

Research Bank

Journal article

Bias reduction in the two-stage method for degradation data analysis

Xu, Ancha, Wang, You-Gan, Zheng, Shurong and Cai, Fengjing

This is the accepted manuscript version. For the publisher's version please see:

Xu, A., Wang, Y.-G., Zheng, S. and Cai, F. (2020). Bias reduction in the two-stage method for degradation data analysis. *Applied Mathematical Modelling*, 77(Part 2), pp. 1413-1424. <https://doi.org/10.1016/j.apm.2019.08.024>

This work © 2020 is licensed under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/).

Bias reduction in the two-stage method for degradation data analysis

Ancha Xu¹, You-Gan Wang^{2,*}, Shurong Zheng³ and Fengjing Cai^{4,*}

¹Zhejiang Gongshang University, Zhejiang, China. Email: xuancha@wzu.edu.cn

^{2,*} School of Mathematical Sciences, Queensland University of Technology, Queensland, Australia

³ Northeast China Normal University, Jilin, China. Email: zhengsr@nenu.edu.cn

^{4,*} Wenzhou University, Zhejiang, China. Email: caifj7704@wzu.edu.cn

Abstract

Degradation data are usually collected for assessing the reliability of the product. We propose a new two-stage method to analyze degradation data. The degradation path is fitted by the nonlinear mixed effects model in the first stage, and the parameters in lifetime distribution are estimated by maximizing the asymptotic marginal distribution of pseudo lifetimes in the second stage. The new method has many advantages: (i). it does not require the distributions on random effects, (ii). the historical information about lifetime distribution of the product can be incorporated easily, and thus the estimated lifetime distribution has a closed form, (iii). bias correction term is automatically embedded into the asymptotic marginal distribution of pseudo lifetime. Finally, simulation studies and real data analysis are performed for illustration.

Keywords: Degradation data; Random effects; Stochastic process; General path model; Bias reduction.

1 Introduction

1.1 Degradation models review

Degradation data are often collected in assessing reliability and life span of industrial products. The models for analyzing degradation data are widely developed in past two decades. For example, general path models [1, 2], stochastic degradation models [3, 4], Markov and semi-Markov models [5, 6], data-driven models [7], and so on. However, the stochastic degradation model and general path model are the two most used degradation models. [8] gave a comprehensive review on the two kinds of degradation models, while [9] focused the review work on the Wiener process and its related models. Recently, [10] used the objective Bayesian method to analyze constant stress accelerated degradation test based on inverse Gaussian (IG) process. [11] considered the IG process model with nonsymmetric random effects using skew-normal distribution. [12] proposed generalized inference method for constant stress accelerated degradation test based on gamma process, which is efficient for the case of small sample size. [13] presented a linear degradation model with skewness and heavy tailed behaviors. [14] introduced a hierarchical Bayesian bi-exponential degradation model with application to a fuel cell stack system. The advantages of stochastic degradation models are that the mathematical derivation is analytical and the physical explanation of the performance degradation can be well justified. While the general path model is to describe longitudinal nature of the degradation process for individual units, and therefore it has both fixed and random effects with fixed effects for the population levels and random effects for individual variability. [8] indicated that the stochastic degradation models are capable of explaining the randomness of the degradation over time due to unobserved environmental factors, while the general path model is much restrictive by assuming that the inherent degradation is deterministic, and thus is applicable when the randomness among the unobserved environmental factors is small enough.

1.2 Motivation

In fact, the difference for deriving the cumulative distribution function (CDF) of the lifetime of the product exists between the two kinds of degradation models. For the stochastic degradation model, the CDF of the lifetime can be achieved by the theory of first passage time for a stochastic process, while for the general path model, the CDF of the lifetime stems from the random-effect variables. Thus, the stochastic degradation models are preferable when the information for the CDF of the lifetime is not available. However, the historical data, the experts' experience or lifetime information of a similar product are usually obtainable in practice. We refer to these information as prior information. The prior information may contradict lifetime distribution deriving from the assumption of a stochastic degradation model. For example, assume that the prior information implies the product's lifetime following log-normal distribution, while the product's lifetime can be shown to follow IG distribution if the Wiener process is assumed for the degradation process. Thus, the stochastic degradation model may be not suitable when prior information exists. For the general path model, the prior information could be used directly or be incorporated into the random-effect variables. Besides, CDF based on stochastic degradation model with random effects is usually not analytical, especially for the cases of involving change points and multiple degradation paths [15, 16, 17]. The general path model may also encounter the problem when the degradation path is nonlinear and there exists several random effects, which will be discussed below.

Assume that y_{i1}, \dots, y_{im_i} are the degradation values measured at times x_{i1}, \dots, x_{im_i} for the i -th unit, $i = 1, \dots, n$. The general path model is expressed as

$$y_{ij} = \eta(x_{ij} | \boldsymbol{\beta}, \mathbf{b}_i) + e_{ij}, j = 1, \dots, m_i, i = 1, \dots, n, \quad (1)$$

where $\eta(\cdot)$ is some known function describing the degradation path, $\boldsymbol{\beta}$ is a vector of fixed-effect parameters, \mathbf{b}_i is a vector of random effects for unit i , and e_{ij} represents the unexplained random error including the measurement error. The random errors are assumed to be inde-

pendent of each other and e_{ij} is assumed to have a constant variance $V(e_{ij}) = \sigma_e^2$. Assume that the joint distribution for the random effects is specified as $\pi(\mathbf{b}_i|\boldsymbol{\omega})$ ($\boldsymbol{\omega}$ is a parameter vector). Based on the data $y = \{y_{ij}, j = 1, \dots, m_i, i = 1, \dots, n\}$, the likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\omega})$ is

$$L = \prod_{i=1}^n \int \prod_{j=1}^{m_i} f(y_{ij}|\boldsymbol{\beta}, \mathbf{b}_i)\pi(\mathbf{b}_i|\boldsymbol{\omega})d\mathbf{b}_i.$$

The distribution of \mathbf{b}_i or the transformed (reparametrized) \mathbf{b}_i is usually assumed to be multivariate normal. If \mathbf{b}_i is p -dimensional, the above integral may induce substantial computational complexity, which is often the case in nonlinear mixed models. With recent computational development for nonlinear mixed models ([18]), numerical solutions are readily available from statistical packages in SAS, S-PLUS and R. The normality assumption of the random effects may not always be realistic, because the random effects are not observed, which may be difficult to verify. [19] have indicated that the parameter estimation of $\boldsymbol{\beta}$ can be sensitive to the misspecification of the joint distribution of random effects. However, correct specification of the joint distribution of the random effects reflecting the true stochastic nature is often difficult, if not impossible. On the other side, the aim of analyzing degradation data is often, not to characterize the path model itself, but to provide information on lifetime distribution. Without loss of generalization, we assume that $\eta(\cdot)$ is a strictly monotone increasing function of measurement time. Let η_c be the threshold level for the degradation path. Then from [the Equation 13.9 on page 329 in \[20\]](#), the lifetime for the unit is defined as

$$T = \inf\{t|\eta(t|\boldsymbol{\beta}, \mathbf{b}) \geq \eta_c\}. \quad (2)$$

If \mathbf{b} is a univariate, T can be expressed as $\eta^{-1}(\mathbf{b})$. That is, \mathbf{b} and T is one-to-one corresponding, and thus the prior information about T can be incorporated into \mathbf{b} easily. However, in the case of multivariate \mathbf{b} , such one-one correspondence between \mathbf{b} and T breaks down. It is difficult to transmit the prior information about T to \mathbf{b} . Two-stage method could solve the problem for multivariate \mathbf{b} ([21, 22]), and the method has high computational efficiency

and simple implementation procedure. However, the first stage will bring estimation bias, and [21] did not address this bias well. In this paper, we will propose a new procedure for two-stage method. The new procedure has many advantages. Firstly, we do not need to assume the joint distribution of random effects; then the prior information of T can be used directly, finally and the most importantly, the estimation bias in the first stage could be corrected.

The paper is organized as follows. In section 2, we introduce the two-stage method, and a new procedure for the two-stage method is proposed in section 3. Simulation studies are performed to compare the new procedure with the other two-stage methods in section 4. Real data analysis is carried out for illustration in Section 5. Finally, we give a conclusion of this paper.

2 Two-stage methods

The two-stage algorithm is fully described in [21], who also reported that this method performs well comparing with other more computationally intensive methods. We briefly summarize the two steps as

- (I). In the first stage, obtain the individual estimates $(\hat{\beta}_i, \hat{\sigma}_{ei}^2, \hat{\mathbf{b}}_i)$ of $(\beta, \sigma_e^2, \mathbf{b}_i)$ for the i -th unit using the least-squares method based on the observed data $(x_{i1}, y_{i1}), \dots, (x_{im_i}, y_{im_i})$.
- (II). In the second stage, combine these individual estimates to obtain the final parameter estimates of the population parameters (β, σ_e^2) and $\boldsymbol{\omega}$ in the joint distribution $\pi(\mathbf{b}|\boldsymbol{\omega})$ using moment approach or maximum likelihood method.

There are a few advantages of this method: it is computationally simple and it does not require normality of the random effects. [21] suggested some transformation may be necessary for the estimates of \mathbf{b} so that their distribution is approximately multivariate normal. This is equivalent to reparametrize \mathbf{b} in the path model. In fact, the approaches used in the

second stage do not require normality because the sample mean and covariance estimators are unbiased. Although the individual estimates $(\hat{\boldsymbol{\beta}}_i, \hat{\sigma}_{e_i}^2, \hat{\mathbf{b}}_i)$ in the first stage are consistent, the estimation efficiency may be reduced since only part of the observed data are used to estimate the population parameters $(\boldsymbol{\beta}, \sigma_e^2)$. Besides, the CDF of T is not usually analytic. [21] proposed to a Monte Carlo method to obtain the CDF of T . However, the engineers often have some historical experiences about the CDF of T , and prefer to use explicit CDF in practice.

If the objective is to infer on the lifetime distribution of the product, we may wish to impose directly a parametric model for it [23]. Such parametric model may be chosen to meet certain desirable properties based on experience. One obvious advantage of this approach is that it is one-dimensional. Then [22] modified the two-stage method by assuming that T follows some lifetime distribution with probability density function (PDF) $g(t|\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is an unknown parameter vector. Their method is summarized as follows:

- (I). In the first stage, obtain the estimates $(\hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2, \hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_n)$ of $(\boldsymbol{\beta}, \sigma_e^2, \mathbf{b}_1, \dots, \mathbf{b}_n)$ using the nonlinear least-squares method based on the observed data $\{(x_{i1}, y_{i1}), \dots, (x_{im_i}, y_{im_i}), i = 1, \dots, n\}$. Then by solving the equation $\eta(t|\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i) = \eta_c$ for unit i , obtain the pseudo lifetimes

$$\hat{T}_i = h(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i), i = 1, \dots, n.$$

- (II). The second stage is an iterative algorithm that includes three steps:

Initialization. Obtain the maximum likelihood estimate (MLE) of $\boldsymbol{\xi}$ by using the pseudo lifetimes

$$\hat{T}_i, i = 1, \dots, n. \text{ We denote the initial estimate as } \hat{\boldsymbol{\xi}}^{(1)}.$$

Imputation. At the k -th iteration, predict T_i by \tilde{T}_i , and

$$\tilde{T}_i = \frac{\int_{T_i^L}^{T_i^U} t \cdot g(t|\hat{\boldsymbol{\xi}}^{(k)}) dt}{\int_{T_i^L}^{T_i^U} g(t|\hat{\boldsymbol{\xi}}^{(k)}) dt},$$

where $T_i^L = \max \left\{ t_{ij}, \widehat{T}_i - Z_{\alpha/2} \sqrt{V(\widehat{T}_i)} \right\}$, $T_i^U = \min \left\{ t_{ij+1}, \widehat{T}_i + Z_{\alpha/2} \sqrt{V(\widehat{T}_i)} \right\}$, t_{ij} denotes the maximum of the x_{ij} s that are less than or equal to \widehat{T}_i , $t_{ij'+1} = \infty$ if $\widehat{T}_i > x_{im_i}$, and $Z_{\alpha/2}$ is the $(1 - \alpha/2)$ -th quantile of the standard normal distribution.

Maximization. Maximize the loglikelihood function $l_{cz} = \sum_{i=1}^n \log \left[g \left(\widetilde{T}_i | \boldsymbol{\xi} \right) \right]$ with respect to $\boldsymbol{\xi}$, and obtain $\widehat{\boldsymbol{\xi}}^{(k+1)}$.

– Repeat Imputation step and Maximization step until convergence occurs.

The two-stage method of [22] is an iterative algorithm, and we call it “[iterative two-stage method](#)”. If $\alpha = 1$, the idea is similar to [21]. Then the above imputation step can be omitted and $\widehat{\boldsymbol{\xi}}^{(1)}$ is the final estimate. We call the two-stage method with $\alpha = 1$ “[direct two-stage method](#)”. For $0 < \alpha < 1$, [22] suggested that for large $\widehat{V(\widehat{T}_i)}$, α should be also chosen largely, and vice versa. They also showed via simulation studies that the their two-stage method performs better than direct two-stage method. However, iterative two-stage method is an iterative algorithm, and thus it is not clear that how the method could improve the precision of parameter estimation.

3 Bias reduction two-stage method

In the first stage of iterative two-stage method, \widehat{T}_i can be treated as the estimate of the true lifetime of the i -th unit T_i . Under some regularity conditions, and $m_i \rightarrow \infty$ ([22]),

$$\mathbb{E}(\widehat{T}_i) = T_i + O(m_i^{-1}). \quad (3)$$

Our interest lies in $g(t|\boldsymbol{\xi})$, and T_i is unobservable. From (3), we know that further inference on the unknown parameter vector $\boldsymbol{\xi}$ can just rely on \widehat{T}_i s. Thus, we suggest to maximize the asymptotic likelihood function for \widehat{T}_i s, and the asymptotic likelihood function for \widehat{T}_i can be

written as

$$L_i = \int_0^\infty f(\widehat{T}_i|T_i = t)g(t|\boldsymbol{\xi})dt. \quad (4)$$

The first term $f(\widehat{T}_i|T_i = t)$ is a conditional PDF of \widehat{T}_i given T_i , since \widehat{T}_i is an estimate of T_i , and the magnitude of \widehat{T}_i may rely on T_i . The second term $g(t|\boldsymbol{\xi})$ is the PDF of T_i . Thus, L_i leads to the marginal PDF of \widehat{T}_i . From [24], we know that conditional on T_i , \widehat{T}_i asymptotically follows a normal distribution with mean T_i and variance u_i^2 . We have

$$u_i^2 = \left[\nabla h \left(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i \right) \right]' V_i \left[\nabla h \left(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i \right) \right],$$

where $\nabla h \left(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i \right) = \left. \frac{\partial h(\boldsymbol{\beta}, \mathbf{b}_i)}{\partial (\boldsymbol{\beta}, \mathbf{b}_i)'} \right|_{(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i)'}$, and V_i is the asymptotic covariance-variance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i)'$ evaluated at $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i)'$. One may therefore assume that $f(\widehat{T}_i|T_i = t)$ is a normal density function with mean t and variance u_i^2 . This leads to

$$L_i = \int_0^\infty \frac{1}{u_i} \phi \left(\frac{\widehat{T}_i - t}{u_i} \right) g(t|\boldsymbol{\xi})dt,$$

where $\phi(\cdot)$ is the PDF of the standard normal distribution. To gain more insight into the likelihood function L_i , we expand $g(t|\boldsymbol{\xi})$ at $t = \widehat{T}_i$ and obtain that

$$\begin{aligned} L_i &= \int_0^\infty \frac{1}{u_i} \phi \left(\frac{\widehat{T}_i - t}{u} \right) g(t|\boldsymbol{\xi})dt \\ &\approx \int_0^\infty \frac{1}{u_i} \phi \left(\frac{\widehat{T}_i - t}{u} \right) \{g(\widehat{T}_i|\boldsymbol{\xi}) + (t - \widehat{T}_i)g'(\widehat{T}_i|\boldsymbol{\xi}) + 0.5(t - \widehat{T}_i)^2 g''(\widehat{T}_i|\boldsymbol{\xi})\}dt \\ &= a_{i1}g(\widehat{T}_i|\boldsymbol{\xi}) + a_{i2}g'(\widehat{T}_i|\boldsymbol{\xi}) + a_{i3}g''(\widehat{T}_i|\boldsymbol{\xi}), \end{aligned}$$

where $a_{i1} = 1 - \Phi(-C_i)$, $a_{i2} = u_i \phi(-C_i)$, $a_{i3} = [1 - \Phi(-C_i) - C_i \phi(-C_i)] u_i^2 / 2$, $C_i = \widehat{T}_i / u_i$, and $\Phi(\cdot)$ is the CDF of the standard normal distribution. Thus, based on the pseudo lifetimes

\widehat{T}_i s, the likelihood function of $\boldsymbol{\xi}$ can be approximated as

$$\begin{aligned} L_{M2S} &= \prod_{i=1}^n \left[a_{i1}g(\widehat{T}_i|\boldsymbol{\xi}) + a_{i2}g'(\widehat{T}_i|\boldsymbol{\xi}) + a_{i3}g''(\widehat{T}_i|\boldsymbol{\xi}) \right] \\ &= \prod_{i=1}^n g(\widehat{T}_i|\boldsymbol{\xi}) \prod_{i=1}^n \left[a_{i1} + a_{i2}g'(\widehat{T}_i|\boldsymbol{\xi})/g(\widehat{T}_i|\boldsymbol{\xi}) + a_{i3}g''(\widehat{T}_i|\boldsymbol{\xi})/g(\widehat{T}_i|\boldsymbol{\xi}) \right]. \end{aligned} \quad (5)$$

Remark 1: As one can see, the first product term is the likelihood function if \widehat{T}_i s are the truly observed failure time, and the second product term arises because \widehat{T}_i is an estimate. As $m_i \rightarrow \infty$, we have $u_i \rightarrow 0$ and $C_i \rightarrow \infty$. Then $a_{i1} \rightarrow 1$, $a_{i2} \rightarrow 0$ and $a_{i3} \rightarrow 0$. Thus, L_{M2S} is reduced to be $\prod_{i=1}^n g(\widehat{T}_i|\boldsymbol{\xi})$, which is direct two-stage method. For $u_i \rightarrow 0$, the imputation step in iterative two-stage method can be neglected, which will also be reduced to be direct two-stage method. Thus, the three two-stage method will lead to the same result. For small or moderate m_i , \widehat{T}_i may be a biased estimate. Then the second product term can be treated as a correction of the bias, and can not be neglected.

Remark 2: The parameter estimate of $\boldsymbol{\xi}$, $\hat{\boldsymbol{\xi}}$, can be obtained by maximizing $l_{M2S} = \log L_{M2S}$. Under certain regularity conditions, $m_i \rightarrow \infty, i = 1, \dots, n$ and $n \rightarrow \infty$, we have

$$\sqrt{n}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \sim N(0, I^{-1}(\boldsymbol{\xi})), \quad (6)$$

where $I(\boldsymbol{\xi})$ is the Fisher information matrix of T . The proof is trivial, because the condition of $m_i \rightarrow \infty, i = 1, \dots, n$ makes that \widehat{T}_i converges to T_i in probability, and then the normality of MLE assures that (6) holds. Then for any continuous function of $\boldsymbol{\xi}$, $q(\boldsymbol{\xi})$, we have

$$\sqrt{n} \left(q(\hat{\boldsymbol{\xi}}) - q(\boldsymbol{\xi}) \right) \sim N(0, V_q(\boldsymbol{\xi})),$$

where $V_q(\boldsymbol{\xi}) = [\nabla q(\boldsymbol{\xi})]' I^{-1}(\boldsymbol{\xi}) [\nabla q(\boldsymbol{\xi})] |_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}$ and $\nabla q(\boldsymbol{\xi}) = \frac{\partial q(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}$. Then the $100(1 - \alpha)\%$ confidence interval of $q(\boldsymbol{\xi})$ can be constructed as

$$\left[q(\hat{\boldsymbol{\xi}}) - Z_{\alpha/2} \sqrt{V_q(\hat{\boldsymbol{\xi}})}, q(\hat{\boldsymbol{\xi}}) + Z_{\alpha/2} \sqrt{V_q(\hat{\boldsymbol{\xi}})} \right].$$

Remark 3: If there are some historical experiences about $g(t|\boldsymbol{\xi})$, then the statistical inference can be done for the determined lifetime distribution. Otherwise, several candidate lifetime distributions can be selected, and the distribution with the smallest Akaike's Information Criterion (AIC) value is chosen as the lifetime distribution.

Example: Lognormal distribution

Suppose that T follows a lognormal distribution ($LN(\mu, \sigma^2)$) with PDF

$$g(t|\mu, \sigma) = \frac{1}{\sqrt{2\pi}t\sigma} \exp \left\{ -\frac{(\log t - \mu)^2}{2\sigma^2} \right\}, t, \sigma > 0, \mu \in \mathcal{R}.$$

Simple algebra computations lead to

$$\begin{aligned} \frac{g'(t|\mu, \sigma)}{g(t|\mu, \sigma)} &= -\frac{1}{t} - \frac{\log t - \mu}{\sigma^2}, \\ \frac{g''(t|\mu, \sigma)}{g(t|\mu, \sigma)} &= -\frac{2}{t^2} - \frac{1}{t\sigma^2} + \frac{\log t - \mu}{t\sigma^2} + \frac{\log t - \mu}{t^2\sigma^2} + \frac{(\log t - \mu)^2}{t\sigma^4}. \end{aligned}$$

Then the procedure for the case of lognormal distribution is implemented as follows.

1. For the observed data $\{(x_{i1}, y_{i1}), \dots, (x_{im_i}, y_{im_i}), i = 1, \dots, n\}$, fit the nonlinear least-squares method to obtain the estimates $(\hat{\boldsymbol{\beta}}, \hat{\sigma}_e^2, \hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_n)$ of $(\boldsymbol{\beta}, \sigma_e^2, \mathbf{b}_1, \dots, \mathbf{b}_n)$ and the corresponding covariance-variance matrix.
2. Obtain \hat{T}_i by solving the equation $\eta(t|\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}_i) = \eta_c$, and compute its variance estimate $u_i^2, i = 1, \dots, n$.

3. Then $(\hat{\mu}, \hat{\sigma})$ can be obtained by maximizing the loglikelihood function

$$l_{M2S} = \sum_{i=1}^n \log g(\hat{T}_i | \mu, \sigma) + \sum_{i=1}^n \log \left[a_{i1} - \frac{a_{i2}}{\hat{T}_i} - \frac{2a_{i3}}{\hat{T}_i^2} - \frac{a_{i3}}{\hat{T}_i \sigma^2} + \frac{(a_{i3}\hat{T}_i + a_{i3} - a_{i2}\hat{T}_i^2)(\log \hat{T}_i - \mu)}{\hat{T}_i^2 \sigma^2} + \frac{(\log \hat{T}_i - \mu)^2}{\hat{T}_i \sigma^4} \right].$$

In fact, the first two steps are general for any lifetime distributions, and the last step depends on the assumed lifetime distribution. For gamma distribution, Weibull distribution, IG distribution and Birnbaum-Saunders (BS) distribution, we also list the results of $g'(t|\boldsymbol{\xi})/g(t|\boldsymbol{\xi})$ and $g''(t|\boldsymbol{\xi})/g(t|\boldsymbol{\xi})$, and the likelihood functions can be written similarly.

For the gamma distribution ($gamma(\beta, \lambda)$) with PDF

$$g(t|\beta, \lambda) = \frac{\lambda^\beta t^{\beta-1}}{\Gamma(\beta)} \exp\{-\lambda t\}, t, \beta, \lambda > 0,$$

we have

$$\begin{aligned} \frac{g'(t|\beta, \lambda)}{g(t|\beta, \lambda)} &= \frac{\beta - 1}{t} - \lambda, \\ \frac{g''(t|\beta, \lambda)}{g(t|\beta, \lambda)} &= [(\beta - 1)/t - \lambda]^2 + \frac{1 - \beta}{t^2}. \end{aligned}$$

For the Weibull distribution ($Wei(\gamma, \omega)$) with PDF

$$g(t|\gamma, \omega) = \frac{\gamma t^{\gamma-1}}{\omega^\gamma} \exp\{-(t/\omega)^\gamma\}, t, \gamma, \omega > 0,$$

after some calculations, we have

$$\begin{aligned} \frac{g'(t|\gamma, \omega)}{g(t|\gamma, \omega)} &= \frac{\gamma - 1}{t} - \frac{\gamma t^{\gamma-1}}{\omega^\gamma}, \\ \frac{g''(t|\gamma, \omega)}{g(t|\gamma, \omega)} &= [(\gamma - 1)/t - \gamma t^{\gamma-1}/\omega^\gamma]^2 + \frac{1 - \beta}{t^2} - \frac{\gamma(\gamma - 1)t^{\gamma-2}}{\omega^\gamma}. \end{aligned}$$

For the IG distribution ($IG(a, b)$) with PDF

$$g(t|a, b) = \left[\frac{b}{2\pi t^3} \right]^{1/2} \exp \left\{ -\frac{b(t-a)^2}{2a^2 t} \right\}, t, a, b > 0,$$

we have

$$\begin{aligned} \frac{g'(t|a, b)}{g(t|a, b)} &= -\frac{3}{2t} - \frac{bt^{1/2}}{2} \left(\frac{t}{a^2} - \frac{1}{t} \right), \\ \frac{g''(t|a, b)}{g(t|a, b)} &= \frac{15}{4t^2} - bt^{-3/2} + \frac{bt^{1/2}}{2} \left[\frac{3}{2t^{5/2}} + \frac{b(a^{-2} - t^{-2})}{2} \right] \left(\frac{t}{a^2} - \frac{1}{t} \right). \end{aligned}$$

For the BS distribution ($BS(c, d)$) with PDF

$$g(t|c, d) = \frac{1}{2\sqrt{2\pi cd}} \left[\left(\frac{d}{t} \right)^{1/2} + \left(\frac{d}{t} \right)^{3/2} \right] \exp \left\{ -\frac{t/d + d/t - 2}{2c^2} \right\}, t, c, d > 0,$$

we have

$$\begin{aligned} \frac{g'(t|c, d)}{g(t|c, d)} &= - \left[\frac{1}{2t} + \frac{3d}{2t^2} + \frac{t^{1/2}}{2c^2 d^{3/2}} - \frac{d^{1/2}}{2c^2 t^{3/2}} \right] \left[1 + \frac{d}{t} \right]^{-1}, \\ \frac{g''(t|c, d)}{g(t|c, d)} &= \left[\frac{3}{4t^2} + \frac{15d}{4t^3} - \frac{d^{1/2}}{c^2 t^{5/2}} + \frac{1}{2c^2} \left(\frac{1}{2t} + \frac{3d}{2t^2} + \frac{t^{1/2}}{2c^2 d^{3/2}} - \frac{d^{1/2}}{2c^2 t^{3/2}} \right) \left(\frac{1}{d} - \frac{d}{t^2} \right) \right] \left[1 + \frac{d}{t} \right]^{-1}. \end{aligned}$$

4 Simulation studies

In this section, we investigate the performance of the modified two-stage method. Firstly, we compare the proposed two-stage method with the other two methods under different scenarios when the lifetime distribution of the product is known. Then when the lifetime distribution of the product is unknown, and even not in the set of candidate lifetime distributions, the performance of the proposed method is studied.

4.1 The efficiency of the proposed method

The sample size n is chosen as 20 and 30, and each unit is measured m times where $m = 10$ and 20. The lifetime distribution of T is assumed to be log-normal, and the parameters are $\mu = 1$ and $\sigma = 0.25$. The path model is $y_{ij} = \theta_{1i} t_j^{\theta_{2i}} + \epsilon_{ij}$, and $\epsilon_{ij} \sim N(0, \sigma_e^2)$ for $i = 1, \dots, n$ and $j = 1, \dots, m$. The values of σ_e are assumed to be 3. Due to the heterogeneity among the products, we assume that θ_{1i} and θ_{2i} are random. The critical value η_c is taken as 50. The degradation data are generated as follows.

1. The measurement times are $(x_1, x_2, \dots, x_m) = (q_{0.3}/m, 2 * q_{0.3}/m, \dots, (m - 1) * q_{0.3}/m, q_{0.3})$, where $q_{0.3}$ denotes the 0.3 quantile of $LN(\mu, \sigma^2)$.
2. Generate n failure times T_1, \dots, T_n from $LN(\mu, \sigma^2)$, and $\theta_{11}, \dots, \theta_{1n}$ from the distribution of θ_{1i} , where three different distribution types of θ_{1i} are examined for illustration. That is, log-normal ($LN(2, 0.1^2)$), gamma distribution ($gamma(100, 13.5)$) and Weibull distribution ($Wei(5, 8)$).
3. The values of $\theta_{21}, \dots, \theta_{2n}$ are determined by $\eta_c = \theta_{1i} * T_i^{\theta_{2i}}$, that is, $\theta_{2i} = \{\log(\eta_c) - \log(\theta_{1i})\} / \log(T_i)$.
4. Generate ϵ_{ij} from $N(0, \sigma_e^2)$; then $y_{ij} = \theta_{1i} * x_j^{\theta_{2i}} + \epsilon_{ij}$, $i = 1, \dots, n$ and $j = 1, \dots, m$.

Each combination of m , n and the distribution of θ_{1i} is repeated 3000 times. In the simulation, we compare the proposed two-stage method with direct two-stage method and iterative two-stage method. The estimates of μ , σ , $q_{0.05}$, $q_{0.95}$ and the mean time to failure (MTTF) are obtained for each sample, where the true values of μ , σ , $q_{0.05}$, $q_{0.95}$ and MTTF are 1.00, 0.25, 1.801, 4.100 and 2.804, respectively. Then the mean, square root of mean squared errors (RMSE) and 95% coverage probabilities (CP) based on 3000 estimators are computed. The results for $\theta_{1i} \sim LN(2, 0.1^2)$ are listed in Tables 1-3. The results for $\theta_{1i} \sim gamma(100, 13.5)$ and $\theta_{1i} \sim Wei(5, 8)$ are similar, and we provide them as the supplementary material. For each combination with 3000 repetitions, the CPU computing times (seconds) for the three

Table 1: The mean estimates of the parameters for different (m, n) based on 3000 replications when $\theta_{1i} \sim LN(2, 0.1^2)$.

(m, n)	Method	μ	σ	$q_{0.05}$	$q_{0.95}$	MTTF
(10,20)	Bias reduction two-stage	1.000	0.247	1.815	4.042	2.805
	Iterative two-stage	1.008	0.252	1.815	4.176	2.838
	Direct two-stage	1.014	0.258	1.809	4.253	2.862
(20,20)	Bias reduction two-stage	1.000	0.250	1.807	4.056	2.805
	Iterative two-stage	1.003	0.251	1.811	4.150	2.824
	Direct two-stage	1.006	0.249	1.822	4.144	2.829
(10,30)	Bias reduction two-stage	0.997	0.247	1.811	4.081	2.801
	Iterative two-stage	1.004	0.258	1.789	4.192	2.828
	Direct two-stage	1.010	0.267	1.776	4.286	2.855
(20,30)	Bias reduction two-stage	0.995	0.252	1.818	4.083	2.803
	Iterative two-stage	0.999	0.252	1.797	4.127	2.809
	Direct two-stage	1.002	0.253	1.800	4.144	2.817

methods (the proposed method, direct two-stage method and iterative two-stage method) are 13.08s, 12.35s and 38.56s, respectively. The time efficiency of the proposed method is close to that of direct two-stage method, and is much better than that of iterative two-stage method. This is not surprising, because iterative two-stage method is an iterative method, and the proposed method and direct two-stage method obtain the estimates by direct maximization.

From Table 1, we see that the mean estimates based on the three methods are all close to the true values for all cases, which indicates that the two-stage methods are efficient. As can be seen in Table 2, for a fixed m or n , RMSE of the estimates are improved greatly as n or m increases. However, the effects of n on the estimates are greater than those of m . For example, compared with $(m, n) = (10, 20)$, the RMSE of the case $(10, 30)$ are much smaller than these of $(20, 20)$. This is reasonable, because the precision of pseudo lifetimes depends on the measurement times m and the precision of the parameter estimation in lifetime distribution of T is correlated with the sample size n . Among the three methods, the proposed two-stage method performs the best, because the RMSE based on the proposed

Table 2: The RMSE of the estimates for different (m, n) based on 3000 replications when $\theta_{1i} \sim LN(2, 0.1^2)$.

(m, n)	Method	μ	σ	$q_{0.05}$	$q_{0.95}$	MTTF
(10,20)	Bias reduction two-stage	0.059	0.049	0.168	0.453	0.178
	Iterative two-stage	0.061	0.049	0.163	0.476	0.188
	Direct two-stage	0.065	0.054	0.165	0.573	0.210
(20,20)	Bias reduction two-stage	0.057	0.046	0.165	0.415	0.168
	Iterative two-stage	0.058	0.049	0.164	0.459	0.177
	Direct two-stage	0.059	0.050	0.165	0.474	0.182
(10,30)	Bias reduction two-stage	0.046	0.041	0.135	0.371	0.138
	Iterative two-stage	0.048	0.041	0.132	0.387	0.146
	Direct two-stage	0.050	0.049	0.140	0.486	0.165
(20,30)	Bias reduction two-stage	0.045	0.035	0.127	0.312	0.130
	Iterative two-stage	0.044	0.037	0.127	0.328	0.130
	Direct two-stage	0.044	0.038	0.129	0.340	0.131

Table 3: The 95% coverage probabilities of the estimates for different (m, n) based on 3000 replications when $\theta_{1i} \sim LN(2, 0.1^2)$.

(m, n)	Method	μ	σ	$q_{0.05}$	$q_{0.95}$	MTTF
(10,20)	Bias reduction two-stage	0.935	0.897	0.920	0.896	0.930
	Iterative two-stage	0.925	0.897	0.917	0.896	0.927
	Direct two-stage	0.920	0.887	0.920	0.886	0.905
(20,20)	Bias reduction two-stage	0.933	0.897	0.940	0.887	0.935
	Iterative two-stage	0.927	0.897	0.935	0.890	0.930
	Direct two-stage	0.927	0.887	0.940	0.883	0.927
(10,30)	Bias reduction two-stage	0.946	0.903	0.969	0.908	0.933
	Iterative two-stage	0.954	0.903	0.961	0.908	0.938
	Direct two-stage	0.946	0.897	0.946	0.897	0.923
(20,30)	Bias reduction two-stage	0.943	0.927	0.937	0.910	0.947
	Iterative two-stage	0.950	0.927	0.933	0.910	0.950
	Direct two-stage	0.953	0.903	0.930	0.907	0.950

method are the smallest for most of the cases, see the bold numbers in Table 2. For example, for the parameter μ with the case $(m, n) = (10, 20)$, the RMSE based on the proposed method is 0.059, which is smaller than that based on the other two methods. Table 3 lists the 95% coverage probabilities of the interval estimates for different values of (m, n) . The coverage probabilities based on the three methods for all cases are close to each other, because the interval estimates are all constructed by the asymptotic normality result (6).

4.2 The effects of the test termination time and misspecification

Under the model setting as the previous subsection, the effects of the test termination time (TTT) on estimation performance are investigated. We choose the TTT as the α quantile of $LN(\mu, \sigma^2)$, where $\alpha = 0.1, 0.2, \dots, 0.9$. The random effect variable $\theta_{1i} \sim LN(2, 0.1^2)$. The sample size n and the measurement times m are 30 and 20, respectively. We repeat the simulation study for each TTT case 3000 times. The results of RMSEs and CPs under different TTTs are shown in Figures 1 and 2. Some interesting results can be summarized from Figures 1 and 2.

1. For all cases, the RMSEs based on the proposed method are smaller than these based on direct two-stage method.
2. The proposed method and iterative two-stage method are comparable. For the cases of small TTTs or small α s, the proposed method performs better than iterative two-stage method according to RMSE, while for large α s, iterative two-stage method is better. However, the TTT is usually much smaller than MTTF in practice, which means that small TTT or α will be chosen. Then the proposed method will be preferred.
3. When $\alpha = 0.1$, the CPs do not perform well for the parameters σ and $q_{0.95}$. However, for the other cases, the CPs are close to the nominal level 0.95.

The above simulation studies are based on the known lifetime distribution of the product. However, we may misspecify the lifetime distribution in practice. The data generation scheme

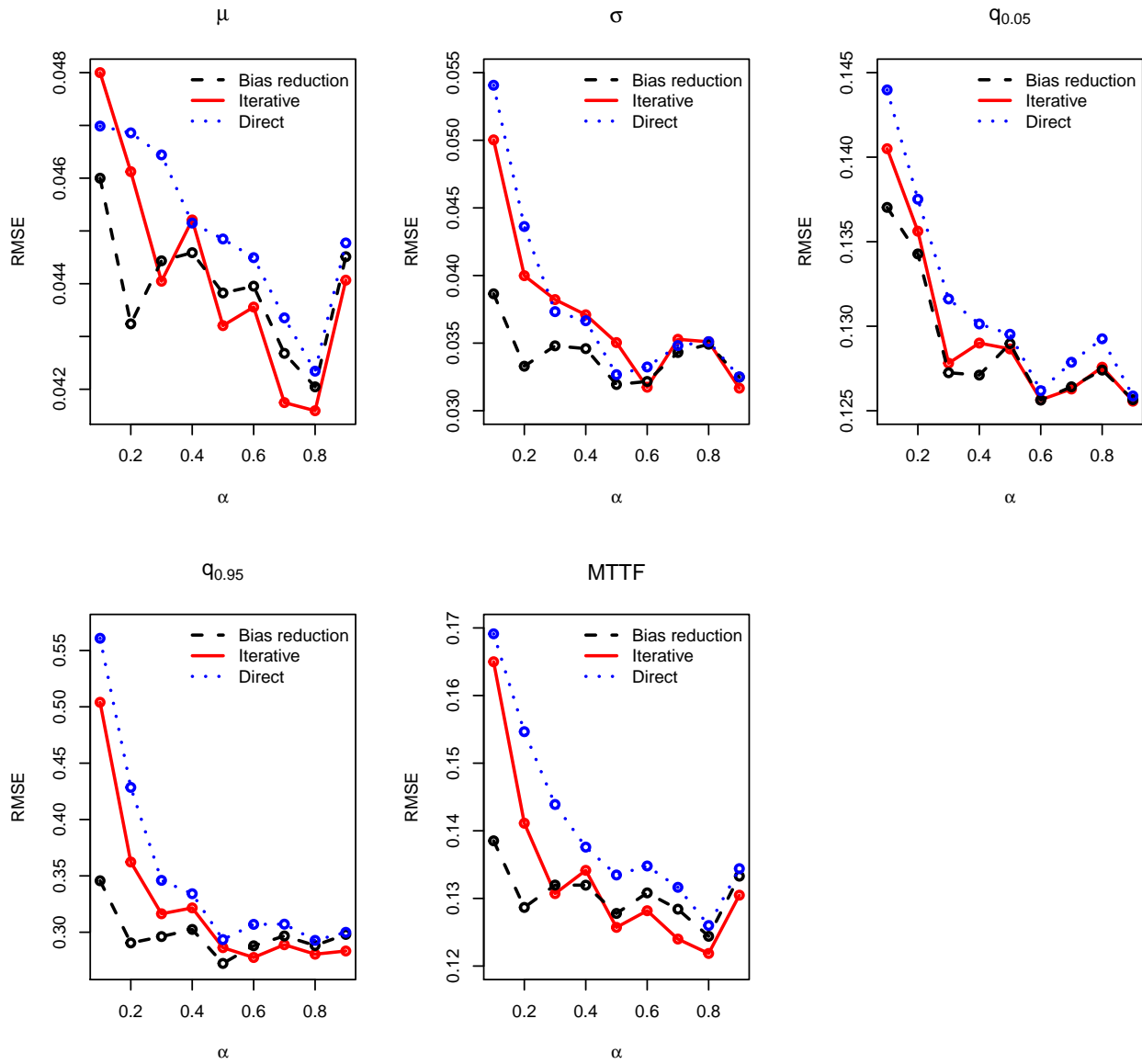


Figure 1: RMSE of the estimates with different test termination times.

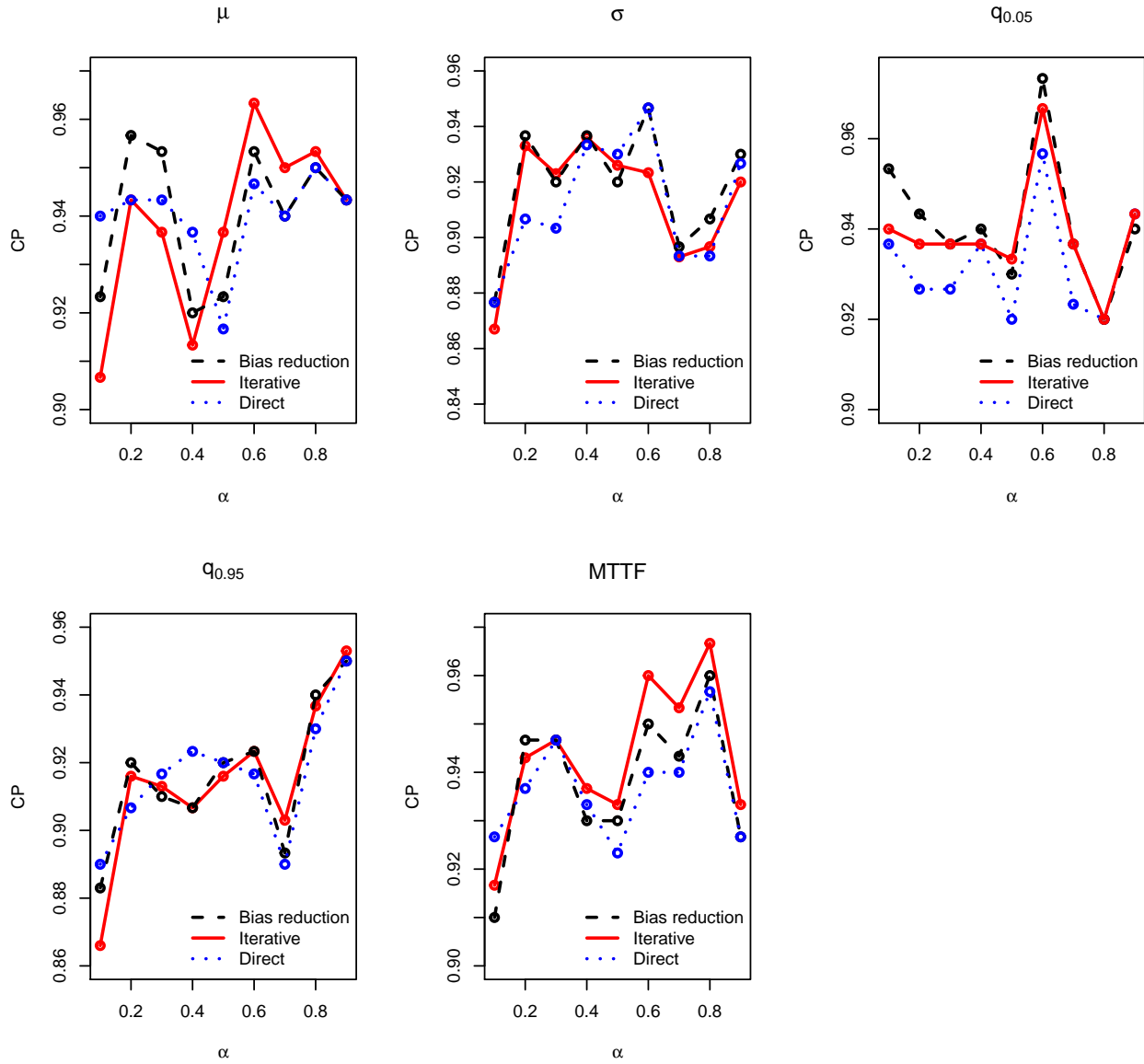


Figure 2: Coverage probabilities of the estimates with different test termination times.

Table 4: The mean and RMSE of the estimates for different test termination times with the true values $(q_{0.05}, q_{0.95}, MTTF)=(1.801, 4.100, 2.804)$.

α	Estimate	$q_{0.05}$	$q_{0.95}$	MTTF
0.1	Mean	1.703	4.101	2.788
	RMSE	0.181	0.490	0.163
0.2	Mean	1.757	4.043	2.798
	RMSE	0.143	0.347	0.139
0.3	Mean	1.751	4.004	2.778
	RMSE	0.135	0.307	0.126
0.4	Mean	1.772	4.023	2.799
	RMSE	0.140	0.324	0.135
0.5	Mean	1.79	3.984	2.794
	RMSE	0.120	0.315	0.127
0.6	Mean	1.787	4.004	2.800
	RMSE	0.132	0.304	0.129
0.7	Mean	1.786	4.027	2.809
	RMSE	0.140	0.293	0.137
0.8	Mean	1.788	3.989	2.795
	RMSE	0.125	0.305	0.129
0.9	Mean	1.779	3.994	2.792
	RMSE	0.122	0.288	0.125

is the same as the previous part ($m = 20, n = 20, \alpha = 0.1, 0.2, \dots, 0.9$, 3000 repetitions). The candidate lifetime distributions of the product are selected as the Weibull distribution, gamma distribution, inverse Gaussian distribution and BS distribution, respectively. Thus, the candidate lifetime distributions do not contain the true lifetime distribution (lognormal distribution). For each data set, we choose the optimal lifetime distribution by the AIC values, then $q_{0.05}$, $q_{0.95}$ and MTTF of the product are estimated. The results are listed in Table 4. From Table 4, we can see that although the true lifetime distribution is not in the set of candidate lifetime distributions, the estimated quantile lifetimes and MTTF are close to the true values for all cases, which means that the proposed method is robust to the choice of candidate lifetime distributions.

5 Real Data Analysis

We illustrate the new method in this section by analyzing two real data sets: (i) the laser degradation data presented in Table C.17 on page 642 in [20], and (ii) the fatigue-crack-growth data in Table 1 on page 162 in [21].

5.1 Laser degradation data

The laser degradation data from [20] is a classic dataset. The data shows the increase in operating current over time for a sample of GaAs lasers tested at $80^\circ C$. When there is a 10% increase in the operating current, the device is considered to have failed. There are totally 15 devices tested, and all the devices are measured every 250 hours. The test terminal time is 4000 hours, and thus $m_1 = \dots = m_{15} = 16$. Following the analysis of [20], we also assume that the path model is

$$y_{ij} = \theta_i x_j + \epsilon_{ij}, i = 1, \dots, 15, j = 1, \dots, 16.$$

The log-likelihoods are computed for the following distributions: Log-normal, Weibull, Gamma, IG and BS, and the AIC values are -121.31, -123.68, -121.39, -121.17 and -121.16, respectively. AIC prefers to choose Weibull distribution, since it leads to the smallest AIC value. Figure 3 show the five fitted distributions and the empirical distribution based on pseudo failure times (black dots). As can be seen in Figure 3, Weibull distribution is apt to fit the last 11 pseudo failure times and neglect the former 4 pseudo failure times, while the other distributions prefer a trade-off between the two parts.

Then based on Weibull distribution, we analyze the data by the proposed modified two-stage method, direct two-stage method and iterative two-stage method. Since $C_i = \hat{T}_i/u_i$ is large enough, the parameter estimation based on the three methods are the same as we indicate in **Remark 1**. The estimates of the parameters are $\hat{\gamma} = 6.599$ and $\hat{\omega} = 5482.46$. The standard deviation of $\hat{\gamma}$ and $\hat{\omega}$ are 1.385 and 220.67, respectively. The 95% confidence

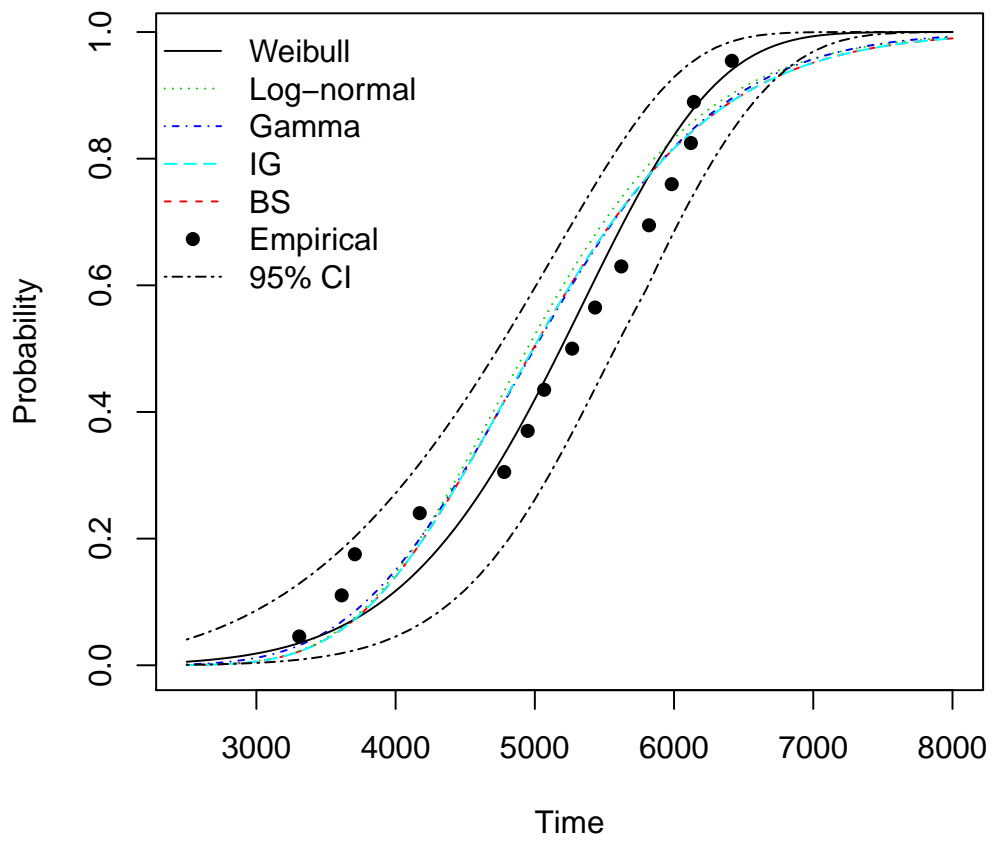


Figure 3: CDF estimates for different distributions, and the associated 95% pointwise confidence interval based on the laser data.

interval (CI) estimates of γ and ω are (3.83, 9.37) and (5041.12, 5923.79), respectively. The 95% pointwise CI of CDF are shown in Figure 3. The 95% pointwise CI covers all the pseudo lifetimes, and thus the Weibull distribution performs well.

5.2 Fatigue-crack-growth data

The fatigue-crack-growth data contains the degradation measurements on 21 test units. The crack length of each unit was recorded once every 0.01 million cycles over a period of 0.12 million cycles. The initial crack length of every unit is the same as 0.9 inches. A unit is considered to have failed if the crack length exceeds 1.6 inches. The path model for the fatigue-crack-growth is derived from the Paris Law in material science and is given below:

$$y_{ij} = \frac{1}{\theta_{2i}} \log (1 - 0.9^{\theta_{2i}} \theta_{1i} \theta_{2i} x_j) + \epsilon_{ij},$$

where $i = 1, \dots, 21$; $j = 1, \dots, m_i$. The fatigue-crack-growth data has been analyzed by many authors using different methods. Lu and Meeker (1993) analyzed the data using their two-stage least squares method. [24] analyzed the data directly using ordinary and weighted least squares methods by treating the path model as a mixed effect regression model. The maximum likelihood method and a Bayesian method for the analysis of the data were explored by [25]. [25] also made comparisons among direct two-stage method, the maximum likelihood method and the Bayesian method. All these methods provided similar conclusions in this application. The common feature of these methods is that an assumption on the distribution of the random parameters θ_{1i} and θ_{2i} was made.

Here, we re-analyze the fatigue-crack-growth data using the new method. The log-likelihoods with parameters estimated by the method-of-moments are computed for the following distributions: Log-normal, Weibull, Gamma, IG and Birnbaum-Saunders. Their AIC values are -96.22, -92.80, -95.82, -95.56 and -95.30, respectively. Since the log-likelihood of the Log-normal distribution is the largest, the Log-normal distribution is taken as the

assumed lifetime distribution. We use the following methods to compute the result.

1. Direct two-stage method: $\hat{\mu} = -2.124$; $\hat{\sigma}^2 = 0.183^2$
2. Iterative two-stage method: $\hat{\mu} = -2.102$; $\hat{\sigma}^2 = 0.182^2$
3. Bias reduction two-stage method: $\hat{\mu} = -2.103$; $\hat{\sigma}^2 = 0.180^2$

The standard deviation of $\hat{\mu}$ and $\hat{\sigma}^2$ are 0.0393 and 9.563×10^{-3} , respectively. Then based on the proposed method, the 95% CI estimates of μ and σ^2 are (-2.180, -2.025) and (0.0137, 0.0511), respectively. In Figure 4, we plotted the three estimated CDF curves and the 95% pointwise CI based on the proposed two-stage method. The 95% pointwise confidence band covers all the pseudo lifetimes, which indicates that the log-normal distribution fits the data well.

6 Conclusion

In this paper, we have proposed a new two-stage method for analyzing the degradation data. The key step is to expand the asymptotic likelihood function for the pseudo lifetimes \hat{T}_i , where the expanded likelihood is simple. Compared with the existing methods (e.g., iterative two-stage method, direct two-stage method), the newly proposed two-stage method has smaller RMSE than the existing two-stage methods. The computational efficiency of the proposed method is close to that of direct two-stage method, and is about two times faster than that of iterative two-stage method. The results based on the new method is robust to the choice of test termination times and candidate lifetime distributions. Moreover, the newly proposed two-stage method does not require any distributional assumptions of the random effects but some regularity conditions. It is not clear if random effects can lead to better estimation. However, the new method also has its limitations. For example, the bias in the first stage depends on the number of measurement time m . For small m , e.g., $m < 5$, the method as well as the other two-stage methods may fail.

Our future work will investigate stochastic degradation model with random effects, where

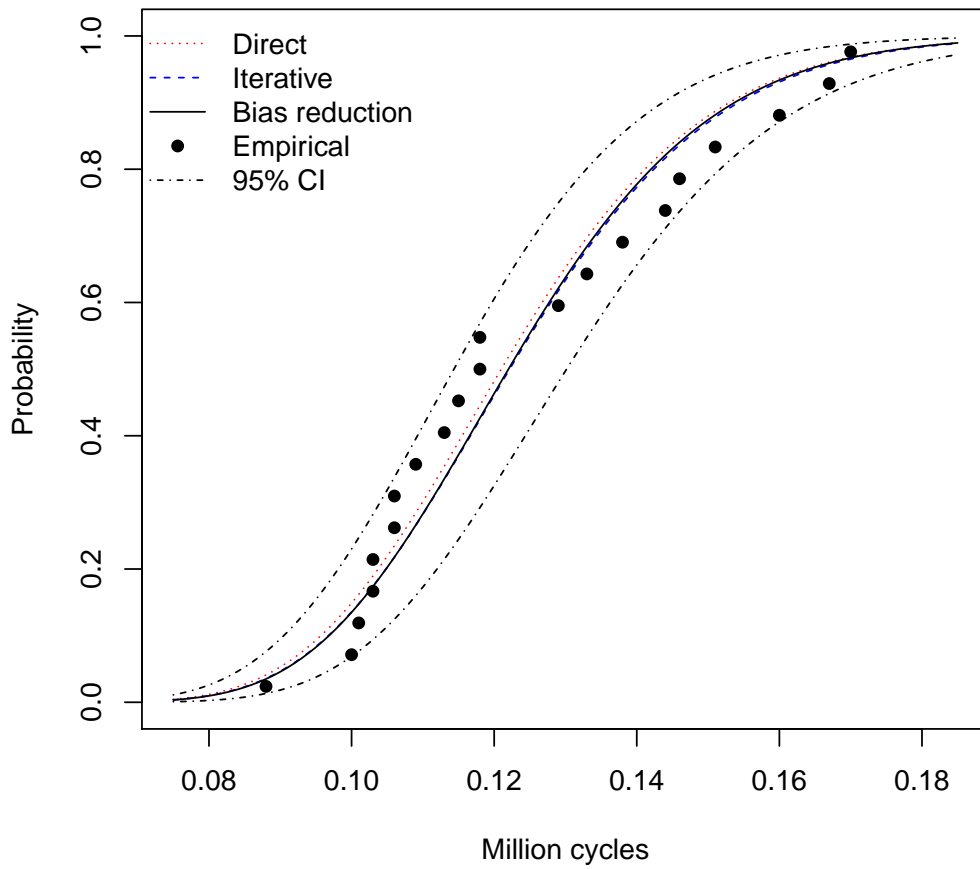


Figure 4: CDF estimates for different methods, and and the associated 95% pointwise CI based on the fatigue-crack-growth data.

the CDF of random effects is unknown. Besides, as we have discussed in the introduction, the CDF derived from the stochastic degradation model may contradict the prior information, then developing an information fusion model is also of interest. More interestingly, degradation models with unknown path function is also a challenging problem for further study.

Acknowledgements

The research is supported by Natural Science Foundation of China under grant number 11671303, Natural Science Foundation of Zhejiang province under grant number Y19A010054.

References

- [1] C. Lu, W. William, L. Escobar, A comparison of degradation and failure time analysis methods for estimating a time-to-failure distribution, *Statistica Sinica*, 6(1996), 531-546.
- [2] Y. Hong, Y. Duan, W. Meeker, D. Stanley, X. Gu, Statistical methods for degradation data with dynamic covariates information and an application to outdoor weathering data, *Technometrics*, 57(2015), 180-193.
- [3] Z. Ye, N. Chen, The inverse Gaussian process as a degradation model, *Technometrics*, 56(2014), 302-311.
- [4] Q. Guan, Y. Tang, A. Xu, Objective Bayesian analysis for accelerated degradation test based on Wiener process models, *Applied Mathematical Modelling*, 40(2016), 2743-2755.
- [5] J. Kharoufeh, C. Solo, M. Ulukus, Semi-Markov models for degradation-based reliability, *IIE Transactions*, 42(2010), 599-612.
- [6] E. Byon, Y. Ding, Season-dependent condition-based maintenance for a wind turbine

- using a partially observed Markov decision process, *IEEE Transactions on Power Systems*, 25(2010), 823-834.
- [7] Y. Xing, E. Ma, K. Tsui, M. Pecht, An ensemble model for predicting the remaining useful performance of lithium-ion batteries, *Microelectronics Reliability*, 53(2013),811-820.
- [8] Z. Ye, M. Xie, Stochastic modelling and analysis of degradation for highly reliable products, *Applied Stochastic Models in Business and Industry*, 31(2015), 16-32.
- [9] Z. Zhang, X. Si, C. Hu, Y. Lei, Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods, *European Journal of Operational Research*, 271(2018), 775-796.
- [10] D. He, Y. Wang, G. Chang, Objective Bayesian analysis for the accelerated degradation model based on the inverse gaussian process, *Applied Mathematical Modelling*, 61(2018), 341-350.
- [11] S. Hao, J. Yang, B. Christophe, Degradation analysis based on an extended inverse Gaussian process model with skew-normal random effects and measurement errors, *Reliability Engineering and System Safety*, 189(2019), 261-270.
- [12] P. Jiang, B. Wang, F. Wu, Inference for constant-stress accelerated degradation test based on Gamma process, *Applied Mathematical Modelling*, 67(2019), 123-134.
- [13] R. Oliveira, R. Loschi, M. Freitas, Skew-heavy-tailed degradation models: an application to train wheel degradation, *IEEE Transactions on Reliability*, 67(2018), 129-141.
- [14] T. Yuan, X. Wu, S. Bae, X. Zhu, Reliability assessment of a continuous-state fuel cell stack system with multiple degrading components, *Reliability Engineering and System Safety*, 189(2019), 157-164.

- [15] P. Wang, Y. Tang, S. Bae, Y. He, Bayesian analysis of two-phase degradation data based on change-point Wiener process, *Reliability Engineering and System Safety*, 170(2018), 244-256.
- [16] W. Peng, Y. Li, Y. Yang, S. Zhu, H. Huang, Bivariate analysis of incomplete degradation observations based on inverse Gaussian processes and copulas, *IEEE Transactions on Reliability*, 65(2016), 624-639.
- [17] A. Xu, L. Shen, B. Wang, Y. Tang, On modeling bivariate Wiener degradation process, *IEEE Transactions on Reliability*, 67(2018), 897-906.
- [18] J. Pinheiro, D. Bates, *Mixed-effects models in S and S-PLUS*, Springer, 2000.
- [19] A. Hartford, M. Davidian, Consequences of misspecifying assumptions in nonlinear mixed effects models, *Computational Statistics and Data Analysis*, 34(2000), 139-164.
- [20] W. Meeker, L. Escobar, *Statistical methods for reliability data*, John Wiley & Sons, 1998.
- [21] C. Lu, W. Meeker, Using degradation measures to estimate a time-to-failure distribution, *Technometrics*, 35(1993), 161-174.
- [22] Z. Chen, S. Zheng, Lifetime distribution based degradation analysis, *IEEE Transactions on Reliability*, 54(2005), 3-10.
- [23] D. Cox, Some Remarks on failure-times, surrogate markers, degradation, wear, and the quality of life, *Lifetime Data Analysis*, 5(1999), 307-314.
- [24] S. Wu, J. Shao, Reliability analysis using the least squares method in nonlinear mixed-effect degradation models, *Statistica Sinica*, 9(1999), 855-877.
- [25] M. Robinson, M. Crowder, Bayesian methods for a growth-curve degradation model with repeated measures, *Lifetime Data Analysis*, 6(2000), 357-374.