and variance compared to results in Experiment #1. Both cases in 2(a-b) have larger variance than the SMLE.

5.2 PCB Example

In Section 2, Figure 1 showed that Weibull distribution is not appropriate for modeling the PCB data from Meeker and LuValle (1995). We consider it further, along with the SMLE method, because it serves as benchmark for testing the SMLE. Following Meeker and LuValle's Weibull mean regression model,

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

where

$$\mu(X_j) = \beta_0 + \beta_1 \text{logit}(X_j), \quad X_j = \text{RH}/100, \quad \text{logit}(p) = \log[p/(1-p)], \tag{29}$$

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 can be justified physically (Meeker and LuValle, 1995). The parametric MLE estimates for model parameters are calculated as $\hat{\beta}_0 = 9.10$, $\hat{\beta}_1 = -3.78$, $\hat{\sigma} = 0.93$, respectively.

Figure 3 gives 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 $Y = \log(T)$ is $\hat{\eta}_p(x) = \hat{\beta}_0 + \hat{\beta}_1 + w_p\hat{\sigma}$, where $w_p = \log[-\log(1-p)]$. Confidence intervals for $\xi_p(x)$ can be obtained by using the large-sample normal approximation with the asymptotic variance calculated from the Fisher information matrix (Lawless 1982, page 305). 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 $(13.54, 15.70)$, respectively.

Figure 4 compares the confidence intervals for the percentile regression coefficients ξ_p and β_1 using SMLE method. Using the delta method, the corresponding point estimate and CI of the 5th percentile lifetime at the normal-use condition are 12.30 and $(11.83, 12.78)$, respectively. Note that the width of this CI is only about 44% of the width for the CI calculated using the Weibull regression model. Back-transform the estimate to the original time scale, the 5th percentile lifetime is predicted as 25 years. Recall that based on the physics-based kinetics model given in Meeker and LuValle (1995), the proportion of product failing is less than 1% under the normal-use condition, and their prediction of the 5th percentile is infinity. Comparing this result against with the SMLE prediction, the SMLE is more conservative and arguably easier to interpret.

To validate the estimating equations pertaining to the log-5th-percentile's linear regression relationship with the stress levels, i.e., $\mathrm{E}(G_j(T_j;\boldsymbol{\theta}, X_j)) = 0$, the SMLE-based LR-test can be conducted according to Corollary 3.5. The LR statistic is $\Lambda_3 = 2.26$ and the p-value is 0.13. In this example, the percentile regression model seems plausible.

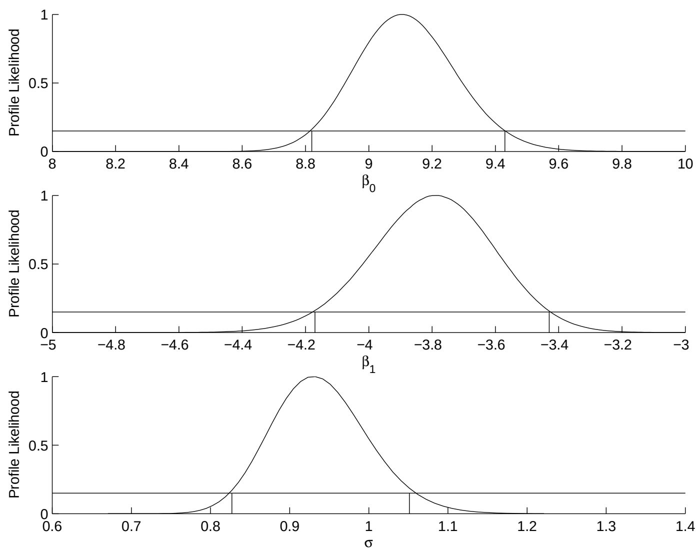

Figure 3: Profile Likelihoods of β_0, β_1 and σ Using the Weibull Regression Model

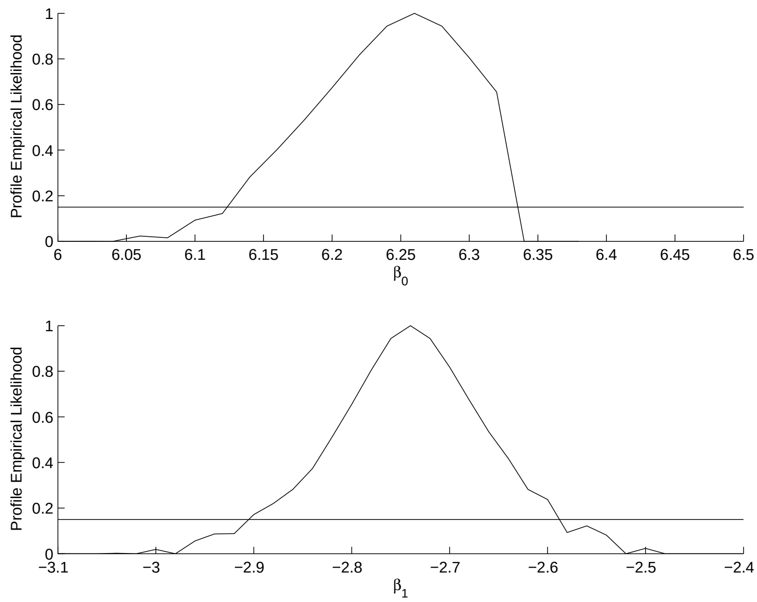

Figure 4: Profile Empirical Likelihoods for β_1 and $\xi_{0.05}$ Using the SMLE Method

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

1. Case (i) − After adjusting the 5th percentiles, lifetime distributions are the same.

2. Case (ii) − After adjusting the 5th percentiles and re-scaling with the quantile-range, 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. The SMLE in Case (i) estimates the 5th percentile regression parameters as $(\beta_0, \beta_1) = (6.2931, -2.6378)$. Correspondingly, the SMLE in Case (ii) leads to $(6.3060, -2.6753)$. Their 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.

6 Concluding Remarks

The SMLE method provides reasonable estimates in a real-life example useful in reliability improvement comparisons. The proposed data-exploration based percentile and quantile-range regressions are effective in overcoming the difficulty of observing mean lifetime in the heavy censored data case for constructing commonly used mean and variance regression models in ALT studies. The LR-based test provides a revenue for validating the model formally. The asymptotic bias and variance studies show that the SMLE is reasonably competitive against the parametric MLE method when the model is correctly specified, and performs much better than the MLE when the model is incorrectly specified. Based on the properties derived in this article, the SMLE method should be a strong candidate for handling challenging data modeling and statistical inference problems.

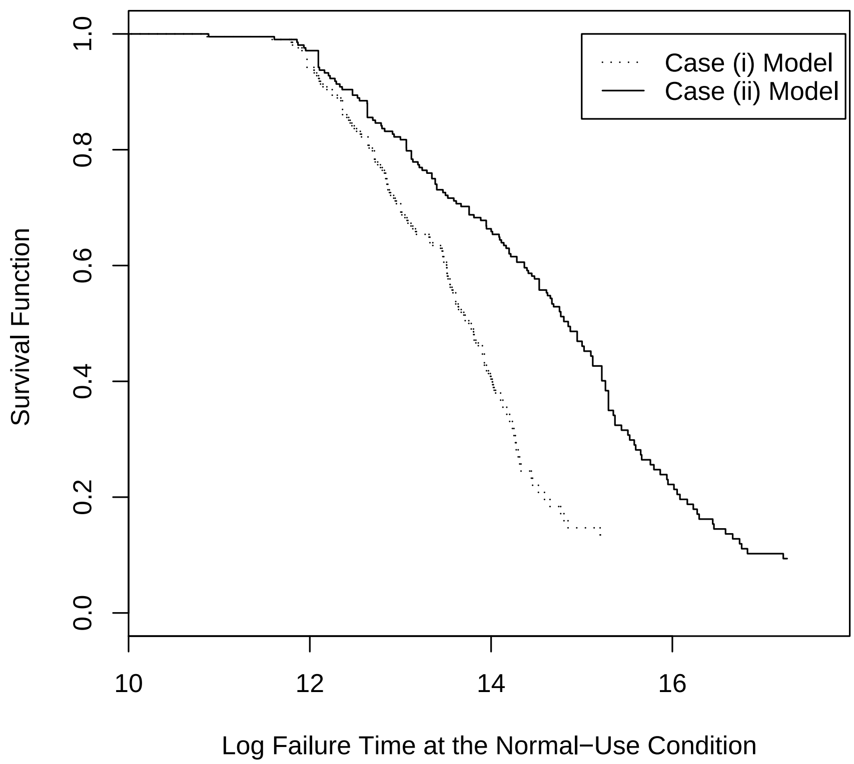

Figure 5: SMLE of Survival Functions Under Different Assumptions.

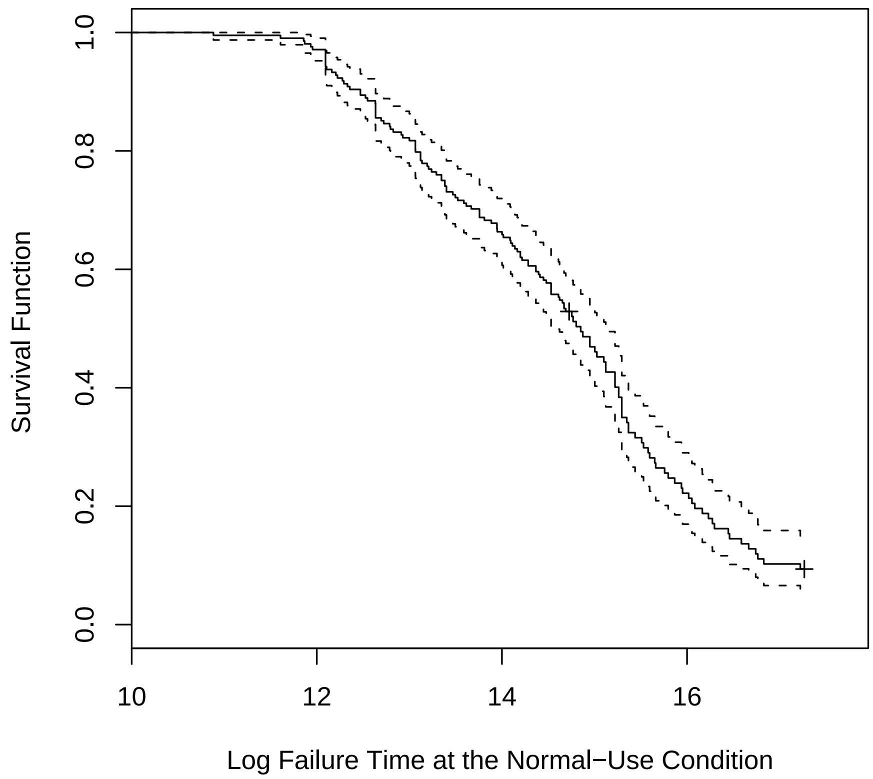

Figure 6: SMLE of Survival Function of the Failure Time

7 Appendix: Some Properties of Gaussian processes

Lemma A.1 *Let $Z(t)$ be a Gaussian process satisfying (1) $Z(0) = 0$ and $E(Z(t)) = 0$ for any t; (2) for any (s,t),*

$$Cov(Z(s), Z(t)) = R(t)R(s) \int_0^{\min(s,t)} h(x)dx.$$

Let $G(t) = (g_1(t), \dots g_r(t))^\top$ and denote

$$\Psi_G = \int_0^\infty Z(x)dG(x) = \left[\int_0^\infty Z(x)dg_1(t), \dots, \int_0^\infty Z(x)dg_r(t)\right]^\top.$$

Suppose that $G(x)$ is differentiable and $R(t) \to 0$ as $t \to \infty$. Then, Ψ_G is distributed $N_m(0, \mathbf{\Sigma}_G)$, where

$$\mathbf{\Sigma}_G = \int_0^\infty h(s) \left[\int_s^\infty (G(s) - G(x))dR(x)\right] \left[\int_s^\infty (G(s) - G(x))dR(x)\right]^\top ds.$$

Proof: We show the normality first. Let $0 = x_0 < x_1 < \cdots < x_n$, and $\Psi_n = \sum_{i=1}^n Z(x_i)(G(x_i) - G(x_{i-1}))$. Then Ψ_n is a series of normal random variables. Suppose that, as $n \to \infty$, $\max_{1 \le i \le n} |x_i - x_{i-1}| \to 0$ and $x_n \to \infty$. Then we know that $\Psi_n \xrightarrow{p} \Psi_G$. It follows from the normality of every Ψ_n that Ψ_G is normal distributed as well, and $\mathrm{E}(\Psi_G) = \int_0^\infty \mathrm{E}(Z(x))dG(x) = 0$.

Next, we calculate the covariance of Ψ_G. Note that

$$\begin{aligned}
\mathbf{\Sigma}_G &= \mathrm{Cov}\left(\int_0^\infty Z(x)dG(x), \int_0^\infty Z(x)dG(x)^\top\right) \\
&= \mathrm{E}\left[\int_s^\infty Z(t)dG(t)\right]\left[\int_s^\infty Z(s)dG(s)\right]^\top \\
&= \int_0^\infty \int_0^\infty \mathrm{E}(Z(t)Z(s))dG(s)dG(t)^\top \\
&= \int_0^\infty \int_0^\infty \left(R(t)R(s)\int_0^{\min(s,t)} h(x)dx\right) dG(s)dG(t)^\top \\
&= \int\int_{t \le s} \left(R(t)R(s)\int_0^t h(x)dx\right) dG(s)dG(t)^\top + \\
&\quad \int\int_{t > s} \left(R(t)R(s)\int_0^s h(x)dx\right) dG(s)dG(t)^\top \\
&= \int_0^\infty h(x)dx \left(\int_x^\infty R(t)\int_t^\infty R(s)dG(s)dG(t)^\top\right) + \\
&\quad \int_0^\infty h(x)dx \left(\int_x^\infty R(s)\int_x^t R(t)dG(t)dG(s)^\top\right) \\
&= \int_0^\infty h(x)dx \left[\int_x^\infty R(t)dG(t)\right]\left[\int_x^\infty R(t)dG(t)\right]^\top.
\end{aligned}$$

Integrating by parts, we know that

$$\begin{aligned}\int_x^\infty R(t)dG(t) &= -R(x)G(x) - \int_x^\infty G(t)dR(t) = \int_x^\infty G(x)dR(t) - \int_x^\infty G(t)dR(t) \\ &= \int_x^\infty (G(x) - G(t)dR(t).\end{aligned}$$

It follows that

$$\begin{aligned}\mathbf{\Sigma}_G &= \mathrm{Cov}\left(\int_0^\infty Z(x)dG(x), \int_0^\infty Z(x)dG(x)^\top\right) \\ &= \int_0^\infty h(s)\left[\int_s^\infty (G(s) - G(x))dR(x)\right]\left[\int_s^\infty (G(s) - G(x))dR(x)\right]^\top ds.\end{aligned}$$

The proof is thus completed. □

Denote $Z_n(t) = \sqrt{n}(\hat{S}_T(t) - S_T(t))$, where $\hat{S}_T(t)$ is the Kaplan-Meier estimator of $S_T(t)$. We know $Z_n(t)$ converges to a Gaussian process $Z(t)$ satisfying $Z(0) = 0$ and $\mathrm{E}(Z(t)) = 0$ with covariance function

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

Replacing $R(t)$ and $h(x)dx$ in (30) by $S_T(t)$ and $S_T(x)^{-2}S_C(x)^{-1}dF_T(x)$, respectively, we obtain the following corollary.

Corollary A.1 *Let $\boldsymbol{A}_n = \int_0^\infty \sqrt{n}(\hat{S}_T(x) - S_T(x))dG(x)$. Then $\boldsymbol{A}_n$ has the asymptotic normal distribution $N(0, \mathbf{\Sigma}_G)$, with covariance matrix*

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

□

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] Pascual, F. G., and Montepiedra, G. (2005), "Lognormal and Weibull Accelerated Life Test Plans Under Distribution Misspecification," *IEEE Transactions On Reliability*, 54(1), 43-52.

[4] Gertsbakh, I. B., and Kordonskiy, K. B. (1969). *Models of Failure (translated from Russian).* New York: Springer-Verlag.

[5] Hutton, J. L., and Monaghan, P. F. (2002), "Choice of Parametric Accelerated Life and Proportional Hazards Models for Survival Data: Asymptotic Results," *Lifetime Data Analysis*, 8, 375-393.

[6] Joseph, V.R., and Yu, I-T, "Reliability Improvement Experiments with Degradation Data", Technical Report, The School of Industrial and Systems Engineering, Georgia Institute of Technology. Paper can be reviewed in http://www.isye.gatech.edu/ brani/isyestat/05-04.pdf

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

[8] Kalbfleish, J. D., and Prentice, R. L, (1980). *The Statistical Analysis of Failure Time Data*, Wiley.

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

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

[11] Lu, C. J., and Meeker, W. Q. Jr. (1993), "Using Degradation Measures to Estimate A Time-to-failure Distribution," *Technometrics*, 35, 161-174.

[12] Lu, J.-C., Park, J., and Yang, Q. (1997), "Statistical Inference of A Time-to-Failure Distribution Derived from Linear Degradation Data," *Technometrics*, 39, 391-400.

[13] 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.

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

[15] Meeker, W. Q., Jr., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*, John Wiley and Sons: New York.

[16] Meeker, W. Q., Jr., and Hahn, J. G., (1985). *How to Plan an Accelerated Life Test – Some Practical Guidelines*, Milwaukee, WI, 1985, vol. 10, ASQC Basic References in Quality Control, Statistical Techniques.

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

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

[19] Qin, J. (2000), "Combining Parametric and Empirical Likelihoods", *Biometrika*, 87, 484-490.

[20] Su, C., Lu, J.-C., Chen, D., and Hughes-Oliver, J. M. (1999), "A Linear Random Coefficient Degradation Model with Random Sample Size," *Lifetime Data Analysis*, 5, 173-183.

[21] Tsiatis, A. (1981), "A Large Sample Study of Cox's Regression mMdel," *The Annals of Statistics*, 9(1), 93-108.

[22] Wang J. G. (1987), "A Note on the Uniform Consistency of the Kaplan-Meier Estimater," *Ann. Statist.* 15, 1313-16.

[23] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-26.

[24] Wang, N., Lu, J.-C., and Kvam, P. H. (2005), "Multi-Level Spatial Modeling and Decision-Making with Application in Logistics Systems," Technical Report, The School of Industrial and Systems Engineering, Georgia Institute of Technology. Paper can be reviewed in http://www.isye.gatech.edu/research/files/jclu-2004-08.pdf

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

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