Statistica Sinica **9**(1999), 855-877

RELIABILITY ANALYSIS USING THE LEAST SQUARES METHOD IN NONLINEAR MIXED-EFFECT DEGRADATION MODELS

Shuo-Jye Wu and Jun Shao

Tamkang University and University of Wisconsin-Madison

Abstract: We develop statistical inference procedures in assessing product reliability based on a nonlinear mixed-effect degradation model and the least squares method. With today's high technology, some life tests result in no or very few failures by the end of test. Thus, it is hard to use the traditional reliability analysis to analyze lifetime data. Since product performance degrades over time, we analyze the degradation data and use the analytical results to estimate percentiles of the failure time distribution. The nonlinear mixed-effect degradation model provides us a way to build the relationship between degradation measurements and time. We establish asymptotic properties of the ordinary and weighted least squares estimators under the nonlinear mixed-effect model. We use these asymptotic results to obtain point estimates and approximate confidence intervals for percentiles of the failure time distribution. Two real data sets are analyzed. Performances of the proposed method are studied by simulation.

Key words and phrases: Asymptotic covariance matrix, asymptotic normality, consistency, failure time distribution, percentile.

1. Introduction

Censoring is very common in life tests. Censoring usually applies when exact lifetimes are known for only a portion of the products and the remainder of the lifetimes are known only to exceed certain values under a life test. With today's high technology, however, many products are designed to work without failure for years. Thus, some life tests result in few or no failures in a short life testing time. In such cases, it is difficult to analyze lifetime data with traditional reliability studies. This circumstance applies to a variety of materials and products such as metals, insulations, semiconductors, building materials, nuclear reactor materials, electrical devices, food and drugs. One approach to solve this problem is to accelerate the life of products by increasing the levels of stress via manipulating use-rate, temperature, voltage, or humidity. Lifetime data obtained under accelerated conditions is used to assess product reliability under normal conditions. A model on the relationship between the lifetime under normal conditions and those under accelerated life conditions is required for this approach.

An alternative approach developed recently is to assess product reliability using degradation measurements of product performance over time. Product performance is usually measured in terms of physical properties. We call these kinds of physical properties, "degradation mechanisms". From an engineering point of view, the common degradation mechanisms include fatigue, creep, cracks, wear, corrosion, oxidation, and weathering. Examples are loss of tread on rubber tires and degradation of the active ingredient of a drug as a result of chemical reactions with oxygen and water, and by microbial growth, etc. For some products, it is possible to obtain degradation measurements over time and these measurements may provide useful information to assess reliability. In a degradation test, one obtains measurements of degradation at various time periods, prespecifies a threshold level of degradation, and identifies failure as the time when the amount of actual degradation for a product unit exceeds the threshold level. Some motivation can be found, for example, in Nelson (1981), Lu and Meeker (1993), Tseng, Hamada and Chiao (1995) and Lu, Park and Yang (1997).

Although DT is better than ALT in analyzing highly reliable products, the literature on DT is not abundant. Nelson ((1990), Chap. 11) reviewed the degradation literature, provided general background on degradation models, and discussed some simple linear degradation models. See also Carey and Koenig (1991), Doksum (1991), Lu and Meeker (1993), Boulanger and Escobar (1994), Shao and Chow (1994), Tseng, Hamada and Chiao (1995) and Lu, Park and Yang (1997).

The main purposes of our study are to describe how the data from degradation tests can be used to assess reliability using the least squares method, and to study theoretical properties of our estimation method.

In Section 2, we describe a nonlinear mixed-effect model for degradation data. Estimators of percentiles of the failure time distribution are obtained using the relation between the failure time distribution and the degradation distribution. Section 3 is devoted to the study of the ordinary least squares estimators of the parameters in the proposed nonlinear mixed-effect degradation model. Under some regularity conditions, we show that the ordinary least squares estimators are consistent and asymptotically normal. Based on the asymptotic results, we construct approximate confidence intervals for percentiles of the failure time distribution. In Section 4, we consider the weighted least squares method, which may provide more accurate results than the ordinary least squares method. In Section 5, the proposed methods are applied to two different data sets. Some simulation results are also presented. Some discussions are in Section 6.

2. Degradation Models

In a degradation test, product performance is obtained as it degrades over time and different product units may have different performance. Thus, a statistical model for a degradation test consists of (1) a relationship between degradation measurement and time, and (2) a distribution that describes an individual product unit's characteristics. The general approach is to model the degradation of the individual components using the same functional form, and the differences between individual components using random effects. Carey and Koenig (1991) considered the following sigmodial growth model: $y_i(t) = \alpha_i[1-\exp(-(\beta_i t)^\gamma)]+\varepsilon_i(t)$, where $y_i(t)$ is the degradation of the ith unit up to time t, α_i and β_i are both random effects, γ is fixed, $\varepsilon_i(t)$ is the measurement error. Lu, Meeker and Escobar (1996) used the following simple path model to describe this phenomenon: $y = \theta + t + \varepsilon$. Here $y = \log(\text{observed degradation})$, $t = \log(\text{time})$, θ is random effect, and ε is the measurement error. Meeker, Escobar and Lu (1998) described degradation reliability models that correspond to physical-failure mechanisms. They proposed that the observed sample degradation path for a unit is a unit's actual degradation path plus error.

In this paper, we will focus on the following nonlinear random effects model considered by Lu and Meeker (1993):

$$y_{ij} = \eta(t_{ij}; \boldsymbol{\alpha}, \boldsymbol{\beta}_i) + \varepsilon_{ij}, \quad i = 1, \ldots, n, \;\; j = 1, \ldots, m_i \le s, \tag{1}$$

where η is the actual level of degradation of the component under study and is a given function nonlinear in $(\boldsymbol{\alpha}, \boldsymbol{\beta}_i)$; t_{ij} is the time of the jth measurement for the ith unit; y_{ij} represents the level of degradation actually observed at time t_{ij}; $\boldsymbol{\alpha}$ denotes the vector of fixed-effect parameters; $\boldsymbol{\beta}_i$ represents the vector of the ith unit random effects; ε_{ij}'s are *i.i.d.* measurement errors with mean 0 and variance σ_ε^2; s is the prespecified largest number of measurements for all units. We assume that $\{\varepsilon_{ij}\}$ and $\{\boldsymbol{\beta}_i\}$ are independent and $E[\eta(t_{ij}; \boldsymbol{\alpha}, \boldsymbol{\beta}_i)\eta(t_{i'j'}; \boldsymbol{\alpha}, \boldsymbol{\beta}_{i'})] < \infty$ for all $i, i' = 1, \ldots, n$, $j, j' = 1, \ldots, m_i$ or $m_{i'}$.

In general, there are three kinds of data in degradation tests:

1. No failure occurs in a test.
2. Some failures occur in a test and the degradation amounts of these failed units can be observed, even if they have already reached the critical value (the example in Lu and Meeker (1993) motivates this).
3. Some failures occur in a test, but the degradation amounts of these failed units cannot be observed. In this case, the sample sizes m_i are random due to the occurrence of failures.

Since degradation tests are designed for products with long lives, the first case occurs most frequently in degradation tests. There is no real difference in our analysis between the first and the second case. For the third case, if the number of failures is small, we consider analysis conditional on the random sample sizes m_i. However, if the number of failures is large, we need to use a model that can accommodate information provided by the failures. Such a model is different from model (1) and will not be discussed in this paper.

Suppose that η in (1) is a continuous and differentiable function for any fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. We assume that the degradation is not reversible. Hence, without loss of generality, we assume that $\eta(t;\boldsymbol{\alpha},\boldsymbol{\beta})$ is a strictly increasing function of t. The component degrades over time and, when actual degradation η reaches a prespecified critical value η_c, failure occurs (e.g., see Nelson (1990), Doksum (1991), Lu and Meeker (1993), Tseng, Hamada and Chiao (1995) and Lu, Meeker and Escobar (1996)). Therefore, the failure time of a product, denoted by $T = T(\eta_c)$, is equal to the solution of $\eta(t;\boldsymbol{\alpha},\boldsymbol{\beta}) = \eta_c$. Furthermore, we assume that there exist constants η_1 and η_2 satisfying $\eta_2 < \eta_c < \eta_1$ such that $P[\ \lim_{t\to 0}\eta(t;\boldsymbol{\alpha},\boldsymbol{\beta}) < \eta_2] = 1$ and $P[\ \lim_{t\to\infty}\eta(t;\boldsymbol{\alpha},\boldsymbol{\beta}) > \eta_1] = 1$; that is, no failure occurs before the test starts and the lifetime of a unit is finite. Under these assumptions, $\eta(t;\boldsymbol{\alpha},\boldsymbol{\beta}) = \eta_c$ has a unique and finite solution for any given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

There is an important relation between the failure time distribution and the degradation distribution that is useful for estimating percentiles of the failure time distribution.

Lemma 1. *Let $F_\eta(x|t)$ be the degradation distribution of η given time t, and $F_T(t|x)$ be the failure time distribution of $T = T(x)$ given degradation value x. Under the previously described assumptions, $F_T(t|x) = 1 - F_\eta(x|t)$.*

The percentiles of the failure time distribution are basic to reliability analysis. Let t_p be the 100pth percentile of the failure time distribution. One can obtain percentiles of the failure time distribution using Lemma 1, *i.e.,* t_p can be obtained by solving $1 - F_\eta(\eta_c|t) = p$ in t. Suppose that the solution is $t_p = g(\eta_c, \boldsymbol{\theta})$, where g is a known function and $\boldsymbol{\theta}$ is a vector of unknown parameters. If $\boldsymbol{\theta}$ is estimated by $\hat{\boldsymbol{\theta}}_n$ based on y_{ij}'s, then t_p is estimated by $\hat{t}_p = g(\eta_c, \hat{\boldsymbol{\theta}}_n)$. In some cases (see Example 1), g is an explicit function of η_c and $\boldsymbol{\theta}$, and thus $\hat{t}_p$ has a closed form. Otherwise, the problem has to be solved numerically.

Example 1. Consider the exponential growth curve $\eta(t;\alpha_0,\alpha_1,\beta) = \alpha_0 - \alpha_1 e^{-\beta t}$, where $\alpha_0 > 0$ and $\alpha_1 > 0$ are fixed effect parameters, β is a random effect distributed as $N(\mu,\sigma^2)$ with $\sigma \ll \mu$ so that $P[\beta \le 0]$ is negligible, t is the measurement time. Then

$$F_\eta(\eta_c|t) = P\left[\beta \le \frac{-1}{t}\ln\left(\frac{\alpha_0 - \eta_c}{\alpha_1}\right)\right] = \Phi\left(\frac{\frac{-1}{t}\ln\left[\frac{\alpha_0-\eta_c}{\alpha_1}\right] - \mu}{\sigma}\right),$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution. Setting $\Phi((-t^{-1}\ln[(\alpha_0 - \eta_c)/\alpha_1] - \mu)/\sigma) = 1 - p$, we find

$$t_p = \frac{\ln\left(\frac{\alpha_1}{\alpha_0-\eta_c}\right)}{\mu + z_p\sigma},$$

where z_p is the upper p percentile of the standard normal distribution. Thus, one can estimate t_p following estimation of α_0, α_1, μ and σ.

It is essential to estimate $\boldsymbol{\theta}$. In the next two sections, we show how this can be done through fitting model (1), using the least squares method.

3. The Ordinary Least Squares Estimators

If the distributions of $\boldsymbol{\beta}_i$ and ε_{ij} in (1) are in parametric families, then we may apply likelihood-based methods. Under model (1), however, likelihood-based methods involve difficult computations. We consider the least squares method, assuming that the distribution of $\boldsymbol{\beta}$ is in a parametric family but without imposing any distributional assumption on ε_{ij}. Suppose that the distribution of $\boldsymbol{\beta}$ is $\pi(\boldsymbol{\beta}|\boldsymbol{\phi})$, where π is a known function and $\boldsymbol{\phi}$ is an unknown $q \times 1$ parameter vector. Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\phi})'$ and $h(t_{ij}; \boldsymbol{\theta}) = E_{\boldsymbol{\beta}}[\eta(t_{ij}; \boldsymbol{\alpha}, \boldsymbol{\beta})]$. Define $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im_i})'$, $\boldsymbol{h}_i(\boldsymbol{\theta}) = (h(t_{i1}; \boldsymbol{\theta}), \ldots, h(t_{im_i}; \boldsymbol{\theta}))'$ and $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{im_i})'$. Then, model (1) can be written as the following heteroscedastic nonlinear model:

$$\boldsymbol{y}_i = \boldsymbol{h}_i(\boldsymbol{\theta}) + \boldsymbol{e}_i, \quad i = 1, \ldots, n. \tag{2}$$

Denote the parameter space by $\boldsymbol{\Theta}$ and the unknown true parameter by $\boldsymbol{\theta}_0$. The ordinary least squares estimator (OLSE) of $\boldsymbol{\theta}_0$ based on data $\{\boldsymbol{y}_i\}_{i=1}^n$ is a vector $\hat{\boldsymbol{\theta}}_n \in \boldsymbol{\Theta}$ minimizing

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{h}_i(\boldsymbol{\theta}))'(\boldsymbol{y}_i - \boldsymbol{h}_i(\boldsymbol{\theta})). \tag{3}$$

Example 1. (continued). Under the exponential growth curve $\eta(t; \alpha_0, \alpha_1, \beta) = \alpha_0 - \alpha_1 e^{-\beta t}$, we have $h(t; \boldsymbol{\theta}) = E_\beta(\alpha_0 - \alpha_1 e^{-\beta t}) = \alpha_0 - \alpha_1 e^{-\mu t + \frac{1}{2}\sigma^2 t^2}$.

Theorem 1.

(i) *Under conditions* (C1)-(C5) *stated in Appendix A,* $\hat{\boldsymbol{\theta}}_n \longrightarrow \boldsymbol{\theta}_0$ *a.s. as* $n \to \infty$.

(ii) *Under conditions* (C1)-(C9) *stated in Appendix A,*

$$\boldsymbol{D}_n^{-1/2}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \longrightarrow N(\boldsymbol{0}, \boldsymbol{I}_{p+q}) \quad \text{in distribution,} \tag{4}$$

where $\boldsymbol{D}_n(\boldsymbol{\theta}) = \boldsymbol{A}_n^{-1}(\boldsymbol{\theta})(\sum_{i=1}^n \boldsymbol{H}_i(\boldsymbol{\theta})\boldsymbol{\Sigma}_i(\boldsymbol{\theta})\boldsymbol{H}_i'(\boldsymbol{\theta}))\boldsymbol{A}_n^{-1}(\boldsymbol{\theta})$, $\boldsymbol{H}_i(\boldsymbol{\theta}) = \partial \boldsymbol{h}_i'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\boldsymbol{A}_n(\boldsymbol{\theta}) = \sum_{i=1}^n \boldsymbol{H}_i(\boldsymbol{\theta})\boldsymbol{H}_i'(\boldsymbol{\theta})$, *and* $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ *is the covariance matrix of* $\boldsymbol{e}_i$.

Proof. The proof of (i) is given in Appendix B. We now show (4). Note that

$$\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{2}{n} \sum_{i=1}^{n} \boldsymbol{H}_i(\boldsymbol{\theta})(\boldsymbol{h}_i(\boldsymbol{\theta}) - \boldsymbol{h}_i(\boldsymbol{\theta}_0) - \boldsymbol{e}_i),$$

and

$$\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{2}{n}\sum_{i=1}^{n}(\boldsymbol{H}_i(\boldsymbol{\theta})\boldsymbol{H}_i'(\boldsymbol{\theta}) + \boldsymbol{G}_i(\boldsymbol{\theta})),$$

where $\boldsymbol{G}_i(\boldsymbol{\theta}) = \sum_{j=1}^{m_i}[h(t_{ij};\boldsymbol{\theta}) - y_{ij}]\boldsymbol{K}_{ij}(\boldsymbol{\theta})$ and $\boldsymbol{K}_{ij}(\boldsymbol{\theta}) = \partial^2 h(t_{ij};\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$. By (i), $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}_0$. Hence, $\partial Q_n(\hat{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta} = 0$ on $\{\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \le \delta\}$ with some $\delta > 0$. By the Mean-Value Theorem,

$$\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \frac{\partial^2 Q_n(\boldsymbol{\theta}_n^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n), \tag{5}$$

where $\boldsymbol{\theta}_n^*$ is a point on the line segment between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$. Thus, substituting the first two derivatives of $Q_n(\boldsymbol{\theta})$ into (5), we have $\sum_{i=1}^{n}\boldsymbol{H}_i(\boldsymbol{\theta}_0)\boldsymbol{e}_i = \boldsymbol{B}_n\boldsymbol{A}_n(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, where $\boldsymbol{B}_n = \boldsymbol{A}_n(\boldsymbol{\theta}_n^*)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0) + \sum_{i=1}^{n}\boldsymbol{G}_i(\boldsymbol{\theta}_n^*)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)$. Under (C7) and the consistency of $\hat{\boldsymbol{\theta}}_n$, the first term of $\boldsymbol{B}_n$ converges to $\boldsymbol{I}_{p+q}$ in probability. The second term of $\boldsymbol{B}_n$ is equal to

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i}[h(t_{ij};\boldsymbol{\theta}_n^*) - h(t_{ij};\boldsymbol{\theta}_0)]\boldsymbol{K}_{ij}(\boldsymbol{\theta}_n^*)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0) - \sum_{i=1}^{n}\sum_{j=1}^{m_i}e_{ij}\boldsymbol{K}_{ij}(\boldsymbol{\theta}_n^*)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0). \tag{6}$$

The first term of (6) converges to $\mathbf{0}$ in probability by (C7), (C8), the Cauchy-Schwarz inequality, and the fact that $\boldsymbol{\theta}_n^* \to \boldsymbol{\theta}_0$ in probability. The second term of (6) is a $(p+q)\times(p+q)$ matrix, whose (r,s)th entry converges to 0 in probability uniformly on $C = \{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le \delta\}$, by (C7), (C8), and Corollary A of Wu (1981). Hence, we showed that $\boldsymbol{B}_n \to \boldsymbol{I}_{p+q}$ in probability.

To complete the proof, we need to show that

$$\boldsymbol{D}_n^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)\sum_{i=1}^{n}\boldsymbol{H}_i(\boldsymbol{\theta}_0)\boldsymbol{e}_i \longrightarrow N(\mathbf{0}, \boldsymbol{I}_{p+q}) \quad \text{in distribution.} \tag{7}$$

Let $\boldsymbol{\ell}$ be a fixed $(p+q)$-vector. It suffices to show that

$$\boldsymbol{\ell}'\boldsymbol{D}_n^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)\sum_{i=1}^{n}\boldsymbol{H}_i(\boldsymbol{\theta}_0)\boldsymbol{e}_i \longrightarrow N(0, \boldsymbol{\ell}'\boldsymbol{\ell}) \qquad \text{in distribution.}$$

Let $\boldsymbol{\ell}_i' = \boldsymbol{\ell}'\boldsymbol{D}_n^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}_i(\boldsymbol{\theta}_0)$. We have $\boldsymbol{\ell}_i'\boldsymbol{\ell}_i = \boldsymbol{\ell}'\boldsymbol{D}_n^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}_i(\boldsymbol{\theta}_0)$ $\boldsymbol{H}_i'(\boldsymbol{\theta}_0)$ $\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)\boldsymbol{D}_n^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{\ell}$, and $\sum_{i=1}^{n}\boldsymbol{\ell}_i'\boldsymbol{\ell}_i = \boldsymbol{\ell}'\boldsymbol{D}_n^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)\boldsymbol{D}_n^{-1/2}(\boldsymbol{\theta}_0)\boldsymbol{\ell}$. By Courant's Theorem (see Graybill (1983), Theorem 12.4.14) and Theorem 3.2.4 in Graybill (1983), we have

$$\max_i \frac{\boldsymbol{\ell}_i'\boldsymbol{\ell}_i}{\sum_{i=1}^{n}\boldsymbol{\ell}_i'\boldsymbol{\ell}_i} \le \max_i\ ch_1[\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}_i(\boldsymbol{\theta}_0)\boldsymbol{H}_i'(\boldsymbol{\theta}_0)].$$

This last converges to 0 because of (C9). Thus Lindberg's condition holds and (7) follows, completing the proof.

The matrix $\boldsymbol{D}_n(\boldsymbol{\theta}_0)$ in Theorem 1 is the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n$. For inference, we need to estimate this matrix. Consider the following estimator of $\boldsymbol{D}_n(\boldsymbol{\theta}_0)$:

$$\hat{\boldsymbol{D}}_n = \boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n) \sum_{i=1}^{n} \boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n) \hat{\boldsymbol{\Sigma}}_i \boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n) \boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n), \tag{8}$$

where $\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{r}_i \boldsymbol{r}_i'$ and $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{h}_i(\hat{\boldsymbol{\theta}}_n)$ is the ith residual vector.

Theorem 2. *Under conditions* (C1)-(C5), (C10) *and* (C11) *stated in Appendix A,*

$$n(\hat{\boldsymbol{D}}_n - \boldsymbol{D}_n(\boldsymbol{\theta}_0)) \longrightarrow \boldsymbol{0} \quad a.s.$$

The proof of Theorem 2 is given in Appendix C.

We now return to degradation analysis, i.e., using degradation data to make inference about percentiles of the failure time distribution. From Lemma 1, t_p is a solution of $1 - F_\eta(\eta_c|t) = p$ and is a function of η_c and $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\phi})'$, say $t_p = g(\eta_c, \boldsymbol{\theta})$. Substituting the OLSE $\hat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}$, we obtain an estimate $\hat{t}_p = g(\eta_c, \hat{\boldsymbol{\theta}}_n)$.

We now consider the problem of constructing a confidence interval for t_p. Since t_p is a function of $\boldsymbol{\theta}$, the asymptotic normality of $\hat{t}_p$ can be derived by the Taylor expansion and the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$. Hence, an approximate $100(1-\alpha)\%$ confidence interval for t_p is

$$\left(\hat{t}_p - z_{\frac{\alpha}{2}} \hat{D}_{n,g}^{1/2} \ , \ \ \hat{t}_p + z_{\frac{\alpha}{2}} \hat{D}_{n,g}^{1/2}\right), \tag{9}$$

where

$$\hat{D}_{n,g} = \left[\frac{\partial g(\eta_c, \hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\right]' \hat{\boldsymbol{D}}_n \left[\frac{\partial g(\eta_c, \hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\right]$$

and $\hat{\boldsymbol{D}}_n$ is defined in (8).

When $F_\eta(\eta_c|t)$ has no closed form, we still can use numerical calculation to compute $\hat{t}_p = g(\eta_c, \hat{\boldsymbol{\theta}}_n)$ and the confidence interval (9).

4. The Weighted Least Squares Estimators

Consider the model at (2). Let $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ denote the covariance matrix of $\boldsymbol{e}_i$. The (j,k)th entry of $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ is

$$\sigma_{ijk}(\boldsymbol{\theta}) = \begin{cases} \sigma_\eta(t_{ij}; \boldsymbol{\theta}) + \sigma_\varepsilon^2 & \text{if } \ j = k \\ \sigma_\eta(t_{ij}, t_{ik}; \boldsymbol{\theta}) & \text{if } \ j \neq k, \end{cases} \tag{10}$$

where $\sigma_\eta(t_{ij};\boldsymbol{\theta}) = \mathrm{Var}_{\boldsymbol{\beta}_i}[\eta(t_{ij};\boldsymbol{\alpha},\boldsymbol{\beta}_i)]$ and $\sigma_\eta(t_{ij},t_{ik};\boldsymbol{\theta}) = \mathrm{Cov}_{\boldsymbol{\beta}_i}[\eta(t_{ij};\boldsymbol{\alpha},\boldsymbol{\beta}_i), \eta(t_{ik};\boldsymbol{\alpha},\boldsymbol{\beta}_i)]$.

Example 1. (continued). Recall that $\eta(t_{ij};\alpha_0,\alpha_1,\beta_i) = \alpha_0 - \alpha_1 e^{-\beta_i t_{ij}}$. Then

$$\begin{aligned}\sigma_\eta(t_{ij},t_{ik};\boldsymbol{\theta}) &= \mathrm{Cov}_{\beta_i}(\alpha_0 - \alpha_1 e^{-\beta_i t_{ij}}\ ,\ \alpha_0 - \alpha_1 e^{-\beta_i t_{ik}}) \\ &= \alpha_1^2 e^{-\mu(t_{ij}+t_{ik})+\frac{1}{2}\sigma^2(t_{ij}^2+t_{ik}^2)}(e^{\sigma^2 t_{ij}t_{ik}} - 1).\end{aligned}$$

Therefore,

$$\sigma_{ijk}(\boldsymbol{\theta}) = \begin{cases} \alpha_1^2 e^{-2\mu t_{ij}+\sigma^2 t_{ij}^2}(e^{\sigma^2 t_{ij}^2} - 1) + \sigma_\varepsilon^2, & \text{if } j = k \\ \alpha_1^2 e^{-\mu(t_{ij}+t_{ik})+\frac{1}{2}\sigma^2(t_{ij}^2+t_{ik}^2)}(e^{\sigma^2 t_{ij}t_{ik}} - 1), & \text{if } j \neq k. \end{cases}$$

Note that the ordinary least squares criterion (3) weights each observation equally. However, (10) indicates that the data do not have equal variances. Under such a case, the OLSE may be improved by a weighted least squares estimator (WLSE) of $\boldsymbol{\theta}_0$. The weighted least squares method involves choosing adequate weights. Suppose we have consistent estimators $\hat{\boldsymbol{\Sigma}}_i$ of the covariance matrices $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)$. Then one can obtain the WLSE $\hat{\boldsymbol{\theta}}_n^w$ of $\boldsymbol{\theta}_0$ by minimizing

$$\hat{Q}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{h}_i(\boldsymbol{\theta}))'\hat{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{h}_i(\boldsymbol{\theta})) \tag{11}$$

over $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Davidian and Giltinan (1993) reviewed some existing methods for estimating parameters in the nonlinear random coefficient model. A simple extension of generalized least squares procedure is also proposed. However, there is no theoretical result discussed in their paper. In this section, we propose two WLSE's and discuss their asymptotic properties.

We first consider a special case. Model (2) is said to be balanced if all test units have the same number of measurements and the measurement times of all units are the same. That is, $m_i = m$ and $t_{ij} = t_j$ for all $i = 1,\ldots,n$. For a balanced model, $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ and $\boldsymbol{h}_i(\boldsymbol{\theta}) = \boldsymbol{h}(\boldsymbol{\theta})$ for all $i = 1,\ldots,n$. Using the ideas in Gallant (1975, 1987) and Phillips (1976), we consider the estimator of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ as follows: $\hat{\boldsymbol{\Sigma}} = n^{-1}\sum_{i=1}^n \boldsymbol{r}_i\boldsymbol{r}_i'$, where $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{h}(\hat{\boldsymbol{\theta}}_n)$ and $\hat{\boldsymbol{\theta}}_n$ is the OLSE of $\boldsymbol{\theta}_0$. Using the law of large numbers and the continuity of $\boldsymbol{h}(\boldsymbol{\theta})$, one can easily show that $\hat{\boldsymbol{\Sigma}}$ is consistent for $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. The weighted least squares estimator $\hat{\boldsymbol{\theta}}_n^w$ is obtained by minimizing

$$\hat{Q}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{h}(\boldsymbol{\theta}))'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}_i - \boldsymbol{h}(\boldsymbol{\theta})).$$

We have the following result.

Theorem 3.

(i) *Under conditions* (C1)-(C5) *and* (C12) *stated in Appendix A,* $\hat{\boldsymbol{\theta}}_n^w \longrightarrow \boldsymbol{\theta}_0$ *a.s.*

(ii) $(\boldsymbol{D}_n^w(\boldsymbol{\theta}_0))^{-1/2}\,(\hat{\boldsymbol{\theta}}_n^w - \boldsymbol{\theta}_0) \longrightarrow N(\mathbf{0}, \boldsymbol{I}_{p+q})$ *in distribution, where* $\boldsymbol{D}_n^w(\boldsymbol{\theta}) = n^{-1}[\boldsymbol{H}(\boldsymbol{\theta})\ \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\boldsymbol{H}'(\boldsymbol{\theta})]^{-1}$ *and* $\boldsymbol{H}(\boldsymbol{\theta}) = \partial \boldsymbol{h}'(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$.

(iii) *Let* $\hat{\boldsymbol{D}}_n^w = n^{-1}[\boldsymbol{H}(\hat{\boldsymbol{\theta}}_n^w)\hat{\boldsymbol{\Sigma}}_w^{-1}\boldsymbol{H}'(\hat{\boldsymbol{\theta}}_n^w)]^{-1}$, *where* $\hat{\boldsymbol{\Sigma}}_w = n^{-1}\sum_{i=1}^n(\boldsymbol{r}_i^w)(\boldsymbol{r}_i^w)'$ *and* $\boldsymbol{r}_i^w = \boldsymbol{y}_i - \boldsymbol{h}(\hat{\boldsymbol{\theta}}_n^w)$. *Then* $n(\hat{\boldsymbol{D}}_n^w - \boldsymbol{D}_n^w(\boldsymbol{\theta}_0)) \longrightarrow \mathbf{0}$ *a.s.*

The proof of Theorem 3 can be found in Wu (1996).

We now consider the unbalanced case. The main difficulty is to obtain a consistent estimator of $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)$. As the method we used for the balanced case cannot be applied.

Let $\boldsymbol{V}_i(\boldsymbol{\theta})$, $i = 1, \ldots, n$, be the $m_i \times m_i$ covariance matrices where the (j,k)th entry of $\boldsymbol{V}_i(\boldsymbol{\theta})$ is $\sigma_\eta(t_{ij}, t_{ik}; \boldsymbol{\theta})$. Then $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}) = \sigma_\varepsilon^2\boldsymbol{I}_{m_i} + \boldsymbol{V}_i(\boldsymbol{\theta}), i = 1, \ldots, n$. We consider the following estimator of $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$:

$$\hat{\boldsymbol{\Sigma}}_i = \hat{\sigma}_\varepsilon^2\boldsymbol{I}_{m_i} + \boldsymbol{V}_i(\hat{\boldsymbol{\theta}}_n), \tag{12}$$

where $\hat{\sigma}_\varepsilon^2 = m^{-1}n^{-1}\sum_{i=1}^n\left[\boldsymbol{r}_i'\boldsymbol{r}_i - tr\left(\boldsymbol{V}_i(\hat{\boldsymbol{\theta}}_n)\right)\right]$, $m = n^{-1}\sum_{i=1}^n m_i$, $\hat{\boldsymbol{\theta}}_n$ is the OLSE, and $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{h}_i(\hat{\boldsymbol{\theta}}_n)$. Under some weak conditions we can show that $\hat{\boldsymbol{\Sigma}}_i \to \boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)$ a.s. uniformly in i (details are given in Wu (1996)).

We propose the weighted least squares estimate $\hat{\boldsymbol{\theta}}_n^w$ of $\boldsymbol{\theta}_0$ obtained by minimizing (11) using $\hat{\boldsymbol{\Sigma}}_i$ in (12). We have the following theorem.

Theorem 4.

(i) *Under conditions* (C5) *and* (C13)-(C16) *stated in Appendix A,* $\hat{\boldsymbol{\theta}}_n^w \longrightarrow \boldsymbol{\theta}_0$ *a.s.*

(ii) *Under conditions* (C5), (C13)-(C15) *and* (C17)-(C20), *we have* $(\boldsymbol{D}_n^w(\boldsymbol{\theta}_0))^{-1/2}$ $(\hat{\boldsymbol{\theta}}_n^w - \boldsymbol{\theta}_0) \longrightarrow N(\mathbf{0}, \boldsymbol{I}_{p+q})$ *in distribution, where* $\boldsymbol{D}_n^w(\boldsymbol{\theta}) = [\sum_{i=1}^n \boldsymbol{H}_i(\boldsymbol{\theta})\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta})$ $\boldsymbol{H}_i'(\boldsymbol{\theta})]^{-1}$.

(iii) *Let* $\hat{\boldsymbol{D}}_n^w = [\sum_{i=1}^n \boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n^w)(\hat{\boldsymbol{\Sigma}}_i^w)^{-1}\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n^w)]^{-1}$, *where* $\hat{\boldsymbol{\Sigma}}_i^w = \hat{\sigma}_{\varepsilon,w}^2\boldsymbol{I}_{m_i} + \boldsymbol{V}_i(\hat{\boldsymbol{\theta}}_n^w)$. *Then* $n(\hat{\boldsymbol{D}}_n^w - \boldsymbol{D}_n^w(\boldsymbol{\theta}_0)) \longrightarrow \mathbf{0}$ *a.s.*

The proof of Theorem 4 can be found in Wu (1996).

For the percentiles of the failure time distribution, we can estimate t_p by $\hat{t}_p^w = g(\eta_c, \hat{\theta}_n^w)$ and obtain the following approximate $100(1-\alpha)\%$ confidence interval for t_p:

$$\left(\hat{t}_p^w - z_{\frac{\alpha}{2}}(\hat{D}_{n,g}^w)^{1/2}\quad,\quad \hat{t}_p^w + z_{\frac{\alpha}{2}}(\hat{D}_{n,g}^w)^{1/2}\right),$$

where

$$\hat{D}_{n,g}^w = \left[\frac{\partial g(\eta_c, \hat{\boldsymbol{\theta}}_n^w)}{\partial\boldsymbol{\theta}}\right]' \hat{\boldsymbol{D}}_n^w \left[\frac{\partial g(\eta_c, \hat{\boldsymbol{\theta}}_n^w)}{\partial\boldsymbol{\theta}}\right].$$

5. Examples and Simulations

We apply the proposed methods to two degradation data sets. For each data set, we also carry out simulations to examine the performance of the proposed methods.

5.1. Data analysis for metal film resistor example

Our first example is the resistance degradation data of metal film resistor from Zhuang (1994). There are 200 resistors in this experiment and each has repeated measurements of resistance taken at five different times: 0, 500, 1000, 2000, and 4600 hours. Figure 1 shows the sample paths of these test units.

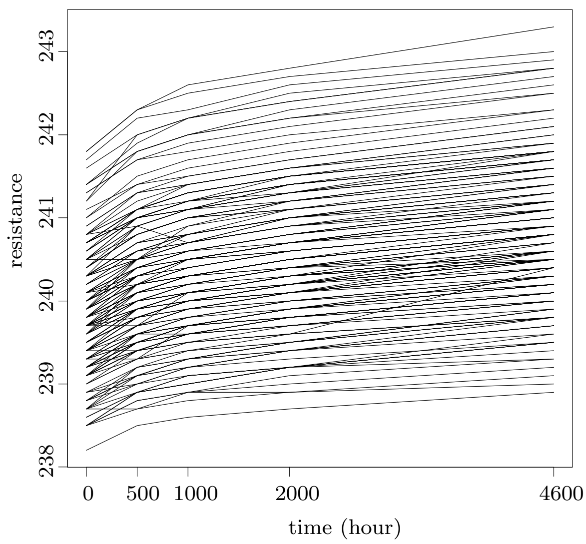

Figure 1. The sample degradation paths of metal film resistors.

Because the resistance of metal film resistor is increasing over time, we say a failure occurs when the ratio of resistance at time t to initial resistance is beyond 1.02, as did Zhuang. We notice that the resistance of metal film resistor degrades very slowly. The maximum observed ratio of $y(t)/y(0)$ in this data set is only 1.0066 after a 4600-hour experiment, where $y(0)$ is the observed initial resistance and $y(t)$ is the observed resistance at time t. There is no failure in this test. But since all resistors were observed at five different times and there are no missing values, we are dealing with a balanced degradation data analysis.

We consider the ratios $y(t_{ij})/y(t_{i1}), i = 1, \ldots, 200, j = 1, \ldots, 5$, where $t_{i1} = 0$. Note that the ratios are always equal to 1 when $j = 1$. Figure 2 is a plot of the ratios versus time, connected by straight lines. We found that the ratios of all test units look like concave functions of time, except the 41st resistor. In addition, the ratio of the 47th resistor decreases between 500 and 1000 hours. This may be a mistake when data were collected. We decided to delete the data on these two test units from our analysis leaving 198 resistors.

We consider the following model, similar to that in Zhuang (1994):

$$y_{ij} = \begin{cases} 0, & i = 1, \ldots, 198, \quad j = 1, \\ \beta_i t_j^{\alpha} + \varepsilon_{ij}, & i = 1, \ldots, 198, \quad j = 2, \ldots, 5, \end{cases} \tag{13}$$

where $y_{ij} = (y(t_{ij})/y(t_{i1})) - 1$, t_j is the measurement time, α is a fixed effect parameter, and $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$. The random effects β_i's are independently distributed as an exponential distribution with mean λ. We can get an explicit form of $h(t; \boldsymbol{\theta})$:

$$h(t_j; \boldsymbol{\theta}) = E_{\beta_i}[\eta(t_j; \beta_i, \alpha)] = \lambda t_j^{\alpha}.$$

Therefore, t_p also has an explicit form:

$$t_p = \left(\frac{-\eta_c}{\lambda \log(p)} \right)^{1/\alpha}.$$

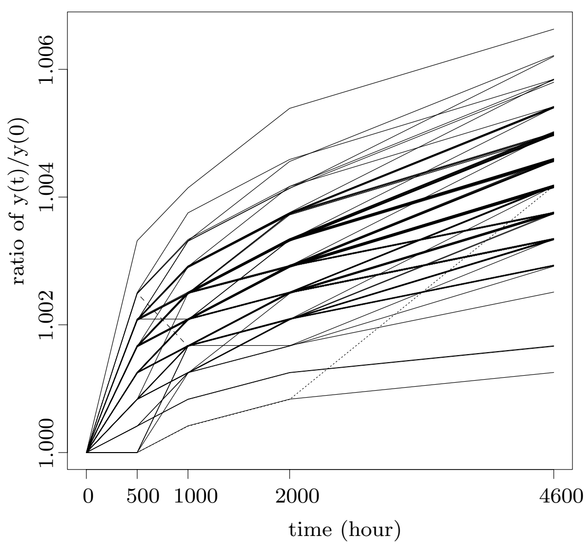

Figure 2. The sample degradation paths of metal film resistors after taking ratios $y(t)/y(0)$.

Following Section 3, the OLSE of λ and α, and the corresponding estimated asymptotic covariance matrix are

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} 0.0021 \\ 0.4626 \end{bmatrix}, \quad \text{and} \quad \hat{\boldsymbol{D}}_n = \begin{bmatrix} 0.012 & -1.233 \\ -1.233 & 272.860 \end{bmatrix} \times 10^{-7},$$

respectively. Table 1 lists point estimates and 90% confidence intervals for some percentiles of the failure time distribution. Figure 3 is the plot of these point estimates and confidence intervals, connected by straight lines. The confidence intervals are narrow when p is close to 0. As p goes to 1, the confidence intervals become wide. This can be explained by the fact that degradation data provide more information about the lower tail of the failure time distribution than the upper tail.

Table 1. Point estimates and 90% confidence intervals for percentiles of the failure time distribution (based on OLSE).

p	$\hat{t}_p$ (thousands)	lower limits (thousands)	upper limits (thousands)
0.05	12.51	11.95	13.07
0.10	22.10	21.05	23.15
0.20	47.94	45.33	50.55
0.30	89.79	84.23	95.34
0.40	162.01	150.68	173.35
0.50	296.19	272.80	319.57
0.60	572.92	521.74	624.10
0.70	1245.39	1118.24	1372.54
0.80	3432.42	3022.83	3842.01
0.90	17381.29	14813.75	19948.83
0.95	82382.92	67925.09	96840.75

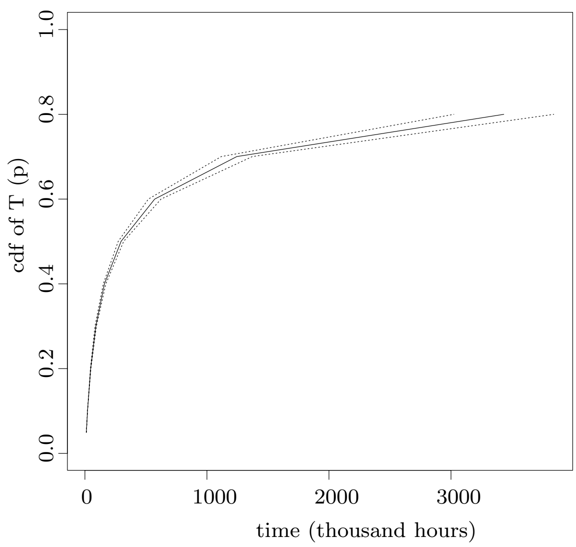

Figure 3. Point estimates and 90% confidence intervals for percentiles of the failure time distribution (based on OLSE). The solid line is the point estimate. The dotted lines are the 90% confidence intervals.

We also analyzed the data by using the weighted least squares method discussed in Section 4. The results are almost the same and are omitted.

5.2. Simulation results for metal film resistor example

It is important to examine how well our proposed methods work for assessing reliability. We simulated 1000 samples from the model at (13). The true parameters (λ, α) are equal to $(0.0021, 0.4626)$, our OLSE in the example. The degradation data of an individual test unit was simulated by generating a random exponential variable with mean 0.0021 and calculating the amount of degradation, at specified times, using this variable. A random normal error was then generated with mean zero and standard deviation 0.0001 (the estimate of standard deviation in our example is smaller than 10^{-5}). The former is then added to the latter to yield the simulation data. The simulation samples were generated by using the UNI and RNOR subroutines in the FORTRAN Library CMLIB.

Table 2. Simulation results for percentiles of the failure time distribution.

OLSE						
p	true t_p	means of $\hat{t}_p$	bias	relative bias	means of the confidence lengths	coverage
0.05	12.18	12.43	0.24	0.0204	6.21	0.896
0.10	21.52	21.96	0.43	0.0204	10.98	0.897
0.20	46.67	47.62	0.94	0.0203	23.85	0.897
0.30	87.42	89.20	1.77	0.0203	44.73	0.897
0.40	157.74	160.95	3.21	0.0203	80.83	0.900
0.50	288.38	294.25	5.87	0.0204	148.02	0.901
0.60	557.84	569.21	11.36	0.0204	286.94	0.902
0.70	1212.66	1237.40	24.73	0.0204	625.53	0.901
0.80	3342.34	3410.69	68.34	0.0204	1731.45	0.900
0.90	16926.30	17274.25	347.95	0.0206	8839.44	0.898
0.95	80231.56	81892.21	1660.65	0.0207	42284.28	0.899

WLSE						
p	true t_p	means of $\hat{t}_p$	bias	relative bias	means of the confidence lengths	coverage
0.05	12.18	12.43	0.24	0.0204	6.18	0.894
0.10	21.52	21.96	0.43	0.0204	10.92	0.895
0.20	46.67	47.62	0.94	0.0203	23.73	0.895
0.30	87.42	89.20	1.77	0.0203	44.50	0.895
0.40	157.74	160.95	3.21	0.0203	80.42	0.895
0.50	288.38	294.25	5.87	0.0204	147.28	0.899
0.60	557.84	569.21	11.36	0.0204	285.49	0.899
0.70	1212.66	1237.40	24.73	0.0204	622.38	0.898
0.80	3342.34	3410.69	68.34	0.0204	1722.72	0.898
0.90	16926.30	17274.25	347.95	0.0206	8794.90	0.894
0.95	80231.56	81892.21	1660.65	0.0207	42071.31	0.898

Table 2 presents the simulation results for some percentiles of the failure time distribution using the ordinary least squares and weighted least squares methods. The means of the confidence lengths of percentiles using the weighted least squares method are slightly smaller than those using the ordinary least squares method. The coverage probabilities of both methods are close to their desired level of 0.9.

5.3. Data analysis for a metal fatigue example

In this section, we will discuss the example of fatigue-crack-growth data from Lu and Meeker (1993), see Figure 4. There are 21 test units in this experiment, each with the same initial crack length of 0.9 inches. We say a failure occurs when the crack length is beyond 1.6 inches as did Lu and Meeker (1993). There are 13 measurement times in this test. However, only 13 units have complete records of crack lengths at 13 different times, an unbalanced case.

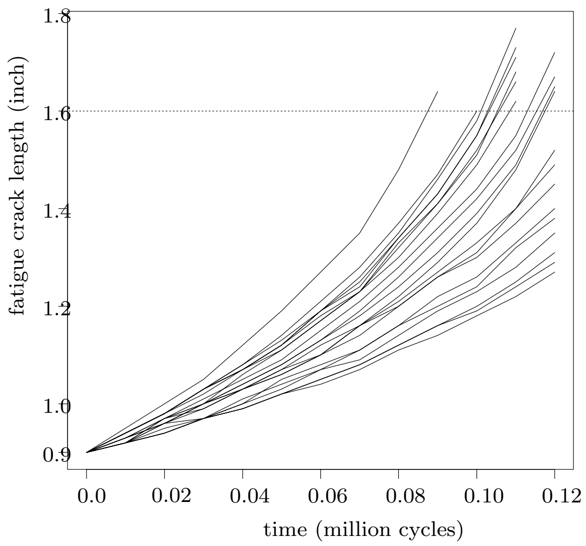

Figure 4. The fatigue crack length data from Lu and Meeker (1993).

Lu and Meeker (1993) suggested the nonlinear mixed-effect model

$$y_{ij} = -\frac{1}{\theta_{2i}}\log(1 - 0.9^{\theta_{2i}}\theta_{1i}\theta_{2i}t_j) + \varepsilon_{ij}, \qquad i = 1, \ldots, 21, \quad j = 1, \ldots, m_i, \quad (14)$$

where y_{ij} = log(observed crack length at time $t_j/0.9$), t_j is the measurement time (in million cycles), the coefficients $(\theta_{1i}, \theta_{2i})'$ are random effects. This model was derived from the Paris Law in material science. In the following analysis, we suppose that θ_{1i} and θ_{2i} are independently distributed as $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively.

Note that $h(t_j; \boldsymbol{\theta}) = E[\eta(t_j; \theta_{1i}, \theta_{2i})]$, where $\boldsymbol{\theta} = (\mu_1, \sigma, \mu_2)'$, does not have a closed form and numerical integration is required to compute the OLSE and WLSE. We used Monte Carlo integration method to compute $h(t_j; \boldsymbol{\theta})$. It required about 35 minutes computer-time to get the following results on a DEC-workstation.

The OLSE of the unknown parameters μ_1, σ and μ_2 and the corresponding estimated asymptotic covariance matrix are

$$\begin{bmatrix} \hat{\mu}_1 \\ \hat{\sigma} \\ \hat{\mu}_2 \end{bmatrix} = \begin{bmatrix} 3.6572 \\ 1.0771 \\ 1.5875 \end{bmatrix}, \quad \text{and} \quad \hat{\boldsymbol{D}}_n = \begin{bmatrix} 22.2862 & -3.3097 & -11.3650 \\ -3.3097 & 73.035 & 144.2751 \\ -11.3650 & 144.2751 & 367.2063 \end{bmatrix} \times 10^{-3},$$

respectively. Table 3 shows point estimates and 90% confidence intervals for some percentiles of the failure time distribution. Figure 5 is the plot of these estimates and confidence intervals. The median of failure time distribution is about 0.1245 million cycles which is close to the median of 0.12 in Lu and Meeker (1993).

Table 3. Point estimates and 90% confidence intervals for percentiles of the failure time distribution (based on OLSE).

p	$\hat{t}_p$	lower limits	upper limits
0.139	0.09	0.0812	0.0988
0.238	0.10	0.0897	0.1103
0.348	0.11	0.0972	0.1228
0.456	0.12	0.1041	0.1359
0.554	0.13	0.1105	0.1495
0.638	0.14	0.1167	0.1633
0.708	0.15	0.1229	0.1771
0.764	0.16	0.1293	0.1907
0.809	0.17	0.1360	0.2040
0.843	0.18	0.1432	0.2168
0.870	0.19	0.1507	0.2293
0.891	0.20	0.1587	0.2413
0.908	0.21	0.1670	0.2530
0.920	0.22	0.1755	0.2645

Since this is an unbalanced case, we use the weights described in (12) to obtain the WLSE and the estimated covariance matrix:

$$\begin{bmatrix} \hat{\mu}_1^w \\ \hat{\sigma}^w \\ \hat{\mu}_2^w \end{bmatrix} = \begin{bmatrix} 3.6712 \\ 0.9818 \\ 1.5680 \end{bmatrix}, \quad \text{and} \quad \hat{\boldsymbol{D}}_n^w = \begin{bmatrix} 47.8279 & 8.0744 & 1.7516 \\ 8.0744 & 80.5848 & 3.4615 \\ 1.7516 & 3.4615 & 61.5018 \end{bmatrix} \times 10^{-3}.$$

From Table 4 and Figure 6, the median of the failure time distribution is about 0.1240 million cycles. It is very close to the median of 0.1245 obtained by using the ordinary least squares method. However, the confidence intervals for percentiles of the failure time distribution based on the ordinary least squares method are wider than those based on the weighted least squares method.

Table 4. Point estimates and 90% confidence intervals for percentiles of the failure time distribution (based on WLSE).

p	$\hat{t}_p$	lower limits	upper limits
0.120	0.09	0.0812	0.0988
0.222	0.10	0.0896	0.1104
0.339	0.11	0.0983	0.1217
0.457	0.12	0.1066	0.1334
0.564	0.13	0.1142	0.1458
0.656	0.14	0.1212	0.1588
0.731	0.15	0.1280	0.1720
0.790	0.16	0.1348	0.1852
0.836	0.17	0.1420	0.1980
0.870	0.18	0.1497	0.2103
0.897	0.19	0.1579	0.2221
0.916	0.20	0.1663	0.2337
0.931	0.21	0.1750	0.2450
0.942	0.22	0.1837	0.2563

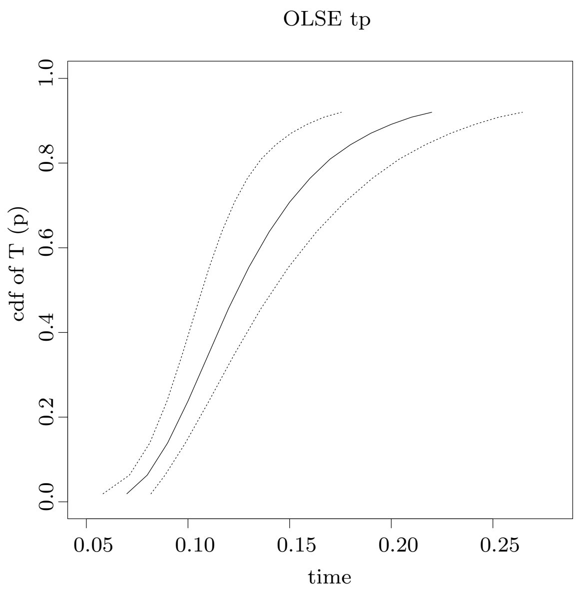

Figure 5. Point estimates and 90% confidence intervals for percentiles of the failure time distribution (based on OLSE). The solid line is the point estimate. The dotted lines are the 90% confidence intervals.

5.4. Simulation results for metal fatigue example

To study the performance of our approach in this example, we simulated 500 samples from (14) with the values of parameters $(\mu_1, \sigma, \mu_2)' = (3.6572, 1.0771, 1.5875)'$. The samples were simulated by using the same method described in Section 5.2, with random exponential variables replaced by independently random normal variables.

The simulation results are shown in Table 5. The coverage probabilities of percentiles of the failure time distribution are close to the desired level of 0.9. Since t_p does not have a closed form, we could get $\hat{t}_p$ by solving $\int_0^{\hat{t}_p} \int_{-\infty}^{\infty} f_{\theta_1,\theta_2} (v(u,t),u)|J|dudt = p$, where $v(u,t) = (1 - e^{-u\eta_c})/(0.9^u tu)$ and $J = (1 - e^{-u\eta_c})/(0.9^u t^2 u)$. We acquired $\hat{t}_p$ first and obtained the p by using numerical integration. Thus, we can get the point estimates and confidence intervals for the t_p. However, the computed p's for a given $\hat{t}_p$ are not the same in these 500 simulated samples, since the estimates of $\boldsymbol{\theta}$ are different. Lengths of confidence intervals are not computed due to complexity of computation.

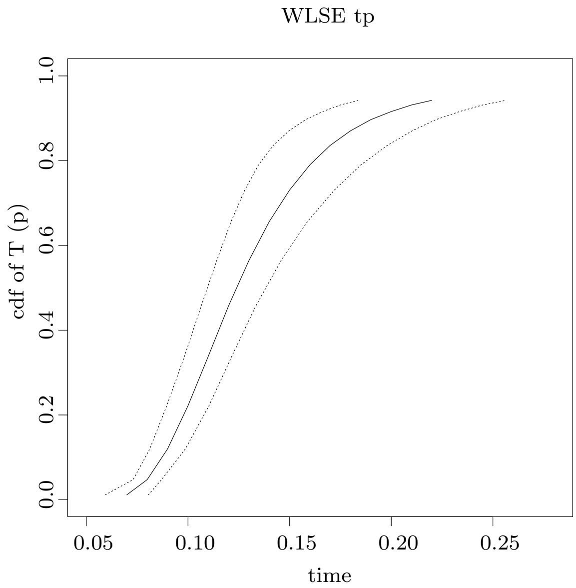

Figure 6. Point estimates and 90% confidence intervals for percentiles of the failure time distribution (based on WLSE). The solid line is the point estimate. The dotted lines are the 90% confidence intervals.

6. Discussion

We use degradation analysis to estimate the failure time distribution and its percentiles in the case where no or very few failures occur in a life test. We mainly cover three basic elements — the nonlinear mixed-effect model, least squares methods, and the asymptotic properties of the least squares estimators.

The nonlinear functional form gives us a flexible way to build the relationship between degradation measurements and times, and the random effects describe individual units' characteristics in a test. The sensitivity of the functional form of η in (1) and the random effects distribution π are not addressed in this paper but will be in a future study.

We adopt ordinary and weighted least squares methods, which are computationally much simpler than likelihood-based methods. The WLSE is asymptotically more efficient than the OLSE but is sensitive to the choice of weights, and its efficiency may be affected by the error in estimating $\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)$ especially in unbalanced cases. Thus, it is a good practice to apply both methods in a given problem.

Table 5. Simulation results for percentiles of the failure time distribution.

p	true t_p	OLSE coverage	WLSE coverage
0.2383	0.10	0.878	0.864
0.3480	0.11	0.894	0.886
0.4559	0.12	0.900	0.894
0.5539	0.13	0.906	0.904
0.6381	0.14	0.920	0.906
0.7078	0.15	0.920	0.910
0.7640	0.16	0.910	0.904
0.8085	0.17	0.902	0.898
0.8433	0.18	0.894	0.882
0.8703	0.19	0.890	0.878

Two real data sets are analyzed in Section 5 to illustrate our approach. From the simulation results in Section 5, the coverage probabilities of our confidence intervals are close to the desired level. This suggests that our method works well if model assumptions are reasonable.

Appendix A. Regularity Conditions

(C1) $\boldsymbol{\Theta}$ is a compact subset of $\Re^{p+q}$.

(C2) The $h(t_{ij};\boldsymbol{\theta})$ are continuous functions in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(C3) The $\boldsymbol{\Sigma}_i(\boldsymbol{\theta})$ are positive definite matrices.

(C4) $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \boldsymbol{h}_i'(\boldsymbol{t})\boldsymbol{h}_i(\boldsymbol{s})$ exists uniformly for all $\boldsymbol{t}$ and $\boldsymbol{s}$ in $\boldsymbol{\Theta}$, and $Q(\boldsymbol{\theta}) = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n (\boldsymbol{h}_i(\boldsymbol{\theta}) - \boldsymbol{h}_i(\boldsymbol{\theta}_0))'(\boldsymbol{h}_i(\boldsymbol{\theta}) - \boldsymbol{h}_i(\boldsymbol{\theta}_0))$ has a unique minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(C5) $\sup_i E\|\boldsymbol{e}_i\|^{2+\delta} < \infty$ for some $\delta > 0$.

(C6) $\boldsymbol{H}_i(\boldsymbol{\theta})$ and $\boldsymbol{K}_{ij}(\boldsymbol{\theta})$ are continuous functions for all i and j.

(C7) (i) For sufficiently large n, the inverse of $\boldsymbol{A}_n(\boldsymbol{\theta}_0)$ exists, $\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0) = O(n^{-1})$, and $\boldsymbol{A}_n(\boldsymbol{\theta}_0) = O(n)$.

(ii) $\boldsymbol{A}_n(\boldsymbol{\theta})\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)$ converges to the identity matrix $\boldsymbol{I}_{p+q}$ as $n \to \infty$ and $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$.

(C8) (i) There exists $\delta > 0$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta} \left(\frac{\partial^2 \boldsymbol{h}_i(\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right)' \left(\frac{\partial^2 \boldsymbol{h}_i(\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right) < \infty,$$

for $r, s = 1, \ldots, p+q$.

(ii) If for a pair (r,s), $r, s = 1, \ldots, p+q$,

$$\sum_{i=1}^{\infty} \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\delta} \left(\frac{\partial^2 \boldsymbol{h}_i(\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right)' \left(\frac{\partial^2 \boldsymbol{h}_i(\boldsymbol{\theta})}{\partial\theta_r\partial\theta_s}\right) = \infty,$$

then there exists an M independent of i such that

$$\sup_{\boldsymbol{v}\neq\boldsymbol{u},\ \boldsymbol{u},\boldsymbol{v}\in C} \frac{\left\| \frac{\partial^2 \boldsymbol{h}_i(\boldsymbol{u})}{\partial\theta_r\partial\theta_s} - \frac{\partial^2 \boldsymbol{h}_i(\boldsymbol{v})}{\partial\theta_r\partial\theta_s} \right\|}{\|\boldsymbol{u}-\boldsymbol{v}\|} \leq M \sup_{\boldsymbol{u}\in C} \left\| \frac{\partial^2 \boldsymbol{h}_i(\boldsymbol{u})}{\partial\theta_r\partial\theta_s} \right\|,$$

for all i, where $C = \{\boldsymbol{\theta} \in \boldsymbol{\Theta},\ \|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| \leq \delta\}$, $\delta > 0$.

(C9) $\max_i\ ch_1(\boldsymbol{A}_n^{-1}(\boldsymbol{\theta}_0)\boldsymbol{H}_i(\boldsymbol{\theta}_0)\boldsymbol{H}_i'(\boldsymbol{\theta}_0)) \to 0$ as $n \to \infty$, where $ch_1(\cdot)$ denotes the largest eigenvalue of a square matrix.

(C10) $\sup_i \sum_{r=1}^{p+q} \|\partial \boldsymbol{h}_i(\boldsymbol{\theta}_0)/\partial\theta_r\| \leq c < \infty$, where c is a positive constant.

(C11) $\{\partial \boldsymbol{h}_i(\boldsymbol{\theta})/\partial\theta_r\}_{i=1}^{\infty}$, $r = 1, \ldots, p+q$, are equicontinuous in $\theta \in C = \{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \delta\}$, where $\delta > 0$.

(C12) $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = (\boldsymbol{h}(\boldsymbol{\theta}_0) - \boldsymbol{h}(\boldsymbol{\theta}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{h}(\boldsymbol{\theta}_0) - \boldsymbol{h}(\boldsymbol{\theta}))$ has a unique minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(C13) The $\sigma_\eta(t_{ij}, t_{ik}; \boldsymbol{\theta})$ are continuous functions in $\boldsymbol{T} \times \boldsymbol{T} \times \boldsymbol{\Theta}$, where $\boldsymbol{T}$ and $\boldsymbol{\Theta}$ are compact sets.

(C14) The $\boldsymbol{h}_i(\boldsymbol{\theta}) = (h(t_{i1}; \boldsymbol{\theta}), \ldots, h(t_{im_i}; \boldsymbol{\theta}))'$ are continuous function in $\boldsymbol{T} \times \boldsymbol{\Theta}$.

(C15) $\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta})$, $i = 1, \ldots, n$, are continuous functions in a compact set $\boldsymbol{T}\times\boldsymbol{T}\times\boldsymbol{\Theta}$.

(C16) $\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \boldsymbol{h}_i'(\boldsymbol{t})\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0)\boldsymbol{h}_i(\boldsymbol{s})$ exist uniformly for all $\boldsymbol{t}$, $\boldsymbol{s}$ and $\boldsymbol{\theta}_0$ in $\boldsymbol{\Theta}$, and

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{h}_i(\boldsymbol{\theta}_0) - \boldsymbol{h}_i(\boldsymbol{\theta}))'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_0)(\boldsymbol{h}_i(\boldsymbol{\theta}_0) - \boldsymbol{h}_i(\boldsymbol{\theta}))$$

has a unique minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

(C17) The first two derivatives of $\boldsymbol{h}_i(\boldsymbol{\theta})$ exist and are continuous in $(\boldsymbol{t}, \boldsymbol{\theta}) \in \boldsymbol{T}\times\boldsymbol{\Theta}$.

(C18) (i) For sufficiently large n, the inverse of $\boldsymbol{D}_n^w(\boldsymbol{\theta}_0)$ exists, $\boldsymbol{D}_n^w(\boldsymbol{\theta}_0) = O(n^{-1})$ and $(\boldsymbol{D}_n^w(\boldsymbol{\theta}_0))^{-1} = O(n)$.

(ii) $\boldsymbol{D}_n^w(\boldsymbol{\theta})\,(\boldsymbol{D}_n^w(\boldsymbol{\theta}_0))^{-1}$ converges to the identity matrix $\boldsymbol{I}_{p+q}$ as $n \to \infty$ and $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$.

(C19) There exist positive constants ν and ρ such that, for all i,

$$\nu \leq (\text{ eigenvalues of } \boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)) \leq \rho.$$

(C20) $\max_i \; tr[\boldsymbol{H}_i'(\boldsymbol{\theta}_0)\boldsymbol{D}_n^w(\boldsymbol{\theta}_0)\boldsymbol{H}_i(\boldsymbol{\theta}_0)] \to 0$ as $n \to \infty$, where $tr(\cdot)$ is the trace of a square matrix.

Appendix B. Proof of Theorem 1 (i)

It follows from (C5) and the strong law of large numbers that

$$\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{e}_i'\boldsymbol{e}_i - tr(\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0))) \longrightarrow 0 \quad \text{a.s.}$$

Let $\omega = \{\boldsymbol{e}_i\}_{i=1}^{\infty}$ be fixed such that $n^{-1}\sum_{i=1}^{n}(\boldsymbol{e}_i'\boldsymbol{e}_i - tr(\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0))) \to 0$, and let $\boldsymbol{\theta}_n = \hat{\boldsymbol{\theta}}_n(\omega)$ be the ordinary least squares estimator. Suppose $\tilde{\boldsymbol{\theta}}$ is a limit point of $\{\boldsymbol{\theta}_n\}_{n=1}^{\infty}$. Then, by the compactness of $\boldsymbol{\Theta}$, there exists a subsequence $\{n_k\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty}\boldsymbol{\theta}_{n_k} = \tilde{\boldsymbol{\theta}}$. Let $\Delta_n = n^{-1}\sum_{i=1}^{n} tr(\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0))$. Then, by (C5), $\{\Delta_n\}_{n=1}^{\infty}$ is bounded. Thus, there exists a subsequence $\{n_j\}_{j=1}^{\infty} \subset \{n_k\}_{k=1}^{\infty}$ such that $\Delta = \lim_{j\to\infty}\Delta_{n_j}$ exists. By extending Theorem 4 of Jennrich (1969), we have $n_j^{-1}\sum_{i=1}^{n_j}(\boldsymbol{h}_i(\boldsymbol{\theta}_0) - \boldsymbol{h}_i(\boldsymbol{\theta}))'\boldsymbol{e}_i \longrightarrow 0$ uniformly for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Then

$$\begin{aligned} Q_{n_j}(\boldsymbol{\theta}) &= \frac{1}{n_j}\sum_{i=1}^{n_j}(\boldsymbol{h}_i(\boldsymbol{\theta}_0) - \boldsymbol{h}_i(\boldsymbol{\theta}))'(\boldsymbol{h}_i(\boldsymbol{\theta}_0) - \boldsymbol{h}_i(\boldsymbol{\theta})) \\ &\qquad + \frac{2}{n_j}\sum_{i=1}^{n_j}(\boldsymbol{h}_i(\boldsymbol{\theta}_0) - \boldsymbol{h}_i(\boldsymbol{\theta}))'\boldsymbol{e}_i + \frac{1}{n_j}\sum_{i=1}^{n_j}\boldsymbol{e}_i'\boldsymbol{e}_i \\ &\longrightarrow Q(\boldsymbol{\theta}) + \Delta \end{aligned}$$

uniformly for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ as $j \to \infty$. Hence, $\lim_{j\to\infty} Q_{n_j}(\boldsymbol{\theta}_{n_j}) = Q(\tilde{\boldsymbol{\theta}}) + \Delta$. Since $\boldsymbol{\theta}_{n_j}$ is an ordinary least squares estimator, we have $Q_{n_j}(\boldsymbol{\theta}_{n_j}) \leq Q_{n_j}(\boldsymbol{\theta}_0) = \frac{1}{n_j}\sum_{i=1}^{n_j}\boldsymbol{e}_i'\boldsymbol{e}_i$. Let $j \to \infty$ to find $Q(\tilde{\boldsymbol{\theta}}) + \Delta \leq \Delta$. Thus, $Q(\tilde{\boldsymbol{\theta}}) = 0$. Since Q has a unique minimum at $\boldsymbol{\theta}_0$, $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$. Hence, $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$ as $n \to \infty$. This result holds for almost every $\{\boldsymbol{e}_i\}_{i=1}^{\infty}$. Therefore, $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ a.s.

Appendix C. Proof of Theorem 2

It suffices to show

$$n\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\sum_{i=1}^{n}\boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{r}_i\boldsymbol{r}_i' - \boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0))\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n)\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n) \longrightarrow \boldsymbol{0} \quad \text{a.s.,} \tag{C.1}$$

$$n\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\sum_{i=1}^{n}[\boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n)\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n)-\boldsymbol{H}_i(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)\boldsymbol{H}_i'(\boldsymbol{\theta}_0)]\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\longrightarrow \mathbf{0}\ \text{ a.s.,} \tag{C.2}$$

and

$$n\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\sum_{i=1}^{n}\boldsymbol{H}_i(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)\boldsymbol{H}_i'(\boldsymbol{\theta}_0)\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n)-n\boldsymbol{D}_n(\boldsymbol{\theta}_0)\longrightarrow \mathbf{0}\quad\text{a.s.} \tag{C.3}$$

Proof of (C.1). First, we prove that

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{e}_i\boldsymbol{e}_i'-\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0))\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n)\longrightarrow \mathbf{0}\quad\text{a.s.} \tag{C.4}$$

It suffices to show that the (r,s)th entry of this $(p+q)\times(p+q)$ matrix,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}(e_{ij}e_{ik}-\sigma_{ijk}(\boldsymbol{\theta}_0))\frac{\partial h(t_{ij};\hat{\boldsymbol{\theta}}_n)}{\partial\theta_r}\frac{\partial h(t_{ik};\hat{\boldsymbol{\theta}}_n)}{\partial\theta_s}\longrightarrow 0\quad\text{a.s.}$$

Under (C5),

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}(e_{ij}e_{ik}-\sigma_{ijk}(\boldsymbol{\theta}_0))\frac{\partial h(t_{ij};\hat{\boldsymbol{\theta}}_n)}{\partial\theta_r}\frac{\partial h(t_{ik};\hat{\boldsymbol{\theta}}_n)}{\partial\theta_s}\\&-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}(e_{ij}e_{ik}-\sigma_{ijk}(\boldsymbol{\theta}_0))\frac{\partial h(t_{ij};\boldsymbol{\theta}_0)}{\partial\theta_r}\frac{\partial h(t_{ik};\boldsymbol{\theta}_0)}{\partial\theta_s}\longrightarrow 0\quad\text{a.s.,}\end{aligned}$$

by strong consistency of $\hat{\boldsymbol{\theta}}_n$ and the Cauchy-Schwarz inequality. Moreover, by (C5), (C10), and the Strong Law of Large Numbers,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}(e_{ij}e_{ik}-\sigma_{ijk}(\boldsymbol{\theta}_0))\frac{\partial h(t_{ij};\boldsymbol{\theta}_0)}{\partial\theta_r}\frac{\partial h(t_{ik};\boldsymbol{\theta}_0)}{\partial\theta_s}\longrightarrow 0\quad\text{a.s.}$$

Hence, (C.4) holds uniformly for $\boldsymbol{\theta}$ in $\boldsymbol{\Theta}$ by (C11). Under (C7), we have

$$n\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\sum_{i=1}^{n}\boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{e}_i\boldsymbol{e}_i'-\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0))\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n)\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\longrightarrow \mathbf{0}\quad\text{a.s.} \tag{C.5}$$

Second, from the Mean Value Theorem, $\boldsymbol{u}_i=\boldsymbol{h}_i(\boldsymbol{\theta}_0)-\boldsymbol{h}_i(\hat{\boldsymbol{\theta}}_n)=\boldsymbol{H}_i'(\boldsymbol{\theta}_n^*)(\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}_n)$, where $\boldsymbol{\theta}_n^*$ is a point on the line segment between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_n$. Thus, the (r,s)th entry of the $m_i\times m_i$ matrix $\boldsymbol{u}_i\boldsymbol{u}_i'$ is

$$\begin{aligned}u_{ir}u_{is}&=\sum_{j=1}^{p+q}\sum_{k=1}^{p+q}(\theta_{0j}-\hat{\theta}_{nj})(\theta_{0k}-\hat{\theta}_{nk})\left[\frac{\partial h(t_{ir};\boldsymbol{\theta}_n^*)}{\partial\theta_j}\quad\frac{\partial h(t_{is};\boldsymbol{\theta}_n^*)}{\partial\theta_k}\right]\\&\le\|\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}_n\|^2\left[\sum_{j=1}^{p+q}\left(\frac{\partial h(t_{ir};\boldsymbol{\theta}_n^*)}{\partial\theta_j}\right)^2\sum_{k=1}^{p+q}\left(\frac{\partial h(t_{is};\boldsymbol{\theta}_n^*)}{\partial\theta_k}\right)^2\right]^{1/2},\end{aligned}$$

by the Cauchy-Schwarz inequality. Thus, under (C10), there is a constant M such that

$$\max_{i\le n}\ u_{ir}u_{is} \le M\|\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n\|^2 \longrightarrow 0 \quad \text{a.s.}$$

Hence,

$$\max_{i\le n} \boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n)\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n) \le M\|\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}_n\|^2 \boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n)\boldsymbol{J}_i\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n) \longrightarrow \mathbf{0} \quad \text{a.s.,}$$

where $\boldsymbol{J}_i$ is an $m_i \times m_i$ matrix with all entries equal to 1, and therefore, under (C7),

$$n\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n)\sum_{i=1}^{n}\boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n)\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n)\boldsymbol{A}_n^{-1}(\hat{\boldsymbol{\theta}}_n) \longrightarrow \mathbf{0} \quad \text{a.s.} \tag{C.6}$$

Thus, (C.1) follows from (C.5), (C.6) and the Cauchy-Schwarz inequality.

Proof of (C.2). Under (C10), for any pair (r, s), $r, s = 1, \ldots, p+q$,

$$\max_i \sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\left|\frac{\partial h(t_{ij};\hat{\boldsymbol{\theta}}_n)}{\partial\theta_r}\ \frac{\partial h(t_{ik};\hat{\boldsymbol{\theta}}_n)}{\partial\theta_s} - \frac{\partial h(t_{ij};\boldsymbol{\theta}_0)}{\partial\theta_r}\ \frac{\partial h(t_{ik};\boldsymbol{\theta}_0)}{\partial\theta_s}\right| \longrightarrow 0 \quad \text{a.s.}$$

Thus, the (r, s) entry of $n^{-1}\sum_{i=1}^{n}[\boldsymbol{H}_i(\hat{\boldsymbol{\theta}}_n)\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)\boldsymbol{H}_i'(\hat{\boldsymbol{\theta}}_n) - \boldsymbol{H}_i(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_i(\boldsymbol{\theta}_0)\boldsymbol{H}_i'(\boldsymbol{\theta}_0)]$,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\sigma_{ijk}(\boldsymbol{\theta}_0)\left[\frac{\partial h(t_{ij};\hat{\boldsymbol{\theta}}_n)}{\partial\theta_r}\ \frac{\partial h(t_{ik};\hat{\boldsymbol{\theta}}_n)}{\partial\theta_s} - \frac{\partial h(t_{ij};\boldsymbol{\theta}_0)}{\partial\theta_r}\ \frac{\partial h(t_{ik};\boldsymbol{\theta}_0)}{\partial\theta_s}\right],$$

converges to 0 since $\{|\sigma_{ijk}(\boldsymbol{\theta})|\}_{i=1}^{\infty}$ is bounded under (C5). This proves (C.2) under (C7).

Proof of (C.3). (C.3) follows from (C7) and $\hat{\boldsymbol{\theta}}_n \longrightarrow \boldsymbol{\theta}_0$ a.s.

References

Boulanger, M. and Escobar, L. A. (1994). Experimental design for a class of accelerated degradation tests. *Technometrics* **36**, 260-272.

Carey, M. B. and Koenig, R. H. (1991). Reliability assessment based on accelerated degradation: a case study. *IEEE Trans. Reliability* **R-40**, 499-506.

Davidian, M. and Giltinan, D. M. (1993). Some simple methods for estimating intraindividual variability in nonlinear mixed effects models. *Biometrics* **49**, 59-73.

Doksum, K. A. (1991). Degradation rate models for failure time and survival data. *CWI Quarterly* **4**, 195-203.

Gallant, A. R. (1975). Seemingly unrelated nonlinear regressions. *J. Econometrics* **3**, 35-50.

Gallant, A. R. (1987). *Nonlinear Statistical Models.* John Wiley, New York.

Graybill, F. A. (1983). *Matrices with Applications in Statistics.* 2nd edition. Wadsworth, Belmont.

Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.

Lu, C. J. and Meeker, W. Q. (1993). Using degradation measures to estimate a time-to-failure distribution. *Technometrics* **35**, 161-174.

Lu, C. J., Meeker, W. Q. and Escobar, L. A. (1996). A comparison of degradation and failure-time analysis methods for estimating a time-to-failure distribution. *Statist. Sinica* **6**, 531-546.

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.

Meeker, W. Q., Escobar, L. A. and Lu, C. J. (1998). Accelerated degradation tests: modeling and analysis. *Technometrics* **40**, 89-99.

Nelson, W. (1981). Analysis of performance-degradation data from accelerated tests. *IEEE Trans. Reliability* **R-30**, 149-155.

Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis.* John Wiley, New York.

Phillips, P. C. B. (1976). The iterated minimum distance estimator and the quasi-maximum likelihood estimator. *Econometrica* **44**, 449-460.

Shao, J. and Chow, S.-C. (1994). Statistical inference in stability analysis. *Biometrics* **50**, 753-763.

Tseng, S.-T., Hamada, M. and Chiao, C.-H. (1995). Using degradation data to improve fluorescent lamp reliability. *Journal of Quality Technology* **27**, 363-369.

Wu, C. F. J. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9**, 501-513.

Wu, S.-J. (1996). Reliability analysis using the least squares method in nonlinear mixed-effect degradation models. Ph. D. dissertation, University of Wisconsin–Madison, Department of Statistics.

Zhuang, D. (1994). Degradation failure model and its statistical analysis. Ph.D. dissertation, East China Normal University, Department of Mathematical Statistics.

Department of Statistics, Tamkang University, Tamsui, Taipei, Taiwan.

E-mail: shuo@stat.tku.edu.tw

Department of Statistics, University of Wisconsin–Madison, U.S.A.

E-mail: shao@stat.wisc.edu

(Received March 1997; accepted December 1998)