

**Analysis of Dynamic Robust Design Experiment
and
Modeling Approach for Degradation Testing**

A Thesis
Presented to
The Academic Faculty

by

Suk Joo Bae

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Industrial and Systems Engineering

School of Industrial and Systems Engineering
Georgia Institute of Technology
October, 2003

Copyright © 2003 by Suk Joo Bae

Analysis of Dynamic Robust Design Experiment
and
Modeling Approach for Degradation Testing

Approved by:

Paul H. Kvam, Committee Chair

Brani Vidakovic

Kowk-Leung Tsui, Advisor

Liang Peng

Jye-Chyi Lu

Date Approved

10/23/03

*To beloved my parents, Ssang Sun Bae and Jung Soon Ham,
my wife, Eunha,
and my adorable daughter, Sumin.*

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisor, Professor Paul H. Kvam for his guidance and encouragement throughout the entire course of my research. His valuable comments and helpful suggestions not only led to original publication of a few papers but also showed me the rigorous altitude of doing science. I would also like to extend my gratitude to my co-advisor Professor Kwok-Leung Tsui for his valuable discussions, inspiration, and friendship.

I am also grateful to Professor Jye-Chyi Lu, Brani Vidakovic, and Liang Peng for serving on my committee and their valuable comments and suggestions on the proposal and thesis.

I would like to give my deepest appreciation to my parents, my parents-in-law, my three sisters, and my wife. It is hard to imagine the work can be completed without them.

Finally, I am sincerely thankful to God for guiding my way.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	vii
SUMMARY	viii
I CONTENTS	1
PART I ANALYSIS OF DYNAMIC ROBUST DESIGN EXPERIMENT	
II INTRODUCTION	3
2.1 Background	3
2.2 Comparisons of RM, LM and RFM approaches	4
2.2.1 Identical Effect Estimates in Intercept and Slope Functions	6
2.2.2 Spurious Significant Effects	8
2.2.3 Biased Effect Estimates in Variance Function	9
2.2.4 Information Loss	9
2.3 Research Motivation	10
2.3.1 Zinc Phosphate Coating Example: No Explicit Noise Factors	10
2.3.2 Cable Actuator Example: No Significant Control-Noise Interactions	12
III GENERALIZED LINEAR MODEL (GLM) - RESPONSE MODEL (RM) APPROACH	15
3.1 Generalized Two-Step Optimization Procedure	15
3.2 Generalized Linear Model (GLM)	17
3.2.1 Maximum likelihood (ML) Estimation for Exponential Family	19
3.2.2 Maximum Likelihood (ML) for Gamma-Multiplicative Model	21
3.2.3 Residual Modeling with GLM	22
3.3 Transformation Method	24
IV DETAILED EXAMPLES	27
4.1 Zinc Phosphate Coating Example	27

4.2	Cable Actuator Example	29
4.3	Simulation Results	31
V	CONCLUDING REMARKS	35
 PART II MODELING APPROACH FOR DEGRADATION TESTING		
VI	DEGRADATION TESTING OF LIGHT DISPLAYS	37
6.1	Burn-in Characteristics of Light Displays	38
6.2	Motivating Example	40
VII	NONLINEAR RANDOM COEFFICIENTS MODEL	43
7.1	Approximation Methods	44
7.1.1	First-Order Approximation	45
7.1.2	Lindstrom-Bates Algorithm	46
7.1.3	Adaptive Importance Sampling Approximation	47
7.1.4	Adaptive Gaussian Quadrature Approximation	48
7.2	Failure-Time Distribution of Random Coefficients Model	50
7.2.1	Linear Random Coefficients Model	50
7.2.2	Nonlinear Random Coefficients Model	53
7.3	VFD Example	54
7.3.1	Comparison of Approximation Methods	54
7.3.2	Comparison of Failure-Time Distributions	58
7.3.3	Monte Carlo Results	61
VIII	ANALYSIS OF A CHANGE-POINT IN NON-MONOTONIC DEGRADATION PATH	63
8.1	Change-Point Regression Model (Non-Bayesian Approach)	63
8.1.1	Maximum Likelihood Estimation of Change-Point Model Parameters	64
8.1.2	Confidence Intervals for an Unknown Change-Point	65
8.2	Bayesian Change-point Regression Model	66
8.2.1	Posterior Distribution of a Change-Point	68
8.2.2	Bayesian Change-point Test	74
8.3	Relating Degradation to Failure Time	75
8.3.1	Random Coefficients Model for Change-Point Regression Model	76

8.3.2	Failure-Time Distribution for Change-Point Regression Model . . .	77
8.4	PDP Degradation	79
8.4.1	Change-Point Analysis	79
8.4.2	Failure-Time Analysis	81
8.5	VFD Degradation	85
8.5.1	Change-Point Analysis	85
8.5.2	Failure-Time Analysis	88
IX	CONCLUDING REMARKS	91
X	RELIABILITY CHARACTERISTICS OF FAILURE-TIME DISTRI- BUTION FOR DEGRADATION DATA	93
10.1	Degradation Process Model	93
10.1.1	Additive Degradation Model	93
10.1.2	Multiplicative Degradation Model	94
10.2	$X(t)$ Is a Random Variable	95
10.2.1	Exponential Distribution	96
10.2.2	Log-logistic Distribution	98
10.3	$X(t)$ is a Gaussian Process	99
10.4	Fatigue Crack Growth Model	102
APPENDIX A	— ZINC PHOSPHATING COATING PROCESS DATA	105
APPENDIX B	— PUSH-PULL CABLE ACTUATOR EXPERIMENT DATA	106
REFERENCES	107
VITA	113

LIST OF TABLES

1	Factors and Levels for Zinc phosphating coating process	11
2	Parameter estimates for individual VFD degradation path	56
3	VFD parameter estimates and error variances for four approximation methods 57	
4	Log-likelihood, average relative bias, and computational efficiency for four approximation methods	58
5	Laplace approximation parameter estimates and their standard errors . . .	59
6	Average relative prediction bias for the LRC and the NRC model	60
7	Quantiles and their bootstrap confidence intervals	61
8	Increase in relative error with respect to truncated log-linear model	62
9	Maximum likelihood (ML) estimates and their standard errors for the parameters in two-phase regression model	82
10	$100(1 - \delta)\%$ confidence intervals of $\hat{\gamma}$	82
11	Quantiles and their 95% bootstrap confidence intervals: PDP example . . .	85
12	Quantiles and their 95% bootstrap confidence intervals: VFD example . . .	90

LIST OF FIGURES

1	Dynamic system	4
2	Linear relationship between coating weight (Response) and geometric area (Signal)	12
3	Boxplot for the residuals from the RM approach	13
4	Flow diagram of generalized two-step optimization procedure	26
5	Boxplot for the residuals from the response model	28
6	Simulated dynamic system with the multiplicative error: $\beta_1 = 1.0, \beta_2 = 0.4$	31
7	Half normal probability plot for 2^3 factorial design	32
8	Half normal probability plot for 3^3 factorial design	34
9	Burn-in characteristics of failure data	39
10	The non-monotonic degradation path caused by incomplete burn-in	41
11	Nonmonotonic degradation paths for VFDs	42
12	Boxplots of residuals from a single degradation model	55
13	Boxplots of residuals from individually fitted models	56
14	LRC and NRC model fit for an individual VFD degradation path	59
15	Estimated CDF of VFD unit lifetime along with 90% (pointwise) confidence intervals	60
16	Two-phase regression fit for an individual PDP	80
17	Likelihood plot as a function of τ . Square marks indicate $\hat{\tau}$'s locations corresponding to the local maxima.	81
18	The mixture of failure-time distributions for the first and the second phase degradation paths	83
19	Comparison of empirical and parametric failure-time distributions: PDP example	84
20	Two-phase regression fit for an individual VFD	86
21	The distribution of change-points mean: VFD example	87
22	Comparison of empirical and parametric failure-time distributions: VFD example	89
23	Failure rate plot of exponential power distribution	97
24	Failure rate plot of 2-parameter distribution II	98
25	Failure rate plot of $bF_{2,2}^{1/2}$ distribution	100

26	Birnbaum-Saunders distribution	102
----	--	-----

SUMMARY

In PART I, optimization method for dynamic robust design experiment is illustrated. The response model (RM) approach allows greater flexibility to investigate the factor effects for the dynamic robust design problem. This article generalizes Tsui's [92] RM approach in the dynamic system. Based on the Generalized Linear Model (GLM), generalized two-step optimization procedure is introduced to further reduce the process variance by identifying the effects of hidden noise variables. Our suggested method (called *GLM-RM approach*) provides more reliable results by modeling the residuals from the fitted response model simultaneously. Simulation is executed to verify such facts. Two practical examples are analyzed with three existing approaches: *RM*, *loss model (LM)*, *response function model (RFM) approach*, and their results are compared with the GLM-RM approach.

In PART II, various modeling approaches are introduced to estimate failure time distributions of degradation testing data precisely. 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. We 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.

As first modeling approach, Random coefficients are used to model this phenomenon in a nonlinear way. In many situations, the relative efficiency of the lifetime estimate is improved over the standard estimators based on transformed linear models.

Secondly, a (log) linear model with a *change-point* is introduced to describe the non-monotonic degradation path. We provide an inference procedure for the parameters of a *change-point* model and introduce methods for estimating the lifetime distribution derived from it.

We show that reliability estimation can be substantially improved by considering *burn-in* characteristics of light displays. Two real examples, plasma display panels (PDPs) and vacuum fluorescent displays (VFDs), are analyzed.



CHAPTER I

CONTENTS

In PART I, analytical method for dynamic robust design experiment is illustrated based on the Generalized Linear Model (GLM) and generalized two-step optimization procedure by identifying the effects of hidden noise variables.

In PART II, various modeling approaches are introduced to estimate failure time distributions of degradation testing data precisely. Two kinds of modeling approach; Nonlinear Random coefficients (NRC) models and a (log) linear model with a *change-point* are used to describe the nonmonotonic degradation path. We show that reliability estimation can be substantially improved by considering *burn-in* characteristics of light displays.

PART I

Analysis of Dynamic Robust Design Experiment

CHAPTER II

INTRODUCTION

2.1 Background

Continuous quality improvement has become widely recognized by every industry as critical to maintain a competitive advantage in the international marketplace. It is also recognized that quality improvement activities are most efficient and cost effective when implemented as part of the design and development stage, through such techniques as statistical design of experiments.

Robust design introduced by Taguchi [86] is a strategic method for improving the performance of a system in the development stage. The main objective of robust design is to reduce the performance variation in products and processes by selecting the setting of easy-to-control factors which make product and process designs insensitive (*i.e.*, robust) to hard-to-control factors (noises). The robust design method has been applied to problems with the static system and the dynamic system. The static system is defined as that for which the desired system output has a fixed target value. In the dynamic system, the response is determined by an input signal set by a system operator, through a transfer function. The transfer function depends on control and noise factors (see Figure 1). Loss is incurred if the response is different from a target that may depend on the input signal. In general, the target value is chosen to reflect a customer's intention. The dynamic system is also called as a "signal-response system" by Miller and Wu [65] or a "multiple-target system" by Joseph and Wu [43].

Although there exists a broad array of robust design applications for the dynamic system, most research has been focused on the static system. Recently, the modeling and analysis methods for the dynamic system have been developed by several researchers (see Lunani *et al.* [58]; Miller and Wu [65]; McCaskey and Tsui [61]; Tsui [90], [91], [92]; and Joseph and Wu [43], [44]).

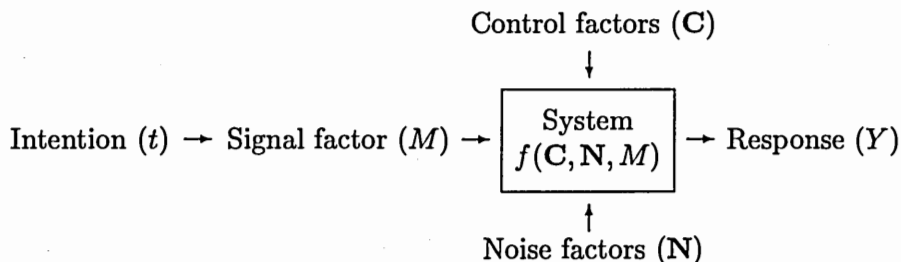


Figure 1: Dynamic system

For the dynamic system, an ideal quality is based on an ideal relationship between the signal and the response. A quality loss is caused by the deviation from this ideal relationship. Significant quality improvement can be achieved by firstly defining a system’s ideal function and then using designed experiments to search for an “optimal” design which minimizes the deviation from this function. The derived optimal solutions, however, might be different according to modeling approaches in the dynamic system. Tsui [90], [91], [92] illustrated such facts thoroughly by comparing the optimization procedure derived from three different approaches: the *response model (RM)*, *loss model (LM)*, and *response function model (RFM) approach*. He also showed the RM approach allows greater flexibility to investigate factor effects for the dynamic robust design experiment.

2.2 Comparisons of RM, LM and RFM approaches

The response model (RM) approach was first proposed in Welch *et al.* [94] and Shoemaker *et al.* [81] for the static system. The RM approach for the dynamic system directly models the response as a function of the control, noise, and signal factor effects. Tsui [92] suggested the following general additive response model:

$$\begin{aligned}
 Y = & \beta_0(\mathbf{C}) + \beta_1(\mathbf{C})f_1(M) + \dots + \beta_k(\mathbf{C})f_k(M) \\
 & + \sum_{h=1}^q [\gamma_h(\mathbf{C}) + \rho_{1h}(\mathbf{C})f_1(M) + \dots + \rho_{kh}(\mathbf{C})f_k(M)]N_h + \epsilon, \quad (2.2.1)
 \end{aligned}$$

where \mathbf{C} is a vector of p control factors, $N_h, h = 1, \dots, q$, is a (explicit) noise factor, which is included and varied systematically in the experiment, and $(f_1(M), \dots, f_k(M))$ is k known

functions of the signal factor M . The parameters $\beta_0(\mathbf{C})$, $\beta_i(\mathbf{C})$, $\gamma_h(\mathbf{C})$, and $\rho_{ih}(\mathbf{C})$ describe the functional relationships between the response and the control factors \mathbf{C} , for $i = 1, \dots, k$. It is assumed that ϵ is *iid* $\mathcal{N}(0, \sigma_\epsilon^2)$ and N_h 's are *iid* $\mathcal{N}(0, \sigma_N^2)$, and ϵ and N_h are independent.

In the RM approach, all the effects in (2.2.1) are first estimated by combining all of control, noise and signal factors in a single design matrix and using overall raw data. The second step is to use these effect estimates to approximate intercept, slope and variance functions, and then determine the control factor settings that optimize the dynamic system. Tsui's [92] two-step optimization procedure is obviously different from that of Taguchi's for optimizing the dynamic system.

Shoemaker *et al.* [81] classified Taguchi's analysis approach as the loss model (LM) approach in which the estimates of control, noise and signal effects are first computed and then a model is fitted to those performance measures to determine the optimal control factors' combination. The choice of performance measures for several different underlying models was discussed in Leon *et al.* [50] and Box [10]. The analysis procedure of the LM approach in the dynamic system is to first estimate the parameters in model (2.2.1) at each fixed value of \mathbf{C} . As these parameters are functions of the control factors, the second step is to treat each of these parameter estimates as a separate response, then estimate the control factor effects and determine the control factor settings robust to the noise factors.

As another modeling approach, Miller and Wu [65] proposed the response function model (RFM) approach to model performance measures as the functions of both control and noise factor effects. At any fixed values of \mathbf{C} and \mathbf{N} , the response model in (2.2.1) can be expressed as a linear regression model

$$Y = \tilde{\alpha}(\mathbf{C}, \mathbf{N}) + \tilde{\beta}_1(\mathbf{C}, \mathbf{N})f_1(M) + \dots + \tilde{\beta}_k(\mathbf{C}, \mathbf{N})f_k(M) + \epsilon, \quad (2.2.2)$$

where

$$\tilde{\alpha}(\mathbf{C}, \mathbf{N}) = \beta_0(\mathbf{C}) + \sum_{h=1}^q \gamma_h(\mathbf{C})N_h,$$

and

$$\tilde{\beta}_i(\mathbf{C}, \mathbf{N}) = \beta_i(\mathbf{C}) + \sum_{h=1}^q \rho_{ih}(\mathbf{C})N_h,$$

for $i = 1, \dots, k$, and $h = 1, \dots, q$.

The assumption for an error term ϵ is same as in the RM approach. The analysis procedure of the RFM approach is to first estimate the parameters in model (2.2.2), $\tilde{\alpha}(\mathbf{C}, \mathbf{N})$, $\tilde{\beta}_1(\mathbf{C}, \mathbf{N}), \dots, \tilde{\beta}_k(\mathbf{C}, \mathbf{N})$ and σ_ϵ^2 , at each fixed value of \mathbf{C} and \mathbf{N} . As these parameters are functions of the control and noise factors, the second step is to treat each of these parameter estimates as a separate response, and estimate the control and noise factor effects, then determine the control factor settings that optimize the dynamic system.

It can be shown that the effect estimates of intercept and slope functions from three different approaches are identical. However, even if the estimates are identical, the significant effects identified in the LM and RFM approaches may be spurious. As for the effect estimates of the variance function, the LM approach may create unnecessarily biased estimates. Both spurious effects and unnecessary biases lead to non-optimal solutions. we will discuss each of these issues in detail at the following.

2.2.1 Identical Effect Estimates in Intercept and Slope Functions

To show that the intercept and slope effect estimates from three different approaches are identical, we need the following Lemma.

Assume that the expected response $E(y)$ at \mathbf{x} and \mathbf{z} can be described by the linear model:

$$E(y) = \sum_{j=1}^k \theta_{0j} f_j(\mathbf{z}) + \sum_{i=1}^h \sum_{j=1}^k \theta_{ij} g_i(\mathbf{x}) f_j(\mathbf{z}), \quad (2.2.3)$$

where $f_1(\mathbf{z}) \equiv 1$. Suppose the design is a product array (see Shoemaker et al. [81], for definition) of an inner design with n runs at $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ and an outer design with m runs at $(\mathbf{z}_1, \dots, \mathbf{z}_m)$. We denote the corresponding response of the design as $\mathbf{y}' = (\mathbf{y}'_1, \dots, \mathbf{y}'_n)$, with \mathbf{y}_i being the $m \times 1$ response vector corresponding to the m outer design runs at the i th run of the inner design. It follows that the regression matrix of the linear model (2.2.3), \mathbf{X} can be expressed as the Kronecker product of the corresponding regression matrices, *i.e.*,

$$\mathbf{X} = (\mathbf{X}_x^* \otimes \mathbf{X}_z) = (\mathbf{1}_n \otimes \mathbf{X}_z \ \mathbf{X}_x \otimes \mathbf{X}_z), \quad (2.2.4)$$

where

$$\mathbf{X}_x^* = \begin{pmatrix} 1 & g_1(\mathbf{x}_1) & \cdots & g_h(\mathbf{x}_1) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 1 & g_1(\mathbf{x}_n) & \cdots & g_h(\mathbf{x}_n) \end{pmatrix} = (\mathbf{1}_n, \mathbf{X}_x), \quad \mathbf{X}_z = \begin{pmatrix} f_1(\mathbf{z}_1) & \cdots & f_k(\mathbf{z}_1) \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ f_1(\mathbf{z}_m) & \cdots & f_k(\mathbf{z}_m) \end{pmatrix},$$

with $\mathbf{1}_n$ being a n -dimensional vector of 1's and $A \otimes B$ being the Kronecker product of matrices A and B .

Lemma 1 *Assume that the column vectors in matrix \mathbf{X}_x^* are orthogonal to each other. Then the least squares estimates of the parameters in model (2.2.3) are identical to the estimates obtained by the following sequential steps of least squares estimation:*

- (1) *For each row of matrix \mathbf{X}_x , compute the least squares estimates using matrix \mathbf{X}_z and corresponding response vector \mathbf{y}_i , i.e.,*

$$\hat{\beta}_i = (\hat{\beta}_{i1}, \dots, \hat{\beta}_{ik})' = (\mathbf{X}'_z \mathbf{X}_z)^{-1} (\mathbf{X}'_z \mathbf{y}_i), \quad i = 1, \dots, n.$$

- (2) *For $j = 1, \dots, k$, compute the least squares estimates of $\theta_j = (\theta_{0j}, \theta_{1j}, \dots, \theta_{hj})$ using matrix \mathbf{X}_x^* and response vector $\hat{\beta}_i^* = (\hat{\beta}_{1j}, \dots, \hat{\beta}_{nj})'$.*

Proof. See the details in Tsui [92] for proof. ■

First, we compare the effect estimates of intercept and slope functions for the LM and RM approaches. As described in section 1.2, the second step of the LM approach is to estimate the control factor effects that affect intercept and slope functions. These effect parameters are corresponding to the control main effect and control-by-signal interaction parameters in model (2.2.1). According to the analysis procedure, effect estimates of intercept and slope functions from the LM approach are obtained through two sequential least squares estimations: (i) At each row of the control array design, consider the noise runs as replications and compute the least squares estimates of the regression model for signal functions, $f_1(M), \dots, f_k(M)$. (ii) For each of the coefficients in the fitted regression models, compute the least squares estimates according to the control array.

Following Lemma 1, matrix \mathbf{X}_x corresponds to the control array and matrix \mathbf{X}_z corresponds to the regression matrix of the fitted regression model. As the column vectors in the control array are orthogonal, the control effect estimates of intercept and slope functions from the LM approach are identical to the least squares estimates of the RM approach.

Next, Comparing the effect estimates of intercept and slope functions for the RFM and RM approaches, effect estimates of intercept and slope functions from the RFM approach are obtained through two sequential least squares estimations: (i) At each row of the control-by-noise product array design, compute the least squares estimates of the regression model for signal functions. (ii) For each of the coefficients in the fitted regression models, compute the least squares estimates according to the control-by-noise product array.

Following Lemma 1, matrix \mathbf{X}_x corresponds to the control-by-noise product array regression matrix and matrix \mathbf{X}_z corresponds to the regression matrix of the fitted regression model. As the column vectors in the product array regression matrix are orthogonal, the control and noise effect estimates of intercept and slope functions from the RFM approach are identical to the least squares estimates of the RM approach.

2.2.2 Spurious Significant Effects

Even if effect estimates of the intercept and slope functions are the same, significant effects identified from three different approaches can be quite different. This is because the groupings of effect estimates are different according to modeling approaches. When error estimate is not available, a half normal probability plot or other non-replicated experiment analysis techniques (*e.g.*, Lenth [49]) are commonly used for identifying significant effects. With the RFM approach, identified significant effects may be spurious and thus result in unreliable error estimates owing to small sized effects in each grouping. For the LM approach, sparse effects problem is more serious as each group contains only the control effects. However, for the RM approach, control and noise effects are incorporated by all the intercept and slope functions together so that the error estimate is more reliable. Besides, it allows us to compare the effect estimates across the intercept and slope functions.

2.2.3 Biased Effect Estimates in Variance Function

As for the control effect estimates of the variance function, there is no identical relationship between these three approaches. In the LM approach, under model (2.2.1) with linear control effects, the experimental design allows confounding between main effects and two-factor interactions or between two-factor interactions themselves in the control array (*i.e.*, resolution III or IV). Consequently, there exists a potential bias problem in estimating the control factor effects that affect the variance function and thus may lead to non-optimal solutions. Whereas, the estimates of control and noise main effects and the control-by-noise interactions in (2.2.1) are unbiased with the RFM or the RM approaches when there are no control-by-control and noise-by-noise interactions in model (2.2.1).

2.2.4 Information Loss

Apart from the potential bias problem, the LM approach suffers from the information loss problem in comparison with the RFM and RM approaches. As the LM approach directly models the variance function, it will not provide the information on how the control factors dampen the variation caused by individual noise factors. On the other hand, the RFM and RM approaches model the control-by-noise interaction effects and thus provide those information.

Based on above properties, Tsui [92] claimed that the RM approach has more potential to reach to an optimal solution in the dynamic system. However, if the error variance σ_e^2 is not independent of control or signal, even (explicit) noise factors, the optimal control settings derived from the RM approach lead to increased variability. In that situation, it is important to model the residual variance via control, noise and signal factors in the RM approach. The residual model provides additional information about the process variation caused by hidden or uncontrollable noise variables. In next chapter, we will illustrate how the residuals from the response model can be used to further reduce the process variance in the dynamic system.

2.3 Research Motivation

Process engineers frequently face the situation where they cannot identify the noise factors or control such factors even if they are identified in a manufacturing process. In such circumstances, the noise factors are not given a priori. In the RM approach, since the control factors which affect the process variance can be selected primarily using the control by noise interactions, the RM approach might not be a good alternative over the LM and the RFM approaches. At the following, we introduce two case studies which provide our research motivation.

2.3.1 Zinc Phosphate Coating Example: No Explicit Noise Factors

Lin and Wen [51] applied the dynamic robust experimental design to obtain the uniform zinc phosphate coating. Eight control factors and a signal factor (geometric area of low-carbon steel plate) were adopted to examine the uniformity for a phosphating process. Level descriptions of control and signal factors were given in Table 1.

The response was the difference in the weight of the phosphate coating before and after stripping. The noise factor was a plated film location where the difference in coating uniformity was present. It cannot be controlled nor observed during a process, therefore, (explicit) noise factors were not given in this experiment. A standard OA18 orthogonal array was used as an experimental design. The response values from OA18 orthogonal array are summarized in **APPENDIX A**.

The objective of this experiment is to determine the best control factor settings that give the uniform plating film thickness, regardless of the unobservable noise variable (plated film location). The ideal relationship between response and signal is assumed to be linear. Figure 2 shows strong linear relationship between signal and response, along with the trend in which response variation is proportional to the signal value.

In the LM approach, the response was first fitted to the signal only for each run of fixed control combinations. Because the intercept was insignificant in the response-signal model, the response was re-fitted into a linear model without intercept. The effects for the slope and error variance were estimated using ordinary least squares (OLS) at each run. All control

Table 1: Factors and Levels for Zinc phosphating coating process

Factors		Level 1	Level 2	Level 3	
A : $Ni(NO_3)_2 \cdot 6H_2O$	(g/L)	2	3		
B : Acid-clean time	(min)	10	15	20	
C : Phosphating temp.	($^{\circ}C$)	70	80	90	
D : Phosphating time	(min)	5	10	15	
E : ZnO	(g/L)	0.5	1.0	1.5	
F : H_3PO_4	(g/L)	3	5	7	
G : $NaNO_2$	(g/L)	0.15	0.3	0.5	
H : NaH_2PO_4	(g/L)	20	25	30	
Signal		Level 1	Level 2	Level 3	Level 4
M : Geometric area (cm^2)		20	50	100	200

factors except one (A) have three levels. We decomposed seven factors' ($B - H$) effects into orthogonal linear (l) and quadratic (q) contrast $\{-1, 0, 1\}$ and $\{1, -2, 1\}$, respectively. The significant control effects on the slope (β) and log variance ($\log \sigma_L^2$) were identified based on a half-normal probability plots and Lenth's [49] method. At 5 % significance level the fitted models were

$$\begin{aligned}\hat{\beta} &= 0.684 + 0.224D_l - 0.122G_l - 0.102H_l - 0.058B_q - 0.062E_q - 0.060H_q, \\ \widehat{\log \sigma_L^2} &= 2.153 - 0.826E_l.\end{aligned}$$

As only the linear effect of E was significant for the process variance, this factor is set at level "1" to reduce the process variance. Next, slope is adjusted to its target; because explicit noise factors are not given in this experiment, control by noise interactions cannot be estimable. Consequently, the optimization procedure of the LM approach is identical to that of the RFM approach.

In the RM approach, all the raw data were first fitted to model (2.2.1) against the control and signal factors. Then we calculated t -statistics by estimating the pure error with the remaining degrees of freedom. All effects at 5% significant level were included in the following OLS model:

$$\begin{aligned}\hat{y} &= (0.684 + 0.049B_l + 0.075C_l + 0.224D_l + 0.053E_l - 0.122G_l - 0.102H_l \\ &\quad - 0.058B_q + 0.014D_q - 0.062E_q - 0.031F_q + 0.036G_q - 0.060H_q)M, \quad (2.3.5)\end{aligned}$$

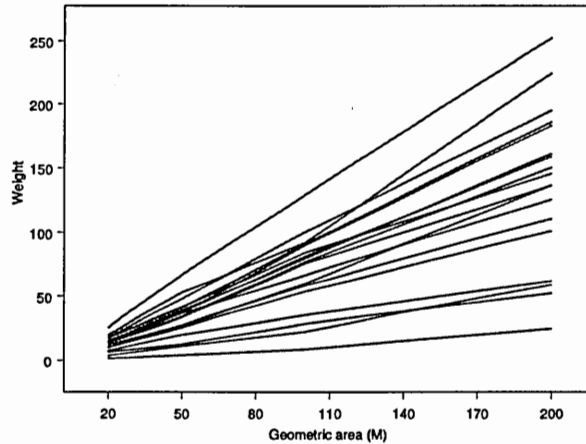


Figure 2: Linear relationship between coating weight (Response) and geometric area (Signal)

with $R^2 = 0.993$. We should first minimize the process variance on the basis of the fitted model (2.3.5). However, because noise factors are not included explicitly in this experiment variance effects cannot be identified using the traditional RM approach. As shown in Figure 3, process variance is not uniform within some control factors, especially within signal factor levels.

2.3.2 Cable Actuator Example: No Significant Control-Noise Interactions

Byrne and Quinlan [12] considered a push-pull cable actuator experiment. This experiment was also analyzed in Tsui [91] and Joseph and Wu [43] under different modeling approaches. In this experiment, the signal is the input displacement and the response is the output displacement. Eleven potential control factors, designated A to K , were identified for use in the experiment and tested at two levels using a standard orthogonal array OA12. One noise factor which represents the settings of four compounded noise factors was used. For each of the 12 rows in the array, three replications at each of two-level noise conditions were collected for each of three signal levels. The experiment, therefore, required 216 ($12 \times 2 \times 3 \times 3$) observations. The data from OA12 experiment are given in **APPENDIX B**.

The objective of this experiment is to first determine the control factor settings that

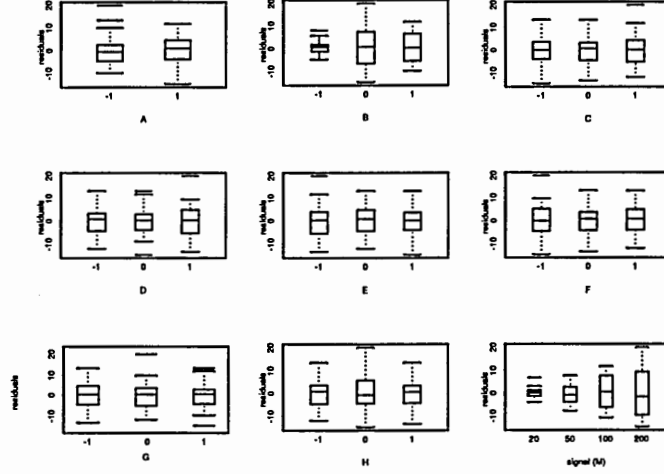


Figure 3: Boxplot for the residuals from the RM approach

minimize the process variance while maintaining the intercept at zero. An ideal relationship is to make no difference between the input and output displacement, and maximize input-output linearity over the range of interest (see Byrne and Quinlan [12] for details).

For the LM approach, we referred to the analysis done in Tsui [91]. Based on his analysis results, the fitted loss models for the intercept (α), the slope (β), and the error variance (σ_L^2) at 10% significance level were

$$\hat{\alpha} = 0.930 - 0.612B + 0.431G,$$

$$\hat{\beta} = 0.511 + 0.089B - 0.145G + 0.065H - 0.031J,$$

$$\hat{\sigma}_L^2 = 0.297 - 0.067B - 0.094D.$$

For the RFM approach, we calculated the intercept (α'), slope (β') and variance (σ_{RFM}^2) estimates from linear regression model using OLS for each of 24 runs as the combination of control and noise factors. At 10% significance level in Lenth's method, the fitted models for these performance measures were

$$\hat{\alpha}' = 0.933 - 0.646B + 0.435G,$$

$$\hat{\beta}' = 0.511 + 0.091B - 0.145G + 0.065H - 0.029J,$$

$$\hat{\sigma}_{RFM}^2 = 0.307 - 0.094B - 0.088D + 0.042G - 0.036H + 0.040J.$$

Following to the RM approach, all terms at 10% significance level are included in the following selected model:

$$\hat{y} = 0.930 - 0.612B + 0.431G + (0.511 + 0.089B - 0.145G + 0.065H - 0.030J)M - 0.011MN, \quad (2.3.6)$$

with $R^2 = 0.987$. Because no control-by-noise interactions are significant in the response model, it is impossible to reduce the process variance using the significant control by noise interactions with the RM approach. When the RM result is compared with the LM and RFM results, the identified significant effects for the intercept and slope functions are identical in three different approaches. As for process variance, control factors B and D are included commonly in the variance function with both the LM and the RFM approach, but only RM approach fails to identify significant control factors for the process variance.

Considering two examples above, it is necessary to consider the effect of hidden noise variables which are not given explicitly in the experiment. Hidden noise variables may interact with control, signal, and even (explicit) noise factors. However, those factors might not be detected by the RM approach since only explicit noise factors can be modeled in the the response model. If we can identify the control factors (called *dispersion effects*) that interact with hidden noise variables with the RM approach, the process variance is substantially reduced by adjusting those control factors. In next chapter, a new method to identify dispersion effects is suggested following Generalized Linear Model (GLM) approach in the dynamic system.

CHAPTER III

GENERALIZED LINEAR MODEL (GLM) - RESPONSE MODEL (RM) APPROACH

3.1 Generalized Two-Step Optimization Procedure

Suppose that the true response model is linear in some functions of p control factors (C), q explicit noise factors (N) and signal factor M of K levels, and also linear in n known functions of the signal factor $f(M) = (f_1(M), \dots, f_n(M))$. All these factors and functions interact with each other. When there are I control runs, J noise runs, and L replications in each one of the signal levels, the response model for $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$, and $l = 1, \dots, L$, is

$$\begin{aligned}
 Y_{ijkl} = & \beta_0(C_i) + e_0(C_i, N_j) + [\beta_1(C_i) + e_1(C_i, N_j)]f_1(M_k) + \dots \\
 & + [\beta_n(C_i) + e_n(C_i, N_j)]f_n(M_k) + \varepsilon_{ijkl},
 \end{aligned} \tag{3.1.7}$$

where $\beta(C_i) = (\beta_0(C_i), \dots, \beta_n(C_i))'$ represents the functional relationship between the response and i^{th} control factors' combination C_i , $e(C_i, N_j) = (e_0(C_i, N_j), \dots, e_n(C_i, N_j))'$ denotes the relationship between the response and j^{th} noise factors' combination N_j only or control-by-noise interactions. When the ideal signal-response function is linear, $f_1(M) = M$ and $f_2(M) = \dots = f_n(M) = 0$.

We assume that the random error ε_{ijkl} 's are *iid* $\mathcal{N}(0, \sigma_{\varepsilon}^2)$. For simplicity, we omit the subscript to illustrate the optimization procedure. In the dynamic robust design, the explicit noise factors N are assumed to be random as the main sources of variation in the manufacturing process. It follows that $e(C, N)$ are *iid* $\mathcal{N}(0, \sigma_e^2(C))$, where $\sigma_e^2(C) = \text{Var}_N[e(C, N)]$. The errors (ε_{ijkl} 's), ϵ are *iid* $\mathcal{N}(0, \sigma_\epsilon^2(C, f(M)))$, where $\sigma_\epsilon^2(C, f(M)) = \text{Var}_\epsilon[e(C, f(M))]$. We assume that e and ϵ are independent. $\sigma_e^2(C)$ represents the variation caused by the explicit noise factors N and $\sigma_\epsilon^2(C, f(M))$ is the variation caused by a hidden noise variable. Possibly, the same control factor can be identified as significant in both

variance functions, but the optimal control level in one conflicts with the other. Thus, the decomposition of the process variance into $\sigma_e^2(C)$ and $\sigma_e^2(C, f(M))$ is essential to secure the optimal solution in the dynamic system.

The conditional mean and variance of the response Y given C, M are, respectively

$$E(Y) = \mu(C, M) = \beta(C)'f_1(M), \quad (3.1.8)$$

$$\begin{aligned} Var(Y) &= Var[e(C, N)'f_1(M)] + Var[\epsilon] \\ &= f_1(M)'\sigma_e^2(C)f_1(M) + \sigma_e^2(C, f(M)), \end{aligned} \quad (3.1.9)$$

where $f_1(M) = (1, f_1(M), \dots, f_n(M))'$.

As $Var(Y)$ depends on the signal factor M , we integrate over M to determine optimal settings of the control factors, assuming that the signal M is uniformly distributed at the range of interest while Joseph and Wu [43] considered a general distribution for a target t . Then the variance of the response Y is

$$\sigma_Y^2(C) = \int_{M_L}^{M_H} [f_1(M)'\sigma_e^2(C)f_1(M) + \sigma_e^2(C, f(M))]dM, \quad (3.1.10)$$

where M_L, M_H denote the low and high limit of the signal, respectively. Suppose the control factors can be divided into three disjointed groups, *i.e.*, $C = (C_1, C_2, C_3)$, and $\sigma_Y^2(C)$ is a function of C_2 and C_3 . C_2 contains the control factors which interact significantly with explicit noise factors and C_3 contains remaining significant control factors except C_2 in σ_e^2 . C_1 is classified as the adjustment factors for the target function. Then the mean function in (3.1.8) can be denoted as

$$\mu(C_1, C_2, C_3, M) = \beta(C_1, C_2, C_3)'f_1(M).$$

If the quadratic loss is believed to be a good approximation, the objective of robust design is to minimize the sum of the process variance and the square of the process bias over an ideal target function $t(M) = t(M, C)$. The average loss for dynamic response model (3.1.7) is

$$\begin{aligned} R(C_1, C_2, C_3) &= E_M\{Var(Y)\} + E_M\{[\mu(C_1, C_2, C_3, M) - t(M)]^2\} \\ &= \sigma_e^2(C_2) + \sigma_e^2(C_2, C_3) + E_M\{[\mu(C_1, C_2, C_3, M) - t(M)]^2\}. \end{aligned} \quad (3.1.11)$$

Under the constraint that the mean function must be adjusted to $t(M)$, we minimize the average loss $R(C_1, C_2, C_3)$. That is,

$$\min_{C_1, C_2, C_3} R(C_1, C_2, C_3)$$

subject to

$$\mu(C_1, C_2, C_3, M) = t(M) \quad \text{for any } M.$$

A generalized two-step optimization procedure is introduced to minimize the loss function $R(C_1, C_2, C_3)$. The procedure is as follows:

Step 1. (a) Find C_2^* that minimize $\sigma_e^2(C_2)$

(b) Find C_3^* that minimize $\sigma_e^2(C_2^*, C_3)$.

Step 2. Find C_1^* so that $\mu(C_1^*, C_2^*, C_3^*, M) = t(M)$ for any M .

If explicit noise factors are not given or no control-by-noise interactions are identified as significant in the response model (that is, C_2 are null), then we skip Step 1-(a) and go directly to Step 1-(b) for reducing the process variation. If σ_e^2 is a function of C_2 and C_3 , i.e., $\sigma_e^2(C_2, C_3)$, then change Step 1 to find (C_2^*, C_3^*) that minimize σ_e^2 . It can be easily shown that (C_1^*, C_2^*, C_3^*) minimizes $R(C_1, C_2, C_3)$ since the first term in (3.1.11) caused by the explicit noises is minimized at C_2^* , the second term is minimized by adjusting the remaining control factors C_3 to C_3^* at fixed C_2^* , and finally the third term drops to zero at (C_1^*, C_2^*, C_3^*) by choosing C_1^* so that mean function is adjusted to the target function $t(M)$, without affecting its variance. The generalized two-step optimization procedure is summarized as a flow diagram in Figure 4.

3.2 Generalized Linear Model (GLM)

Most dynamic systems have been solved under the assumption that the residual error is normally distributed with constant variance. Ordinary least squares (OLS) estimates are obtained from such an assumption and successfully applied in most applications. However, when the variance is a function of the mean as in the exponential family (i.e., binomial,

poisson, and gamma distribution, *etc.*), the inference procedure based on OLS is inaccurate. Introducing a Generalized Linear Model (GLM), the assumption of normality and constant variance for the residual errors is no longer required, although the way in which the variance depends on the mean must be known. In the GLM, there are three important components:

- response distribution
- linear predictor, which is a linear combination of the design variables
- link function which relates the mean response to linear predictor

The response distribution and the link function are not independent of each other because certain types of models are more appropriate for some distributions than for others. We can view the selection of the link function in a vein similar to the choice of a transformation on the response. However, it is important to understand that differences between the two lie in fact that the link function is a transformation of the population mean, not the data. Unlike a transformation, it can take advantage of the natural distribution of the response. Just as not using an appropriate transformation can result in problems with a fitted linear model, improper choice of link function may produce seriously distorted results (see Myers *et al.* [68]).

Recently, GLM has been highlighted in an industry experiment where traditional model assumption is violated. Hamada and Nelder [37] discussed the power of GLM as a method to improve the quality and provided several models which can handle a broad range of applications. Myers and Montgomery [67] presented a tutorial to GLM and analyzed real industrial experiments using statistical software such as SASTM GENMOD procedure.

In the robust design experiment, a GLM was first introduced by Nelder and Lee [69], and extended by Grego [35] and Engel [25] to solve the static system. Their methods are considered as the LM approach in the context of modeling. Engel and Huele [26] adopted GLM to detect the additional dispersion effects with the RM approach in the static system. Lee and Nelder [48] illustrated robust design applications of a GLM with both the LM and the RM approaches. However, most robust design applications of a GLM have been focused

on the static system only. Joseph and Wu [43] applied a GLM implicitly using a different performance measure in the analysis of dynamic robust design experiment.

We introduce a GLM to the dynamic robust design problem for the purpose of reducing the variation caused by the noise variables which are not included explicitly in the experiment.

3.2.1 Maximum likelihood (ML) Estimation for Exponential Family

Maximum likelihood (ML) is a theoretical method for parameter estimation in GLM. Suppose our observation Y is generated from an exponential family. The log-likelihood function is

$$\mathcal{L} = \sum_{i=1}^n \{a(\phi)[Y_i\theta_i - c(\theta_i)] + b(\phi, Y_i)\},$$

where θ is a location parameter and fixed ϕ represents a nuisance parameter such as the variance σ^2 of a normal distribution. The function $a(\phi) > 0$ is generally of the form $a(\phi) = \phi \cdot \delta$ for a known constant δ .

The derivative of the log-likelihood with respect to coefficient vector γ is

$$\frac{\partial \mathcal{L}}{\partial \gamma} = \sum_{i=1}^n \frac{\partial \mathcal{L}}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \zeta_i} \cdot \frac{\partial \zeta_i}{\partial \gamma}, \quad (3.2.12)$$

where ζ is link function for the response Y . Note that

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \sum_{i=1}^n a(\phi)(Y_i - c'(\theta_i)) = \sum_{i=1}^n a(\phi)(Y_i - \mu_i),$$

and

$$\frac{\partial \zeta_i}{\partial \gamma} = \mathbf{x}_i,$$

where \mathbf{x}_i denotes i^{th} predictor vector. Putting this all together, maximum likelihood estimates of the parameters can be obtained by solving the following equations for γ :

$$\frac{\partial \mathcal{L}}{\partial \gamma} = \sum_{i=1}^n (Y_i - \mu_i) \left(\frac{\partial \theta_i}{\partial \zeta_i} \right) \mathbf{x}_i = 0,$$

with constant $a(\phi)$. In matrix form these equations are

$$\mathbf{X}' \Delta (\mathbf{Y} - \boldsymbol{\mu}) = 0, \quad (3.2.13)$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$, $\boldsymbol{\mu} = [\mu_1, \dots, \mu_n]'$, and $\boldsymbol{\Delta} = \text{diag}\{\partial\theta_i/\partial\zeta_i\}$. These are called *maximum likelihood score equations*.

To solve the maximum likelihood score equations, iteratively reweighted least squares (IRLS) can be used. We start by finding a first-order Taylor series approximation in the neighborhood of the solution ζ_i^* , which is

$$Y_i - \mu_i \approx \frac{\partial\mu_i}{\partial\zeta_i}(\zeta_i^* - \zeta_i).$$

Because $\partial\theta_i/\partial\zeta_i = (\partial\theta_i/\partial\mu_i) \cdot (\partial\mu_i/\partial\zeta_i)$, and $\partial\theta_i/\partial\mu_i = \text{Var}_\mu^{-1}$, the score equations (3.2.13) then become

$$\sum_{i=1}^n \frac{(\zeta_i^* - \zeta_i)}{\text{Var}_\mu} \left(\frac{\partial\mu_i}{\partial\zeta_i} \right)^2 \mathbf{x}_i = 0.$$

It is straightforward to show that

$$\text{Var}(\zeta_i^* - \zeta_i) \approx \text{Var}_\mu \cdot \left(\frac{\partial\zeta_i}{\partial\mu_i} \right)^2,$$

and as a result, the score equations are simplified to

$$\sum_{i=1}^n \frac{(\zeta_i^* - \zeta_i)}{\text{Var}(\zeta_i^* - \zeta_i)} \mathbf{x}_i = 0.$$

Let $\mathbf{V}_\zeta = \text{diag}\{\text{Var}(\zeta_i^* - \zeta_i)\}$, then the score equations in matrix form are

$$\mathbf{X}'\mathbf{V}_\zeta^{-1}(\boldsymbol{\zeta}^* - \boldsymbol{\zeta}) = 0. \quad (3.2.14)$$

Since $\boldsymbol{\zeta} = \mathbf{X}\boldsymbol{\gamma}$,

$$\begin{aligned} \mathbf{X}'\mathbf{V}_\zeta^{-1}(\boldsymbol{\zeta}^* - \mathbf{X}\boldsymbol{\gamma}) &= \mathbf{0} \\ \mathbf{X}'\mathbf{V}_\zeta^{-1}\boldsymbol{\zeta}^* - \mathbf{X}'\mathbf{V}_\zeta^{-1}\mathbf{X}\boldsymbol{\gamma} &= \mathbf{0}. \end{aligned}$$

The maximum likelihood estimate (MLE) $\hat{\boldsymbol{\gamma}}$ is

$$\hat{\boldsymbol{\gamma}} = (\mathbf{X}'\mathbf{V}_\zeta^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_\zeta^{-1}\boldsymbol{\zeta}^*, \quad (3.2.15)$$

which is generalized least squares estimate on $\boldsymbol{\zeta}^*$. For unknown $\boldsymbol{\zeta}^*$, we pursue an iterative scheme based on $\tilde{\zeta}_i = \hat{\zeta}_i + (Y_i - \hat{\mu}_i) \frac{\partial\zeta_i}{\partial\mu_i}$. Using iteratively reweighted least squares with the Newton-Raphson method (see McCullagh and Nelder [62]), we find the solution from

$$\hat{\boldsymbol{\gamma}} = (\mathbf{X}'\mathbf{V}_\zeta^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_\zeta^{-1}\tilde{\boldsymbol{\zeta}}. \quad (3.2.16)$$

If model assumptions, including the choice of link function, are correct, $\hat{\gamma}$ is asymptotically unbiased estimator of γ and information matrix $I(\gamma)$ given by the variance of the score is

$$\begin{aligned} I(\hat{\gamma}) &= \text{Var} \{a(\phi)[X' \Delta(Y - \mu)]\} \\ &= [a(\phi)]^2 (X' \Delta V \Delta X), \end{aligned}$$

where $V = \text{diag}\{\text{Var}(Y_i)\}$. Thus the asymptotic variance-covariance matrix of $\hat{\gamma}$ is given by

$$\text{Var}(\hat{\gamma}) = I^{-1}(\hat{\gamma}) = \frac{[X' \Delta V \Delta X]^{-1}}{[a(\phi)]^2}. \quad (3.2.17)$$

3.2.2 Maximum Likelihood (ML) for Gamma-Multiplicative Model

Consider a density function for gamma distribution given as

$$\frac{1}{\Gamma(\nu)} \left(\frac{1}{\lambda}\right)^\nu e^{-y/\lambda} y^{\nu-1}, \quad y \geq 0, \nu > 0, \lambda > 0 \quad (3.2.18)$$

where ν is a scale parameter. When $\nu = 1$, the distribution corresponds to an exponential distribution. The density can be expressed in the form of the exponential family

$$\begin{aligned} \theta &= -\frac{1}{\lambda\nu} = -\frac{1}{\mu}, \\ \mu &= \lambda\nu, \\ \text{Var}_Y &= \frac{\mu^2}{\nu}, \\ a(\phi) &= \nu, \\ c(\theta) &= -\log(-\theta), \\ b(\phi, Y) &= \nu \log \nu - \ln \Gamma(\nu) + (\nu - 1) \log Y. \end{aligned} \quad (3.2.19)$$

Consider the asymptotic covariance matrix using non-canonical link

$$\text{Var}(\hat{\gamma}) = \frac{[X' \Delta V \Delta X]^{-1}}{[a(\phi)]^2}, \quad (3.2.20)$$

If systematic part of the model is multiplicative on the original scale, and thus additive on the log scale, then

$$\zeta = \log \mu = X\gamma,$$

and

$$\theta = -1/\mu = -\exp(-\mathbf{X}\boldsymbol{\gamma}).$$

Consequently

$$\Delta_i = \partial\theta_i/\partial(x_i\boldsymbol{\gamma}) = \exp(-x_i\boldsymbol{\gamma}),$$

and since $\text{Var}_Y = \mu^2/\nu = \exp(2x_i\boldsymbol{\gamma})/\nu$, $\Delta V \Delta$ is reduced to $\text{diag}\{1/\nu, \dots, 1/\nu\}$. From (3.2.20) asymptotic variance of $\boldsymbol{\gamma}$ is

$$\text{Var}(\hat{\boldsymbol{\gamma}}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\frac{1}{\nu} \right), \quad (3.2.21)$$

where $1/\nu$ is the square of coefficient of variation. This result is equivalent to that found in a logarithmic transformation on the response itself.

3.2.3 Residual Modeling with GLM

The additional variation caused by a hidden noise variable can be reduced by solving the dynamic system after modeling the errors (ε_{ijkl}) via control, explicit noise, and signal factors simultaneously. An error from the response model (3.1.7) is written as

$$\varepsilon_{ijkl} = Y_{ijkl} - \psi(C_i, N_j, f(M_k)),$$

where ψ is an unknown true response function. The distribution of squared error is $\varepsilon_{ijkl}^2 \sim \sigma_{ijk}^2 \chi_{l-1}^2$. Since ε_{ijkl} is not an observed residual, the error is approximated by the residual (e_{ijkl}) derived from the fitted response model. The squared residual model can be represented as

$$\zeta\{e_{ijkl}^2\} = \boldsymbol{x}(C_i, N_j, M_k)\boldsymbol{\gamma},$$

where $\boldsymbol{x}(C_i, N_j, M_k)$ represents a predictor vector consisting of the control, explicit noise, and signal factors, $\boldsymbol{\gamma}$ is a coefficient vector, and ζ denotes a link function for the response e_{ijkl}^2 . Note that the element of first column in \boldsymbol{x} is 1. We assume that higher orders of signal functions are negligible, thus only the linear term is included in the residual modeling. Based on the fact that squared residuals are approximately gamma distributed and the variance must remain positive, we consider a gamma-multiplicative model with

$$e_{ijkl}^2 = \exp[\boldsymbol{x}(C_i, N_j, M_k)\boldsymbol{\gamma}]. \quad (3.2.22)$$

The maximum likelihood estimate (MLE) $\hat{\gamma}$ is derived from a log-gamma link function for the mean of squared residuals. Actual implementation of maximum likelihood results in an algorithm based on iteratively reweighted least squares (IRLS). Thus, MLE $\hat{\gamma}$ is obtained by minimizing

$$\sum_i \sum_j \sum_k \sum_l \frac{[e_{ijkl}^2 - \mu(\mathbf{x}_{ijk}\gamma)]^2}{h[\mu(\mathbf{x}_{ijk}\gamma)]},$$

where

$$\mu(\mathbf{x}_{ijk}\gamma) = E[e_{ijkl}^2 | \{C, N, M\}],$$

and

$$h[\mu(\mathbf{x}_{ijk}\gamma)] = Var[e_{ijkl}^2 | \{C, N, M\}].$$

After approximating $\hat{\sigma}_{ijk}^2$ with the squared residuals using a gamma-multiplicative model, the weighted least squares (WLS) method is used to improve the response model. We propose a following iterative procedure:

- Step 1.** Estimate the coefficients of the response model (3.1.7) through ordinary least squares (OLS). Using t -statistics, insignificant effects are pooled, then the residuals are calculated from the fitted model.
- Step 2.** Estimate the variance function $\hat{\sigma}_{ijk}^2$ by modeling the squared residuals derived from Step 1 via the control, explicit noise, and signal factors, using a gamma-log link function. Approximate σ_{ijk}^2 using the fitted residual value \hat{e}_{ijk}^2 . In case of replications, $\sum_{l=1}^L \hat{e}_{ijkl}^2 / L$ is used.
- Step 3.** Re-estimate the coefficients of the response model, applying weighted least squares (WLS) with weights: $w_{ijkl} = 1/\hat{\sigma}_{ijk}^2$.
- Step 4.** Repeat Step 1 - 3 until the effect estimates in (3.1.7) converge. At last, we obtain the final effect estimates in the response model and variance function.

We can optimize the dynamic system based on the effect estimates resulting from this iterative procedure. Hereafter, our procedure for optimizing the dynamic system will be called the *GLM-RM approach*.

3.3 Transformation Method

A common motivation for a transformation is to achieve an appropriately stable variance. When random errors are unlikely follow a normal distribution or variance is a function of the mean, variance-stabilizing transformation is appealing. For a positive valued response, if variance is proportional to the square of the mean (*i.e.*, a constant coefficient of variation), a log-transformation on the response can be an alternative to a gamma-log link function.

It is known that all coefficient estimates are unbiased except intercepts in the log-transformation. In the Taylor series expansion of $\log Y$,

$$E(\log Y) = \log \mu - \frac{(\sigma/\mu)^2}{2},$$

so consequently, the intercept is biased by $(\sigma/\mu)^2/2$.

The (log) transformation approach assumes that a distribution of the response is log-normal. Linear models with constant variance for $\log Y$ and multiplicative model with constant coefficient of variation for Y are closely connected. Suppose ρ^2 (the square of the coefficient of variation) is sufficiently small so that

$$\text{Var}(\log Y) = \rho^2 = \text{Var}_Y/\mu^2.$$

In a linear model for $\log Y$, the variance of parameter estimates $\hat{\gamma}$ is $\rho^2(\mathbf{X}'\mathbf{X})^{-1}$. For the corresponding multiplicative model, the variance estimates of γ is

$$\text{Var}(\gamma) = \rho^2(\mathbf{X}'\mathbf{X})^{-1}.$$

In particular, if \mathbf{X} is an orthogonal design matrix so that parameter estimates are independent, corresponding parameter estimates in the gamma-multiplicative model are independent asymptotically. The property of approximate independence holds for all GLMs whenever the link function is the same as the variance-stabilizing transformation.

When ρ^2 is small enough, it is likely difficult to discriminate between normal linear model for $\log Y$ and gamma-multiplicative model for Y . Atkinson's [4] work confirms this assertion even for ρ^2 as large as 0.6.

Firth [30] compared efficiencies between the gamma model and the lognormal model in the misspecification case. In his conclusion, the gamma model performs slightly better

under reciprocal misspecification. Nevertheless, the difference is unlikely to be a major factor affecting one's choice between the two methods. Other criteria can be considered as important in practice. For example, a least squares analysis for the log-transformed data allows us to use well-developed diagnostic tools, while analysis on the original scale may have advantages for interpretation.

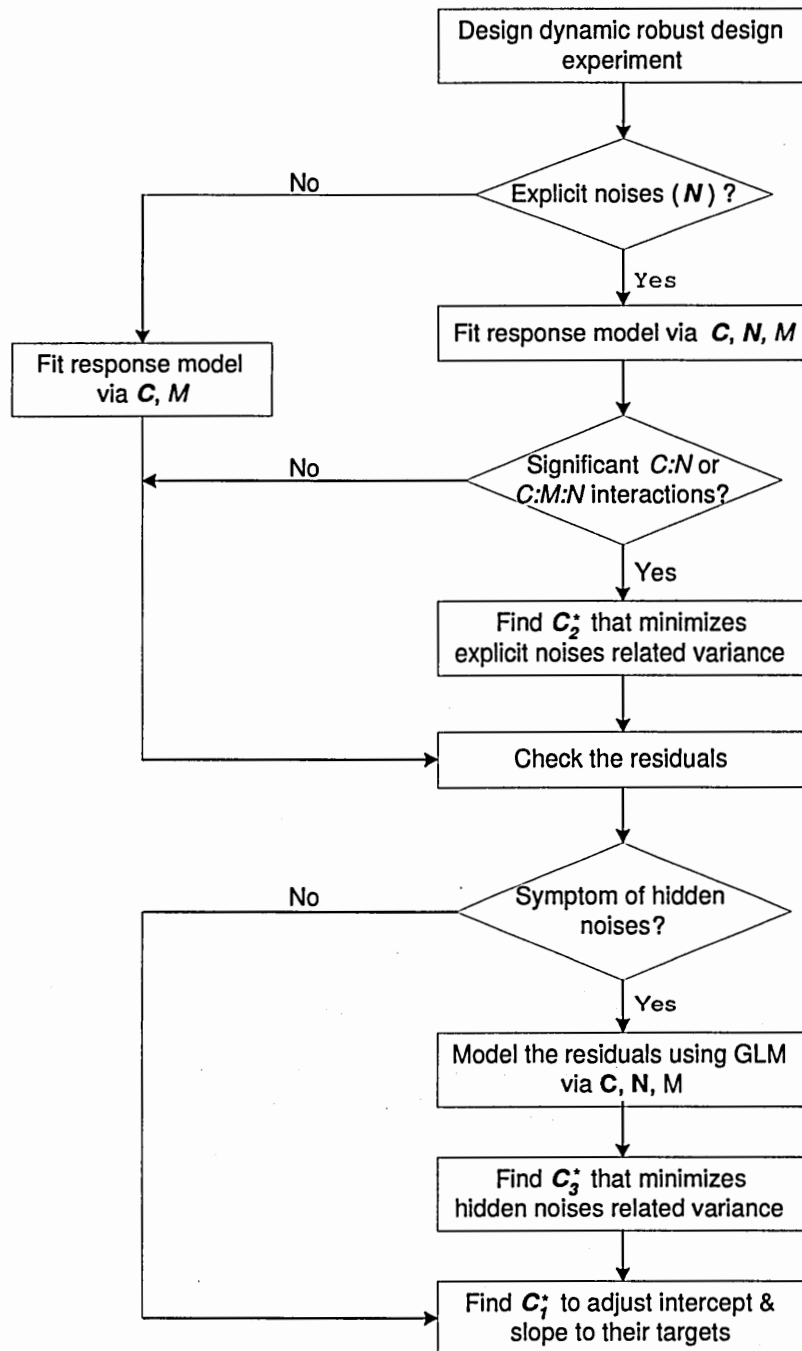


Figure 4: Flow diagram of generalized two-step optimization procedure

CHAPTER IV

DETAILED EXAMPLES

In this chapter, two examples: zinc coating and cable actuator experiments, are re-analyzed following the optimization procedure suggested in chapter II.

4.1 Zinc Phosphate Coating Example

First, the residuals from the fitted response model (2.3.5) were calculated, then the squared residuals were regressed on the control, signal, and control-signal interactions using a gamma-log link function. We used the fitted squared residuals as the weights in weighted least squares (WLS) procedure. We executed the WLS procedure iteratively until the estimated regression coefficients were stabilized. After three iterations, the regression coefficients were stabilized. The final response and GLM variance model with 5% significant level were, respectively

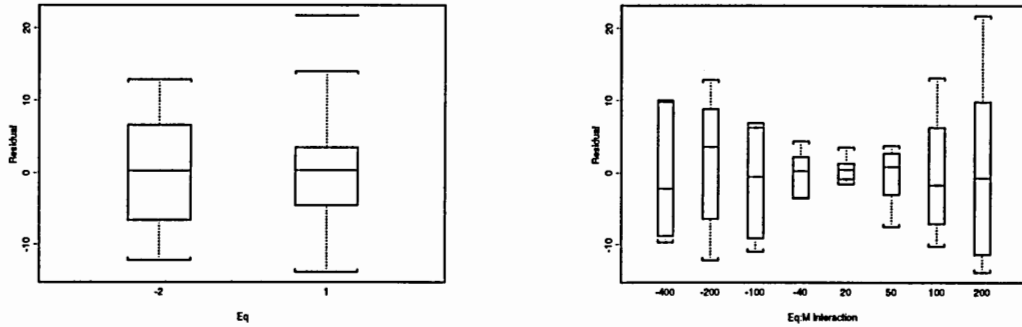
$$\begin{aligned}\hat{Y} = & (0.676 + 0.044B_l + 0.073C_l + 0.208D_l + 0.060E_l - 0.119G_l - 0.113H_l \\ & - 0.057B_q - 0.060E_q - 0.032F_q + 0.037G_q - 0.048H_q)M, \quad (4.1.23)\end{aligned}$$

and

$$\hat{\sigma}_\epsilon^2 = \exp[1.464 + 0.918B_l - 0.913E_q + 0.019M + 0.005E_qM]. \quad (4.1.24)$$

From previous analysis in Section 1.3.1, only E_l was significant for the variance function with the LM approach, but B_l , E_q , M and E_qM interactions were identified as significant in the GLM-RM approach. In the residual plot (Figure 3), the heteroscedasticity is substantial within B and especially M levels, while the variation within E levels is relatively small.

To investigate whether E factor affects the process variation, the residuals of the response model (4.1.23) were plotted with respect to quadratic effect of E (E_q) and E_qM interaction for each (Figure 5). The variance appears nonconstant in both E_q and E_qM levels. As shown in Figure 3, the process variation is largely due to the signal factor, proportional



(a) Eq effect

(b) Eq:Signal interaction effect

Figure 5: Boxplot for the residuals from the response model

to the signal factor's value. Because the E_qM interaction is significant, factor E has a substantial effect on the variance function in the GLM-RM approach.

On the other hand, in the LM approach, variation caused by the signal factor is cancelled out when the response is modeled by the signal factor only for obtaining the separate performance measures at each of control factors' combinations. In addition, when the process variance is dominated by the signal factor, the control factors which interact significantly with the signal factor can only be identifiable by the half normal probability plot or Lenth's method in the LM approach. This illustrates completely the reason why only E_l was identified as significant in the LM approach. We also noted that significant E_qM interaction in the response model (2.3.5) was insignificant in the final response model derived from the GLM-RM approach.

It is known that when the variance is proportional to the square of the mean and the response has a positive value, a log transformation on the response can be an alternative for using the gamma-multiplicative model. Thus, the squared residuals were (log) transformed to achieve an approximately constant variance and linear model was fitted to transformed squared residuals. Consequently, Bl , Eq , and M were identified as dispersion effects. It is noted that the E_qM interaction was not detected in the analysis of the transformed residuals.

At last, we applied the generalized two-step optimization procedure to the response

and variance model estimated from the GLM-RM approach. First, the variance function (4.1.24) should be integrated over M

$$\begin{aligned}\hat{\sigma}_Y^2(\mathbf{C}) &= \int_{M_L}^{M_H} \exp[1.464 + 0.918B_l - 0.913E_q + 0.019M + 0.005E_qM]dM \\ &= \left(\frac{44.701 \cdot \exp[0.087E_q] - 1.462 \cdot \exp[-0.813E_q]}{0.019 + 0.005E_q} \right) \cdot \exp[1.464 + 0.918B_l].\end{aligned}$$

Since no explicit noise factors are given in this experiment, the (explicit) noise related term $\sigma_\epsilon^2(\mathbf{C}_2)$ cannot be estimated. Following the step 1-(b), we consider the choice of control factor values, for $\mathbf{C}_3 = (B_l, E_q)$, the combination of $B = -1$ and $E = 0$ ($E_q = -2$) minimizes the predicted process variance $\hat{\sigma}_Y^2$. In case the target value for the slope target is pre-specified in practice, the values of other control factors $\mathbf{C}_1 = (C, D, F, G, H)$ are selected to shift to its target and remained A level can be chosen by considering an experimental cost. The optimization results from the GLM-RM approach are quite different from the LM approach. Factor E was fitted at level 1 in the LM approach, while $E = 0$ was selected to minimize the process variance in the GLM-RM approach.

4.2 Cable Actuator Example

The cable actuator experiment was re-analyzed with the GLM-RM approach. From the fitted response model (2.3.6), the squared residuals were modeled via control, (explicit) noise, and signal factors (including all second order interactions) with a gamma-log link function. Following iterative procedure Step 1 - 4 in section 2.2.3, the coefficients of all the factors were estimated. After three iterations, the estimates of all the coefficients converged. The final response and variance model with 5% significance level were

$$\begin{aligned}\hat{Y} &= 0.505 - 0.397B + 0.336G + 0.545M \\ &\quad + (0.073B - 0.137G + 0.066H - 0.030J)M - 0.011MN\end{aligned}\quad (4.2.25)$$

and

$$\hat{\sigma}_\epsilon^2 = \exp[-4.584 - 0.308B - 0.300D + 0.176M],\quad (4.2.26)$$

respectively.

Additionally, B , D , and M were identified as significant dispersion effects with 10% significance level in the GLM-RM approach. The OLS estimates for $\ln e^2$ were also obtained and same factors (B , D , and M) were identified as significant. As in the phosphate coating example, the phenomenon of nonconstant variance was substantial within the signal levels in the residual plot (not shown), however the LM and the RFM approach failed to reflect the significant signal factor in the variance model. In contrast, the GLM-RM approach identified the signal factor M as the primary element which influences the process variance.

To perform the generalized two-step optimization procedure, we first integrate the variance model $\hat{\sigma}_Y^2$ over M . Assuming that the explicit noise factor has an unit variance,

$$\begin{aligned}\hat{\sigma}_Y^2(C) &= \int_{M_L}^{M_H} \{(-0.011)^2 M^2 + \exp[-4.584 - 0.308B - 0.300D + 0.176M]\} dM \\ &\approx \exp[-0.308B - 0.300D].\end{aligned}$$

Since only noise-by-signal interaction is significant in the response model, we cannot identify C_2 in the explicit noise related variance σ_e^2 . To illustrate the generalized two-step optimization procedure in detail, suppose that DN interaction with a positive coefficient β_{DN} is included additionally in the response model. Then,

$$\hat{\sigma}_Y^2(C) \approx -\beta_{DN}^* D + \beta_{DN}^2 D^2 + \exp[-0.308B - 0.300D],$$

where $\beta_{DN}^* = (\beta_{DN} \times 0.011) \int_M M dM$. Obviously, $\sigma_e^2(C_2) = -\beta_{DN}^* D + \beta_{DN}^2 D^2$ and $\sigma_e^2(C_2, C_3) = \exp[-0.308B - 0.300D]$.

In the first step, $D (= C_2)$ is fixed at level "1" to minimize σ_e^2 . At $D = 1$, the remaining $B (= C_3)$ in $\hat{\sigma}_Y^2$ is chosen at level "1" to minimize the hidden noise related variance σ_e^2 . Next, to shift both slope and intercept to their target, we use G , which is significant for the intercept but insignificant for the process variance, to adjust the intercept to its target value of zero after minimizing the process variance. At $B = 1$, it follows from (4.2.25) that $G = -0.321$ shift the intercept to zero. At last, we choose the levels of H and J to maximize the input-output linearity. $H = 1$ and $J = -1$ maximize the slope to 0.758. Here, $C_1 = (G, H, J)$.

We can reduce the process variance substantially by identifying the effects of hidden noise variables as well as explicit noise factors.

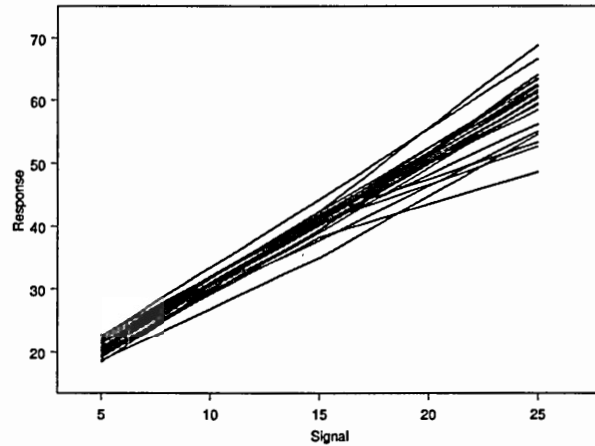


Figure 6: Simulated dynamic system with the multiplicative error: $\beta_1 = 1.0, \beta_2 = 0.4$

4.3 Simulation Results

In the previous two examples, we claimed that when the heteroscedasticity is overwhelming within the signal levels, the LM and the RFM approach identify only control factors that interact significantly with the signal factor as dispersion effects. However, the two approaches might not identify relatively small main control effects as significant for the variance, if ever really significant.

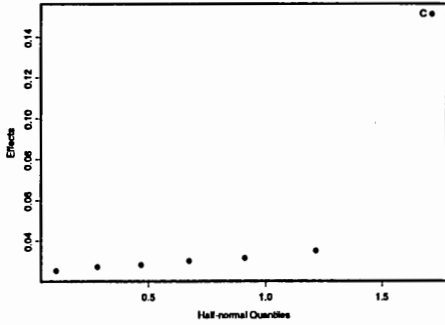
To verify those claims, we simulated a dynamic system with three control factors (A , B and C) and a signal factor of three input levels (5, 15 and 25). A 2^3 factorial design was first used for the control factors and the response data were generated by the equation

$$Y = 10 + 0.8A + 2M + \epsilon,$$

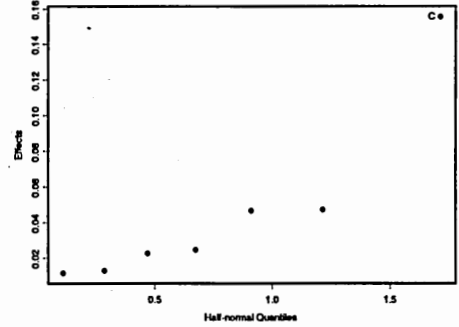
where the error ϵ was generated from the normal distribution with mean 0 and variance $\sigma_\epsilon^2 \propto \exp[\beta_1 B + \beta_2 CM]$. The simulated response was plotted versus the signal in Figure 6.

The main purpose of the simulation is to assess how the LM approach identify the true significant dispersion effects accurately in the dynamic system with a multiplicative error. Following the purpose, we changed β_1 as 0.1, 0.5, 1.0 and β_2 as 0.1, 0.2 and 0.4 to evaluate the power of detecting the real significant effects in variance function.

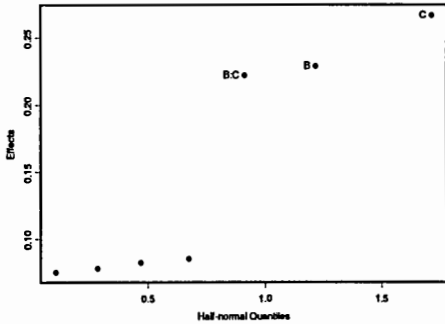
First, we fixed the value of β_2 at 0.1, changed β_1 , then analyzed the simulated dynamic



(a) $\beta_1 = 0.1$



(a) $\beta_1 = 0.5$



(c) $\beta_1 = 1.0$

Figure 7: Half normal probability plot for 2^3 factorial design

system using the LM approach. In this simulation, we limited our interest to the variance function, hence the results for the intercept and the slope function are not shown here. For the 2^3 full factorial design, only C was identified as the significant variance effect when $\beta_1 = 0.1$ and 0.5 . At $\beta_1 = 1.0$, B , C and BC were significant for the variance. Note that the spurious effect BC was significant at $\beta_1 = 1.0$ in LM approach. (see the half normal probability plot in Figure 7). For $\beta_2 = 0.2$ and 0.4 , only effect C was significant when $\beta_1 = 0.1$, and B and BC were significant when $\beta_1 = 0.5$ and 1.0 in the half normal plot (not shown here).

Next, we considered a 3^3 full factorial design for the control factors to confirm the simulation results from 2^3 factorial design. The simulations were restricted to the case with small value of β_2 , *i.e.*, $\beta_2 = 0.1$ to show whether the LM approach identify a significant

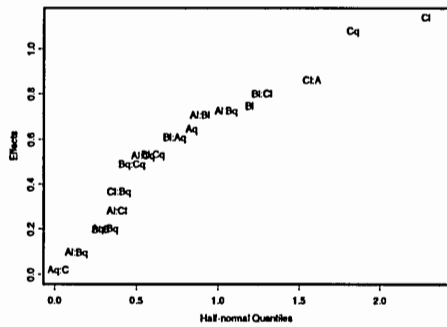
main effect in the variance function. The responses were generated by the equation

$$Y = 10 + A_l + 0.5B_q + 2M + \epsilon, \quad (4.3.27)$$

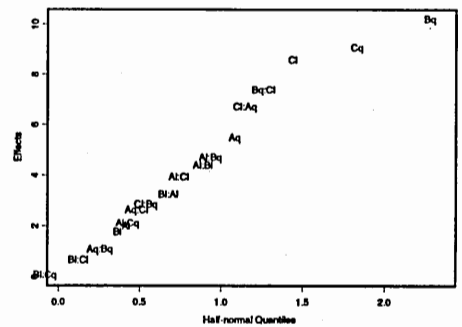
with the same signal inputs as 2^3 factorial design. The assumption for the error was the same except how the error variance $\sigma_\epsilon^2 \propto \exp[\beta_1 B_l + \beta_2 C_q M]$. We varied β_1 as 0.1, 0.5, and 1.0 while fixing β_2 at 0.1.

Similarly, we analyzed the simulated data using the LM approach. In Figure 7, C_l and C_q seem to be significant, but B_l was not significant at $\beta_1 = 0.1$. At $\beta_1 = 0.5$ and 1.0, B_q and C_q effects appeared significantly large in the half normal plot. We repeated the simulation at a number of different values for β_1 and β_2 . When β_1 is relatively small, the main effect is unlikely to be detected in the LM approach since it is dominated by relatively large control-by-signal interactions. However, the control factor that interacts with the signal was identified consistently as the significant variance effect in the LM approach. Furthermore, the LM approach sometimes detects a spurious effect.

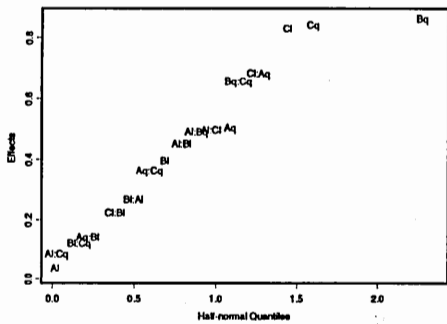
We can easily extend these results to the RFM approach by considering explicit noise factor as an another control factor. With the GLM-RM approach, true dispersion effects were identified consistently by taking the iterative estimation procedure. In addition, the GLM-RM approach provided valuable information about how the control factors interact with the other control and signal factors.



(a) $\beta_1 = 0.1$



(a) $\beta_1 = 0.5$



(c) $\beta_1 = 1.0$

Figure 8: Half normal probability plot for 3^3 factorial design

CHAPTER V

CONCLUDING REMARKS

For the dynamic robust design problem, the RM approach allows greater flexibility to study factor effects and results in experimental cost savings. This article illustrates how we extend the response model approach to more general cases: there are no explicit noise factors and there are no significant control-by-noise interactions. It was also noted that GLM modeling of the residuals from the response model reveals the potential for reducing the variance caused by hidden noise variables in the RM approach. As shown in the examples, both the LM and the RFM approaches might detect spurious dispersion effects because the signal cannot be included explicitly in the variance modeling. We also discussed the efficiency of the LM and RFM approaches using the simulated data. In conclusion, we suggested the GLM-RM approach and generalized two-step optimization procedure as the tool for providing more reliable results.

However, the ability to identify the dispersion effects greatly depends on the fitted response model in GLM-RM approach and if the response model is handled in a sloppy fashion, it can cause serious problem in detecting the dispersion effects. We should, then, check outliers and model inadequacies before implementing a GLM for the residuals. In general, the ideal relationship between the signal and the response is assumed to be known by the engineers in advance. In the case that ideal relationship is of unknown nonlinear form, the engineer needs to identify the ideal relationship and then model the dynamic system via the nonlinear random coefficients model (NRCM). Wolfinger and Tobias [96] considered the linear random coefficient model (LRCM) for quality improvement experiments, but NRCM application to the dynamic robust design is not abundant at all.

PART II

Modeling Approach for Degradation Testing

CHAPTER VI

DEGRADATION TESTING OF LIGHT DISPLAYS

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 covariate 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.* [56] found that, in terms of asymptotic efficiency, degradation models are superior if degradation data are available.

Along with degradation modeling, additional reliability information can be gained by testing units in a harsh environment where product aging is hastened so that degradation is accelerated and product lifetimes decrease, relative to the normal-use environment. 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.

Degradation modeling has a rich history in the electronics manufacturing, materials testing and various fields of engineering. Nelson ([70], chapter 11) and Meeker and Escobar [63]

review the degradation literature, survey applications and describe basic analytical methods on ADT models. Meeker and Escobar [64] provide a practical guide for ADT modeling based on transformed linear degradation models along with standard ALT formulas.

There are numerous model developments for more specific degradation test experiments. Lu, *et al.* [57] derive a transformed linear model for hot-carrier-induced degradation of metallized and oxidized semiconductors. By considering the sample size from degradation paths as random, Su *et al.* [85] develop a random coefficient model and used the maximum likelihood estimation (MLE) method to handle inconsistency problems found in least squares estimation (LSE). Shiau and Lin [80] derive a nonparametric model for describing degradation of LEDs. Tseng and Wen [92] propose a step-stress accelerated degradation test method to reduce experimental costs in assessing the lifetime distribution of LEDs. Yu and Tseng [98] suggest on-line procedure for determining an appropriate termination time for an ADT. Fagerstorm [27] 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, [31]), Fluorescent Lamps (Tseng *et al.*, [87]), computer disks and compact disks (Murray, [66]), and power supplies (Chang, [14]).

6.1 Burn-in Characteristics of Light Displays

In the manufacture of light display devices, chemical processing produces impurities inside the device's vacuum tube. Impurities remain throughout the manufacturing process, degrading the display quality. Manufacturers burn off harmful impurities before shipping them to customers, hence the burn-in procedure is essential in the manufacturing process. From the manufacturer's viewpoint, the burn-in costs (including the operating cost of burn-in equipment), marketing losses caused by delivery delay, along with the costs associated with testing and failure monitoring during burn-in are crucial. Furthermore, the current trend of shortened product life-cycles has driven manufacturers to reduce the burn-in time during manufacturing. However, if the burn-in period is incomplete, the light display will initially experience a rapid decrease in light intensity until the impurities are burned out

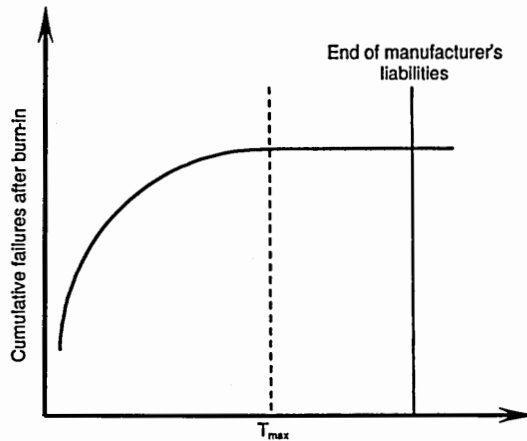


Figure 9: Burn-in characteristics of failure data

completely, at which time the light degradation will continue at a slower, more expected rate.

It is known that the display luminosity for several types of VFDs, fluorescent lamps and plasma display panels (PDPs) decreases exponentially over most of the usage period so the degradation path can be expressed as

$$\ln \Lambda(t) = \phi - \lambda t \quad (6.1.28)$$

where $\Lambda(t)$ denotes the luminosity at time t , ϕ is the initial luminosity, and $\lambda > 0$ is the rate of degradation. Naturally, this model does not hold for the initial time period in incomplete burn-in, when the degradation is unstable.

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 fit the model (6.1.28) to the degradation path with an initial wear-in, one might ignore the initial wear-in period, then derive a failure-time model based on the truncated data.

The general burn-in characteristic is illustrated by the failure data in Figure 9. T_{max} denotes the maximum burn-in time period showing the levelling-off of early failures. Compared to the failure frequency before T_{max} , the number of test period failures is small enough to be ignored.

For the purpose of customer warranty, the degradation testing in the field lasts at most 1,000 hours. Reliability engineers infer the device's lifetime largely by extrapolating the time when the degradation path and a predefined threshold cross. However, in the case of incomplete burn-in, the degradation path decreases rapidly at first, causing the extrapolated degradation curve to cross the failure threshold at a much earlier time, resulting in a predicted lifetime that is much smaller than the actual expected device failure time. For the guidelines of a burn-in procedure, see Jensen and Peterson [42].

6.2 Motivating Example

This research was motivated by degradation data on VFDs. 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. Electrons emitted from a cathode (filament) are accelerated by a grid, and collide with a phosphor coated surface of an anode electrode, producing luminescence.

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. The critical characteristic of VFD is the luminosity or brightness, which is measured in candela and failure can be defined in terms of the amount of degradation in the luminosity.

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 harmful impurities before shipping them to customers, hence the burn-in procedure is essential in the manufacturing process. However, owing to incomplete burn-in (called "aging" in the industry) in the manufacturing process, 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 (type I and II in Figure 10).

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

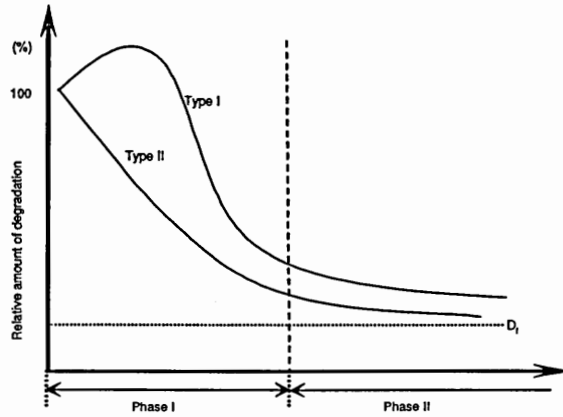


Figure 10: The non-monotonic degradation path caused by incomplete burn-in

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 before following its inherent degradation path (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 up to an unknown point in time before the degradation becomes more gradual. Eventually, after several hundred of hours of operation, the cathode and phosphor on 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 11 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 non-monotonic 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.* [87] and Chiao and Hamada [16] truncate the first several hundred hours assuming that after this much test time has been completed, the degradation path will be monotonic. However,

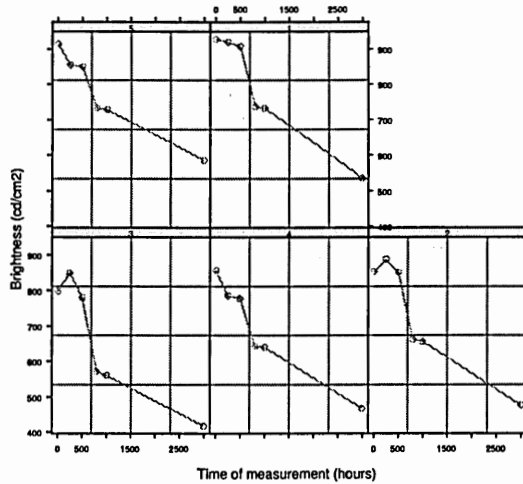


Figure 11: Nonmonotonic degradation paths for VFDs

this assumption is not suitable for the light displays. Monotonic model fitting to truncated data ignores the effects mentioned above that can cause one or two dramatic changes in the degradation curve, including the loss of monotonicity.

In addition, owing to shorter life-cycle of VFD, reliability is often evaluated based on a relatively small amount of degradation data measured for a short period of time. These early degradation measurements can comprise a significant amount of the data in VFD degradation testing. As a consequence, we risk losing efficiency in the reliability analysis if these early degradation observations are ignored. The characteristic of non-monotonicity and the sparseness of measurements for VFD degradation provides motivation for an improved degradation model in the following chapter.

CHAPTER VII

NONLINEAR RANDOM COEFFICIENTS MODEL

Nonlinear random coefficient 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, [20]). 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.

Nonlinear random coefficient 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 nonlinear random coefficient models have occurred in recent years; see, for example, Beal and Sheiner ([5], [7], [8]), Lindstrom and Bates [54], Ramos and Pantula [76] and Kim [47]. Davidian and Giltinan [20] and Vonesh and Chinchilli [93] provide thorough overviews as well as some general theoretical developments for nonlinear random coefficient models.

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

$$y_{ij} = f(\mathbf{x}_{ij}, \boldsymbol{\beta}_i) + e_{ij} \quad i = 1, \dots, m, \quad j = 1, \dots, n_i, \quad (7.0.29)$$

where y_{ij} is the j^{th} response on the i^{th} individual, \mathbf{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 \mathbf{Y}_i and \mathbf{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 $\mathbf{f}_i(\boldsymbol{\beta}_i) = (f(\mathbf{x}_{i1}, \boldsymbol{\beta}_i), \dots, f(\mathbf{x}_{in_i}, \boldsymbol{\beta}_i))'$ for the i^{th} individual test

item, depending on the $(p \times 1)$ individual-specific regression parameter β_i . Suppose that $E(\mathbf{e}_i|\beta_i) = \mathbf{0}$ and $Cov(\mathbf{e}_i|\beta_i) = \sigma^2 \mathbf{G}_i^{1/2}(\beta_i) \mathbf{G}_i^{1/2}(\beta_i) = \sigma^2 \mathbf{G}_i(\beta_i)$, where $\mathbf{G}_i^{1/2}(\beta_i)$ denotes a $(n_i \times n_i)$ positive definite matrix which depends on i only through its dimension. The elements are equal to the square root of those of $\mathbf{G}_i(\beta_i)$, characterizing within-individual variance structure. The $(k \times 1)$ vector \mathbf{b}_i represents a vector of random effects, and β is $(r \times 1)$ vector of fixed effects.

Based on the set-up in Sheiner and Beal [79] and Davidian and Giltinan [20], the general two-stage nonlinear random coefficient model can be written

Stage 1 (*within-individual variation*)

$$\mathbf{y}_i = \mathbf{f}_i(\beta_i) + \mathbf{e}_i \quad (7.0.30)$$

Stage 2 (*between-individual variation*)

$$\beta_i = \mathbf{A}_i \beta + \mathbf{B}_i \mathbf{b}_i \quad (7.0.31)$$

where \mathbf{b}_i are independent and identically distributed as $\text{Normal}(\mathbf{0}, \mathbf{D})$, and $\mathbf{A}_i, \mathbf{B}_i$ are known design matrices of size $(p \times r)$ and $(p \times k)$ for the fixed and random effects, respectively. \mathbf{D} is a $(k \times k)$ covariance matrix and β_i in (7.0.31) is specific to the i^{th} test item through \mathbf{b}_i , allowing the conditional moments of \mathbf{e}_i to be expressed as $\mathbf{e}_i|\mathbf{b}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{G}_i(\beta_i))$.

Within the framework of (7.0.30) and (7.0.31), the random effects are unobserved quantities, and maximum likelihood estimation in (7.0.30) is based on the marginal density of $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_m)'$

$$p(\mathbf{y}|\beta, \mathbf{D}, \sigma^2) = \int p(\mathbf{y}|\mathbf{b}, \beta, \mathbf{D}, \sigma^2) p(\mathbf{b}) d\mathbf{b}. \quad (7.0.32)$$

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

7.1 Approximation Methods

In this section, we consider four different approximations to the log-likelihood corresponding to (7.0.32): a first-order method (Beal and Sheiner, [5], [6]), Lindstrom and Bates's [54]

algorithm, adaptive importance sampling (Pinheiro and Bates, [72]), and adaptive Gaussian quadrature (Pinheiro and Bates, [72]). For two-stage nonlinear models, the marginal distribution of \mathbf{y}_i is based on the random effects \mathbf{b}_i and the individual error vectors \mathbf{e}_i . The random vectors \mathbf{b}_i and \mathbf{e}_i are assumed to be independent, normally distributed and enter the model in an additive, linear fashion. Suppose that

$$\mathbf{e}_i = \mathbf{G}_i^{1/2}(\boldsymbol{\beta}_i)\boldsymbol{\epsilon}_i, \quad (7.1.33)$$

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

$$\mathbf{y}_i = \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{b}_i) + \mathbf{G}_i^{1/2}(\boldsymbol{\beta}, \mathbf{b}_i)\boldsymbol{\epsilon}_i. \quad (7.1.34)$$

In many practical applications involving inference on the degradation of individuals, within-individual heteroscedasticity and correlation are dominated by between-individual variations \mathbf{b}_i , thus can be ignored. $\mathbf{G}_i(\boldsymbol{\beta}_i) = \mathbf{I}_{n_i}$ 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” \mathbf{y}_i response. This issue has been discussed by Zeger, *et al.* [99], 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 [54]).

7.1.1 First-Order Approximation

The Taylor series expansion of (7.1.34) about $E(\mathbf{b}_i) = \mathbf{0}$ is

$$\mathbf{y}_i \approx \mathbf{f}_i(\boldsymbol{\beta}, \mathbf{0}) + \boldsymbol{\Lambda}_i(\boldsymbol{\beta}, \mathbf{0})\boldsymbol{\Delta}_{\mathbf{b}_i}(\boldsymbol{\beta}, \mathbf{0})\mathbf{b}_i + \mathbf{G}_i^{1/2}(\boldsymbol{\beta}, \mathbf{0})\boldsymbol{\epsilon}_i \quad (7.1.35)$$

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

Defining the $(n_i \times k)$ matrix $Z_i(\beta, \mathbf{0}) = \Lambda_i(\beta, \mathbf{0})\Delta_{b_i}(\beta, \mathbf{0})$ and $\mathbf{e}_i^* = \mathbf{G}_i^{1/2}(\beta, \mathbf{0})\boldsymbol{\epsilon}_i$, equation (7.1.35) may be shortened to

$$\mathbf{y}_i \approx \mathbf{f}_i(\beta, \mathbf{0}) + Z_i(\beta, \mathbf{0})\mathbf{b}_i + \mathbf{e}_i^*. \quad (7.1.36)$$

In terms of (7.1.36), the marginal mean and covariance of \mathbf{y}_i can be specified as

$$\begin{aligned} E(\mathbf{y}_i) &\approx \mathbf{f}_i(\beta, \mathbf{0}) \\ Cov(\mathbf{y}_i) &\approx \sigma^2 \mathbf{G}_i(\beta, \mathbf{0}) + Z_i(\beta, \mathbf{0})\mathbf{D}Z_i'(\beta, \mathbf{0}) \\ &\equiv \boldsymbol{\Sigma}_i(\beta, \mathbf{0}, \boldsymbol{\omega}) \end{aligned} \quad (7.1.37)$$

where $\boldsymbol{\omega}$ is a vector containing the unknown covariance parameters σ^2 and \mathbf{D} .

Beal and Sheiner ([5], [6]) utilize extended least squares (ELS) to estimate β and $\boldsymbol{\omega}$. Under normality assumptions, ELS is equivalent to joint maximum likelihood estimation of $(\beta, \boldsymbol{\omega})$, which is known as the *first-order method* and minimizes the objective function:

$$Q_{FO}(\beta, \boldsymbol{\omega}) = \sum_{i=1}^m \{ \log |\boldsymbol{\Sigma}_i(\beta, \mathbf{0}, \boldsymbol{\omega})| + (\mathbf{y}_i - \mathbf{f}_i(\beta, \mathbf{0}))' \boldsymbol{\Sigma}_i^{-1}(\beta, \mathbf{0}, \boldsymbol{\omega}) (\mathbf{y}_i - \mathbf{f}_i(\beta, \mathbf{0})) \} \quad (7.1.38)$$

They note that ELS estimators are consistent and asymptotically normal under some regularity conditions provided that the first and second moments of \mathbf{y}_i are correctly specified. To minimize (7.1.38), Beal and Sheiner [5] employ a derivative-free quasi-Newton algorithm which avoids the need for specifying $\partial \boldsymbol{\Sigma}_i(\beta, \mathbf{0}, \boldsymbol{\omega}) / \partial \beta$.

7.1.2 Lindstrom-Bates Algorithm

Lindstrom and Bates [54] suggest the alternative to first-order method by considering a linearization of (7.1.34) in the random effects about some value \mathbf{b}_i^* that is closer to \mathbf{b}_i than its expectation $\mathbf{0}$. The Taylor series expansion about $\mathbf{b}_i = \mathbf{b}_i^*$ yields

$$\mathbf{y}_i \approx \mathbf{f}_i(\beta, \mathbf{b}_i^*) + \Lambda_i(\beta, \mathbf{b}_i^*)\Delta_{b_i}(\beta, \mathbf{b}_i^*)(\mathbf{b}_i - \mathbf{b}_i^*) + \mathbf{G}_i^{1/2}(\beta, \mathbf{b}_i^*)\boldsymbol{\epsilon}_i. \quad (7.1.39)$$

where $\Lambda_i(\cdot)$, $\Delta_{b_i}(\cdot)$ were defined in (7.1.35). This notation makes explicit the fact that the expansion in \mathbf{b}_i has been taken about \mathbf{b}_i^* .

In order to construct an estimation procedure for β and $\boldsymbol{\omega}$, reasonable values of \mathbf{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 β and ω . Lindstrom and Bates [54] propose the following iterated two-step estimation algorithm:

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

$$\sum_{i=1}^m \{ \log |\hat{D}| + \mathbf{b}'_i \hat{D}^{-1} \mathbf{b}_i + \log |\hat{\sigma}^2 \mathbf{G}_i(\hat{\xi})| + \hat{\sigma}^{-2} (\mathbf{y}_i - \mathbf{f}_i(\beta, \mathbf{b}_i))' \mathbf{G}_i^{-1}(\hat{\xi}) (\mathbf{y}_i - \mathbf{f}_i(\beta, \mathbf{b}_i)) \}. \quad (7.1.40)$$

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

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

$$Q_{LB}(\beta, \omega) = \sum_{i=1}^m \{ |\log \Sigma_i(\hat{\beta}_0, \hat{\mathbf{b}}_i, \omega)| + \mathbf{r}'_i \Sigma_i^{-1}(\hat{\beta}_0, \hat{\mathbf{b}}_i, \omega) \mathbf{r}_i \} \quad (7.1.41)$$

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

As an alternative to (7.1.41), one can use a restricted maximum likelihood (REML) approach which minimizes

$$Q_{LB,REML}(\beta, \omega) = Q_{LB}(\beta, \omega) + \log |\mathbf{X}'_i(\hat{\beta}_0, \hat{\mathbf{b}}_i) \Sigma_i^{-1}(\hat{\beta}_0, \hat{\mathbf{b}}_i, \omega) \mathbf{X}_i(\hat{\beta}_0, \hat{\mathbf{b}}_i)|, \quad (7.1.42)$$

where $\mathbf{X}_i(\beta_0, \hat{\mathbf{b}}_i) = \Lambda_i(\beta_0, \hat{\mathbf{b}}_i) \Delta_{\beta_i}(\beta_0, \hat{\mathbf{b}}_i)$, and $\Delta_{\beta_i}(\beta, \hat{\mathbf{b}}_i)$ is the $(p \times r)$ matrix of derivatives of β_i with respect to β at $\mathbf{b}_i = \hat{\mathbf{b}}_i$.

The algorithm alternates between the PD and LME steps until some convergence criterion is met, with final estimates denoted by $(\hat{\beta}_{LB}, \hat{\omega}_{LB}, \hat{\mathbf{b}}_{i, LB})$. The $\hat{\mathbf{b}}_{i, LB}$ may be regarded as empirical Bayes estimates of the random effects; these estimates are approximate posterior modes for \mathbf{b}_i , where the fixed parameters β and ω are replaced by estimates. Wolfinger [95] also showed that LME approximation to the restricted log-likelihood (7.1.42) is equivalent to a Laplacian approximation to the integral (7.0.32) when a flat prior is assumed for β .

7.1.3 Adaptive Importance Sampling Approximation

Importance sampling is a simple and efficient method for performing Monte Carlo integration. Geweke [33] 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. The marginal density should be integrated to obtain this distribution, but instead an easily sampled approximation is used. From (7.0.32), the marginal density of \mathbf{y}_i can be written as

$$p(\mathbf{y}_i|\beta, \mathbf{D}, \sigma^2) = \int (2\pi\sigma^2)^{-n_i/2} |\mathbf{D}|^{-1/2} \exp(-v_i(\beta, \mathbf{b}_i, \mathbf{D})/2) d\mathbf{b}_i, \quad \text{where}$$

$$v_i(\beta, \mathbf{b}_i, \mathbf{D}) = \mathbf{b}_i' \mathbf{D}^{-1} \mathbf{b}_i + \sigma^{-2} (\mathbf{y}_i - \mathbf{f}_i(\beta, \mathbf{b}_i))' \mathbf{G}_i^{-1}(\gamma) (\mathbf{y}_i - \mathbf{f}_i(\beta, \mathbf{b}_i)).$$

For the nonlinear random coefficients model, the integrand of $\exp(-v_i(\beta, \mathbf{D}, \mathbf{b}_i)/2)$ 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 \mathbf{z}^* and calculating the sample of random effects as $\mathbf{b}_i^* = \hat{\mathbf{b}}_i^* + \mathbf{V}_i(\beta, \mathbf{D})^{-1/2} \mathbf{z}^*$, where $\hat{\mathbf{b}}_i^*$ is the value of \mathbf{b}_i minimizing $v_i(\beta, \mathbf{b}_i, \mathbf{D})$ and $\mathbf{V}_i(\beta, \mathbf{D})$ is the approximation for the second derivatives of $v_i(\beta, \mathbf{b}_i, \mathbf{D})$ with respect to $\hat{\mathbf{b}}_i^*$. If we let $\mathbf{V}_i(\beta, \mathbf{D})^{-1/2}$ denote the inverse of the Cholesky factor of $\mathbf{V}_i(\beta, \mathbf{D})$, the importance sampling approximation to the log-likelihood of \mathbf{y} can be written as

$$Q_{IS}(\beta, \mathbf{D}, \sigma^2|\mathbf{y}) = -\frac{1}{2} \left(N \log(2\pi\sigma^2) + m \log |\mathbf{D}| + \sum_{i=1}^m \log |\mathbf{V}_i(\beta, \mathbf{D})| \right)$$

$$+ \sum_{i=1}^m \log \left(\sum_{j=1}^{N_{IS}} \exp\{-v_i(\beta, \mathbf{D}, \mathbf{b}_{ij}^*)/2 + \|\mathbf{z}_j^*\|^2/2\} / N_{IS} \right),$$

where $N = \sum_{i=1}^m n_i$. In the case the model function is linear in \mathbf{b} , the right side of (7.0.32) is $p(\mathbf{y}_i|\mathbf{b}_i, \beta, \mathbf{D}, \sigma^2)p(\mathbf{b}_i) = p(\mathbf{y}_i|\beta, \mathbf{D}, \sigma^2)\phi(\hat{\mathbf{b}}_i, \mathbf{V}_i(\beta, \mathbf{D})^{-1})$ where ϕ is the standard Gaussian density. In this case, the results are exact and the importance sampling weights are equal to $p(\mathbf{y}_i|\beta, \mathbf{D}, \sigma^2)$.

7.1.4 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. Gaussian quadrature rules for multiple integrals can be numerically complex (Davis and Rabinowitz, [21]), but due to the simple structure of the integrand in the nonlinear mixed-effects model, we can transform the problem into successive applications of simple one-dimensional Gaussian quadrature rules. We consider an importance sample version of the Gaussian quadrature rule, which we denote by *adaptive Gaussian quadrature*.

For $j = 1, \dots, N_{GQ}$, 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, [34]; Abramowitz and Stegun, [1]). The adaptive Gaussian quadrature is then given by

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

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

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

When $N_{GQ} = 1$, adaptive Gaussian quadrature approximation is equivalent to Laplacian approximation, because in this case $z_1 = 0$ and $w_1 = 1$. This result 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 \mathbf{b} . In practice $N_{GQ} \leq 7$ generally suffices and $N_{GQ} = 1$ often provides a reasonable approximation (Pinheiro and Bates, [72]).

7.2 Failure-Time Distribution of Random Coefficients Model

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

$$F_T(t) = Pr(T \leq t) = Pr[\tau(t; \beta, \mathbf{b}) \leq \tau_f]. \quad (7.2.43)$$

The failure-time distribution depends on the distribution of the random coefficient \mathbf{b} , which is determined by the variance-covariance matrix \mathbf{D} .

In the following, 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.

7.2.1 Linear Random Coefficients Model

In many cases, the LRC model reveals a closed-form expression $F_T(t)$ and the computations are straightforward. Tseng *et al.* [87] and Chiao and Hamada [16] modeled (transformed) luminosity degradation for fluorescent lamps and LEDs using a linear random coefficient model after truncating the first few hundred hours on the test when the degradation is unstable. Following that approach, the failure-time distribution as to the nonmonotonic degradation paths is derived from the LRC model.

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, $i = 1, \dots, m$, $j = 1, \dots, n_i$, is

$$\tau(t_{ij}; \beta, \mathbf{b}_i) = \log(y(t_{ij})) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})(t_{ij} - t_0) + e_{ij}, \quad (7.2.44)$$

with fixed effect $\beta = (\beta_0, \beta_1)$ and the random coefficient $\mathbf{b}_i = (b_{0i}, b_{1i})$ characterizes the individual item variation. We assume $\mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D})$ with elements $D_{00} = \sigma_0^2$, $D_{11} = \sigma_1^2$,

and $D_{01} = D_{10} = \rho\sigma_0\sigma_1$, and b_i is independent of the error term $e_{ij} \sim \mathcal{N}(0, \sigma^2)$. The failure-time distribution for model (7.2.44) is

$$\begin{aligned} F_T(t) &= Pr[(\beta_0 + b_{0i}) + (\beta_1 + b_{1i})(t - t_0) \leq \tau_f] \\ &\approx \Phi \left\{ \frac{[\beta_0 + \beta_1(t - t_0)] - \tau_f}{[\sigma_0^2 + (t - t_0)^2\sigma_1^2 + 2(t - t_0)\rho\sigma_0\sigma_1]^{1/2}} \right\} \quad t > t_0, \end{aligned} \quad (7.2.45)$$

where $\Phi\{\cdot\}$ is the cdf of the standard normal distribution. 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

$$z_p = \frac{(\beta_0 + \beta_1 t'_p) - \tau_f}{[\sigma_0^2 + t_p'^2\sigma_1^2 + 2t'_p\rho\sigma_0\sigma_1]^{1/2}}, \quad (7.2.46)$$

where $t'_p = t_p - t_0$ and z_p is the p^{th} quantile of the standard normal distribution.

Example 1: A degradation path model for the relative luminosity for a VFD. The actual degradation path of (log-transformed) relative luminosity for a VFD is a special case of (6.1.28), given by

$$\mathcal{D}(t; \Theta) = \phi - \vartheta t, \quad \vartheta > 0, t > 0,$$

where the parameter ϕ represents the fixed initial amount of degradation at the beginning of the test and the degradation rate ϑ varies from unit to unit. By industry standard, the VFD is considered to fail when the relative luminosity crosses $\mathcal{D}_f = \log 0.5$. Suppose that ϑ follows a Weibull distribution with cumulative distribution function (cdf) $G(\vartheta) = 1 - \exp[-(\vartheta/\eta)^\zeta]$, where $\eta > 0$ is a scale parameter and $\zeta > 0$ is a shape parameter. After transforming as $T = (\mathcal{D}_f - \phi)/\vartheta$,

$$\begin{aligned} F_T(t) &= Pr[T \leq t] = Pr \left[\frac{\phi - \mathcal{D}_f}{\vartheta} \leq t \right] \\ &= Pr \left[\vartheta \geq \frac{\phi - \log 0.5}{t} \right] \\ &= 1 - G \left(\frac{\phi - \log 0.5}{t} \right) = \exp \left[- \left(\frac{\phi - \log 0.5}{\eta t} \right)^\zeta \right], \quad t > 0, \end{aligned}$$

and the distribution of T is called as the *reciprocal Weibull* distribution. Similarly, if ϑ follows a lognormal (μ, σ^2) , then

$$F_T(t) = 1 - \Phi \left\{ \frac{[\log(\phi - \log 0.5) - \log t] - \mu}{\sigma} \right\},$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution, hence the failure-time T also follows a lognormal distribution.

Example 2: A degradation path model of the standardized light intensity for a LED. The degradation path model of a light intensity depends heavily on light generating materials or methods. The light generating principle of a light emitting diode (LED) is quite different from that of VFDs, Fluorescent lamps, or PDPs, which give forth a brightness by means of a phosphor. Consequently, the degradation pattern of luminosity for a LED may not be of the form (6.1.28). Yu and Tseng [98] described the degradation path model of the standardized light intensity for a LED as

$$\mathcal{D}(t; \Theta) = \exp \left[-\alpha t^\beta \right], \quad \alpha > 0, t > 0,$$

where α and β are unknown parameters. The transformation of the observed light intensity is

$$\log (-\log \mathcal{D}(t; \Theta)) = \log \alpha + \beta \log t = \vartheta_1 + \vartheta_2 \log t, \quad t > 0.$$

Suppose that $\vartheta_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $\vartheta_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and ϑ_1 is independent of ϑ_2 . Then $\log (-\log \mathcal{D}(t; \Theta)) \sim \mathcal{N}(\mu_1 + \mu_2 \log t, \sigma_1^2 + \sigma_2^2 (\log t)^2)$, assuming both $Pr(\vartheta_1 < 0)$ and $Pr(\vartheta_2 \leq 0)$ are negligible. Since $T = \exp[(\log (-\log \mathcal{D}_f) - \vartheta_1)/\vartheta_2]$, with a value $\mathcal{D}_f = 0.5$ we have

$$\begin{aligned} F_T(t) &= Pr \left\{ \exp \left[\frac{\log (-\log \mathcal{D}_f) - \vartheta_1}{\vartheta_2} \right] \leq t \right\} \\ &\approx 1 - \Phi \left\{ \frac{-0.367 - (\mu_1 + \mu_2 \log t)}{\sqrt{\sigma_1^2 + \sigma_2^2 (\log t)^2}} \right\}. \end{aligned}$$

The failure-time distribution in the above is called *Bernstein* distribution and is discussed in Ahmad and Sheikh [2].

The MLEs for the parameters in the failure-time distribution can be computed by using the Newton-Raphson procedure described in Lindstrom and Bates [53]. MLEs of the failure-time distribution $\hat{F}_T(t)$ and the quantile \hat{t}_p are then constructed by replacing the model parameters in (4.17) and (7.2.46), 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. As a result, a numerical method such as 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. 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.

7.2.2 Nonlinear Random Coefficients Model

In the NRC model, 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 (see Meeker and Escobar [64]). For this evaluation, we first use the model parameter estimates $\hat{\beta}$, $\hat{\mathbf{b}}$, and $\hat{\mathbf{D}}$ (obtained from the m sample paths) to generate the N simulated realizations $\tilde{\beta}$, $\tilde{\mathbf{b}}$. From N values of $\tilde{\beta}$ and $\tilde{\mathbf{b}}$, compute the N failure times \tilde{t} by substituting $\tilde{\beta}$ and $\tilde{\mathbf{b}}$ into $\tau(t; \beta, \mathbf{b})$, and then solve for τ_f . For any desired values of t , $F_T(t)$ is estimated from the simulated empirical distribution

$$\hat{F}_T(t) \approx \frac{\text{number of } \tilde{t} \leq t}{N}. \quad (7.2.47)$$

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

Step 1. From the estimates $\hat{\beta}$, $\hat{\mathbf{b}}$, $\hat{\mathbf{D}}$, and $\hat{\sigma}^2$ (hereafter, assuming that $G_i(\beta_i) = I_{n_i}$) obtained by using ML or approximation methods, generate m simulated realizations of β_i^* , \mathbf{b}_i^* , $i = 1, \dots, m$.

Step 2. Compute m simulated degradation pathes

$$y_{ij}^* = \tau(t_{ij}; \beta_i^*, \mathbf{b}_i^*) + e_{ij}^* \quad (7.2.48)$$

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 $\hat{\beta}^*$, $\hat{\mathbf{b}}^*$, and $\hat{\mathbf{D}}^*$.

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

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

Step 5. Following Efron and Tibshirani [24], determine the lower and upper bounds of point-wise $100(1 - \alpha)\%$ confidence intervals for the distribution function $F_T(t)$:

$$[\underline{F}_T(t), \overline{F}_T(t)] = [\hat{F}_T^*(t)_{[l]}, \hat{F}_T^*(t)_{[u]}], \quad (7.2.49)$$

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

$$q = \frac{\text{number of } \hat{F}_T^*(t)_b \leq \hat{F}_T(t)}{B}, \quad b = 1, \dots, B. \quad (7.2.50)$$

7.3 VFD Example

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 VFD lifetime analyses, the information from the link function is not necessary to illustrate the nonlinear random coefficients model. In this analysis we model the degradation only at the test level, rather than at the milder regular-use conditions. Degradation tests of VFD luminosity are generally executed up to 1,000 hours for customer warranty purposes. In a special extended test, we also measured the luminosity 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. The nonlinear modeling with random coefficients model is slightly more difficult to illustrate with the smaller degradation data set taken in such a restricted way.

7.3.1 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. Based on a general nonlinear model described in Vonesh and

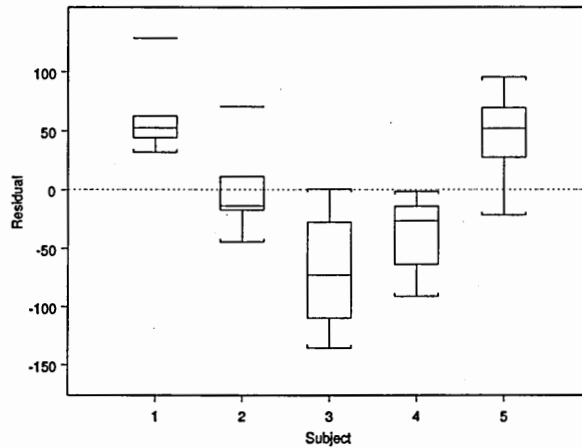


Figure 12: Boxplots of residuals from a single degradation model

Chinchilli [93] for a similar longitudinal data problem, we fit the following four-parameter (nonmonotonic) model to the degradation data:

$$E[y_t] = \frac{\beta_1 t^{\beta_2}}{\exp[\beta_3 t^{\beta_4}]}, \quad t > 0, \quad (7.3.51)$$

where $E[y_t]$ denotes expected brightness at measurement time t . The four-parameter model was deemed necessary since models in the form of (7.3.51) that have fewer than four parameters fail to characterize the nonmonotonic degradation of the VFDs. 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

$$E[y_t] = \frac{11259.7 \cdot t^{0.416}}{\exp[2.560 \cdot t^{0.116}]}, \quad t > 0. \quad (7.3.52)$$

Figure 12 shows boxplots for the computed residuals resulting from fitting (7.3.52) 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, [73], p. 279).

To accommodate this between-unit variation, a separate nonlinear model was fitted

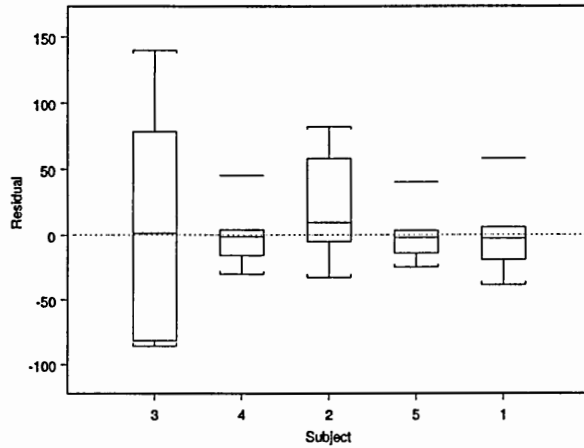


Figure 13: Boxplots of residuals from individually fitted models

to each unit; the individual specific nonlinear model shows an improved fit; the resulting boxplots of the residuals are displayed in Figure 13.

Despite the improved fit, Table 2 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 [55], 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 include variability among and within individual units, which is modeled most effectively using random coefficients.

Table 2: Parameter estimates for individual VFD degradation path

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

Parameters of the nonlinear random coefficient 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 SASTM NL MIXED procedure).

Because VFD units were inspected at identical test times, the parameters (β_2, β_4) in model (7.3.51) are constant across all VFD units. The model with two random coefficients provides a better fit than the simpler model with just one, with average prediction error 10 times larger for the simpler model. The final selected nonlinear model, then, is based on four unknown parameters $(\beta_1, \dots, \beta_4)$ and two random effects (b_1, b_2) :

$$y_{ij} = \frac{(\beta_1 + b_{i1}) \cdot t_{ij}^{\beta_2}}{\exp[(\beta_3 + b_{i2}) \cdot t_{ij}^{\beta_4}]} + e_{ij}, \quad (7.3.53)$$

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}) , $i = 1, \dots, 5$ are *i.i.d.* $\mathcal{N}(\mathbf{0}, \sigma_b^2 \mathbf{I}_2)$, and independent of the e_{ij} .

Table 3 shows the estimation results using the first-order approximation, the Lindstrom-Bates (LB) algorithm, the Laplacian approximation, the adaptive importance sampling approximation, and the adaptive Gaussian quadrature approximation. All of approximation methods are done by NL MIXED procedure and NLINMIX macro program in SASTM. 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 approximation method are based on the adaptive Gaussian quadrature with $N_{GQ} = 1$. The

Table 3: VFD parameter estimates and error variances for four approximation methods

Approximation method	β_1	β_2	β_3	β_4	$\sigma_{b_1}^2$	$\sigma_{b_2}^2$	σ_e^2
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

LB 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 3 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 4. 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 (Laplace, Adaptive Importance Sampling, Adaptive Gaussian Quadrature), the Laplace 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 confidence bands for brightness degradation, based on the Laplace approximation, are tabulated in Table 5. The standard errors of each parameters in Table 5 are constructed from the variance-covariance matrix computed as the inverse Hessian matrix. In the Laplace approximation, Hessian matrix $\partial^2 v_i(\beta, \mathbf{D}, \mathbf{b}_i) / \partial \mathbf{b}_i \partial \mathbf{b}_i'$

Table 4: Log-likelihood, average relative bias, and computational efficiency 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

is approximated by

$$\simeq \frac{\partial f_i(\beta, b_i)}{\partial b_i} \Big|_{\hat{b}_i} \frac{\partial f_i(\beta, b_i)}{\partial b'_i} \Big|_{\hat{b}_i} + \sigma^2 D^{-1}, \quad (7.3.54)$$

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

7.3.2 Comparison of Failure-Time Distributions

For the LRC model in (7.2.45), 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: $\hat{\beta} = (6.7, -1.92 \times 10^{-4})'$, $\hat{D}_{00} = 2.38 \times 10^{-3}$, $\hat{D}_{11} = 4.99 \times 10^{-10}$, $\hat{D}_{01} = \hat{D}_{10} = -5.13 \times 10^{-8}$, $\sigma^2 = 6.88 \times 10^{-3}$. In Figure 14, even after discarding the initial burn-in stage, the LRC model (dot line) fails to accurately fit an individual degradation path for VFD. The main reason is that the LRC model ignores the poisoning effect accompanied inevitably after initial burn-in stage. In contrast, the NRC model fit (based on parameter estimates resulted from the Laplace approximation) appears to successfully captures the burn-in characteristic of VFDs.

The predicted values for response, along with the average relative prediction bias are obtained by plugging the estimates into the LRC model along with the estimates in Table 5 from the NRC model. The resulting relative prediction bias values are summarized in Table 6. Prediction bias is much higher in the LRC model, especially for sample 2 and 3 which have an obvious nonmonotonic behavior. This shows that the prediction from the

Table 5: Laplace approximation parameter estimates and their standard errors

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
$\sigma_{b_1}^2$	10.0056	0.02023
$\sigma_{b_2}^2$	8.411×10^{-6}	2.458×10^{-6}
σ_e^2	100.10	4.5227

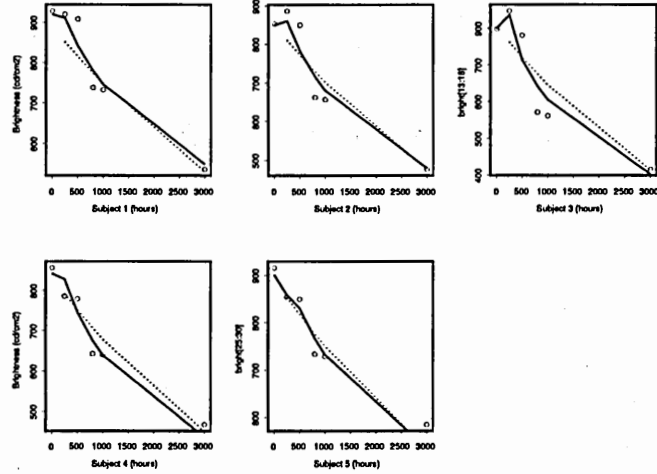


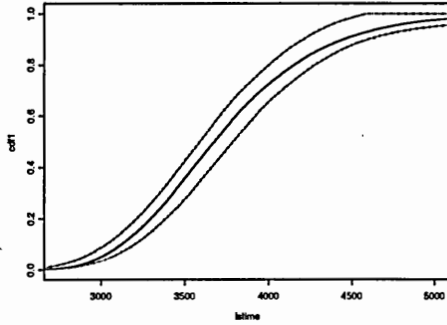
Figure 14: LRC and NRC model fit for an individual VFD degradation path

NRC model is substantially more reliable.

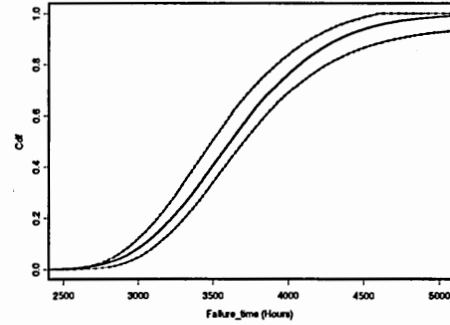
For comparison, using Monte Carlo simulation (with $N = 50,000$), the failure-time distributions using were derived for both LRC and NRC models. Both failure-time distributions were based on the threshold τ_f with the extended lines of five degradation paths censored at 3,000 hours. Applied to the procedure in section 6.2.2, $\hat{F}_T(t)$ and its 90% bootstrap confidence intervals are plotted in Fig 15. The point estimates and 90% confidence intervals (in parenthesis) of p^{th} quantiles of failure-time distribution are summarized in Table 7, for $p = .05, .1, \text{ and } .5$. The intervals are based on $B = 2,000$ bootstrap samples.

Table 6: Average relative prediction bias for the LRC and the NRC model

Sample	LRC	NRC
1	0.0513	0.0320
2	0.0695	0.0397
3	0.1028	0.0562
4	0.0472	0.0382
5	0.0427	0.0302
Average	0.0627	0.0393



(a) LRC model



(b) NRC model

Figure 15: Estimated CDF of VFD unit lifetime along with 90% (pointwise) confidence intervals

7.3.3 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 (7.3.53) introduced in the last section

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

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

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

Table 7: Quantiles and their bootstrap confidence intervals

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

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

where $y_{ij} = \log(y_{ij}^*)$, y_{ij}^* represents the real luminosity at j^{th} measurement on the i^{th} individual and $I[\cdot]$ is an index function for time interval. Model II is simulated on the basis of the intercept and slope estimates obtained by fitting five degradation paths in separated (transformed) linear model in corresponding time intervals. Model II is constructed in such a way that neither the NRC nor LRC will fit the model particularly well. For model I, $\beta = (900, 0.06, 0.04, 0.424)$, $b_{i1} \sim N(0, \sigma_1^2)$, $b_{i2} \sim N(0, \sigma_2^2)$, and $\text{Cov}(b_{i1}, b_{i2}) = 0$. It reaches a peak luminosity at 19.7 hours, on average. Model II is generated with $\beta = (6.7, 2.0 \times 10^{-4}, 6.9, -5.0 \times 10^{-4}, 6.65, -2.0 \times 10^{-4})$, and has two random effects as Model I, with different values of σ_1^2 and σ_2^2 . $\sigma_\epsilon^2 = 100$, same for model I and II. Expected peak luminosity for Model II occurs at 250 hours.

To compare the NRC model to the LRC model, we fit (7.3.53) 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 8 lists the percentage increase in average relative bias when the (truncated) log-linear model is used with varying random coefficients. 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.

Table 8: Increase in relative error with respect to truncated log-linear model

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%

CHAPTER VIII

ANALYSIS OF A CHANGE-POINT IN NON-MONOTONIC DEGRADATION PATH

8.1 *Change-Point Regression Model (Non-Bayesian Approach)*

For a tested light display, let Y_n be the measured log-luminosity at the n^{th} (fixed) time period, $n = 1, 2, \dots, N$. Suppose Y_1, Y_2, \dots, Y_N are normally and independently distributed, then

$$Y_n = \begin{cases} \alpha_1 + \beta_1 t_n + \epsilon_{1n} & n = 1, \dots, \tau \\ \alpha_2 + \beta_2 t_n + \epsilon_{2n} & n = \tau + 1, \dots, N, \end{cases} \quad (8.1.55)$$

where ϵ_{in} 's are *iid* $\mathcal{N}(0, \sigma_i^2)$ for $i = 1, 2$ and the parameter τ is an unknown change-point. The model (8.1.55) is also called “two-phase” or “switching” regression model. We assume that $(\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)$ so that there is exactly one change in the relationship between Y_n and t_n , the parameters $\Theta = (\alpha_i, \beta_i, \sigma_i^2)_{i=1,2}$ are unknown, and $t_1 < t_2 \dots < t_N$. Additionally, under the continuity assumption for the degradation path, the intersection point γ of two regression models is $\gamma = (\alpha_1 - \alpha_2)/(\beta_2 - \beta_1)$ satisfying the constraint $\mathcal{C}_\tau = \{\gamma : t_\tau \leq \gamma \leq t_{\tau+1}\}$.

The change-point problem for linear models has been developed by several authors including Quandt ([74], [75]), Hawkins [38], and Kim [46]. Beam [9] applied a two-phase regression to the study of strategy shifts in a mental rotation task using the self-developed software *Segcurve*, which we apply to analyze practical examples. Csörgő and Horváth [19] and Chen and Gupta [15] provide several references pertaining to different aspects of change point problems.

8.1.1 Maximum Likelihood Estimation of Change-Point Model Parameters

Using the model described in model (8.1.55), it is possible to derive maximum likelihood estimators (MLEs) for $\alpha_i, \beta_i, \sigma_i^2, i = 1, 2$. The log likelihood is given by

$$\begin{aligned} \mathcal{L}(\alpha_i, \beta_i, \sigma_i^2, \tau) \approx & -\tau \log \sigma_1^2 - (n - \tau) \log \sigma_2^2 - \frac{1}{\sigma_1^2} \sum_{n=1}^{\tau} (y_n - \alpha_1 - \beta_1 t_n)^2 \\ & - \frac{1}{\sigma_2^2} \sum_{n=\tau+1}^N (y_n - \alpha_2 - \beta_2 t_n)^2. \end{aligned} \quad (8.1.56)$$

The procedure to obtain MLEs for the parameters in (8.1.56) along with the corresponding estimate of change-point τ can be summarized as:

Find $\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i^2$, and $\hat{\tau}$ such that

$$\mathcal{L}(\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i^2, \hat{\tau}) \equiv \sup \mathcal{L}(\alpha_i, \beta_i, \sigma_i^2, \tau), \quad i = 1, 2$$

subject to

$$\hat{\gamma} = \frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\hat{\beta}_2 - \hat{\beta}_1},$$

and

$$t_{\hat{\tau}} \leq \hat{\gamma} < t_{\hat{\tau}+1}.$$

Under all the constraints above, it was proved by Hudson (1966) that $(\hat{\alpha}_i, \hat{\beta}_i)_{i=1,2}$ are unconstrained local least squares estimates and $\hat{\gamma}$ is the real point of intersection of the two regression lines if $\hat{\gamma} \in [t_{\hat{\tau}}, t_{\hat{\tau}+1})$. The MLEs $\hat{\alpha}_i$ and $\hat{\beta}_i, i = 1, 2$ are

$$\begin{aligned} \hat{\beta}_1 &= \frac{\hat{\tau} \sum_{n=1}^{\hat{\tau}} t_n y_n - (\sum_{n=1}^{\hat{\tau}} t_n) \cdot (\sum_{n=1}^{\hat{\tau}} y_n)}{\hat{\tau} \sum_{n=1}^{\hat{\tau}} t_n^2 - (\sum_{n=1}^{\hat{\tau}} t_n)^2}, \\ \hat{\alpha}_1 &= \bar{y}_{\hat{\tau}} - \hat{\beta}_1 \bar{t}_{\hat{\tau}}, \\ \hat{\beta}_2 &= \frac{(N - \hat{\tau}) \sum_{n=\hat{\tau}+1}^N t_n y_n - (\sum_{n=\hat{\tau}+1}^N t_n) \cdot (\sum_{n=\hat{\tau}+1}^N y_n)}{(N - \hat{\tau}) \sum_{n=\hat{\tau}+1}^N t_n^2 - (\sum_{n=\hat{\tau}+1}^N t_n)^2}, \\ \hat{\alpha}_2 &= \bar{y}_{N-\hat{\tau}} - \hat{\beta}_2 \bar{t}_{N-\hat{\tau}}, \end{aligned} \quad (8.1.57)$$

where $\bar{z}_{\hat{\tau}} = \sum_{i=1}^{\hat{\tau}} z_i / \hat{\tau}$, $\bar{z}_{N-\hat{\tau}} = \sum_{i=\hat{\tau}+1}^N z_i / (N - \hat{\tau})$. The estimators for σ_1^2 and σ_2^2 are given as

$$\begin{aligned} \hat{\sigma}_1^2 &= \frac{\sum_{n=1}^{\hat{\tau}} (y_n - \hat{\alpha}_1 - \hat{\beta}_1 t_n)^2}{\hat{\tau}}, \\ \hat{\sigma}_2^2 &= \frac{\sum_{n=\hat{\tau}+1}^N (y_n - \hat{\alpha}_2 - \hat{\beta}_2 t_n)^2}{N - \hat{\tau}}, \end{aligned} \quad (8.1.58)$$

and, finally, the MLE $\hat{\gamma} = (\hat{\alpha}_1 - \hat{\alpha}_2)(\hat{\beta}_2 - \hat{\beta}_1)$ is the value of $t_{\hat{\tau}} \leq \hat{\gamma} < t_{\hat{\tau}+1}$ which minimizes

$$\mathcal{L}(\gamma) \approx \hat{\tau} \log \hat{\sigma}_1^2 + (n - \hat{\tau}) \log \hat{\sigma}_2^2. \quad (8.1.59)$$

Since γ is a function of parameters $\theta = [\theta_1, \theta_2]' = [(\alpha_1, \beta_1)', (\alpha_2, \beta_2)']'$ and τ , we can evaluate $\mathcal{L}(\gamma)$ by calculating the parameters $\theta(\gamma)$ as a function of γ , using different values of τ . At last, we can find $\hat{\gamma}$ that minimizes $\mathcal{L}(\gamma)$, while simultaneously satisfying all the constraints for γ so that $\theta(\hat{\gamma}) = \hat{\theta}$. Even when the length of the interval is relatively large, or the number of measurements is small, (8.1.59) ensures a unique solution of γ .

8.1.2 Confidence Intervals for an Unknown Change-Point

Confidence intervals for $\hat{\theta}$ and $\hat{\gamma}$ can be constructed by using a normal approximation based on the asymptotic distribution. Feder [29] considered the asymptotic distribution of $\hat{\theta}$ and $\hat{\gamma}$ in the change-point regression problem. He derived the asymptotic distribution of $\hat{\theta}$ as

$$\hat{\theta} \sim \mathcal{MN}(\theta, \sigma^2 \Sigma_{\theta}^{-1}),$$

where \mathcal{MN} represents the multivariate normal function and the asymptotic correlation matrix Σ_{θ} can be estimated by

$$\hat{\Sigma}_{\theta} = \frac{1}{N} \sum_{i=1}^N \frac{\partial Y(\hat{\theta}, \hat{\gamma})}{\partial \theta} \frac{\partial Y(\hat{\theta}, \hat{\gamma})}{\partial \theta'}.$$

Since $\hat{\Sigma}_{\theta}$ is a diagonal block matrix with 2 blocks, $\hat{\theta}_i$ can be treated as independent multivariate normal distributions with

$$\hat{\theta}_i \sim \mathcal{MN}(\theta_i, \hat{\sigma}^2 (X_i' X_i)^{-1}), \quad i = 1, 2,$$

where $\hat{\sigma}^2 = [\hat{\tau} \hat{\sigma}_1^2 + (N - \hat{\tau}) \hat{\sigma}_2^2] / N$ and X_1, X_2 are the respective design matrices for the first τ and the second $N - \tau$ observations. By applying the so-called delta method, the MLE $\hat{\gamma}$ is asymptotically normal with mean γ and variance

$$\text{var}(\hat{\gamma}) \approx \left[\frac{\partial \hat{\gamma}}{\partial \theta} \right]' \hat{\Sigma}_{\theta}^{-1} \left[\frac{\partial \hat{\gamma}}{\partial \theta} \right].$$

In the two-phase regression model, $\partial \hat{\gamma} / \partial \theta = (1, \hat{\gamma}, -1, -\hat{\gamma})'$, hence $\text{var}(\hat{\gamma})$ is reduced to

$$(\hat{\beta}_2 - \hat{\beta}_1)^{-2} \left\{ [\text{var}(\hat{\alpha}_1) + \text{var}(\hat{\alpha}_2)] + 2\hat{\gamma} [\text{cov}(\hat{\alpha}_1, \hat{\beta}_1) + \text{cov}(\hat{\alpha}_2, \hat{\beta}_2)] + \hat{\gamma}^2 [\text{var}(\hat{\beta}_1) + \text{var}(\hat{\beta}_2)] \right\}. \quad (8.1.60)$$

As pointed out in Hinkley [39], for moderate-sized samples, asymptotic normal distributions of $\hat{\theta}_i$'s serve as effective approximations but the asymptotic normal distribution of $\hat{\gamma}$ may be inadequate. Both Hinkley [39] and Feder [28] suggested using the refined asymptotic distribution of $\hat{\gamma}$ which has better small-sample properties. However, Hinkley [39] recommended use of the likelihood-ratio statistic for inferences about γ or joint inferences about (θ, γ) .

Feder [29] proved that under ordinary regularity conditions, the log-likelihood ratio

$$-2\log \mathcal{L} = N \log \frac{\mathcal{S}(\hat{\theta}, \hat{\gamma})}{\mathcal{S}(\tilde{\theta}_0, \tilde{\gamma}_0)}$$

is approximately χ_1^2 distributed under $H_0 : \gamma = \gamma_0$. Here $\mathcal{S}(\cdot)$ denotes the residual sum of squares and $(\tilde{\theta}_0, \tilde{\gamma}_0)$ denote the MLEs of (θ, γ) subject to the constraints imposed by H_0 . He noted that the distribution of $-2\log \mathcal{L}$ can be better approximated in small or medium sized samples by a multiple of the (asymptotically equivalent) $F_{1, N-p}$ distribution, where p is the dimension of the parameter space.

Following Hinkley's [40] approximate F (AF) method, we can obtain a $100(1 - \delta)\%$ confidence interval for γ by inverting the F -test for H_0 , namely

$$\left\{ \gamma : \frac{\mathcal{S}(\tilde{\theta}_0, \tilde{\gamma}_0) - \mathcal{S}(\hat{\theta}, \hat{\gamma})}{\mathcal{S}(\hat{\theta}, \hat{\gamma})/(N - 4)} \leq F_{1, N-4}^\delta \right\}. \quad (8.1.61)$$

Siegmund and Zhang [82] discussed several procedures (including the AF method) for constructing confidence regions in change-point regression model. They suggested the conditional likelihood ratio (CLR) statistic which gives more conservative confidence intervals in the nonlinear change-point regression model. However, the AF method is much easier to employ than the CLR method. In addition, confidence intervals obtained by the AF method are close to those of the CLR method in two-phase regression model, hence the AF method is applied to construct the confidence intervals for change-point γ .

8.2 Bayesian Change-point Regression Model

In this section, a "Bayesian" method for detecting a change-point in the degradation path is introduced. Bayesian method can be especially important in nonlinear paths that can

be motivated by physical principles and paths with change-points, when a change in the degradation mechanism causes the degradation path to change functionally.

We consider a general regression model with one change-point

$$Y_i = \begin{cases} \mathbf{x}'_i \beta_1 + \epsilon_i & i = 1, \dots, \tau \\ \mathbf{x}'_i \beta_2 + \epsilon_i & i = \tau + 1, \dots, n, \end{cases} \quad (8.2.62)$$

where \mathbf{x}_i denotes $(p \times 1)$ vector of predictor variables, $\beta_i, i = 1, 2$ are $(p \times 1)$ vectors of coefficient parameters of the first and second linear regression model, and the ϵ_i 's are *iid* $\mathcal{N}(0, \sigma^2)$. We assume that $\beta_1 \neq \beta_2$ and all the parameters $\beta_1, \beta_2, \sigma^2, \tau$ are unknown, and $\mathbf{x}_1 < \mathbf{x}_2, \dots < \mathbf{x}_n$ to denote time sequential variables. In addition, we assume the intersection hyperplane $\gamma \in \mathcal{R}^{p-1}$ of two regression models satisfies the constraint $\mathbf{x}_\tau \leq \gamma < \mathbf{x}_{\tau+1}$. For instance, in two-phase model consisting of two simple linear lines, the intersection point $\gamma = (\alpha_1 - \alpha_2)/(\beta_2 - \beta_1)$, where α_i, β_i for $i = 1, 2$ are i^{th} intercept and slope respectively.

Bayesian change-point problem for the regression model has been adopted by Chin Choy and Broemeling [17] without a continuity constraint and Smith and Cook [83] with the continuity at the change-point. Carlin *et al.* [13] presented hierarchical Bayesian change-point models and used a Gibbs sampler to obtain posterior distributions.

When the model (8.2.62) is represented as a standard regression model

$$\mathbf{Y} = \mathbf{X}(\tau)\beta + \mathbf{e}, \quad (8.2.63)$$

where $\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2)'$, $\beta = (\beta'_1, \beta'_2)' \in \mathcal{R}^{2p}$, $\mathbf{e} = (\mathbf{e}'_1, \mathbf{e}'_2)'$ and

$$\mathbf{X}(\tau) = \begin{bmatrix} \mathbf{X}_1(\tau) & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2(\tau) \end{bmatrix}.$$

$\mathbf{Y}_1, \mathbf{X}_1(\tau), \beta_1$ and \mathbf{e}_1 denote the response vector, design matrix, coefficient vector and random error vector, respectively, for the first τ observations. Similarly, $\mathbf{Y}_2, \mathbf{X}_2(\tau), \beta_2$ and \mathbf{e}_2 correspond to the same parameters for the remaining $n - \tau$ observations, and $\mathbf{0}$ denotes zero matrix.

We assign the prior information concerning β, σ^2, τ as:

(a) The conditional distribution of $\beta|r = 1/\sigma^2$ is $2p$ -dimensional multivariate normal distribution with mean β^* and covariance matrix $r^{-1}\Sigma$, where Σ is a given symmetric $2p \times 2p$ positive definite matrix.

(b) The prior distribution on R is a gamma distribution with parameters a and b ($a > 0, b > 0$), that is

$$\pi_0(r) = \frac{b^a}{\Gamma(a)} r^{a-1} e^{-br}, \quad r > 0.$$

(c) change-point τ , independent of β and R , is uniformly distributed over \mathcal{I}_{n-3} , that is, $\pi_0(\tau) = 1/(n-3)$ for $\tau = 2, \dots, n-2$, and 0 otherwise.

From (a), the condition prior of β given $R = r$

$$\pi_0(\beta|R=r) = \frac{r^p}{(2\pi)^p |\Sigma|^{1/2}} \exp\left[-\frac{r}{2}(\beta - \beta^*)' \Sigma^{-1}(\beta - \beta^*)\right].$$

The likelihood function of the data given the parameter is

$$\begin{aligned} \mathcal{L}(\beta, \tau, r) &= f(\mathbf{y}|\beta, \tau, r) \\ &= \left(\frac{r}{2\pi}\right)^{n/2} \exp\left\{-\frac{r}{2}(\mathbf{y} - \mathbf{X}(\tau)\beta)'(\mathbf{y} - \mathbf{X}(\tau)\beta)\right\}. \end{aligned} \quad (8.2.64)$$

8.2.1 Posterior Distribution of a Change-Point

A posterior distribution of τ is derived in two cases: with and without a continuity constraint.

8.2.1.1 Without a Continuity Constraint

Deriving a posterior distribution of τ is much simpler than deriving with a continuity constraint. The posterior distribution of τ $p(\tau|\mathbf{y})$ can be obtained from the following theorem:

Theorem 1 Denote the MLE as $\hat{\beta} = (\mathbf{X}(\tau)' \mathbf{X}(\tau))^{-1} \mathbf{X}(\tau)' \mathbf{y}$, $\hat{\mathbf{y}} = \mathbf{X}(\tau) \hat{\beta}$, $\Sigma_{(\tau)} = \mathbf{X}(\tau)' \mathbf{X}(\tau) + \Sigma^{-1}$, and $\tilde{\beta} = \Sigma_{(\tau)}^{-1}(\mathbf{X}(\tau)' \mathbf{y} + \Sigma^{-1} \beta^*)$. Then the posterior distribution of change-point τ

$$p(\tau|\mathbf{y}) \propto \frac{1}{|\Sigma|^{1/2} |\Sigma_{(\tau)}|^{1/2}} \cdot \frac{\Gamma(b + \Lambda(\tau)/2) \cdot b^a}{\Gamma(a) \cdot (b + \Lambda(\tau)/2)^{a+n/2}}, \quad (8.2.65)$$

where $\Lambda(\tau) = (\mathbf{y} - \mathbf{X}(\tau) \hat{\beta})'(\mathbf{y} - \mathbf{X}(\tau) \hat{\beta}) + (\hat{\beta} - \tilde{\beta})' \Sigma^{-1}(\hat{\beta} - \tilde{\beta}) + (\hat{\beta} - \tilde{\beta})' \mathbf{X}(\tau)' \mathbf{X}(\tau)(\hat{\beta} - \tilde{\beta})$.

Proof. The joint posterior density of all the parameters

$$\begin{aligned} p(\beta, \tau, r | \mathbf{y}) &= \frac{\mathcal{L}(\beta, \tau, r) f(\beta, \tau, r)}{f(\mathbf{y})} \\ &\propto \mathcal{L}(\beta, \tau, r) \pi_0(\beta | \tau, r) \pi_0(r | \tau) \pi_0(\tau). \end{aligned}$$

The posterior mass function of change point τ is obtained by integrating $\pi(\beta, \tau, r | \mathbf{y})$ with respect to β and R .

$$\mathcal{L}(\beta, \tau, r) \pi_0(\beta | \tau, r) \propto \frac{r^{p+n/2}}{|\Sigma|^{1/2}} \exp \left\{ -\frac{r}{2} [(\mathbf{y} - \mathbf{X}(\tau)\beta)'(\mathbf{y} - \mathbf{X}(\tau)\beta) + (\beta - \beta^*)' \Sigma^{-1} (\beta - \beta^*)] \right\},$$

Note that $(\mathbf{y} - \mathbf{X}(\tau)\beta)'(\mathbf{y} - \mathbf{X}(\tau)\beta) + (\beta - \beta^*)' \Sigma^{-1} (\beta - \beta^*) = \Lambda(\tau) + (\beta - \tilde{\beta})' \Sigma_{(\tau)} (\beta - \tilde{\beta})$, and hence

$$\begin{aligned} \int \mathcal{L}(\beta, \tau, r) \pi_0(\beta | \tau, r) d\beta &\propto \frac{r^{p+n/2}}{|\Sigma|^{1/2}} \exp \left(-\frac{r\Lambda(\tau)}{2} \right) \\ &\quad \times \int \exp \left[-\frac{r}{2} (\beta - \tilde{\beta})' \Sigma_{(\tau)} (\beta - \tilde{\beta}) \right] d\beta \\ &= \frac{r^{n/2}}{|\Sigma|^{1/2} |\Sigma_{(\tau)}|^{1/2}} \exp \left(-\frac{r\Lambda(\tau)}{2} \right), \end{aligned}$$

since this integral is with respect to $(2p)$ -dimensional normal density for β with mean $\tilde{\beta}$ and covariance $r^{-1} \Sigma_{(\tau)}^{-1}$. Then

$$\begin{aligned} \int \mathcal{L}(\beta, \tau, r) \pi_0(\beta | \tau, r) \pi_0(r | \tau) d\beta dr &\propto \frac{1}{|\Sigma|^{1/2} |\Sigma_{(\tau)}|^{1/2}} \cdot \frac{b^a}{\Gamma(a)} \\ &\quad \times \int r^{a-1+n/2} \exp \left\{ -r \left(b + \frac{\Lambda(\tau)}{2} \right) \right\} dr \\ &= \frac{1}{|\Sigma|^{1/2} |\Sigma_{(\tau)}|^{1/2}} \cdot \frac{\Gamma(b + \Lambda(\tau)/2) \cdot b^a}{\Gamma(a) \cdot (b + \Lambda(\tau)/2)^{a+n/2}}. \end{aligned}$$

Since $\pi_0(\tau)$ is constant, the posterior distribution of τ is

$$p(\tau | \mathbf{y}) \propto \frac{1}{|\Sigma|^{1/2} |\Sigma_{(\tau)}|^{1/2}} \cdot \frac{\Gamma(b + \Lambda(\tau)/2)}{\Gamma(a)} \cdot \frac{b^a}{(b + \Lambda(\tau)/2)^{a+n/2}}.$$

This completes the proof. ■

8.2.1.2 With a Continuity Constraint

To derive a posterior distribution of τ when there is a continuity constraint, first use the following lemma:

Lemma 2 *The conditional posterior distribution of R given τ*

$$p(r|\tau, \mathbf{y}) \sim G\left(a + \frac{n}{2}, b + \frac{\Lambda(\tau)}{2}\right),$$

where $G(\cdot)$ denotes gamma distribution.

Proof. The posterior of R given τ is

$$p(r|\tau, \mathbf{y}) = \frac{p(\tau, r|\mathbf{y})}{p(\tau|\mathbf{y})},$$

where

$$\begin{aligned} p(\tau, r|\mathbf{y}) &= \int p(\beta, \tau, r|\mathbf{y}) d\beta \\ &\propto \frac{1}{|\Sigma|^{1/2} |\Sigma(\tau)|^{1/2}} \cdot \frac{b^a}{\Gamma(a)} r^{a+n/2-1} \exp\left\{-r\left(b + \frac{\Lambda(\tau)}{2}\right)\right\}. \end{aligned}$$

$p(\tau, r|\mathbf{y})$ is divided by (8.2.65), and then

$$p(r|\tau, \mathbf{y}) = \frac{\Gamma(a + \frac{n}{2})}{(b + \frac{\Lambda(\tau)}{2})^{a+n/2}} r^{a+n/2-1} \exp\left\{-r\left(b + \frac{\Lambda(\tau)}{2}\right)\right\}, \quad (8.2.66)$$

which represents a gamma density with parameter $(a + n/2)$ and $(b + \Lambda(\tau)/2)$. ■

Our primary concern is for estimating the meeting point of two switching lines in order to provide the information about optimal “burn-in” policy. Previously, we restricted the intersection of the hyperplane γ to be in the interval $[\mathbf{x}_\tau, \mathbf{x}_{\tau+1}]$. Introducing the index function for the constraint $I(C) = I\{C : \gamma \in [\mathbf{x}_\tau, \mathbf{x}_{\tau+1}]\}$, then the conditional posterior on γ given the constraint is

$$\begin{aligned} p(\gamma|C, \mathbf{y}) &= \sum_{\tau} p(\gamma, \tau|C, \mathbf{y}) \\ &\propto \sum_{\tau} \frac{I(C|\gamma, \tau, \mathbf{y}) p(\gamma, \tau|\mathbf{y})}{p(I(C)|\mathbf{y})}, \end{aligned}$$

where $I(C|\gamma, \tau, \mathbf{y}) = 1$ if $\gamma \in (\mathbf{x}_\tau, \mathbf{x}_{\tau+1})$ and 0 otherwise. The posterior on C is

$$\begin{aligned} p(I(C)|\mathbf{y}) &= \sum_{\tau} \left\{ \int p(I(C)|\gamma, \tau, \mathbf{y}) p(\gamma|\tau, \mathbf{y}) d\gamma \right\} p(\tau|\mathbf{y}) \\ &= \sum_{\tau} \left\{ \int_{\mathcal{I}_C} p(\gamma|\tau, \mathbf{y}) d\gamma \right\} p(\tau|\mathbf{y}) \\ &= \sum_{\tau} \int_{\mathcal{I}_C} p(\gamma, \tau|\mathbf{y}) d\gamma, \end{aligned}$$

where $\mathcal{I}_C \in (\mathbf{x}_\tau, \mathbf{x}_{\tau+1})$.

We need to derive the posterior densities as to each elements of (8.2.67) to obtain $p(\gamma|C, \mathbf{y})$. Using basic conditional probability rule

$$p(\gamma, \tau|\mathbf{y}) = p(\gamma|\tau, \mathbf{y})p(\tau|\mathbf{y}), \quad (8.2.67)$$

where $p(\gamma|\tau, \mathbf{y}) = \int p(\gamma|\tau, r, \mathbf{y})p(r|\tau, \mathbf{y})dr$ and $p(\tau|\mathbf{y})$ is given in (8.2.65).

To derive $p(\gamma|\tau, \mathbf{y})$, we need only calculate $p(\gamma|\tau, r, \mathbf{y})$. Bayesian analysis of γ for multiple predictors is complicated and for repeated measurement data such as VFD and PDP degradation paths, only one predictor, *i.e.*, inspection time is used to model the response, hence for simplicity, we will develop a Bayesian procedure of γ based on a simple two-phase regression. Define the differences of two of intercepts and slopes in two-phase regression model $\theta_\alpha = \alpha_1 - \alpha_2$, $\theta_\beta = \beta_2 - \beta_1$ respectively and let $\gamma = \theta_\alpha/\theta_\beta$, $\eta = \theta_\beta$. Then we have

$$p(\gamma|\tau, r, \mathbf{y}) = \int p(\gamma, \eta|\tau, r, \mathbf{y})d\eta.$$

The joint density of (γ, η) is easily derived from that of $(\theta_\alpha, \theta_\beta)$ using a change of variables. To derive the joint posterior density of $\theta = (\theta_\alpha, \theta_\beta)'$ given τ, r, \mathbf{y} , first we note:

Lemma 3 *The posterior distribution of β given τ, r , and \mathbf{y} $p(\beta|\tau, r, \mathbf{y})$ has a multivariate density with mean $\tilde{\beta}$ and variance $r^{-1}\Sigma_{(\tau)}^{-1}$.*

Proof.

$$\begin{aligned} p(\beta|\tau, r, \mathbf{y}) &\propto \mathcal{L}(\beta, \tau, r)\pi_0(\beta|\tau, r) \\ &\propto \frac{r^{p+n/2}}{|\Sigma|^{1/2}} \exp \left\{ -\frac{r}{2} [\Lambda(\tau) + (\beta - \tilde{\beta})' \Sigma_{(\tau)} (\beta - \tilde{\beta})] \right\} \\ &\propto \frac{r^{p+n/2}}{|\Sigma_{(\tau)} - \mathbf{X}(\tau)' \mathbf{X}(\tau)|^{-1/2}} \exp \left\{ -\frac{r}{2} [(\beta - \tilde{\beta})' \Sigma_{(\tau)} (\beta - \tilde{\beta})] \right\} \\ &\equiv \mathcal{MN}(\tilde{\beta}, r^{-1}\Sigma_{(\tau)}^{-1}), \end{aligned}$$

where $\mathcal{MN}(\cdot)$ denotes multivariate normal density. ■

For $(\theta_\alpha, \theta_\beta)$ which is the linear combination of β , the joint density given τ, r, \mathbf{y} is also multivariate normal with mean $\tilde{\theta} = (\tilde{\theta}_\alpha, \tilde{\theta}_\beta)'$ and covariance matrix $(r\Sigma_{(\theta)})^{-1}$, where the

entries $\delta(\tau)_{ij}, i, j = 1, 2$ of $\Sigma_{(\theta)}$ can be identified straightforwardly by manipulating an inverse of $\Sigma_{(\tau)}^{-1}$. Thus

$$p(\theta|\tau, r, \mathbf{y}) \propto r|\Sigma_{(\theta)}|^{1/2} \exp \left[- \left(\frac{r}{2} \right) \begin{pmatrix} \theta_\alpha - \tilde{\theta}_\alpha \\ \theta_\beta - \tilde{\theta}_\beta \end{pmatrix}' \Sigma_{(\theta)} \begin{pmatrix} \theta_\alpha - \tilde{\theta}_\alpha \\ \theta_\beta - \tilde{\theta}_\beta \end{pmatrix} \right].$$

The joint density of (γ, η) given τ, r, \mathbf{y} can be obtained by transforming θ and using Jacobian term $|\eta|$

$$\begin{aligned} p(\gamma, \eta|\tau, r, \mathbf{y}) &\propto r|\eta||\Sigma_{(\theta)}|^{1/2} \exp \left[- \left(\frac{r}{2} \right) \begin{pmatrix} \gamma\eta - \tilde{\theta}_\alpha \\ \eta - \tilde{\theta}_\beta \end{pmatrix}' \begin{bmatrix} \delta_{11}(\tau) & \delta_{12}(\tau) \\ \delta_{12}(\tau) & \delta_{22}(\tau) \end{bmatrix} \begin{pmatrix} \gamma\eta - \tilde{\theta}_\alpha \\ \eta - \tilde{\theta}_\beta \end{pmatrix} \right] \\ &\equiv r|\eta||\Sigma_{(\theta)}|^{1/2} \exp \left[- \left(\frac{r\psi_1(\gamma, \tau)}{2} \right) \left(\eta - \frac{\psi_2(\gamma, \tau)}{\psi_1(\gamma, \tau)} \right)^2 \right] \\ &\quad \times \exp \left[- \left(\frac{r}{2} \right) \left(\psi_3(\gamma, \tau) - \frac{\psi_2(\gamma, \tau)^2}{\psi_1(\gamma, \tau)} \right) \right], \end{aligned}$$

where

$$\begin{aligned} \psi_1(\gamma, \tau) &= \gamma^2 \delta_{11}(\tau) + 2\gamma\delta_{12}(\tau) + \delta_{22}(\tau), \\ \psi_2(\gamma, \tau) &= \gamma(\tilde{\theta}_\alpha \delta_{11}(\tau) + \tilde{\theta}_\beta \delta_{12}(\tau)) + \tilde{\theta}_\alpha \delta_{12}(\tau) + \tilde{\theta}_\beta \delta_{22}(\tau), \\ \psi_3(\gamma, \tau) &= \tilde{\theta}_\alpha^2 \delta_{11}(\tau) + 2\tilde{\theta}_\alpha \tilde{\theta}_\beta \delta_{12}(\tau) + \tilde{\theta}_\beta^2 \delta_{22}(\tau). \end{aligned}$$

Integrating $p(\gamma, \eta|\tau, r, \mathbf{y})$ over η , we obtain the explicit form of $p(\gamma|\tau, r, \mathbf{y})$;

$$\begin{aligned} p(\gamma|\tau, r, \mathbf{y}) &\propto r|\Sigma_{(\theta)}|^{1/2} \exp \left[- \left(\frac{r}{2} \right) \left(\psi_3(\gamma, \tau) - \frac{\psi_2(\gamma, \tau)^2}{\psi_1(\gamma, \tau)} \right) \right] \times \\ &\quad \int |\eta| \exp \left[- \left(\frac{r\psi_1(\gamma, \tau)}{2} \right) \left(\eta - \frac{\psi_2(\gamma, \tau)}{\psi_1(\gamma, \tau)} \right)^2 \right] d\eta. \end{aligned} \quad (8.2.68)$$

Because the integral in (8.2.68) is related to the expectation of $|\eta|$ and η is normally distributed with mean $\mu(\gamma, \tau) = \psi_2(\gamma, \tau)/\psi_1(\gamma, \tau)$ and variance $\nu^2 = [r\psi_1(\gamma, \tau)]^{-1}$, $p(\gamma|\tau, r, \mathbf{y})$ is proportional to

$$\begin{aligned} &[\psi_1(\gamma, \tau)]^{-1/2} |\Sigma_{(\theta)}|^{1/2} \exp \left[- \left(\frac{r}{2} \right) \left(\psi_3(\gamma, \tau) - \frac{\psi_2(\gamma, \tau)^2}{\psi_1(\gamma, \tau)} \right) \right] \times \\ &\quad \left\{ (2\pi)^{1/2} \mu(\gamma, \tau) \left(2\Psi \left(\frac{\mu(\gamma, \tau)}{\nu} \right) - 1 \right) + 2\nu \cdot \exp \left(- \frac{\mu(\gamma, \tau)^2}{2\nu^2} \right) \right\}, \end{aligned} \quad (8.2.69)$$

where $\Psi(\cdot)$ denotes the standard normal function. Combining (8.2.66) and (8.2.69), $\pi(\gamma|\tau, \mathbf{y})$ is proportional to the integral with respect to r as

$$2\sqrt{\pi}\kappa_1(\gamma, \tau)\Psi\left[\zeta(\gamma, \tau)r^{1/2}\right] \cdot r^{a+n/2-1}\exp\left[-r\left(b+\frac{\Lambda(\tau)}{2}\right)\right] - \sqrt{\pi}\kappa_1(\gamma, \tau)r^{a+n/2-1} \times \\ \exp\left[-r\left(b+\frac{\Lambda(\tau)}{2}\right)\right] + \sqrt{2}\kappa_2(\gamma, \tau)r^{a+n/2-3/2}\exp\left[-\frac{\zeta(\gamma, \tau)^2r}{2}\right], \quad (8.2.70)$$

where

$$\kappa_1(\gamma, \tau) = \frac{\Gamma(a + \frac{n}{2})}{(b + \frac{\Lambda(\tau)}{2})^{a+n/2}} \cdot \frac{\psi_2(\gamma, \tau)}{\psi_1(\gamma, \tau)^{3/2}} |\Sigma(\theta)|^{1/2} \exp\left[-\left(\frac{r}{2}\right)\left(\psi_3(\gamma, \tau) - \frac{\psi_2(\gamma, \tau)^2}{\psi_1(\gamma, \tau)}\right)\right], \\ \kappa_2(\gamma, \tau) = \kappa_1(\gamma, \tau) \cdot \frac{\psi_1(\gamma, \tau)^{1/2}}{\psi_2(\gamma, \tau)}, \\ \zeta(\gamma, \tau) = \frac{\psi_2(\gamma, \tau)}{\psi_1(\gamma, \tau)^{1/2}}. \quad (8.2.71)$$

Let κ_i, ζ denote $\kappa_i(\gamma, \tau)$ and $\zeta(\gamma, \tau)$ respectively for convenience. For $a_1 = (a + n/2)$, $b_1 = (b + \Lambda(\tau)/2)$, the integral of first term in the above over r is

$$Q = \kappa_1 \int_0^\infty r^{a_1-1} \exp(-b_1 r) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\zeta r^{1/2}} \exp\left(-\frac{z^2}{2}\right) dz \right) dr, \quad (8.2.72)$$

after letting $y = z/(\zeta r^{1/2})$ and changing the order of integration,

$$Q = \frac{\kappa_1}{\sqrt{2\pi}} \int_{-\infty}^1 \zeta \left\{ \int_0^\infty r^{a_1-1/2} \exp\left[-r\left(b_1 + \frac{\zeta^2 y^2}{2}\right)\right] dr \right\} dy \\ = \frac{\kappa_1}{\sqrt{2\pi}} \cdot \Gamma\left(a_1 + \frac{1}{2}\right) b_1^{-(a_1+1/2)} \zeta \int_{-\infty}^1 \left(1 + \frac{\zeta^2 y^2}{2b_1}\right)^{-(a_1+1/2)} dy, \quad (8.2.73)$$

again let $x = \zeta[(a_1 - 1)/b_1]^{1/2}y$, then

$$Q = \frac{\kappa_1}{\sqrt{2\pi}} \cdot \Gamma\left(a_1 + \frac{1}{2}\right) (a_1 - 1)^{-1/2} b_1^{-a_1} \int_{-\infty}^{\zeta[(a_1-1)/b_1]^{1/2}} \left(1 + \frac{x^2}{2(a_1 - 1)}\right)^{-[2(a_1-1)+1]} dx \\ \propto \kappa_1 \frac{\Gamma(a_1 - 1)}{b_1^{a_1}} \cdot t_{2(a_1-1)}\left(\zeta[(a_1 - 1)/b_1]^{1/2}\right), \quad (8.2.74)$$

where $t_{\lambda(\cdot)}$ denotes t -distribution function with λ degree of freedom. The posterior density of γ given τ, \mathbf{y} is, therefore

$$p(\gamma|\tau, \mathbf{y}) \propto \kappa_1 \Gamma(a_1 - 1) b_1^{-a_1} \cdot t_{2(a_1-1)}\left(\zeta[(a_1 - 1)/b_1]^{1/2}\right) - \kappa_1 \Gamma(a_1) b_1^{-a_1} \\ + \kappa_2 \Gamma\left(a_1 - \frac{1}{2}\right) \zeta^{-2(a_1-1/2)}, \quad (8.2.75)$$

and from (8.2.67) we can obtain $p(\gamma, \tau|\mathbf{y})$ by multiplying (8.2.65) and (8.2.75). This results are used to calculate $p(I(C)|\mathbf{y})$ and then $p(\gamma|C, \mathbf{y})$ can be obtained by calculating the terms in equation (8.2.67). In addition, we can have the posterior density of τ given C, \mathbf{y} as

$$\begin{aligned} p(\tau|C, \mathbf{y}) &= \int p(\gamma, \tau|C, \mathbf{y})d\gamma \\ &= \int_{x_\tau}^{x_{\tau+1}} p(\gamma, \tau|\mathbf{y})d\gamma / \sum_{\tau} \int_{x_\tau}^{x_{\tau+1}} p(\gamma, \tau|\mathbf{y})d\gamma. \end{aligned} \quad (8.2.76)$$

8.2.2 Bayesian Change-point Test

In this section we discuss the null and alternative hypothesis for two-phase regression and provide the test statistics using Bayesian approach, which will be called ‘‘Bayesian change-point test (BCT)’’ hereafter. Under the null hypothesis of no change-point, the model is

$$H_0: \quad \mathbf{Y} = \mathbf{X}\beta_0 + \mathbf{e}_0, \quad (8.2.77)$$

where \mathbf{Y} is $(n \times 1)$ vector of the response, \mathbf{X} is $(n \times p)$ matrix of predictors, β_0 is $(p \times 1)$ vector of coefficient parameters, and $\mathbf{e}_0 \sim \mathcal{N}(0, \sigma_0^2 \mathbf{I}_n)$ is $(n \times 1)$ random error vector. Under the alternative hypothesis of one change with change-point τ , the model is

$$H_1: \quad \mathbf{Y} = \mathbf{X}(\tau)\beta + \mathbf{e}_1, \quad (8.2.78)$$

where all the parameters \mathbf{Y} , $\mathbf{X}(\tau)$, and β were same as those defined in section 2. We denote $\mathbf{e}_1 \sim \mathcal{N}(0, \sigma_1^2 \mathbf{I}_n)$ as random errors under H_1 . Under H_0 and H_1 , the likelihood function is

$$\mathcal{L}(\beta_0, \sigma_0^2, H_0) \propto (\sigma_0^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_0^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right\},$$

and

$$\begin{aligned} \mathcal{L}(\beta_0, \sigma_1^2, \tau, H_1) &\propto (\sigma_1^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_1^2}(\mathbf{y} - \mathbf{X}(\tau)\beta)'(\mathbf{y} - \mathbf{X}(\tau)\beta)\right\} \\ &= (\sigma_1^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_1^2}[(\mathbf{y}_1 - \mathbf{X}_1(\tau)\beta_1)'(\mathbf{y}_1 - \mathbf{X}_1(\tau)\beta_1) \right. \\ &\quad \left. + (\mathbf{y}_2 - \mathbf{X}_2(\tau)\beta_2)'(\mathbf{y}_2 - \mathbf{X}_2(\tau)\beta_2)]\right\}, \end{aligned}$$

respectively. Let the prior distributions of the null and alternative hypothesis are $\pi_0(H_0) = p_0$ and $\pi_0(H_1) = p_1$ for each. Additionally, we assume that $\beta_0|\tau_0 \sim \mathcal{N}(\beta_0^*, \tau_0^{-1}\Sigma_0)$ and

$r_0 = 1/\sigma_0^2 \sim G(a_0, b_0)$. The joint probability density of \mathbf{Y} and H_0 is

$$\begin{aligned} f(\mathbf{y}, H_0) &= p_0 f(\mathbf{y}|H_0) \\ &= p_0 \int \mathcal{L}(\beta_0, r_0, H_0) \pi_0(\beta_0|r_0, H_0) \pi_0(r_0|H_0) d\beta_0 dr_0 \\ &\propto \frac{p_0}{|\Sigma_0|^{1/2} |\Sigma_{(p)}|^{1/2}} \cdot \frac{\Gamma(b_0 + \Lambda_0/2) \cdot b_0^{a_0}}{\Gamma(b_0) \cdot (b_0 + \Lambda_0/2)^{a_0+n/2}}, \end{aligned}$$

where $\Sigma_{(p)} = \mathbf{X}'\mathbf{X} + \Sigma_0^{-1}$ and Λ_0 corresponds to all the components in $\Lambda(\tau)$ in section 2 except using subscript "0" for denoting the model parameters under H_0 . Similarly, The joint probability density of \mathbf{Y} and H_1 is

$$f(\mathbf{y}, H_1) \propto \sum_{\tau=2}^{n-2} \frac{p_1}{n-3} \cdot \frac{\Gamma(b + \Lambda(\tau)/2)}{|\Sigma|^{1/2} |\Sigma_{(\tau)}|^{1/2} \Gamma(b)} \cdot \frac{b^a}{(b + \Lambda(\tau)/2)^{a+n/2}}.$$

By Bayes theorem, the posterior distribution of hypothesis is

$$p(H_i|\mathbf{y}) = \frac{p_i f(\mathbf{y}|H_i)}{p_0 f(\mathbf{y}, H_0) + p_1 f(\mathbf{y}, H_1)},$$

for $i = 0$ and 1 . Using posterior odds in Zellner (1987) as a test of two-phase regression,

$$K_{01} = \frac{p(H_0|\mathbf{y})}{p(H_1|\mathbf{y})} = \frac{p_0}{p_1} \cdot \frac{f(\mathbf{y}|H_0)}{f(\mathbf{y}|H_1)} = \frac{f(\mathbf{y}, H_0)}{f(\mathbf{y}, H_1)},$$

that is, the ratio of joint probability density of \mathbf{Y} and H_0 to that of \mathbf{Y} and H_1 . The result for K_{01} is

$$K_{01} = C_0 \cdot \sum_{\tau=2}^{n-2} \frac{|\Sigma_{(\tau)}|^{1/2} (b + \Lambda(\tau)/2)^{a+n/2}}{\Gamma(b + \Lambda(\tau)/2)},$$

where C_0 is constant as

$$C_0 = \frac{p_0}{p_1} \cdot \frac{(n-3) |\Sigma|^{1/2}}{|\Sigma_0|^{1/2} |\Sigma_{(p)}|^{1/2}} \cdot \frac{b_0^{a_0} \Gamma(b) \Gamma(b_0 + \Lambda_0/2)}{b^a \Gamma(b_0) (b_0 + \Lambda_0/2)^{a_0+n/2}}.$$

In the case of testing the hypothesis of two-phase regression, the null hypothesis of no change-point can be rejected if K_{01} is smaller than the critical value.

8.3 Relating Degradation to Failure Time

The failure-time distribution in the two-phase degradation model as (8.1.55) is

$$F_T(t) = [Pr(\alpha_1 + \beta_1 t \leq \mathcal{D}_f) - Pr(\alpha_2 + \beta_2 t \leq \mathcal{D}_f)] \cdot I_{[0, \gamma]}(t) + Pr(\alpha_2 + \beta_2 t \leq \mathcal{D}_f) \cdot I_{[0, \infty)}(t), \quad (8.3.79)$$

where $I(t)$ denotes the indicator function. Because (α_i, β_i) can vary from unit to unit, we can represent them as random coefficients, and (8.3.79) becomes a (linear) random coefficients model.

8.3.1 Random Coefficients Model for Change-Point Regression Model

Under the assumption that the intercept and slope coefficients in the degradation path vary between test items, the response vector \mathbf{y}_i for i^{th} subject is

$$\mathbf{y}_i = \mathcal{D}(t; \Theta) = \mathbf{X}(\tau_i)\boldsymbol{\mu} + \mathbf{X}(\tau_i)\boldsymbol{\vartheta}_i + \mathbf{e}_i, \quad i = 1, \dots, M, \quad (8.3.80)$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iN_i})$. The vector $\boldsymbol{\mu} = (\mu_{\alpha_1}, \mu_{\beta_1}, \mu_{\alpha_2}, \mu_{\beta_2})'$ denotes the fixed effects, the random coefficient $\boldsymbol{\vartheta}_i = (\vartheta_{\alpha_1 i}, \vartheta_{\beta_1 i}, \vartheta_{\alpha_2 i}, \vartheta_{\beta_2 i})'$ characterizes the between-item variation, and τ_i represents a change-point for i^{th} individual. Fixed ($\boldsymbol{\mu}$) and random ($\boldsymbol{\vartheta}_i$) effects are divided into two parts according to the estimate of τ_i , that is, $(\mu_{\alpha_1}, \mu_{\beta_1})$, $(\vartheta_{\alpha_1 i}, \vartheta_{\beta_1 i})$ and $(\mu_{\alpha_2}, \mu_{\beta_2})$, $(\vartheta_{\alpha_2 i}, \vartheta_{\beta_2 i})$, denoting the parameter sets before and after the change-point τ_i , respectively. Therefore, the $(N_i \times 4)$ covariate matrix $\mathbf{X}(\tau_i)$ is of the form

$$\mathbf{X}(\tau_i) = \begin{bmatrix} \mathbf{T}_1(\tau_i) & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{T}_2(\tau_i) \end{bmatrix},$$

where

$$\mathbf{T}_1(\tau_i) = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_{\tau_i} \end{bmatrix}', \quad \mathbf{T}_2(\tau_i) = \begin{bmatrix} 1 & \cdots & 1 \\ t_{\tau_i+1} & \cdots & t_{N_i} \end{bmatrix}'$$

and $\mathbf{0}_1, \mathbf{0}_2$ are $(\tau_i \times 2)$ and $((N_i - \tau_i) \times 2)$ zero matrices, respectively. We assume that the error term \mathbf{e}_i is iid $\mathcal{N}(\mathbf{0}, \Omega_i)$, where $\Omega_i = (\sigma_1^2 \mathbf{I}_{\tau_i}, \sigma_2^2 \mathbf{I}_{N_i - \tau_i})'$, the random effects vector $\boldsymbol{\vartheta}_i$ has independent components and has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ , with $\boldsymbol{\vartheta}_i$ independent of \mathbf{e}_i . If $(\vartheta_{\alpha_1}, \vartheta_{\beta_1})'$ and $(\vartheta_{\alpha_2}, \vartheta_{\beta_2})'$ are independent of each other, then on the condition that τ_i is known, Σ can be written as

$$\Sigma = \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix},$$

where $\Sigma_k, k = 1, 2$ denotes the (2×2) covariance matrix with elements $\Sigma_{00,k} = \delta_{0,k}^2$, $\Sigma_{11,k} = \delta_{1,k}^2$ and $\Sigma_{01,k} = \Sigma_{10,k} = \rho_k \delta_{1,k} \delta_{2,k}$. In the last section, it was noted that the

individual MLEs $(\hat{\alpha}_{1i}, \hat{\beta}_{1i})'$ and $(\hat{\alpha}_{2i}, \hat{\beta}_{2i})'$ are independent and (asymptotically) bivariate normally distributed. The diagonal blocked form of Σ above is, hence, reasonably defined. The independence between $(\vartheta_{\alpha_1}, \vartheta_{\beta_1})'$ and $(\vartheta_{\alpha_2}, \vartheta_{\beta_2})'$ makes the computation of $F_T(t)$ straightforward because we can treat the random coefficients $(\vartheta_{\alpha_1}, \vartheta_{\beta_1})'$ and $(\vartheta_{\alpha_2}, \vartheta_{\beta_2})'$ independently. To estimate all the parameters, including random coefficients, we suggest the following two-stage approach:

Stage 1: For each degradation path, use the MLE procedure in Section 3 to estimate the two-phase model parameters and obtain $\hat{\theta}_i = (\hat{\theta}_{1i}, \hat{\theta}_{2i})'$, $i = 1, \dots, M$.

Stage 2: Use the average of the individual $\hat{\theta}_i$ as the final fixed-effect parameter; for the random coefficients, treat the path estimates as “pseudo-data” and estimate the parameters in a multivariate normal distribution.

In stage 2, for the fixed-effect estimate $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)'$, $\hat{\mu}_k = (\sum_i^M \hat{\theta}_{ki})/M$, $k = 1, 2$. The random coefficients can be estimated using a moment estimation procedure; following the “variance decomposition” theory in Lu and Meeker [55], the block diagonal covariance estimate for the random coefficients is

$$\hat{\Sigma}_k = \frac{1}{M-1} \sum_{i=1}^M (\hat{\theta}_{ki} - \hat{\mu}_k)(\hat{\theta}_{ki} - \hat{\mu}_k)' - \frac{1}{M} \sum_{i=1}^M \text{Var}(\hat{\theta}_{ki}), \quad k = 1, 2, \quad (8.3.81)$$

where $\text{Var}(\hat{\theta}_{ki}) = \hat{\sigma}_i^2 [\mathbf{X}(\tau_i)' \mathbf{X}(\tau_i)]^{-1}$. Here $\hat{\sigma}_i^2 = [\hat{\tau}_i \hat{\sigma}_{1i}^2 + (N_i - \hat{\tau}_i) \hat{\sigma}_{2i}^2]/N_i$. Although $\hat{\Sigma}$ may be negative definite, Amemiya’s [3] procedure guarantees the nonnegative definite form of $\hat{\Sigma}$ by using a modified estimator (see Lu and Meeker [55] for details).

8.3.2 Failure-Time Distribution for Change-Point Regression Model

Using the fixed-effect and random coefficient estimates derived from the above two-stage approach, we can derive a failure-time distribution estimate for the degradation data. We assume that $(\vartheta_{\alpha_1}, \vartheta_{\beta_1})'$ and $(\vartheta_{\alpha_2}, \vartheta_{\beta_2})'$ are independent, hence the failure-time distribution for the two-phase degradation model can be derived easily by considering each phase in turn. If \mathcal{D}_f crosses the degradation path before the estimated change-point $\hat{\gamma}$, the MLE of

failure-time distribution is

$$\begin{aligned}\hat{F}_1(t) &= Pr \left[\frac{\mathcal{D}_f - (\hat{\mu}_{\alpha_1} + \vartheta_{\alpha_1})}{\hat{\mu}_{\beta_1} + \vartheta_{\beta_1}} \leq t \right] \\ &= \Phi \left\{ \frac{\mathcal{D}_f - [\hat{\mu}_{\alpha_1} + \hat{\mu}_{\beta_1} t]}{\sqrt{\hat{\delta}_{0,1}^2 + \hat{\delta}_{1,1}^2 t^2 + 2\hat{\rho}_1 \hat{\delta}_{0,1} \hat{\delta}_{1,1}}} \right\} \quad t \leq \hat{\gamma}.\end{aligned}\quad (8.3.82)$$

Similarly, $\hat{F}_2(t)$ can be derived in the case that \mathcal{D}_f and the degradation path cross after $\hat{\gamma}$. The distribution function (8.3.82) is considered as the generalized form of Bernstein's model (see Gertsbakh and Kordonskiy [32]) and discussed in Lu *et al.* [57]. The p^{th} quantile of the failure-time distribution (8.3.82) is the value of t_p such that $Pr(T \leq t_p) = p$, so that

$$\hat{z}_{1p} = \frac{\mathcal{D}_f - [\hat{\mu}_{\alpha_1} + \hat{\mu}_{\beta_1} t_p]}{\sqrt{\hat{\delta}_{0,1}^2 + \hat{\delta}_{1,1}^2 t_p^2 + 2t_p \hat{\rho}_1 \hat{\delta}_{0,1} \hat{\delta}_{1,1}}}, \quad (8.3.83)$$

where \hat{z}_{1p} is the p^{th} quantile of the standard normal distribution in (8.3.82).

The resulting failure-time distribution for the change-point degradation model is a mixture of the failure-time distributions for the first and the second phase regressions. Suppose the degradation path and \mathcal{D}_f crosses before $\hat{\gamma}$ with p probability and cross after $\hat{\gamma}$ with $(1 - p)$ probability. Then the failure-time distribution for the composite degradation paths is

$$\begin{aligned}\hat{F}_T(t) &= p\hat{F}(t|t \leq \hat{\gamma}) + (1 - p)\hat{F}(t|t > \hat{\gamma}) \\ &= p\hat{F}_1(\hat{\gamma}) + (1 - p)[\hat{F}_1(\hat{\gamma}) + \hat{F}_2(t) - \hat{F}_2(\hat{\gamma})] \\ &= \hat{F}_1(\hat{\gamma}) + (1 - p)[\hat{F}_2(t) - \hat{F}_2(\hat{\gamma})].\end{aligned}\quad (8.3.84)$$

From the observed degradation data, \hat{p} is simply the number of subjects which cross \mathcal{D}_f within $[t_1, t_\tau]$ over the total number of subjects.

Up to this point, we discussed a procedure to estimate the change-point in a nonmonotonic degradation path. Incorporating the parameter estimates in a change-point regression model for each component, a random coefficients model is established to derive the failure time distribution for all of the test items.

8.4 PDP Degradation

Plasma display panels (PDPs) are new self-emissive flat panel displays with excellent image quality and large screen sizes viewable in virtually any environment. PDPs are displayed by an array of cells called pixels, which are composed of three sub-pixels corresponding to the colors of red, green and blue. The operational principle of the plasma display is that an address electrode causes gas to change to a plasma state, then gas in plasma state reacts with phosphors in the discharge region, causing each sub-pixel to produce red, green, and blue light.

It is known that the display phosphor in a vacuum tube degrades exponentially. However, the impure gas produced during a manufacturing process (that was not burned out) disturbs how the plasma reacts with phosphor, which results in decreased luminosity during its initial life.

We focus on estimating the unknown time change-point within a degradation path rather than modeling a degradation path more fully for the purpose of providing information related to burn-in in the manufacturing process.

8.4.1 Change-Point Analysis

To model the degradation paths of the PDPs, for each component, we individually fitted a linear regression model for (log) transformed luminosity. In Figure 16, the simple linear fit (dashed line) consistently underestimates the true luminosity at the latter part of degradation testing by failing to reflect the nonmonotonic characteristics of the PDP degradation paths. As a result, the failure times in the linear model are generally underestimated.

To capture the nonmonotonicity of the PDP, a two-phase regression model was fitted to the six PDP degradation paths. The parameters of the two-phase model were estimated by maximum likelihood. Following the procedures from (8.1.57) to (8.1.59), the log-likelihood $\mathcal{L}(\gamma)$ was evaluated sequentially, using different values of τ .

The resulting likelihood values for six PDP degradation paths are plotted against measurement points ($n = 1, \dots, N$) in Figure 17. The log-likelihood values reach at the maximum around 115 hours for all of the PDPs. However, the MLE $\hat{\gamma} = (\hat{\alpha}_1 - \hat{\alpha}_2)/(\hat{\beta}_2 - \hat{\beta}_1)$

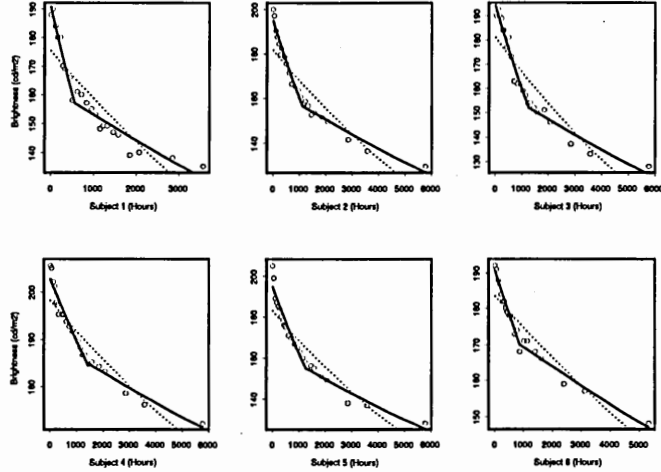


Figure 16: Two-phase regression fit for an individual PDP

must satisfy the continuity constraint $t_{\hat{\tau}} \leq \hat{\gamma} < t_{\hat{\tau}+1}$. Consequently, $\hat{\gamma}$ is a unique *local* maximum within that interval.

The MLEs of change-point ($\hat{\gamma}$) and parameters of two-phase model are summarized (with corresponding standard errors in parentheses) in Table 9. The standard errors of the two-phase model coefficients (θ) were derived from an asymptotic distribution of $\hat{\theta}$. The two-phase regressions (solid lines in Figure 16) provide a better fit to the PDP degradation path, especially the latter observations. In hypotheses test for $H_0 : \theta_1 = \theta_2$ against $H_1 : \theta_1 \neq \theta_2$, popular Chow's [18] F -statistic has p-value less than 0.01 for every six PDP degradation paths, strongly supporting the validity of two-phase regression model.

Next, we draw inference on the change-point γ , which is a parameter of primary interest in the PDP example. The AN confidence intervals, based on asymptotic normal (AN) proportion as the approximate F (AF) method, were constructed by inserting all ML estimates into (8.1.60). To derive the AF confidence intervals using $\hat{\gamma}$, we sequentially split the data into two parts using different values of τ and fit a separate regression line for each set of data, then solve the quadratic function (8.1.61) with respect to γ . If the lower bound is a negative, we follow the Hinkley's [40] sequential lower bound procedure. The resulting $100(1 - \delta)\%$ confidence intervals are summarized in Table 10. The AF confidence intervals are narrower than the AN intervals in the PDP example, but not symmetric about $\hat{\gamma}$. This

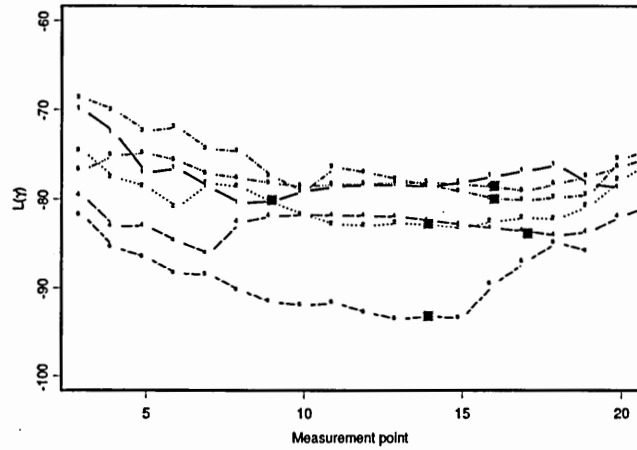


Figure 17: Likelihood plot as a function of τ . Square marks indicate $\hat{\tau}$'s locations corresponding to the local maxima.

phenomenon was also observed by Hinkley [40] and Siegmund and Zhang [82].

8.4.2 Failure-Time Analysis

In field testing, it is not uncommon to estimate the display lifetime by extrapolating the first phase of degradation. The resulting estimator can be dramatically biased, which is illustrated in the following analysis.

Using the two-stage approach in Section 7.3.1, all the fixed and random effects were estimated for PDP degradation data. Taking an average along with the estimates in Table 9 from the change-point analysis at the first stage, the fixed effect estimate was calculated: $\hat{\boldsymbol{\mu}} = (\hat{\mu}_{\alpha_1}, \hat{\mu}_{\beta_1}, \hat{\mu}_{\alpha_2}, \hat{\mu}_{\beta_2})' = (5.275, -1.99 \times 10^{-4}, 5.122, -4.48 \times 10^{-5})'$. The error variances were estimated as $\hat{\sigma}_1^2 = 2.044 \times 10^{-4}$ and $\hat{\sigma}_2^2 = 2.937 \times 10^{-4}$. Following the procedure (8.3.81), we obtain the elements of covariance matrix $\hat{\boldsymbol{\Sigma}}$:

$$\hat{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} 6.80 \times 10^{-4} & 1.07 \times 10^{-6} \\ 1.07 \times 10^{-6} & 5.95 \times 10^{-9} \end{bmatrix} \text{ and } \hat{\boldsymbol{\Sigma}}_2 = \begin{bmatrix} 1.88 \times 10^{-3} & 2.68 \times 10^{-7} \\ 2.68 \times 10^{-7} & 6.21 \times 10^{-11} \end{bmatrix}.$$

The estimates of failure-time distribution $\hat{F}_T(t)$ and the quantile \hat{t}_p are then constructed by replacing the model parameters in (8.3.82) and (8.3.83), respectively, with their estimates derived from the two-stage approach.

Table 9: Maximum likelihood (ML) estimates and their standard errors for the parameters in two-phase regression model

Subject	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\sigma}_1^2$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\sigma}_2^2$	$\hat{\gamma}$
1	5.253 (0.011)	-3.5×10^{-4} (3.9×10^{-5})	1.81×10^{-4}	5.091 (0.011)	-6.0×10^{-5} (6.0×10^{-6})	4.74×10^{-4}	557.15
2	5.274 (0.006)	-1.98×10^{-4} (1.1×10^{-5})	1.82×10^{-4}	5.104 (0.009)	-4.5×10^{-5} (3.0×10^{-6})	2.19×10^{-4}	1107.50
3	5.271 (0.007)	-1.95×10^{-4} (1.1×10^{-5})	2.35×10^{-4}	5.080 (0.013)	-4.4×10^{-5} (4.0×10^{-6})	4.35×10^{-4}	1264.52
4	5.326 (0.006)	-1.37×10^{-4} (8.0×10^{-6})	2.35×10^{-4}	5.191 (0.012)	-4.0×10^{-5} (4.0×10^{-6})	1.71×10^{-4}	1387.63
5	5.273 (0.008)	-1.82×10^{-4} (1.2×10^{-5})	3.53×10^{-4}	5.103 (0.014)	-4.7×10^{-5} (5.0×10^{-6})	3.54×10^{-4}	1263.29
6	5.251 (0.004)	-1.32×10^{-4} (9.0×10^{-6})	4.06×10^{-5}	5.165 (0.005)	-3.3×10^{-5} (2.0×10^{-6})	1.09×10^{-4}	859.31

We considered an aging effect for $\hat{F}_T(t)$, which is a mixture of $\hat{F}_1(t)$ and $\hat{F}_2(t)$. To illustrate the role of the burn-in process, we varied the portion p of $\hat{F}_1(t)$ and plotted the $\hat{F}_T(t)$ with respect to the different p values. Figure 18 clearly shows the importance of a burn-in process. The product reliability can be greatly improved by aging completely ($p = 0$) so that the brightness degradation follows an inherent degradation mechanism. If we evaluate the reliability largely based on the degradation paths censored before the change-point $\hat{\gamma}$, we are prone to underestimate the reliability.

Next, we compare the empirical failure-time distribution with two parametric failure-time distributions: Weibull and lognormal. The empirical failure-time distribution is generated using Monte Carlo simulation with $N = 50,000$. To introduce Monte Carlo evaluation of $F_T(t)$, we first use the two-phase model parameter estimates $\hat{\mu}$, $\hat{\Omega} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2)'$, and $\hat{\Sigma}$ obtained from the six sample paths to generate the N simulated realizations $\tilde{\mu}$.

Table 10: $100(1 - \delta)\%$ confidence intervals of $\hat{\gamma}$

Subject	AN			AF		
	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
1	(466.0, 648.3)	(448.5, 665.8)	(414.4, 699.9)	(535.6, 594.1)	(507.9, 612.6)	(404.9, 668.9)
2	(997.5, 1,217.5)	(976.5, 1,238.5)	(935.3, 1,279.7)	(966.5, 1,135.1)	(902.1, 1,146.7)	(841.1, 1,209.8)
3	(1,129.9, 1,399.1)	(1,104.2, 1,424.9)	(1,053.8, 1,475.2)	(1,206.3, 1,297.0)	(1,152.4, 1,312.2)	(1,090.0, 1,423.3)
4	(1,203.9, 1,571.4)	(1,168.7, 1,606.6)	(1,100.0, 1,675.3)	(1,343.8, 1,444.0)	(1,289.3, 1,470.3)	(1,143.9, 1,553.2)
5	(1,100.7, 1,425.9)	(1,069.5, 1,457.0)	(1,008.7, 1,517.9)	(1,187.2, 1,435.1)	(1,105.6, 1,459.2)	(1,062.2, 1,527.5)
6	(754.2, 964.4)	(734.1, 984.6)	(694.8, 1,023.9)	(796.6, 892.5)	(776.0, 902.3)	(695.9, 1,044.7)

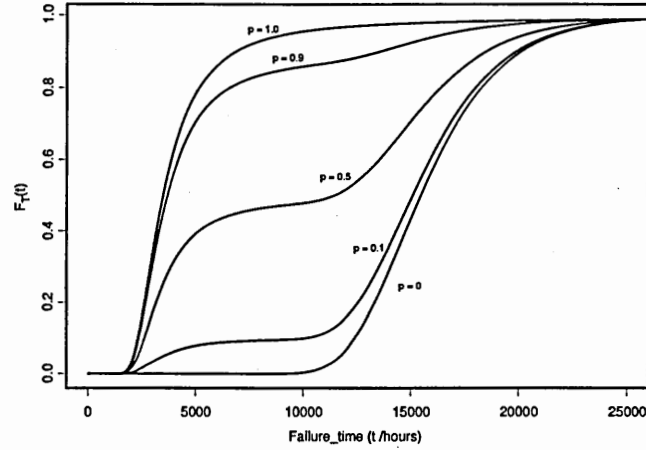


Figure 18: The mixture of failure-time distributions for the first and the second phase degradation paths

The change-point estimate $\hat{\gamma}$ need not be considered in the simulation since it is determined by $\hat{\mu}$. From N values of $\tilde{\mu}$, we compute the N failure times \tilde{t} which are cross-times between the realized degradation paths and \mathcal{D}_f . For any desired values of t , $F_T(t)$ is estimated from the simulated empirical distribution

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

Two empirical distributions were generated by following the above Monte Carlo simulation procedure. One is based on the simple regression fit and the other on the two-phase regression fit for 6 PDP degradation paths. Additionally, the parametric failure-time distributions were derived from the PDP failure-time data. Following the approximation method in Meeker and Escobar ([64], chapter 13.9), we predict “pseudo failure times” for the six degradation paths individually, and then obtain the ML estimates by fitting the six failure-times to Weibull and Lognormal distributions.

Because PDP degradation paths are relatively simple and the two-phase model is approximately correct, the approximation method is adequate for this case. The ML estimates of the parameters of the Weibull distribution are $\hat{\eta} = 13448.08$ (1421.28), $\hat{\zeta} = 4.1129$ (1.2131) and for the lognormal distribution the estimates are $\hat{\mu} = 9.387$ (0.0919) and $\hat{\sigma} = 0.225$ (0.0650) with the standard errors in parentheses. We estimate $F_T(t)$ based on these ML

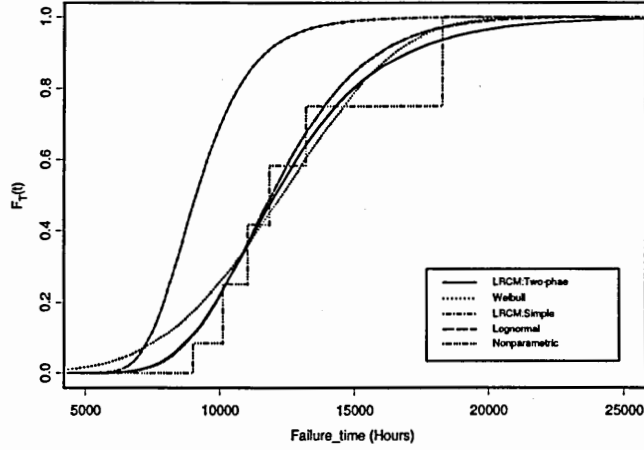


Figure 19: Comparison of empirical and parametric failure-time distributions: PDP example

estimates.

The empirical and parametric failure-time distributions are displayed along with the nonparametric estimate of $F_T(t)$ in Figure 19. The nonparametric estimate can be calculated as $\hat{F}_T^{NP}(t) = [(number\ of\ failures\ up\ to\ t) - 0.5]/6$. Figure 19 shows that the failure-time distributions derived from the two-phase regression and lognormal model provide a good fit to the nonparametric estimate $\hat{F}_T^{NP}(t)$. As mentioned before, the lifetime estimates based on the standard regression fit are much shorter than those derived from two-phase regression model.

For comparison, the point estimates and $100(1-\delta)\%$ confidence intervals of p^{th} quantiles of failure-time distribution were derived. The intervals are based on the simulation procedure in section 6.2.2 for constructing parametric bootstrap confidence intervals of p^{th} quantiles for two empirical failure-time distributions. In this case $B = 4,000$ bootstrap samples were generated. The confidence intervals of the Weibull and lognormal quantiles were constructed using the parametric bootstrap sampling procedure (with $B = 10,000$) in Meeker and Escobar ([64], Chapter 9.2.2). As the distribution of (log) transformed p^{th} quantile estimate ($\log \hat{t}_p$) has better properties for bootstrap confidence intervals rather than \hat{t}_p , we calculated the confidence intervals from the bootstrap samples of $\log \hat{t}_p$. The resulting

Table 11: Quantiles and their 95% bootstrap confidence intervals: PDP example

	Quantile		
	0.1	0.5	0.9
LRCM:Simple	6,751.9 (5,971.7, 6,923.3)	8,221.1 (6,880.9 , 10,246.0)	10,378.9 (9,882.3 , 13,132.2)
LRCM:Two-phase	8,901.2 (7,508.9 , 10,555.6)	12,054.8 (10,398.6 , 13,298.5)	16,992.6 (14,857.5 , 18,539.2)
Weibull	7,781.1 (2,728.1 , 8,908.0)	12,301.5 (9,351.7 , 19,988.5)	16,471.3 (13,756.2 , 62,606.3)
Lognormal	8,937.1 (5,587.9 , 9,974.7)	11,926.3 (9,762.3 , 17,704.5)	15,915.3 (13,030.3 , 42,013.7)

point estimates and 95% bootstrap confidence intervals of p^{th} quantiles are summarized in Table 11, for $p = 0.1, 0.5,$ and 0.9 .

The confidence intervals of two empirical quantiles are consistently shorter than those calculated on the basis of the Weibull and lognormal parameter estimates. We observed that the distributions of $\log \hat{t}_p$ as well as \hat{t}_p were highly skewed in the Weibull and lognormal distributions. Consequently, the resulting bootstrap confidence intervals were not symmetric about the corresponding point estimates of the quantiles and greatly biased to one limit.

8.5 VFD Degradation

We re-analyze the VFD data used in nonlinear random coefficients modeling with change-point analysis. Because of incomplete burn-in in the VFD manufacturing, the degradation path of the luminosity is unstable during its earlier stage as shown in Figure 11. We introduce a two-phase regression model in order to reflect the nonmonotonic characteristic of the VFD degradation.

8.5.1 Change-Point Analysis

For estimating a change-point in the degradation path, a two-phase regression model was individually fitted to five nonmonotonic degradation paths of the VFDs. Every degradation path includes six observations measured at identical times. Following the general change-point estimation procedure in the continuous case, change-point γ is determined as 1,000

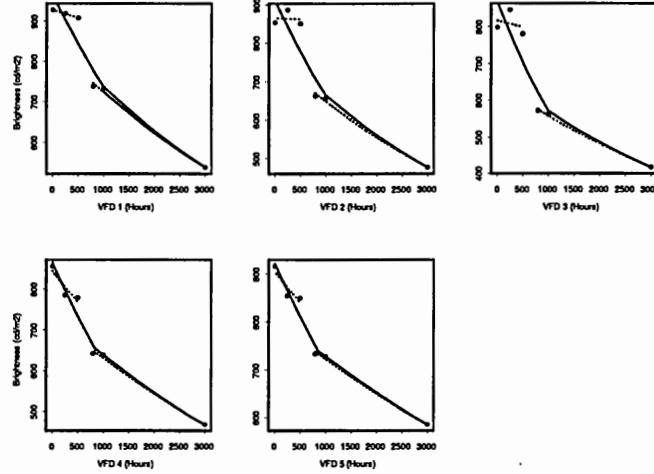


Figure 20: Two-phase regression fit for an individual VFD

hours when $t_{\hat{\tau}} = 1,000$ and $t_{\hat{\tau}+1} = t_N = 3,000$ hours so that a two-phase regression model can be applied.

Under the continuity constraint, all the change-points were estimated around 1,000 hours. However, the continuous two-phase model (solid line Figure 20) poorly fits to the earlier observations of the VFD's degradation paths. Without the continuity constraint, the two-phase regression model provides a better fit to the VFD degradation path (dashed line in Figure 20). Nevertheless, the discontinuous two-phase regression is inadequate for modeling the degradation path representing a continuous deterioration status of the performance. This modeling problem is largely due to the fact that the measurements are sparse in the VFD degradation path.

To investigate how the estimation of change-points is changed in the case of large and equally spaced measurements, we generate “pseudo” data from the nonlinear random coefficients (NRC) model introduced in section 6.3.1 to capture the nonmonotonic characteristic of VFD degradation paths:

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

with the fixed effects $\beta = (\beta_1, \beta_2, \beta_3, \beta_4) = (900.02, 0.081, 0.062, 0.3828)$, the random effects $b_{i1} \sim \mathcal{N}(0, 10.0056)$, $b_{i2} \sim \mathcal{N}(0, 8.411 \times 10^{-6})$, and the random error within the subject

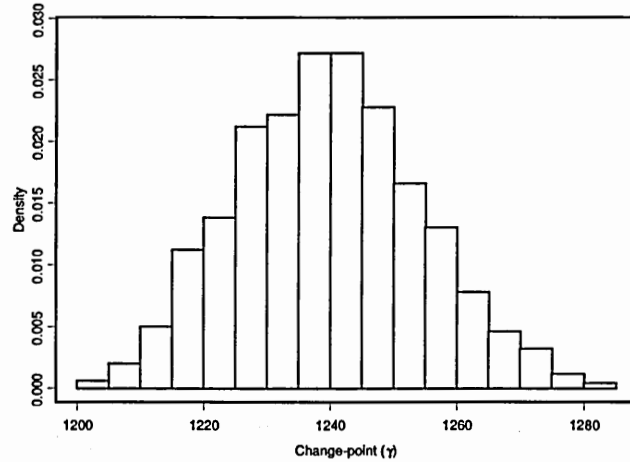


Figure 21: The distribution of change-points mean: VFD example

$$e_{ij} \sim \mathcal{N}(0, 100.10).$$

We simulated five degradation paths based on the parameter values of the above NRC model $N = 1,000$ times. The measurements were taken equally spaced at every 100 hours up to 3,000 hours. A two-phase regression model was refitted to the simulated degradation paths in order to estimate a change-point for each individual degradation path. We averaged the five change-point estimates for $N = 1000$ simulated runs. The distribution of this average is presented in Figure 21.

The change-point mean is 1,240.0 hours and its 90% and 95% confidence intervals are (1,216.1 , 1,263.9) and (1,211.5 , 1,268.5). All of change-points estimated from the NRC model lie in outside the field testing duration (1,000 hours). However, the change-points estimated from the constrained two-phase model were distributed around 1,000 hours. With no measurements between 1,000 and 3,000 hours, we can not effectively estimate the exact change-point location with the constrained two-phase model when the change-point extends beyond 1,000 hours. The change-points derived from two different models (NRC and constrained two-phased regression model) are considered as identically distributed.

Summarizing this change-point analysis for the VFDs, in field testing the VFD reliability is evaluated mainly by extrapolating the degradation path caused by incomplete burn-in during manufacturing. As a result, the VFD reliability is significantly underestimated.

From a viewpoint of modeling, the constrained two-phase regression model fails to fit the earlier degradation adequately because the number of measurements in the degradation path are small. If more measurements are taken before 1,000 hours of testing, the constrained two-phase model can provide an improved fit to the nonmonotonic degradation path of the VFDs.

8.5.2 Failure-Time Analysis

For the VFD degradation paths, the two-stage approach proposed previously fails to effectively estimate the random effects and error variance in the second phase. Instead, the Lindstrom and Bates algorithm [54] can be used to estimate both fixed and random effects by incorporating all the measurements of every subject simultaneously. It provides improved estimates for modest samples of repeated measurements, such as the VFD example.

The resulting fixed effects estimate $\hat{\mu} = (6.808, -3.02 \times 10^{-4}, 6.632, -1.44 \times 10^{-4})'$ with error estimates $\hat{\sigma}_1^2 = 5.25 \times 10^{-5}$ and $\hat{\sigma}_2^2 = 1.47 \times 10^{-2}$. The elements of covariance matrix $\hat{\Sigma}$ are

$$\hat{\Sigma}_1 = \begin{bmatrix} 6.81 \times 10^{-4} & -7.23 \times 10^{-7} \\ -7.23 \times 10^{-7} & 2.76 \times 10^{-9} \end{bmatrix} \text{ and } \hat{\Sigma}_2 = \begin{bmatrix} 9.85 \times 10^{-3} & 3.78 \times 10^{-7} \\ 3.78 \times 10^{-7} & 4.11 \times 10^{-10} \end{bmatrix}.$$

As in the PDP failure-time analysis, two empirical failure-time distributions were compared with the Weibull and lognormal failure-time distributions. The Weibull ML estimates are $\hat{\eta} = 4018.82$ (296.22), $\hat{\zeta} = 6.469$ (2.083) and lognormal ML estimates are $\hat{\mu} = 8.224$ (0.062) and $\hat{\sigma} = 0.138$ (0.044) (standard errors in parentheses).

To study the failure-time distributions, empirical distributions were generated using $N = 50,000$ Monte Carlo simulation. Assuming that early readings are largely due to lack of stability, the degradation measurements earlier than 240 hours (10 days) was discarded to fit a simple regression model, following the approach in Tseng *et al.* [87] and Chiao and Hamada [16], then the fixed and random effects were estimated. Four failure-time distributions are given along with the nonparametric estimate of $F_T(t)$ in Figure 22.

The p^{th} quantiles and their confidence intervals for the four distributions are also obtained using the same bootstrap procedures in the PDP example, and are given in Table

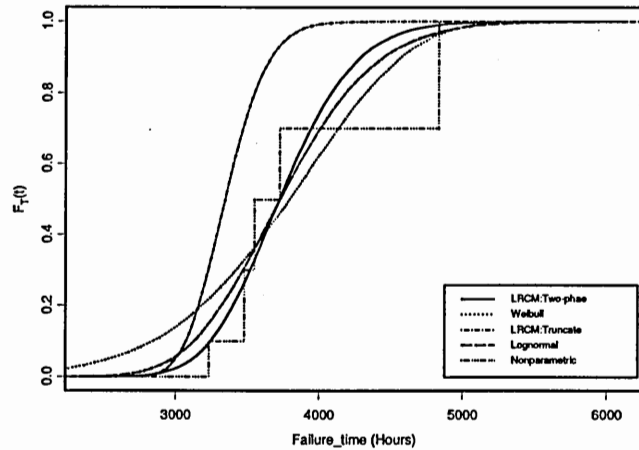


Figure 22: Comparison of empirical and parametric failure-time distributions: VFD example

12. The estimated lifetimes derived from the (truncated) simple regression are consistently shorter than those based on the two-phase regression fit. Because the arbitrary truncation point is not necessarily the true change-point caused by incomplete burn-in, the simple regression model underestimates the lifetimes by failing to correctly reflect the burn-in effects into the model. Moreover, the average relative prediction bias, that is the absolute difference between the real response and prediction values divided by the real response, is 0.0342 for the two-phase model and 0.0733 for (truncated) simple regression model. The two-phase model substantially improves the relative efficiency of prediction over the standard estimators based on (truncated) simple regression model.

As observed in the PDP example, the bootstrap confidence intervals for failure-time distribution quantiles are shorter than both lognormal and Weibull intervals (see Table 12). The analysis based on degradation data provides less conservative confidence intervals than the failure-time analysis when the number of failures is relatively small. The general characteristics between the degradation and failure-time analysis are discussed in Lu *et al.* [56].

Table 12: Quantiles and their 95% bootstrap confidence intervals: VFD example

	Quantile		
	0.1	0.5	0.9
LRCM:Truncate	3,073.9 (2,879.8 , 3,863.1)	3,346.4 (3,000.9 , 4,531.9)	3,665.5 (3,097.7 , 5,097.1)
LRCM:Two-phase	3,245.5 (2,977.1 , 3,359.3)	3,724.7 (3,345.7 , 4,193.1)	4,286.3 (4,103.0 , 4,734.3)
Weibull	2,838.0 (694.3, 3,012.0)	3,797.4 (3038.7 , 7,368.5)	4,571.9 (4,013.3 , 27,049.22)
Lognormal	3,124.7 (1,664.7 , 3,264.7)	3,727.5 (3,317.2 , 5,359.7)	4,446.6 (3,875.1 , 12,752.7)

CHAPTER IX

CONCLUDING REMARKS

We provide the motivation for using a nonlinear random coefficients model to describe a nonmonotonic degradation path of light displays' testing data. A nonlinear fit without random coefficients fails to reflect the variation between individuals. Estimates of the individual-item regression parameters are unstable with sparse data in the as VFD example, and consequently the two-stage estimation method is not well suited. We suggest an alternative method to estimate the parameters in the nonlinear random coefficient model more efficiently based on an approximate method that simultaneously uses all the individual measurements. Several approximation methods were applied to estimate fixed and random effects in the nonlinear random coefficient model and Laplace approximation was selected based on the comparison criteria. The failure-time distributions and quantiles, along with their bootstrap confidence intervals, were compared under two different modeling approaches: linear and nonlinear random coefficient model.

The VFD degradation analysis suggests that the nonlinear random coefficient model improves the relative efficiency of prediction over the standard estimators based on transformed linear model. The simulation results also confirm that the nonlinear random coefficient model be more robust under different model assumptions and various sampling schemes. The nonlinear random coefficient model provides more reliable results when inspection resources are more concentrated during first phase of VFD's life.

As another modeling approach, we introduced a linear degradation model with an unknown change-point to characterize the nonmonotonic degradation paths representing incomplete burn-in during the manufacturing process. The change-point model fits the nonmonotonic degradation data well, and provides valuable information about the burn-in process for light displays by detecting the transition time between the burn-in related

phase and the inherent degradation phase. Not only do models ignoring or failing to correctly capture the nonmonotonic characteristics lose the efficiency of prediction, but they also risk underestimating the true reliability. If an initial unstabilized degradation stage caused by incomplete burn-in can be eliminated by extending the burn-in time or using a higher stress, the reliability estimation can be greatly improved by extrapolating the lifetimes from inherent degradation paths which are degrading slowly. Consequently this will allow manufacturer to consider longer-term warranties to gain advantage in the market. With small sample sizes and unequally spaced measurements (as in the VFD example), fitting a nonmonotonic degradation model is challenging if not controversial. For the PDP degradation, with more degradation data, a continuous change-point model effectively fits the degradation path characterizing incomplete burn-in during the manufacturing.

An effective burn-in strategy that considers burn-in cost and reliability simultaneously can be developed for this kind of degradation data. Optimal burn-in policies have been based on failure-time data, but relatively undeveloped for analyses using degradation data. As another approach for analyzing a change-point, Bayesian methods are considered using the prior information accumulated from a field degradation testing.

CHAPTER X

RELIABILITY CHARACTERISTICS OF FAILURE-TIME DISTRIBUTION FOR DEGRADATION DATA

10.1 Degradation Process Model

We investigate reliability characteristics of a failure-time distribution resulted from what a related performance degrades.

10.1.1 Additive Degradation Model

First, we define a general additive degradation model

$$\mathcal{D}(t; X, \Theta) = \eta(t; \Theta) + X(t), \quad (10.1.85)$$

where $\eta(t; \Theta)$ is a deterministic mean degradation path with fixed effect parameters Θ for time $t \geq 0$. It may be monotonic or non-monotonic function of t , however we consider monotonic case only. $\{X(t) : t \geq 0\}$ represents a random variation process about mean degradation level $\eta(t; \Theta)$ and has a cumulative distribution function (Cdf) $F_{X(t)}$ and a probability density function (Pdf) $f_{X(t)}$. In a large number of cases, it will be sufficient to represent the degradation path in the form (10.1.85).

We assume failure occurs when degradation process reaches at pre-determined threshold value (\mathcal{D}_f). Then for monotone decreasing degradation path (MDDP) $\eta(t; \Theta)$, a failure is defined as the time that degradation level decreases below threshold, *i.e.*, $\mathcal{D}(t; X, \Theta) < \mathcal{D}_f$ and $\mathcal{D}(t; X, \Theta) > \mathcal{D}_f$ for monotone increasing degradation path (MIDP).

First, We will derive several reliability functions for a general degradation path. Let $G_{\mathcal{D}}(t)$ denote a failure-time distribution of the MDDP $\mathcal{D} = \{\mathcal{D}(t; X, \Theta); t \geq 0\}$. Then

$$\begin{aligned} G_{\mathcal{D}}(t) &= Pr[\mathcal{D}(t; X, \Theta) < \mathcal{D}_f] = Pr[\eta(t; \Theta) + X(t) < \mathcal{D}_f] \\ &= Pr[X(t) < \mathcal{D}_f - \eta(t; \Theta)] \\ &= F_{X(t)}(\mathcal{D}_f - \eta(t; \Theta)), \end{aligned} \quad (10.1.86)$$

with the survival function, $\bar{G}_D(t) = 1 - G_D(t)$. For the MIDP, a failure-time distribution $G_I(t) = 1 - F_{X(t)}(\mathcal{D}_f - \eta(t; \Theta))$. First, it must be checked that the distribution G is properly defined, that is, $\lim_{t \rightarrow 0^+} G(t) = 0$ and $\lim_{t \rightarrow \infty} G(t) = 1$.

The failure rate of the MDDP $\eta(t) = \eta(t; \Theta)$ is

$$\begin{aligned} r_D(t) &= \frac{g_D(t)}{1 - G_D(t)} = \frac{F'_{X(t)}(\mathcal{D}_f - \eta(t))}{1 - F_{X(t)}(\mathcal{D}_f - \eta(t))} \\ &= -\eta'(t) \cdot \frac{f_{X(t)}(\mathcal{D}_f - \eta(t))}{1 - F_{X(t)}(\mathcal{D}_f - \eta(t))} \\ &= -\eta'(t) \cdot r_{X(t)}(\mathcal{D}_f - \eta(t)), \end{aligned} \quad (10.1.87)$$

where $r_{X(t)}(\cdot)$ denotes the failure rate of $X(t)$.

Assuming that G is proper, $\eta'(t; \Theta) \leq 0$ for a continuous and monotonic decreasing degradation path $\eta(t; \Theta)$, consequently $r_D(t) \geq 0$ for $r_{X(t)}(\cdot) \geq 0$. The failure rate of \mathcal{D} is closely related to that of $X(t)$ and functional form of $\eta(t; \Theta)$. The failure rate of the MIDP

$$\begin{aligned} r_I(t) &= \frac{g_I(t)}{1 - G_I(t)} \\ &= \frac{F'_{X(t)}(\mathcal{D}_f - \eta(t))}{F_{X(t)}(\mathcal{D}_f - \eta(t))} \\ &= -\log [F_{X(t)}(\mathcal{D}_f - \eta(t))]'. \end{aligned} \quad (10.1.88)$$

The survival function of MDDP can be expressed as a form of failure rate:

$$\bar{G}_D(t) = \exp[-R_D(t)],$$

where $R_D(t) = \int_0^t r_D(x) dx$ is called a cumulative failure rate. Cumulative failure rate of the MDDP

$$R_D(t) = -\log \bar{G}_D(t) = -\log [1 - F_{X(t)}(\mathcal{D}_f - \eta(t))], \quad (10.1.89)$$

and

$$R_I(t) = -\log [F_{X(t)}(\mathcal{D}_f - \eta(t))], \quad (10.1.90)$$

for the MIDP.

10.1.2 Multiplicative Degradation Model

A general multiplicative degradation model is given by

$$\mathcal{D}(t; X, \Theta) = X(t) \cdot \eta(t; \Theta). \quad (10.1.91)$$

Then, a failure-time distribution of the MDDP

$$\begin{aligned}
 G_D(t) &= Pr[X(t) \cdot \eta(t) < \mathcal{D}_f] \\
 &= Pr \left[X(t) < \frac{\mathcal{D}_f}{\eta(t)} \right] \\
 &= F_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right),
 \end{aligned} \tag{10.1.92}$$

and for the MIDP, a failure-time distribution $G_I(t) = 1 - F_{X(t)}\left(\frac{\mathcal{D}_f}{\eta(t)}\right)$.

The failure rate of the MDDP is

$$\begin{aligned}
 r_D(t) &= \frac{F'_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right)}{1 - F_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right)} \\
 &= -\mathcal{D}_f \eta(t)^{-2} \eta'(t) \cdot \frac{f_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right)}{1 - F_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right)} \\
 &= - \left(\frac{\mathcal{D}_f}{\eta(t)} \right) \log[\eta(t)]' \cdot r_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right),
 \end{aligned} \tag{10.1.93}$$

and for MIDP,

$$\begin{aligned}
 r_I(t) &= - \frac{F'_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right)}{F_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right)} \\
 &= - \log \left[F_{X(t)} \left(\frac{\mathcal{D}_f}{\eta(t)} \right) \right]'.
 \end{aligned} \tag{10.1.94}$$

Sometimes, the multiplicative model can be exchanged to an additive model. For example, when $X(t)$ is a lognormal process in the multiplicative degradation model, $\log X(t)$ is a normal process in the additive degradation model.

In another case, a random effect can be included in the $\eta(t; \Theta)$ of the multiplicative degradation model. Then the degradation path \mathcal{D} is given by

$$\mathcal{D}(t; X, \Theta) = X(t) \cdot \eta(t; X(t), \Theta). \tag{10.1.95}$$

At that situation, it is difficult to derive explicit form of G and numerical method such as Monte Carlo simulation can be used.

10.2 $X(t)$ Is a Random Variable

Consider that $X(t)$ is a random variable, that is, $X(t) = X$ is independently identical distributed with a Cdf F_X . Using the relationship between a distributional assumption

for X and a deterministic path $\eta(t; \Theta)$, we will derive a failure-time distribution for the degradation path \mathcal{D} under several distributional assumptions with respect to F_X .

10.2.1 Exponential Distribution

Suppose that X is a random effect which follows an exponential distribution with a parameter λ . then a failure-time distribution of \mathcal{D} for MDDP $\eta(t)$ in the additive degradation model is

$$\begin{aligned} G_D(t) &= F_{X(t)}(\mathcal{D}_f - \eta(t)) \\ &= 1 - \exp[-\lambda(\mathcal{D}_f - \eta(t))], \quad t \geq 0 \end{aligned} \quad (10.2.96)$$

and for MIDP, a failure-time distribution $G_I(t) = \exp[-\lambda(\mathcal{D}_f - \eta(t))]$. To G_D be a cumulative distribution function, $\eta(t)$ should be defined so that $\eta(\infty) = -\infty$ and $\eta(0) = \mathcal{D}_f$. On the other hand, G_I is not a distribution for exponential random variable X .

In the multiplicative degradation model, the failure-time distribution of \mathcal{D} is

$$G_D(t) = 1 - \exp \left[-\lambda \left(\frac{\mathcal{D}_f}{\eta(t)} \right) \right], \quad (10.2.97)$$

for MDDP, and for MIDP $\eta(t)$, $G_I(t) = \exp \left[-\lambda \left(\frac{\mathcal{D}_f}{\eta(t)} \right) \right]$.

For example, consider following monotone decreasing degradation path (MDDP)

$$\mathcal{D}(t; X, \Theta) = \theta_3 \cdot \exp[-\exp\{(\theta_1 t)^{\theta_2}\}], \quad \theta_1, \theta_2 > 0, \theta_3 \geq 1$$

and after logarithmic transformation

$$\log \mathcal{D}(t; X, \Theta) = \log \theta_3 - \exp\{(\theta_1 t)^{\theta_2}\}. \quad (10.2.98)$$

If $X = \log \theta_3 \geq 0$ is a random effect which follows an exponential distribution with a parameter λ , for $\theta_i, i = 1, 2$ satisfying $\eta(t) = -\exp\{(\theta_1 t)^{\theta_2}\} \leq \log \mathcal{D}$, a failure-time distribution of the additive degradation model (10.1.85) is

$$G_D(t) = 1 - \exp[-\lambda(\mathcal{D}_f + \exp\{(\theta_1 t)^{\theta_2}\})], \quad (10.2.99)$$

and the failure rate is

$$\begin{aligned} r_D(t) &= -\eta'(t) \cdot r_{X(t)}(\mathcal{D}_f - \eta(t)) \\ &= \lambda \theta_1 \theta_2 (\theta_1 t)^{\theta_2 - 1} \exp\{(\theta_1 t)^{\theta_2}\}. \end{aligned} \quad (10.2.100)$$

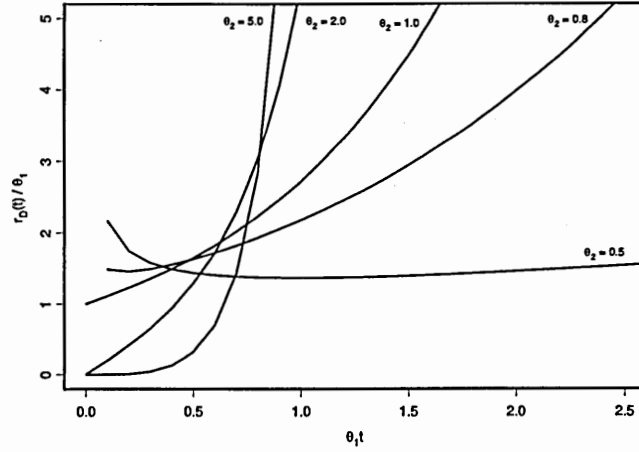


Figure 23: Failure rate plot of exponential power distribution

For a standard exponential distribution ($\lambda = 1$), if a failure is defined as $\mathcal{D}_f (= \log \mathcal{D}) = -1$, a survivor function is given by

$$\bar{G}_D(t) = \exp[1 - \exp\{(\theta_1 t)^{\theta_2}\}]. \quad (10.2.101)$$

This is a survivor function of a *exponential power* distribution. The exponential power distribution is one of the few tractable 2-parameter distributions which possess a bathtub-shaped failure rate. The failure rate is shown in Figure 23 for various values of θ_2 .

If $\theta_2 < 1$, its failure rate is bathtub shaped, achieving a minimum at $\{(1 - \theta_2)/(\theta_1 \theta_2)\}^{1/\theta_2}$. When $\theta_2 = 1$, exponential power distribution is reduced to an extreme value distribution (see Dhillon [22]).

Next, consider following monotone decreasing degradation path

$$\mathcal{D}(t; X, \Theta) = \theta_4 \cdot \{\log[\theta_1 t + \theta_2]\}^{-\theta_3}, \quad \theta_1, \theta_2, \theta_4 > 0, \theta_3 \geq 1. \quad (10.2.102)$$

If a random effect $X = \theta_4$ is exponentially distributed with the parameter λ , then a survivor function is given by

$$\begin{aligned} \bar{G}_D(t) &= \exp[-(\lambda \mathcal{D}_f) \cdot \eta(t)^{-1}] \\ &= \exp[-(\lambda \mathcal{D}_f) \cdot \{\log[\theta_1 t + \theta_2]\}^{\theta_3}]. \end{aligned} \quad (10.2.103)$$

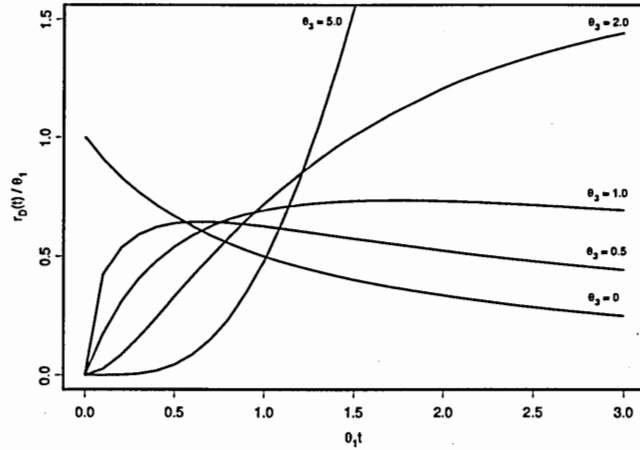


Figure 24: Failure rate plot of 2-parameter distribution II

Specially, for $\lambda D_f = 1$, $\theta_2 = 1$, and if θ_3 can be represented as $\tilde{\theta}_3 + 1$, where $\tilde{\theta}_3 \geq 0$, the survivor function

$$\bar{G}_D(t) = \exp[-\{\log(\theta_1 t + 1)\}^{\tilde{\theta}_3 + 1}]$$

is called *2-parameter distribution II* in Dhillon [22]. The failure rate of two-parameter distribution II

$$r_D(t) = \frac{\theta_1(\tilde{\theta}_3 + 1) \cdot (\log[\theta_1 t + 1])^{\tilde{\theta}_3}}{\theta_1 t + 1}$$

has increasing and decreasing failure rates, as depicted in Figure 24 at various values of $\tilde{\theta}_3$.

10.2.2 Log-logistic Distribution

The Pdf of a standard logistic distribution is

$$f_Y(y) = \frac{e^y}{(1 + e^y)^2}, \quad -\infty < x < \infty$$

and the Cdf is

$$F_Y(x) = (1 + e^y)^{-1}.$$

With the transformation $Y = \alpha + \beta \log X$ for $X > 0$, the Pdf of X is given by

$$f_X(x) = \frac{\beta e^{\alpha} x^{\beta-1}}{(1 + e^{\alpha} x^{\beta})^2}, \quad x > 0, \beta > 0, \quad (10.2.104)$$

and the Cdf

$$F_X(x) = 1 - \frac{1}{1 + e^{\alpha} x^{\beta}}. \quad (10.2.105)$$

This is called a *log-logistic* or a *Weibull-exponential* distribution in Dubey [23]. It is a special case of Burr's type XII family of distribution.

Consider following monotone decreasing degradation path

$$\mathcal{D}(t; X, \Theta) = \theta_3(\theta_1 t)^{-\theta_2}, \quad \theta_i > 0, \quad i = 1, 2, 3.$$

If a random effect $X = \theta_3$ follows the log-logistic distribution with the parameters α and β , then a failure-time distribution of \mathcal{D} is

$$\begin{aligned} G_D(t) &= 1 - \frac{1}{1 + e^{\alpha[\mathcal{D}_f(\theta_1 t)^{\theta_2}]^\beta}} \\ &= 1 - \frac{1}{1 + (e^\alpha \mathcal{D}_f^\beta \theta_1^{\theta_2 \beta}) t^{\theta_2 \beta}}. \end{aligned} \quad (10.2.106)$$

By fixing $\theta_2 \beta = 2$ and letting $b = e^{-\alpha/2} \mathcal{D}_f^{-1/\theta_1} \theta_1^{-1}$, the failure-time distribution in (10.2.106) is

$$G_D(t) = 1 - \left[1 + \left(\frac{t}{b} \right)^2 \right]^{-1}, \quad b > 0, t \geq 0. \quad (10.2.107)$$

This is the distribution of $b(F_{2,2}^{1/2})$, where $F_{2,2}$ denotes a (central) F random variable with (2,2) degrees of freedom. The Pdf of $Z = bF_{\nu_1, \nu_2}^a$ for $a, b > 0$ is derived by Malik [59] as

$$f_Z(z) = \frac{(\nu_1/\nu_2)^{\nu_1/2} (ab)^{-1} (z/b)^{\nu_1/(2a)-1}}{B(\nu_1/2, \nu_2/2) \{1 + (\nu_1/\nu_2)(z/b)^{1/a}\}^{(\nu_1+\nu_2)/2}},$$

where $B(\cdot)$ denotes a beta function.

The failure rate of \mathcal{D} can be obtained by

$$r_D(t) = \frac{2t}{b^2 + t^2}. \quad (10.2.108)$$

As shown in Figure 25, it has always an unimodal failure rate.

10.3 $X(t)$ is a Gaussian Process

First, we consider a stationary Gaussian stochastic process for $\{X(t) : t \geq 0\}$.

Definition 1 A stochastic process $\{X(t) : t \geq 0\}$ is called a *Gaussian* or a *normal process* if

$X(t_1), \dots, X(t_n)$ has a multivariate normal distribution for all t_1, \dots, t_n .

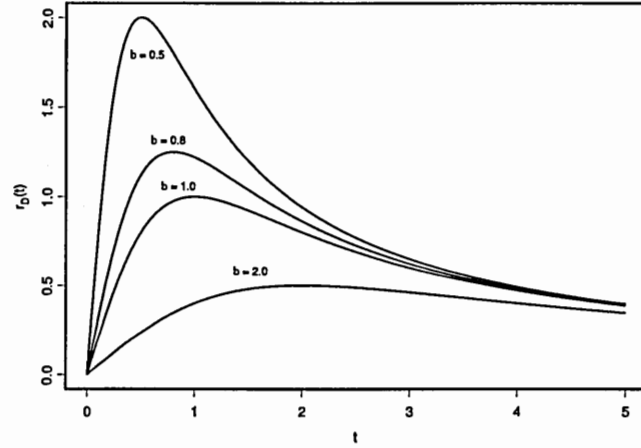


Figure 25: Failure rate plot of $bF_{2,2}^{1/2}$ distribution

Definition 2 A stochastic process $\{X(t) : t \geq 0\}$ is (weakly) stationary if

- (i) $E[X(t)] = \mu_X = \text{constant} < \infty$,
- (ii) $\text{Cov}[X(t), X(t + \tau)] = \gamma_X(t, t + \tau) = \gamma_X(\tau)$. That is, $\gamma_X(t, t + \tau)$ is independent of t for each $\tau \geq 0$.

The popular one of Gaussian process is a Wiener process.

Definition 3 A stochastic process $\{W(t) : t \geq 0\}$ is said to be a Wiener process if

- (i) $\text{Pr}[W(0) = 0] = 1$,
- (ii) $\{W(t) : t \geq 0\}$ has stationary and independent increments,
- (iii) for every $t > 0$, $W(t)$ is normally distributed with mean 0 and variance $\sigma_W^2 t$.

The covariance of the Wiener process is defined as:

$$\gamma_W(s, t) = \sigma_W^2 \cdot (s \wedge t),$$

where ' \wedge ' means the minimum value of the two. When $\sigma_W^2 = 1$, the process is a *standard Wiener process*.

We investigate the relationship between $G(t)$ and $F_X(t)$ for a Wiener Process $\{W(t) : t \geq 0\}$. In considering a Wiener process, we shall always assume that variance of $W(t)$ is $\sigma_w^2 t$.

For MIDP, a failure-time distribution of the degradation path

$$\begin{aligned} G_I(t) &= Pr[\eta(t) + X(t) > \mathcal{D}_f] \\ &= 1 - F_X(t)(\mathcal{D}_f - \eta(t)) \\ &= \frac{1}{\sqrt{2\pi t}\sigma_w} \int_{\mathcal{D}_f - \eta(t)}^{\infty} \exp\left[-\frac{x^2}{2t\sigma_w^2}\right] dx. \end{aligned} \quad (10.3.109)$$

By letting $y = x/\sqrt{t}\sigma_w$,

$$\begin{aligned} G_I(t) &= \frac{1}{\sqrt{2\pi}} \int_{\frac{\mathcal{D}_f - \eta(t)}{\sqrt{t}\sigma_w}}^{\infty} \exp\left[-\frac{y^2}{2}\right] dy \\ &= 1 - \Phi\left\{\frac{\mathcal{D}_f - \eta(t)}{\sqrt{t}\sigma_w}\right\}, \end{aligned} \quad (10.3.110)$$

where $\Phi\{\cdot\}$ denotes the Cdf of the standard normal distribution. If $\eta(t) = \theta t$ for $\theta > 0$, that is $\mathcal{D}(t)$ is of the form of the Wiener process with a linear ramp, then

$$G_I(t) = 1 - \Phi\left\{\frac{\mathcal{D}_f}{\sqrt{t}\sigma_w} - \frac{\theta\sqrt{t}}{\sigma_w}\right\}. \quad (10.3.111)$$

If

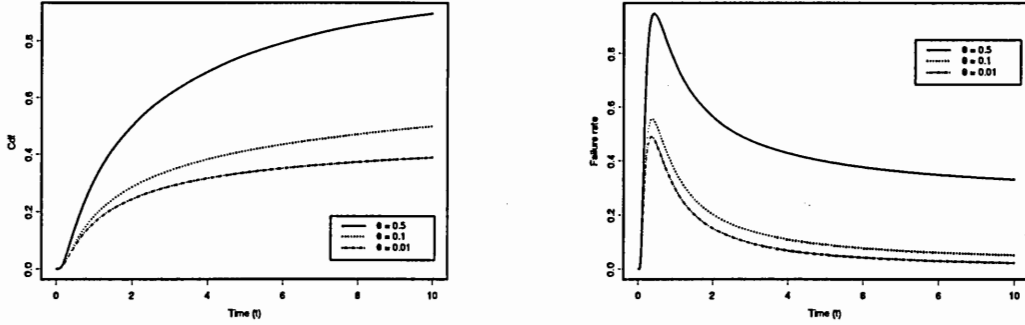
$$\alpha = \frac{\sigma_w}{\sqrt{\mathcal{D}_f\theta}} \quad \text{and} \quad \beta = \frac{\mathcal{D}_f}{\theta},$$

then $G_I(t)$ is reduced to

$$\begin{aligned} G_I(t) &= 1 - \Phi\left\{\frac{1}{\alpha}\left[\sqrt{\frac{\beta}{t}} - \sqrt{\frac{t}{\beta}}\right]\right\} \\ &= \Phi\left\{\frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right\}. \end{aligned} \quad (10.3.112)$$

It is noted that the derived failure-time distribution (10.3.112) is a Birnbaum-Saunders distribution with a shape parameter $\alpha > 0$ and a scale parameter $\beta > 0$. $\beta = \mathcal{D}_f/\theta$ represents a degradation rate of the MIDP $\mathcal{D}(t; X, \Theta)$. The Pdf of (10.3.112) is

$$g_I(t) = \frac{\sqrt{\frac{t}{\beta}} + \sqrt{\frac{\beta}{t}}}{2\alpha t} \cdot \phi\left\{\frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right\}, \quad (10.3.113)$$



(a) Cdf plot

(b) Failure rate plot

Figure 26: Birnbaum-Saunders distribution

where $\phi\{\cdot\}$ represents the Pdf of the standard normal distribution. The failure rate of $G_I(t)$

$$r_I(t) = \frac{g_I(t)}{1 - G_I(t)} = \frac{\frac{\sqrt{\frac{t}{\beta}} + \sqrt{\frac{\beta}{t}}}{2\alpha t} \cdot \phi\left\{\frac{1}{\alpha} \left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right\}}{1 - \Phi\left\{\frac{1}{\alpha} \left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right\}}. \quad (10.3.114)$$

The Cdf and the failure rate of Birnbaum-Saunders distribution are plotted in Figure 26, with respect to the changing values of θ with $\mathcal{D}_f = 1$, $\sigma_w = 1$ (that is, a standard Wiener process).

Through the simulations with a wide range of θ , \mathcal{D}_f , and σ_w , the failure rate $r_I(t)$ is always unimodal. It is also observed that $r_I(0) = 0$ and for relatively large β (smaller degradation rate),

$$\lim_{t \rightarrow \infty} r_I(t) = \frac{1}{2\alpha^2\beta} \rightarrow 0.$$

Some properties of Birnbaum-Saunders distribution are discussed in Mann *et al.* [60].

10.4 Fatigue Crack Growth Model

The prediction of stochastic crack growth accumulation is crucial for the reliability and durability analyses of fatigue critical components. Yang *et al.* [97] and Sobczyk [84] proposed the following crack growth rate model

$$\frac{d\alpha(t)}{dt} = X(t) \cdot \eta(\Delta K(\alpha)), \quad (10.4.115)$$

where $\alpha(t)$ is the crack size at time t , $\Delta K(\alpha)$ is the stress intensity range which is a function of crack size α , and $\eta(\cdot)$ represents any deterministic crack growth rate function. $\eta(\cdot)$ is

based on the principles of fracture mechanics. Some commonly used crack growth rate functions are the Paris-Erdogan relationship [71], *i.e.*, $\eta(\Delta K(\alpha)) = C(\Delta K)^m$, where C and m are assumed constants for a given material.

The deterministic crack growth rate equation $\frac{d\alpha(t)}{dt} = \eta(\Delta K(\alpha))$ represents the mean crack growth rate, whereas the gaussian random process $\{X(t) : t > 0\}$ accounts for the statistical variability of the crack growth accumulation.

Yang *et al.* [97] assumed that $X(t)$ is a stationary lognormal random process with a median value of unity. In other words, the process $Z(t) = \log X(t)$ is a gaussian process with zero mean. The stationary lognormal random process $X(t)$ is defined by the autocovariance function

$$\gamma_X(\tau) = E[X(t)X(t + \tau)],$$

and $\gamma_X(0) = Var[X(t)] = \sigma_X^2$. The autocovariance function plays a significant role in random process analysis and specifies the statistical behavior of the random process.

From the physical standpoint, the autocovariance function of the crack growth rate should decrease as time difference τ increases. According to Yang *et al.* [97], the autocovariance function $\gamma_X(\tau)$ of the random fatigue crack growth process $X(t)$ is an exponential decaying function of time difference τ as

$$\gamma_X(\tau) = \sigma_X^2 \cdot \exp[-\zeta|\tau|], \quad \tau > 0 \quad (10.4.116)$$

where ζ^{-1} is a measure of the correlation time for $X(t)$.

Bsed on the mean crack growth rate $\eta(\Delta K(\alpha))$ and the lognormal random process $\{X(t) : t > 0\}$ with autocovariance (10.4.116), we will derive a failure-time distribution of a fatigue degradation process.

Denote a deterministic initial crack size as α_0 at $t_0 = 0$ and a crack size at any service time τ as $\alpha(\tau)$. Then the mean service time to reach $\alpha(\tau)$ from α_0 is

$$\int_{\alpha_0}^{\alpha(\tau)} \frac{d\alpha}{d\eta(\Delta K(\alpha))} = \int_0^\tau X(t)dt, \quad (10.4.117)$$

since $K(\alpha)$ is a function of the crack size α . Let $\omega(\tau) = \int_0^\tau X(t)dt$. , It is noted from (10.4.117) that the crack size $\alpha(\tau)$ is a monotone increasing function of $\omega(\tau)$. The integration

of the lognormal random process $X(t)$, $\{\omega(\tau) : \tau > 0\}$ is a also lognormal random process (see Ross [77]). We assume that $\omega(\tau)$ has a mean $\mu_{\omega(\tau)}$ and a variance $\sigma_{\omega(\tau)}^2$. Then, the mean and the variance of $\omega(\tau)$ is expressed as the parameters of $X(t)$, that is

$$\mu_{\omega(\tau)} = \int_0^\tau E[X(t)]dt = \mu_X \tau, \quad (10.4.118)$$

and

$$\begin{aligned} \sigma_{\omega(\tau)}^2 &= \int_0^\tau \int_0^\tau E[X(t_1)X(t_2)]dt_1dt_2 \\ &= \int_0^\tau \int_0^\tau \sigma_X^2 \cdot \exp[-\zeta|t_2 - t_1|]dt_1dt_2 \\ &= 2 \left(\frac{\sigma_X^2}{\zeta} \right) \cdot (\exp[-\zeta\tau] + \zeta\tau - 1). \end{aligned} \quad (10.4.119)$$

A failure-time of a fatigue process is defined as the time that a crack size increases upper a pre-determined threshold value α_f . Then a failure-time of $\alpha(\tau)$ is

$$\begin{aligned} Pr[\alpha(\tau) \geq \alpha_f] &= Pr[\omega(\tau) \geq \omega_f] \\ &= 1 - G_{\omega(\tau)}(\omega_f) \\ &= 1 - \Phi \left[\frac{\log \omega_f - \tilde{\mu}_{\omega(\tau)}}{\tilde{\sigma}_{\omega(\tau)}} \right], \end{aligned} \quad (10.4.120)$$

where $G_{\omega(\tau)}$ represents the Cdf of $\omega(\tau)$ and $\omega_f = \int_{\alpha_0}^{\alpha_f} \frac{d\alpha}{d\eta(\Delta K(\alpha))}$. $\tilde{\mu}_{\omega(\tau)}$ and $\tilde{\sigma}_{\omega(\tau)}$ can be obtained by solving the following equation

$$\begin{aligned} \mu_{\omega(\tau)} &= \exp \left[\tilde{\mu}_{\omega(\tau)} + \tilde{\sigma}_{\omega(\tau)}^2 \right], \\ \sigma_{\omega(\tau)}^2 &= \exp \left[2\tilde{\mu}_{\omega(\tau)} + \tilde{\sigma}_{\omega(\tau)}^2 \right] \cdot (\exp[\tilde{\sigma}_{\omega(\tau)}^2] - 1). \end{aligned}$$

APPENDIX A

ZINC PHOSPHATING COATING PROCESS DATA

Run	Control								Signal			
	A	B	C	D	E	F	G	H	M1	M2	M3	M4
1	1	1	1	1	1	1	1	1	7.6	20.1	35.1	62.2
2	1	1	2	2	2	2	2	2	13.6	34.0	76.9	150.4
3	1	1	3	3	3	3	3	3	12.1	27.5	58.6	110.9
4	1	2	1	1	2	2	3	3	10.0	25.6	53.9	101.2
5	1	2	2	2	3	3	1	1	18.0	47.9	100.2	195.2
6	1	2	3	3	1	1	2	2	19.9	52.9	91.8	224.7
7	1	3	1	2	1	3	2	3	1.5	4.1	8.7	25.0
8	1	3	2	3	2	1	3	1	14.9	36.6	89.7	183.5
9	1	3	3	1	3	2	1	2	15.5	38.8	79.4	161.3
10	2	1	1	3	3	2	2	1	14.8	37.2	81.1	158.9
11	2	1	2	1	1	3	3	2	3.7	11.1	22.6	59.2
12	2	1	3	2	2	1	1	3	17.8	41.1	84.0	145.8
13	2	2	1	2	3	1	3	2	14.1	36.3	66.7	126.0
14	2	2	2	3	1	2	1	3	17.5	40.8	90.9	186.3
15	2	2	3	1	2	3	2	1	10.3	26.5	60.1	136.8
16	2	3	1	3	2	3	1	2	25.6	66.9	129.3	252.5
17	2	3	2	1	3	1	2	3	6.5	12.6	28.6	52.3
18	2	3	3	2	1	2	3	1	13.3	33.2	75.5	136.6

APPENDIX B

PUSH-PULL CABLE ACTUATOR EXPERIMENT DATA

Run	Control											Noise	Signal								
	A	B	C	D	E	F	G	H	I	J	K	N	M=8			M=16			M=24		
1	-----											-1	4.97	5.19	5.41	10.05	10.49	10.94	13.28	13.95	14.62
												+1	4.69	4.9	5.11	9.48	9.9	10.32	12.53	13.16	13.79
2	-----++++											-1	4.0	4.27	4.54	7.2	7.74	8.27	8.56	9.37	10.18
												+1	3.84	4.1	4.35	6.9	7.42	7.93	8.21	8.98	9.76
3	--+++----++											-1	4.46	4.65	4.84	9.04	9.41	9.79	11.78	12.35	12.91
												+1	4.4	4.48	4.66	8.7	9.06	9.42	11.34	11.89	12.43
4	-+-++-+-+--											-1	5.0	5.14	5.28	9.59	9.86	10.14	13.66	14.07	14.48
												+1	4.89	5.03	5.16	9.38	9.65	9.92	13.36	13.76	14.17
5	-++-+-+--+--											-1	6.0	6.27	6.54	12.1	12.64	13.18	18.2	19.01	19.82
												+1	5.75	6.01	6.27	11.6	12.12	12.64	17.45	18.23	19.0
6	-++++-+-+---											-1	3.86	4.0	4.14	7.7	7.98	8.26	10.5	10.92	11.33
												+1	3.72	3.85	3.99	7.42	7.68	7.95	10.11	10.51	10.91
7	+--+-+--+--+--											-1	4.04	4.23	4.42	7.29	7.66	8.03	8.72	9.28	9.84
												+1	3.96	4.14	4.32	7.13	7.49	7.86	8.53	9.08	9.62
8	+-+--++++--+											-1	3.39	4.61	3.83	6.36	6.8	7.25	6.96	7.63	8.29
												+1	3.2	3.41	3.62	6.0	6.42	6.84	6.57	7.2	7.83
9	+---+++-+--+--											-1	6.13	6.27	6.41	11.98	12.25	12.53	16.53	16.94	17.36
												+1	6.0	6.13	6.27	11.72	11.99	12.26	16.17	16.58	16.98
10	+++-----+--+											-1	6.54	6.76	6.98	13.18	13.62	14.06	19.82	20.48	21.14
												+1	6.27	6.48	6.69	12.64	13.06	13.48	19.0	19.64	20.27
11	++-+-+-----++											-1	4.89	5.08	5.27	10.28	10.65	11.03	15.15	15.71	16.27
												+1	4.71	4.89	5.07	9.9	10.26	10.62	14.58	15.12	15.66
12	++--+-+--+--+--											-1	3.28	3.55	3.82	6.53	7.07	7.61	8.73	9.54	10.36
												+1	3.09	3.35	3.60	6.16	6.67	7.19	8.23	8.23	9.77

REFERENCES

- [1] Abramowitz, M. and Stegun, I. A. (1964), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York.
- [2] Ahmad, M., and Sheikh, A. K. (1984), "Bernstein Reliability Model: Derivation and Estimation of Parameters", *Reliability Engineering*, 8, 131-148.
- [3] Amemiya, Y. (1985), "What Should Be Done When an Estimated Between Group Covariance Matrix is not Nonnegative Definite?", *The American Statistician*, 39, 112-117.
- [4] Atkinson, A. C. (1982), "Regression Diagnostics, Transformation and Constructed Variables", (with discussion) *Journal of the Royal Statistical Society*, B 44, 1-36.
- [5] Beal, S. L. and Sheiner, L. B. (1982), "Estimating Population Kinetics", *CRC Critical Reviews in Biomedical Engineering*, 8, 195-222.
- [6] Beal, S. L. and Sheiner, L. B. (1985), "Methodology of population pharmacokinetics", *Drug Fate and Metabolism - Methods and Techniques*, Eds. E. R. Garret and J. L. Hirtz, Marcel Dekker, New York.
- [7] Beal, S. L. and Sheiner, L. B. (1988), "Heteroscedastic Nonlinear Regression", *Technometrics*, 30, 327-338.
- [8] Beal, S. L. and Sheiner, L. B. (1992), *NONMEM User's Guides*, NONMEM Project Group, University of California, San Francisco.
- [9] Beam, A. L. (1995), "A program for Fitting Two-Phase Segmented-Curve Models with an Unknown Change Point, with an Application to the Analysis of Strategy Shifts in a Cognitive Task", *Behavior Research Methods, Instruments, & Computers*, 27, 392-399.
- [10] Box, G. E. P. (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations", (with discussion) *Technometrics*, 30, 1-40.
- [11] Box, G. E. P., and Jones, S. (1990), "Designs for Minimizing the Effects of Environmental Variables", Technical Report, University of Wisconsin.
- [12] Byrne, D., and Quinlan, J. (1993), "Robust Function for Attaining High Reliability at Low Cost", *IEEE Proceedings of the Annual Reliability and Maintainability Symposium*, 183-191.
- [13] Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992), "Hierarchical Bayesian Analysis of Changepoint Problems", *Applied Statistics*, 41, 389-405.
- [14] Chang, D. S. (1992), "Analysis of Accelerated Degradation Data in a Two-way Design", *Reliability Engineering and System Safety*, 39, 65-69.

- [15] Chen, J., and Gupta, A. K. (2000), *Parametric Statistical Change Point Analysis*, Birkhäuser, Boston.
- [16] 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.
- [17] Chin Choy, J. H., and Broemeling, L. D. (1980), "Some Bayesian Inferences for a Changing Linear Model", *Technometrics*, 22, 71–78.
- [18] Chow, G. (1960), "Tests of the Equality Between Two Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 561–605.
- [19] Csörgő, M., and Horváth, L. (1997), *Limit Theorems in Change-Point Analysis*, John Wiley & Sons, New York.
- [20] Davidian, M. and Giltinan, D. M. (1995), *Nonlinear models for repeated measurement data*, Chapman and Hall, London.
- [21] Davis, P. J. and Rabinowitz, P. (1984), *Methods of Numerical Integration*, 2nd ed., Academic Press, New York.
- [22] Dhillon, B. S. (1981), "Life Distributions", *IEEE Transactions on Reliability*, 30, 457–459.
- [23] Dubey, S. D. (1966), "Transformation for Estimation of Parameters", *Journal of the Indian Statistical Association*, 4, 109–124.
- [24] Efron, B. and Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman & Hall, New York.
- [25] Engel, J. (1992), "Modelling Variation in Industrial Experiment", *Applied Statistics*, 41, 579–593.
- [26] Engel, J., and Huele, A. F. (1996), "A Generalized Linear Modeling Approach to Robust Design", *Technometrics*, 38, 365–373.
- [27] 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).
- [28] Feder, P. I. (1975a), "On Asymptotic Distribution Theory in Segmented Regression Problems-Identified Case", *The Annals of Statistics*, 3, 49–83.
- [29] Feder, P. I. (1975b), "The Log Likelihood Ratio in Segmented Regression", *The Annals of Statistics*, 3, 84–97.
- [30] Firth, D. (1988), "Multiplicative Errors: Log-normal or Gamma ?", *Journal of the Royal Statistical Society*, B 50, 266–268.
- [31] Fukuda, M. (1991), *Reliability and Degradation of Semiconductor Lasers and LEDs*, Artech House, Boston.

- [32] Gertsbakh, I. B., and Kordonskiy, Kh. B. (1969), *Models of Failure*, English translation from the Russian version, Springer-verlag, New York.
- [33] Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration", *Econometrica*, 57, 1317–1339.
- [34] Golub, G. H. and Welsch, J. H. (1969), "Calculation of Gaussian Quadrature Rules", *Mathematical Computing*, 23, 221–230.
- [35] Grego, J. M. (1993), "Generalized Linear Models and Process Variation", *Journal of Quality Technology*, 25, 288–295.
- [36] Hamada, M. S., and Wu, C. F. J. (1992), "Analysis of Designed Experiments with Complex Aliasing", *Journal of Quality Technology*, 24, 130–137.
- [37] Hamada, M., and Nelder, J. A. (1997), "Generalized Linear Models for Quality-Improvement Experiments", *Journal of Quality Technology*, 29, 292–304.
- [38] Hawkins, D. L. (1989), "A U-I Approach to Retrospective Testing for Shift Parameters in a Linear Model", *Communications in Statistics: Theory and Methods*, 18, 3117–3134.
- [39] Hinkley, D. V. (1969), "Inference about the Intersection in Two-Phase Regression", *Biometrika*, 57, 1–17.
- [40] Hinkley, D. V. (1971), "Inference in Two-Phase Regression", *Journal of the American Statistical Association*, 66, 736–743.
- [41] Hudson, D. J. (1966), "Fitting Segmented Curves Whose Join Points Have to be Estimated", *Journal of the American Statistical Association*, 61, 1097–1129.
- [42] Jensen, F., and Petersen, N. E. (1982), *Burn-in: An Engineering Approach to Design and Analysis of Burn-in Procedures*, John Wiley & Sons, New York.
- [43] Joseph, V. R., and Wu, C. F. J. (2002a), "Robust Parameter Design of Multiple-Target Systems", *Technometrics*, 44, 338–346.
- [44] Joseph, V. R., and Wu, C. F. J. (2002b) "Performance Measures in Dynamic Parameter Design", *Journal of Japanese Quality Engineering Society*, 10, 82–86.
- [45] Kendall, R. A., and Stuart, A. (1977), *The Advanced Theory of Statistics*, Griffin, London.
- [46] Kim, D. (1994), "Tests for a change-point in Linear Regression", *IMS Lecture Notes, Monograph Series*, 23, 170–176.
- [47] 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.
- [48] Lee, Y., and Nelder, J. A. (2003), "Robust Design via Generalized Linear Models", *Journal of Quality Technology*, 35, 2–12.
- [49] Lenth, R. V. (1989), "Quick and Easy Analysis of Unreplicated Experiments", *Technometrics*, 31, 469–473.

- [50] Leon, R. V., Shoemaker, A. C., and Kackar, R. N. (1987), "Performance Measure Independent of Adjustment: An Explanation and Extension of Taguchi's Signal to Noise Ratio", *Technometrics*, 29, 253–285.
- [51] Lin, S. M., and Wen, T. C. (1994), "Experimental Strategy: Application of Taguchi's Quality Engineering Method to Zinc Phosphate Coating Uniformity", *Plating and Surface Finishing*, 59–64.
- [52] Lin, W. Z. (1976), *Fundamentals of Fluorescent Lamp*, Taiwan Fluorescent Lamp Co. Corp., Taiwan.
- [53] 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.
- [54] Lindstrom M. J. and Bates, D. M. (1990), "Nonlinear Mixed Effects Models for Repeated Measures Data", *Biometrics*, 46, 673–687.
- [55] Lu, C. J. and Meeker, W. Q. (1993), "Using Degradation Measures to Estimate of Time-to-failure Distribution", *Technometrics*, 35, 161–176.
- [56] 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.
- [57] 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.
- [58] Lunani, M., Nair, V. N., and Wasserman, G. S. (1995), "Graphical Methods for Robust Design with Dynamic Characteristics", *Journal of Quality Technology*, 29, 327–338.
- [59] Malik, H. J. (1967), "Exact Distribution of the Quotient of Independent Generalized Gamma Variables", *Canadian Mathematical Bulletin*, 10, 463–465.
- [60] Mann, N. R., Schafer, R. E., and Singpurwalla, N. D. (1974), *Methods for Statistical Analysis of Reliability and Life Data*, Wiley, New York.
- [61] McCaskey, S. D., and Tsui, K. L. (1997), "Analysis of Dynamic Robust Design Experiments", *International Journal of Production Research*, 35, 1561–1574.
- [62] McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Model*, Chapman and Hall, New York.
- [63] 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.
- [64] Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, Wiley, New York.
- [65] Miller, A., and Wu, C. F. J. (1996), "Parameter Design for Signal-Response Systems: A Different Look at Taguchi's Dynamic Parameter Design", *Statistical Science*, 11, 122–136.

- [66] 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.
- [67] Myers, R. H., and Montgomery, D. C. (1997), "A Tutorial on Generalized Linear Models", *Journal of Quality Technology*, 29, 274–291.
- [68] Myers, R. H., Montgomery, D. C., and Vining, G. G. (2002), *Generalized Linear Models with Applications in Engineering and the Sciences*, Wiley, New York.
- [69] Nelder, J. A., and Lee, Y. (1991), "Generalized Linear Models for The Analysis of Taguchi-Type Experiments", *Applied Stochastic Models and Data Analysis*, 7, 107–120.
- [70] Nelson, W. (1990), *Accelerated Testing: Statistical Models, Test Plans and Data Analysis*, Wiley, New York.
- [71] Paris, P. C., and Erdogan, F. (1963), "A Critical Analysis of Crack Propagation Laws", *Journal of Basic Engineering Transactions, ASME, Series D*, 85, 528–534.
- [72] 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.
- [73] Pinheiro, J. C. and Bates, D. M., (2000), *Mixed-Effects Models in S and S-Plus*, Springer, New York.
- [74] Quandt, R. E. (1958), "The Estimation of the Parameters of a Linear Regression System Obeys Two Separate Regimes", *Journal of the American Statistical Association*, 53, 873–880.
- [75] Quandt, R. E. (1960), "Test of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes", *Journal of the American Statistical Association*, 55, 324–330.
- [76] Ramos, R. Q. and Pantula, S. G. (1995), "Estimation of Nonlinear Random Coefficient Models", *Statistics & Probability letters*, 24, 49–56.
- [77] Ross, S. M. (1997), *Probability Models*, six ed., Academic Press, New York.
- [78] SAS Institute, Inc. (1999), *SAS OnlineDoc, Version 8.0*, SAS Inc., Cary, NC.
- [79] Sheiner, L. B. and Beal, S. L. (1985), "Pharmacokinetic Parameter Estimates from Several Least Squares Procedures: Superiority of Extended Least Squares", *Journal of Pharmacokinetics and Biopharmaceutics*, 13, 185–201.
- [80] Shiau, J. H. and Lin, H. (1999), "Analyzing Accelerated Degradation Data by Non-parametric Regression", *IEEE Transactions on Reliability*, 48, 149–158.
- [81] Shoemaker, A. C., Tsui, K. L., and Wu, C. F. J. (1991), "Economical Experimentation Methods for Robust Parameter Design", *Technometrics*, 33, 415–427.
- [82] Siegmund, D. O., and Zhang, H. (1994), "Confidence Regions in Broken Line Regression", *IMS Lecture Notes, Monograph Series*, 23, 292–316.

- [83] Smith, A. F. M., and Cook, D. G. (1980), "Straight Lines with a Change-point: a Bayesian Analysis of Some Renal Transplant Data", *Applied Statistics*, 29, 180–189.
- [84] Sobczyk, K. (1984), "Stochastic Modeling of Fatigue Crack Growth", in Proceedings of the IUTAM Symposium on 'Probabilistic Methods in Mechanics of Solids and Structures', Stockholm, Sweden, Springer-Verlag, Berlin, 111–119.
- [85] 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.
- [86] Taguchi, G. (1987), *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*, White Plains, NY: Unipub/Kraus International.
- [87] 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.
- [88] Tseng S. T. and Wen, J. C. (2000), "Step-stress Accelerated Degradation Analysis for Highly Reliable Products", *Journal of Quality Technology*, 32, 209–216.
- [89] Tsui, K. L. (1994), "Avoiding Unnecessary Bias in Robust Design Analysis", *Computational Statistics and Data Analysis*, 18, 535–546.
- [90] Tsui, K. L. (1998), "Alternatives of Taguchi's Approach for Dynamic Robust Design Problems", *International Journal of Reliability, Quality and Safety Engineering*, 5, 115–131.
- [91] Tsui, K. L. (1999), "Response Model Analysis of Dynamic Robust Design Experiments", *IIE Transactions*, 31, 115–131.
- [92] Tsui, K. L. (2000), "Modeling and Analysis of Dynamic Robust Design Experiments", *Frontiers in SQC*, 31, 115–131.
- [93] Vonesh, E. F. and Chinchilli, V. M. (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, Inc., New York.
- [94] Welch, W. J., Kang, S. M., and Sacks, J. (1990), "Computer Experiments for Quality Control by Parameter Design", *Journal of Quality Technology*, 22, 1–12.
- [95] Wolfinger, R. (1993), "Laplace's Approximation for Nonlinear Mixed Models", *Biometrika*, 80, 791–795.
- [96] Wolfinger, R. D., and Tobias, R. D. (1998), "Joint Estimation of Location, Dispersion, and Random Effects in Robust Design", *Technometrics*, 40, 62–71.
- [97] Yang, J. N., His, W. H., and Manning, S. D. (1987), "Stochastic Crack Growth Models for Applications to Aircraft Structures", *Probabilistic Fracture Mechanics and Reliability*, Provan, Ed., Martinus Nijhoff Publisher, Chapter 4, 171–211.
- [98] Yu, H. F. and Tseng, S. T. (1998), "On-line Procedure for Terminating an Accelerated Degradation Test", *Statistica Sinica*, 8, 207–220.
- [99] Zeger, S. L., Liang, K.Y. and Albert, P. S. (1988), "Models for Longitudinal Data: A Generalized Estimating Equation Approach", *Biometrics*, 44, 1049–1060.

VITA

Suk Joo Bae was born on July 30, 1970, in Mokpo, Korea. He is the only son in the family and has three sisters. He earned B. S. and M. S. in Industrial Engineering at Hanyang University, Seoul, Korea, in February 1994, and February 1996. His M. S. Thesis title is "The Diagnosis for Life Data in Accelerated Life Testing".

From February 1996 to January 1999, he worked in reliability research group in Research Laboratory of Samsung SDI Inc., Korea. He engaged in developing accelerated life testing method of Cathode Lay Tubes (CRTs), Liquid Crystal Displays (LCDs), and Vacuum Fluorescent Displays (VFDs) and establishing reliability infrastructures in Samsung SDI.

In August, 1999, he entered Georgia Institute of Technology to pursue Ph. D. degree in the School of Industrial and Systems Engineering. Since summer, 2000, he has worked as a research assistant in the School of Industrial and Systems Engineering at Georgia Institute of Technology.

His primary research interests include modeling of physical degradation mechanism for light displays using statistical methods and developing dynamic robust design experiments. He has published several technical papers in reputed journals and conference proceedings.

He is a member of Institute of Industrial Engineers (IIE), American Statistical Association (ASA), Institute for Operations Research and the Management Sciences (INFORMS), and Institute of Mathematical Statistics (IMS).