

Advancements in Degradation Modeling, Uncertainty Quantification, and Spatial Variable Selection

Yimeng Xie

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Yili Hong, Chair
Xinwei Deng
Inyoung Kim
William H. Woodall

Apr 27, 2016
Blacksburg, Virginia

Key Words: ADDT; Degradation Model; Spatial Variable Selection; SPI/SPB.

Copyright 2016, Yimeng Xie

Advancements in Degradation Modeling, Uncertainty Quantification and Spatial Variable Selection

Yimeng Xie

ABSTRACT

This dissertation focuses on three research projects: 1) construction of simultaneous prediction intervals/bounds for at least k out of m ($k \leq m$) future observations; 2) semi-parametric degradation model for accelerated destructive degradation test (ADDT) data; and 3) spatial variable selection and application to Lyme disease data in Virginia. Followed by the general introduction in Chapter 1, the rest of the dissertation consists of three main chapters.

Chapter 2 presents the construction of two-sided simultaneous prediction intervals (SPIs) or one-sided simultaneous prediction bounds (SPBs) to contain at least k out of m ($k \leq m$) future observations, based on complete or right censored data from (log)-location-scale family of distributions. SPI/SPB calculated by the proposed procedure has exact coverage probability for complete and Type II censored data. In Type I censoring case, it has asymptotically correct coverage probability and reasonably good results for small samples. The proposed procedures can be extended to multiply-censored data or randomly censored data.

Chapter 3 focuses on the analysis of ADDT data. We use a general degradation path model with correlated covariance structure to describe ADDT data. Monotone B-splines are used to modeling the underlying degradation process. A likelihood based iterative procedure for parameter estimation is developed. The confidence intervals of parameters are calculated using the nonparametric bootstrap procedure. Both simulated data and real datasets are used to compare the semi-parametric model with the existing parametric models.

Chapter 4 studies the Lyme disease emergence in Virginia. The objective is to find important environmental and demographical covariates that are associated with Lyme disease emergence. To address the high-dimensional integral problem in the loglikelihood function, we consider the penalized quasi loglikelihood and the approximated loglikelihood based on Laplace approximation. We impose the adaptive elastic net penalty to obtain sparse estimation of parameters and thus to achieve variable selection of important variables. The proposed methods are investigated in simulation studies. We also apply the proposed methods to Lyme disease data in Virginia.

Finally, Chapter 5 contains general conclusions and discussions for future work.

Advancements in Degradation Modeling, Uncertainty Quantification and Spatial Variable Selection

Yimeng Xie

GENERAL AUDIENCE ABSTRACT

This dissertation focuses on simultaneous prediction intervals/bounds, accelerated destructive degradation test (ADDT) data analysis, spatial variable selection and an application to Virginia Lyme disease emergence. Our simultaneous prediction intervals/bounds (SPIs/SPBs) are designed to qualify the uncertainties of at least k out of m ($k \leq m$) future observations. We propose a simulation-based approach to compute SPIs/SPBs based on completely observed or censored data (the event time is not observed) from (log)-location-scale family of distributions. The coverage probabilities of SPIs/SPBs are studied. The second topic is related with ADDT data. ADDT is useful for evaluating the product's long-term properties. It is important for companies to ensure that the reliability requirements of products are met. In Chapter 3, we propose a general and flexible semi-parametric model for ADDT data. Compared to existing parametric models for ADDT in literature, our semi-parametric model requires less assumptions and can be applied to different kinds of products. The third project studies the Lyme disease emergence in Virginia. We observed that the number of reported Lyme disease cases increased in recent years. Therefore, it is meaningful and helpful to find factors that are related with Lyme disease incidence. In Chapter 4, we develop variable selection procedures for spatial correlated data and apply the techniques to Lyme disease data to find important factors. Our findings as well as the results from past studies are compared and discussed.

To my family.

Acknowledgments

I would like first of all to express my sincere gratitude and appreciation to my adviser, Dr. Yili Hong, for his excellent guidance and continued support through my research. I would like to thank my committee members, Dr. Xinwei Deng, Dr. Inyoung Kim, and Dr. William Woodall for their valuable advice and persistent help. I would also like to thank Dr. Laura Sands for her guidance and patience.

Many thanks to all the graduate students in Dr. Hong's research group, Dr. Yuanyuan Duan, Dr. Khaled Bedair, Dr. Caleb King, Dr. Zhibing Xu, Miao Yuan, I-Chen Lee and Zhongnan Jin for their comments and suggestions.

Special thanks go to my family for their unconditional love and support.

Contents

1	General Introduction	1
1.1	Background	1
1.1.1	Reliability Analysis	1
1.1.2	Spatial Epidemiology	3
1.2	Motivation	6
1.2.1	Simultaneous Prediction Interval	6
1.2.2	Semi-parametric Model for ADDT Data	7
1.2.3	Spatial Variable Selection	7
1.3	Outline of the Dissertation	7
2	Simultaneous Prediction Intervals for the (Log)-Location-Scale Family of Distributions	12
2.1	Introduction	13
2.1.1	Motivation	13
2.1.2	Literature Review and Contributions of This Work	14
2.1.3	Overview	15
2.2	Data, Model, and Maximum Likelihood Estimation	15
2.2.1	Data	15
2.2.2	Model	16
2.2.3	Maximum Likelihood Estimation	17
2.3	Simultaneous Prediction Intervals and Bounds	18
2.3.1	Two-sided Simultaneous Prediction Intervals	18
2.3.2	One-sided Simultaneous Prediction Bounds	20
2.4	Computations of the Simultaneous Prediction Intervals/Bounds	22
2.4.1	Complete and Type II Censored Data	22
2.4.2	Two-sided SPI with Equal Tail Probability	24

2.4.3	Type I Censored Data	24
2.4.4	Illustrative Examples	25
2.5	Simulation Study for Type I censoring	26
2.6	Applications	28
2.6.1	Nozzle Failure Time Data	28
2.6.2	Aircraft Component Failure Time Data	30
2.6.3	Vinyl Chloride Data	30
2.7	Concluding Remarks and Areas for Future Research	31
2.A	Proof of Equation (2.2)	32
2.B	Approximate Pivotal Property for Type I Censored Data	34
2.C	Variance of Estimated Coverage Probability	35
3	A Semi-parametric Model for Accelerated Destructive Degradation Test	
	Data Analysis	39
3.1	Introduction	40
3.1.1	Motivation	40
3.1.2	Related Literature	41
3.1.3	Overview	42
3.2	The Semi-parametric Model	43
3.2.1	General Setting	43
3.2.2	The Scale Acceleration Model	44
3.2.3	Nonparametric Form for Baseline Degradation Path	45
3.3	Estimation and Inference	47
3.3.1	Parameter Estimation	47
3.3.2	Reliability Measures	49
3.3.3	Spline Knots Selection	50
3.3.4	Statistical Inference	51
3.4	Simulation Study	52
3.4.1	Performance of Parameter Estimators	53
3.4.2	Performance under Model Misspecification	54

3.5	Applications	61
3.5.1	ADDT Datasets and Parametric Models	61
3.5.2	Comparisons of Parametric and Semi-parametric Models	64
3.5.3	Illustration of failure time distribution	67
3.6	Conclusions and Areas for Future Work	67
4	Spatial Variable Selection via Elastic Net and an Application to Virginia	
	Lyme Disease Emergence	73
4.1	Introduction	74
4.1.1	Background	74
4.1.2	Lyme Disease Emergence	74
4.1.3	Related Literature	75
4.1.4	Overview	77
4.2	Data, Model, Likelihood and Penalty Function	77
4.2.1	Data and Model	77
4.2.2	Likelihood Function and Penalty	78
4.3	Estimation Procedures	79
4.3.1	The Estimation Problem	79
4.3.2	Penalized Quasi-likelihood	80
4.3.3	PQL with Adaptive Elastic Net Penalty	82
4.3.4	Laplace Approximated Loglikelihood with Elastic Net Penalty	84
4.3.5	Tuning Parameter Selection	85
4.4	Simulation Study	85
4.4.1	Setting	86
4.4.2	Results and Discussions	87
4.5	Application to Lyme Disease Data	91
4.5.1	Data Description	91
4.5.2	Estimation and Findings	94
4.6	Concluding Remarks	95
4.A	Quadratic Approximation to PQL	97

4.B	A Description of the Variable Selection Procedure using PQL	98
4.C	A Description of the Variable Selection Procedure using Laplace Approximated Loglikelihood	99
5	General Conclusions and Areas for Future Work	104

List of Figures

2.1	CP curves of one-sided upper SPBs for $n = 20, r = 5, 10, 15$ and $20, k = 4$, and $m = 5$ based on Algorithm 1	26
2.2	(a) Contour plot of the CP as a function of u_L and u_U for $n = 20, r = 8, k = m = 3$, based on 100,000 simulations. (b) The contour lines of equation (2.9).	27
2.3	Estimated actual CP versus nominal confidence level for fixed $p_f = 0.25$, when $k = 4$ and $m = 5$. (a) Lower SPB. (b) Upper SPB. (c) Two-sided SPI.	29
2.4	Probability plots for the aircraft data. (a) Weibull fit. (b) Lognormal fit.	31
2.5	Probability plots for the vinyl chloride data. (a) Weibull fit. (b) Lognormal fit.	32
3.1	Spline bases and baseline degradation path used in simulation study.	55
3.2	Empirical bias and MSE of parameter estimators for $(\beta, \sigma, \rho)'$	56
3.3	Empirical pointwise MSE for the estimator of the baseline degradation path.	57
3.4	CP of the CI procedures for parameters $(\beta, \sigma, \rho)'$, using quantile-based and bias-corrected methods, respectively.	58
3.5	Pointwise CP of the CI procedure for baseline degradation path, using quantile-based and bias-corrected methods, respectively.	59
3.6	Plot of simulated data and fitted degradation paths based on the true and incorrect parametric models, and semi-parametric model.	60
3.7	Fitted degradation paths of the Adhesive Bond B data.	62
3.8	Residual analysis for the Adhesive Bond B data.	62
3.9	Fitted degradation paths of the Seal Strength data.	63
3.10	Residual analysis for the Seal Strength data.	64
3.11	Fitted degradation paths of the Adhesive Formulation K data.	65
3.12	Residual analysis for the Adhesive Formulation K data.	65
3.13	Estimates and CIs of quantile functions at different temperature levels for Adhesive Bond B data.	68

4.1 Number of cases and incidence rates of each census tract in Virginia (2006-2011). (a) Case counts. (b) Incidence rates. 93

List of Tables

2.1	The pdfs and cdfs of different commonly-used members of the standard location-scale and log-location-scale distributions.	17
3.1	Selected temperature levels and time points for the simulation studies.	53
3.2	Empirical IBias, root IVar (RIVar), and root IMSE (RIMSE) for the true model (3.12), incorrect model (3.13), and the semi-parametric model.	59
3.3	Empirical mean, bias, SD, and root MSE (RMSE) of the MTTF estimators based on the true model (3.12), incorrect model (3.13), and the semi-parametric model.	60
3.4	Parameter estimates and corresponding CI for the semi-parametric models for the three applications.	66
3.5	Estimated MTTF and CI at normal use condition based on parametric and semi-parametric models for the three applications (time in weeks).	66
3.6	Log likelihood and AIC values of parametric and semi-parametric models for the ADDT data from the three applications.	66
4.1	Model selection results based on simulated samples. The parameters are $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$, $\theta = (0.1, 5)'$	88
4.2	Model selection results based on simulated samples. The parameters are $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$, $\theta = (0.5, 5)'$	89
4.3	Model selection results based on simulated samples. The parameters are $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$, $\theta = (0.1, 10)'$	89
4.4	Model selection results based on simulated samples. The parameters are $\beta = (0.2, 0.3, 0.4, 0.5, 0.7, 0.8, -0.1, -0.6, -0.9, -1, \mathbf{0}_{10})'$, $\theta = (0.1, 5)'$	90
4.5	Model selection results based on simulated samples. The parameters are $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{20})'$, $\theta = (0.1, 5)'$	90
4.6	Description of covariates in Lyme disease data.	92

4.7 Table of coefficients (separate models for two subregions). 96

Chapter 1 General Introduction

1.1 Background

1.1.1 Reliability Analysis

In today's expanding and highly competitive market, there is a growing need for new generations of high-technology products. Only by ensuring high quality and reliability of products can make the company competitive and attractive in the market. For companies to build their reputation, the reliability of products plays a significant role. A reliable product may not remarkably improve customer satisfaction, while an unreliable product will most certainly lead to dissatisfactions. Therefore, high reliability is an obligatory prerequisite to acquire customer satisfaction. Additionally, unreliable products may also negatively affect profits because of the replacement, repair and maintenance costs.

Reliability is a characteristic that reflects the ability of a product to carry out its intended function for a specified time interval under declared conditions. There are two main types of reliability data: laboratory test data and field data. Before a company releases a new product, laboratory tests are usually performed to obtain the time to failure information of the product based on well-designed experiments. For traditional time to failure data, log-location-scale family of distributions (e.g., the Weibull and lognormal distributions) are often used to describe their distributions. Likelihood approaches are usually used for estimation. Predictions of reliability of products are usually needed. The predictions can help to evaluate the reliability of product and improve the product design.

1.1.1.1 Censoring

Censored data are common in reliability analysis. For highly reliable products, it may be hard to obtain failures for all tested units in a limited time period even under adverse

working conditions. Censoring schemes include right censoring, left censoring, and interval censoring. Right censoring happens when the test unit still works at the stopping time of the experiment. In this case, we only know the service life of the unit is longer than the experiment time, but the exact time of failure is unknown. Much research has been done on censored data, one may refer to Escobar and Meeker (1998), and Lawless (2003) for a class of statistical methods for censored data.

1.1.1.2 Accelerated Testing

For products designed to have high reliability, it is often unpractical to observe failures of products under normal operating conditions during the laboratory test period, which makes the statistical analysis challenging. In order to observe failures in a relatively short time period, laboratory tests are usually conducted in accelerated test conditions. That is, the reliability information is obtained by testing units under high levels of stress variables. The stress variables usually include temperature, humidity, usage frequency, work load, etc. Extrapolation is often used to acquire reliability information under normal use conditions. One may refer to Escobar and Meeker (2006) for a review of accelerated testing models.

1.1.1.3 Degradation Data

Degradation data offer an important resource for assessing reliability. For a wide variety of products, the failure mechanisms are related with certain underlying degradation processes. An example is that the tensile strength of fibre composite degrades over time in a normal use environment, and a failure can be considered to occur if the tensile strength reaches a threshold value. Compared to the traditional time to failure data, degradation data contain a sequence of measurements of some characteristic of product that is directly related to the failure of the product. We can not only obtain the failure information, but also obtain a sense of how the performance of the product degrades over time. General degradation path models and stochastic models are two different approaches for degradation data. For work

related to general degradation path models, one may see Lu and Meeker (1993), Wang and Coit (2007) and Hong et al. (2015) for reference. For using stochastic processes to analyze degradation data, one may refer to Padgett and Tomlinson (2004), and Park and Padgett (2006).

If the products haven't reached the threshold value at the end of the experiment, extrapolation can be used to estimate the failure time based on the estimated degradation path. Therefore, degradation analysis can be used for prediction of the lifetime of highly reliable products, with a few number of observed failures or even no failures.

1.1.1.4 Accelerated Destructive Degradation Test

Degradation data can be obtained under accelerated conditions. For example, we can expose fibre composite under a high temperature to expedite the degradation process. There are situations where the test unit will be destroyed during the measurement process, in other words, the measurement is destructive. In that case, we can only obtain one measurement per test unit. This type of test refers to as accelerated destructive degradation test (ADDT).

1.1.2 Spatial Epidemiology

Spatial epidemiology focuses on the description and examination of disease risk or incidence and its geographic variations. From a public health perspective, detecting disease patterns, discovering crucial factors associated with disease transmission, and making predictions of disease incidence rates are important for better understanding of the diseases, enforcing decisions that contribute to prevention, and inhibiting the spread of disease. It is appealing to use statistical methods for epidemiological research (Lawson, 2013). It is important to note that the incidences of disease at proximal locations have positive or negative correlation. Accounting for spatial dependency in analysis is important but also challenging.

1.1.2.1 Generalized Linear Mixed Models

The complexity of spatial epidemiology data requires models which can describe discrete outcomes (e.g., case counts of disease) and capture correlations among observations due to different spatial locations. Generalized linear mixed models (GLMMs) are an extension of linear mixed models and generalized linear models. GLMMs have wide applications in the cases where responses have different distributions other than normal distributions (e.g., Poisson and logistic distributions) and data are correlated rather than independent.

Denote \mathbf{y}_i as the vector of responses for cluster i , $i = 1, \dots, n$, where n is number of clusters. And \mathbf{b}_i is a vector of random effects with multivariate distribution $N(\mathbf{0}, \mathcal{D})$. \mathcal{D} is the variance-covariance matrix. Conditional on \mathbf{b}_i , the \mathbf{y}_i 's are independent. We assume the conditional mean and variance of \mathbf{y}_i have the following forms:

$$E(\mathbf{y}_i|\mathbf{b}_i) = \boldsymbol{\mu}_i^b,$$

$$\text{Var}(\mathbf{y}_i|\mathbf{b}_i) = \phi a_i v(\boldsymbol{\mu}_i^b),$$

where ϕ is a scale parameter, a_i is a known constant, and $v(\cdot)$ is a function of conditional mean $\boldsymbol{\mu}_i^b$. The conditional mean is linked with the linear predictor $\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{b}_i$ by a known function $\eta(\cdot)$. That is, $\eta(\boldsymbol{\mu}_i^b) = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{b}_i$. Here \mathbf{x}_i and \mathbf{z}_i are covariates associated with fixed effects $\boldsymbol{\beta}$ and random effects \mathbf{b}_i , respectively.

The likelihood function can be expressed as

$$L(\boldsymbol{\beta}, \phi, \mathcal{D}|\mathbf{y}_1, \dots, \mathbf{y}_n) = \prod_{i=1}^n \int f(\mathbf{y}_i|\boldsymbol{\beta}, \phi, \mathbf{b}_i) f(\mathbf{b}_i|\mathcal{D}) d\mathbf{b}_i,$$

where $f(\mathbf{y}_i|\boldsymbol{\beta}, \phi, \mathbf{b}_i)$ is the distribution of \mathbf{y}_i conditional on \mathbf{b}_i , and $f(\mathbf{b}_i|\mathcal{D})$ is the distribution of \mathbf{b}_i .

The likelihood approach for estimating unknown parameters in GLMMs involves integrals over random effects \mathbf{b}_i . There is no general analytic expressions for solutions and numerical

approximations are needed. Details regarding estimation and inference of GLMMs can be found in Molenberghs and Verbeke (2005, Chapter 14) and Stroup (2012).

Bayesian methodology for estimation and inference is also popular because of the hierarchical model formulation of GLMM. By assuming a prior distribution $f(\boldsymbol{\beta}, \phi, \mathcal{D})$, the posterior distribution of $(\boldsymbol{\beta}, \phi, \mathcal{D})'$ is

$$f(\boldsymbol{\beta}, \phi, \mathcal{D} | \mathbf{y}_1, \dots, \mathbf{y}_n) \propto L(\boldsymbol{\beta}, \phi, \mathcal{D} | \mathbf{y}_1, \dots, \mathbf{y}_n) f(\boldsymbol{\beta}, \phi, \mathcal{D}).$$

For methods of sampling from the posterior distribution, one may refer to Ripley (2009), Zeger and Karim (1991), and Gamerman (1997).

1.1.2.2 Variable Selection Methods

In practice, many explanatory variables may be collected and included in the model. However, it is often the case that only a subset of predictors contribute to the outcome of interest. By identifying the important variables, we can obtain a good model fit, improve the prediction performance, and enhance the model interpretation. Thus, variable selection is often necessary and critical. In spatial epidemiology, discovering important factors associated with disease incidence can inform decision-making on disease control and prevention.

There is a considerable amount of research on variable selection methods in regression models. Classical variable selection methods include backward, forward, stepwise, and all subset selection. Despite of the fact that these subset selection methods are widely accepted and extensively used, they still have many drawbacks. Backward, forward, stepwise procedures may not select the same model (i.e., these procedures are only sub-optimal). All subset selection procedure is time-consuming if there is a large number of covariates. See Ratner (2010) for more details.

Alternative methods for variable selection are based on likelihood with penalty functions. Examples include the ridge penalty (Hoerl and Kennard, 1970), the least absolute selection

and shrinkage operator (LASSO) (Tibshirani, 1996), the elastic net (Zou and Hastie, 2005), the adaptive LASSO (Zou, 2006), the adaptive elastic net (Zou and Zhang, 2009), and the smoothly clipped absolute deviation penalty (SCAD) (Fan and Li, 2001). Information theoretic criteria such as Akaike’s information criterion (AIC), Schwarz’s Bayesian information criterion (BIC) and deviance information criterion (DIC) are available for model comparisons. Predictive measures may also be used to select the best model.

Bayesian variable selection methods are also popular. Posterior model probabilities are comparable and have straightforward and meaningful interpretations. Therefore, it’s natural and reasonable to select the model with the highest posterior model probability.

Regarding the use of penalty functions to perform variable selection in GLMM, one may refer to Schelldorfer, Meier, and Bühlmann (2014), Groll and Tutz (2014), Yang (2007) and Cui (2011). Besides, Cai and Dunson (2008) described the Bayesian variable selection methods in GLMM.

1.2 Motivation

This section describes the motivations for the three projects in the dissertation.

1.2.1 Simultaneous Prediction Interval

In many applications, we are more interested in the prediction intervals which account for the uncertainty instead of point predictions. There also arises a situation when a company produces a batch of bulbs, the question of interest may not be the prediction interval of the mean lifetime of the batch of bulbs, but rather the simultaneous prediction interval/bound that contains the lifetime of at least 95% of bulbs in the batch with a specified confidence level. We present a statistical procedure for computing simultaneous prediction intervals/bounds for at least k out of m ($k \leq m$) future observations in Chapter 2.

1.2.2 Semi-parametric Model for ADDT Data

Current research regarding ADDT data usually assumes parametric models for degradation path, while in general, the parametric models may not be suitable for a wide class of products. However, due to the flexibility of splines, semi-parametric models based on splines can be applied to different kinds of products. We propose a semi-parametric model for ADDT data and describe the estimation and inference procedures in Chapter 3.

1.2.3 Spatial Variable Selection

This project is motivated by the Lyme disease data in Virginia from 2006 to 2011. Lyme disease emergence has attracted considerable attention recently. There is an increasing trend in number of Lyme disease cases in Virginia after 2006. Jackson, Hilborn, and Thomas (2006), Allan, Keesing, and Ostfeld (2003) and Yahner (1992) showed that Lyme disease is associated with percentage of forest, number of small forested fragments, etc. Seukep et al. (2015) considered a spatial Poisson regression model to examine the relationship between forest fragmentation, land cover types, land cover change, demographic information and the incidence of Lyme disease in Virginia. Conditional autoregressive model (CAR) was used to describe the spatially correlated random effects in their study. Seukep et al. (2015) applied the principal components analysis to attain important components and overcame the multicollinearity problem in the original set of predictors. Their approach did not involve selection of covariates. In Chapter 4, we present variable selection procedures in GLMM while taking spatial correlation and multicollinearity problem into account. Our approaches are applied to the Lyme disease data to discover the crucial factors.

1.3 Outline of the Dissertation

The rest of this dissertation is organized as follows. In Chapter 2, we develop a procedure to construct simultaneous prediction intervals/bounds for at least k out of m ($k \leq m$) future

observations. The procedure can be applied to (log)-location-scale family of distributions with complete or censored data. Chapter 2 is based on Xie et al. (2015a). Chapter 3 proposes a semi-parametric degradation model to analyze the ADDT data. Monotone B-splines are used to model the degradation path. Chapter 3 is mainly based on Xie, King, and Hong (2015b). Chapter 4 introduces the Lyme disease emergence in Virginia and describes variable selection methods for GLMM while accounting for spatial correlations. Chapter 4 is based on Xie et al. (2016). Finally, Chapter 5 contains general conclusions and areas for future work.

Bibliography

- B. F. Allan, F. Keesing, and R. S. Ostfeld. Effect of forest fragmentation on Lyme disease risk. *Conservation Biology*, 17(1):267–272, 2003.
- B. Cai and D. B. Dunson. Bayesian variable selection in generalized linear mixed models. In *Random Effect and Latent Variable Model Selection*, pages 63–91. Springer, 2008.
- R. Cui. *Variable selection methods for longitudinal data*. PhD thesis, Harvard University, 2011.
- L. A. Escobar and W. Q. Meeker. Fisher information matrices with censoring, truncation, and explanatory variables. *Statistica Sinica*, 8(1):221–237, 1998.
- L. A. Escobar and W. Q. Meeker. A review of accelerated test models. *Statistical Science*, 21(4):552–577, 2006.
- J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.
- D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68, 1997.
- A. Groll and G. Tutz. Variable selection for generalized linear mixed models by L_1 -penalized estimation. *Statistics and Computing*, 24(2):137–154, 2014.
- A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- Y. Hong, Y. Duan, W. Q. Meeker, D. L. Stanley, and X. Gu. Statistical methods for degradation data with dynamic covariates information and an application to outdoor weathering data. *Technometrics*, 57(2):180–193, 2015.

- L. E. Jackson, E. D. Hilborn, and J. C. Thomas. Towards landscape design guidelines for reducing Lyme disease risk. *International Journal of Epidemiology*, 35(2):315–322, 2006.
- A. B. Lawson. *Statistical methods in spatial epidemiology*. John Wiley & Sons, 2013.
- C. J. Lu and W. O. Meeker. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2):161–174, 1993.
- G. Molenberghs and G. Verbeke. *Models for Discrete Longitudinal Data. Springer Series in Statistics*. Springer, 2005.
- W. Padgett and M. A. Tomlinson. Inference from accelerated degradation and failure data based on Gaussian process models. *Lifetime Data Analysis*, 10(2):191–206, 2004.
- C. Park and W. J. Padgett. Stochastic degradation models with several accelerating variables. *IEEE Transactions on Reliability*, 55(2):379–390, 2006.
- B. Ratner. Variable selection methods in regression: Ignorable problem, outing notable solution. *Journal of Targeting, Measurement and Analysis for Marketing*, 18(1):65–75, 2010.
- B. D. Ripley. *Stochastic simulation*. John Wiley & Sons, 2009.
- J. Schelldorfer, L. Meier, and P. Bühlmann. Glmmlasso: An algorithm for high-dimensional generalized linear mixed models using L_1 -penalization. *Journal of Computational and Graphical Statistics*, 23(2):460–477, 2014.
- S. E. Seukey, K. N. Kolivras, Y. Hong, J. Li, S. P. Prisley, J. B. Campbell, D. N. Gaines, and R. L. Dymond. An examination of the demographic and environmental variables correlated with Lyme disease emergence in Virginia. *EcoHealth*, 12(4):634–644, 2015.
- W. W. Stroup. *Generalized linear mixed models: modern concepts, methods and applications*. CRC Press, 2012.

- R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.
- P. Wang and D. W. Coit. Reliability and degradation modeling with random or uncertain failure threshold. In *Reliability and Maintainability Symposium, 2007. RAMS'07. Annual*, pages 392–397. IEEE, 2007.
- Y. Xie, Y. Hong, L. A. Escobar, and W. Q. Meeker. Simultaneous prediction intervals for the (Log)-Location-Scale family of distributions. Under review, 2015a.
- Y. Xie, C. B. King, and Y. Hong. A semi-parametric model for accelerated destructive degradation test data analysis. Under review, 2015b.
- Y. Xie, J. Li, X. Deng, Y. Hong, and K. Kolivras. Spatial variable selection via elastic net and an application to Virginia Lyme disease emergence. In progress, 2016.
- R. H. Yahner. Dynamics of a small mammal community in a fragmented forest. *American Midland Naturalist*, 127(2):381–391, 1992.
- H. Yang. *Variable selection procedures for generalized linear mixed models in longitudinal data analysis*. PhD thesis, North Carolina State University, 2007.
- S. L. Zeger and M. R. Karim. Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):79–86, 1991.
- H. Zou. The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.
- H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67(2):301–320, 2005.
- H. Zou and H. H. Zhang. On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, 37(4):1733–1751, 2009.

Chapter 2 Simultaneous Prediction Intervals for the (Log)-Location-Scale Family of Distributions

Abstract

Making predictions of future realized values of random variables based on currently available data is a frequent task in statistical applications. In some applications, the interest is to obtain a two-sided simultaneous prediction interval (SPI) to contain at least k out of m future observations with a certain confidence level based on n previous observations from the same distribution. A closely related problem is to obtain a one-sided upper (or lower) simultaneous prediction bound (SPB) to exceed (or be exceeded) by at least k out of m future observations. In this chapter, we provide a general approach for constructing SPIs and SPBs based on data from a member of the (log)-location-scale family of distributions with complete or right censored data. The proposed simulation-based procedure can provide exact coverage probability for complete and Type II censored data. For Type I censored data, the simulation results show that our procedure provides satisfactory results in small samples. We use three applications to illustrate the proposed simultaneous prediction intervals and bounds.

Key Words: Censored Data; Coverage Probability; k out of m , Lognormal; Simulation; Weibull.

2.1 Introduction

2.1.1 Motivation

Prediction intervals are used to quantify the uncertainty associated with future realized values of random variables. In predicting future outcomes, one might be interested in point predictions. Often, however, the focus is on whether the future observations will fall within a prediction interval (PI) or conforming to a one-sided prediction bound (PB) obtained from the available data and a pre-specified confidence level.

In some applications, it is desirable to obtain a two-sided simultaneous prediction interval (SPI) or a one-sided simultaneous prediction bound (SPB) for at least k out of m future observations, where $1 \leq k \leq m$. For example, Fertig and Mann (1977) consider time to failure of turbine nozzles subject to a certain load. The company had manufactured 50 nozzles. Based on the failure times in a life test of 10 of those nozzles, they obtained a 95% lower prediction bound to be exceeded by at least 90% of the remaining 40 nozzles (i.e., 36 out of 40). In another study, Fertig and Mann (1977) use failure times (in hours) based on a life test of aircraft components to obtain an SPI to contain the failure times of all 10 future components.

Much research has been done for statistical prediction for a single future observation. Details and additional references can be found in Mee and Kushary (1994) and Escobar and Meeker (1999). There has been some work for the SPIs/SPBs for at least k out of m future observations. Those procedures, however, have been developed only for specific distributions (e.g., normal and Weibull distributions). Hence, it is desirable to have a general approach to generate SPIs/SPBs for a general class of distributions. In this chapter, we develop a general procedure to obtain SPIs and SPBs for the location-scale family and the log-location-scale family of distributions. The proposed procedures can be used with complete or censored data and can be extended, in an approximate manner, to other distributions.

2.1.2 Literature Review and Contributions of This Work

There is some previous work on the construction of SPIs/SPBs to contain/bound at least k out of m future observations. Danziger and Davis (1964) described and provided tables of coverage probabilities for non-parametric SPIs to contain k out of m future observations (which they refer to as tolerance intervals) and corresponding one-sided SPBs. Hahn (1969) considered the special case of $k = m$ based on observations from a normal distribution. Hahn (1969) gave the factors to calculate two-sided SPIs. One-sided SPBs were considered in Hahn (1970). Fertig and Mann (1977) presented factors for constructing one-sided SPBs to contain at least k out of m future observations for a normal distribution. Odeh (1990) provided a method for generating k out of m two-sided SPIs for a normal distribution. Due to computational limitations, these papers only provided factors for a limited number of combinations of n, k, m and for some specified confidence levels. In the area of environmental monitoring, some articles considered the use of SPIs/SPBs for at least k out of m future observations at p locations. Davis and McNichols (1987) studied this type of problem for one-sided prediction bounds and for observations from a normal distribution. Krishnamoorthy, Lin, and Xia (2009) constructed one-sided upper prediction bounds for the Weibull distribution based on generalized pivotal quantities. Bhaumik and Gibbons (2006) developed an approximate upper SPB for samples from a gamma distribution. Bhaumik (2008) constructed a one-sided SPB for left-censored normal random variables. Beran (1990) gives theoretical results on the coverage properties of the prediction regions based on simulation. There are no methods in the literature for two-sided SPIs for the Weibull distribution.

None of existing literature proposes a general procedure for the location-scale (e.g., the smallest extreme value, normal, and largest extreme value distributions) or the related log-location-scale family of distributions (e.g., the Weibull, lognormal, and Fréchet distributions). In this chapter, we develop methods for constructing such intervals/bounds based on a general procedure. The methods are exact (except for Monte Carlo error) for complete and Type II censored data. Type I censoring is commonly in life tests. We use simulation to

study the coverage properties for the approximate intervals/bounds under Type I censoring.

2.1.3 Overview

The rest of this chapter is organized as follows. Section 2.2 introduces the data and model setting for the problem. Section 2.3 gives the formal definition of the proposed SPI procedure. Section 2.4 proposes a general procedure to obtain an SPI, followed by illustrative examples. Section 2.5 describes simulation studies on the performance of the proposed procedure for Type I censored data. Section 2.6 illustrates the use of the proposed method with applications. Section 2.7 contains concluding remarks and some discussion about related extensions and applications of the methods.

2.2 Data, Model, and Maximum Likelihood Estimation

2.2.1 Data

We consider situations in which n independent experimental units are under study. At the moment of doing the analysis, the data consist of: (a) r exact observations and (b) a set of $(n - r)$ right-censored observations at x_c , where x_c is larger or equal to the maximum of the exact observations. Three important special cases of these data structure are: (a) complete data, when $r = n$; (b) Type II censored data, when r ($2 \leq r \leq n$) is pre-specified and x_c is equal to the maximum of the exact observations. Note that in this case x_c is random; (c) Type I censored data, when x_c is pre-specified and x_c exceeds the maximum of the exact observations. Note that in the case of Type I censoring, r ($1 \leq r \leq n$) is random (if $r = 0$ the maximum likelihood (ML) estimate does not exist).

To be precise, let $\mathbf{X} = (X_1, \dots, X_n)$ denote the random variables for the observations

from the n units, where $-\infty < X_i < \infty, i = 1, \dots, n$. Define

$$\delta_i = \begin{cases} 1, & \text{if } X_i \text{ is an exact observation} \\ 0, & \text{if } X_i \text{ is a right-censored observation.} \end{cases}$$

For Type I and Type II censoring, we observe $x_i = \min(X_i, x_c)$ and $\delta_i, i = 1, \dots, n$. The observed values are denoted by $\mathbf{x} = (x_1, \dots, x_n)$. This data structure is general and includes data from reliability and lifetime studies with right-censored data from a positive response. In this case all the components of \mathbf{X} take positive values.

2.2.2 Model

To construct an SPI for a set of future observations, we use a statistical model to describe the population of interest. In this chapter, we assume the observations have a distribution in the family of the location-scale or log-location-scale family of distributions. A location-scale distribution has a location parameter μ and a scale parameter σ . The parameters μ and σ are typically unknown and need to be estimated. The probability density function (pdf) and the cumulative distribution function (cdf) of a location-scale distribution are

$$f(x) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) \quad \text{and} \quad F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

respectively. The definitions of the standard pdf $\phi(\cdot)$ and cdf $\Phi(\cdot)$ functions for the different members of this family are given in Table 2.1.

The pdf and cdf of the log-location-scale family are

$$f(t) = \frac{1}{\sigma t} \phi\left[\frac{\log(t) - \mu}{\sigma}\right] \quad \text{and} \quad F(t) = \Phi\left[\frac{\log(t) - \mu}{\sigma}\right],$$

respectively. The Weibull, lognormal, Fréchet, and log-logistic distributions are members of the log-location-scale family. For these distributions, σ is a shape parameter and $\exp(\mu)$ is

Table 2.1: The pdfs and cdfs of different commonly-used members of the standard location-scale and log-location-scale distributions.

Location-Scale	Log-Location-Scale	pdf $\phi(x)$	cdf $\Phi(x)$
Normal	Lognormal	$\frac{\exp(-x^2/2)}{\sqrt{2\pi}}$	$\int_{-\infty}^x \phi(w) dw$
Logistic	Loglogistic	$\frac{\exp(x)}{[1 + \exp(x)]^2}$	$\frac{\exp(x)}{1 + \exp(x)}$
Largest extreme value	Fréchet	$\exp[-x - \exp(-x)]$	$\exp[-\exp(-x)]$
Smallest extreme value	Weibull	$\exp[x - \exp(x)]$	$1 - \exp[-\exp(x)]$

a scale parameter. In the remainder of this chapter, however, we will refer to μ and σ as location and scale parameters, respectively.

This chapter focuses on the construction of SPIs and SPBs containing at least k of m future observations $\mathbf{Y} = (Y_1, \dots, Y_m)$ from a previously sampled population. The sample data are denoted by \mathbf{X} and the assumptions are that \mathbf{Y} and \mathbf{X} are independent and random samples from the same distribution.

2.2.3 Maximum Likelihood Estimation

We use maximum likelihood (ML) to estimate the unknown parameters (μ, σ) . Under the independent and identically distributed (i.i.d.) assumptions in Section 2.2, the likelihood of the right censored data has the form

$$L(\mu, \sigma) = \mathcal{C} \prod_{i=1}^n [f(x_i; \mu, \sigma)]^{\delta_i} [1 - F(x_i; \mu, \sigma)]^{1-\delta_i},$$

where \mathcal{C} is a constant that does not depend on μ or σ , $f(x_i; \mu, \sigma)$ is the assumed pdf, and $F(x_i; \mu, \sigma)$ is the corresponding cdf. The ML estimates can be obtained by finding the values of μ and σ that maximize the likelihood function. In general, there is no closed-form expression for the ML estimates, which are denoted by $(\hat{\mu}, \hat{\sigma})$. Consequently, numerical methods are used to find the ML estimates.

2.3 Simultaneous Prediction Intervals and Bounds

2.3.1 Two-sided Simultaneous Prediction Intervals

This section shows how to construct an SPI $[L(\mathbf{x}, 1 - \alpha), U(\mathbf{x}, 1 - \alpha)]$ that will contain at least k out of m independent future observations from the sampled distribution, with a specified confidence level $1 - \alpha$. Conditioning on the observed data $\mathbf{X} = \mathbf{x}$, the conditional coverage probability (CP) of the interval $[L(\mathbf{x}, 1 - \alpha), U(\mathbf{x}, 1 - \alpha)]$ with nominal confidence level $1 - \alpha$ is

$$\begin{aligned} \text{CP}(\boldsymbol{\theta} | \mathbf{X} = \mathbf{x}) &= \Pr\{\text{at least } k \text{ of } m \text{ values lie in } [L(\mathbf{x}, 1 - \alpha), U(\mathbf{x}, 1 - \alpha)] | \mathbf{X} = \mathbf{x}\} \\ &= \sum_{j=k}^m \binom{m}{j} p^j (1-p)^{m-j}, \end{aligned} \quad (2.1)$$

where $\boldsymbol{\theta} = (\mu, \sigma)$ is the vector of unknown parameters and

$$p = \Pr\{\text{a future observation is in } [L(\mathbf{x}, 1 - \alpha), U(\mathbf{x}, 1 - \alpha)] | \mathbf{X} = \mathbf{x}\}.$$

The conditional CP is unobservable because it depends on the unknown parameters and varies from sample to sample because it depends on the data. Following standard procedure, to evaluate the prediction interval procedure, we use the unconditional CP

$$\text{CP}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{X}} \left[\sum_{j=k}^m \binom{m}{j} p^j (1-p)^{m-j} \right],$$

where expectation is taken with respect to the joint distribution of the data \mathbf{X} .

Because $(Y_i - \hat{\mu})/\hat{\sigma}, i = 1, \dots, m$ are pivotal quantities, one can construct a two-sided $100(1 - \alpha)\%$ SPI to contain at least k out of m future observations with the following form

$$[\hat{\mu} + u_L(k, m; \alpha)\hat{\sigma}, \hat{\mu} + u_U(k, m; \alpha)\hat{\sigma}],$$

for the location-scale family of distributions. Here $u_L(k, m; \alpha)$ and $u_U(k, m; \alpha)$ are factors to be chosen so that the SPI will have CP equal to $1 - \alpha$. For notational simplicity, we let $u_L = u_L(k, m; \alpha)$ and $u_U = u_U(k, m; \alpha)$. In particular, the factors (u_L, u_U) satisfy the equation

$$1 - \alpha = \int_0^\infty \int_{-\infty}^\infty \sum_{j=k}^m \binom{m}{j} [\Phi(a) - \Phi(b)]^j [1 - \Phi(a) + \Phi(b)]^{m-j} f_{\mathbf{Z}}(z_1, z_2) dz_1 dz_2, \quad (2.2)$$

where $a = z_1 + u_U z_2$, $b = z_1 + u_L z_2$, $\mathbf{Z} = (Z_1, Z_2)$, $f_{\mathbf{Z}}(z_1, z_2)$ is the joint pdf of $Z_1 = (\hat{\mu} - \mu)/\sigma$ and $Z_2 = \hat{\sigma}/\sigma$, and $\Phi(\cdot)$ is the standard cdf of X .

The proof of (2.2) is given in Appendix 2.A. Note that (2.2) can be written as

$$1 - \alpha = \mathbf{E}_{\mathbf{Z}} \left[\sum_{j=k}^m \binom{m}{j} [\Phi(A) - \Phi(B)]^j [1 - \Phi(A) + \Phi(B)]^{m-j} \right], \quad (2.3)$$

where $A = Z_1 + u_U Z_2$, $B = Z_1 + u_L Z_2$, and $\mathbf{E}_{\mathbf{Z}}(\cdot)$ is the expectation with respect to the joint distribution of \mathbf{Z} .

For distributions in the log-location-scale family, the corresponding two-sided $100(1 - \alpha)\%$ SPI to contain at least k out of m future observations has the form $[\exp(\hat{\mu} + u_L \hat{\sigma}), \exp(\hat{\mu} + u_U \hat{\sigma})]$. Thus, (2.2) is still used to obtain a prediction interval for distributions in the log-location-scale family.

For Type II censored data or complete data from the location-scale/log-location-scale family of distributions, Lawless (2003, pages 217 and 262) describes the pivotal property of \mathbf{Z} . That is, the distribution of \mathbf{Z} does not depend on unknown parameters. For Type I censoring, the pivotal property of \mathbf{Z} no longer holds. The quantity \mathbf{Z} , however, can be treated as being approximately pivotal. Thus we can still use (2.2) to get the approximate asymptotically correct SPIs under Type I censoring, and other types of non-informative censoring.

2.3.2 One-sided Simultaneous Prediction Bounds

There are similar CP statements for one-sided simultaneous prediction bounds. In particular, for a one-sided lower simultaneous prediction bound, the conditional CP is

$$\begin{aligned} \text{CP}_L(\boldsymbol{\theta}|\mathbf{X} = \mathbf{x}) &= \Pr[\text{at least } k \text{ of } m \text{ values are larger than } L(\mathbf{x}, 1 - \alpha)|\mathbf{X} = \mathbf{x}] \\ &= \sum_{j=k}^m \binom{m}{j} p^j (1-p)^{m-j}, \end{aligned} \quad (2.4)$$

where $p = \Pr[\text{a single future observation is larger than } L(\mathbf{x}, 1 - \alpha)|\mathbf{X} = \mathbf{x}]$.

The unconditional CP is

$$\text{CP}_L(\boldsymbol{\theta}) = \mathbf{E}_{\mathbf{X}} \left[\sum_{j=k}^m \binom{m}{j} p^j (1-p)^{m-j} \right].$$

For the location-scale family of distributions, a one-sided lower simultaneous prediction bound to be exceeded by at least k out of m future observations can be expressed as $L(\mathbf{x}, 1 - \alpha) = \hat{\mu} + u'_L(k, m; \alpha)\hat{\sigma}$, where $u'_L(k, m; \alpha)$ is a factor to be chosen so that the interval will give a CP of $1 - \alpha$. Let $u'_L = u'_L(k, m; \alpha)$ and note that u'_L satisfies the equation

$$\begin{aligned} 1 - \alpha &= \int_0^\infty \int_{-\infty}^\infty \sum_{j=k}^m \binom{m}{j} [1 - \Phi(b)]^j [\Phi(b)]^{m-j} f_{\mathbf{Z}}(z_1, z_2) dz_1 dz_2 \\ &= \mathbf{E}_{\mathbf{Z}} \left[\sum_{j=k}^m \binom{m}{j} [1 - \Phi(B)]^j [\Phi(B)]^{m-j} \right], \end{aligned} \quad (2.5)$$

where $b = z_1 + u'_L z_2$ and $B = Z_1 + u'_L Z_2$. When $k = m$, one obtains the lower prediction bound to contain all m new additional observations.

Similarly, for a one-sided upper simultaneous prediction bound, the conditional CP is

$$\begin{aligned} \text{CP}_U(\boldsymbol{\theta}|\mathbf{X} = \mathbf{x}) &= \Pr[\text{at least } k \text{ of } m \text{ values are less than } L(\mathbf{x}, 1 - \alpha)|\mathbf{X} = \mathbf{x}] \\ &= \sum_{j=k}^m \binom{m}{j} p^j (1 - p)^{m-j}, \end{aligned} \quad (2.6)$$

where

$$p = \Pr[\text{a single future observation is less than } L(\mathbf{x}, 1 - \alpha)|\mathbf{X} = \mathbf{x}].$$

The unconditional CP is

$$\text{CP}_U(\boldsymbol{\theta}) = \mathbf{E}_{\mathbf{X}} \left[\sum_{j=k}^m \binom{m}{j} p^j (1 - p)^{m-j} \right].$$

A one-sided upper simultaneous prediction bound to exceed at least k out of m future observations for the location-scale family of distributions is $U(\mathbf{x}, 1 - \alpha) = \hat{\mu} + u'_U(k, m; \alpha)\hat{\sigma}$, where $u'_U(k, m; \alpha)$ is a factor to be chosen so that the interval will give a CP equal to $1 - \alpha$. Let $u'_U = u'_U(k, m; \alpha)$ and note that u'_U satisfies the equation

$$\begin{aligned} 1 - \alpha &= \int_0^\infty \int_{-\infty}^\infty \sum_{j=k}^m \binom{m}{j} [\Phi(a)]^j [1 - \Phi(a)]^{m-j} f_{\mathbf{Z}}(z_1, z_2) dz_1 dz_2 \\ &= \mathbf{E}_{\mathbf{Z}} \left[\sum_{j=k}^m \binom{m}{j} [\Phi(A)]^j [1 - \Phi(A)]^{m-j} \right], \end{aligned} \quad (2.7)$$

where $a = z_1 + u'_U z_2$ and $A = Z_1 + u'_U Z_2$.

For the log-location-scale family of distributions, the lower and upper SPBs have the form $L(\mathbf{x}, 1 - \alpha) = \exp(\hat{\mu} + u'_L \hat{\sigma})$ and $U(\mathbf{x}, 1 - \alpha) = \exp(\hat{\mu} + u'_U \hat{\sigma})$, respectively. The factors u'_L and u'_U are obtained as solutions of (2.5) and (2.7), respectively.

2.4 Computations of the Simultaneous Prediction Intervals/Bounds

In this section, we introduce a general procedure for finding the factors so that the two-sided SPIs and one-sided SPBs will have the correct CP. The computing procedure requires solving equations (2.3), (2.5), and (2.7). In general, there is no closed-form expression for the solution of these equations. The exact distribution of \mathbf{Z} can be complicated, especially with censored data. Therefore, we use Monte Carlo simulation to obtain the distribution of \mathbf{Z} and evaluate the expectation based on the simulated samples.

2.4.1 Complete and Type II Censored Data

The two-sided SPI for complete or Type II censored data can be obtained from the following algorithm.

Algorithm 1:

1. Draw a complete or Type II censored sample of size n from a (log)-location-scale family of distributions with $(\mu, \sigma) = (0, 1)$. Detailed discussion on efficient simulation of censored samples can be found in Meeker and Escobar (1998, Section 4.13).
2. Repeat step 1 B_1 times and compute ML estimates $(\hat{\mu}_l^*, \hat{\sigma}_l^*)$ for each simulated sample, $l = 1, \dots, B_1$.

To save computing time, these $(\hat{\mu}_l^*, \hat{\sigma}_l^*)$ values are stored and used to compute all the SPIs and SPBs for the particular censoring specification (n, r) as shown below.

3. For every (u_L, u_U) , in a collection of chosen values, compute

$$\text{CP}^*(u_L, u_U) = \frac{1}{B_1} \sum_{l=1}^{B_1} \left\{ \sum_{j=k}^m \binom{m}{j} p_l(u_L, u_U)^j [1 - p_l(u_L, u_U)]^{m-j} \right\}, \quad (2.8)$$

where $p_l(u_L, u_U) = \Phi(\hat{\mu}_l^* + u_U \hat{\sigma}_l^*) - \Phi(\hat{\mu}_l^* + u_L \hat{\sigma}_l^*)$ and $u_L < u_U$.

4. Find (u_L, u_U) such that $\text{CP}^*(u_L, u_U) = 1 - \alpha$.

Note that the choice of $(\mu, \sigma) = (0, 1)$ in Step 1 above is justified because for the Type II censored and complete data case, the **Algorithm 1** procedure does not depend on unknown parameters due to the pivotal property of \mathbf{Z} .

Finding (u_L, u_U) such that $\text{CP}^*(u_L, u_U) = 1 - \alpha$ is a two-dimensional root-finding problem and there are multiple solutions. An additional constraint on u_L and u_U is needed for a unique solution. For symmetric distributions, $u_L = -u_U$ is an appropriate constraint and leads to two-sided SPIs with equal tail probabilities. For non-symmetric distributions, the two-sided SPI with equal tail probabilities is appealing from a practical point of view. The computation, however, is more complicated. Detailed discussion of the computation is given in Section 2.4.2.

For one-sided SPBs, modifications to the algorithm are needed. Specifically, for the lower SPB, replace (2.8) by

$$\text{CP}_L^*(u'_L) = \frac{1}{B_1} \sum_{l=1}^{B_1} \left\{ \sum_{j=k}^m \binom{m}{j} p_l(u'_L)^j [1 - p_l(u'_L)]^{m-j} \right\},$$

where $p_l(u'_L) = 1 - \Phi(\widehat{\mu}_l^* + u'_L \widehat{\sigma}_l^*)$. Then find the unique value of u'_L such that $\text{CP}_L^*(u'_L) = 1 - \alpha$. For the upper SPB, we need to replace (2.8) by

$$\text{CP}_U^*(u'_U) = \frac{1}{B_1} \sum_{l=1}^{B_1} \left\{ \sum_{j=k}^m \binom{m}{j} p_l(u'_U)^j [1 - p_l(u'_U)]^{m-j} \right\},$$

where $p_l(u'_U) = \Phi(\widehat{\mu}_l^* + u'_U \widehat{\sigma}_l^*)$. Then find the unique value of u'_U such that $\text{CP}_U^*(u'_U) = 1 - \alpha$. For one-sided prediction bounds, we use linear interpolation to obtain lower or upper limits based on the CP curve ($1 - \alpha$ versus u'_L or u'_U , respectively) for desired confidence levels.

2.4.2 Two-sided SPI with Equal Tail Probability

In applications, even involving a non-symmetric distribution, it is preferable to have a two-sided prediction interval with equal tail probabilities. For this purpose, we define the tail probability as the tail probability of the one-sided bound. Therefore, the equal tail probability implies that $CP_L(u_L) = CP_U(u_U)$. Except for the special case of $k = 1$ (i.e., a prediction interval for exactly one new observation), combining a one-sided lower $100(1 - \alpha_1)\%$ prediction bound and a one-sided upper $100(1 - \alpha_2)\%$ prediction bound will not provide a two-sided $100(1 - \alpha_1 - \alpha_2)\%$ SPI. Thus, a special procedure for a two-sided SPI with equal tail probabilities is needed. For a given confidence level $1 - \alpha$, we can obtain u_L and u_U by solving numerically the equations

$$CP(u_L, u_U) = 1 - \alpha \quad \text{and} \quad CP_L(u_L) - CP_U(u_U) = 0. \quad (2.9)$$

To find the solutions to (2.9), one finds numerically the contour lines of $CP(u_L, u_U)$ and $CP_L(u_L) - CP_U(u_U)$. Then use interpolation to locate the intersecting point of the contours. It is also possible to re-express the two-sided CP as a function of the one-sided tail probability to reduce the dimension of root-finding, and then find the common tail probability that gives the desired two-sided CP. Illustration of this method is given in Section 2.4.4.

2.4.3 Type I Censored Data

For Type I censored data, the statistics Z_1 and Z_2 are only approximately pivotal. The simulation procedure will depend on the censoring time (or more precisely, the estimated expected fraction failing). The expected fraction failing p_f is defined as $\Phi[(x_c - \mu)/\sigma]$. For Type I censoring, we use the following algorithm.

Algorithm 2:

1. For the observed Type I data, calculate the ML estimates $(\hat{\mu}, \hat{\sigma})$ and then compute

estimated expected fraction failing $\hat{p}_f = \Phi[(x_c - \hat{\mu})/\hat{\sigma}]$.

2. Draw a censored sample of size n from the (log)-location-scale family of distributions with $(\mu, \sigma) = (0, 1)$ and censoring time is $x_c = \Phi_{\hat{p}_f}^{-1}$.
3. Follow steps 2 to 4 in **Algorithm 1**.

As the sample size increases, the CP of the SPIs/SPBs for Type I censoring data computed by **Algorithm 2** will approach the nominal confidence level (see proof in Appendix 2.B). In Section 2.5, we study finite sample CPs for SPIs and SPBs obtained using **Algorithm 2**.

Appendix 2.C shows that the variance of estimated coverage probability $CP^*(u_L, u_U)$ (2.8) is a function of $(u_L, u_U, k, m, n, r/x_c, \Phi(\cdot), B_1)$. Hence, for given distribution and values of $(u_L, u_U, k, m, n, r/x_c)$, we can compute B_1 that provides desired precision.

2.4.4 Illustrative Examples

2.4.4.1 Illustration A: Upper SPB for Type II Censoring and Complete Data

For purpose of illustration, we generate the CP curve for a one-sided upper SPB for at least 4 out of 5 future observations from a previous sampled Weibull distribution. The sample size is $n = 20$ and we consider the Type II censored configurations corresponding to $r = 5, 10, 15$, and 20 (complete data case). The number of simulations B_1 is set to be 100,000 so that the results are stable (i.e., negligible Monte Carlo error). Figure 2.1 shows the CP as a function of u'_U and r . For a desired coverage level, say $1 - \alpha = 0.95$ and a specific value of r , the value of u'_U is determined from the CP curve corresponding to the specified r value.

2.4.4.2 Illustration B: Two-sided SPI with Equal Tail Probability

Here the objective is to construct a two-sided SPI with equal tail probabilities from a previously sampled Weibull distribution. Again, B_1 is chosen to be 100,000. The contour plot of the CP as a function of u_L and u_U is shown in Figure 2.2(a). To obtain the prediction interval with equal probability in each tail, we solve the equations in (2.9). Figure 2.2(b) shows the

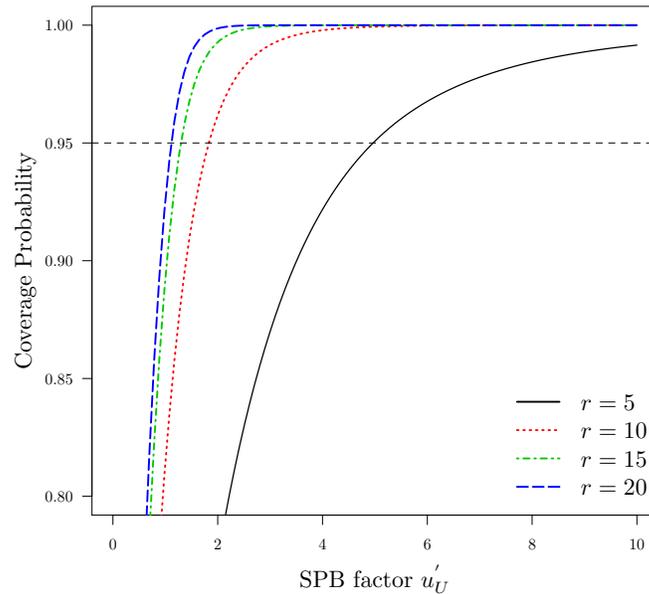
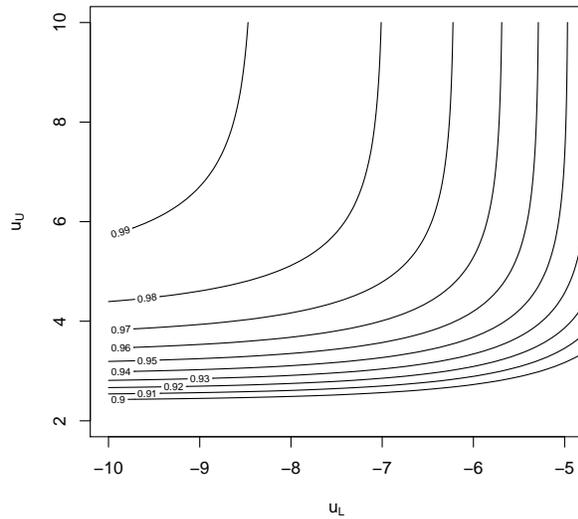


Figure 2.1: CP curves of one-sided upper SPBs for $n = 20$, $r = 5, 10, 15$ and 20 , $k = 4$, and $m = 5$ based on **Algorithm 1**.

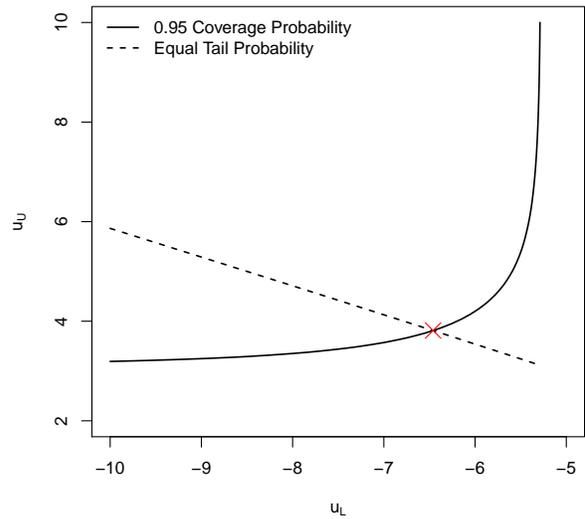
contour lines of the two equations. The upper and lower limits u_U and u_L of the 95% SPI with equal tail probabilities are the coordinates of the intersection point of the two non-linear curves in Figure 2.2(b). In Figure 2.2(b), the coordinates $(u_L, u_U) = (-6.46, 3.82)$ produce both the 0.95 overall coverage probability and the equal tail probabilities.

2.5 Simulation Study for Type I censoring

This section studies the CP properties of the simulation-based procedure proposed in Section 2.4.3. For the Type I censored data case, the procedure properties will depend on unknown parameters through the censoring time (or, more generally, the expected fraction failing). The CP of the SPIs/SPBs, however, will converge to the nominal confidence level as the expected number failing increases to infinity. Here we study the effect of the expected number of failures $r_f (= n \times p_f)$ on the CP of the SPIs/SPBs in small samples. Similar simulation designs can be found in Vander Weil and Meeker (1990) and Jeng and Meeker



(a) Coverage Probability Plot



(b) 0.95 Coverage and Equal Tail Probability Plot

Figure 2.2: (a) Contour plot of the CP as a function of u_L and u_U for $n = 20, r = 8, k = m = 3$, based on 100,000 simulations. (b) The contour lines of equation (2.9).

(2001). In **Algorithm 2**, we calculated the SPIs/SPBs based on the ML estimates $(\hat{\mu}, \hat{\sigma})$, which are determined from the observed data. To evaluate the performance of **Algorithm 2**, we simulate the data many times and average over the results. The detailed simulation plan is as follows.

1. Simulate $\mathbf{X} = (X_1, X_2, \dots, X_n)$ with the pre-determined censoring time. Without loss of generality, we simulate samples from the Weibull distribution with parameters $(\mu, \sigma) = (0, 1)$. Then, calculate the ML estimates of (μ, σ) for each simulated sample.
2. Use **Algorithm 2** to obtain the SPIs/SPBs. For example, we can obtain the one-sided upper SPB by computing u'_U .
3. Use (2.1), (2.4), and (2.6) to compute the conditional CP for the SPI, the lower SPB, and the upper SPB, respectively.
4. Repeat the steps 1 – 3 B_2 times and obtain the estimates of the unconditional CP for

the SPIs/SPBs by averaging over the conditional CPs.

Because the focus is on the CP for small sample sizes, we simulate datasets with the expected number of failures equal to $r_f = 5, 7, 10, 25$ and the expected fraction failing equal to $p_f = 0.25$. Here we chose $B_2 = 500$ for the purpose of controlling the computational cost while maintaining a reasonably small Monte Carlo error.

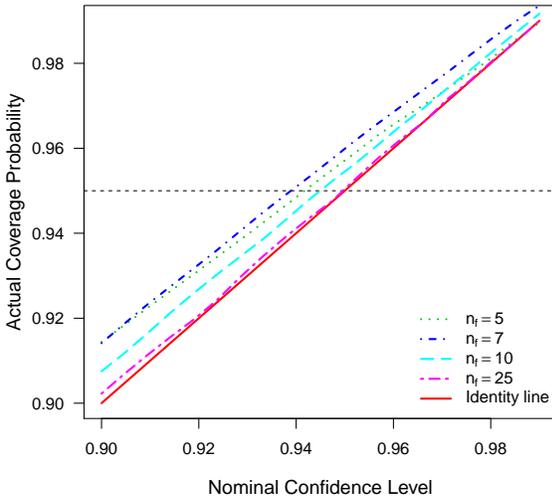
Figure 2.3 displays the estimated actual CP versus the nominal confidence level for the one-sided lower and upper SPBs, and the two-sided SPI. Figure 2.3 shows that there are some deviations from the nominal CP when the expected number of failures r_f is small (around 10). The estimated actual CP is close to the nominal confidence level when r_f is large enough (e.g., around 25). In the case of $r_f = 25$, the corresponding line is nearly the same as the identity line. When r_f is large, the observed data tends to have more failures, thus the estimates are more accurate and the SPIs/SPBs have better CP. We also note that the two-sided SPI tends to perform better than one-sided SPBs when r_f is small. As indicated earlier, we used $(\mu, \sigma) = (0, 1)$ in the simulation. For other values of (μ, σ) , the simulation results are similar because they depend on the expected number of failures. Overall, **Algorithm 2** provides satisfactory results for Type I censoring in finite samples when the expected number of failures is at least 5.

2.6 Applications

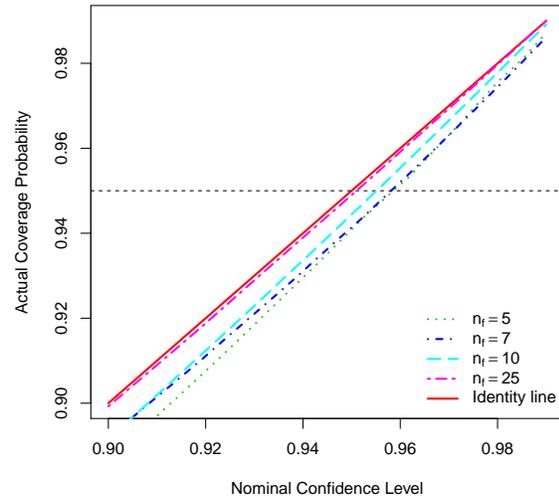
In this section, we use three examples to illustrate the applicability of the proposed procedure.

2.6.1 Nozzle Failure Time Data

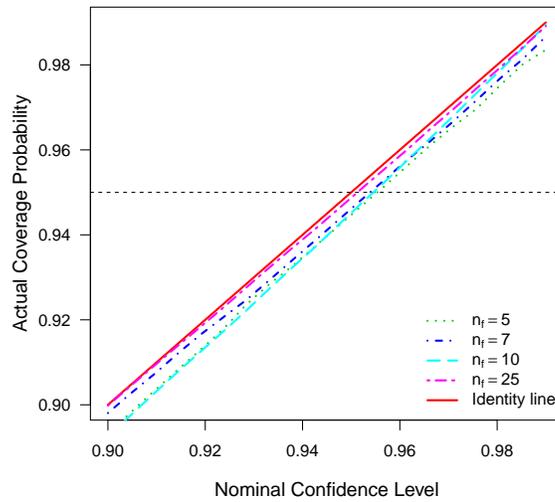
This example is adopted from the application described in Fertig and Mann (1977). They wanted to compute a 95% lower prediction bound (they called a “warranty period”) of the failure times of at least 36 or 40 out of 40 nozzles. They provided the sample mean and sample standard derivation of the logarithm of failure times (which they assumed to have



(a) Lower SPB



(b) Upper SPB



(c) Two-sided SPI

Figure 2.3: Estimated actual CP versus nominal confidence level for fixed $p_f = 0.25$, when $k = 4$ and $m = 5$. (a) Lower SPB. (b) Upper SPB. (c) Two-sided SPI.

normal distribution) of 10 nozzles, which are $\hat{\mu} = 3.850$ and $\hat{\sigma} = 0.034$, respectively. Applying **Algorithm 1**, we found that the lower SPBs for at least 36 and at least 40 out of 40 nozzles to be 43.35 and 40.96 hours (based on 100,000 Monte Carlo trials), respectively.

2.6.2 Aircraft Component Failure Time Data

Mann and Fertig (1973) describes a study yielding ten failure times out of 13 aircraft components that were tested. The failure times were 0.22, 0.50, 0.88, 1.00, 1.32, 1.33, 1.54, 1.76, 2.50, and 3.00 hours. The three right censored observations occurred at 3.00 hours. Both Mann and Fertig (1973) and Hsieh (1996) state that it is reasonable to assume a Weibull model for the data. The Weibull probability plot in Figure 2.4(a) corroborates the adequacy of the Weibull model. Based on Figure 2.4(b), the lognormal distribution, however, is also suitable to describe the failure-time distribution of the aircraft component. Using **Algorithm 1** one obtains 95% lower SPBs of the failure times of all 10 future aircraft components, which are 0.003 hours and 0.04 hours for the Weibull and lognormal distributions, respectively. Also we found that the 95% upper SPBs are 39.789 hours and 107.465 hours for the Weibull and lognormal distributions, respectively. The large difference is due to the implied extrapolation, especially into the upper tail of the failure-time distribution.

2.6.3 Vinyl Chloride Data

This application uses data consisting of vinyl chloride concentrations (in $\mu\text{g}/L$) from clean upgradient ground-water monitoring wells. The data were given in Bhaumik and Gibbons (2006). The probability plot in Bhaumik and Gibbons (2006) indicates that the gamma distribution fits the data well. Figures 2.5(a) and 2.5(b) indicate that the Weibull and lognormal distributions also provide good fit to the vinyl chloride data. Bhaumik and Gibbons (2006) wanted to obtain a 95% upper SPB to exceed at least $k = 1$ out of $m = 2$ future observations. For the gamma distribution, the 95% upper SPB is $2.931 \mu\text{g}/L$. Using **Algorithm 1**, for the Weibull distribution, the 95% upper SPB is $\exp(0.635 + 0.464 \times 0.99) = 2.989 \mu\text{g}/L$; for the

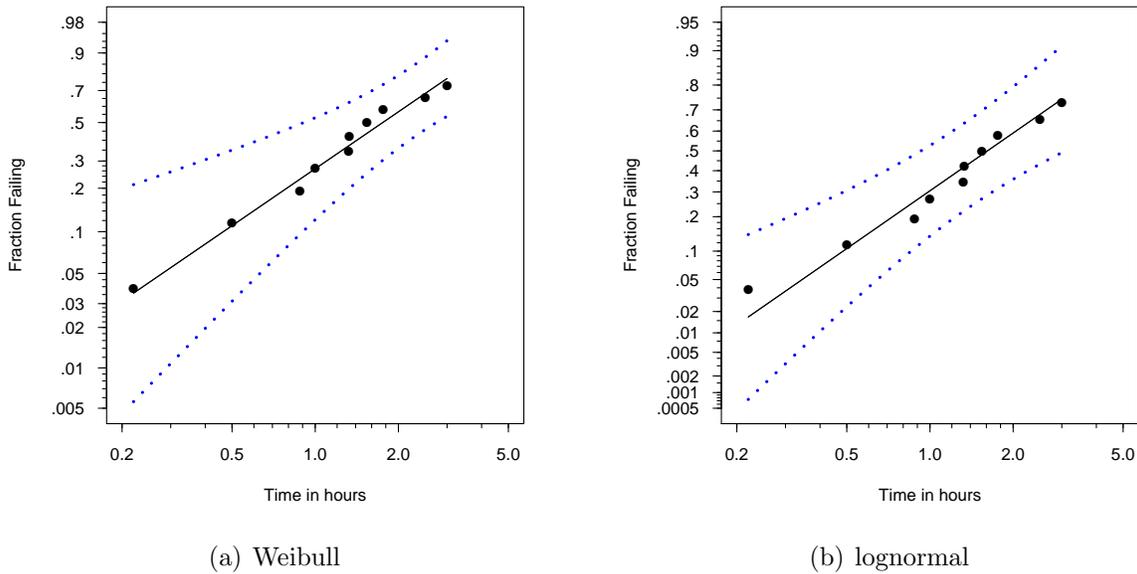


Figure 2.4: Probability plots for the aircraft data. (a) Weibull fit. (b) Lognormal fit.

lognormal distribution, the 95% upper SPB is $\exp(0.092 + 0.829 \times 1.120) = 2.773 \mu\text{g}/L$. For this application, the 95% upper SPBs for the gamma, Weibull, and lognormal distributions are closed to each other. This is because extrapolation is not required to construct this interval.

2.7 Concluding Remarks and Areas for Future Research

In this chapter, we propose a general method for constructing simultaneous two-sided prediction intervals for at least k out of m future observations as well as the corresponding one-sided bounds for the (log)-location-scale family of distributions. For the Type II censored or complete data cases, the method provides a procedure with CP equal to the nominal confidence level (ignoring Monte Carlo error that can be made arbitrarily small). For Type I censored data, the approximate procedure provides coverage probabilities that are close to the nominal confidence level if the expected number of failures is not too small.

The procedures in this chapter can also be extended to data involving multiple censor-

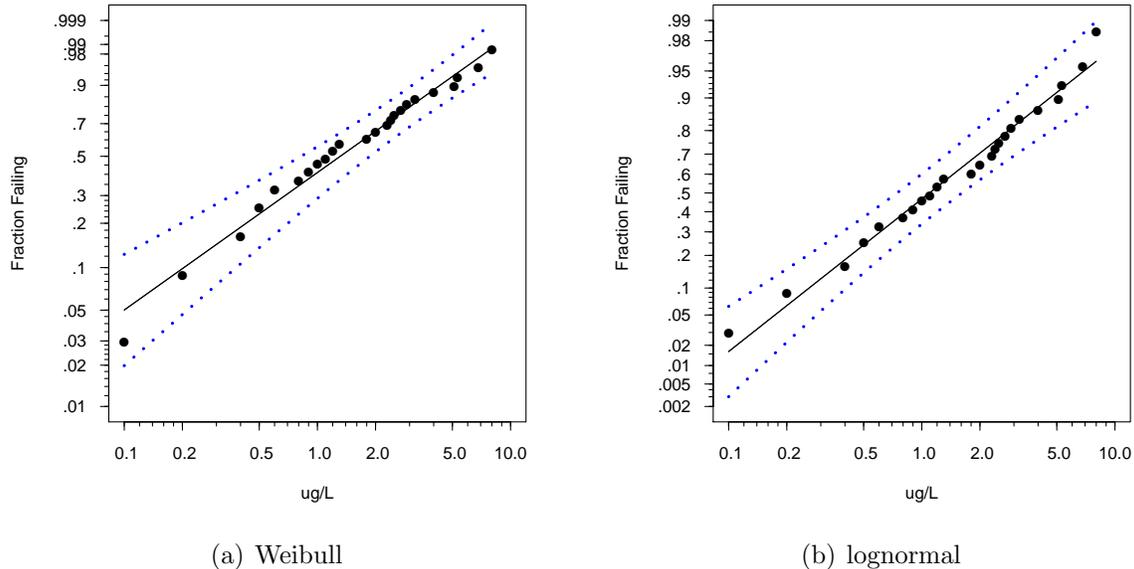


Figure 2.5: Probability plots for the vinyl chloride data. (a) Weibull fit. (b) Lognormal fit.

ing or random censoring. With complete data, the extension of the proposed methods to regression case is straightforward because the pivotal properties still hold (Lawless, 2003, Appendix E4). As long as the pivotal property holds, the proposed procedure can be easily extended to give exact prediction intervals. When the pivotal property no longer holds (e.g., with regression and censoring), the approximate pivotal approach can be applied.

Appendix

2.A Proof of Equation (2.2)

Let A_j be the event that exactly j of the future observations \mathbf{Y} lie in the prediction interval $[\hat{\mu} + u_L \hat{\sigma}, \hat{\mu} + u_U \hat{\sigma}]$. To compute $\Pr(A_j)$, first we compute the conditional probability $\Pr(A_j | \hat{\mu}, \hat{\sigma})$ and then average this conditional probability over the sampling distribution of the ML estimates $(\hat{\mu}, \hat{\sigma})$.

Now we proceed to compute $\Pr(A_j | \hat{\mu}, \hat{\sigma})$. Define the indicator variables

$$I_j = \begin{cases} 1 & \text{if } Y_j \in [\hat{\mu} + u_L \hat{\sigma}, \hat{\mu} + u_U \hat{\sigma}] \\ 0 & \text{otherwise,} \end{cases}$$

where $j = 1, \dots, m$.

The I_j variables are i.i.d. because the Y_j are i.i.d.. The I_j are Bernoulli(p) distributed where the p parameter is given in (2.10). Consequently, the number of future observations, say $S = \sum_{j=1}^m I_j$, contained by the conditional prediction interval $[\hat{\mu} + u_L \hat{\sigma}, \hat{\mu} + u_U \hat{\sigma}]$ is Binomial(m, p) distributed.

The parameter p is

$$\begin{aligned} p &= P(I_j = 1 | \hat{\mu}, \hat{\sigma}) = \Pr(\hat{\mu} + u_L \hat{\sigma} \leq Y_j \leq \hat{\mu} + u_U \hat{\sigma}) \\ &= \Pr(Y_j \leq \hat{\mu} + u_U \hat{\sigma}) - \Pr(Y_j \leq \hat{\mu} + u_L \hat{\sigma}) \\ &= \Phi\left(\frac{\hat{\mu} - \mu + u_U \hat{\sigma}}{\sigma}\right) - \Phi\left(\frac{\hat{\mu} - \mu + u_L \hat{\sigma}}{\sigma}\right) \\ &= \Phi\left(\frac{\hat{\mu} - \mu}{\sigma} + u_U \frac{\hat{\sigma}}{\sigma}\right) - \Phi\left(\frac{\hat{\mu} - \mu}{\sigma} + u_L \frac{\hat{\sigma}}{\sigma}\right) \\ &= \Phi(a) - \Phi(b) \end{aligned} \tag{2.10}$$

where $a = z_1 + u_U z_2$, $b = z_1 + u_L z_2$, with z_1 and z_2 being realizations of the pivotals $Z_1 = (\hat{\mu} - \mu)/\sigma$ and $Z_2 = \hat{\sigma}/\sigma$, respectively. The value of p is the same for all the variables I_j , $j = 1, \dots, m$, because its value does not depend on the variable Y_j chosen for the probability computation in (2.10).

Thus

$$\Pr(A_j | \hat{\mu}, \hat{\sigma}) = \Pr(S = j) = \binom{m}{j} p^j (1-p)^{m-j}$$

and the unconditional probability for A_j is

$$\Pr(A_j) = \int_0^\infty \int_{-\infty}^\infty \binom{m}{j} p^j (1-p)^{m-j} f_{(L,S)}(\hat{\mu}, \hat{\sigma}) d\hat{\mu} d\hat{\sigma}$$

where $f_{(L,S)}(\hat{\mu}, \hat{\sigma})$ is the sampling distribution of $(\hat{\mu}, \hat{\sigma})$.

Define M to be the number of future observations contained by the prediction interval $[\hat{\mu} + u_L \hat{\sigma}, \hat{\mu} + u_U \hat{\sigma}]$. Then the probability that the prediction interval contains at least k out of m future observations is

$$\begin{aligned} \Pr(M \geq k) &= \sum_{j=k}^m \Pr(A_j) \\ &= \int_0^\infty \int_{-\infty}^\infty \sum_{j=k}^m \binom{m}{j} [\Phi(a) - \Phi(b)]^j [1 - \Phi(a) + \Phi(b)]^{m-j} f_{(L,S)}(\hat{\mu}, \hat{\sigma}) d\hat{\mu} d\hat{\sigma} \\ &= \mathbf{E} \left[\sum_{j=k}^m \binom{m}{j} [\Phi(A) - \Phi(B)]^j [1 - \Phi(A) + \Phi(B)]^{m-j} \right]. \end{aligned}$$

Using (2.2), (u_L, u_U) can be chosen (selected/computed) to ensure that CP is equal to $(1 - \alpha)$.

2.B Approximate Pivotal Property for Type I Censored Data

The loglikelihood of the right censored data has the following form,

$$l(\mu, \sigma) = \sum_{i=1}^n \{ \delta_i \log f(x_i; \mu, \sigma) + (1 - \delta_i) \log [1 - F(x_i; \mu, \sigma)] \},$$

up to a constant.

Based on the loglikelihood function, the ML estimators $(\hat{\mu}, \hat{\sigma})$ satisfies the following score

equations:

$$\sum_{i=1}^n \left\{ -\delta_i \frac{\phi' \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)}{\phi \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)} + (1 - \delta_i) \frac{\phi \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)}{\left[1 - \Phi \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \right]} \right\} = 0, \quad (2.11)$$

$$\sum_{i=1}^n \left\{ -\delta_i \left[1 + \frac{\phi' \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)}{\phi \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)} \right] + (1 - \delta_i) \frac{\phi \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)}{1 - \Phi \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)} \right\} = 0. \quad (2.12)$$

For Type I censored data, x_c is fixed. Hence, $\delta_i = I(x_i \leq x_c) = I[(x_i - \mu)/\sigma \leq (x_c - \mu)/\sigma]$, where $I(\cdot)$ is an indicator function.

Notice that

$$\begin{aligned} \frac{x_i - \hat{\mu}}{\hat{\sigma}} &= \frac{x_i - \mu}{\sigma} \left(\frac{\hat{\sigma}}{\sigma} \right)^{-1} - \frac{\hat{\mu} - \mu}{\sigma} \left(\frac{\hat{\sigma}}{\sigma} \right)^{-1}, \\ \frac{x_c - \mu}{\sigma} &= \Phi^{-1}(p_f). \end{aligned}$$

Because the distribution of $(x_i - \mu)/\sigma$ is not related with (μ, σ) , therefore the left side of score equations are only the function of $Z_1 = (\hat{\mu} - \mu)/\sigma$, $Z_2 = \hat{\sigma}/\sigma$ and p_f . In **Algorithm 2**, we use \hat{p}_f to approximate p_f .

If we fix the censoring time x_c and let sample size n increases to infinity, \hat{p}_f will converge to p_f , Z_1 and Z_2 are asymptotic pivotal statistics, thus the CP of SPIs/SPBs computed by Algorithm 2 will approach the nominal confidence level.

2.C Variance of Estimated Coverage Probability

In this section, we want to calculate the variance of bootstrap estimated coverage probability $\text{CP}^*(u_L, u_U)$ (i.e., equation (2.8)). For each simulated sample $l = 1, \dots, B_1$, ML estimates $\hat{\boldsymbol{\theta}}^* = (\hat{\mu}_l^*, \hat{\sigma}_l^*)$ are independent. Moreover,

$$\hat{\boldsymbol{\theta}}^* \sim \text{N}[\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta})],$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix. In the simulation, we take $\boldsymbol{\theta} = (0, 1)'$. Denote the corresponding Fisher information matrix as I_o . A discussion on the calculation of Fisher information matrix with censored data can be found in Escobar and Meeker (1998). They showed that I_o is only the function of n, x_c , and $\Phi(\cdot)$ for type I censored data; I_o is the function of n, r , and $\Phi(\cdot)$ for type II censored data.

Denote $g(u_L, u_U) = \sum_{j=k}^m \binom{m}{j} p_l(u_L, u_U)^j [1 - p_l(u_L, u_U)]^{m-j}$, therefore we have

$$\text{Var}[\text{CP}^*(u_L, u_U)] = \frac{1}{B_1} \text{Var}[g(u_L, u_U)].$$

The first derivatives of $g(u_L, u_U)$ with respect to $(\widehat{\mu}_l^*, \widehat{\sigma}_l^*)$ are

$$\begin{aligned} \frac{\partial g}{\partial \widehat{\mu}_l^*} &= m \binom{m-1}{k-1} p_l^{k-1} (1-p_l)^{m-k} \frac{\partial p_l(u_L, u_U)}{\partial \widehat{\mu}_l^*}, \\ \frac{\partial g}{\partial \widehat{\sigma}_l^*} &= m \binom{m-1}{k-1} p_l^{k-1} (1-p_l)^{m-k} \frac{\partial p_l(u_L, u_U)}{\partial \widehat{\sigma}_l^*}, \end{aligned}$$

where

$$\begin{aligned} \frac{\partial p_l(u_L, u_U)}{\partial \widehat{\mu}_l^*} &= \phi(\widehat{\mu}_l^* + u_U \widehat{\sigma}_l^*) - \phi(\widehat{\mu}_l^* + u_L \widehat{\sigma}_l^*), \\ \frac{\partial p_l(u_L, u_U)}{\partial \widehat{\sigma}_l^*} &= \phi(\widehat{\mu}_l^* + u_U \widehat{\sigma}_l^*) u_U - \phi(\widehat{\mu}_l^* + u_L \widehat{\sigma}_l^*) u_L. \end{aligned}$$

According to Delta method, we have

$$\text{Var}[g(u_L, u_U)] = \begin{pmatrix} \frac{\partial g}{\partial \widehat{\mu}_l^*} \\ \frac{\partial g}{\partial \widehat{\sigma}_l^*} \end{pmatrix}_{(\widehat{\mu}_l^*, \widehat{\sigma}_l^*)=(0,1)}^T I_o^{-1} \begin{pmatrix} \frac{\partial g}{\partial \widehat{\mu}_l^*} \\ \frac{\partial g}{\partial \widehat{\sigma}_l^*} \end{pmatrix}_{(\widehat{\mu}_l^*, \widehat{\sigma}_l^*)=(0,1)}.$$

Therefore the variance of estimated coverage probability $\text{CP}^*(u_L, u_U)$ is a function of $(u_L, u_U, k, m, n, x_c, \Phi(\cdot), B_1)$ for type I censored data, and a function of $(u_L, u_U, k, m, n, r, \Phi(\cdot), B_1)$ for type II censored data.

Bibliography

- R. Beran. Calibrating prediction regions. *Journal of the American Statistical Association*, 85(411):715–723, 1990.
- D. K. Bhaumik. One-sided simultaneous prediction limits for left-censored normal random variables. *The Indian Journal of Statistics*, 70(2):248–266, 2008.
- D. K. Bhaumik and R. D. Gibbons. One-sided approximate prediction intervals for at least p of m observations from a Gamma population at each of r locations. *Technometrics*, 48(1):112–119, 2006.
- L. Danziger and S. A. Davis. Tables of distribution-free tolerance limits. *Annals of Mathematical Statistics*, 35(3):1361–1365, 1964.
- C. B. Davis and R. J. McNichols. One-sided intervals for at least p of m observations from a normal population on each of r future occasions. *Technometrics*, 29(3):359–370, 1987.
- L. A. Escobar and W. Q. Meeker. Fisher information matrices with censoring, truncation, and explanatory variables. *Statistica Sinica*, 8(1):221–237, 1998.
- L. A. Escobar and W. Q. Meeker. Statistical prediction based on censored life data. *Technometrics*, 41(2):113–124, 1999.
- K. W. Fertig and N. R. Mann. One-sided prediction intervals for at least p out of m future observations from a normal population. *Technometrics*, 19(2):167–177, 1977.
- G. J. Hahn. Factors for calculating two-sided prediction intervals for samples from a normal distribution. *Journal of the American Statistical Association*, 64(327):878–888, 1969.
- G. J. Hahn. Additional factors for calculating prediction intervals for samples from a normal distribution. *Journal of the American Statistical Association*, 65(332):1668–1676, 1970.

- H. K. Hsieh. Prediction intervals for Weibull observations, based on early-failure data. *IEEE Transactions on Reliability*, 45(4):666–670, 1996.
- S.-L. Jeng and W. Q. Meeker. Simultaneous parametric confidence bands for cumulative distributions from censored data. *Technometrics*, 43(4):450–461, 2001.
- K. Krishnamoorthy, Y. Lin, and Y. Xia. Confidence limits and prediction limits for a Weibull distribution based on the generalized variable approach. *Journal of Statistical Planning and Inference*, 139(8):2675–2684, 2009.
- J. F. Lawless. *Statistical models and methods for lifetime data, 2nd Edition*. John Wiley and Sons, New York, 2003.
- N. R. Mann and K. W. Fertig. Tables for obtaining Weibull confidence bounds and tolerance bounds based on best linear invariant estimates of parameters of the extreme-value distribution. *Technometrics*, 15(1):87–101, 1973.
- R. W. Mee and D. Kushary. Prediction limits for the Weibull distribution utilizing simulation. *Computational Statistics & Data Analysis*, 17(3):327–336, 1994.
- W. Q. Meeker and L. A. Escobar. *Statistical methods for reliability data*. John Wiley & Sons, 1998.
- R. E. Odeh. Two-sided prediction intervals to contain at least k out of m future observations from a normal distribution. *Technometrics*, 32(2):203–216, 1990.
- S. A. Vander Weil and W. Q. Meeker. Accuracy of approx confidence bounds using censored Weibull regression data from accelerated life tests. *IEEE Transactions on Reliability*, 39(3):346–351, 1990.

Chapter 3 A Semi-parametric Model for Accelerated Destructive Degradation Test Data Analysis

Abstract

Accelerated destructive degradation tests (ADDT) are widely used in industry to evaluate materials' long term properties. Even though there has been tremendous statistical research in nonparametric methods, the current industrial practice is still to use application-specific parametric models to describe ADDT data. The challenge of using a nonparametric approach comes from the need to retain the physical meaning of degradation mechanisms and also perform extrapolation for predictions at the use condition. Motivated by this challenge, we propose a semi-parametric model to describe ADDT data. We use monotonic B-splines to model the degradation path, which not only provides flexible models with few assumptions, but also retains the physical meaning of degradation mechanisms (e.g., the degradation path is monotonically decreasing). Parametric models, such as the Arrhenius model, are used for modeling the relationship between the degradation and accelerating variable, allowing for extrapolation to the use conditions. We develop an efficient procedure to estimate model parameters. We also use simulation to validate the developed procedures and demonstrate the robustness of the semi-parametric model under model misspecification. Finally, the proposed method is illustrated by multiple industrial applications.

Key Words: Acceleration model; ADDT; Arrhenius model; Degradation model; Long-term property evaluation; Polymeric materials.

3.1 Introduction

3.1.1 Motivation

It is important for manufacturers to understand the lifetime of their products in order to ensure accurate marketing and determine areas for improvement. While lifetime testing is the most common approach, for many materials it is more informative to observe the degradation of some performance characteristic, such as the tensile strength of an adhesive bond, over time. The lifetime is determined by a “soft failure” when the characteristic drops below a predetermined level. This form of testing is known as degradation testing.

Several varieties of degradation testing have been developed to accommodate unique circumstances. Due to the long service life of many new materials, degradation testing under normal use conditions is often not feasible. By exposing the material to a more harsh environment, such as higher levels of temperature or humidity compared to the use conditions, degradation data can be collected more efficiently. Thus, an accelerating variable is often used in degradation tests. In some applications, measurements of the degradation level are destructive. That is, the units being tested are destroyed or the physical characteristics changed in a significant manner. An example could be determining the strength of a material by measuring the force needed to break it. This form of testing, combined with an accelerating variable, is referred to as accelerated destructive degradation testing (ADDT). Because of the nature of the testing, ADDT must be analyzed differently from other common forms of degradation testing, such as repeated-measures degradation testing (RMDT), in which multiple measurements can be taken from the same unit.

Current procedures for analyzing ADDT data involve an assumed parametric model for the degradation path over time and a parametric form for the accelerating-variable effect. The predominance of parametric models is mainly due to the need for extrapolation in two aspects; extrapolation in time and extrapolation to the use conditions. For example, an ADDT may cover only 100-70% of the original material’s strength and be performed at an

elevated temperature range ($60-80^{\circ}\text{C}$), but interest lies at strengths 50-70% of the original at a temperature of 30°C . These parametric models tend to be material-specific and at present there seems to be no general model that can be applied to a wide variety of materials.

Even though there has been tremendous statistical research in nonparametric methods, the current industrial practice is still to use application-specific parametric models to describe ADDT data. Motivated by multiple industrial applications, we aim to bridge this gap between the statistical research and current industrial practice. Instead of a case-by-case parametric modeling approach, we propose a general and flexible semi-parametric model to describe ADDT data. The challenge of using a nonparametric approach comes from the need to retain the physical meaning of degradation mechanisms and performing extrapolations for predictions at the use condition. To overcome those challenges, the semi-parametric model consists of a nonparametric model for the degradation path and a parametric form for the accelerating-variable effect. In order to preserve the monotonic nature of many degradation paths, the nonparametric model portion will be constructed based on monotonic spline methods. For the parametric model portion, commonly used models, such as the Arrhenius relationship for temperature, will be used for extrapolation. Parameter estimation and inference procedures will also be developed.

3.1.2 Related Literature

The literature on accelerated degradation data modeling and analysis can be divided into two areas: RMDT and ADDT. In the pioneering work, Lu and Meeker (1993) used RMDT data to estimate failure-time distribution via the framework on mixed-effects models. Meeker et al. (1998) introduced nonlinear mixed-effects models for RMDT data, which were derived from physical-failure mechanisms. Introductory level description of degradation models can be found in Gorjian et al. (2010), and Meeker et al. (2011). Ye and Xie (2015) provided a comprehensive review of the state-of-art methods in modeling RMDT data.

In the area of ADDT data modeling and analysis, Nelson (1990, Chapter 11) used ADDT

data from an insulation to estimate performance degradation. Escobar et al. (2003) provided a parametric model and method to analyze the ADDT data collected from an adhesive bond. Tsai et al. (2013) considered the problem of designing an ADDT with a nonlinear model motivated by a polymer dataset. Li and Doganaksoy (2014) used a parametric model to model ADDT data collected from a temperature accelerated test to study the degradation of seal strength. In all existing methods for analyzing ADDT data, the parametric method is the most popular.

Compared to parametric models of degradation data, spline functions tend to be more flexible and require less assumptions regarding the model formulation. Because the degradation path is often monotonic in nature, monotone splines are suitable for modeling degradation paths. Ramsay (1988) suggested using a basis of I-splines (integrated splines) for semi-parametric modeling. He and Shi (1998) considered the use of B-splines with L_1 optimization. Meyer (2008) extended the work in Ramsay (1988) by proposing cubic monotone splines. Leitenstorfer and Tutz (2007) considered the use of monotone B-splines in generalized additive models. For other applications of monotone B-splines, one can refer to Kanungo et al. (1995) and Fengler and Hin (2015). In addition, Eilers and Marx (1996) proposed a flexible class of P-splines. Bollaerts et al. (2006), Hofner et al. (2011), and Hofner et al. (2016) considered the estimation of monotonic effects with P-splines.

Related to RMDT models, Ye et al. (2014) considered semi-parametric estimation of Gamma processes. Hong et al. (2015), and Xu et al. (2015) used shape-restricted splines to model the effects of time-varying covariates on the degradation process. There is little literature, however, on the use of semi-parametric models in ADDT data modeling and analysis.

3.1.3 Overview

The rest of this chapter is organized as follows. Section 3.2 introduces some general notation for ADDT data. It also presents in detail the construction of the semi-parametric model using

monotonic B-splines. In Section 3.3, we present a procedure for estimating the unknown parameters as well as procedures for conducting inference on ADDT data based on this model. We conduct simulation studies in Section 3.4 to investigate the performance of the semi-parametric method with special consideration of model misspecification. In Section 3.5, we apply the model to data from several published datasets and provide comparisons with other well-known parametric models. Finally, Section 3.6 contains conclusions and areas for future research.

3.2 The Semi-parametric Model

3.2.1 General Setting

Let y_{ijk} be the degradation measurement for the k th sample at level i of the accelerating variable \mathcal{AF}_i and the j th observation time point t_{ij} , $i = 1, \dots, I$, $j = 1, \dots, J_i$, and $k = 1, \dots, n_{ij}$, where n_{ij} denotes the sample size at t_{ij} . Let $n = \sum_{i=1}^I \sum_{j=1}^{J_i} n_{ij}$ be the total number of observations. A general form of the degradation model is

$$y_{ijk} = \mathcal{D}(t_{ij}, x_i; \boldsymbol{\theta}) + \varepsilon_{ijk}, \quad (3.1)$$

where $x_i = h(\mathcal{AF}_i)$ is a function of the accelerating variable, $\boldsymbol{\theta}$ is a vector of unknown parameters in the degradation path, and ε_{ijk} is an error term that describes unit-to-unit variability. For the purposes of illustration, we will assume that the degradation path is monotone decreasing with time. The model can easily be generalized to paths that are increasing with time. We will also be considering temperature as the accelerating factor as it is the most common form of acceleration encountered in ADDT. However, the model can easily incorporate other types of acceleration, such as voltage.

For temperature-accelerated processes, the Arrhenius model is often used to describe the relationship between degradation and temperature. This model uses a transformed

temperature level given as

$$x_i = \frac{-11605}{\text{Temp}_i + 273.16}. \quad (3.2)$$

Here, Temp_i is in degrees Celsius, and 11605 is the reciprocal of the Boltzmann's constant (in units of eV). The value 273.16 in the denominator is used to convert to the Kelvin temperature scale.

3.2.2 The Scale Acceleration Model

We propose the following semi-parametric functional forms for the degradation model in (3.1).

$$\mathcal{D}(t_{ij}, x_i; \boldsymbol{\theta}) = g[\eta_i(t_{ij}; \beta); \boldsymbol{\gamma}], \quad (3.3)$$

$$\eta_i(t; \beta) = \frac{t}{\exp(\beta s_i)}, \quad s_i = x_{\max} - x_i, \quad (3.4)$$

$$\varepsilon_{ijk} \sim N(0, \sigma^2), \quad \text{and} \quad \text{Corr}(\varepsilon_{ijk}, \varepsilon_{ijk'}) = \rho, \quad k \neq k'. \quad (3.5)$$

Here, $g(\cdot)$ is a monotone decreasing function with unknown parameter vector $\boldsymbol{\gamma}$, β is an unknown parameter associated with the accelerating variable, and $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \beta, \sigma, \rho)'$ is the vector containing all of the unknown parameters. The quantity $x_{\max} = -11605/[\max_i(\text{Temp}_i) + 273.16]$ is defined to be the transformed value of the highest level of the accelerating variable.

The model in (3.3) falls within the class of scale acceleration models. When the acceleration level is at its highest, $s_{\max} = x_{\max} - x_{\max} = 0$. In this case, $\eta_i(t; \beta) = t$ implies that the degradation path no longer relies on β , and

$$\mathcal{D}(t, x_{\max}; \boldsymbol{\theta}) = g(t; \boldsymbol{\gamma}).$$

Thus, the function $g(\cdot)$ can be interpreted as the baseline degradation path for the scale acceleration model in (3.3). For other specific stress level i , $\mathcal{D}(t, x_i; \boldsymbol{\theta})$ is a decreasing function

of time t , in which β controls the degradation rate through time-scale factor $\exp(\beta s_i)$ in (3.4). A small time-scale factor corresponds to a rapid decrease in degradation. By assuming the function form of $\eta_i(t; \beta)$ in (3.4), we assume that the time to reach certain degradation level at stress level x_i is the time needed for the baseline degradation path multiply by the factor $\exp(\beta s_i)$. The distribution of error terms ε_{ijk} is specified in (3.5) with parameters σ and ρ . In particular, we consider a compound symmetric correlation structure for measurements taken on the same temperature and time point. Measurements at different temperatures and times are assumed to be independent. There exists situation where the samples that are tested on the same temperature and time point are from the same batch. There may exists large with-in-batch correlation, which is modeled by ρ . If there is no with-in-batch correlation, then the covariance structure can be simplified. Hence, the correlation specified in (3.5) is a more general case.

Let y_M be the lowest degradation level present in the observed data. Then the scale-acceleration model and the monotonicity of $g(\cdot)$ will allow one to extrapolate the degradation level to y_M for any given acceleration level. Let \mathcal{D}_f be the failure threshold. Then, if $y_M < \mathcal{D}_f$, one can use the semi-parametric model to obtain failure information at the use conditions through this extrapolation. This is particularly useful since, in general, measurements may be available below \mathcal{D}_f for only some of the highest levels of the accelerating variable. In fact, some industrial standards require that tests be run until the degradation level drops below \mathcal{D}_f for several acceleration levels. However, extrapolation beyond y_M is not possible due to the nonparametric construction of the $g(\cdot)$, which is the tradeoff for this kind of model flexibility.

3.2.3 Nonparametric Form for Baseline Degradation Path

We use nonparametric methods to estimate the baseline degradation path $g(\cdot)$. Specifically, we use monotonic B-splines to model the baseline degradation path. This not only provides flexible models, but also retains the physical meaning of degradation mechanisms (e.g., the

degradation path is monotonically decreasing).

Consider a set of interior knots $d_1 \leq \dots \leq d_N$, and two boundary points d_0 and d_{N+1} . The entire set of ordered knots are

$$d_{-q} = \dots = d_0 \leq d_1 \leq \dots \leq d_N \leq d_{N+1} = \dots = d_{N+q+1},$$

where the lower and upper boundary points are appended q times and q is the polynomial degree. For notational simplicity, we rewrite the subscripts in the ordered knot sequences as d_1, \dots, d_{N+2q+2} . The total number of basis functions is $p = N + q + 1$. The l th B-spline basis function of degree q evaluated at z can be recursively obtained in the following formulas:

$$B_{0,l}(z) = \mathbf{1}(d_l \leq z < d_{l+1}),$$

$$B_{q,l}(z) = \frac{z - d_l}{d_{l+q} - d_l} B_{q-1,l}(z) + \frac{d_{l+q+1} - z}{d_{l+q+1} - d_{l+1}} B_{q-1,l+1}(z),$$

where $l = 1, \dots, p$, and $\mathbf{1}(\cdot)$ is an indicator function. The degradation model can then be expressed as

$$y_{ijk} = \sum_{l=1}^p \gamma_l B_{q,l}[\eta_i(t_{ij}; \beta)] + \varepsilon_{ijk}, \quad (3.6)$$

where γ_l 's are the coefficients.

To ensure the degradation path is monotone decreasing, we require the first derivative of $\mathcal{D}(\tau_{ij}, x_i; \boldsymbol{\theta})$ be negative. For B-spline basis functions, De Boor (2001) proved that the derivative of $\mathcal{D}(t, x_i; \boldsymbol{\theta})$ with respect to $\eta_i(t; \beta)$ is

$$\frac{d\mathcal{D}(t, x_i; \boldsymbol{\theta})}{d\eta_i(t; \beta)} = \sum_{l=2}^p (q-1) \frac{(\gamma_l - \gamma_{l-1})}{d_{l+q+1} - d_l} B_{q-1,l}[\eta_i(t; \beta)].$$

As B-spline basis functions are nonnegative, it follows that $\gamma_l \leq \gamma_{l-1}$ for all $2 \leq l \leq p$ gives a sufficient condition for a monotone decreasing degradation path. However, except for basis

functions with degree $q = 1, 2$, it is not a necessary condition. Fritsch and Carlson (1980) derived the necessary conditions for cubic splines ($q = 3$), though for higher order splines necessary conditions are as yet unclear.

3.3 Estimation and Inference

3.3.1 Parameter Estimation

Let $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijn_{ij}})'$, $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \dots, \varepsilon_{ijn_{ij}})'$, $\mathbf{y} = (y'_{11}, \dots, y'_{1J_1}, \dots, y'_{I1}, \dots, y'_{IJ_I})'$, $\boldsymbol{\varepsilon} = (\varepsilon'_{11}, \dots, \varepsilon'_{1J_1}, \dots, \varepsilon'_{I1}, \dots, \varepsilon'_{IJ_I})'$ and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)'$. The degradation model in (3.6) can be written as

$$\mathbf{y} = \mathbf{X}_\beta \boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (3.7)$$

where

$$\mathbf{X}_\beta = \begin{bmatrix} B_{q,1}[\eta_1(t_{11}; \beta)] & \cdots & B_{q,p}[\eta_1(t_{11}; \beta)] \\ B_{q,1}[\eta_1(t_{12}; \beta)] & \cdots & B_{q,p}[\eta_1(t_{12}; \beta)] \\ \vdots & \ddots & \vdots \\ B_{q,1}[\eta_I(t_{IJ_I}; \beta)] & \cdots & B_{q,p}[\eta_I(t_{IJ_I}; \beta)] \end{bmatrix},$$

and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Here, $\boldsymbol{\Sigma} = \text{Diag}(\boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{1J_1}, \dots, \boldsymbol{\Sigma}_{I1}, \dots, \boldsymbol{\Sigma}_{IJ_I})$ and $\boldsymbol{\Sigma}_{ij} = \sigma^2[(1 - \rho)I_{n_{ij}} + \rho P_{n_{ij}}]$, where $I_{n_{ij}}$ is an $n_{ij} \times n_{ij}$ identity matrix and $P_{n_{ij}}$ is an $n_{ij} \times n_{ij}$ matrix of 1's. We can also rewrite $\boldsymbol{\Sigma} = \sigma^2 \mathbf{R}$, where $\mathbf{R} = \text{Diag}(\mathbf{R}_{11}, \dots, \mathbf{R}_{1J_1}, \dots, \mathbf{R}_{I1}, \dots, \mathbf{R}_{IJ_I})$ and $R_{ij} = (1 - \rho)I_{n_{ij}} + \rho P_{n_{ij}}$.

We use likelihood-based methods to estimate the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \beta, \sigma, \rho)'$. For now, we consider estimation of $\boldsymbol{\theta}$ with a given number of knots and knot locations. We will give a discussion on knot selection in Section 3.3.3. A particular challenge to the estimation comes from the constraints on $\boldsymbol{\gamma}$, namely that $\gamma_l \leq \gamma_{l-1}$, $2 \leq l \leq p$. We also note that, for a given β , \mathbf{X}_β is known, in which case (3.7) becomes a linear model with a correlated covariance structure. Thus, we proceed by first deriving estimates of $(\boldsymbol{\gamma}', \sigma, \rho)'$ given β and then use a profile likelihood approach to estimate β .

The estimates of $\boldsymbol{\gamma}$ and $(\sigma, \rho)'$ are obtained using an iterative procedure. In particular, at the m th iteration, given estimates $(\hat{\boldsymbol{\sigma}}^{(m-1)}, \hat{\boldsymbol{\rho}}^{(m-1)})'$, the value of $\hat{\boldsymbol{\gamma}}^{(m)}$ is obtained by minimizing

$$Q(\boldsymbol{\gamma}) = (\mathbf{y} - \mathbf{X}_\beta \boldsymbol{\gamma})' \left(\hat{\boldsymbol{\Sigma}}^{(m-1)} \right)^{-1} (\mathbf{y} - \mathbf{X}_\beta \boldsymbol{\gamma})$$

subject to $\gamma_l \leq \gamma_{l-1}, 2 \leq l \leq p.$ (3.8)

Equation (3.8) is a quadratic object function with linear constraints and so can be solved with quadratic programming techniques. Given $\hat{\boldsymbol{\gamma}}^{(m)}$, one can then obtain $(\hat{\boldsymbol{\sigma}}^{(m)}, \hat{\boldsymbol{\rho}}^{(m)})'$ using restricted maximum likelihood (REML) so long as $\hat{\boldsymbol{\gamma}}^{(m)}$ does not take values on the boundary of the linear constraints. If the solution of equation (3.8) does take values on the boundary of the linear constraints, we can still consider approximate REML to obtain these estimates. Let $\hat{\boldsymbol{\gamma}}_u^{(m)}$ represent all of the unique values in $\hat{\boldsymbol{\gamma}}^{(m)}$ and p_u be the length of $\hat{\boldsymbol{\gamma}}_u^{(m)}$. For each unique value $\hat{\gamma}_{i,u}^{(m)}$, let $\mathbf{x}_{i,\beta u}$ be the sum of the corresponding columns in \mathbf{X}_β . Then we have $\mathbf{X}_\beta \hat{\boldsymbol{\gamma}}^{(m)} = \mathbf{X}_{\beta u} \hat{\boldsymbol{\gamma}}_u^{(m)}$, where $\mathbf{X}_{\beta u} = (\mathbf{x}_{1,\beta u}, \dots, \mathbf{x}_{p_u,\beta u})$. The approximate REML log-likelihood is then

$$\mathcal{L}_{\text{REML}}(\sigma, \rho | \hat{\boldsymbol{\gamma}}^{(m)}) = -\frac{1}{2} \left\{ \log |\boldsymbol{\Sigma}| + \log |\mathbf{X}'_{\beta u} \boldsymbol{\Sigma}^{-1} \mathbf{X}_{\beta u}| + (\mathbf{y} - \mathbf{X}_\beta \hat{\boldsymbol{\gamma}}^{(m)})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}_\beta \hat{\boldsymbol{\gamma}}^{(m)}) \right\}. \quad (3.9)$$

The covariance parameter estimates $(\hat{\boldsymbol{\sigma}}^{(m)}, \hat{\boldsymbol{\rho}}^{(m)})'$ are those values that maximize equation (3.9). In particular, after some calculation it can be shown that $\hat{\boldsymbol{\sigma}}^{(m)}$ has the following closed-form expression

$$\hat{\boldsymbol{\sigma}}^{(m)} = \left[\frac{(\mathbf{y} - \mathbf{X}_\beta \hat{\boldsymbol{\gamma}}^{(m)})' (\hat{\mathbf{R}}^{(m-1)})^{-1} (\mathbf{y} - \mathbf{X}_\beta \hat{\boldsymbol{\gamma}}^{(m)})}{n - p_u} \right]^{\frac{1}{2}}.$$

Thus, $\hat{\rho}^{(m)}$ can be obtained from a one dimensional optimization problem. That is,

$$\hat{\rho}^{(m)} = \underset{\rho}{\operatorname{argmax}} \left\{ -\log |(\hat{\sigma}^{(m)})^2 \mathbf{R}| - \log |(\hat{\sigma}^{(m)})^{-2} \mathbf{X}'_{\beta u} \mathbf{R}^{-1} \mathbf{X}_{\beta u}| \right. \\ \left. - (\hat{\sigma}^{(m)})^{-2} (\mathbf{y} - \mathbf{X}_{\beta} \hat{\gamma}^{(m)})' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}_{\beta} \hat{\gamma}^{(m)}) \right\}.$$

Upon convergence, the estimates of $(\hat{\gamma}', \hat{\sigma}, \hat{\rho})'$ are obtained for a given β , denoted by $(\hat{\gamma}'_{\beta}, \hat{\sigma}_{\beta}, \hat{\rho}_{\beta})'$.

The initial values $(\hat{\sigma}^{(0)}, \hat{\rho}^{(0)})'$ can be easily obtained by fitting a non-constrained model.

The profile log-likelihood for β is given as

$$\mathcal{L}(\beta, \hat{\gamma}_{\beta}, \hat{\sigma}_{\beta}, \hat{\rho}_{\beta}) = \log \left\{ \frac{1}{\sqrt{2\pi} |\hat{\Sigma}_{\beta}|^{1/2}} \exp \left[-\frac{(\mathbf{y} - \mathbf{X}_{\beta} \hat{\gamma}_{\beta}) \hat{\Sigma}_{\beta}^{-1} (\mathbf{y} - \mathbf{X}_{\beta} \hat{\gamma}_{\beta})}{2} \right] \right\}.$$

In practice, one can first estimate $(\gamma', \sigma, \rho)'$ for a specified range of values of β , then compute $\mathcal{L}(\beta, \hat{\gamma}_{\beta}, \hat{\sigma}_{\beta}, \hat{\rho}_{\beta})$ as a function of β . The estimate $\hat{\beta}$ is the value that maximizes this function. The final estimates are denoted by $\hat{\theta} = (\hat{\gamma}', \hat{\beta}, \hat{\sigma}, \hat{\rho})'$. To specify the initial range of β , we first fit a polynomial line for each temperature level. By comparing the time reaches to certain degradation level (for example, failure threshold) at different temperature levels, we can get several estimates of β . The range of the estimates $(\hat{\beta}_{min}^{(0)}, \hat{\beta}_{max}^{(0)})$ might be expected to contain β . Sometimes, multiplying a factor ϕ ($\phi = 1, 2, 3, \dots$), that is considering the interval $(\phi^{-1} \hat{\beta}_{min}^{(0)}, \phi \hat{\beta}_{max}^{(0)})$ is necessary to find the estimation of β .

3.3.2 Reliability Measures

Once the model parameters have been estimated, other parameters related to reliability can then be estimated. For example, the mean time to failure (MTTF), denoted by m_f , is one of many ways to evaluate the reliability of a product/material. Based on the semi-parametric model, we can derive an estimate \hat{m}_f at a use condition x_f and failure threshold \mathcal{D}_f by

solving

$$\sum_{l=1}^p \hat{\gamma}_l B_{q,l} \left(\frac{\hat{m}_f}{\exp[\hat{\beta}(x_{\max} - x_f)]} \right) = \mathcal{D}_f.$$

We can also derive the failure time distribution from the semi-parametric model. The event of failure time T is less than t is equivalent to having the degradation measurement at time t is less than the failure threshold D_f , that is the cumulative distribution function of failure time can be calculated as

$$F_T(t) = P(T < t) = P(y_t < D_f) = \Phi \left(\frac{D_f - g \left[\frac{t}{\exp(\beta s)}; \gamma \right]}{\sigma} \right), t \geq 0.$$

Hence, the quantile function can also be calculated. The α quantile is $t_\alpha = F_T^{-1}(\alpha)$. In the case of no available closed-form expression, numerical result is searched.

3.3.3 Spline Knots Selection

The number of knots and knot locations are a key component to using B-splines to model the degradation path. In addition, it is also necessary to determine the maximum degree of the B-splines. For knot selection, we first fix the degree of the B-splines and then find the optimum knot locations. Optimality is determined by a variation of the Akaike information criterion:

$$\text{AIC} = -2 \log \left\{ \frac{1}{\sqrt{2\pi} |\hat{\Sigma}|^{1/2}} \exp \left[-\frac{(\mathbf{y} - \mathbf{X}_{\hat{\beta}} \hat{\gamma}) \hat{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}_{\hat{\beta}} \hat{\gamma})}{2} \right] \right\} + 2 \times edf, \quad (3.10)$$

where edf is the effective degrees of freedom in γ plus three for the parameters $(\beta, \sigma, \rho)'$. Wang et al. (2013) and Meyer (2012) discussed constrained spline regression for both independent and correlated error cases. In particular, they showed how to calculate the effective degrees of freedom for a constrained fit through the use of a cone projection, which is the trace of the projection matrix. Because we have $p-1$ linear constraints, the effective degrees

of freedom in γ has a value from 1 to p , where p corresponds to a unconstrained fit. Letting q denote the degree of the B-spline functions, the procedure for knot selection is as follows:

1. Determine the number of interior knots $N_{\text{opt},q}$ which minimizes the AIC. The default knot locations are equally-spaced sample quantiles. That is, if number of interior knots is N , the default knot locations are $b/N, b = 1, \dots, N - 1$.
2. Delete each of the internal knots in sequence. The knot whose deletion leads to the greatest reduction in AIC is removed. Repeat until no more existing knots can be removed.

The whole procedure is to be repeated for different B-spline degrees to determine the knot sequence. This knot selection procedure is similar to the procedure in He and Shi (1998). The sample size for an ADDT is typically small and so a low degree of spline ($q \leq 4$) and a small number of interior knots ($1 \leq N \leq 5$) are usually sufficient to provide a good fit to the data.

3.3.4 Statistical Inference

Inference based on the semi-parametric model in (3.7) can rely on either asymptotic theory or a bootstrap procedure. Because the bootstrap method is straightforward and easy to implement, we use a nonparametric bootstrap to calculate confidence intervals (CI) for the parameters and pointwise CI for the degradation path. The error term in model (3.7) can be written as

$$\varepsilon_{ijk} = u_{ij} + e_{ijk},$$

where $u_{ij} \sim N(0, \sigma_u^2)$, $e_{ijk} \sim N(0, \sigma_e^2)$, $\text{Corr}(u_{ij}, e_{ijk}) = 0$, $\sigma_u^2 = \rho\sigma^2$, and $\sigma_e^2 = (1 - \rho)\sigma^2$. That is, the error term in model (3.7) can be written as the sum of a random effect term u_{ij} and an independent error term e_{ijk} . To obtain the CI, one could resample from the estimated random effect term \hat{u}_{ij} and the estimated independent error term \hat{e}_{ijk} separately. However,

Carpenter et al. (2003) showed that directly resampling from \hat{u}_{ij} and \hat{e}_{ijk} will cause bias. Therefore, we adjust \hat{u}_{ij} and \hat{e}_{ijk} prior to bootstrapping. That is,

$$\hat{u}_{ij}^c = \left[\sum_{ij} \hat{u}_{ij}^2 / (nJ_n) \right]^{-1/2} \hat{\sigma}_u \hat{u}_{ij}, \quad \text{and} \quad \hat{e}_{ijk}^c = \left[\sum_k \hat{e}_{ijk}^2 / (n_{ij}) \right]^{-1/2} \hat{\sigma}_e \hat{e}_{ijk}.$$

The specific steps of nonparametric bootstrap are described as follows:

For $m = 1, \dots, B$,

1. Sample $u_{ij}^{(m)c}$ with replacement from \hat{u}_{ij}^c and sample $e_{ijk}^{(m)c}$ with replacement from \hat{e}_{ijk}^c .
2. Compute $y_{ijk}^{(m)} = x_{ij}' \hat{\gamma} + u_{ij}^{(m)c} + e_{ijk}^{(m)c}$.
3. Fit the semi-parametric model to the bootstrapped sample $y_{ijk}^{(m)}$.

The CI with confidence level $1 - \alpha$ for a parameter of interest, θ , is calculated by taking the lower and upper $\alpha/2$ quantiles of the bootstrap estimates. If the bootstrap sampling distribution is not symmetric, we can use the bias-corrected CI. For a sequence of bootstrap estimates $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(B)}$, a bias-corrected CI, proposed by Efron and Tibshirani (1994), can be computed by taking the $B\Phi(2z_q + z_{\alpha/2})$ and $B\Phi(2z_q + z_{1-\alpha/2})$ ordered values, where q denotes the proportion of bootstrap values less than $\hat{\theta}$, $\Phi(\cdot)$ is the cumulative distribution function and $z_{(\cdot)}$ is the quantile function of the standard normal distribution.

3.4 Simulation Study

The objective of the simulation study is to investigate the performance of the proposed parameter estimation and inference procedures. We will examine the bias, standard derivation (SD), and mean square error (MSE) of the parameter estimators and the estimated baseline degradation function. We also will investigate the coverage probability (CP) of the bootstrap-based CI procedure in Section 3.3.4. An additional simulation study will be conducted to investigate the performance of our semi-parametric model under model misspecification.

Table 3.1: Selected temperature levels and time points for the simulation studies.

Settings	Number of Temp. Levels (n)	Temperature Levels ($^{\circ}\text{C}$)
Temperature setting 1	3	50, 65, 80
Temperature setting 2	6	30, 40, 50, 60, 70, 80
	Number of Time Points (J_n)	Measuring Times (Hours)
Time point setting 1	5	8, 25, 75, 130, 170
Time point setting 2	10	5, 10, 30, 50, 70, 90, 110, 130, 150, 170
Time point setting 3	15	10, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 170

3.4.1 Performance of Parameter Estimators

3.4.1.1 Simulation Settings

We consider two different sets of $n = \{3, 6\}$ temperature levels and three different sets of $J_n = \{5, 10, 15\}$ measuring times. The specific settings are summarized in Table 3.1. Ten samples are tested at each combination of temperature level and measuring times. The data are simulated from the following model:

$$y_{ijk} = \sum_{l=1}^p r_l B_{q,l}[\eta_i(t_{ij}; \beta)] + \varepsilon_{ijk}, \quad (3.11)$$

where the degree of the B-splines is $q = 2$, and number of interior knots is $N = 3$. The knot locations are the sample quantiles. Figure 3.1 gives the spline basis functions and the baseline degradation function for scenario $n = 3$, $J_n = 5$. The true parameters in the model are $\beta = 0.83$, $\gamma = (1, 0.9, 0.8, 0.7, 0.6, 0.6)'$, and $(\sigma, \rho)' = (0.019, 0.2)'$.

For each scenario, 500 datasets are generated and the bias, SD, and MSE of the parameter estimators and baseline degradation curves are calculated. The quantile and bias-corrected CI are computed based on $B = 1,000$ bootstrap samples and the CP is also computed.

3.4.1.2 Simulation Results

Figure 3.2 shows the bias and MSE of parameter estimators. Figure 3.3 shows the pointwise MSE curves of baseline degradation curves. We found out that MSE of point estimators and baseline degradation curves decrease as either number of temperature levels or time points increases. Even when the number of temperature levels and time points are both small, biases of β and σ are small, while bias of ρ is large. However, when either number of temperature levels or time points is large, the estimates of β, σ and ρ are all close to the true values.

Figures 3.4 and 3.5 present the CP for quantile-based CI and bias-corrected CI of the parameter estimators and baseline degradation curves. The performance of bias-corrected CI seems to be similar for β , and better for σ, ρ and baseline degradation curve compared to quantile-based CI. For the parameter estimators, the CP of bias-corrected CI of β is good when n or J_n is small. However, the CP of bias-corrected CI of $(\sigma, \rho)'$ are overall slightly less than the desired confidence level. For the baseline degradation function, the CP of pointwise bias-corrected CI are poor when $n = 3$ and $J_n = 5$. The performance of pointwise bias-corrected CI improve as n and J_n increases. Overall, the results show that the performance of the estimation and inference procedures are good.

3.4.2 Performance under Model Misspecification

3.4.2.1 Simulation Settings

In this simulation study, the data are simulated according to a parametric model, but the semi-parametric model is fit to the data. The temperature levels are set at 50°C, 65°C, 80°C and the measuring times are set at 192, 600, 1800, 3120, and 4320 hours. There are 10 measurements at time 0 and 5 measurements at all other measuring times. The data are

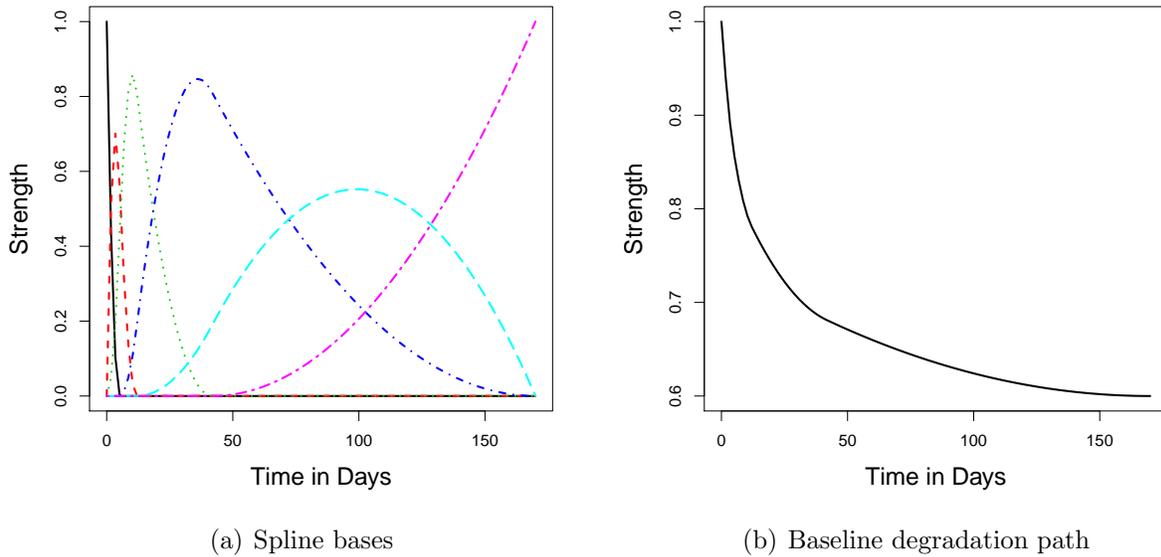


Figure 3.1: Spline bases and baseline degradation path used in simulation study.

simulated from the model

$$y_{ijk} = \beta_0 + \beta_1 \exp(\beta_2 x_i) t_j + \varepsilon_{ijk}, \quad (3.12)$$

where $t_j = \text{Hour}_j$, $x_i = -11605/(\text{Temp}_i + 273.15)$. The true parameters are $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (1, -3.5, 0.3)'$, and $(\sigma, \rho)' = (0.02, 0)'$. It is rare for the true model to be known exactly, so we also consider the case when a different parametric model from the true one is fit to the data. The incorrect parametric model, adapted from Vaca-Trigo and Meeker (2009), is given by

$$y_{ijk} = \frac{\alpha}{1 + \left[\frac{t_{ij}}{\exp(\beta_0 + \beta_1 x_i)} \right]^\gamma} + \varepsilon_{ijk}, \quad (3.13)$$

with parameters $(\alpha, \beta_0, \beta_1, \gamma)'$ in the mean structure. We fit the true model (3.12), the incorrect parametric model (3.13), and our semi-parametric model (3.3) to the simulated data. Figure 3.6 shows one case of the simulated data and the fitted degradation paths.

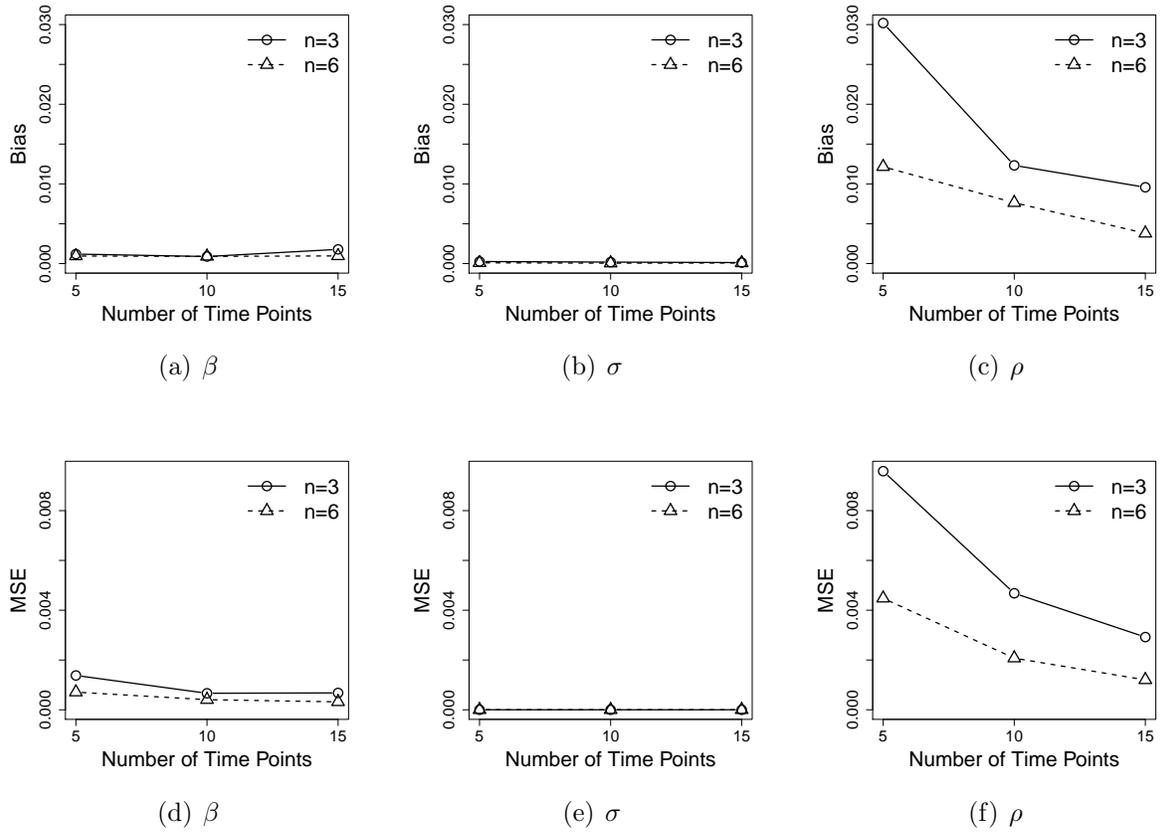


Figure 3.2: Empirical bias and MSE of parameter estimators for $(\beta, \sigma, \rho)'$.

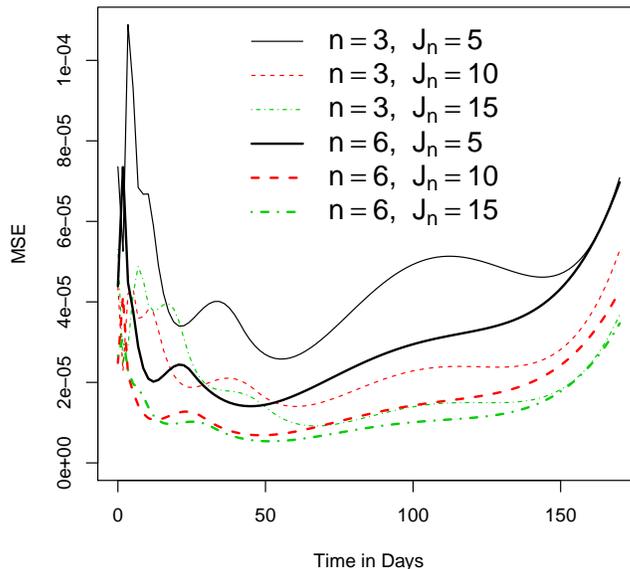


Figure 3.3: Empirical pointwise MSE for the estimator of the baseline degradation path.

3.4.2.2 Simulation Results

To assess the fit of our semi-parametric model, we compare the fitted degradation path to the true degradation path using the integrated mean square error (IMSE) of the baseline degradation function, which is defined as

$$\begin{aligned} \text{IMSE} &= \int_0^{t_m} \mathbf{E} \{ [\hat{g}(t; \gamma) - g(t; \gamma)]^2 \} dt \\ &= \int_0^{t_m} \{ \mathbf{E} [\hat{g}(t; \gamma)] - g(t; \gamma) \}^2 dt + \int_0^{t_m} \text{Var} [\hat{g}(t; \gamma)] dt = \text{IBias}^2 + \text{IVar}, \end{aligned}$$

where t_m is the maximum time under the maximum level of the accelerating variable. As there is no closed-form expressions for IMSE, IBias and IVar, we report the empirical results. Table 3.2 presents these results, which indicate that the performance of our semi-parametric model is good. The largest contribution to the root IMSE comes from the variance component. Thus, it is not surprising that the incorrect parametric model (3.13) performs the

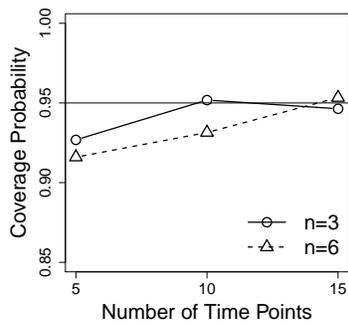
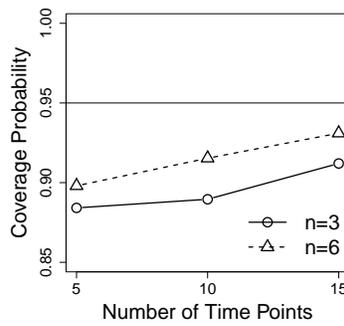
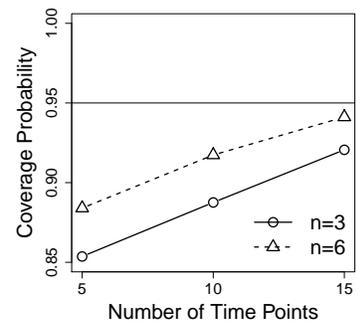
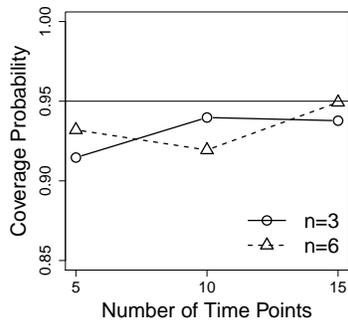
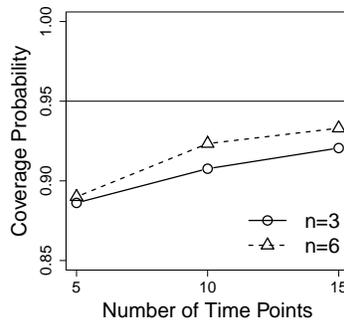
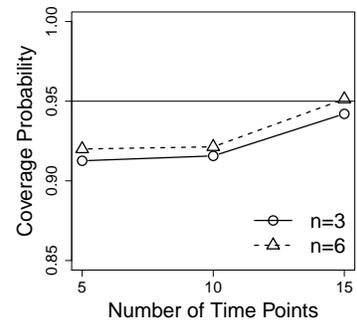
(a) β , Quantile-based CI(b) σ , Quantile-based CI(c) ρ , Quantile-based CI(d) β , Bias-corrected CI(e) σ , Bias-corrected CI(f) ρ , Bias-corrected CI

Figure 3.4: CP of the CI procedures for parameters $(\beta, \sigma, \rho)'$, using quantile-based and bias-corrected methods, respectively.

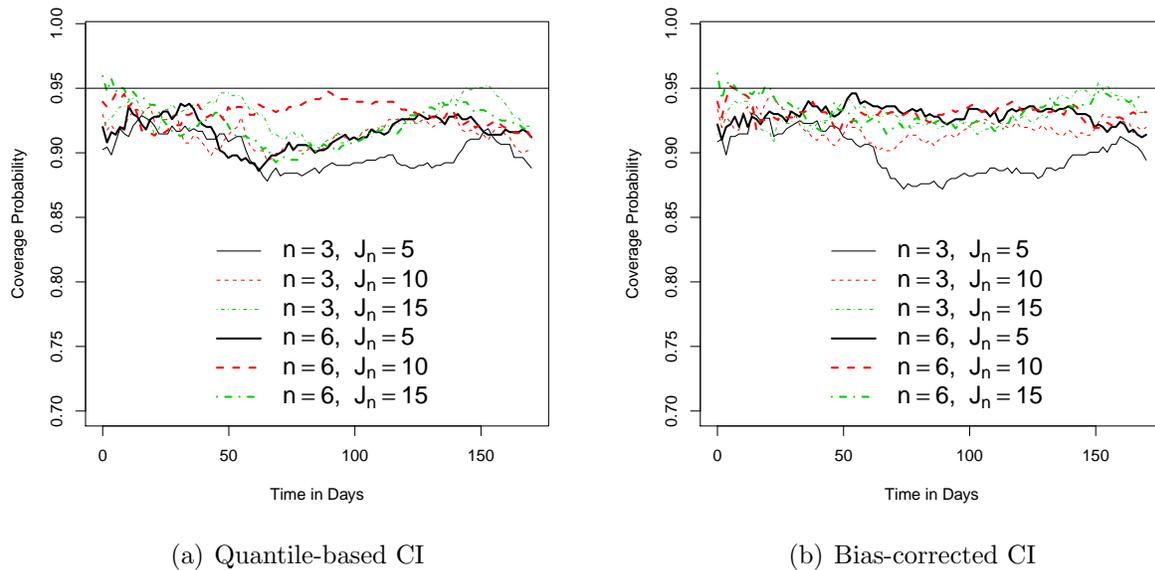


Figure 3.5: Pointwise CP of the CI procedure for baseline degradation path, using quantile-based and bias-corrected methods, respectively.

Table 3.2: Empirical IBias, root IVar (RIVar), and root IMSE (RIMSE) for the true model (3.12), incorrect model (3.13), and the semi-parametric model.

Models	IBias	RIVar	RIMSE
True Model	0.0003	0.0043	0.0043
Incorrect Model	0.0267	0.0060	0.0274
Semi-parametric Model	0.0003	0.0091	0.0091

worst in capturing the true degradation path.

For each simulated dataset, the MTTF (here failure is defined as the measurement drops below 0.5) at 30°C is calculated based on the true parametric model (3.12), incorrect parametric model (3.13) and the semi-parametric model. The mean, bias, standard derivation and root MSE of the MTTF for each of the different models based on 600 datasets are summarized in Table 3.3. The results indicate that the estimate of MTTF from our semi-parametric model is close to the true values, but with larger variance. The estimated MTTF from the incorrect parametric model (3.13) has the largest bias. The results indicate our semi-parametric model performs quite well.

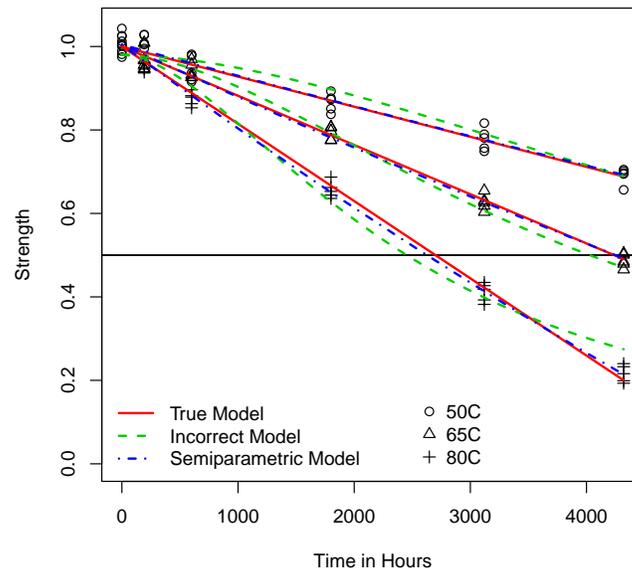


Figure 3.6: Plot of simulated data and fitted degradation paths based on the true and incorrect parametric models, and semi-parametric model.

Table 3.3: Empirical mean, bias, SD, and root MSE (RMSE) of the MTTF estimators based on the true model (3.12), incorrect model (3.13), and the semi-parametric model.

Models	Mean	Bias	SD	RMSE
True Model	82.60	0.01	2.99	2.99
Incorrect Model	85.82	3.20	3.75	4.93
Semi-parametric Model	82.77	0.16	4.22	4.22

3.5 Applications

To help motivate the use of our semi-parametric model, we selected three published datasets from well-known examples of ADDT. The data for each example are summarized below.

3.5.1 ADDT Datasets and Parametric Models

3.5.1.1 Adhesive Bond B Data

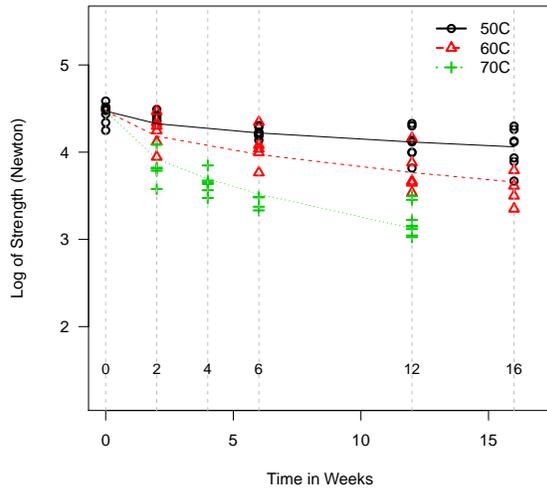
Escobar et al. (2003) discussed an experiment that measured the strength of an adhesive bond (Adhesive Bond B) over time. Eight units were measured at the beginning of the experiment under normal temperature to serve as the baseline strength. The remaining measurements were taken at selected weeks (2, 4, 6, 12, and 16) for three accelerated temperature levels (50°C, 60°C, and 70°C). A scatter plot of Adhesive Bond B dataset is presented in Figure 3.7(a). The degradation model used by Escobar et al. (2003) is

$$y_{ijk} = \beta_0 + \beta_1 \exp(\beta_2 x_i) \sqrt{\text{Week}_j} + \varepsilon_{ijk}, \quad (3.14)$$

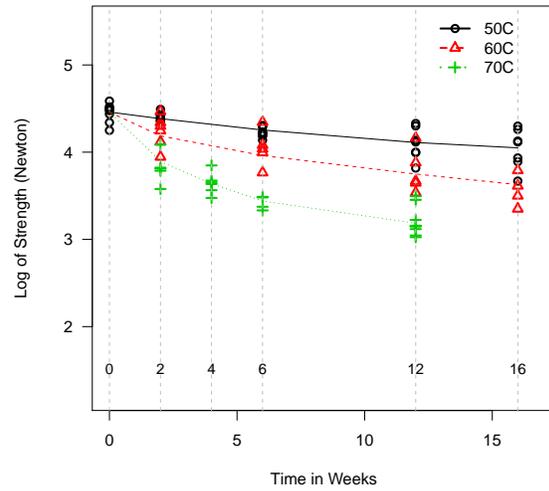
where y_{ijk} is the strength of Adhesive Bond B in log Newtons, $x_i = -11605 / (\text{Temp}_i + 273.15)$ is the Arrhenius-transformed temperature, and $\varepsilon_{ijk} \sim N(0, \sigma^2)$. The estimates are $\hat{\beta}_0 = 4.4713$, $\hat{\beta}_1 = -8.6384 \times 10^8$, $\hat{\beta}_2 = 0.6364$ and $\hat{\sigma} = 0.1609$.

3.5.1.2 Seal Strength Data

Seal strength data were considered by Li and Doganaksoy (2014). At the start of the experiment, a batch of 10 seals were measured at the use temperature level of 100°C. A batch of 10 seal samples were then tested at selected weeks (5, 10, 15, 20, and 25) for four temperature levels (200°C, 250°C, 300°C, and 350°C). A scatter plot of seal strength data is shown in Figure 3.9(a). Though one would expect the seal strength to decrease under higher temperature, some batches of seal samples yielded higher strengths in later weeks compared

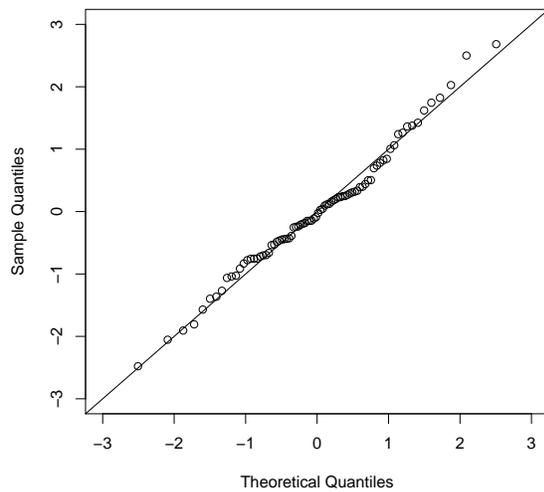


(a) Parametric model

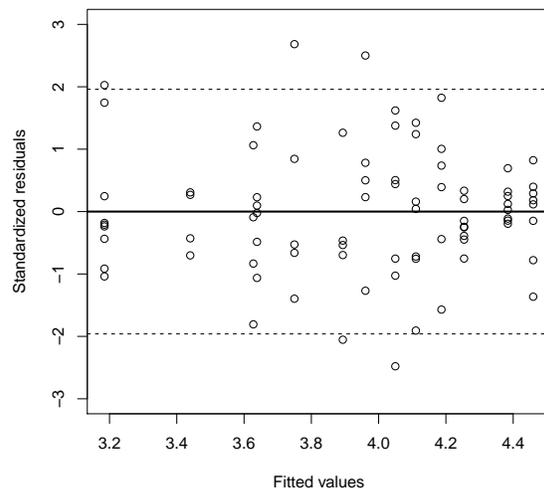


(b) Semi-parametric model

Figure 3.7: Fitted degradation paths of the Adhesive Bond B data.



(a) Normal probability plot



(b) Standardized residuals versus the fitted values

Figure 3.8: Residual analysis for the Adhesive Bond B data.

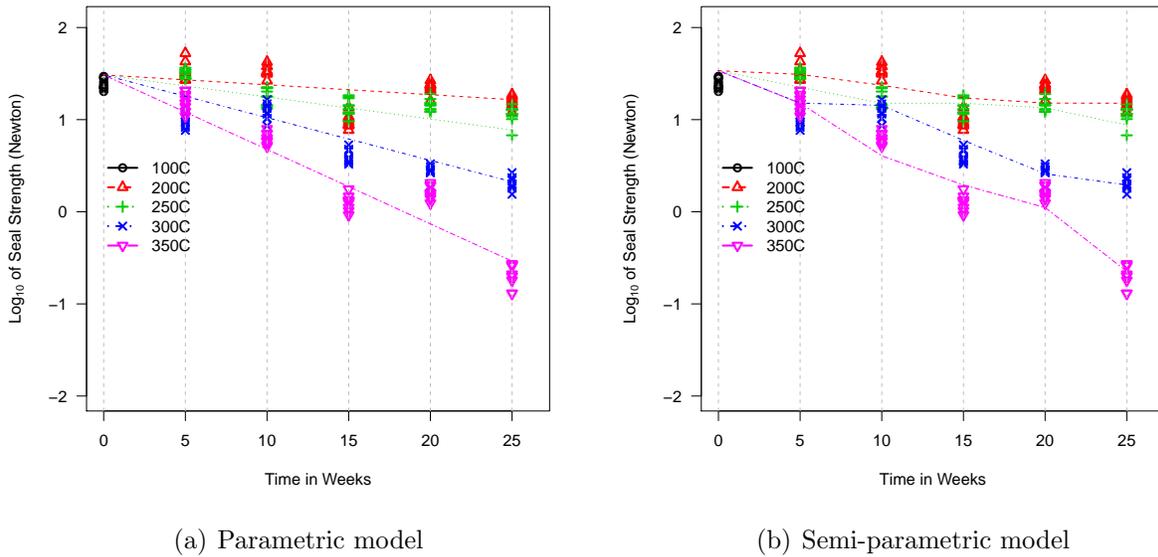


Figure 3.9: Fitted degradation paths of the Seal Strength data.

with the initial measurements. This suggests a large batch-to-batch variability which must be incorporated into the model. Thus, Li and Doganaksoy (2014) considered the following nonlinear mixed model:

$$y_{ijk} = \beta_0 - \beta_1 \exp(\beta_2 x_i) \text{Week}_j + \delta_{ij} + \varepsilon_{ijk}, \quad (3.15)$$

where y_{ijk} is the \log_{10} strength of seal sample, and $x_i = -11605/(\text{Temp}_i + 273.15)$. The random variable $\delta_{ij} \sim N(0, \sigma_\delta^2)$ represents batch variability, $\varepsilon_{ijk} \sim N(0, \sigma^2)$, and δ_{ij} and ε_{ijk} are independent. The estimates are $\hat{\beta}_0 = 1.4856$, $\hat{\beta}_1 = 47.2166$, $\hat{\beta}_2 = 0.3420$, $\hat{\sigma} = 0.1603$, and $\hat{\sigma}_\delta = 0.0793$.

3.5.1.3 Adhesive Formulation K Data

A new adhesive (Formulation K) was developed and tested at 40°C, 50°C, and 60°C. The strength of 10 units were measured at the beginning of the experiment and a specified number of samples were tested at 3, 6, 12, 18, and 24 weeks. Figure 3.11(a) is a scatter plot of the

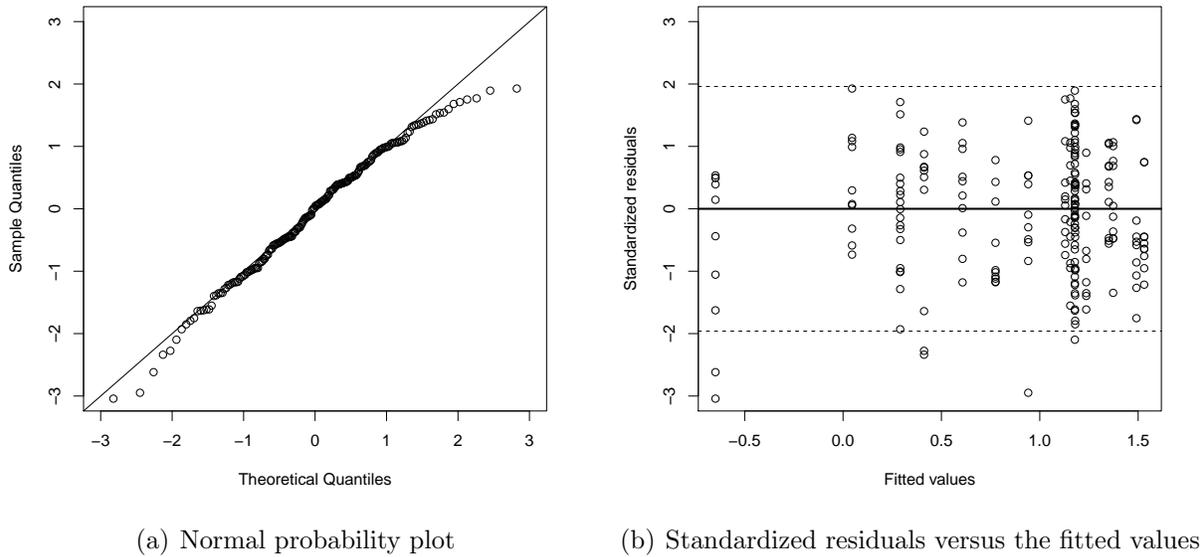


Figure 3.10: Residual analysis for the Seal Strength data.

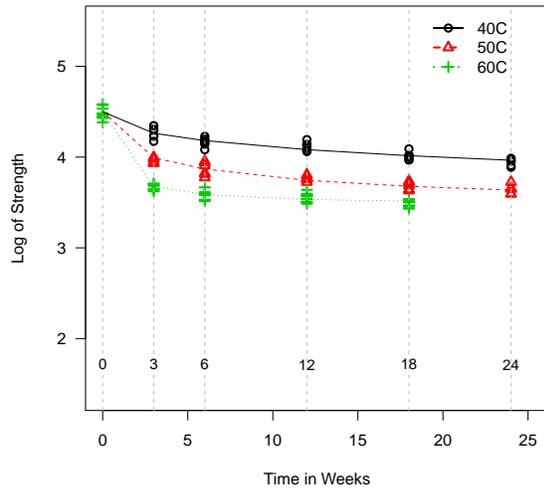
data. The nonlinear degradation model is

$$y_{ijk} = \log(90) + \beta_0(1 - \exp \left\{ -\beta_1 \exp [\beta_2(x_i - x_2)] \sqrt{\text{Week}_j} \right\}) + \varepsilon_{ijk}, \quad (3.16)$$

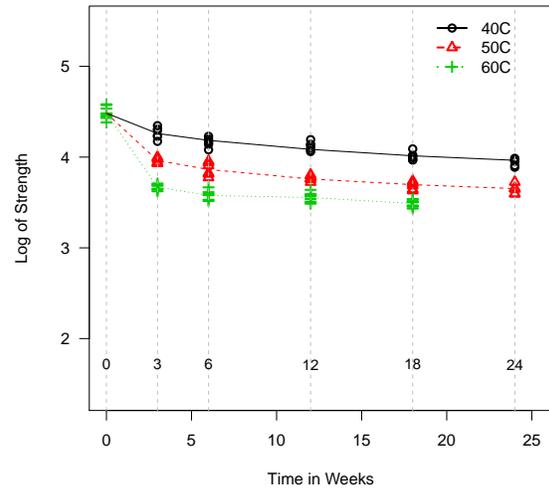
where y_{ijk} is the strength of Adhesive Formulation K in log Newtons, $x_i = -11605/(\text{Temp}_i + 273.15)$, $x_2 = -11605/(50 + 273.15)$, and $\varepsilon_{ijk} \sim N(0, \sigma^2)$. The estimates are $\hat{\beta}_0 = -0.9978$, $\hat{\beta}_1 = 0.4091$, $\hat{\beta}_2 = 0.8371$, and $\hat{\sigma} = 0.0501$.

3.5.2 Comparisons of Parametric and Semi-parametric Models

In order to assess the fit of the semi-parametric model, we applied it to each of the datasets and compared it with the corresponding parametric model chosen by the respective applications. We applied the knot selection technique in Section 3.3.3 for each application. We also checked whether $\rho = 0$ which informs the selection of appropriate covariance structure. The parameter estimates and CI, as well as the MTTF at the normal use condition are presented in Tables 3.4 and 3.5. The AIC defined in Section 3.3.3 can also be used to compare the

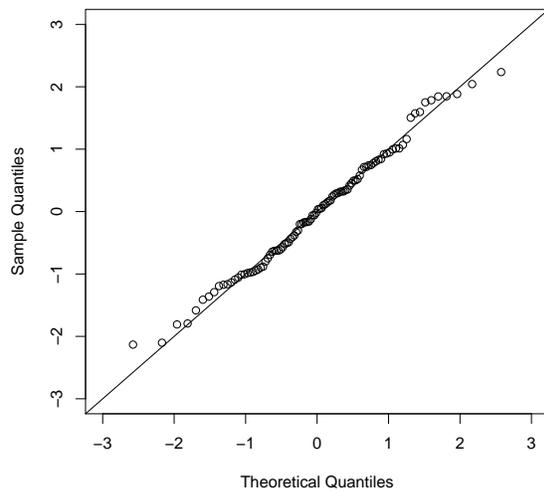


(a) Parametric model

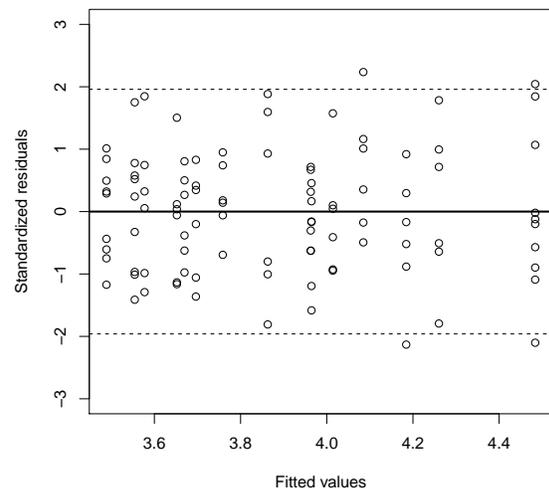


(b) Semi-parametric model

Figure 3.11: Fitted degradation paths of the Adhesive Formulation K data.



(a) Normal probability plot



(b) Standardized residuals versus the fitted values

Figure 3.12: Residual analysis for the Adhesive Formulation K data.

Table 3.4: Parameter estimates and corresponding CI for the semi-parametric models for the three applications.

Applications	Parameter	Estimate	Quantile-based CI	
			95% lower	95% upper
Adhesive Bond B	β	1.3422	1.1071	1.6165
	σ	0.1537	0.1265	0.1787
Seal Strength	β	0.3235	0.2451	0.5194
	σ	0.1610	0.1192	0.1904
	ρ	0.7573	0.5465	0.8307
Adhesive Formulation K	β	1.8221	1.6575	2.3658
	σ	0.0484	0.0419	0.0544

Table 3.5: Estimated MTTF and CI at normal use condition based on parametric and semi-parametric models for the three applications (time in weeks).

Applications	Failure Threshold	Normal Use Conditions	Parametric Models	Semi-parametric Models
Adhesive Bond B	70%	30°C	270	306
Seal Strength	70%	100°C	222	127
Adhesive Formulation K	70%	30°C	68	86

parametric and semi-parametric models. In the calculation of AIC, the log-likelihood is the marginal log-likelihood for the parametric models.

Table 3.6 contains the log-likelihood values, edf , and AIC for each model and dataset. For all three datasets, the semi-parametric models possessed a lower AIC as compared to the parametric models. The fitted degradation paths for the parametric and semi-parametric models are presented in Figures 3.7, 3.9, and 3.11. All three figures show that the semi-parametric models provide a good fit to the data. We can see that the proposed model is flexible in fitting ADDT data from different applications.

Table 3.6: Log likelihood and AIC values of parametric and semi-parametric models for the ADDT data from the three applications.

Applications	Parametric Models			Semi-parametric Models		
	Loglik	df	AIC	Loglik	edf	AIC
Adhesive Bond B	34.9665	4	-61.9330	38.7264	5	-67.4418
Seal Strength	194.9907	5	-379.9814	199.7454	6	-387.4909
Adhesive Formulation K	158.9508	4	-309.9016	163.9898	8	-311.9797

Checking the semi-parametric model assumptions is very important. We did some graphical checks based on the standardized residuals. Take the adhesive bond B as an example, Figure 3.8(a) is a normal probability plot for the standardized residuals and indicates that the normal assumption is appropriate; Figure 3.8(b) is a scatter plot of standardized residuals versus the fitted values. No unusual pattern is exhibited in Figure 3.8(b). Residual analysis of the seal strength data and adhesive formulation K data are summarized in Figures 3.10 and 3.12. Both figures show that the model assumptions are satisfied.

3.5.3 Illustration of failure time distribution

For each application, the quantile functions and corresponding confidence intervals can be calculated. We use the adhesive bond B data as an example. Figure 3.13 shows the quantiles and 95% pointwise bootstrap CIs for four temperature levels. The dotted lines are the pointwise confidence intervals. From the model specification, the α quantile at level x_i is the α quantile at baseline degradation level times the factor $\exp(\beta x_i)$, therefore the quantiles in Figure 3.13 are parallel.

3.6 Conclusions and Areas for Future Work

In this chapter, we propose a new semi-parametric degradation model for ADDT data based on monotone B-splines. We develop estimation and inference procedures for the proposed model as well as methods for selecting knot locations for the B-splines. Our simulation results indicate that the proposed estimation procedures for our semi-parametric model perform well. Compared to parametric models, our semi-parametric approach is more flexible and can be applied to a wide range of applications and may be best suited as a generic method for ADDT data analysis for industrial standards. In addition, the semi-parametric model is more robust to model misspecification than a parametric model approach.

One key application of our semi-parametric model could be for test planning. A test plan based on this model would be general enough for application to a variety of materials

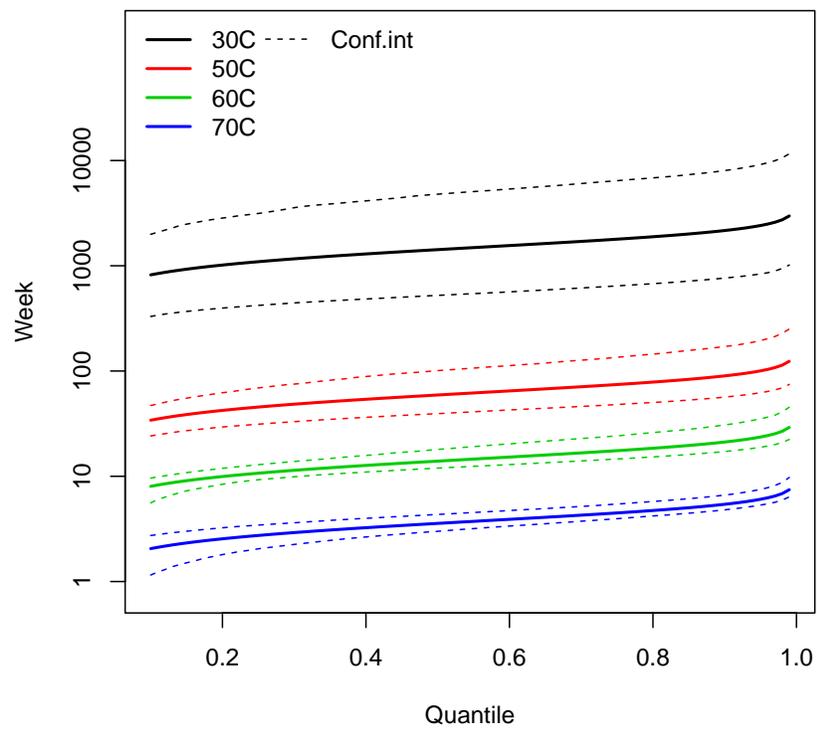


Figure 3.13: Estimates and CIs of quantile functions at different temperature levels for Adhesive Bond B data.

and also allow for testing of different models. Our model can be served as a starting ground from which to test models against the data gathered rather than having to assume a given model prior to data collection. This would certainly serve as an interesting topic for future research.

The models considered here were solely scale-acceleration models. However, for certain types of products, a model with both scale and shape acceleration may describe the degradation path more appropriately. For example, Tsai et al. (2013) considered a parametric model with both scale and the shape acceleration in test planning. Estimation and inference procedures for the semi-parametric model would certainly be more complex with the introduction of a shape acceleration parameter. It would be of great interest to pursue this in future research.

Bibliography

- K. Bollaerts, P. H. Eilers, and I. Mechelen. Simple and multiple P-splines regression with shape constraints. *British Journal of Mathematical and Statistical Psychology*, 59(2):451–469, 2006.
- J. R. Carpenter, H. Goldstein, and J. Rasbash. A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4):431–443, 2003.
- C. De Boor. *A practical guide to splines*. Springer-Verlag, New York, 2001.
- B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. CRC Press, 1994.
- P. H. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121, 1996.
- L. A. Escobar, W. Q. Meeker, D. L. Kugler, and L. L. Kramer. Accelerated destructive degradation tests: Data, models, and analysis. *Mathematical and Statistical Methods in Reliability*, 7:319–338, 2003.
- M. R. Fengler and L.-Y. Hin. A simple and general approach to fitting the discount curve under no-arbitrage constraints. *Finance Research Letters*, 15:78–84, 2015.
- F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.
- N. Gorjian, L. Ma, M. Mittinty, P. Yarlagadda, and Y. Sun. A review on degradation models in reliability analysis. In *Engineering Asset Lifecycle Management*, pages 369–384. Springer, 2010.

- X. He and P. Shi. Monotone B-spline smoothing. *Journal of the American Statistical Association*, 93(442):643–650, 1998.
- B. Hofner, J. Müller, and T. Hothorn. Monotonicity-constrained species distribution models. *Ecology*, 92(10):1895–1901, 2011.
- B. Hofner, T. Kneib, and T. Hothorn. A unified framework of constrained regression. *Statistics and Computing*, 26(1-2):1–14, 2016.
- Y. Hong, Y. Duan, W. Q. Meeker, D. L. Stanley, and X. Gu. Statistical methods for degradation data with dynamic covariates information and an application to outdoor weathering data. *Technometrics*, 57(2):180–193, 2015.
- T. Kanungo, D. M. Gay, and R. M. Haralick. Constrained monotone regression of roc curves and histograms using splines and polynomials. In *Image Processing, 1995. Proceedings., International Conference on*, pages 292–295. IEEE, 1995.
- F. Leitenstorfer and G. Tutz. Generalized monotonic regression based on B-splines with an application to air pollution data. *Biostatistics*, 8(3):654–673, 2007.
- M. Li and N. Doganaksoy. Batch variability in accelerated-degradation testing. *Journal of Quality Technology*, 46(2):171–180, 2014.
- C. J. Lu and W. O. Meeker. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2):161–174, 1993.
- W. Q. Meeker, L. A. Escobar, and C. J. Lu. Accelerated degradation tests: modeling and analysis. *Technometrics*, 40(2):89–99, 1998.
- W. Q. Meeker, Y. Hong, and L. A. Escobar. Degradation models and data analyses. In *Encyclopedia of Statistical Sciences*. Wiley, 2011.
- M. C. Meyer. Inference using shape-restricted regression splines. *The Annals of Applied Statistics*, 2(3):1013–1033, 2008.

- M. C. Meyer. Constrained penalized splines. *Canadian Journal of Statistics*, 40(1):190–206, 2012.
- W. B. Nelson. *Accelerated testing: statistical models, test plans, and data analysis*. John Wiley & Sons, 1990.
- J. O. Ramsay. Monotone regression splines in action. *Statistical Science*, 3(4):425–441, 1988.
- C.-C. Tsai, S.-T. Tseng, N. Balakrishnan, and C.-T. Lin. Optimal design for accelerated destructive degradation tests. *Quality Technology and Quantitative Management*, 10(3):263–276, 2013.
- I. Vaca-Trigo and W. Q. Meeker. A statistical model for linking field and laboratory exposure results for a model coating. In *Service Life Prediction of Polymeric Materials*, pages 29–43. Springer, 2009.
- H. Wang, M. C. Meyer, and J. D. Opsomer. Constrained spline regression in the presence of AR (p) errors. *Journal of Nonparametric Statistics*, 25(4):809–827, 2013.
- Z. Xu, Y. Hong, and R. Jin. Nonlinear general path models for degradation data with dynamic covariates. *Applied Stochastic Models in Business and Industry*, 2015.
- Z. Ye and M. Xie. Stochastic modelling and analysis of degradation for highly reliable products. *Applied Stochastic Models in Business and Industry*, 31(1):16–32, 2015.
- Z.-S. Ye, M. Xie, L.-C. Tang, and N. Chen. Semiparametric estimation of Gamma processes for deteriorating products. *Technometrics*, 56(4):504–513, 2014.

Chapter 4 Spatial Variable Selection via Elastic Net and an Application to Virginia Lyme Disease Emergence

Abstract

Lyme disease is an infectious disease that is caused by a bacteria called *Borrelia burgdorferi*. In US, the major endemic areas are New England, Mid-Atlantic, East-North Central, South Atlantic, and West North-Central. Virginia is on the front-line of spreading of the disease from north east to south. One of the research goal is to identify if there are any environmental and economical variables that are associated with the emergency of the disease. In this chapter, we develop spatial variable selection procedures to address this problem. A general linear mixed model (GLMM) is used to describe the spatial data. We impose the adaptive elastic net penalty to select important covariates. The performance of the proposed procedures is evaluated via simulation study. Then we apply the variable selection methods to the Virginia Lyme disease data and compare our findings with literature.

Key Words: GLMM; Laplace Approximation; Multicollinearity; Poisson Regression; PQL; Spatial Count Data.

4.1 Introduction

4.1.1 Background

Spatial data modeling attracts great attention in recent years. It has applications in ecology, epidemiology, agriculture, sociology and so on. For spatial data, the correlation induced by distances among locations is typically unnegligible and a proper way to incorporate such correlation into the model is needed. One way to model the spatial correlations among locations is through random effects. Moreover, in many cases, the interested measurement is non-Gaussian, for instance, the incidence rates, number of counts, and binary indicators of having heart diseases.

In application, it is possible that there are many variables available for modeling. It is often reasonable to assume only a subset of predictors is actually related to the outcome of interest. Having many irrelevant and/or redundant variables in the model can cause unstable estimation procedure, waste of computational time, difficulty of interpretation, unsatisfactory prediction performance, etc. Thus, it is important to do a variable selection.

4.1.2 Lyme Disease Emergence

An motivating example of this chapter is the Lyme disease data in Virginia from 2006 to 2011. This dataset is also introduced in Seukep et al. (2015), Duan (2014) and Li et al. (2014). Lyme disease is an infectious disease and transmitted via a bite of tick. Tick is a vector of many diseases, including Lyme disease, babesiosis, among others. Duan (2014) found that the Lyme disease spread from the northern part of Virginia to the southwestern part over the past decade, along with increasing number of cases. Furthermore, the number of Lyme disease cases has a increase after 2006. This motivates us to study the mechanism behind the disease, and discover crucial factors associated with emergence of Lyme disease.

The Lyme disease dataset for this chapter contains case data, demographic data and land cover data in Virginia. Lyme disease case data were collected by Virginia Department of

Health. The demographic data (e.g., population density, median income, average age) were from 2010 census. The land cover data were obtained from the Multi-Resolution Land Cover Consortium for 2006.

Our goal is to develop a method that can identify a subset of the explanatory variables that are important for the case counts of Lyme disease. As we can see from Lyme disease data (see Section 4.5 for more details), there exists spatial correlation and strong multicollinearity among explanatory variables. Variable selection while account for spatial dependence and multicollinearity are the challenging aspects.

4.1.3 Related Literature

Diggle et al. (1998) and Zhang (2002) employed generalized linear mixed model (GLMM) for spatial data with non-Gaussian outcomes. The GLMM is an extension of generalized linear model (GLM), with extra flexibility to capture the subject dis-similarities and correlations among observations by adding a random effect term to the linear predictor. It is a useful and popular tool for correlated observations. In this chapter, we model the spatial count data using GLMM. Various approaches to estimate the parameters in GLMM have been developed. An overview on current methods can be found in McCulloch et al. (2008).

A wide class of variable selection approaches have been developed. Classical approaches include backward, forward, stepwise, and all subsets selection procedures. A modern way to do variable selection is shrinkage methods. That is, we add a penalty term $P_\lambda(\boldsymbol{\beta})$ to the residuals sum of squares or loglikelihood, where $P_\lambda(\boldsymbol{\beta})$ is a function of $\boldsymbol{\beta}$ (vector of regression parameters) with regularization parameter vector $\boldsymbol{\lambda}$. The family of L^q penalties ($q > 0$, Frank and Friedman, 1993) are commonly used penalties. The least absolute shrinkage and selection operator (LASSO) penalty ($q = 1$, $P_\lambda(\boldsymbol{\beta}) = \boldsymbol{\lambda} \sum_i |\beta_i|$) is studied in Tibshirani (1996) to solve the regression type problem. It is shown that LASSO does parameter estimation and variable selection simultaneously due to the shrinkage property of L^1 penalty. The ridge penalty ($q = 2$, $P_\lambda(\boldsymbol{\beta}) = \boldsymbol{\lambda} \sum_i \beta_i^2$) introduced in Hoerl and Kennard (1970) always includes all the

covariates. If there are a group of highly correlated covariates, the ridge penalty shrinks coefficients to each other but not zero. Conversely, LASSO picks one covariate and assigns all weights to this covariate. In other words, ridge penalty tends to select the entire group, while LASSO tends to randomly pick only one covariate (Tibshirani, 1996). Zou and Hastie (2005) proposed the elastic net penalty, which is a linear combination of the LASSO penalty and the ridge penalty. For a group of highly correlated covariates, the combination of ridge and LASSO results in the trend of in and out together. Thus, the elastic net penalty has the property of automatic variable selection and continuous shrinkage. However, LASSO doesn't have the oracle property and can be inconsistent unless certain conditions are satisfied. In light of the drawbacks, the adaptive LASSO (Zou, 2006) and adaptive elastic net (Zou and Zhang, 2009) were developed. Beside the above work, Fan and Li (2001) developed the smoothly clipped absolute deviation (SCAD) penalty. The SCAD penalty has the properties of unbiasedness, sparsity and continuity. It can be used for the variable selection in linear models, and as an extension, in GLMs and GLMMs. One may refer to Fan and Lv (2010) for a comprehensive review of variable selection methods.

In addition to the above work, Bayesian methods are also popular for variable selection. Bayesian variable selection methods assign posterior probability to each model and automatically pick the one with the largest posterior probability. A review and comparison of Bayesian variable selection methods is available in O'Hara and Sillanpää (2009).

In terms of implementation, Efron et al. (2004) proposed the least-angle regression (LARS) method to efficiently calculate the solution path of LASSO penalty in linear models. Park and Hastie (2007) extended the concept of LARS algorithm to GLM. Their approach is also efficient and flexible. An algorithm named elastic net penalized least squares (LARS-EN) in Zou and Hastie (2005) is proposed for linear models with elastic net penalty. The LARS-EN algorithm works by transforming the elastic net penalty into a LASSO penalty, which can be solved by LARS.

Variable selection in GLMM, especially for large n (number of observations) or large p

(number of predictors) case, however, is still of difficulty and there is only a limit number of research on this. Schelldorfer et al. (2014) presented a GLMMLasso for high-dimensional GLMM with LASSO penalty. Their approach addressed the problem when number of observations is less than number of covariates, (i.e., $n \ll p$). The corresponding R package “`glmmixedlasso`” (Schelldorfer et al., 2012) is available online. Groll and Tutz (2014) also considered this type of problem and a gradient descent algorithm is proposed to maximize the penalized loglikelihood function. A associated R package “`glmmLasso`” (Groll, 2014) can be downloaded from website. A number of variable selection procedures for GLMMs with longitudinal data settings are studied in Yang (2007) and Cui (2011). Besides, Cai and Dunson (2006) proposed a fully Bayesian method to selection fixed and random effects in GLMM.

4.1.4 Overview

The rest of this chapter is organized as follows. In Section 4.2, we introduce the data, model, likelihood and penalty functions. In Section 4.3, we present two approximations to the penalized loglikelihood function, and explain the estimation procedures in detail. In Sections 4.4 and 4.5, we illustrate the methods with simulated data and Lyme disease data. Section 6 contains discussions and areas for future work.

4.2 Data, Model, Likelihood and Penalty Function

4.2.1 Data and Model

Consider a situation where we have observations at n spatial locations indexed by $i = 1, \dots, n$. Let y_i be the count of cases at location i , which takes values in $\{0, 1, 2, \dots\}$. The explanatory variables are denoted by $\mathbf{x}_i = (x_{i1}, \dots, x_{ij}, \dots, x_{ip})$, where p is the number of explanatory variables and x_{ij} is the value of the j th covariate at location i . Denote $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ as the vector of observations and X is the matrix with \mathbf{x}_i as its i th row.

Let m_i be the population of location i .

We use a spatial Poisson regression model with random effect to describe the spatial count data. That is,

$$y_i|b_i \sim \text{Poisson}(\mu_i), \text{ and} \\ \eta_i = \log(\mu_i) = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + b_i + \log(m_i). \quad (4.1)$$

In model (4.1), the responses y_i 's conditional on random effects b_i are independent for $i = 1, \dots, n$. μ_i 's are the conditional means. m_i 's are the offset term.

Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$, and $\mathbf{b} = (b_1, b_2, \dots, b_n)'$. The spatial correlations among locations are captured through random effects \mathbf{b} . We use multivariate normal distribution to model the random effect \mathbf{b} . That is, $\mathbf{b} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\theta}})$. The variance-covariance matrix of \mathbf{b} is $\Sigma_{\boldsymbol{\theta}} = \sigma^2\Omega$, and the ij th element of the Ω is $\rho(d_{ij}; \boldsymbol{\theta})$. Here $\rho(\cdot)$ is a spatial correlation function and $\boldsymbol{\theta}$ are parameters in $\Sigma_{\boldsymbol{\theta}}$. Note that d_{ij} is the distance between two locations i and j .

4.2.2 Likelihood Function and Penalty

Let $f(\mathbf{y}|\boldsymbol{\beta}, \mathbf{b})$ be the distribution of \mathbf{y} given \mathbf{b} , and $f(\mathbf{b}|\boldsymbol{\theta})$ be the distribution of random effects \mathbf{b} . The likelihood function of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ is

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \int_{\mathbb{R}^n} f(\mathbf{y}|\boldsymbol{\beta}, \mathbf{b})f(\mathbf{b}|\boldsymbol{\theta}) d\mathbf{b} \\ &= \int_{\mathbb{R}^n} \left[\prod_{i=1}^n \exp(-\mu_i) \frac{\mu_i^{y_i}}{y_i!} \right] \left[(2\pi)^{-\frac{n}{2}} |\Sigma_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{b}'\Sigma_{\boldsymbol{\theta}}^{-1}\mathbf{b}\right) \right] d\mathbf{b} \\ &= (2\pi)^{-\frac{n}{2}} |\Sigma_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \int_{\mathbb{R}^n} \exp\left\{ \sum_{i=1}^n [-\mu_i + y_i \log(\mu_i) - \log(y_i!)] - \frac{1}{2}\mathbf{b}'\Sigma_{\boldsymbol{\theta}}^{-1}\mathbf{b} \right\} d\mathbf{b}. \quad (4.2) \end{aligned}$$

Hence the loglikelihood is $l(\boldsymbol{\beta}, \boldsymbol{\theta}) = \log[L(\boldsymbol{\beta}, \boldsymbol{\theta})]$.

To perform variable selection, we add adaptive elastic net penalty term for fixed effects

$\boldsymbol{\beta}$ to the loglikelihood function. That is, we consider the following penalized loglikelihood function

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}) = -l(\boldsymbol{\beta}, \boldsymbol{\theta}) + P_{\boldsymbol{\lambda}}(\boldsymbol{\beta}), \quad (4.3)$$

where $P_{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \lambda_1 \left[\lambda_2 \sum_{j=1}^p \hat{w}_j |\beta_j| + (1 - \lambda_2) \sum_{j=1}^p \beta_j^2 \right]$ is the adaptive elastic net penalty. Here λ_1, λ_2 are regularization parameters. Note that $0 \leq \lambda_2 \leq 1$, $\lambda_2 = 1$ is the case of LASSO penalty and $\lambda_2 = 0$ is the case of ridge penalty. $\hat{w}_j = |\hat{\boldsymbol{\beta}}_{\text{cnst}}|^{-r}$ is the adaptive weight, $r > 0$ and $\hat{\boldsymbol{\beta}}_{\text{cnst}}$ is a consistent estimate of $\boldsymbol{\beta}$. The unpenalized estimate of $\boldsymbol{\beta}$ is a good choice of $\hat{\boldsymbol{\beta}}_{\text{cnst}}$. Note that we don't impose penalty on β_0 , that is the intercept term is always included in the model.

4.3 Estimation Procedures

4.3.1 The Estimation Problem

Our objective is to obtain the parameter estimation via optimizing the penalized loglikelihood in (4.3). That is

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) := \operatorname{argmin}_{\boldsymbol{\beta}, \boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}).$$

The joint likelihood function (4.2) contains intractable integrals over distribution of random effects. If the random effects are of low dimension, we may use Gaussian quadrature to do numerical integral. However, in spatial Poisson regression model with random effect, the dimension of random effects is the same as the number of observations. That is, the dimension of integrals is typically so large that the Gaussian quadrature or other low-dimensional methods may not work.

4.3.2 Penalized Quasi-likelihood

Because the full likelihood approach is infeasible, if not impossible, approximated likelihood is considered. Several approaches have been established to approximate the likelihood function. One popular method named penalized quasi-likelihood (PQL) is introduced in Breslow and Clayton (1993). The approximation of likelihood function can be derived by employing the result of Laplace approximation. In general, the Laplace approximation (Laplace, 1986) of multi-dimensional integrals of the form

$$\int_{\mathbb{R}^n} \exp[h(\mathbf{b})] d\mathbf{b} \approx (2\pi)^{\frac{n}{2}} | -h''(\tilde{\mathbf{b}}) |^{-\frac{1}{2}} \exp[h(\tilde{\mathbf{b}})],$$

where $\tilde{\mathbf{b}}$ is the maximizer of function $h(\mathbf{b})$.

In our case, $h(\mathbf{b}) = \sum_{i=1}^n [-\mu_i + y_i \log(\mu_i) - \log(y_i!)] - \mathbf{b}'\Sigma^{-1}\mathbf{b}/2$. $\tilde{\mathbf{b}}$ can be calculated via iteratively weighted least square (IWLS) algorithm (see Bates, 2008 for details). Applying the Laplace approximation to likelihood function (4.2), we obtain

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx \left| \Sigma_{\boldsymbol{\theta}} W + I_n \right|^{-\frac{1}{2}} \exp \left\{ \sum_{i=1}^n [-\mu_i + y_i \log(\mu_i) - \log(y_i!)] - \frac{1}{2} \tilde{\mathbf{b}}' \Sigma_{\boldsymbol{\theta}}^{-1} \tilde{\mathbf{b}} \right\}, \quad (4.4)$$

where $W = \text{Diag}\{\boldsymbol{\mu}\}$ and I_n is $n \times n$ diagonal matrix. Notice that $\tilde{\mathbf{b}}$ depends on parameters $(\boldsymbol{\beta}, \boldsymbol{\theta})$. The PQL proposed in Breslow and Clayton (1993) ignores the dependency of the first term in (4.4) on $\boldsymbol{\beta}$ (i.e., treat W is not related with $\boldsymbol{\beta}$), and yields the following approximated loglikelihood:

$$l_{\text{PQL}}(\boldsymbol{\beta}, \mathbf{b}|\boldsymbol{\theta}) = \sum_{i=1}^n [-\mu_i + y_i \log(\mu_i) - \log(y_i!)] - \frac{1}{2} \mathbf{b}' \Sigma_{\boldsymbol{\theta}}^{-1} \mathbf{b}. \quad (4.5)$$

Breslow and Clayton (1993) described an Fisher scoring algorithm to obtain estimates of $(\boldsymbol{\beta}, \mathbf{b})$ from (4.5) simultaneously. They showed that it is equivalently to fit a normal linear

mixed model (LMM):

$$\mathbf{y}^* = X\boldsymbol{\beta} + \mathbf{b} + \boldsymbol{\epsilon}, \text{ with } \mathbf{b} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\theta}}), \text{ and } \boldsymbol{\epsilon} \sim N(\mathbf{0}, W^{-1}). \quad (4.6)$$

Here $\mathbf{y}^* = (y_1^*, \dots, y_n^*)$ is the working response vector with $y_i^* = \mathbf{x}'_i \boldsymbol{\beta} + b_i + (y_i - \mu_i)/\mu_i$. Based on the model formulation, we have $\text{Var}(\mathbf{y}^*) = V = W^{-1} + \Sigma_{\boldsymbol{\theta}}$. We update $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{b}}$ in the following formulas iteratively until converge.

$$\begin{aligned} \tilde{\boldsymbol{\beta}} &= (X'V^{-1}X)^{-1} X'V^{-1}\mathbf{y}^*, \\ \tilde{\mathbf{b}} &= \Sigma_{\boldsymbol{\theta}}V^{-1}(\mathbf{y}^* - X\tilde{\boldsymbol{\beta}}). \end{aligned}$$

Based on current estimates $(\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}})$, linear mixed model theory can also be used to estimate covariance parameters $\boldsymbol{\theta}$. Breslow and Clayton (1993) suggested the following restricted maximum likelihood (REML) version of loglikelihood for estimation of $\boldsymbol{\theta}$:

$$l_{\text{R}}(\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}) = -\frac{1}{2} \log |V| - \frac{1}{2} \log |X'V^{-1}X| - \frac{1}{2} (\mathbf{y}^* - X\tilde{\boldsymbol{\beta}})' V^{-1} (\mathbf{y}^* - X\tilde{\boldsymbol{\beta}}). \quad (4.7)$$

The first derivative of $l_{\text{R}}(\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}})$ is

$$l'_{\text{R}}(\boldsymbol{\theta}|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}) = \frac{1}{2} \text{tr} \left(P \frac{\partial V}{\partial \theta_j} \right) - \frac{1}{2} (\mathbf{y}^* - X\tilde{\boldsymbol{\beta}})' V^{-1} \frac{\partial V}{\partial \theta_j} V^{-1} (\mathbf{y}^* - X\tilde{\boldsymbol{\beta}}), \quad (4.8)$$

and the (j, k) th component of Fisher Information matrix has the form

$$-\frac{1}{2} \left(P \frac{\partial V}{\partial \theta_j} P \frac{\partial V}{\partial \theta_k} \right).$$

Here $P = V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}$. Therefore, we can use Fisher scoring algorithm to solve $\boldsymbol{\theta}$. We emphasize that the dependency of W on $\boldsymbol{\theta}$ is ignored in $\partial V/\partial \theta_j$.

4.3.3 PQL with Adaptive Elastic Net Penalty

4.3.3.1 Estimation of β

To estimate β with adaptive elastic net penalty, we expand $l_{\text{PQL}}(\beta|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}})$ in a way similar to Friedman et al. (2010). Notice that $l_{\text{PQL}}(\beta|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}})$ is a concave function of β given $\tilde{\boldsymbol{\theta}}$ and $\tilde{\mathbf{b}}$. Given $\tilde{\beta}$ that maximizes $l_{\text{PQL}}(\beta|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}})$, we form a quadratic approximation to $l_{\text{PQL}}(\beta|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}})$ around $\tilde{\beta}$, then we have

$$l_{\text{PQL}}^Q(\beta|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}}) \approx -\frac{1}{2} \sum_{i=1}^n \mu_i (z_i - \mathbf{x}'_i \beta)^2,$$

where $z_i = \mathbf{x}'_i \tilde{\beta} - 1 + y_i/\mu_i$ (details regarding to the calculation are in Appendix 4.A).

Incorporating the penalty function, we have

$$\mathcal{L}_{\text{PQL}}^Q(\beta|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}}) = \frac{1}{2} \sum_{i=1}^n \mu_i (z_i - \mathbf{x}'_i \beta)^2 + P_\lambda(\beta). \quad (4.9)$$

The minimizer $\tilde{\beta}_Q$ of (4.9) can be achieved as a penalized weighted least squares problem by “*glmnet*” package in R.

In fact, estimating β via quadratic approximated PQL (4.9) is equivalent to consider the penalized loglikelihood (PL) of the linear mixed model (4.6) in Section 4.3.2:

$$l_{\text{LMM.PL}}(\beta, \mathbf{b}|\boldsymbol{\theta}) = -\frac{1}{2}(\mathbf{y}^* - X\beta - \mathbf{b})'W(\mathbf{y}^* - X\beta - \mathbf{b}) - \frac{1}{2}\mathbf{b}'\Sigma_{\boldsymbol{\theta}}^{-1}\mathbf{b} - P_\lambda(\beta). \quad (4.10)$$

To see the connection between (4.9) and (4.10), we estimate $\tilde{\mathbf{b}}$ first and let $\mathbf{y}^{**} = \mathbf{y}^* - \tilde{\mathbf{b}}$. Then again we have a weighted linear regression with elastic net penalty problem. That is, we want to minimize

$$\frac{1}{2} \sum_{i=1}^n \mu_i (y_i^{**} - \mathbf{x}'_i \beta)^2 + P_\lambda(\beta),$$

which is equivalent to minimizing (4.9). That is, we can transform the GLMM with penalty

problem into a LMM with penalty problem.

4.3.3.2 Estimation of θ

The adaptive elastic net penalty function $P_\lambda(\boldsymbol{\beta})$ has the singularity at the origin, therefore we consider an approximation of it. Based on the work of Fan and Li (2001), the penalty function $P_\lambda(\boldsymbol{\beta})$ can be approximated by

$$P_\lambda(\boldsymbol{\beta}) \approx \frac{1}{2} \tilde{\boldsymbol{\beta}}_\lambda^T \boldsymbol{\Sigma}_\lambda(\tilde{\boldsymbol{\beta}}) \tilde{\boldsymbol{\beta}}_\lambda,$$

where

- $\tilde{\boldsymbol{\beta}}_\lambda$ only contains nonzero elements $\tilde{\beta}_1, \dots, \tilde{\beta}_m$ of $\tilde{\boldsymbol{\beta}}$.
- $\boldsymbol{\Sigma}_\lambda(\tilde{\boldsymbol{\beta}}) = \text{Diag}\{P'_\lambda(|\tilde{\beta}_1|)/|\tilde{\beta}_1|, \dots, P'_\lambda(|\tilde{\beta}_m|)/|\tilde{\beta}_m|\}$.

Also, define X_L be the matrix corresponding to the nonzero $\tilde{\boldsymbol{\beta}}$'s.

Cui (2011) showed that the approximate REML estimator for θ can be calculated by maximizing

$$l_{\text{R,P}}(\theta|\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}) = -\frac{1}{2} \log |V| - \frac{1}{2} \log |X'_L V^{-1} X_L + \boldsymbol{\Sigma}_\lambda(\tilde{\boldsymbol{\beta}})| - \frac{1}{2} (\mathbf{y}^* - X_L \tilde{\boldsymbol{\beta}})' V^{-1} (\mathbf{y}^* - X_L \tilde{\boldsymbol{\beta}}). \quad (4.11)$$

Compared to (4.7), we have an extra term $\boldsymbol{\Sigma}_\lambda(\tilde{\boldsymbol{\beta}})$ in the logarithm of determinant that is adjusted for the penalty function of $\boldsymbol{\beta}$. An algorithm of the estimation procedure is presented in Appendix 4.B.

4.3.4 Laplace Approximated Loglikelihood with Elastic Net Penalty

In Section 4.3.2, we apply the Laplace approximation to the integrals in the likelihood function of GLMM and obtain the following approximated penalized function (APL)

$$\mathcal{L}_{\text{APL}}(\boldsymbol{\beta}, \boldsymbol{\theta}|\tilde{\mathbf{b}}) = \frac{1}{2} \log |\Sigma_{\boldsymbol{\theta}} W + I_n| - \sum_{i=1}^n (-\mu_i + y_i \eta_i) + \frac{1}{2} \tilde{\mathbf{b}}' \Sigma_{\boldsymbol{\theta}}^{-1} \tilde{\mathbf{b}} + P_{\lambda}(\boldsymbol{\beta}). \quad (4.12)$$

Breslow and Clayton (1993) ignored the first term in (4.12) and yielded the PQL. If we consider the dependency of W on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, we can apply the block coordinate gradient descent (BCGD) method proposed in Tseng and Yun (2009).

The solution of an adaptive elastic net penalty problem can be solved by transforming into a LASSO type of problem. Specifically, minimizing equation (4.12) with respect to $\boldsymbol{\beta}$ is equivalent to minimizing

$$f(\boldsymbol{\beta}|\mathbf{b}, \boldsymbol{\theta}) + \lambda_1 \lambda_2 \sum_j \hat{w}_j |\beta_j|,$$

where

$$f(\boldsymbol{\beta}|\mathbf{b}, \boldsymbol{\theta}) = \frac{1}{2} \log |\Sigma_{\boldsymbol{\theta}} W + I_n| - \sum_{i=1}^n [-\mu_i + y_i \eta_i] + \frac{1}{2} \mathbf{b}' \Sigma_{\boldsymbol{\theta}}^{-1} \mathbf{b} + \lambda_1 (1 - \lambda_2) \sum_j \beta_j^2. \quad (4.13)$$

It is important to note that $f(\boldsymbol{\beta}|\mathbf{b}, \boldsymbol{\theta})$ is a non-convex but differentiable function, and $\sum_j \hat{w}_j |\beta_j|$ is a convex but non-differentiable function. To apply the BCGD algorithm, we update only one component of $\boldsymbol{\beta}$ at one time. For sth component of $\boldsymbol{\beta}$ (denoted as β_s), we first obtain $\tilde{\mathbf{b}}$ based on current estimates $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\theta}}$, then we update the sth component by $\tilde{\beta}_s + d_s$. Here

$$d_s = \text{median} \left\{ \frac{\hat{w}_s \lambda_1 \lambda_2 - f'_s(\boldsymbol{\beta}|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}})}{h_{s,s}}, -\tilde{\beta}_s, \frac{-\hat{w}_s \lambda_1 \lambda_2 - f'_s(\boldsymbol{\beta}|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}})}{h_{s,s}} \right\},$$

where $f'_s(\boldsymbol{\beta}|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}})$ is the sth component of the first derivative of $f(\boldsymbol{\beta}|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}})$ and $h_{s,s}$ is the

sth diagonal element of H . Moreover,

$$f'(\boldsymbol{\beta}|\tilde{\mathbf{b}}, \tilde{\boldsymbol{\theta}}) = X^T(\boldsymbol{\mu} - \mathbf{y}) + 2\lambda_1(1 - \lambda_2)\boldsymbol{\beta} + \frac{1}{2}\text{tr} \left\{ (\Sigma_{\boldsymbol{\theta}}W + I_n)^{-1}\Sigma_{\boldsymbol{\theta}} \frac{\partial W}{\partial \boldsymbol{\beta}} \right\},$$

$$H = X^TWX + 2\lambda_1(1 - \lambda_2)I_p.$$

The estimate of $\boldsymbol{\theta}$ is updated by minimizing equation (4.12) with current estimates of $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\theta}}$. A description of an algorithm for the estimation procedure is in Appendix 4.C.

4.3.5 Tuning Parameter Selection

Popular methods to choose the tuning parameters (λ_1, λ_2) include cross-validation and criterion-based approaches. In our study, we use the Bayesian Information Criterion (BIC) to select the tuning parameter. The calculation of exact loglikelihood for GLMM is complicated, thus the approximated loglikelihood based on Laplace method is used instead. For notation simplicity, we also use $\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \hat{\boldsymbol{\theta}}$ to represent estimates obtained from penalized approximated likelihood. BIC is defined by

$$-2 \left\{ -\frac{1}{2} \log |\Sigma_{\boldsymbol{\theta}}W + I_n| + \sum_{i=1}^n [-\hat{\mu}_i + y_i \hat{\eta}_i - \log(y_i!)] - \frac{1}{2} \hat{\mathbf{b}}' \Sigma_{\boldsymbol{\theta}}^{-1} \hat{\mathbf{b}} \right\} + \log(n)df,$$

where df is the number of nonzero parameters in $\hat{\boldsymbol{\beta}}$ plus the number of parameters in $\hat{\boldsymbol{\theta}}$.

4.4 Simulation Study

In this section, we study the performance of methods proposed in Sections 4.3.3 and 4.3.4 through simulations. PQL and Laplace approximation of the integrals are both commonly used in the framework of GLMM.

4.4.1 Setting

In our simulation study, we consider the following model:

$$y_i|b_i \sim \text{Poisson} [\exp(\mathbf{x}_i^T \boldsymbol{\beta} + b_i)],$$

$$\mathbf{b} \sim \text{N}(\mathbf{0}, \Sigma_{\boldsymbol{\theta}}).$$

We choose the covariance matrix that has the form $(\Sigma_{\boldsymbol{\theta}})_{ij} = \sigma^2 e^{d_{ij}/d}$ ($d > 0$). That is, we consider an exponential correlation function. Each dataset consists of $n = 225$ equal spaced data points that are simulated on a $[1, 10] \times [1, 10]$ regular grid. The distance d_{ij} between data point i and j is the great circle distance. The \mathbf{x}_i 's are randomly simulated from multivariate normal distribution with mean 0 and variance 0.5.

We consider three setting of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$:

(i) $\boldsymbol{\beta} = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$

(ii) $\boldsymbol{\beta} = (0.2, 0.3, 0.4, 0.5, 0.7, 0.8, -0.1, -0.6, -0.9, -1, \mathbf{0}_{10})'$

(iii) $\boldsymbol{\beta} = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{20})'$.

Here $\mathbf{0}_N$ is a vector of zeros with length N . Let ν represents the length of $\boldsymbol{\beta}$. And $\boldsymbol{\theta} = (\sigma^2, d)' = (0.1, 5)', (0.5, 5)'$ or $(0.1, 10)'$.

For the model matrix, we consider the following five cases.

1. all covariates are independent.
2. $\text{Corr}(X_k, X_l) = \omega^{|k-l|}$, $k = 1, \dots, 5, l = 1, \dots, 5$ with $\omega = 0.8$; the other covariates are independent.
3. $\text{Corr}(X_k, X_l) = \omega^{|k-l|}$, $k = 1, 2, 3, l = 1, 2, 3$ with $\omega = 0.8$ and $\text{Corr}(X_4, X_5) = 0.8$; the other covariates are independent.

4. $\text{Corr}(X_k, X_l) = \omega^{|k-l|}$, $k = 1, 2, 3$, $l = 1, 2, 3$ with $\omega = 0.8$ and $\text{Corr}(X_4, X_5) = 0.5$; the other covariates are independent.
5. $\text{Corr}(X_k, X_l) = \omega^{|k-l|}$, $k = 1, \dots, 5$, $l = 1, \dots, 5$ with $\omega = 0.8$; $\text{Corr}(X_k, X_l) = \omega^{|k-l|}$, $k = \nu - 4, \dots, \nu$, $l = \nu - 4, \dots, \nu$ with $\omega = 0.8$; the other covariates are independent.

We consider the following criteria for variable selection performance: (a) *aver.size*: average model size; (b) *corr.coef*: average number of coefficients set to 0 correctly; (c) *mis.coef*: average number of coefficients set to 0 mistakenly.

4.4.2 Results and Discussions

For each case, we simulate 300 datasets and the covariate matrices are all centered and standardized. For simplicity, we assume there is no intercept term in the model. For each simulated dataset, we apply the methods described in Sections 4.3.3 (PQL.elatnet) and 4.3.4 (LP.elatnet) to obtain estimates of parameters and variable selection. We also consider the case of $(\Sigma_{\theta})_{ij} = \sigma^2$, that is the spatial correlation induced by the distance was ignored. BIC is used to select the appropriate tuning parameters. For the weight factor in the adaptive elastic net penalty, we choose $\hat{\beta}_{\text{cnst}}$ to be the maximum likelihood estimate without any penalty on β . And for both methods, we use $r = 1$.

Table 4.1 reports the *aver.size*, *corr.coef* and *mis.coef* for the setting of $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$, $\theta = (0.1, 5)'$. There is no big difference among five cases of model matrix, which suggests that the adaptive elastic net penalty performs well for correlated covariates. Based on the same table, using PQL or Laplace approximated loglikelihood yields similar results. Moreover, considering spatial correlation or ignoring spatial correlation also yields similar results.

Table 4.2 summarizes the results of considering $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$, $\theta = (0.5, 5)'$. In comparing Table 4.1 and Table 4.2, using Laplace approximated loglikelihood provides reasonably good results, while using PQL gives slightly worse results. For the use of PQL, average number of coefficients set to 0 correctly is lower and average model size

Table 4.1: Model selection results based on simulated samples. The parameters are $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$, $\theta = (0.1, 5)'$.

Method	Cases	Consider spatial correlation			Ignore spatial correlation		
		<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>	<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>
True value		5	10	0	5	10	0
PQL.elatnet	Case 1	5.24	9.76	0.00	5.46	9.54	0.00
	Case 2	5.25	9.71	0.04	5.53	9.43	0.04
	Case 3	5.39	9.60	0.00	5.50	9.50	0.00
	Case 4	5.37	9.63	0.00	5.52	9.48	0.00
	Case 5	5.35	9.62	0.03	5.68	9.31	0.01
LP.elatnet	Case 1	5.04	9.96	0.00	5.18	9.82	0.00
	Case 2	5.01	9.88	0.12	5.25	9.72	0.03
	Case 3	5.08	9.92	0.01	5.23	9.76	0.00
	Case 4	5.06	9.94	0.00	5.27	9.73	0.00
	Case 5	5.10	9.81	0.09	5.40	9.58	0.01

is larger as compared to Table 4.2. That is to say, using PQL has larger active set. If σ^2 increases, which means the random effects account for greater proportion of variation in the dependent variable, using PQL tends to include more irrelevant covariates, while the performance of using Laplace approximated loglikelihood is less affected.

Table 4.3 shows the results of increasing d (the spatial correlation is stronger). The performance of adopting PQL or Laplace approximated loglikelihood are both good.

Tables 4.4 and 4.5 show the results of varying number of coefficients. The *mis.coef* in Table 4.4 is larger compared to Table 4.1. In Table 4.4, the value of fixed-effect parameters is changed. Some of the values are quite small (for example, -0.1), and increases the difficulty of picking the correct model. The weak signals are sometimes failed to capture by the algorithms. By looking at Table 4.5, we notice that the variable selection performance is not affected if number of noise variables is increased.

In general, the variable selection methods described in Sections 4.3.3 and 4.3.4 have reasonably good performance for independent or correlated covariates, different settings of fixed-effect and random effect parameters. However, using PQL has less computing time compared to the use of Laplace approximated loglikelihood.

Table 4.2: Model selection results based on simulated samples. The parameters are $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$, $\theta = (0.5, 5)'$.

Method	Cases	Consider spatial correlation			Ignore spatial correlation		
		<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>	<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>
True value		5	10	0	5	10	0
PQL.elatnet	Case 1	5.84	9.16	0.00	5.89	9.11	0.00
	Case 2	5.59	9.39	0.02	6.06	8.89	0.05
	Case 3	6.27	8.73	0.00	5.86	9.09	0.05
	Case 4	6.12	8.88	0.00	6.08	8.88	0.03
	Case 5	5.52	9.45	0.03	5.82	9.10	0.08
LP.elatnet	Case 1	5.05	9.95	0.00	5.35	9.65	0.00
	Case 2	4.92	9.90	0.18	5.46	9.44	0.10
	Case 3	5.05	9.94	0.02	5.29	9.65	0.06
	Case 4	4.99	9.95	0.05	5.35	9.61	0.04
	Case 5	5.02	9.82	0.16	5.55	9.35	0.10

Table 4.3: Model selection results based on simulated samples. The parameters are $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{10})'$, $\theta = (0.1, 10)'$.

Method	Cases	Consider spatial correlation			Ignore spatial correlation		
		<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>	<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>
True value		5	10	0	5	10	0
PQL.elatnet	Case 1	5.19	9.81	0.00	5.42	9.58	0.00
	Case 2	5.23	9.73	0.04	5.70	9.28	0.02
	Case 3	5.40	9.60	0.00	5.54	9.46	0.00
	Case 4	5.24	9.76	0.00	5.60	9.40	0.00
	Case 5	5.35	9.61	0.04	5.71	9.28	0.01
LP.elatnet	Case 1	5.07	9.93	0.00	5.16	9.84	0.00
	Case 2	5.11	9.83	0.07	5.33	9.65	0.02
	Case 3	5.11	9.89	0.00	5.26	9.74	0.00
	Case 4	5.05	9.95	0.00	5.23	9.77	0.00
	Case 5	5.16	9.73	0.10	5.41	9.58	0.01

Table 4.4: Model selection results based on simulated samples. The parameters are $\beta = (0.2, 0.3, 0.4, 0.5, 0.7, 0.8, -0.1, -0.6, -0.9, -1, \mathbf{0}_{10})'$, $\theta = (0.1, 5)'$.

Method	Cases	Consider spatial correlation			Ignore spatial correlation		
		<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>	<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>
True value		10	10	0	10	10	0
PQL.elatnet	Case 1	9.99	9.54	0.47	10.11	9.34	0.55
	Case 2	9.86	9.58	0.56	9.76	9.45	0.79
	Case 3	9.62	9.63	0.75	9.65	9.49	0.86
	Case 4	9.72	9.58	0.70	9.71	9.51	0.78
	Case 5	9.95	9.49	0.56	9.78	9.43	0.78
LP.elatnet	Case 1	9.40	9.86	0.74	9.85	9.58	0.57
	Case 2	9.31	9.88	0.82	9.63	9.62	0.75
	Case 3	9.07	9.89	1.04	9.48	9.63	0.89
	Case 4	9.19	9.84	0.97	9.68	9.59	0.73
	Case 5	9.33	9.81	0.86	9.64	9.60	0.76

Table 4.5: Model selection results based on simulated samples. The parameters are $\beta = (-0.5, 0.75, 1, -0.75, -1, \mathbf{0}_{20})'$, $\theta = (0.1, 5)'$.

Method	Cases	Consider spatial correlation			Ignore spatial correlation		
		<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>	<i>aver.size</i>	<i>corr.coef</i>	<i>mis.coef</i>
True value		5	20	0	5	20	0
PQL.elatnet	Case 1	5.44	19.56	0.00	6.03	18.97	0.00
	Case 2	5.69	19.27	0.04	6.62	18.35	0.03
	Case 3	5.79	19.21	0.00	6.09	18.91	0.00
	Case 4	5.64	19.36	0.00	6.28	18.72	0.00
	Case 5	5.76	19.20	0.03	6.52	18.46	0.02
LP.elatnet	Case 1	5.12	19.88	0.00	5.31	19.69	0.00
	Case 2	5.16	19.74	0.10	5.48	19.48	0.03
	Case 3	5.11	19.88	0.00	5.37	19.63	0.00
	Case 4	5.11	19.88	0.01	5.42	19.58	0.00
	Case 5	5.14	19.70	0.16	5.59	19.39	0.02

4.5 Application to Lyme Disease Data

4.5.1 Data Description

The Lyme disease cases were aggregated into census tracts for a couple of reasons: 1) the demographic information and land cover data are available for each census tract; 2) the tract borders are based on certain features (e.g., rivers, roads) that are potential barriers to the movement of tick or Lyme disease reservoirs (e.g., white-footed mice or deer). The response of interest is the summarization of case counts from 2006 and 2011 in each census tract. The total population counts in each census tract are included into the model as an offset term. Figure 4.1 shows the number of Lyme disease cases and incidence rates (number of cases divided by the size of the population) for each census tract.

The dissimilarities in economic and demographic characteristics may affect the incidence of Lyme disease. Based on that, we select a list of covariates that may contribute to the case counts of Lyme disease. To understand the transmission of Lyme disease, it is important to study the living environment of tick and Lyme disease reservoirs. In past studies, white-footed mice or deer are shown to be very important hosts of ticks. Forested and herbaceous/scrub areas are ideal habitats for white-footed mice or deer. Jackson et al. (2006) and Allan et al. (2003) studied the effect of forest fragmentation on Lyme disease and showed that the percent of forested areas and number of small forest fragments (<2 ha) within each polygon are associated with incidence rate of Lyme disease. In our study, we consider two types of forest fragmentation variables: percent of small forest fragments (<2 ha) and percent of perimeters of the small forest fragments (<2 ha) within each census tract. The percentages of four land cover types (developed land, forest, scrub and water) within each tract are also considered.

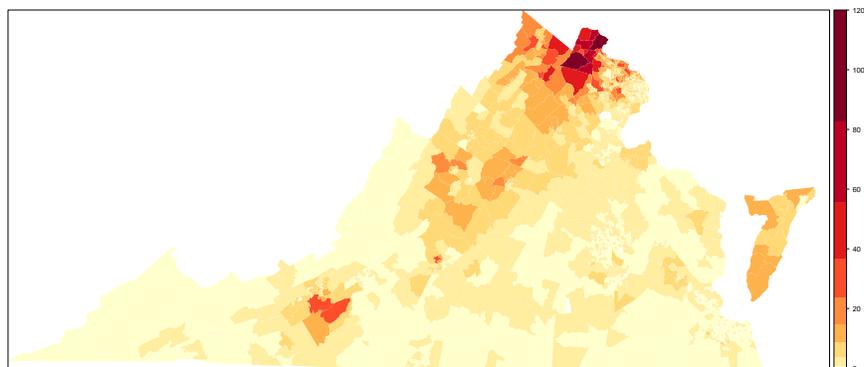
The mixture of land cover types might also be important factors. For example, the boundary between forest and residential areas raises the risk for the interaction between tick or disease reservoirs and human, which may lead to an increase in the incidence rate. We

Table 4.6: Description of covariates in Lyme disease data.

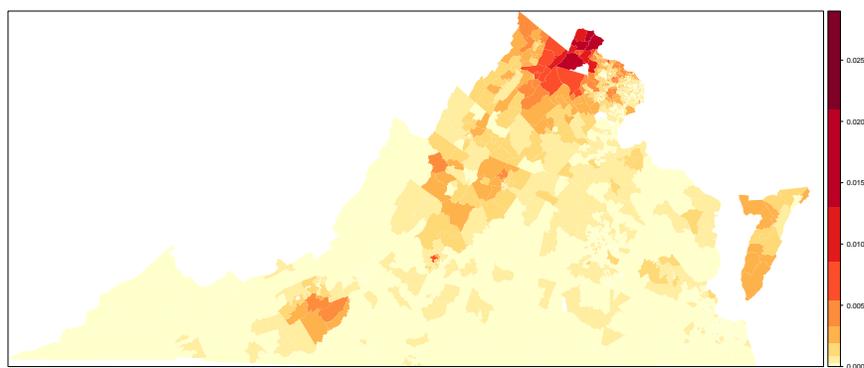
Variable	Description
Dvlpd_NLCD06	Percentage of developed land in each census tract
Forest_NLCD06	Percentage of forest in each census tract
Scrub_NLCD06	Percentage of scrub in each census tract
Tract_Frag06	Sum of area of forested fragments in each census tract divided by the total area
FragPerim06	Sum of forest fragment perimeters in each census tract divided by the total area
CWED_DF06	CWED of developed-forest edge
TECI_DF06	TECI of developed-forest edge
CWED_FS06	CWED of forest-scrub edge
TECI_FS06	TECI of forest-scrub edge
CWED_SD06	CWED of scrub-developed edge
TECI_SD06	TECI of scrub-developed edge
Pop_den	Tract population density in 2010
Median_age	Median age at each census tract in 2010
Mean_income	Mean income (inflation adjusted) at each census tract in 2010
Eco_id	Eco_id = 1 represents the Piedmont, Middle Atlantic Coastal Plain, and Southeastern Plains areas; Eco_id = 0 represents the Northern Piedmont, Blue Ridge, Ridge and Valley and Central Appalachian areas

consider two indexes that characterize the mixture of land cover types: Contrast Weighted Edge Density (CWED) and Total Edge Contrast Index (TECI).

Based on the level III ecoregion map of Virginia (<https://www.hort.purdue.edu/newcrop/cropmap/virginia/maps/VAeco3.html>), Virginia can be divided into two subregions. One subregion consists of Piedmont, Middle Atlantic Coastal Plain, and Southeastern Plains areas, while the other subregion includes the Northern Piedmont, Blue Ridge, Ridge and Valley and Central Appalachian areas. The population density, median age, and mean income in 2010 are also considered as potential factors and included in the study. Summary of selected covariates is in Table 4.6.



(a) Case counts



(b) Incidence rates

Figure 4.1: Number of cases and incidence rates of each census tract in Virginia (2006-2011). (a) Case counts. (b) Incidence rates.

4.5.2 Estimation and Findings

To fit the GLMM to the Lyme disease data, we consider an exponential correlation function. The Northern Piedmont, Blue Ridge, Ridge and Valley and Central Appalachian areas (Eco.id=0) reported larger number of Lyme disease cases than the remaining part. It also appears that the two subregions have different demographic characteristics. Therefore, separate models are fitted for the two subregions. We apply the algorithm described in Section 4.3.3 to the Lyme disease data. Unpenalized estimate is taken to be the weight factor and r is set to be 1. We use BIC as the criterion to select the tuning parameters.

Table 4.7 documents the estimates of the selected covariates as well as the estimates of the covariance parameters. The results show that the factors that affect the Lyme disease case counts are different for two subregions. For the Piedmont, Middle Atlantic Coastal Plain, Southeastern Plains areas, only the percentage of developed land (Dvlpd_NLCD06), and mean income (Mean_income) are selected. Particularly, percent developed has a negative correlation with the number of Lyme disease cases. This is expected as previous studies showed that areas of developed (for example, cities) had a lower risks of having Lyme disease. The mean income was also found out to be a significant variable.

For areas of Northern Piedmont, Blue Ridge, Ridge and Valley and Central Appalachian, the selected variables are percentage of forest (Forest_NLCD06), percentage of scrub (Scrub_NLCD06), forest-scrub edge (CWED_FS06), scrub-developed edge (CWED_SD06) and mean income (Mean_income). Percentage of forest cover and percentage of scrub cover have a positive relationships with Lyme disease case counts, which is in consistent with the findings in Jackson et al. (2006). Forested or scrub areas could provide good living environment for deer and mice. The interaction of human with the hosts leads to a greater risk of having Lyme disease. It is not surprising to find out the forest-scrub edge is important. The mixture of forest and scrub areas is appealing for some host animals. Therefore, the interspersions of forest and scrub land has a positive relationship with Lyme disease incidence. However, the interspersions of scrub and developed areas has a negative correlation with Lyme

disease incidence and the reason behind is not clear.

Our findings are slightly different from the results in Jackson et al. (2006) and Li et al. (2014). Jackson et al. (2006) suggested that forest fragmentation variables are important for Lyme disease incidence. However, in our study, `Tract_Frag06` (percent of forested fragments) and `FragPerim06` (percent of forest fragment perimeters) are not included in the final models for either ecoregions. Li et al. (2014) fitted a spatial model using Lyme disease data without considering the ecoregion variable. We show that two eco-regions have different patterns. In addition, the estimates of covariance parameters are different in two ecoregions. \hat{d} is quite small in the subregion of `Eco.id=1`, which implies that the spatial correlation is weak in that subregion. As for the other subregion (`Eco.id=0`), the estimated d is 33.62. If the distance between two census tracts is 33.62 Kilometre (KM), then the correlation is 0.37. Unlike \hat{d} , the estimated $\hat{\sigma}^2$ in two subregions are close. In Li et al. (2014), population density and median age were significant variables, while population density and median age are not selected in either ecoregions. Moreover, Li et al. (2014) found that percent forest was not significant, while our findings support the results in Jackson et al. (2006) and suggest that percent forest is an important variable.

4.6 Concluding Remarks

In this chapter, we consider the problem of variable selection in GLMM with spatial correlated data. By introducing the adaptive elastic net penalty, we perform variable selection and parameter estimation simultaneously. We consider PQL with penalty (Section 4.3.3) and Laplace approximated loglikelihood with penalty (Section 4.3.4). Simulation study in Section 4.4 shows that both methods perform reasonably good and quite similar. We then apply our method to select important variables associated with the Lyme disease emergence in Virginia.

In this chapter, we use Laplace approximation to the integrals in likelihood functions. Alternatively, Bayesian methods can also be used to approximate integrals. We may use

Table 4.7: Table of coefficients (separate models for two subregions).

Covariate (Eco_id=0) ($n = 583$)	Coef.	Covariate (Eco_id=1) ($n = 1275$)	Coef.
Intercept	0.15	Intercept	-0.39
Dvlpd_NLCD06	0	Dvlpd_NLCD06	-0.34
Forest_NLCD06	0.16	Forest_NLCD06	0
Scrub_NLCD06	0.14	Scrub_NLCD06	0
Tract_Frag06	0	Tract_Frag06	0
FragPerim06	0	FragPerim06	0
CWED_DF06	0	CWED_DF06	0
TECL_DF06	0	TECL_DF06	0
CWED_FS06	0.08	CWED_FS06	0
TECL_FS06	0	TECL_FS06	0
CWED_SD06	-0.01	CWED_SD06	0
TECL_SD06	0	TECL_SD06	0
Pop_den	0	Pop_den	0
Median_age	0	Median_age	0
Mean_income	0.22	Mean_income	0.42
$\hat{\sigma}^2$	0.53	$\hat{\sigma}^2$	0.45
\hat{d} (in KM)	33.62	\hat{d} (in KM)	1.40

Gibbs sampler, Metropolis-Hastings algorithm, MCMC, importance sampling, to name a few. However, this is usually time-consuming.

In this chapter, we consider Poisson regression model with random effect and dispersion parameter ϕ equals to one. If over-dispersion appears in the data, we can add the dispersion parameter into model formulation and obtain estimates of $(\boldsymbol{\beta}, \boldsymbol{\theta}, \phi)$ simultaneously.

In some cases, we may encounter a dataset with large n . Kaufman et al. (2008) developed the covariance tapering method for large irregularly spaced data or missing data on lattice. By taking the inner product of a covariance matrix with a positive definite and compactly supported correlation matrix, one can obtain the “tapered” covariance matrix with sparsity. Future research can be incorporating the covariance tapering method for large n case to achieve computational efficiency.

Appendix

4.A Quadratic Approximation to PQL

Given current estimates of $\boldsymbol{\theta}$ and \mathbf{b} , which are denoted by $\tilde{\boldsymbol{\theta}}$ and $\tilde{\mathbf{b}}$, respectively, (4.5) reduces to

$$l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}}) = \sum_{i=1}^n (-\mu_i + y_i \mathbf{x}'_i \boldsymbol{\beta}), \quad (4.14)$$

up to a constant that is independent of $\boldsymbol{\beta}$.

Next form a quadratic approximation to $l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}})$ around current estimate $\tilde{\boldsymbol{\beta}}$, that is

$$l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}}) \approx l_{\text{PQL}}(\tilde{\boldsymbol{\beta}}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}}) + \frac{\partial l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}})}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \frac{1}{2} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' \frac{\partial^2 l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}),$$

where

$$\begin{aligned} \frac{\partial l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n (-\mu_i \mathbf{x}'_i + y_i \mathbf{x}'_i), \\ \frac{\partial^2 l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= \sum_{i=1}^n (-\mu_i \mathbf{x}_i \mathbf{x}'_i) \end{aligned}$$

are the first and second derivatives for $l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}})$ with respect to $\boldsymbol{\beta}$, respectively.

Therefore,

$$\begin{aligned}
l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}}) &\approx \frac{\partial l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}})}{\partial \boldsymbol{\beta}} \boldsymbol{\beta} + \frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' \frac{\partial^2 l_{\text{PQL}}(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{b}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + c \\
&= \sum_{i=1}^n (y_i - \mu_i) \mathbf{x}'_i \boldsymbol{\beta} + \frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' \left[\sum_{i=1}^n (-\mu_i \mathbf{x}_i \mathbf{x}'_i) \right] (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + c \\
&= -\frac{1}{2} \sum_{i=1}^n \mu_i \left[2 \left(1 - \frac{y_i}{\mu_i} \right) \mathbf{x}'_i \boldsymbol{\beta} + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' \mathbf{x}_i \mathbf{x}'_i (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) \right] + c \\
&= -\frac{1}{2} \sum_{i=1}^n \mu_i \left(\mathbf{x}'_i \tilde{\boldsymbol{\beta}} - 1 + \frac{y_i}{\mu_i} - \mathbf{x}'_i \boldsymbol{\beta} \right)^2 + c \\
&= -\frac{1}{2} \sum_{i=1}^n \mu_i (z_i - \mathbf{x}'_i \boldsymbol{\beta})^2 + c,
\end{aligned}$$

where $z_i = \mathbf{x}'_i \tilde{\boldsymbol{\beta}} - 1 + y_i/\mu_i$ (working response), c is some constant unrelated to $\boldsymbol{\beta}$.

4.B A Description of the Variable Selection Procedure using PQL

In the following, we present an algorithm for the method proposed in Section 4.3.3.

Algorithm 1:

For a collection of values of (λ_1, λ_2) :

1. Initialize $\boldsymbol{\beta}^{(0)}, \mathbf{b}^{(0)}, \boldsymbol{\theta}^{(0)}$.

2. For k th iteration:

(i) Find $\boldsymbol{\beta}, \mathbf{b}$ that maximize (4.5). Define the working response $\mathbf{y}^{*(k)} = \mathbf{x}'_i \tilde{\boldsymbol{\beta}}^{(k-1)} + \tilde{\mathbf{b}}^{(k-1)} + (\mathbf{y} - \tilde{\boldsymbol{\mu}}^{(k-1)})/\tilde{\boldsymbol{\mu}}^{(k-1)}$, and update $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{b}}$ iteratively until convergence.

The estimates obtained are denoted as $\tilde{\boldsymbol{\beta}}^{(k)}, \tilde{\mathbf{b}}^{(k)}$.

(ii) Given the current estimates $\tilde{\boldsymbol{\beta}}^{(k)}, \tilde{\mathbf{b}}^{(k)}$, and $\tilde{\boldsymbol{\theta}}^{(k-1)}$, we perform a quadratic approx-

imation to the PQL and solve

$$\mathcal{L}_{\text{PQL}}^Q(\boldsymbol{\beta}|\tilde{\boldsymbol{\theta}}^{(k-1)}, \tilde{\mathbf{b}}^{(k)}) = \frac{1}{2} \sum_{i=1}^n \mu_i (z_i - \mathbf{x}'_i \boldsymbol{\beta})^2 + P_\lambda(\boldsymbol{\beta}), \quad (4.15)$$

where z_i, μ_i are evaluated at $\tilde{\boldsymbol{\beta}}^{(k)}, \tilde{\mathbf{b}}^{(k)}$, and $\tilde{\boldsymbol{\theta}}^{(k-1)}$. The estimate obtained is denoted by $\tilde{\boldsymbol{\beta}}_Q^{(k)}$.

- (iii) Obtain the estimates of covariance parameters $\tilde{\boldsymbol{\theta}}^{(k)}$ by maximizing (4.11) with $\tilde{\boldsymbol{\beta}}_Q^{(k)}$ and $\tilde{\mathbf{b}}^{(k)}$.

3. Repeat step 2 until convergence. The estimates obtained are denoted by $\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \hat{\boldsymbol{\theta}}$.

4.C A Description of the Variable Selection Procedure using Laplace Approximated Loglikelihood

In this section, we present the block coordinate gradient descent algorithm for estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in (4.12).

Algorithm 2:

For a collection of values (λ_1, λ_2) :

1. Initialize $\boldsymbol{\beta}^{(0)}, \mathbf{b}^{(0)}, \boldsymbol{\theta}^{(0)}$.

2. For k th iteration:

- (i) Repeat for $s = 1, \dots, p$:

For sth component of $\boldsymbol{\beta}$, find $\tilde{\mathbf{b}}^{(k,s)}$ that maximizes $h(\boldsymbol{\beta})$ with given $\tilde{\boldsymbol{\beta}}^{(k,s)} = (\tilde{\boldsymbol{\beta}}_1^{(k)}, \dots, \tilde{\boldsymbol{\beta}}_s^{(k-1)}, \dots, \tilde{\boldsymbol{\beta}}_p^{(k-1)})$ and $\tilde{\boldsymbol{\theta}}^{(k-1)}$, then we obtain the Laplace approximated loglikelihood (4.12). Let $\tilde{\boldsymbol{\beta}}_s^{(k)} = \tilde{\boldsymbol{\beta}}_s^{(k-1)} + d_s$, where d_s is defined in Section 4.3.4.

- (ii) $\tilde{\boldsymbol{\theta}}^{(k)}$ is obtained by minimizing equation (4.12) with $\tilde{\boldsymbol{\beta}}^{(k)}$.

3. Repeat step 2 until convergence. The estimates obtained are denoted by $\hat{\beta}, \hat{b}, \hat{\theta}$.

Bibliography

- B. F. Allan, F. Keesing, and R. S. Ostfeld. Effect of forest fragmentation on Lyme disease risk. *Conservation Biology*, 17(1):267–272, 2003.
- D. Bates. Fitting mixed-effects models using the lme4 package in r. In *International Meeting of the Psychometric Society*, 2008.
- N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.
- B. Cai and D. B. Dunson. Bayesian covariance selection in generalized linear mixed models. *Biometrics*, 62(2):446–457, 2006.
- P. Diggle, R. A. Moyeed, and J. A. Tawn. Model-based geostatistics. *Journal of the Royal Statistical Society, Series C*, 47(3):299–350, 1998.
- Y. Duan. *Statistical Predictions Based on Accelerated Degradation Data and Spatial Count Data*. PhD thesis, Virginia Tech, 2014.
- B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–499, 2004.
- J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.
- J. Fan and J. Lv. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101–148, 2010.
- I. E. Frank and J. H. Friedman. A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135, 1993.

- J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.
- A. Groll. *glmmLasso: Variable selection for generalized linear mixed models by L_1 -penalized estimation*, 2014. <http://CRAN.R-project.org/package=glmmLasso>.
- A. Groll and G. Tutz. Variable selection for generalized linear mixed models by L_1 -penalized estimation. *Statistics and Computing*, 24(2):137–154, 2014.
- A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.
- L. E. Jackson, E. D. Hilborn, and J. C. Thomas. Towards landscape design guidelines for reducing Lyme disease risk. *International Journal of Epidemiology*, 35(2):315–322, 2006.
- C. Kaufman, M. Schervish, and D. Nychka. Covariance tapering for likelihood based estimation in large spatial datasets. *Journal of the American Statistical Association*, 103(484):1545–1555, 2008.
- P. S. Laplace. Memoir on the probability of the causes of events. *Statistical Science*, 1(3):364–378, 1986.
- J. Li, K. N. Kolivras, Y. Hong, Y. Duan, S. E. Seukep, S. P. Prisley, J. B. Campbell, and D. N. Gaines. Spatial and temporal emergence pattern of Lyme disease in Virginia. *The American Journal of Tropical Medicine and Hygiene*, 91(6):1166–1172, 2014.
- C. E. McCulloch, S. R. Searle, and J. M. Neuhaus. *Generalized, Linear and Mixed Models. 2nd Edition*. John Wiley and Sons, New Jersey, 2008.
- R. B. O’Hara and M. J. Sillanpää. A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis*, 4(1):85–117, 2009.
- M. Y. Park and T. Hastie. L_1 -regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society, Series B*, 69(4):659–677, 2007.

- J. Schelldorfer, L. Meier, and P. Bühlmann. *Generalized Linear Mixed Models with Lasso*, 2012. https://r-forge.r-project.org/R/?group_id=984.
- J. Schelldorfer, L. Meier, and P. Bühlmann. Glmmlasso: An algorithm for high-dimensional generalized linear mixed models using L_1 -penalization. *Journal of Computational and Graphical Statistics*, 23(2):460–477, 2014.
- S. E. Seukep, K. N. Kolivras, Y. Hong, J. Li, S. P. Prisley, J. B. Campbell, D. N. Gaines, and R. L. Dymond. An examination of the demographic and environmental variables correlated with Lyme disease emergence in Virginia. *EcoHealth*, 12(4):634–644, 2015.
- R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*, 58(1):267–288, 1996.
- H. Yang. *Variable selection procedures for generalized linear mixed models in longitudinal data analysis*. PhD thesis, North Carolina State University, 2007.
- H. Zhang. On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 58(1):129–136, 2002.
- H. Zou. The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.
- H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67(2):301–320, 2005.
- H. Zou and H. H. Zhang. On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, 37(4):1733–1751, 2009.

Chapter 5 General Conclusions and Areas for Future Work

In Chapter 2, we propose a general method for constructing two-sided simultaneous prediction intervals (SPIs) as well as the corresponding one-sided prediction bounds (SPBs) for at least k out of m ($k \leq m$) future observations. Our study focuses on the (log) location-scale family of distributions. SPI/SPB computed by the proposed method provides exact coverage probability for the case of complete or type II censored data. However, for type I censored data, the coverage probability of the proposed SPI/SPB approaches to the nominal confidence level asymptotically. The method of constructing simultaneous two-sided prediction intervals or one-sided bounds in Chapter 2 can be extended to other censoring schemes.

In Chapter 3, we propose a semi-parametric degradation model to analyze ADDT data. Our degradation path is constructed using monotonic B-splines. Estimation procedure and inference procedure are also developed in this chapter. In the simulation study, we examine the bootstrap confidence interval procedure. We also compare our model to the parametric models. Our simulation results show that the semi-parametric model performs well in the aspects of both fitting and reliability measures. The applications to published ADDT datasets show that the proposed semi-parametric degradation model is flexible and provides good fit. Chapter 3 only considers the scale-accelerated degradation model, which works well in applications. However, the shape acceleration may be appropriate and need to be included in some situations. It would be interesting to consider a shape-scale-accelerated semi-parametric degradation model, and develop estimation and inference procedures. Though challenging, it would also be interesting to develop a formal statistical test that helps to choose the suitable model.

In Chapter 4, we study the variable selection problem in the context of GLMM with spatial count data. For estimation of parameters, we apply the Laplace approximation of integrals and obtain approximated loglikelihood. We also consider PQL, which further

approximates the loglikelihood function by ignoring the dependence of W on parameters. We adopt the adaptive elastic net penalty function for selecting the fixed effects. We then perform a comprehensive simulation study to examine the performance of using Laplace approximated loglikelihood and PQL. Results of the simulation study demonstrate that both methods work well. Last, we apply variable selection techniques to the Lyme disease data and discover factors that affect the Lyme disease case counts. It would be interesting and useful to extend our method for the large sample case.