2004

A Nonlinear Random Coefficients Model for Degradation Testing

Suk Joo Bae

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

A Nonlinear Random Coefficients Model for Degradation Testing

Suk Joo Bae **Paul H. Kvam**

Georgia Institute of Technology

April 29, 2004

Abstract

As an alternative to traditional life testing, degradation tests can be effective in assessing product reliability when measurements of degradation leading to failure can be observed. This article presents a degradation model for highly reliable light displays, such as plasma display panels (PDPs) or vacuum fluorescent displays (VFDs). Standard degradation models fail to capture the burn-in characteristics of VFDs, when emitted light actually increases up to a certain point in time before it decreases (or degrades) continuously. Random coefficients are used to model this phenomenon in a nonlinear way which allows for a nonmonotonic degradation path. In many situations, the relative efficiency of the lifetime estimate is improved over the standard estimators based on transformed linear models.

Keywords: Adaptive Gaussian Quadrature Approximation, Plasma Display Panel, Vacuum Fluorescent Displays, Reliability.

1 Introduction

Product testing presents a significant challenge to manufacturers of highly reliable components, such as integrated circuits, semiconductors, fiber optics in high speed computer networks or communication systems, plasma display panels (PDPs), vacuum fluorescent displays (VFDs), light emitting diodes (LEDs) and numerous other dependable systems. While products now are developed to last longer and perform more reliably, product test times have been reduced to meet the time-to-market requirements and to maximize a company's profit and competitiveness. Data gathered from component failures cannot always guarantee warranty contracts or safety standards will be satisfied for the system that contains that component.

If the experimenter can measure useful information about the product, such as *performance degradation*, product reliability inference can be greatly improved over regular failure-time data analysis. In a comparison of simple degradation modeling versus traditional failure-time analysis, Lu, *et al.* (1996) found that, in terms of asymptotic efficiency, degradation models are superior if degradation data are available. Analogous to accelerated life testing (ALT), accelerated degradation testing (ADT) provides the experimenter with more opportunities to draw inference on highly reliable test items, provided there is a known functional link that relates the harsh testing environment to the normal use environment.

1.1 Degradation Testing Literature

Degradation modeling has a rich history in the electronics manufacturing, materials testing and various fields of engineering. Nelson (1990, chapter 11) and Meeker and Escobar (1993) review the degradation literature, survey applications and describe basic analytical methods on ADT models. Meeker and Escobar (1998) provide a practical guide for ADT modeling based on transformed linear degradation models (e.g., log-linear models) along with standard ALT formulas.

There are numerous model developments for more specific degradation test experiments. Lu, *et al.* (1997) derive a transformed linear model for hot-carrier-induced degradation of metalized and oxidized semiconductors. By considering the sample size from degradation paths as random,

Su *et al.* (1999) develop a random coefficients model and used the maximum likelihood estimation (MLE) method to handle inconsistency problems found in least squares estimation (LSE). Shiau and Lin (1999) derive a nonparametric model for describing degradation of LEDs. Tseng and Wen (2000) propose a step-stress accelerated degradation test method to reduce experimental costs in assessing the lifetime distribution of LEDs. Yu and Tseng (1998) suggest an on-line procedure for determining an appropriate termination time for an ADT. Fagerstorm (1991) derives a differential equation model to estimate the time-to-failure distribution for lasers.

Important research applications of degradation models include the testing of LEDs (Fukada, 1991), Fluorescent Lamps (Tseng *et al.*, 1995), computer disks and compact discs (Murray, 1994), and power supplies (Chang, 1992).

1.2 Motivating Example

This research was motivated by degradation data on the PDP, a relatively new "emmissive" flat panel display which provides rich, accurate color fidelity. However, because the analyzed PDP data are proprietary, we instead focus on an analogous analysis based on VFDs which share degradation characteristics with PDPs (see Section 5). A VFD is a variation of the Triode Vacuum Tube under a high vacuum condition in a glass envelope, and is composed of three basic electrodes: the cathode (filament), anode (phosphor) and grid. The basic structure of the most commonly used VFDs is shown schematically in Figure 1. Electrons emitted from a cathode (filament) are accelerated by a grid, and collide with a phosphor coated surface of an anode electrode, producing luminescence.

PDPs and VFDs have several advantages over other display devices, including excellent visual recognition, operation at low voltage with lower power consumption, high reliability and a wide viewing angle. Luminosity (or brightness), measured in candela, is the critical characteristic of light display quality and effectiveness.

In the manufacture of VFDs, chemical processing produces impurities inside the device's vacuum tube. Impurities remain throughout the manufacturing process, degrading the display quality. Manufacturers burn off some harmful impurities before shipping them to customers. However,

owing to incomplete burn-in (called "aging" in the industry), the degradation path is not monotonic. Generally, degradation paths caused by incomplete burn-in appear in two kinds of patterns: a three-phase and two-phase pattern.

In a three-phase degradation, the accelerated grid powered by high-voltage serves to eliminate traces of impurities in the VFD, and as a result, the light intensity actually increases up to a certain unknown point in time. Then, due to a temporary "poisoning effect" of impurities on the cathode surface, the light intensity decreases rapidly for a short time period (phase two) before following its inherent path of slow degradation (phase three).

In testing environments where degradation is accelerated by using higher voltage, the first phase may not be discernable unless frequent early degradation measurements are taken. In this case, the light intensity decreases rapidly until an unknown point in time (phase one) before the degradation becomes more gradual (phase two). Eventually, after several hundred of hours of operation, the filament and phosphor anode will degrade, and the emitted light level decreases below a fixed threshold, when it is considered to have failed. The current VFD industry's standard defines this threshold as the time when a VFD's luminosity falls below 50% of its initial luminosity. The degradation measurements in Figure 2 illustrate the lack of stability in the initial "burn-in" stage of a VFD lifetime with two kinds of non-monotonic patterns.

This phenomenon presents a difficulty in modeling, because current parametric models are based on the assumption that degradation curves are simply monotonic. If the nonmonotonic behavior in the degradation data is ignored, the degradation model will likely be poorly fit, and the estimated lifetime distributions can be highly inaccurate.

To compensate for the lack of stability in observed degradation, Tseng *et al.* (1995) and Chiao and Hamada (1996) truncate the first several hundred hours assuming that after this much test time has been completed, the degradation path will be monotonic. However, this assumption is not suitable for these light displays. The monotonic model ignores the effects mentioned above that can cause one or two dramatic changes in the degradation curve, including the loss of monotonicity. Due to short life-cycles, the reliability evaluation is based on a relatively small amount of degradation

data measured for a short period of time, and these early truncated measurements can comprise a significant amount of the data in the experiment. As a consequence, we risk losing efficiency in the reliability analysis.

The paper is organized as follows. In Section 2, a general nonlinear random coefficients model is introduced. Estimation is based on the likelihood function of the degradation data. No simple, closed form expressions result and Section 3 compares various approximation methods. In Section 4, the failure time distribution for the VFD is derived, based on the estimated degradation curve. Confidence statements for the distribution function are constructed through bootstrap techniques. We feature the VFD data in Section 5, and compare the nonlinear random coefficients model to the more standard models in Section 6. In that section, we show conditions in which the nonmonotonic degradation model is superior.

2 Nonlinear Random coefficients Model

Random coefficients models provide a powerful tool for analyzing repeated-measurement data that arise in various fields of application, such as economics and pharmacokinetics (Davidian and Giltinan, 1995). Repeated-measurement data are generated by observing a number of individuals (or test units) repeatedly under differing experimental conditions, where the individuals are assumed to constitute a random sample from a population of interest. A common type of repeated-measurement data are *longitudinal data*, which are ordered by time or spatial position.

Random coefficients models are intuitively appealing because they allow for flexible variance-covariance structures of the response vector as well as linear random coefficients. Many developments for linear and nonlinear random coefficients models have occurred in recent years; see, for example, Beal and Sheiner (1982, 1988, 1992), Lindstrom and Bates (1990), Ramos and Pantula (1995) and Kim (1997). Davidian and Giltinan (1995) and Vonesh and Chinchilli (1997) provide thorough overviews as well as some general theoretical developments for nonlinear random coefficients models.

Although current methods for degradation testing use random coefficients models to handle

nonlinear forms of degradation (Meeker and Escobar, 1998), those models are linearized through one-to-one mappings such as the natural logarithm. This affects the model of measurement error but otherwise preserves the simplicity of a linear model. In this section, we consider nonlinear functions that cannot be simplified to linear regressions (i.e., nonmonotonic models). While the resulting degradation curves fit the data more closely, it is up for debate whether the introduction of a complex model will be justifiable for a particular application. In the case of light displays, we will see that the nonmonotonic behavior in the degradation supports such a model.

A general, nonlinear random coefficients model for the j^{th} response on the i^{th} individual test item can be defined as

$$y_{ij} = f(\boldsymbol{x}_{ij}, \boldsymbol{\beta}_i) + e_{ij} \qquad i = 1, \ldots, m, \qquad j = 1, \ldots, n_i, \tag{2.1}$$

where y_{ij} is the j^{th} response on the i^{th} individual, $\boldsymbol{x}_{ij}$ is the covariate vector for the j^{th} measurement on the i^{th} individual, $f(\cdot)$ is a nonlinear function of the covariate vector and parameter vector $\boldsymbol{\beta}_i$ and e_{ij} is a normally distributed random error term.

Modeling the i^{th} individual response is accomplished by letting $\boldsymbol{y}_i$ and $\boldsymbol{e}_i$ be the $(n_i \times 1)$ vectors of responses and random within-individual errors for individual i. Define the $(n_i \times 1)$ mean response vector $\boldsymbol{f}_i(\boldsymbol{\beta}_i) = (f(\boldsymbol{x}_{i1}, \boldsymbol{\beta}_i), \ldots, f(\boldsymbol{x}_{in_i}, \boldsymbol{\beta}_i))'$ for the i^{th} individual test item, depending on the $(p \times 1)$ individual-specific regression parameter $\boldsymbol{\beta}_i$. Suppose that $E(\boldsymbol{e}_i|\boldsymbol{\beta}_i) = \mathbf{0}$ and $Cov(\boldsymbol{e}_i|\boldsymbol{\beta}_i) = \sigma^2 \boldsymbol{G}_i^{1/2}(\boldsymbol{\beta}_i)\boldsymbol{G}_i^{1/2}(\boldsymbol{\beta}_i) = \sigma^2 \boldsymbol{G}_i(\boldsymbol{\beta}_i)$, where $\boldsymbol{G}_i^{1/2}(\boldsymbol{\beta}_i)$ denotes a $(n_i \times n_i)$ positive definite matrix which depends on i only through its dimension. The $(k \times 1)$ vector $\boldsymbol{b}_i$ represents a vector of random effects, and $\boldsymbol{\beta}$ is $(r \times 1)$ vector of fixed effects.

Based on the set-up in Sheiner and Beal (1985) and Davidian and Giltinan (1995), the general two-stage nonlinear random coefficients model can be written as follows:

Stage 1 (*within-individual variation*)

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}_i) + \boldsymbol{e}_i \tag{2.2}$$

Stage 2 (*between-individual variation*)

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i \tag{2.3}$$

where $\boldsymbol{b}_i$ are independent and identically distributed as Normal$(\mathbf{0}, \boldsymbol{D})$, and $\boldsymbol{A}_i$, $\boldsymbol{B}_i$ are known design matrices of size $(p \times r)$ and $(p \times k)$ for the fixed and random effects, respectively. $\boldsymbol{D}$ is a $(k \times k)$ covariance matrix and $\boldsymbol{\beta_i}$ in (2.3) is specific to the i^{th} test item through $\boldsymbol{b}_i$, allowing the conditional moments of $\boldsymbol{e}_i$ to be expressed as $\boldsymbol{e}_i|\boldsymbol{b}_i \sim N(\mathbf{0}, \sigma^2 \boldsymbol{G}_i(\boldsymbol{\beta}_i))$.

Within the framework of (2.2) and (2.3), the random effects are unobserved quantities, and maximum likelihood estimation in (2.2) is based on the marginal density of $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_m')'$

$$\boldsymbol{p}(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) = \int \boldsymbol{p}(\boldsymbol{y}|\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}, \sigma^2)\boldsymbol{p}(\boldsymbol{b})d\boldsymbol{b}. \tag{2.4}$$

In general this integral does not have a closed-form expression when the model function f is nonlinear in $\boldsymbol{b}$, and approximation methods are used to help solve the estimations.

3 Approximation Methods

In this section, we consider four different approximations to the log-likelihood corresponding to (2.4): a first-order method (Beal and Sheiner, 1982, 1985), Lindstrom and Bates's (1990) algorithm, adaptive importance sampling (Pinheiro and Bates, 1995), and adaptive Gaussian quadrature (Pinheiro and Bates, 1995). For two-stage nonlinear models, the marginal distribution of $\boldsymbol{y}_i$ is based on the random effects $\boldsymbol{b}_i$ and the individual error vectors $\boldsymbol{e}_i$. Suppose that

$$\boldsymbol{e}_i = \boldsymbol{G}_i^{1/2}(\boldsymbol{\beta}_i)\boldsymbol{\epsilon}_i, \tag{3.5}$$

where the random error $\boldsymbol{\epsilon}_i$ is assumed to be independent of $\boldsymbol{b}_i$, and normally distributed with mean zero and covariance matrix $\sigma^2 \boldsymbol{I}_{n_i}$. $\boldsymbol{G}_i^{1/2}(\boldsymbol{\beta}_i)$ is the Cholesky decomposition of $\boldsymbol{G}_i(\boldsymbol{\beta}_i)$, representing the heteroscedasticity or correlation among within-individual errors. Then, the first-stage model may be rewritten as

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i) + \boldsymbol{G}_i^{1/2}(\boldsymbol{\beta}, \boldsymbol{b}_i)\boldsymbol{\epsilon}_i. \tag{3.6}$$

In many practical applications involving inference on the degradation of individuals, within-individual heteroscedasticity and correlation are dominated by between-individual variation $\boldsymbol{b}_i$ and

thus can be ignored. $\boldsymbol{G}_i(\boldsymbol{\beta}_i) = \boldsymbol{I}_{n_i}$ is the common specification of uncorrelated within-individual errors with constant variance.

Estimation of the population parameters $\boldsymbol{\beta}$ can be interpreted as the "typical" fixed-effect value, and also represents the parameter values producing a "typical" $\boldsymbol{y}_i$ response. This issue has been discussed by Zeger, *et al.* (1988), who introduced the terminology of population-average models, where attention focuses on inference in the marginal distribution and subject-specific parameters, which pertain to inference in the conditional distribution for a given subject (see also Lindstrom and Bates (1990)).

3.1 First-Order Approximation

The Taylor series expansion of (3.6) about $E(\boldsymbol{b}_i) = \boldsymbol{0}$ is

$$\boldsymbol{y}_i \approx \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{0}) + \boldsymbol{\Lambda}_i(\boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{\Delta}_{\boldsymbol{b}_i}(\boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{b}_i + \boldsymbol{G}_i^{1/2}(\boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{\epsilon}_i \tag{3.7}$$

where $\boldsymbol{\Lambda}_i(\boldsymbol{\beta}, \boldsymbol{0})$ is the $(n_i \times p)$ matrix of derivatives of $\boldsymbol{f}_i(\boldsymbol{\beta}_i)$ with respect to $\boldsymbol{\beta}_i$ and $\boldsymbol{\Delta}_{\boldsymbol{b}_i}(\boldsymbol{\beta}, \boldsymbol{0})$ is the $(p \times k)$ matrix of derivatives of $\boldsymbol{\beta}_i$ with respect to $\boldsymbol{b}_i$, evaluated at $\boldsymbol{b}_i = \boldsymbol{0}$.

Defining the $(n_i \times k)$ matrix $\boldsymbol{Z}_i(\boldsymbol{\beta}, \boldsymbol{0}) = \boldsymbol{\Lambda}_i(\boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{\Delta}_{\boldsymbol{b}_i}(\boldsymbol{\beta}, \boldsymbol{0})$ and $\boldsymbol{e}_i^* = \boldsymbol{G}^{1/2}(\boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{\epsilon}_i$, equation (3.7) may be shortened to $\boldsymbol{y}_i \approx \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{0}) + \boldsymbol{Z}_i(\boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{b}_i + \boldsymbol{e}_i^*$. The marginal mean and covariance of $\boldsymbol{y}_i$ can be specified as $\mathrm{E}(\boldsymbol{y}_i) \approx \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{0})$, $\mathrm{Cov}(\boldsymbol{y}_i) \approx \sigma^2\boldsymbol{G}_i(\boldsymbol{\beta}, \boldsymbol{0}) + \boldsymbol{Z}_i(\boldsymbol{\beta}, \boldsymbol{0})\boldsymbol{D}\boldsymbol{Z}_i'(\boldsymbol{\beta}, \boldsymbol{0}) \equiv \boldsymbol{\Sigma}_i(\boldsymbol{\beta}, \boldsymbol{0}, \boldsymbol{\omega})$ where $\boldsymbol{\omega}$ is a vector containing the unknown covariance parameters σ^2 and $\boldsymbol{D}$.

Beal and Sheiner (1982, 1985) utilize extended least squares (ELS) to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$. Under the previous normality assumptions, ELS is equivalent to joint maximum likelihood estimation of $(\boldsymbol{\beta}, \boldsymbol{\omega})$, which is known as the *first-order method* and minimizes the objective function:

$$Q_{FO}(\boldsymbol{\beta}, \boldsymbol{\omega}) = \sum_{i=1}^{m} \{\log|\boldsymbol{\Sigma}_i(\boldsymbol{\beta}, \boldsymbol{0}, \boldsymbol{\omega})| + (\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{0}))'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{0}, \boldsymbol{\omega})(\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{0}))\}. \tag{3.8}$$

They note that ELS estimators are consistent and asymptotically normal under some regularity conditions provided that the first and second moments of $\boldsymbol{y}_i$ are correctly specified. To minimize (3.8), Beal and Sheiner (1982) employ a derivative-free quasi-Newton algorithm which avoids the need for specifying $\partial\boldsymbol{\Sigma}_i(\boldsymbol{\beta}, \boldsymbol{0}, \boldsymbol{\omega})/\partial\boldsymbol{\beta}$.

3.2 Lindstrom-Bates Algorithm

Lindstrom and Bates (1990) suggest an alternative to the first-order method by considering a linearization of (3.6) in the random effects about some value $\boldsymbol{b}_i^*$ that is closer to $\boldsymbol{b}_i$ than its expectation $\mathbf{0}$. The Taylor series expansion about $\boldsymbol{b}_i = \boldsymbol{b}_i^*$ yields

$$\boldsymbol{y}_i \approx \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i^*) + \boldsymbol{\Lambda}_i(\boldsymbol{\beta}, \boldsymbol{b}_i^*)\boldsymbol{\Delta}_{\boldsymbol{b}_i}(\boldsymbol{\beta}, \boldsymbol{b}_i^*)(\boldsymbol{b}_i - \boldsymbol{b}_i^*) + \boldsymbol{G}_i^{1/2}(\boldsymbol{\beta}, \boldsymbol{b}_i^*)\boldsymbol{\epsilon}_i. \tag{3.9}$$

where $\boldsymbol{\Lambda}_i(\cdot), \boldsymbol{\Delta}_{\boldsymbol{b}_i}(\cdot)$ are defined as in (3.7). This notation makes explicit the fact that the expansion in $\boldsymbol{b}_i$ has been taken about $\boldsymbol{b}_i^*$.

In order to construct an estimation procedure for $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$, reasonable values of $\boldsymbol{b}_i^*$ must first be selected. Then, treating the estimate as fixed, either generalized least squares (GLS) or maximum likelihood (ML) can be used to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$. Lindstrom and Bates (1990) propose the following iterated two-step estimation algorithm:

1. *Pseudo-data (PD) step:* Given the current estimate $\hat{\boldsymbol{\omega}}$ of $\boldsymbol{\omega}$, minimize

$$\sum_{i=1}^{m}\{\log|\hat{\boldsymbol{D}}| + \boldsymbol{b}_i'\hat{\boldsymbol{D}}^{-1}\boldsymbol{b}_i + \log|\hat{\sigma}^2\boldsymbol{G}_i(\hat{\omega})| + \hat{\sigma}^{-2}\left(\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\right)'\boldsymbol{G}_i^{-1}(\hat{\omega})\left(\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\right)\} \tag{3.10}$$

   with respect to $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$. The resulting estimates are denoted by $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{\beta}}_0$.

2. *Linear mixed effects (LME) step:* Estimate $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ with the values $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\omega}}$ that minimize

$$Q_{LB}(\boldsymbol{\beta}, \omega) = \sum_{i=1}^{m}\{|\log\boldsymbol{\Sigma}_i(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i, \boldsymbol{\omega})| + \boldsymbol{r}_i'\boldsymbol{\Sigma}_i^{-1}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i, \boldsymbol{\omega})\boldsymbol{r}_i\} \tag{3.11}$$

   where $\boldsymbol{r}_i \equiv \boldsymbol{r}_i(\boldsymbol{\beta}, \hat{\boldsymbol{b}}_i, \hat{\boldsymbol{\beta}}_0) = \boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \hat{\boldsymbol{b}}_i) + \boldsymbol{Z}_i(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i)\hat{\boldsymbol{b}}_i$.

As an alternative to (3.11), one can use a restricted maximum likelihood (REML) approach which minimizes

$$Q_{LB,REML}(\boldsymbol{\beta}, \omega) = Q_{LB}(\boldsymbol{\beta}, \omega) + \log|\boldsymbol{X}_i'(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i)\boldsymbol{\Sigma}_i^{-1}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i, \hat{\boldsymbol{\omega}})\boldsymbol{X}_i(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i)|, \tag{3.12}$$

where $\boldsymbol{X}_i(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i) = \boldsymbol{\Lambda}_i(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i)\boldsymbol{\Delta}_{\boldsymbol{\beta}_i}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i)$, and $\boldsymbol{\Delta}_{\boldsymbol{\beta}_i}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{b}}_i)$ is the $(p \times r)$ matrix of derivatives of $\boldsymbol{\beta}_i$ with respect to $\boldsymbol{\beta}$ at $\boldsymbol{b}_i = \hat{\boldsymbol{b}}_i$.

The algorithm alternates between the PD and LME steps until some convergence criterion is met, with final estimates denoted by $(\hat{\boldsymbol{\beta}}_{LB}, \hat{\boldsymbol{\omega}}_{LB}, \hat{\boldsymbol{b}}_{i,LB})$. The $\hat{\boldsymbol{b}}_{i,LB}$ may be regarded as empirical Bayes estimates of the random effects; these estimates are approximate posterior modes for $\boldsymbol{b}_i$, where the fixed parameters $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ are replaced by estimates. Wolfinger (1993) also showed that LME approximation to the restricted log-likelihood (3.12) is equivalent to a Laplacian approximation to the integral (2.4) when a flat prior is assumed for $\boldsymbol{\beta}$.

3.3 Adaptive Importance Sampling Approximation

Importance sampling is a simple and efficient method for performing Monte Carlo integration. Geweke (1989) developed the methods for the systematic application of Monte Carlo integration with importance sampling to Bayesian inference in econometric models. Successful convergence relies mainly upon the choice of the importance distribution from which the sample is drawn and the importance weights are calculated. If this distribution is hard to find by integrating the marginal density, an easily sampled approximation can be used. From (2.4), the marginal density of $\boldsymbol{y}_i$ can be written as

$$\boldsymbol{p}(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) = \int (2\pi\sigma^2)^{-n_i/2}|\boldsymbol{D}|^{-1/2}\exp(-\boldsymbol{v}_i(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{D})/2)d\boldsymbol{b}_i, \quad \text{where}$$
$$\boldsymbol{v}_i(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{D}) = \boldsymbol{b}_i'\boldsymbol{D}^{-1}\boldsymbol{b}_i + \sigma^{-2}(\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i))'\boldsymbol{G}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{b}_i)(\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)).$$

For the nonlinear random coefficients model, this integrand is approximately normally distributed, giving a natural choice for the importance distribution.

Define N_{IS} as the number of importance samples to be chosen. Importance samples can be generated by selecting a standard normal vector $\boldsymbol{z}^*$ and calculating the sample of random effects as $\boldsymbol{b}_i^* = \hat{\boldsymbol{b}}_i^* + \boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})^{-1/2}\boldsymbol{z}^*$, where $\hat{\boldsymbol{b}}_i^*$ is the value of $\boldsymbol{b}_i$ minimizing $\boldsymbol{v}_i(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{D})$ and $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})$ is the approximation for the second derivatives of $\boldsymbol{v}_i(\boldsymbol{\beta}, \boldsymbol{b}_i, \boldsymbol{D})$ with respect to $\hat{\boldsymbol{b}}_i^*$. If we let $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})^{-1/2}$ denote the inverse of the Cholesky factor of $\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})$, the importance sampling approximation to the log-likelihood of $\boldsymbol{y}$ can be written as

$$Q_{IS}(\beta, \boldsymbol{D}, \sigma^2|\boldsymbol{y}) = -\frac{1}{2}\left(N\log(2\pi\sigma^2) + m\log|\boldsymbol{D}| + \sum_{i=1}^{m}\log|\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})|\right) + \sum_{i=1}^{m}\log\left(\sum_{j=1}^{N_{IS}}\exp\{-\boldsymbol{v}_i(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{b}_{ij}^*)/2 + \|\boldsymbol{z}_j^*\|^2/2\}/N_{IS}\right),$$

where $N = \sum_{i=1}^{m} n_i$. In the case the model function is linear in $\boldsymbol{b}$, the right side of (2.4) is $\boldsymbol{p}(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\beta}, \boldsymbol{D}, \sigma^2)\boldsymbol{p}(\boldsymbol{b}_i) = \boldsymbol{p}(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2)\phi(\hat{\boldsymbol{b}}_i, \boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})^{-1})$ where ϕ is the Gaussian density function. In this case, the results are exact and the importance sampling weights are equal to $\boldsymbol{p}(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2)$.

3.4 Adaptive Gaussian Quadrature Approximation

The Gaussian quadrature approximates the integral of a function with respect to a given kernel by a weighted sum over predefined abscissas for the random effects. Unlike other numerical integration techniques, the abscissas are unevenly spaced throughout the interval of integration. With a modest number of quadrature points, along with appropriate centering and scaling of the abscissas, the Gaussian quadrature approximation can be highly effective (see Abramowitz and Stegun, 1964 for details). We consider an importance sample version of the Gaussian quadrature rule, which we denote by *adaptive Gaussian quadrature*.

For $j = 1, \ldots, N_{GQ}$, suppose that (z_j^*, w_j) are, respectively, the standard Gauss-Hermite abscissas and weights for the Gaussian quadrature rule with N_{GQ} points (Golub and Welsch, 1969). The adaptive Gaussian quadrature is then given by

$$\begin{aligned}\int \exp[-\boldsymbol{v}_i(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{b}_i)/2]d\boldsymbol{b}_i &= \int |\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})|^{-1/2} e^{-\boldsymbol{v}_i[\boldsymbol{\beta}, \boldsymbol{D}, \hat{\boldsymbol{b}}_i^* + \boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})^{-1/2}\boldsymbol{z}^*]/2 + \|\boldsymbol{z}^*\|^2/2} e^{-\|\boldsymbol{z}^*\|^2/2} d\boldsymbol{z} \\ &\simeq |\boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})|^{-1/2} \sum_{j_1=1}^{N_{GQ}} \cdots \sum_{j_k=1}^{N_{GQ}} e^{-\boldsymbol{v}_i[\boldsymbol{\beta}, \boldsymbol{D}, \hat{\boldsymbol{b}}_i^* + \boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})^{-1/2}\boldsymbol{z}_j^*]/2 + \|\boldsymbol{z}^*\|^2/2} \prod_{q=1}^{k} w_{j_q},\end{aligned}$$

where $\boldsymbol{z}_j^* = [z_{j_1}^*, \ldots, z_{j_k}^*]'$. The objective function to minimize corresponds to negative log-likelihood:

$$Q_{AGQ}(\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) = -\frac{1}{2}\left(N\log(2\pi\sigma^2) + m\log|\boldsymbol{D}| + \sum_{i=1}^{m}\log|bV_i(\boldsymbol{\beta}, \boldsymbol{D})|\right) + \sum_{i=1}^{m}\log\left(\sum_{j=1}^{N_{GQ}}\exp\left\{-\boldsymbol{v}_i[\boldsymbol{\beta}, \boldsymbol{D}, \hat{\boldsymbol{b}}_i^* + \boldsymbol{V}_i(\boldsymbol{\beta}, \boldsymbol{D})^{-1/2}\boldsymbol{z}_j^*]/2 + \|\boldsymbol{z}_j\|^2/2\right\}\prod_{q=1}^{k} w_{j_q}\right).$$

When $N_{GQ} = 1$, adaptive Gaussian quadrature approximation is equivalent to Laplacian approximation, which serves as a Bayesian method for estimating marginal posterior densities and predictive distributions, because in this case $z_1 = 0$ and $w_1 = 1$. This Adaptive Gaussian quadrature approximation is similar to the approximation obtained from adaptive importance sampling; the basic difference is that the former uses fixed abscissas and weights, but the latter allows them to be determined by a pseudo-random mechanism. As with the importance sampling approximation, the adaptive Gaussian quadrature produces the exact log-likelihood when the model function is linear in random effects $\boldsymbol{b}$. In practice, $N_{GQ} \leq 7$ generally suffices and $N_{GQ} = 1$ often provides a reasonable approximation (Pinheiro and Bates, 1995).

4 Failure-Time Distribution

To derive the failure-time distribution and its quantiles, define failure time T as the time that the actual degradation path $\tau(t; \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{e})$ reaches the prespecified degradation level τ_f. Then the distribution of the failure time is

$$F_T(t) = Pr(T \leq t) = Pr[\tau(t; \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{e}) \leq \tau_f]. \tag{4.13}$$

The failure-time distribution depends on the distribution of the random coefficient $\boldsymbol{b}$, which is determined by the variance-covariance matrix $\boldsymbol{D}$. In this section, the failure-time distributions under two different approaches to degradation modeling are considered: the linear random coefficients (LRC) model and the nonlinear random coefficients (NRC) model.

In many cases, the LRC model reveals a closed-form expression $F_T(t)$ and the computations are straightforward. Tseng *et al.* (1995) and Chiao and Hamada (1996) modeled (transformed) luminosity degradation for fluorescent lamps and LEDs using a linear random coefficients model after truncating the first few hundred hours on the test when degradation was unstable.

Let $y(t)$ be the luminosity at time t and $y(t_0)$ be the baseline luminosity measured at the burn-in

time t_0. The LRC model for observed degradation, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$, is

$$\tau(t_{ij}; \boldsymbol{\beta}, \boldsymbol{b}_i, e_{ij}) = \log(y(t_{ij})) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})(t_{ij} - t_0) + e_{ij}, \tag{4.14}$$

with fixed effect $\boldsymbol{\beta} = (\beta_0, \beta_1)$ and random coefficient $\boldsymbol{b}_i = (b_{0i}, b_{1i})$ which characterizes the item-to-item variation. We assume $\boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{D})$ with elements $D_{00} = \sigma_0^2$, $D_{11} = \sigma_1^2$, and $D_{01} = D_{10} = \rho\sigma_0\sigma_1$, and $\boldsymbol{b}_i$ is independent of the error term $e_{ij} \sim \mathcal{N}(0, \sigma^2)$. The failure-time distribution for model (4.14) is

$$\begin{aligned} F_T(t) &= Pr[(\beta_0 + b_{0i}) + (\beta_1 + b_{1i})(t - t_0) \leq \tau_f] \\ &\approx \Phi\left\{ \frac{[\beta_0 + \beta_1(t - t_0)] - \tau_f}{[\sigma_0^2 + (t - t_0)^2\sigma_1^2 + 2(t - t_0)\rho\sigma_0\sigma_1]^{1/2}} \right\} \qquad t > t_0, \end{aligned} \tag{4.15}$$

where $\Phi\{\cdot\}$ is the standard normal distribution function. The p^{th} quantile of the failure-time distribution is the value of t_p such that $Pr(T \leq t_p) = p$, so that

$$z_p = \frac{(\beta_0 + \beta_1 t_p') - \tau_f}{[\sigma_0^2 + t_p'^2\sigma_1^2 + 2t_p'\rho\sigma_0\sigma_1]^{1/2}}, \tag{4.16}$$

where $t_p' = t_p - t_0$ and z_p is the p^{th} quantile of the standard normal distribution.

The MLEs for the parameters in the failure-time distribution can be computed by using the Newton-Raphson procedure described in Lindstrom and Bates (1988). MLEs of the $F_T(t)$ and t_p are then found by replacing the model parameters in (4.15) and (4.16), respectively, with their estimates. In the case τ_f is a relative threshold (e.g., set to 50% of the starting degradation level), estimation of $F_T(t)$ is complicated due to τ_f depending on the degradation so Monte Carlo simulation is used to evaluate $\hat{F}_T(t)$. Pointwise confidence intervals for $F_T(t)$ can be obtained by using a parametric bootstrap procedure, similar to Meeker and Escobar (1998). Confidence intervals for quantiles are established from these methods. The procedure for generating $\hat{F}_T(t)$, and confidence intervals for $F_T(t)$ are similar as that in the NRC model, illustrated below.

In the NRC model (2.1), if there is no closed-form expression for $\hat{F}_T(t)$ or if the numerical transformation methods are overly complicated, we can choose to evaluate $\hat{F}_T(t)$ using Monte Carlo simulation. For this evaluation, we first use the model parameter estimates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{b}}$, and $\hat{\boldsymbol{D}}$ (obtained from the m sample paths) to generate the N simulated realizations $\tilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{b}}$. From N values

of $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{b}}$, compute the N failure times $\tilde{t}$ by substituting $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{b}}$ into $\tau(t;\boldsymbol{\beta},\boldsymbol{b})$, and then solve for τ_f. For any desired values of t, $F_T(t)$ is estimated from the simulated empirical distribution

$$\hat{F}_T(t) \approx \frac{\text{number of } \tilde{t} \leq t}{N}. \tag{4.17}$$

The procedure for constructing parametric bootstrap confidence intervals is implemented with the following steps.

Step 1. From the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{D}}$, and $\hat{\sigma}^2$ (hereafter, assuming that $\boldsymbol{G}_i(\boldsymbol{\beta}_i) = \boldsymbol{I}_{n_i}$) obtained by using ML or approximation methods, generate m simulated realizations of $\boldsymbol{\beta}_i^*$, $\boldsymbol{b}_i^*, i = 1, \ldots, m$.

Step 2. Compute m simulated degradation pathes

$$y_{ij}^* = \tau(t_{ij}; \boldsymbol{\beta}^*, \boldsymbol{b}_i^*) + e_{ij}^* \tag{4.18}$$

up to the specified stopping time t_c, where the e_{ij}^* values are generated from $\mathcal{N}(0, \hat{\sigma}^2)$, giving bootstrap estimates $\hat{\boldsymbol{\beta}}^*$, $\hat{\boldsymbol{b}}^*$, and $\hat{\boldsymbol{D}}^*$.

Step 3. Following the Monte Carlo simulation mentioned above, compute the bootstrap estimates $\hat{F}_T^*(t)$ at desired values of t with $\hat{\boldsymbol{\beta}}^*$, $\hat{\boldsymbol{b}}^*$, and $\hat{\boldsymbol{D}}^*$.

Step 4. Repeat Step 1-3 B times ($B \geq 1,000$), then obtain the bootstrap estimates $\hat{F}_T^*(t)_1$,$\hat{F}_T^*(t)_2$, ..., $\hat{F}_T^*(t)_B$. Sort the B bootstrap estimates in increasing order giving $\hat{F}_T^*(t)_{[b]}, b = 1, \ldots, B$.

Step 5. Following Efron and Tibshirani (1993), determine the lower and upper bounds of pointwise $100(1-\alpha)\%$ confidence intervals for the distribution function $F_T(t)$:

$$[\underline{F_T}(t), \overline{F_T}(t)] = [\hat{F}_T^*(t)_{[l]}, \hat{F}_T^*(t)_{[u]}], \tag{4.19}$$

where $l = B \times \Phi[2\Phi^{-1}(q) + \Phi^{-1}(\alpha/2)]$, $u = B \times \Phi[2\Phi^{-1}(q) + \Phi^{-1}(1-\alpha/2)]$ and

$$q = \frac{\text{number of } \hat{F}_T^*(t)_b \leq \hat{F}_T(t)}{B}, \qquad b = 1, \ldots, B. \tag{4.20}$$

5 VFD example

As mentioned in Section 1.2, this NRC model (2.1) is motivated by degradation of PDPs, but due to the proprietary nature of the experiment, we focus instead on VFDs, which have analogous degradation characteristics. Unfortunately, the VFD data are relatively sparse compared to the more recent PDP data that motivated the study (more PDP units were tested at a larger variety of test levels). As a consequence, the example below admits a casual disregard for parsimony for the sake of illustration.

The VFD degradation was accelerated by using both increased voltage and temperature during product testing. Because the ALT link function was assumed to be completely known in the manufacturer's lifetime analyses, the information from the link function is not necessary to illustrate the nonlinear random coefficients model. In this example degradation is observed at a single test level. Degradation tests of VFD luminosity are generally executed up to 1,000 hours for customer warranty purposes. In a special extended test, luminosity was also measured at 3,000 hours to check the accuracy of the failure-time estimates derived from the field degradation tests that are censored at 1,000 hours.

5.1 Comparison of Approximation Methods

In this subsection, the four approximation methods are compared using the VFD testing data. The individual VFD degradation paths consist of measurements of VFD luminosity (at six different time points) for five VFDs, taken over a period of 3,000 hours. Model fitting is especially challenging here due to the sparseness of the data and the complexities required of a nonlinear model.

The lifetime of VFDs is limited by the evaporation of electrons deposited on the cathode. The degradation (evaporation) rate is constant over time, that is, $dA/dt = -\lambda$, where $\lambda > 0$. Consequently the amount of degradation in luminosity at time t is $A = \beta_1 \exp[-\beta_2 t], \quad \beta_1, \beta_2, t > 0$, where β_1 denotes initial luminosity. However, some impurities remain during the initial operation, degrading the display quality. Because the impurities decrease gradually during operation ("operation burn-in"), the brightness can increase in time. Combining the cathode degradation and the

effect of burn-in, the degradation model of a luminosity can be expressed by

$$E[y_t] = \frac{\beta_1 t^{\beta_2}}{\exp[\beta_3 t]}. \tag{5.21}$$

The impurity burn-in rate and the cathode degradation rate are not separable; to model this conflictive behavior in luminosity degradation explicitly, we introduce the following four-parameter (nonmonotonic) model to the degradation data:

$$E[y_t] = \frac{\beta_1 t^{\beta_2}}{\exp[\beta_3 t^{\beta_4}]}, \qquad t > 0. \tag{5.22}$$

With both β_2 and β_4 in the model, (5.22) provides great flexibility in describing two and three-phase degradation patterns.

This model has precedence in analyzing longitudinal data as described in Vonesh and Chinchilli (1997). As a practical consideration, models with fewer than four parameters failed to characterize the nonmonotonic degradation in VFD and PDP data. Furthermore, a fixed effects model (with four parameters) also fails to adequately describe the degradation; if we ignore the grouping of brightness measurements according to individual units and fit a single model to the collective VFD paths, then using nonlinear least squares, we would estimate $E[y_t] = (11259.7 \cdot t^{0.416})/(\exp[2.560 \cdot t^{0.116}])$, $t > 0$. Figure 3 shows boxplots for the computed residuals resulting from fitting this model for each VFD, and clearly indicates inadequate model fit. If a single brightness curve represents the collective VFDs, the differences between units are confounded with the measurement error in the residuals, thus inflating the residual standard error estimate (see Pinheiro and Bates, 2000, *p.* 279). To accommodate this between-unit variation, a separate nonlinear model was fitted to each unit; the individual specific nonlinear model shows an improved fit; the resulting boxplots of the residuals are displayed in Figure 4.

Despite the improved fit, Table 1 shows that from this separate nonlinear model fitting to each VFD, the parameter estimates for the second and third subjects are highly unstable and resulting residual variance is higher, which is a consequence of overfitting sparse data with a complex model. As a result, the two-stage estimation method suggested by Lu and Meeker (1993), under the assumption that individual-specific regression parameters are estimable for each subject, is less

suited for the VFD example. Instead, an approximation method should be used to compute the nonlinear model.

To recap, the data suggest that the VFD degradation model should include variability among and within individual units, which is modeled most effectively using random coefficients. Parameters of the nonlinear random coefficients model can be estimated using one of the approximation methods outlined in the last section. An adaptive Gaussian quadrature (with $N_{GQ} = 1$) was used to decide which of the coefficients in the model require random effects to account for between-unit variation (this is the default method in SAS^{TM} NLMIXED procedure).

The degradation model must simultaneously reflect the variation in initial luminosity as well as the degradation rate, thus initial luminosity (β_1 by assuming $t = 1$) varies from unit to unit. Further analysis revealed that a model with two random coefficients provides a superior fit to the model with just one; the average prediction error was 10 times larger for the simpler model. The final selected nonlinear model, then, is based on four unknown parameters $(\beta_1, ..., \beta_4)$ and two random effects (b_1, b_2):

$$y_{ij} = \frac{(\beta_1 + b_{i1}) \cdot t_{ij}^{\beta_2}}{\exp[(\beta_3 + b_{i2}) \cdot t_{ij}^{\beta_4}]} + e_{ij}, \tag{5.23}$$

where y_{ij} and t_{ij} represent the j^{th} luminosity response and measurement time on the i^{th} VFD, respectively. The random effects (b_{i1}, b_{i2}), $i = 1, \ldots, 5$ are *i.i.d.* $\mathcal{N}(\mathbf{0}, \sigma_b^2 \boldsymbol{I}_2)$, and independent of the error term.

Table 2 shows estimation results using the first-order approximation, the Laplacian approximation, the adaptive importance sampling approximation, and the adaptive Gaussian quadrature approximation. The NLMIXED procedure and NLINMIX macro program in SAS^{TM} were used for all of the approximations. The adaptive Gaussian approximation method is based on three quadrature points; when $N_{GQ} > 5$, the method fails to converge consistently. The parameter estimates from the Laplacian method are based on the adaptive Gaussian quadrature with $N_{GQ} = 1$. The Lindstrom-Bates algorithm does not converge in some regions of the parameter space, so results from the LB approximation are not listed in the table. The correlation between random effects was computed to be negligible over all approximation methods.

For this example, Table 2 shows that the results are similar for all approximations except the first-order approximation, for which the variance of the error term is 17 times larger than the other approximations. Estimates of the fixed-effects in the VFD degradation model are dramatically different for the first-order approximation because the random-effect parameters (b_{i1}, b_{i2}) do not enter the model linearly.

Three criteria for selecting the most suitable approximation are listed in Table 3. Comparisons are based on log-likelihood, relative bias and the computational efficiency. The relative bias is the absolute difference between the real response and prediction value divided by the real response. Computational efficiency is measured in terms the number of function evaluations until convergence is achieved. Of the three approximation methods still under consideration (Laplacian, Adaptive Importance Sampling, Adaptive Gaussian Quadrature), the Laplacian approximation is considered the best under each of the three criteria. Obviously, the computational efficiency for the first-order approximation is not directly comparable to other approximations. The final parameter estimates and their standard errors, based on the Laplacian approximation, are tabulated in Table 4. The standard errors are constructed from the variance-covariance matrix computed as the inverse Hessian matrix. In the Laplacian approximation, Hessian matrix $\partial^2 v_i(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{b}_i)/\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i'$ is approximated by

$$\simeq \frac{\partial \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}\Big|_{\hat{\boldsymbol{b}}_i} \frac{\partial \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)}{\partial \boldsymbol{b}_i'}\Big|_{\hat{\boldsymbol{b}}_i} + \sigma^2 \boldsymbol{D}^{-1}, \tag{5.24}$$

where $\hat{\boldsymbol{b}}_i = \arg\min_{\boldsymbol{b}_i} \boldsymbol{v}_i(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{b}_i)$. The statistical inference about the parameters and their asymptotic properties are demonstrated in detail by Pinheiro and Bates (2000).

5.2 Comparison of Degradation Models

For the LRC model in (4.14), degradation measurements earlier than $t_0 = 240$ hours (10 days) are discarded to increase the chance that the resulting degradation process is monotonically decreasing. Parameters were estimated using maximum likelihood: $\hat{\boldsymbol{\beta}} = (6.7, -1.92 \times 10^{-4})'$, $\hat{D}_{00} = 2.38 \times 10^{-3}$, $\hat{D}_{11} = 4.99 \times 10^{-10}$, $\hat{D}_{01} = \hat{D}_{10} = -5.13 \times 10^{-8}$, $\sigma^2 = 6.88 \times 10^{-3}$.

The LRC model failed to accurately fit an individual degradation path, even after the initial

measurements from the burn-in stage are discarded. This is because the LRC model ignores the poisoning effect (with steep degradation) that follows the initial burn-in stage. In contrast, the NRC model fit (based on parameter estimates resulting from the Laplacian approximation) successfully captured the burn-in characteristic of VFDs. The predicted values for response, along with the average relative prediction biases are obtained by plugging the estimates into the LRC model along with the estimates in Table 4 from the NRC model. Prediction bias is much higher in the LRC model, especially for samples 2 and 3 which have an obvious nonmonotonic behavior. The average prediction bias for the LRC model (0.0627) is 60% higher than that of the NRC model (0.0393), which shows that the prediction from the NRC model is substantially more reliable.

For comparison, using Monte Carlo simulation (with $N = 50,000$), the failure-time distributions using were derived for both LRC and NRC models. Both failure-time distributions were based on the threshold τ_f with the extended lines of five degradation paths censored at $3,000$ hours. Applied to the procedure in Section 4, $\hat{F}_T(t)$ and its 90% bootstrap confidence intervals are plotted in Figure 5. The point estimates and 90% confidence intervals (in parenthesis) of p^{th} quantiles of failure-time distribution are summarized in Table 5, for $p = .05$, .1, and .5. The intervals are based on $B = 2,000$ bootstrap samples.

6 Monte Carlo Results

In this section, we use simulated VFD data to compare the analytical results of the random coefficients degradation model to more standard degradation models. Data are generated from two different nonlinear models: the random coefficients model (5.23) introduced in the last section

$$\text{Model} \quad \text{I}: \qquad y_{ij} = \frac{(\beta_1 + b_{i1}) \cdot t_{ij}^{\beta_2}}{\exp[(\beta_3 + b_{i2}) \cdot t_{ij}^{\beta_4}]} + \epsilon_{ij}$$

and an alternative nonlinear degradation model based on a mixture of three simpler degradation functions:

$$\text{Model} \quad \text{II}: \qquad y_{ij} = [\beta_1 + (\beta_2 + b_{i2})t_{ij}] \cdot I(0, 250] + [(\beta_3 + b_{i1}) + (\beta_4 + b_{i2})t_{ij}] \cdot I(250, 1000)$$

$$+[\beta_5 + \beta_6 t_{ij}] \cdot I[1000, 3000] + \epsilon_{ij},$$

where y_{ij} represents the measured (natural) logarithm of the luminosity at j^{th} measurement on the i^{th} individual and $I[\cdot]$ is an index function for time. Model II is simulated on the basis of the intercept and slope estimates obtained by fitting five degradation paths in separate (transformed) linear models in corresponding time intervals. Model II is constructed in such a way that neither the NRC nor LRC could fit the model particularly well. For model I, $\boldsymbol{\beta} = (900, 0.06, 0.04, 0.424)$, $b_{i1} \sim N(0, \sigma_1^2)$, $b_{i2} \sim N(0, \sigma_2^2)$, and $\text{Cov}(b_{i1}, b_{i2}) = 0$. It reaches a peak luminosity at 19.7 hours, on average. Model II is generated with $\boldsymbol{\beta} = (6.7, 2.0 \times 10^{-4}, 6.9, -5.0 \times 10^{-4}, 6.65, -2.0 \times 10^{-4})$, and has two random effects as Model I, with different values of σ_1^2 and σ_2^2. The error variance $\sigma_\epsilon^2 = 100$ is the same for both models.

To compare the NRC model to the LRC model, we fit (5.23) to the generated data, so that the NRC model will fit data generated from Model I by design. For the LRC model, degradation measurements earlier than $t = 240$ hours (10 days) are discarded to increase the chance that the resulting degradation process is monotonically decreasing.

Three sampling schemes are based on six, seven and nine measurement points. Sample I uses the same time points (in hours) from the VFD example: $(0, 250, 500, 800, 1000, 3000)$. Sample II includes one additional time point at 125 hours. Sample III consists of three different additional time points: $(1500, 2000, 2500)$. Table 6 lists the percentage increase in average relative bias when the (truncated) log-linear model is used with varying random coefficients. The absolute relative bias, $|y - \hat{y}|/y$, is aggregated over every simulated degradation point y in the simulation.

The efficiency gained from NRC model is significant, especially at large values of (σ_1^2, σ_2^2) in Model II. In addition, when the data were generated from Model II, the NRC model is more robust, especially in sample II, when an earlier degradation measurement is taken. The results suggest that sampling more degradation measurements at the beginning of the test has potential

to improve model precision dramatically. The simulations are obviously computationally intensive, and each table value is based on just 250 simulations.

7 Conclusion and Future Research

The nonlinear random coefficients model was derived to characterize the nonmonotonic behavior of light display degradation. Admittedly, most degradation phenomenon can be described with simpler linear regressions. Only in unusual circumstances, such as the problems with unstable light display degradation described in Section 1, are more complex models justified. In this paper we found the NRC model to adequately describe this effect of burn-in and showed the data analysis with the refined model can be significantly more efficient than a simpler model that ignores the initial measurements during the burn-in period. This improvement is apparent in simulation studies that contrast linear and nonlinear random coefficients models under different types of degradation.

Further study is needed to determine when and how frequently the units should be inspected during that time interval. With the nonmonotonic model, measurement spacing is not apparent from standard regression results; extra measurements close to time of phase change might be more valuable than previously assumed. Estimating and testing this unknown phase change time can be useful for studying the "burn-in" process, in particular.

Acknowledgements

The authors wish to thank Samsung SDI Co., LTD for providing the degradation data from the Samsung Plasma Display Device Division. We are especially thankful to Dr. W. Y. Soh, Director of Samsung SDI PDP Team. Research for both authors was supported by NSF Grant DMI – 01149003. The authors are grateful to the referees for their careful reading and helpful suggestions.

References

[1] Abramowitz, M. and Stegun, I. A. (1964), *Handbook of Mathematical Functions with Fomulas, Graphs, and Mathematical Tables*, Dover, New York.

[2] Beal, S. L. and Sheiner, L. B. (1982), "Estimating Population Kinetics ", *CRC Critical Reviews in Biomedical Engineering*, 8, 195–222.

[3] Beal, S. L. and Sheiner, L. B. (1988), "Heteroscedastic Nonlinear Regression ", *Technometrics*, 30, 327–338.

[4] Beal, S. L. and Sheiner, L. B. (1992), *NONMEM User's Guides*, NONMEM Project Group, University of California, San Francisco.

[5] Chang, D. S. (1992) "Analysis of accelerated degradation data in a two-way design ", *Reliability Engineering & System Safety*, 39, 65–69.

[6] Chiao, C. H. and Hamada, M. S. (1996), "Using Degradation Data from an Experiment to Achieve Robust Reliability for Light Emitting Diodes", *Quality and Reliability Engineering International*, 12, 89–94.

[7] Davis, P. J. and Rabinowitz, P. (1984), *Methods of Numerical Integration*, 2nd ed., Academic Press, New York.

[8] Efron, B. and Tibshirani (1993), *An Introduction to the Bootstrap* , Chapman & Hall, New York.

[9] Fagerstorm, R. (1991), "Stochastic Differential Equation Modeling of Laser Degradation", Paper presented at the Institute of Mathematical Statistics Special Topics Meeting on Statistics in Industry, Philadelphia, PA, June 1991. (Abstract in Institute of Mathematical Statistics Bulletin 20, page 159).

[10] Fukuda, M. (1991), *Reliability and Degradation of Semiconductor Lasers and LEDs*, Artech House, Boston.

[11] Golub, G. H. and Welsch, J. H. (1969), "Calculation of Gaussian Quadrature Rules ", *Mathematical Computing*, 23, 221–230.

[12] Jeger, S. L., Liang, K.Y. and Albert, P. S. (1988), "Models for Longitudinal Data: A Generalized Estimating Equation Approach ", *Biometrics*, 44, 1049–1060.

[13] Kim, S. Y. (1997), "Extended Least Squares Estimator using Monte Carlo Method in Nonlinear Random Coefficient Models.", Ph.D dissertation. Department of Statistics, North Carolina State University, unpublished.

[14] Lindstrom, M. J. and Bates, D. M. (1988), "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated Measures Data ", *Journal of the American Statistical Association*, 83, 1014–1022.

[15] Lindstrom M. J. and Bates, D. M. (1990), "Nonlinear Mixed Effects Models for Repeated Measures Data ", *Biometrics*, 46, 673–687.

[16] Lu, C. J. and Meeker, W. Q. (1993), "Using Degradation Measures to Estimate of Time-to-failure Distribution", *Technometrics*, 35, 161–176.

[17] Lu, C. J., Meeker, W. Q. and Escobar, L. A. (1996), "A Comparison of Degradation and Failure-time Methods for Estimating a Time-to-failure Distribution ", *Statistica Sinica*, 6, 531–546.

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

[19] Meeker, W. Q. and Escobar, L. A. (1993), "A Review of Recent Research and Current Issues in Accelerated Testing", *International Statistical Review*, 61, 147–168.

[20] Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, Wiley, New York.

[21] Murray, W. P. (1994), "Accelerated Service Life Prediction of Compact Disks", *Accelerated and Outdoor Durability Testing of Organic Materials*, ASTM STP 1202, W. D. Ketola and D. Grossman, editors. Philadelphia: American Society of Testing and Materials, 263–271.

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

[23] Pinheiro, J. C. and Bates, D. M. (1995), "Approximations to the Log-likelihood Function in the Nonlinear Mixed-effects Model ", *Journal of Computational and Graphical Statistics*, 4, 12–35.

[24] Pinheiro, J. C. and Bates, D. M., (2000), *Mixed-Effects Models in S and S-Plus*, Springer, New York.

[25] Ramos, R. Q. and Pantula, S. G. (1995), "Estimation of Nonlinear Random Coefficient Models ", *Statistics & Probability letters*, 24, 49–56.

[26] SAS Institute, Inc. (1999), *SAS OnlineDoc, Version 8.0*, SAS Inc., Cary, NC.

[27] Sheiner, L. B. and Beal, S. L. (1985), "Pharmacokinetic Paramete Estimates from Several Least Squares Procedures: Superiority of Extended Least Squares ", *Journal of Pharmacokinetics and Biopharmaceutics*, 13, 185–201.

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

[29] Tseng, S. T., Hamada, M. S. and Chiao, C. H. (1995), "Using Degradation Data to Improve Fluorescent Lamp Reliability ", *Journal of Quality Technology*, 27, 363–369.

[30] Tseng S. T. and Wen, J. C. (2000), "Step-stress Accelerated Degradation Analysis for Highly Reliable Products", *Journal of Quality Technology*, 32, 209–216.

[31] Vonesh, E. F. and Chinchilli, V. M. (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, Inc., New York.

[32] Wolfinger, R. (1993), "Laplace's Approximation for Nonlinear Mixed Models ", *Biometrika*, 80, 791–795.

[33] Yu, H. F. and Tseng, S. T. (1998), "On-line Procedure for Terminating an Accelerated Degradation Test", *Statistica Sinica*, 8, 207–220.

[34] Zeger, S.L., Liang, K.-Y., Paul, S. (1988), "Models for longitudinal data: A generalized estimating equation approach, ", *Biometrics*, 44, 1049–1060.

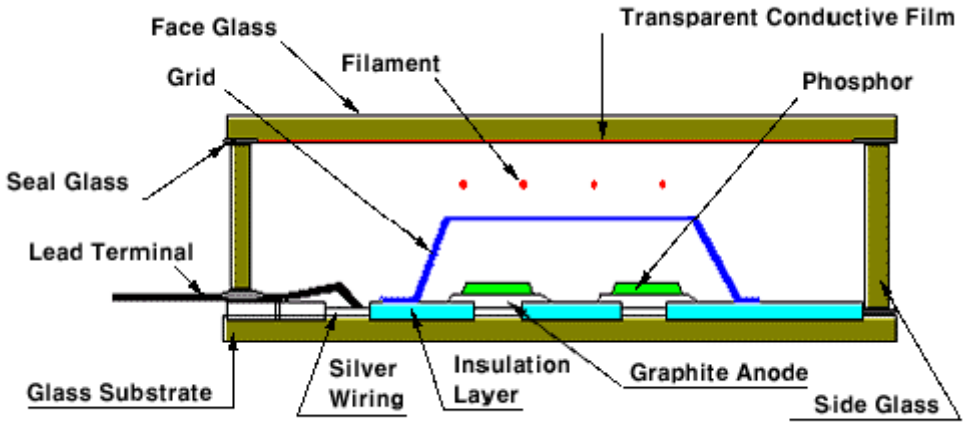

Figure 1: Basic structure of a Vacuum Fluorescent Display

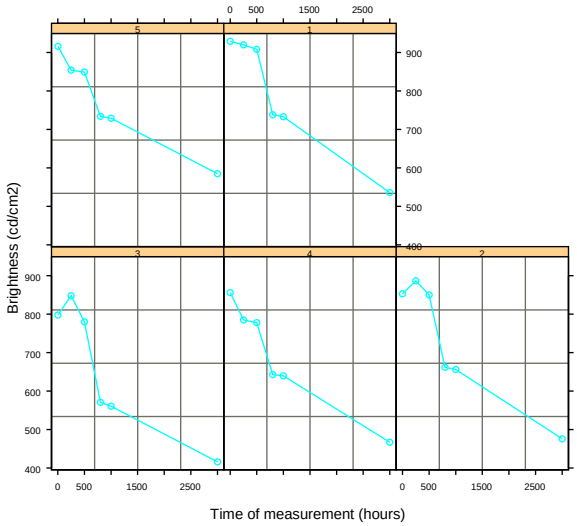

Figure 2: Nonmonotonic degradation paths for VFDs

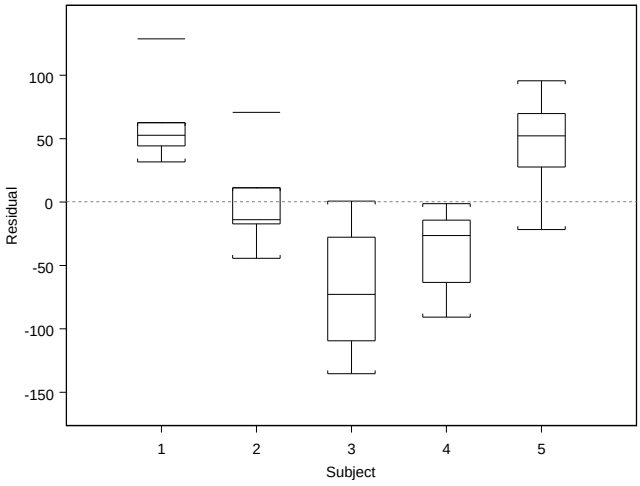

Figure 3: Boxplots of residuals from a fixed-effects degradation model

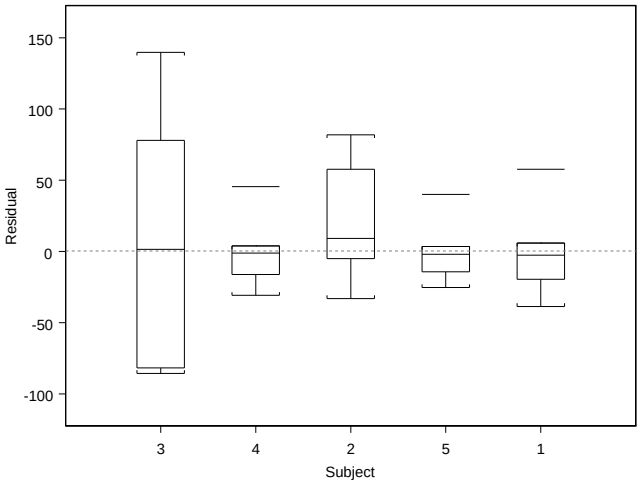

Figure 4: Boxplots of residuals from individually fitted models

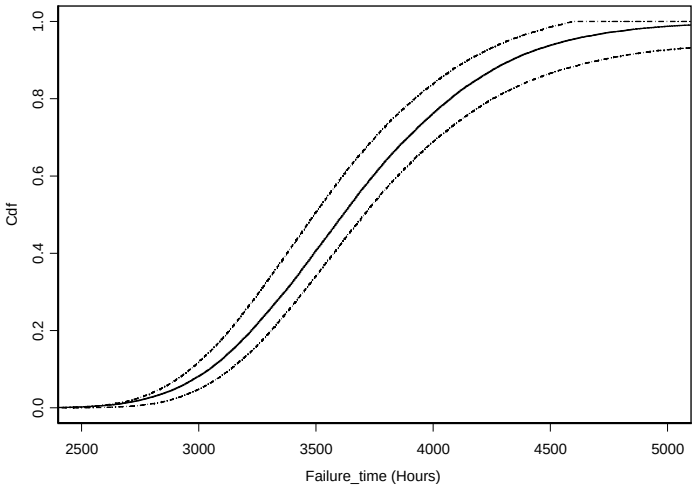

Figure 5: Estimated CDF of VFD unit lifetime along with 90% (pointwise) confidence intervals

Subject	β_1	β_2	β_3	β_4
1	1091.8	0.1203	0.1617	0.2928
2	2.9×10^7	0.8578	10.5036	0.0664
3	8.0×10^7	0.7450	9.4013	0.0636
4	887.3	0.0475	0.0362	0.4182
5	980.0	0.0542	0.0677	0.3306

Table 1: Parameter estimates for individual VFD degradation paths

Approximation method	β_1	β_2	β_3	β_4	$\sigma^2_{b_1}$	$\sigma^2_{b_2}$	σ^2_e
First-order	1825.96	0.2324	0.7376	0.1824	253.75	5×10^{-5}	1729.24
Laplace	900.02	0.0810	0.0620	0.3828	10.0056	8.411×10^{-6}	100.10
Importance sampling	900.80	0.0634	0.0409	0.4244	10.4036	1.4×10^{-5}	102.90
Adap. Gaussian	901.09	0.0660	0.0415	0.4185	10.7038	5.4×10^{-6}	104.20

Table 2: VFD parameter estimates and error variances for four approximation methods

Approximation method	$\log L$	Average relative bias	# of evaluations
First-order	-156.55	0.0823	2855
Laplace	-363.65	0.0476	427
Importance sampling	-356.35	0.0478	632
Adap. Gaussian	-364.65	0.0476	1924

Table 3: Log-likelihood, average relative bias, and computational efficiency for four approximation methods

Parameter	Estimate	Standard error
β_1	900.02	0.6759
β_2	0.081	0.0038
β_3	0.062	0.0053
β_4	0.3828	0.0073
$\sigma^2_{b_1}$	10.0056	0.02023
$\sigma^2_{b_2}$	8.411×10^{-6}	2.458×10^{-6}
σ^2_e	100.10	4.5227

Table 4: Laplace approximation parameter estimates and their standard errors

	quantile		
	.05	.1	.5
LRC model	3,000	3,125	3,675
	(2,892.5 , 3,062.5)	(3,042.5 , 3,197.5)	(3,587.5 , 3,787.5)
NRC model	2,905	3,045	3,615
	(2,842.5 , 3,007.5)	(2,967.5 , 3,137.5)	(3,497.5 , 3,707.5)

Table 5: Quantiles and their bootstrap confidence intervals for estimated failure-time distributions.

Model	σ_1^2	σ_2^2	Sample I	Sample II	Sample III
I	5	7×10^{-6}	+18%	+16%	+17%
I	10	1.4×10^{-5}	+9%	+6%	+8%
I	20	2.8×10^{-5}	-4%	+14%	+17%
II	0.01	2.5×10^{-11}	+3%	+4%	+4%
II	0.02	1.0×10^{-10}	+23%	+30%	+13%
II	0.04	4.0×10^{-10}	+46%	+62%	+19%

Table 6: Increase in absolute relative error with respect to truncated log-linear model