Methods For Planning Repeated Measures Degradation Studies

Brian P. Weaver

Statistical Sciences Group

Los Alamos National Laboratory

Los Alamos, New Mexico 87545

William Q. Meeker

Department of Statistics

Iowa State University

Ames, IA, 50010

Luis A. Escobar

Department of Experimental Statistics

Louisiana State University

Baton Rouge, LA, 70803

Joanne Wendelberger

Statistical Sciences Group

Los Alamos National Laboratory

Los Alamos, New Mexico 87545

April 16, 2012

Abstract

Repeated measures degradation studies are used to assess product or component reliability when there are few or even no failures expected during a study. Such studies are often used to assess the shelf life of materials, components, and products. We show how to evaluate the properties of proposed test plans. Such evaluations are needed to identify statistically efficient tests. We consider test plans for applications where parameters related to the degradation distribution or the related lifetime distribution are to be estimated. We use the approximate large-sample variance-covariance matrix of the parameters of a mixed effects linear regression model for repeated measures degradation data to assess the effect of sample size (number of units and number of measurements within the units) on estimation precision of both degradation and failure-time distribution quantiles. We also illustrate the complementary use of simulation-based methods for evaluating and comparing test plans. These test-planning methods are illustrated with two examples.

Keywords: Repeated Measures Planning, Aging and Degradation, Lifetime Distributions, Degradation Distributions

1 Introduction

1.1 Motivating Examples

Engineers often need to quantify the failure-time distribution of highly reliable items. Traditional life tests, where the response is time to failure, typically yield few or no failures. Instead engineers can sometimes use methods that measure the degradation of an item, providing more information than the traditional life tests. One such method is to use non-destructive repeated measurements over time on the degradation of each item. Given a degradation model and a relationship between degradation and failure, a failure-time distribution can be established. Before the test is performed, however, the engineers need to decide how many items should be measured and how often should these measurements be made in order to achieve a certain level of precision.

This work is motivated by two different applications that we have encountered. The first application involved a long-term shelf-life study on the chemical degradation of a certain compound in a particular environment. A sample of 12 items were randomly selected from a much larger population of items in storage. The engineers would then make annual measurements of the concentration of the chemical compound in units of parts per million volume (ppmv). Because of the importance of the application the available data would be analyzed and a summary report would be prepared annually. Since the data were sensitive and not available for release, Figure 1 shows data that were simulated on a modified scale to mimic the original study. The question asked by the engineers was, "Given the pattern of the observations in Figure 1 (from a previous similar study), how should the next shelf-life study be performed?"
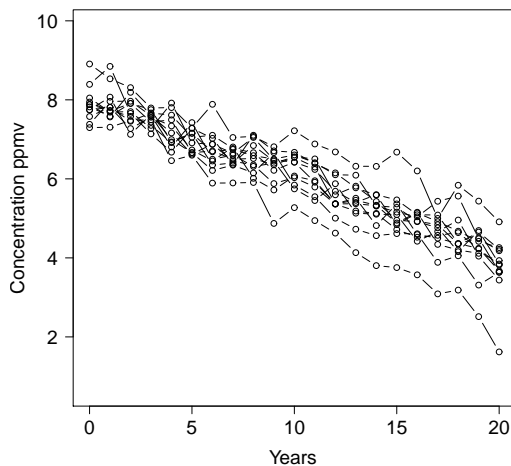

Figure 1: Simulated shelf-life degradation data for $n = 12$ units.

The second application involves a study involving inkjet printer heads. The engineers involved in this example were interested in performing a system reliability study for which the print heads were a compo-

2

nent. The engineers wanted an estimate of the failure-time distribution where failure-time depends on the degradation level of the print head. Degradation was defined to be the amount of diffusion of an ink-related substance in the printheads. As time progresses, if this substance reaches a certain location in the printhead, a failure will soon follow.

In the experiment, measurements were taken periodically on a sample of 12 units. At each inspection time, the units were measured to determine how far this substance had moved (in millimeters) after a certain amount of time. Figure 2 shows a scatterplot of the print head degradation data. Again, the data were scaled to protect proprietary information. The first point in time (time point zero) is considered the point for which the printhead had been initially loaded with ink. According to the coordinate system used, failure will occur when the degradation level reaches 60 mm.
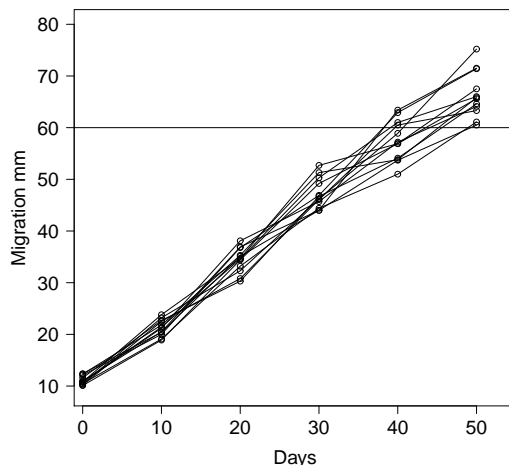

Figure 2: Scatterplot of printhead migration data. The horizontal line indicates point at which a failure is declared.

1.2 Related Work

This section reviews some of the literature on degradation test plans or related applications. Yu and Tseng (1999) discuss the optimization of degradation plans under the constraint of total experimental cost and the assumption that the lifetime distribution is lognormal. Lu and Meeker (1993) derive an analytical form of the lifetime distribution under different models and assumptions on the model parameters. Lenth (2006), on his website, has several Java programs that allow design of repeated measures experiments based on controlling power. Diggle, Heagerty, Liang, and Zeger (2002) give sample size calculations for longitudinal data models where the number of measurements per unit is specified. Boulanger and Escobar (1994) discuss experimental design for accelerated degradation tests where the amount of degradation over time levels off to a plateau. Vickers (2003) discusses how adding more measurements in a repeated measures study can affect the power of the test (i.e., the probability of detecting a difference caused by a treatment when a difference

is truly present).

1.3 Overview

The remainder of this paper is organized as follows. Section 2 describes the linear degradation model used in our work. Section 3 gives the likelihood function and the large-sample approximate variance-covariance matrix of the maximum likelihood estimator, followed by a discussion of parameter estimation. Section 4 gives the degradation distribution quantile function and shows the use of the Fisher information matrix for inference on this function. Section 5 illustrates the use of the Fisher information matrix and simulation for degradation test planning and for comparing test plans. Section 6 describes test plans that focus on estimating quantities of the failure-time distribution induced by the degradation model. Section 7 presents results from a simulation study which assesses the accuracy of the large-sample approximate standard error used in the test planning relative to the empirical standard errors obtained from simulation. Section 8 gives conclusions and describes possible areas for future related research.

2 Repeated Measures Degradation Model

2.1 Model and Data

Let y_{ij} be the observed degradation at time t_{ij} on unit i where $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. The linear degradation random effects model is

$$y_{ij} = \mathcal{D}_{ij} + \epsilon_{ij}, \tag{1}$$

where the actual degradation path is

$$\mathcal{D}_{ij} = b_{0i} + b_{1i}t_{ij}. \tag{2}$$

The intercept b_{0i} and the slope b_{1i} are modeled as random realizations from the bivariate-normal distribution $(b_0, b_1)^T \sim \text{BVN}(\boldsymbol{\beta}, \mathbf{V})$, where the elements of $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ are fixed terms representing the population's mean intercept and slope and

$$\mathbf{V} = \begin{pmatrix} \sigma_{b_0}^2 & \rho\sigma_{b_0}\sigma_{b_1} \\ \rho\sigma_{b_0}\sigma_{b_1} & \sigma_{b_1}^2 \end{pmatrix}$$

is the covariance matrix.

Collecting into $\mathbf{Y}_i = (y_{i1}, \ldots, y_{im_i})^T$ the observations from unit i, an equivalent expression for the linear degradation model in (1) is

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i^* + \boldsymbol{\epsilon}_i, \tag{3}$$

where $\mathbf{b}_i^* = (b_{0i}^*, b_{1i}^*)^T$ is modeled as $(b_0^*, b_1^*)^T \sim \text{BVN}(\mathbf{0}, \mathbf{V})$, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are matrices of explanatory variables

defined by

$$\mathbf{X}_i = \mathbf{Z}_i = \begin{pmatrix} 1 & t_{i1} \\ \vdots & \vdots \\ 1 & t_{im_i} \end{pmatrix}$$

and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$.

Assuming independence between $\boldsymbol{\epsilon}_i$ and $\mathbf{b}_i^*$ and that the components of $\boldsymbol{\epsilon}_i$ are independent and jointly normal distributed, that is $\boldsymbol{\epsilon}_i \sim \mathrm{MVN}(0, \sigma^2 \mathbf{I}_i)$ where $\mathbf{I}_i$ is a $m_i \times m_i$ identity matrix, it follows that $\mathbf{Y}_i \sim \mathrm{MVN}(\mathbf{X}_i\boldsymbol{\beta}, \Sigma_i)$ with

$$\Sigma_i = \mathrm{Var}(\mathbf{Y}_i) = \mathrm{Var}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i) = \mathbf{Z}_i\mathbf{V}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_i. \tag{4}$$

See Jenrich and Schluchter (1986) for more details. Notice that the independence assumption among the components of $\boldsymbol{\epsilon}$ implies that the error terms are not autocorrelated which is a reasonable assumption when spacing between observations is not too small.

3 Model Likelihood Function and Fisher Information Matrix

3.1 Likelihood

Suppose that $\mathbf{y}_1, \ldots, \mathbf{y}_n$ are n independent observations from $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, respectively. The log-likelihood for observational unit i is

$$\mathcal{L}_i = -\frac{1}{2}\log\left[\det(\Sigma_i)\right] - \frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\Sigma_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}). \tag{5}$$

The total log-likelihood for n units is

$$\mathcal{L} = \sum_{i=1}^n \mathcal{L}_i = -\frac{1}{2}\sum_{i=1}^n \log\left[\det(\Sigma_i)\right] - \frac{1}{2}\sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\Sigma_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}). \tag{6}$$

3.2 Variance Covariance Matrix

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\vartheta}^T)^T$ be the parameter vector where $\boldsymbol{\vartheta} = (\sigma_{b_0}, \sigma_{b_1}, \rho, \sigma)^T$. Recall that the Fisher information matrix is defined as $\mathcal{I}(\boldsymbol{\theta}) = \mathrm{E}(-\partial^2\mathcal{L}/\partial\boldsymbol{\theta}^2)$. From large sample theory, the large-sample approximate covariance matrix of the maximum likelihood (ML) estimators is

$$\mathrm{AVar}(\widehat{\boldsymbol{\theta}}) = [\mathcal{I}(\boldsymbol{\theta})]^{-1}. \tag{7}$$

$\mathrm{AVar}(\widehat{\boldsymbol{\theta}})$ can be estimated by evaluating (7) at the ML estimates $\widehat{\boldsymbol{\theta}}$. We denote this estimator by $\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\theta}})$. The derivation of the information matrix is given in the appendix.

3.3 Linear Mixed-Model Parameter Estimation

Our focus is on test planning, but it is necessary to mention how mixed effects model parameters can be estimated. Laird and Ware (1982) discuss ML and restricted ML (REML) parameter estimation for a general class of linear mixed effects models, which includes our repeated measures model, using the EM algorithm. Jenrich and Schluchter (1986) derive the derivatives and second derivatives needed in a Newton-Rhapson algorithm for ML estimation of parameters for a general class of models that includes linear mixed effects models. Lindstrom and Bates (1988) extended the work of Laird and Ware (1982) and Jenrich and Schluchter (1986) and developed efficient algorithms for computation of both ML and REML estimates for mixed-effects models.

Faraway (2006) is a useful reference for methods to estimate model parameters using the `R` software package. In particular, he focuses on using the package `lme4` with the built-in function `lmer` to perform all the analysis and find the estimates of both the fixed and random effects. For more information on the package `lme4` and its functions, see Bates, Maechler, and Bolker (2011). Another model fitting function in `R` is `lme` found in the package `nlme`. This package contains functions that also allow for interval estimation of both fixed and random effects model parameters as well as best linear unbiased predictors of the response. We use this function to calculate ML estimates of the model parameters. See Pinheiro, Bates, Debroy, Sarkar and the `R` Development Core Team (2012) for additional information on `nlme` and its underlying functions. Although both `lme` and `lmer` both produce ML and REML estimates, there are some differences between the two packages. To name a few, `lmer` is a quicker function than `lme` and can also handle crossed random effects. The function `lme` can only handle nested random effects. `lme` is a more stable function and is easier for handling heteroscedasticity than `lmer` and provides p-values for significance of effects. Finally, the `SAS` software procedure Proc Mixed (SAS Institute Inc. 2012) is widely used for the fitting and estimation of mixed effects models. Littell, Milliken, Stroup, Wolfinger, and Shabenberber (2006) give many examples of fitting mixed effects models in `SAS` as well as the theory involved.

4 Estimating Quantiles of the Degradation Distribution

4.1 The Quantile of the Degradation Distribution

From the model in Section 2.1, it follows that the degradation at time t is given by $\mathcal{D} = b_0 + b_1 t$. When $(b_0, b_1)^T$ has a bivariate normal distribution, $\mathcal{D}$ is normally distributed with $\mathrm{E}(\mathcal{D}) = \mathrm{E}(b_0 + b_1 t) = \beta_0 + \beta_1 t$ and $\mathrm{Var}(\mathcal{D}) = \mathrm{Var}(b_0 + b_1 t) = \sigma_{b_0}^2 + \sigma_{b_1}^2 t^2 + 2t\rho\sigma_{b_0}\sigma_{b_1}$. The p quantile of the degradation distribution at time t is

$$
\begin{aligned}
d_p(t) &= \mathrm{E}(\mathcal{D}) + \Phi_{\mathrm{nor}}^{-1}(p)\sqrt{\mathrm{Var}(\mathcal{D})} \\
&= \beta_0 + \beta_1 t + \Phi_{\mathrm{nor}}^{-1}(p)\sqrt{\sigma_{b_0}^2 + \sigma_{b_1}^2 t^2 + 2t\rho\sigma_{b_0}\sigma_{b_1}},
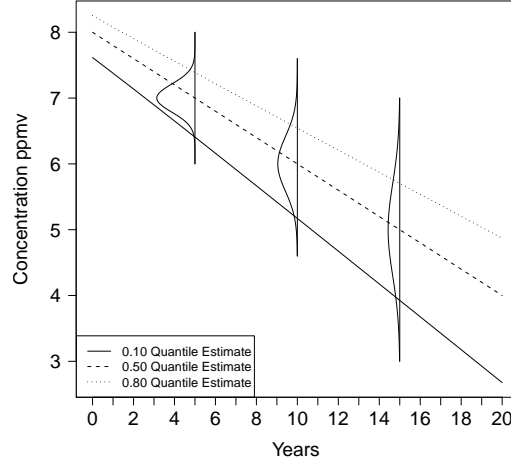\end{aligned}
\tag{8}
$$

Figure 3: Estimated 0.10 (solid line), 0.50 (dashed line), and 0.80 (dotted line) quantiles of the degradation distribution at different points in time.

where $\Phi_{\text{nor}}^{-1}(p)$ is the inverse standard normal cumulative distribution function. The ML estimate of d_p can be computed by evaluating (8) at the ML estimates $\widehat{\boldsymbol{\theta}}$.

Example 1 *Consider the simulated shelf-life data set in Figure 1. The data were simulated using (1) and the parameter values $\beta_0 = 8$, $\beta_1 = -0.2$, $\sigma_{b_0} = 0.3$, $\sigma_{b_1} = 0.04$, $\rho = 0.7$, and $\sigma = 0.3$ for $n = 12$ items and for a length of time of 20 years. The R function* `lme` *provides the ML estimates of these parameters as $\widehat{\beta}_0 = 7.98$, $\widehat{\beta}_1 = -0.19$, $\widehat{\sigma}_{b_0} = 0.35$, $\widehat{\sigma}_{b_1} = 0.05$, $\widehat{\rho} = 0.9$, and $\widehat{\sigma} = 0.28$. For given values of p and t, the ML estimate of the degradation quantile is (8) evaluated at the ML estimates of $\boldsymbol{\theta}$. This is illustrated in Figure 3 for $p = 0.10$, 0.50, and 0.80 and at different points in time.*

4.2 Standard Error for the Maximum Likelihood Estimator of the p Quantile

This section deals with the estimation of the standard error of the ML estimator of the degradation quantile d_p in (8). This quantile is a function of the parameters $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma_{b_0}, \sigma_{b_1}, \rho, \sigma)^T$. Using the invariance property of ML estimators, the ML estimator $\widehat{d}_p$ of d_p is obtained by evaluating (8) at $\widehat{\boldsymbol{\theta}}$. The formula for the approximate standard error of $\widehat{d}_p$ was derived using the delta method. Let $\mathbf{c}$ be a vector with elements $c_i = \partial d_p / \partial \theta_i$, $i = 1, \ldots, 6$. Then by the delta method, the large-sample approximate variance of $\widehat{d}_p$ is

$$\text{AVar}(\widehat{d}_p) = \mathbf{c}^T \text{AVar}(\widehat{\boldsymbol{\theta}})\mathbf{c}. \tag{9}$$

The standard error of $\widehat{d}_p$ is $\text{ASE}_{\widehat{d}_p} = \sqrt{\text{AVar}(\widehat{d}_p)}$ which is estimated by evaluating (9) at $\widehat{\boldsymbol{\theta}}$ giving $\widehat{\text{SE}}_{\widehat{d}_p} = \sqrt{\widehat{\text{Var}}(\widehat{d}_p)}$. The explicit forms of the partial derivatives are given in the appendix.
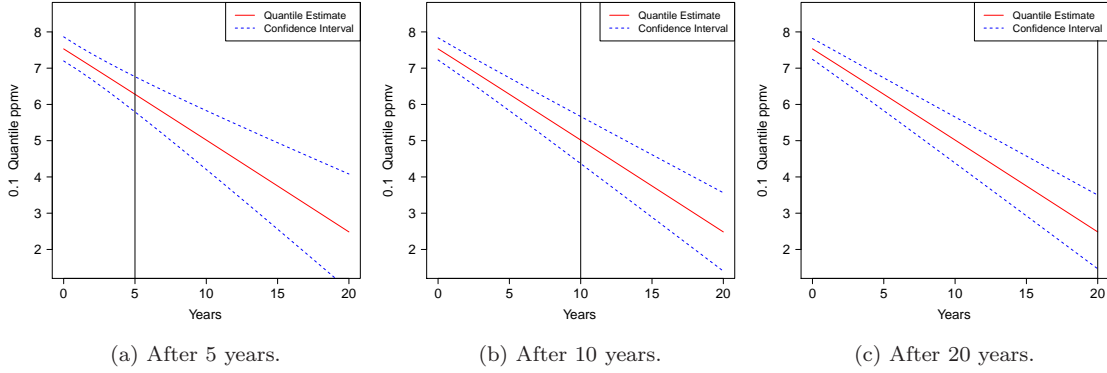
7

Figure 4: ML estimates of $d_{0.10}$ as a function of years in service, along with approximate 95% confidence intervals. The vertical lines indicate the point where the data analysis was done after 5 years (a), after 10 years (b), and after 20 years (c).

4.3 Confidence Interval for the Degradation Distribution Quantiles

A large-sample approximate $100\%(1-\alpha)$ confidence interval for d_p is

$$[\underset{\sim}{d_p}, \quad \widetilde{d_p}] = \widehat{d}_p \pm z_{(1-\alpha/2)}\widehat{\mathrm{SE}}_{\widehat{d}_p}, \tag{10}$$

where $z_{(1-\alpha/2)}$ is the $1-\alpha/2$ standard normal quantile.

Example 2 *Returning to Example 1, Figures 4a, 4b, and 4c show 95% confidence intervals when extrapolations were made to estimate degradation based on data available at 5 years, 10 years, and 20 years respectively. As expected, the width of the confidence intervals decreases as more information becomes available.*

5 Degradation Test Planning

This section describes planning methods for repeated measures degradation tests. The test plans and test plan properties described in this section and in Section 6 depend, however, on the true model and its parameters. In order to illustrate or describe the results of a proposed test plan, one must have "planning information" for the model parameters. Ideally, this information would come from design specification, expert opinion, or previous experience. Because this planning information does not correspond to the true values of the model parameters, it is recommended to perform a sensitivity evaluation over a range of the unknown values. Additionally, the results of such an evaluation could be used to help select a more robust or conservative (i.e., a plan that will meet experimental goals with high probability) test plan. A more formal way to do this would be to assume a prior distribution for the unknown parameter values and take a Bayesian approach, as has been done for accelerated life tests (e.g., Chaloner and Larntz 1992). In this paper, the superscript □ (an open box) on a parameter is used to denote planning information for the unknown model

parameters (i.e., $\sigma^{\square}$ is a planning value for the unknown parameter σ).

Additionally, when the degradation paths deviate from linearity, one should first consider transformations of the response and/or time that could lead to an approximately linear degradation path. In some applications, certain transformations are suggested by previous experience with the same degradation mechanism or physical/chemical knowledge of the failure mechanism that would make the degradation linear in time. The test planning would then be based on the transformed relationship. See, for example, Box and Cox (1964) for very general family of transformations. Some degradation processes are inherently nonlinear and cannot be transformed to linearity (e.g., a degradation process described by first order kinetics, leading to a degradation path with an asymptote). In such cases, the general ideas presented in this paper could still be used. The simulation approach would be straightforward to implement using `nlme` in R (see Section 22.5 of Meeker and Escobar 1998 for an example of this). Boulanger and Escobar 1994 developed methods for designing accelerated degradation tests for degradation processes that approach an asymptote. Certainly large-sample approximate variances could also be derived, but this would have to be done on a case-by-case basis, depending on the particular degradation path model.

The rest of Section 5 is organized as follows: Section 5.1 shows a simple, graphical approach for test planning that assumes all units have the same inspection schedule. Section 5.2 describes an approach that allows for different schedules for different units. Section 5.3 illustrates a simulation-based approach that complements the analytical evaluations and can also be applied to the more general settings. Finally Section 5.4 describes an approach for minimizing the cost of a test, subject to a constraint on estimation precision.

5.1 Simple Degradation Test Plans

In a simple degradation test plan all units are measured using the same schedule. Evaluation of statistical test-plan properties help to determine the number of units to measure in the study and how many measurements should be made over time. We use the large-sample approximate standard error $\mathrm{ASE}_{\widehat{d_p}}$ to quantify and compare the precision provided by alternative test plans. In particular, we obtain a contour plot of the $\mathrm{ASE}_{\widehat{d_p}}$ values obtained over a grid of $n = 3, 4, \ldots, 10$ experimental units and $m = 3, 4, \ldots, 10$ measurements per unit. Test-plan decisions and recommendations are based on the actual values of $\mathrm{ASE}_{\widehat{d_p}}$ calculated over the grid and the corresponding contour plot. In the following two examples, we use rather extreme levels for the measurement error variability planning values to illustrate the strong effect that this parameter can have on degradation test plans.

Example 3 *Suppose that the objective is to assess the trade-off, in terms of variance, between the number of measurements per unit and the number of units being used in the study. The shelf-life study is expected to run for 20 years. The parameter values from Example 1 are used as the planning information and they are denoted by $\beta_0^{\square} = 8$, $\beta_1^{\square} = -0.2$, $\sigma_{b_0}^{\square} = 0.3$, $\sigma_{b_1}^{\square} = 0.04$, $\rho^{\square} = 0.7$, and $\sigma^{\square} = 0.3$. Figure 5a shows the results for some simple test plans. This plot shows that for the proposed planning values, the smallest standard error that could be obtained is less than 0.46 (corresponding to $n = 10$ and $m = 10$). The plot*

suggests that a trade-off could be made by choosing a small number of units, say 6, and measuring them 7 times without losing much in terms of precision (in this region $\mathrm{ASE}_{\widehat{d}_p} \approx 0.59$). *Because the measurement error is relatively small in this example, increasing the number of measurements over time will not have a large effect on estimation precision.*

In the next example, we use a much larger planning value for the measurement error variability to illustrate its effect on estimation precision.

Example 4 *Now suppose that the planning information value for* σ *is increased to* $\sigma^{\square} = 3$. *Figure 5b shows the large-sample approximate standard error* $\mathrm{ASE}_{\widehat{d}_{0.10}}$ *for different combinations of n and m with the new planning information. In this case, the plot shows that a test should be chosen from the North–East region where* $\mathrm{ASE}_{\widehat{d}_{0.10}}$ *is less than 1.5. In the South–West corner of the plot, however,* $\mathrm{ASE}_{\widehat{d}_{0.10}}$ *reaches values larger than 5.5 . In summary, to compensate for the large variability in measurements (i.e., large* σ), *the test plan requires more units and more measurements per unit to achieve a smaller standard error when compared to* $\sigma^{\square} = 0.3$ *in Example 3.*
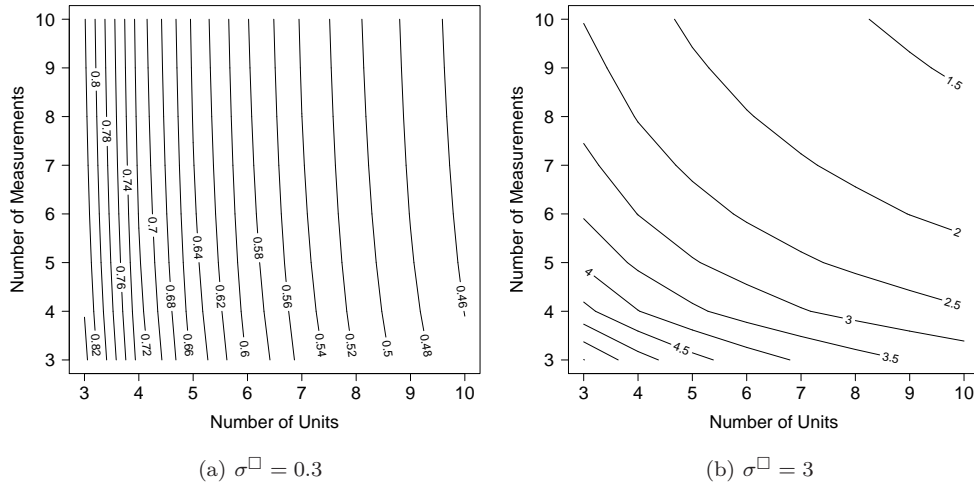

(a) $\sigma^{\square} = 0.3$ (b) $\sigma^{\square} = 3$

Figure 5: Contour plot of the large-sample approximate standard error $\mathrm{ASE}_{\widehat{d}_{0.10}}$ as a function of n and m. For 5a the measurement error variability $\sigma^{\square}$ is small and for 5b the measurement error variability is large.

5.2 Degradation Test Plans with Differing Schedules

The use of different inspection schedules for groups of units is motivated by two concerns of test planners:

- Inspections can be expensive and there can be substantial savings if some units are sampled less frequently.
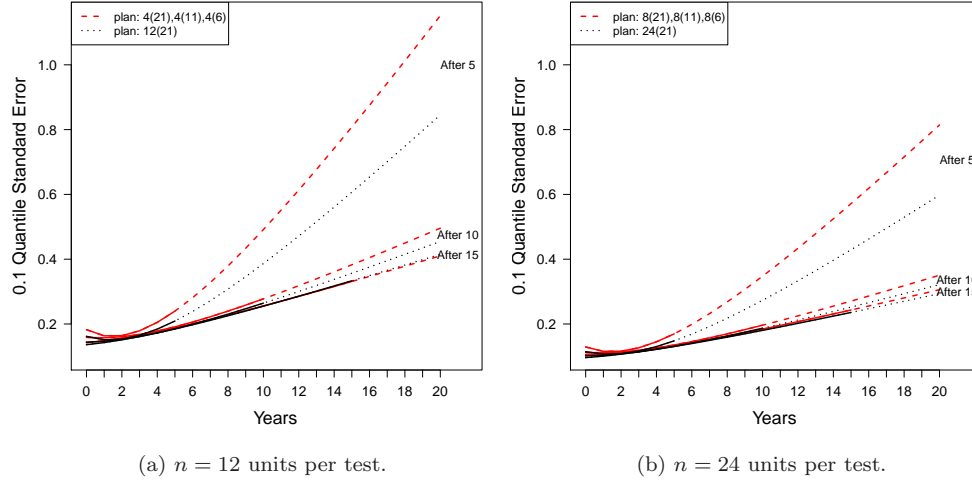
(a) $n = 12$ units per test.

(b) $n = 24$ units per test.

Figure 6: Comparison of large-sample approximate standard errors of $\widehat{d}_{0.10}$ for 2 different test plans with extrapolation out to 20 years. The point where the line changes from solid to dashed or dotted is the time at which extrapolation begins.

- There was concern that the measurements could have an effect on the degradation process. Having groups of units on different inspection schedules can provide information to detect and model the effect of such changes, if they exist.

Example 5 *Again consider a shelf-life study that is to be performed over a period of 20 years. Periodic evaluation of a sample of units is scheduled for 5, 10, and 20 years. Some questions of interest are "What is the current state of the units in the larger population of units from which the sample was taken?" and "Can the future state of the population of units be predicted?" Two different plans will be compared for this study, 4(21),4(11),4(6) and 12(21). This notation means that this plan will use 12 units of which 4 units are measured 21 times (i.e., every year starting at time 0), 4 are measured 11 times (i.e., every other year), and 4 units are measured 6 times (i.e., every 4 years). In the alternative plan, all 12 units will be measured every year. Notice that the first of these plans will involve 152 measurements and the second will have 252 measurements. Thus if the first plan gives adequate information, it would be preferred because it costs much less than the second plan. Figure 6a shows a comparison of these two plans. There is a large difference in the standard errors for plans 12(21) and 4(21),4(11),4(6) when extrapolation to 20 years is performed after 5 years of observations. After 10 years of observations, however, there is little difference between the two plans. These results suggest that the sampling plan 4(21),4(11),4(6), especially after 10 years of inspection could result in large savings in both time and money with little loss of precision.*

If the number of units tested could be doubled from 12 to 24, estimation precision would be improved. Figure 6b shows the results from the two plans 24(21) and 8(21),8(11),8(6). It is easy to show that standard errors for the 24(21) test plan are the same as those from the 12(21) plan, divided by $\sqrt{2}$. This is only

approximately so for the 8(21),8(11),8(6) plan with respect to the 4(21), 4(11), 4(6) plan.

5.3   Using Simulation to Evaluate Test Plans

This section describes a complementary simulation-based method for comparing test plans. After methods based on large-sample approximate variance are used to find a candidate test plan, we generally recommend the use of simulation to study the plan. Simulation provides visualization of sampling variability and insight into the test planning process. Simulation results, presented graphically, are particularly useful when communicating with engineers. In general, simulation methods for evaluating and comparing test plans are also useful in situations where the delta method might not provide a good approximation, when it is difficult to derive an analytical method, or when there is not enough time to derive an analytical method. The simulation algorithm for repeated measures degradation testing is as follows:

1. With a given test plan and planning values, simulate data vectors $\mathbf{Y}_i^*$ from the model in (3) where $i = 1, \ldots, B$ and B being a large number, say 10,000.

2. For each simulated data set $\mathbf{Y}_i^*$, calculate the ML estimates of $\boldsymbol{\theta}_i^*$, say $\widehat{\boldsymbol{\theta}}_i^*$.

3. Calculate functions of $\widehat{\boldsymbol{\theta}}_i^*$ that are of interest, say $g(\widehat{\boldsymbol{\theta}}_i^*)$ (e.g., $\widehat{d}_{p,i}^*$).

4. Plot the estimates to illustrate the trial to trial variability.

5. Estimate the standard error of the components of $\widehat{\boldsymbol{\theta}}_i^*$ or $g(\widehat{\boldsymbol{\theta}}_i^*)$ by calculating the sample standard deviation of the simulated estimates.

Figure 7 shows an example of Step 3 for $\widehat{d}_{0.10}^*$ in the above algorithm for the planning information given in Section 5.1. Figure 8 shows a comparison between the simulation method and the large sample method for the plan 12(21) when considering again the standard error for the 0.10 quantile of the degradation distribution. Notice that the two approaches agree with each other and are similar in shape and numerical values. The simulation approach, however, indicates smaller standard errors than the large-sample approximation value for the 0.10 quantile for the case when extrapolation began after 5 years. This can also be seen in the simulation results presented in Figures 13b and 14a in Section 7.

5.4   Selecting a Test Plan Under a Cost Constraint

This section describes the selection of a degradation test plan when there is a constraint on $\mathrm{ASE}_{\widehat{d}_p}$ and there is a desire to minimize the cost of running the experiment. Suppose the cost of the experiment is

$$\mathrm{cost}(n, m) = c_1 + c_2 n + c_3 nm,$$

where c_1 denotes the fixed cost of running the experiment, c_2 is the cost of testing an experimental unit, and c_3 is the cost of a measurement on an item. Although the approach is general we will use the exact same
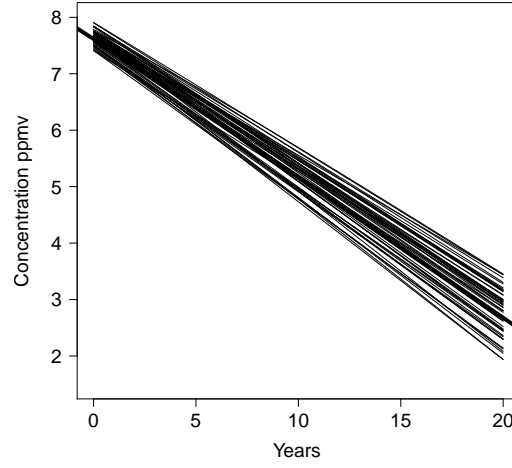
Figure 7: Simulations of ML estimates of the 0.10 quantile of the degradation level using the planning values in Example 3. The longer thick line represents the 0.10 quantile under the planning information.
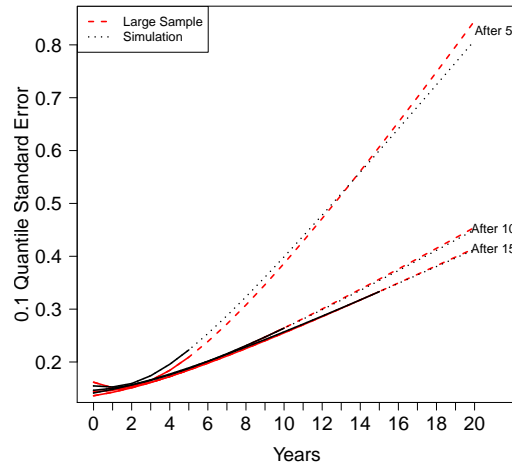

Figure 8: large-sample approximate standard errors of the estimator of the 0.1 quantile of the degradation level for the 12(21) test plan using simulation (solid lines) and large sample approach (dashed lines) for comparing different test plans.
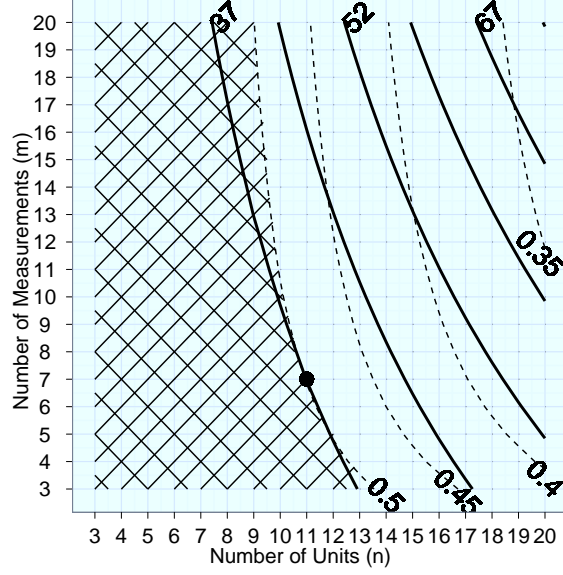
Figure 9: Contour plot of cost (thick solid lines) and $\text{ASE}_{\widehat{d}_{0.10}}$ (dashed lines) for Example 6. The large dot indicates the constrained optimum test plan. The cross-hatched region corresponds to the pairs (n, m) that do not satisfy the constraint $\text{ASE}_{\widehat{d}_{0.10}} \leq 0.50$. The labels for the cost contours have been multiplied by 10^{-3} for readability.

measurement schedule for each unit and equally spaced inspections. Let γ denote the maximum acceptable value of $\text{ASE}_{\widehat{d}_p}$. Then we wish to find the values of n and m, say n^* and m^* such that $\text{ASE}_{\widehat{d}_p} \leq \gamma$ and

$$\text{cost}(n^*, m^*) = \min_{n,m}[\text{cost}(n, m)].$$

Example 6 *Consider a shelf-life study that is to be performed for 20 years with the planning information given in Example 3. The information after 10 years of observations will, however, be used to make important predictions at 20 years. This study has a limited budget and a test plan is to be chosen so that the cost of the study is to be minimized subject to the constraint $\text{ASE}_{\widehat{d}_{0.10}} \leq 0.50$. The individual cost components of the study are $c_1 = \$15,000$, $c_2 = \$1,500$, and $c_3 = \$75$. Figure 9 shows the results of this optimization. In the cross-hatched region $\text{ASE}_{\widehat{d}_{0.10}} > 0.50$. The large dot on the plot corresponds to the constrained optimum test plan. The plot indicates that $n = 11$ items should be measured at $m = 7$ equally-spaced times. For this test plan, $\text{ASE}_{\widehat{d}_{0.10}} = 0.498$ and $\text{cost}(n, m) = \$37,275$.*

6   Failure-Time Distribution

This section derives the failure-time distribution implied by the linear degradation model in (2) and a specification of the degradation level $\mathcal{D}_{\text{f}}$ for failure. Chapter 13 of Meeker and Escobar (1998) provides a more general discussion of failure-time distributions that are implied by a degradation model.

6.1 Relationship Between Degradation and Failures

We assume a degradation process with soft failures. That is, the failure time for a unit is defined to be the time at which the degradation level reaches the specified degradation level $\mathcal{D}_{\mathrm{f}}$. Let T define the random variable associated with the unit's time to failure.

For a fixed t, $b_0 + b_1 t \sim \mathrm{NOR}(\beta_0 + t\beta_1, \sigma_{b_0}^2 + t^2\sigma_{b_1}^2 + 2t\rho\sigma_{b_0}\sigma_{b_1})$. First, consider the case of increasing degradation. In this case

$$
\begin{aligned}
\Pr(T \leq t) = F(t; \boldsymbol{\theta}) &= \Pr(b_0 + b_1 t \geq \mathcal{D}_{\mathrm{f}}) \\
&= 1 - \Pr(b_0 + b_1 t \leq \mathcal{D}_{\mathrm{f}}) = 1 - \Phi_{\mathrm{nor}}\left[\kappa(\boldsymbol{\theta})\right],
\end{aligned} \tag{11}
$$

where $\kappa(\boldsymbol{\theta}) = (\mathcal{D}_{\mathrm{f}} - \beta_0 - t\beta_1)/\sqrt{\sigma_{b_0}^2 + t^2\sigma_{b_1}^2 + 2t\rho\sigma_{b_0}\sigma_{b_1}}$ and $\Phi_{\mathrm{nor}}(z)$ is the cdf for a standard normal distribution evaluated at z. Similarly, if failure occurs when the degradation level decreases to $\mathcal{D}_{\mathrm{f}}$, then

$$
\Pr(T \leq t) = F(t; \boldsymbol{\theta}) = \Pr(D = b_0 + b_1 t \leq \mathcal{D}_{\mathrm{f}}) = \Phi_{\mathrm{nor}}\left[\kappa(\boldsymbol{\theta})\right]. \tag{12}
$$

When $\rho = 0$, $F(t; \boldsymbol{\theta})$ is known as the Bernstein distribution (e.g., Gertsbakh and Kordonskiy 1969, Ahmad and Sheikh 1984, and Lu and Meeker 1993). The ML estimator of the failure-time distribution is $F(t; \widehat{\boldsymbol{\theta}})$ where $\widehat{\boldsymbol{\theta}}$ is the ML estimator of $\boldsymbol{\theta}$. Meeker and Escobar (1998), page 330, describes a numerical integration and a simulation based approach to evaluate the failure-time distribution for more complicated models where a closed form solution for the cdf $F(t; \boldsymbol{\theta})$ does not exist.

Example 7 *The printhead of an inkjet cartridge is a component in a larger series system for a printer. Estimation of its lifetime distribution was needed to estimate the lifetime distribution for the entire system. As described in Section 1.1, the failure mechanism was diffusion of a failure-causing substance. The engineers defined a degradation level of $\mathcal{D}_{\mathrm{f}} = 60$ mm to be a failure. This degradation level is represented by the horizontal line in Figure 2. The ML estimates of the model parameters for the degradation model are*

$$
\widehat{\beta}_0 = 11.22, \ \widehat{\beta}_1 = 1.14, \ \widehat{\sigma}_{b_0} = 0.45, \ \widehat{\sigma}_{b_1} = 0.07, \ \widehat{\rho} = -0.82, \ \text{and} \ \widehat{\sigma} = 2.6. \tag{13}
$$

Figure 10 gives the ML estimate of the failure-time distribution for the print head degradation data.

6.2 Estimating the p Quantile of the Failure-Time Distribution

From inverting (11) the p quantile of the failure-time distribution is

$$
t_p = \frac{-(k\sigma_{b_0 b_1} + h\beta_1) \pm \sqrt{k^2\sigma_{b_0 b_1}^2 + k\sigma_{b_0}^2\beta_1^2 + h^2 k\sigma_{b_1}^2 - k^2\sigma_{b_0}^2\sigma_{b_1}^2 + 2hk\beta_1\sigma_{b_0 b_1}}}{k\sigma_{b_1}^2 - \beta_1^2} \tag{14}
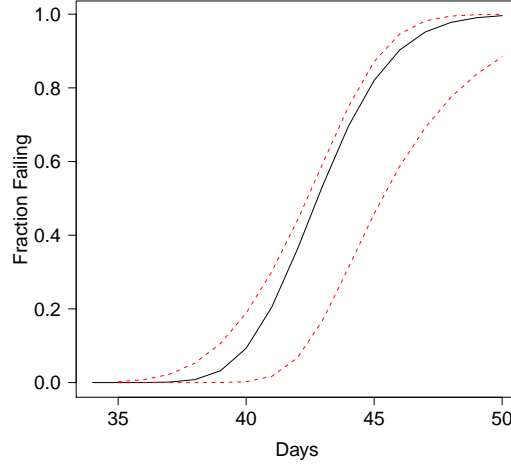$$

Figure 10: Estimated failure-time distribution based on the printhead data. The dashed lines represent pointwise approximate 95 % confidence intervals.

where $\sigma_{b_0 b_1} = \rho \sigma_{b_0} \sigma_{b_1}$ is the covariance between b_0 and b_1, $h = \mathcal{D}_{\mathrm{f}} - \beta_0$ and $k = \left[\Phi_{\mathrm{nor}}^{-1} (1 - p) \right]^2$ or $k = \left[\Phi_{\mathrm{nor}}^{-1} (p) \right]^2$ depending on whether a failure is declared when $\mathcal{D} \geq \mathcal{D}_{\mathrm{f}}$ or $\mathcal{D} \leq \mathcal{D}_{\mathrm{f}}$, respectively. The derivation of (14) is given in the appendix. If $0 < p < 0.5$, t_p is the root where the radical is added. If $0.5 < p < 1$, t_p is given by the root where the radical is subtracted. An estimate of t_p can be computed by evaluating (14) at the ML estimates $\widehat{\boldsymbol{\theta}}$.

6.3 Standard Error for the Maximum Likelihood Estimator of the Failure-Time Quantile

Let $\mathbf{c}$ be the gradient vector with elements $c_i = \partial t_p / \partial \theta_i$, $i = 1, \ldots, 6$. Using the delta method, the large-sample approximate variance of $\widehat{t}_p$ is

$$\mathrm{AVar}(\widehat{t}_p) = \mathbf{c}^T \mathrm{AVar}(\widehat{\boldsymbol{\theta}}) \mathbf{c}. \tag{15}$$

The large-sample approximate standard error of $\widehat{t}_p$ is $\mathrm{ASE}_{\widehat{t}_p} = \sqrt{\mathrm{AVar}(\widehat{t}_p)}$ which is estimated by evaluating (15) at $\widehat{\boldsymbol{\theta}}$ giving $\widehat{\mathrm{SE}}_{\widehat{t}_p} = \sqrt{\widehat{\mathrm{Var}}(\widehat{t}_p)}$. The explicit forms of the partial derivatives in $\mathbf{c}$ are given in the appendix.

6.4 Degradation Test Planning Using t_p

This section applies the test planning techniques described in Sections 5.1 and 5.2 to t_p. This work is motivated by the inkjet cartridge example. The engineers were interested in estimating $t_{0.10}$, the time at which 10% of the items in the population would fail. They were interested in performing other degradation tests in the future on similar parts and wanted to know how many items should be measured and how many measurements should be made on each item. The ML estimates obtained in Example 7 will be used as the planning information (i.e., $\beta_0^\square = 11.22$, $\beta_1^\square = 1.14$, $\sigma_{b_0}^\square = 0.45$, $\sigma_{b_1}^\square = 0.07$, $\rho^\square = -0.82$, and $\sigma^\square = 2.6$).

First, we consider the simple test plans described in Section 5.1 where each unit is measured the same number of times. Figure 11 shows several simple degradation test plans using the planning information given above. As expected, the best plan is the 12(21) on the North–East corner for which $\text{ASE}_{\widehat{t}_{0.10}} \approx 0.83$.

A drawback to the plan 12(21) is that it might be expensive or time consuming to complete because it requires 252 measurements. Thus, it is of interest to entertain other degradation test plans that involve different measurement sequences on the units. For example, Table 1 shows the $\text{ASE}_{\widehat{t}_{0.10}}$ for the degradation plans 12(21) and 4(21),4(11),4(6), and 3(21), 3(11), 3(6), 3(3) respectively. Notice that there is not a large difference in the estimation precision for the three different plans. One could achieve savings for both time and money at the sacrifice of only a small amount of estimation precision if the plan 3(21), 3(11), 3(6), 3(3) (with 123 measurements) is selected over the plan 12(21).

Plan	Number of Measurements	Standard Error
12(21)	252	0.83
4(21),4(11),4(6)	152	0.88
3(21),3(11),3(6),3(3)	123	0.92

Table 1: large-sample approximate standard error $\text{ASE}_{\widehat{t}_{0.10}}$ for three different degradation test plans.
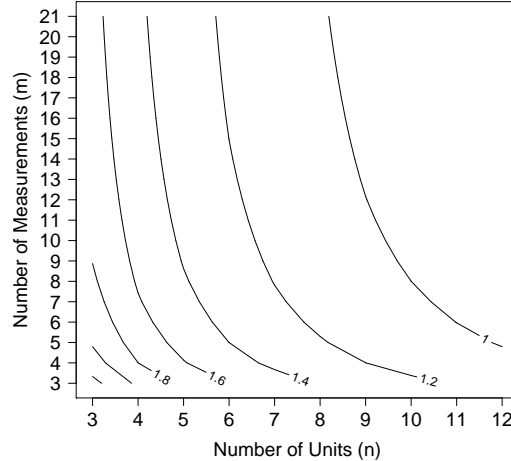

Figure 11: Contour plot of the large-sample approximate standard error $\text{ASE}_{\widehat{t}_{0.10}}$ as a function of n and m.

As in Section 5.4, test planning to minimize cost under the constraint $\text{ASE}_{\widehat{t}_p} \leq \alpha$ can be applied to the failure-time distribution quantile. Consider again the cost structure from Example 6 but now with the constraint $\text{ASE}_{\widehat{t}_{0.1}} \leq 0.80$ for a test that is going to run for 50 hours. Figure 12 shows the constrained optimum plan to be $n = 16$ and $m = 8$. The cost associated with this test plan is \$48,600 with $\text{ASE}_{\widehat{t}_{0.1}} = 0.791$.
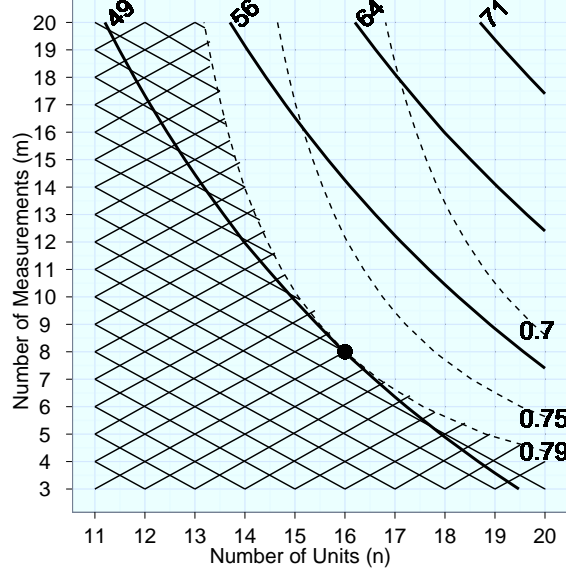
Figure 12: Contour plot of cost (thick solid lines) and $\text{ASE}_{\widehat{t}_{0.10}}$ (dashed lines) . The large dot indicates the constrained optimum test plan. The cross-hatched region corresponds to the pairs $(n,\ m)$ that do not satisfy the constraint $\text{ASE}_{\widehat{t}_{0.1}} \leq 0.80$. The labels for the cost contours have been multiplied by 10^{-3} for readability.

7   Accuracy of Approximate Standard Errors

We conducted a simulation study to assess the accuracy of the large-sample approximate standard errors (ASE) relative to empirical standard errors (ESE) obtained from Monte Carlo simulation. We focus on results for the ASE of the 0.10 degradation distribution quantile. The experiment used the four factors that have the most influence on the ASE for a given fixed time schedule. These factors are σ_{b_0}, σ_{b_1}, the length of the test (TL), and the point in time where extrapolation begins (EP). Additionally, the model parameters were scaled so that they are unit-less (i.e., are free of a unit of measurement). In particular, the model parameters β_0, β_1, σ_{b_0}, σ_{b_1}, and σ were divided by σ (removing the degradation units) and β_1 and σ_{b_1} were both multiplied by the time corresponding to the end of the test (eliminating the time units). Additionally, the time vector was divided by the time corresponding to the end of the test so that all time values are in between zero and one.

We used a 2^4 factorial design with the factors described in the previous paragraph. For each factor-level combination of the 2^4 factorial design and for a combination of n (the number of units) and m (the number of measurements), data was simulated from (1). For each simulated data set, the ML estimate of $d_{0.10}$ was calculated. This procedure was repeated 300,000 times providing a distribution of $\widehat{d}_{0.10}$. The Monte Carlo standard error was then calculated by taking the sample standard deviation of the simulated values of $\widehat{d}_{0.10}$. The large number of Monte Carlo trials was needed in order to reduce the Monte Carlo error to be less than some specified constant at the factor-level with the larger values of σ_{b_0}, σ_{b_1}, and TL.

Figures 13 and 14 provide a summary of a subset of the factor level combinations that give a general

18

picture of the relationship between the ASE and the corresponding ESE as a function of the number of experimental units (n) and the number of equally-spaced measurements within each experimental unit (m). In each plot there are four horizontal lines, of different line types and symbols, which represent the different equally-spaced measurement schedules. These levels of are $m = 10$, 25, 50, and 200. The horizontal line at the top of each plot corresponds to $m = 10$ and the horizontal line at the bottom of each plots corresponds to $m = 200$.

Figure 13a shows that for small m (the solid lines), there is a large difference between the ESE and the ASE. For small m, the ESE curve generally has an asymptote that is well below the ASE. For large m ($m = 200$), represented by the horizontal line at the bottom of the plot, the ESE and ASE are close to each other, indicating that n does not have to be too large for good agreement of the ASE when m is reasonably large.

Figures 13a and 13b combined illustrate the effect that changing TL and EP have on the ESE and ASE. The levels of σ_{b_0} and σ_{b_1} are the same in Figures 13a and 13b whereas the levels of TL and EP are different (Figure 13b has a larger levels of TL and a larger amount of extrapolation). We see that for smaller values of n and m the standard errors (both ESE and ASE) are smaller for Figure 13b relative to Figure 13a, corresponding to the larger TL and more extrapolation, whereas for larger values of n and m, the standard errors are smaller in Figure 13a. The reason for this is that for the smaller levels of n and m, the sources of error that are contributing to the ASE and ESE is the sampling error from the population, sampling error for the model parameter estimators, and the error amplification associated with the extrapolations. For smaller levels of n and m, the measurements made at larger time values for the items in Figure 13b are helping to decrease the error (relative to Figure 13a) associated with the model parameter estimation. As both n and m increase, the total error contributed from the model parameter estimation decreases to where the contribution of error amplified by the extrapolation plays a more dominant role. Comparing the two figures, Figure 13a has less extrapolation, given smaller values of the ASA and ESE for the larger levels of n and m.

Figure 14a illustrates the case where the variance components are at their lowest levels and both TL and EP are at their largest levels. Notice that the standard error values are at their lowest values among the four plots. We observe good agreement between the ESE and ASE in this figure.

The purpose of Figure 14b is to show the effects of large values of σ_{b_1} (the unit-to-unit variability in slopes). To emphasize the effects of σ_{b_1}, the levels of the other factors are all set at their least influential values, i.e., small σ_{b_0} and large TL and EP (which is in contrast to Figure 14a where σ_{b_1} is at its lowest level). Notice that the standard error values have increased compared to Figure 14a . There still seems to be good agreement, however, across values of m as the different paths are still similar.

Based on this simulation study, for situations where the ASE performs poorly, it is recommended to use simulation methods for test planning. In particular, if large extrapolations are to be made or if the number of measurements per item is small, then simulation methods would be preferred.
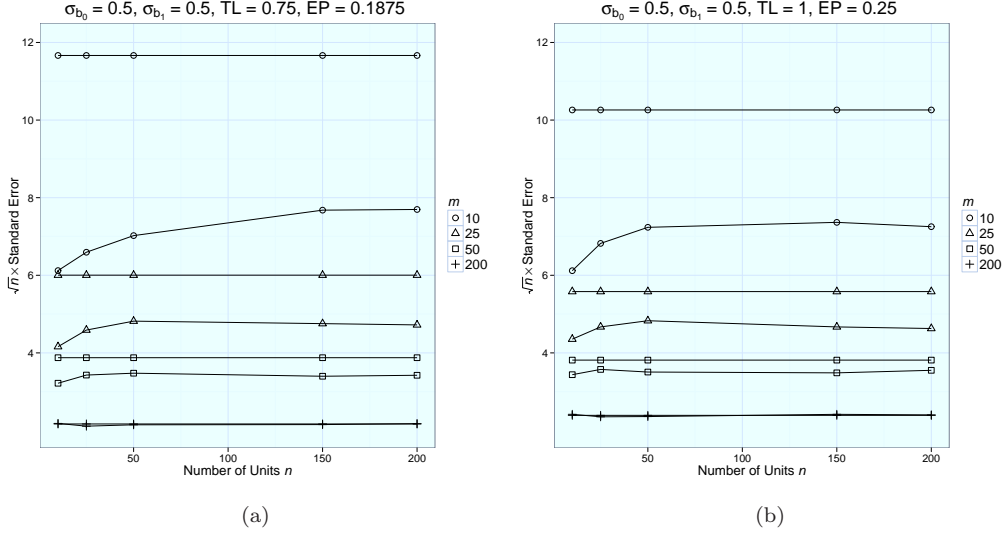
Figure 13: Factor-level combinations to assess the effects of extrapolation and test length. The horizontal lines corresponds to the ASE and the different symbols represent different numbers of measurement.
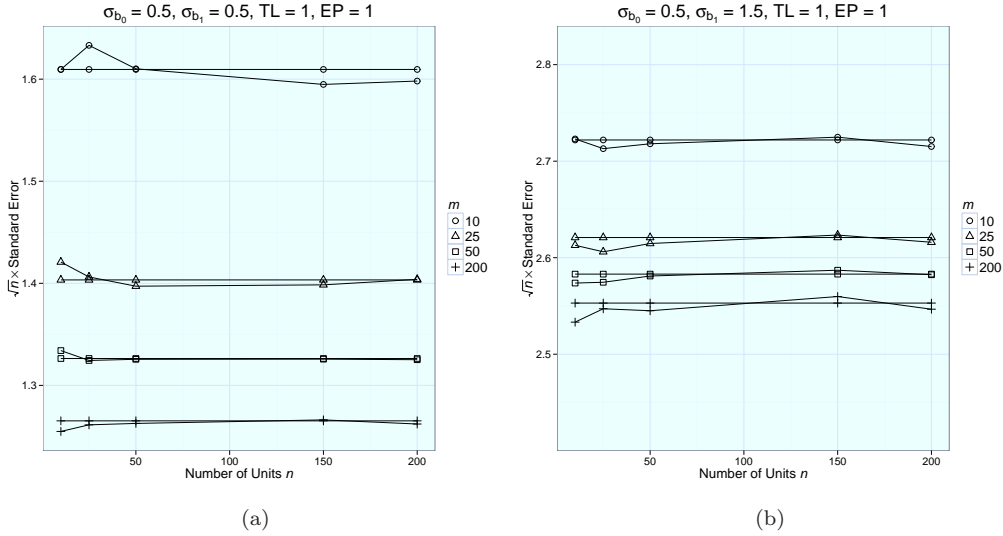

Figure 14: Factor-level combinations to assess the effects of the slope-to-slope variability. The horizontal lines corresponds to the ASE and the different symbols represent different numbers of measurement.

8   Conclusions and Areas for Future Research

Nondestructive repeated measures degradation tests are useful in understanding the material or performance degradation of a product or components over time. It is important to plan these tests carefully in order to acquire the desired level of precision while working within resource constraints (time, number of units, and number of measurements). The methodology presented in this paper can be extended to more complicated

situations. The following list suggests future work:

- Extend to models with more complicated covariance structures such as autocorrelations which might be needed when one has smaller spacing between measurements.

- In some applications *accelerated* repeated measures degradation testing is needed (e.g., when using a regression model to describe the effect of temperature on degradation rates). For examples, see Chapter 21 of Meeker and Escobar (1998).

- Bayesian methods are often useful when there is prior knowledge (e.g., from physics of failure or previous experience with similar products). When such information is available, it should be incorporated into both the analysis and test planning.

9   Acknowledgments

10   Appendix

10.1   Derivation of the Information Matrix in Section 3.2

Using equation (4) of Jenrich and Schluchter (1986), it can be shown that, using our notation from Section 2, the Hessian Matrix, $\mathbf{H}_i$, for unit i , is given by

$$
\mathbf{H}_i = \begin{pmatrix} \mathbf{H}_{\boldsymbol{\beta\beta},i} & \mathbf{H}_{\boldsymbol{\beta\phi},i} \\ \mathbf{H}_{\boldsymbol{\phi\beta},i} & \mathbf{H}_{\boldsymbol{\phi\phi},i} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}} & \dfrac{\partial^2 \mathcal{L}_i}{\partial\boldsymbol{\beta}\partial\boldsymbol{\phi}} \\ \dfrac{\partial^2 \mathcal{L}_i}{\partial\boldsymbol{\phi}\partial\boldsymbol{\beta}} & \dfrac{\partial^2 \mathcal{L}_i}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}} \end{pmatrix}.
$$

Then the information matrix can be expressed as

$$
I_i(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{X}_i^T \Sigma_i^{-1} \mathbf{X}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_i \end{pmatrix},
$$

where $\mathbf{M}_i$ is a 4×4 symmetric matrix with elements

$$
M_{jk}^i = \frac{1}{2}\mathrm{tr}(\Sigma_i^{-1}\dot{\Sigma}_{ij}\Sigma_i^{-1}\dot{\Sigma}_{ik}),\ j = 1,\ldots,4;\ k = 1,\ldots,4,
$$

and

$$
\dot{\Sigma}_{ij} = \frac{\partial \Sigma_i}{\partial \phi_j},\ j = 1,\ldots,4.
$$

From equation (4), it follows that

$$\dot{\Sigma}_{i1} = \frac{\partial \Sigma_i}{\partial \sigma_{b_0}} = \mathbf{Z}_i \begin{pmatrix} 2\sigma_{b_0} & \rho\sigma_{b_1} \\ \rho\sigma_{b_1} & 0 \end{pmatrix} \mathbf{Z}_i^T, \qquad \dot{\Sigma}_{i2} = \frac{\partial \Sigma_i}{\partial \sigma_{b_1}} = \mathbf{Z}_i \begin{pmatrix} 0 & \rho\sigma_{b_0} \\ \rho\sigma_{b_0} & 2\sigma_{b1} \end{pmatrix} \mathbf{Z}_i^T,$$

$$\dot{\Sigma}_{i3} = \frac{\partial \Sigma_i}{\partial \rho} = \mathbf{Z}_i \begin{pmatrix} 0 & \sigma_{b1}\sigma_{b_0} \\ \sigma_{b1}\sigma_{b_0} & 0 \end{pmatrix} \mathbf{Z}_i^T, \qquad \dot{\Sigma}_{i4} = \frac{\partial \Sigma_i}{\partial \sigma} = 2\sigma \mathbf{I}_i.$$

Then the information matrix for all n units is $\mathbf{I}(\boldsymbol{\theta}) = \sum_{i=1}^{n} I_i(\boldsymbol{\theta})$.

10.2 Forms of the Partial Derivatives in Section 4.2

The individual elements of $\mathbf{c}$ in (9) are $\frac{\partial d_p}{\partial \beta_0} = 1$, $\frac{\partial d_p}{\partial \beta_1} = t$, $\frac{\partial d_p}{\partial \sigma_{b_0}} = \zeta(2\sigma_{b_0} + 2t\rho\sigma_{b_1})$, $\frac{\partial d_p}{\partial \sigma_{b_1}} = \zeta(2t^2\sigma_{b_1} + 2t\rho\sigma_{b_0})$,
$\frac{\partial d_p}{\partial \rho} = \zeta(2t\sigma_{b_0}\sigma_{b_1})$, and $\frac{\partial d_p}{\partial \sigma} = 0$, where

$$\zeta = \frac{\Phi_{\text{nor}}^{-1}(p)}{2\sqrt{\sigma_{b_0}^2 + \sigma_{b_1}^2 t^2 + 2t\rho\sigma_{b_0}\sigma_{b_1}}}.$$

10.3 Derivation of t_p in Section 6.2

Let F denote the CDF of the random variable T, corresponding to the time to crossing definition in Section 6 (i.e., failure occurs when $b_0 + b_1 t \geq \mathcal{D}_f$).

$$F(t_p) = 1 - \Phi_{\text{nor}} \left(\frac{\mathcal{D}_f - \beta_0 - t_p\beta_1}{\sqrt{\sigma_{b_0}^2 + t_p^2\sigma_{b_1}^2 + 2t_p\sigma_{b_0 b_1}}} \right) = p$$

$$\Phi_{\text{nor}} \left(\frac{\mathcal{D}_f - \beta_0 - t_p\beta_1}{\sqrt{\sigma_{b_0}^2 + t_p^2\sigma_{b_1}^2 + 2t_p\sigma_{b_0 b_1}}} \right) = 1 - p$$

$$\frac{\mathcal{D}_f - \beta_0 - t_p\beta_1}{\sqrt{\sigma_{b_0}^2 + t_p^2\sigma_{b_1}^2 + 2t_p\sigma_{b_0 b_1}}} = \Phi_{\text{nor}}^{-1}(1 - p)$$

$$\frac{(\mathcal{D}_f - \beta_0 - t_p\beta_1)^2}{\sigma_{b_0}^2 + t_p^2\sigma_{b_1}^2 + 2t_p\sigma_{b_0 b_1}} = \left[\Phi_{\text{nor}}^{-1}(1 - p) \right]^2. \tag{16}$$

Let $k = \left[\Phi_{\text{nor}}^{-1}(1 - p) \right]^2$, $h = \mathcal{D}_f - \beta_0$, and $l = k\sigma_{b_0}^2$. Then

$$k\left(\sigma_{b_0}^2 + t_p^2\sigma_{b_1}^2 + 2t_p\sigma_{b_0 b_1}\right) = (h - t_p\beta_1)^2$$

$$l + t_p^2 k\sigma_{b_1}^2 + 2t_p k\sigma_{b_0 b_1} = h^2 - 2h\beta_1 t_p + \beta_1^2 t_p^2$$

$$t_p^2\left(k\sigma_{b_1}^2 - \beta_1^2\right) + 2t_p\left(k\sigma_{b_0 b_1} + h\beta_1\right) + \left(l - h^2\right) = 0.$$

22

Let $a = \left(k\sigma_{b_1}^2 - \beta_1^2\right)$, $b = 2\left(k\sigma_{b_0 b_1} + h\beta_1\right)$, $c = \left(l - h^2\right)$. Then this equation is of the form:

$$at_p^2 + bt_p + c = 0 \tag{17}$$

with the following solutions for t_p:

$$
\begin{aligned}
t_p &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\[2mm]
&= \frac{-2\left(k\sigma_{b_0 b_1} + h\beta_1\right) \pm \sqrt{4\left(k\sigma_{b_0 b_1} + h\beta_1\right)^2 - 4\left(k\sigma_{b_1}^2 - \beta_1^2\right)\left(l - h^2\right)}}{2\left(k\sigma_{b_1}^2 - \beta_1^2\right)} \\[2mm]
&= \frac{-\left(k\sigma_{b_0 b_1} + h\beta_1\right) \pm \sqrt{\left(k\sigma_{b_0 b_1} + h\beta_1\right)^2 - \left(k\sigma_{b_1}^2 - \beta_1^2\right)\left(l - h^2\right)}}{k\sigma_{b_1}^2 - \beta_1^2} \\[2mm]
&= \frac{-\left(k\sigma_{b_0 b_1} + h\beta_1\right) \pm \sqrt{k^2\sigma_{b_0 b_1}^2 + k\sigma_{b_0}^2\beta_1^2 + h^2 k\sigma_{b_1}^2 - k^2\sigma_{b_0}^2\sigma_{b_1}^2 + 2hk\beta_1\sigma_{b_0 b_1}}}{k\sigma_{b_1}^2 - \beta_1^2}.
\end{aligned}
$$

The derivation is similar when the failure definition is $b_0 + b_1 t \le \mathcal{D}_{\mathrm{f}}$, using

$$F(t_p) = \Phi_{\mathrm{nor}}\left(\frac{\mathcal{D}_{\mathrm{f}} - \beta_0 - t_p\beta_1}{\sqrt{\sigma_{b_0}^2 + t_p^2\sigma_{b_1}^2 + 2t_p\sigma_{b_0 b_1}}}\right).$$

10.4   Forms of the Partial Derivatives in Section 6.3

Let $\psi = \sqrt{k\beta_1^2\sigma_{b_0}^2 + k\sigma_{b_1}^2\left(\beta_0 - \mathcal{D}_{\mathrm{f}}\right)^2 - k^2\sigma_{b_0}^2\sigma_{b_1}^2 + k^2\rho^2\sigma_{b_0}^2\sigma_{b_1}^2 - 2k\rho\beta_1\sigma_{b_0}\sigma_{b_1}\left(\beta_0 - \mathcal{D}_{\mathrm{f}}\right)}$. Then,

$$\frac{\partial}{\partial\beta_0}t_p = \frac{\beta_1}{k\sigma_{b_1}^2 - \beta_1^2} \pm \frac{k\sigma_{b_1}^2\left[\beta_0 - \mathcal{D}_{\mathrm{f}}\right) - k\rho\beta_1\sigma_{b_0}\sigma_{b_1}}{\left(k\sigma_{b_1}^2 - \beta_1^2\right)\psi}.$$

$$\frac{\partial}{\partial\beta_1}t_p = 2\frac{\beta_1}{\left(\beta_1^2 - k\sigma_{b_1}^2\right)^2}\left(\beta_0\beta_1 - \beta_1\mathcal{D}_{\mathrm{f}} \pm \psi - k\rho\sigma_{b_0}\sigma_{b_1}\right) - \frac{1}{\beta_1^2 - k\sigma_{b_1}^2}\left[\beta_0 - \mathcal{D} \pm \frac{k\beta_1\sigma_{b_0}^2 - k\rho\sigma_{b_0}\sigma_{b_1}\left(\beta_0 - \mathcal{D}_{\mathrm{f}}\right)}{\psi}\right].$$

$$\frac{\partial}{\partial\sigma_{b_0}}t_p = \frac{1}{\beta_1^2 - k\sigma_{b_1}^2}\left[\sigma_{b_1}k\rho \pm \frac{\sigma_{b_0}\sigma_{b_1}^2 k^2 - \sigma_{b_0}k\beta_1^2 - \sigma_{b_0}\sigma_{b_1}^2 k^2\rho^2 + \sigma_{b_1}k\rho\beta_1\left(\beta_0 - \mathcal{D}_{\mathrm{f}}\right)}{\psi}\right].$$

$$\frac{\partial}{\partial\sigma_{b_1}}t_p = \frac{1}{\beta_1^2 - \sigma_{b_1}^2 k}\left[\sigma_{b_0}k\rho \pm \frac{\sigma_{b_1}k\left(\beta_0 - \mathcal{D}_{\mathrm{f}}\right)^2 - \sigma_{b_0}^2\sigma_{b_1}k^2 + \sigma_{b_0}^2\sigma_{b_1}k^2\rho^2 - \sigma_{b_0}k\rho\beta_1\left(\beta_0 - \mathcal{D}_{\mathrm{f}}\right)}{\psi}\right]$$
$$- 2\sigma_{b_1}\frac{k}{\left(\beta_1^2 - \sigma_{b_1}^2 k\right)^2}\left(\beta_0\beta_1 - \beta_1\mathcal{D}_{\mathrm{f}} - \sigma_{b_0}\sigma_{b_1}k\rho - \psi\right).$$

$$\frac{\partial}{\partial\rho}t_p = \frac{1}{\beta_1^2 - k\sigma_{b_1}^2}\left[k\sigma_{b_0}\sigma_{b_1} \pm \frac{k^2\rho\sigma_{b_0}^2\sigma_{b_1}^2 - k\beta_1\sigma_{b_0}\sigma_{b_1}\left(\beta_0 - \mathcal{D}_{\mathrm{f}}\right)}{\psi}\right].$$

$$\frac{\partial}{\partial\sigma}t_p = 0.$$

For all cases, excluding the partial derivative with respect to ρ, replace $\pm$ with "+" if $0 < p < 0.5$ and with "−" if $0.5 \le p \le 1$. For the partial derivative with respect to ρ, replace $\pm$ with "−" if $0 < p < 0.5$ and with

"+" if $0.5 \leq p \leq 1$.

References

[1] Ahmad, M., and Sheikh, A. K. (1984), "Bernstein Reliability Model: Derivation and Estimation of Parameters," *Reliability Engineering*, 8, 131-148.

[2] Bates, D., Maechler, M., and Bolker, B. (2009), *lme4: Linear Mixed-effects Models Using S4 Classes*, R package version 0.999375-42. Available at http://CRAN.R-project.org/package=lme4.

[3] Boulanger, M., and Escobar, L. A. (1994), "Experimental Design for a Class of Accelerated Degradation Tests," *Technometrics*, 36, 260-272.

[4] Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, 26, 211-252.

[5] Casella, G., and Berger, R. (2002), *Statistical Inference 2nd Edition*, Pacific Grove, CA: Duxbury.

[6] Chaloner, K., and Larntz, K. (1992), "Bayesian Design for Accelerated Life Testing," *Journal of Statistical Planning and Inference*, 33, 245-259.

[7] Diggle, P., Heagerty, P., Liang, K. Y., and Zeger, S. (2002), *Analysis of Longitudinal Data, Second Edition*, New York, NY: Oxford University Press.

[8] Faraway, J. J. (2006), *Extending the Linear Model with R*, London: Chapman & Hall.

[9] Gertsbakh, I. B., and Kordonskiy, Kh. B. (1969), *Models of Failure*, New York: Springer-Verlag, pg. 87.

[10] Hong, Y., Meeker, W. Q., and Escobar, L. A. (2008), "Avoiding Problems with Normal Approximation Confidence Intervals for Probabilities," *Technometrics*, 50, 64-68.

[11] Jenrich, R. I., and Schluchter, M. D. (1986), "Unbalanced Repeated-Measures Models with Structured Covariance Matrices," *Biometrics*, 42, 805-820.

[12] Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963-974.

[13] Lenth, R. V. (2006), "Java Applets for Power and Sample Size [Computer software]," Retrieved April 28, 2008. Available at http://www.stat.uiowa.edu/~rlenth/Power.

[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] Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D., Shabenberber (2006), *SAS for Mixed Models, Second Edition*, Cary, NC: SAS publishing.

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

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

[18] Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and the R Core Team (2012), *nlme: Linear and Nonlinear Mixed Effects Models*, R package version 3.1-103.

[19] R Development Core Team (2012), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org/.

[20] SAS software. Copyright, SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

[21] Vickers, A. J. (2003), "How Many Repeated Measures in Repeated Measures Designs? Statistical Issues for Comparative Trials," *BMC Medical Research Methodology*, 3, 3-22.

[22] Yu, H., and Tseng, S. (1999), "Designing a Degradation Experiment," *Naval Research Logistics*, 46, 689-706.