## DISSERTATION

## BAYESIAN SHAPE-RESTRICTED REGRESSION SPLINES

Submitted by<br>Amber J. Hackstadt<br>Department of Statistics

In partial fulfillment of the requirements for the Degree of Doctor of Philosophy Colorado State University Fort Collins, Colorado

Fall 2011

Doctoral Committee:
Advisor: Jennifer Hoeting
Co-Advisor: Mary Meyer
Jean Opsomer
Kate Huyvaert

Copyright by Amber Hackstadt 2011
All Rights Reserved


#### Abstract

\section*{BAYESIAN SHAPE-RESTRICTED REGRESSION SPLINES}


Semi-parametric and non-parametric function estimation are useful tools to model the relationship between design variables and response variables as well as to make predictions without requiring the assumption of a parametric form for the regression function. Additionally, Bayesian methods have become increasingly popular in statistical analysis since they provide a flexible framework for the construction of complex models and produce a joint posterior distribution for the coefficients that allows for inference through various sampling methods. We use non-parametric function estimation and a Bayesian framework to estimate regression functions with shape restrictions. Shape-restricted functions include functions that are monotonically increasing, monotonically decreasing, convex, concave, and combinations of these restrictions such as increasing and convex. Shape restrictions allow researchers to incorporate knowledge about the relationship between variables into the estimation process. We propose Bayesian semi-parametric models for regression analysis under shape restrictions that use a linear combination of shape-restricted regression splines such as $I$-splines or $C$-splines. We find function estimates using Markov chain Monte Carlo (MCMC) algorithms. The Bayesian framework along with MCMC allows us to perform model selection and produce uncertainty estimates much more easily than in the frequentist paradigm. Indeed, some of the work proposed in this dissertation has not been developed in parallel in the frequentist paradigm.

We begin by proposing a semi-parametric generalized linear model for regression analysis under shape-restrictions. We provide Bayesian shape-restricted regression
spline (Bayes SRRS) models and MCMC estimation algorithms for the normal errors, Bernoulli, and Poisson models. We propose several types of inference that can be performed for the normal errors model as well as examine the asymptotic behavior of the estimates for the normal errors model under the monotone shaperestriction. We also examine the small sample behavior of the proposed Bayes SRRS model estimates via simulation studies. We then extend the semi-parametric Bayesian shape-restricted regression splines to generalized linear mixed models. We provide a MCMC algorithm to estimate functions for the random intercept model with normal errors under the monotone shape restriction. We then further extend the semi-parametric Bayesian shape-restricted regression splines to allow the number and location of the knot points for the regression splines to be random and propose a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm for regression function estimation under the monotone shape restriction. Lastly, we propose a Bayesian shape-restricted regression spline change-point model where the regression function is shape-restricted except at the change-points. We provide RJMCMC algorithms to estimate functions with change-points where the number and location of interior knot points for the regression splines are random. We provide a RJMCMC algorithm to estimate the location of an unknown change-point as well as a RJMCMC algorithm to decide between a model with no change-points and model with a change-point.

## ACKNOWLEDGEMENTS

I would first like to thank my family, especially my parents, Amy and Gene Hackstadt, for their love and encouragement. I really appreciate their help throughout my college career. I would also like to especially thank my husband, Ian Breunig, for his unwavering support and understanding and my sister, Annette Banks, whose work ethic has inspired me throughout my graduate studies.

I would also like thank my committee especially my co-advisors Jennifer Hoeting and Mary Meyer for their support throughout this dissertation. My co-advisors have been great instructors, advisors, and research partners. I would like to thank them not only for their help with my dissertation but for their valuable mentoring throughout these past few years. Furthermore, I would like to thank Ann Hess and the Center for Bioinformatics at Colorado State University for valuable research opportunities in Biostatistics. I would also like to express my gratitude to my fellow graduate students, the faculty, and the staff in the Statistics Department at Colorado State University for their help over the years.

Finally, I gratefully acknowledge the National Science Foundation for support for this work (DMS-0905656).

## DEDICATION

To my family and Ian.

## TABLE OF CONTENTS

1 Introduction ..... 1
1.1 Introduction to Regression Splines ..... 1
1.2 Basis Functions ..... 3
1.3 Function Estimation ..... 4
1.4 Shape-restricted Regression ..... 5
1.4.1 Monotone Shape-restricted Regression Splines ..... 8
1.4.2 Other Shape-restricted Regression Splines ..... 11
1.5 Bayesian Inference and Shape-restricted Regression ..... 13
2 Bayesian Shape-restricted Regression Splines Model ..... 17
2.1 Motivation for Model ..... 17
2.2 Shape-restricted Regression Spline Model ..... 18
2.3 Knot Selection ..... 19
2.4 Normal Errors Model ..... 21
2.4.1 Bayesian Model ..... 21
2.4.2 Additive Regression Model ..... 26
2.4.3 Non-Additive Regression Model ..... 26
2.5 Other Generalized Linear Models ..... 27
2.5.1 Bernoulli Model ..... 28
2.5.2 Poisson Model ..... 31
2.6 Inference ..... 33
2.6.1 Credible and Prediction Intervals ..... 34
2.6.2 Inference using the Posterior Distribution ..... 34
2.6.3 Model Selection ..... 35
2.7 Consistency ..... 36
2.7.1 Background ..... 36
2.7.2 Consistency of the Unrestricted Regression Coefficients ..... 37
2.7.3 Consistency of the Restricted Regression Coefficients ..... 39
2.7.4 Consistency of Precision ..... 42
2.8 Simulation Study ..... 44
2.8.1 Method Comparisons ..... 44
2.8.2 Parallel-curves Regression Model ..... 48
2.8.3 Non-parallel Curves Regression Model ..... 50
2.8.4 Model Selection ..... 52
2.8.4.1 Constant versus Increasing Regression Function ..... 53
2.8.4.2 Linear versus Monotone Regression Function ..... 54
2.8.4.3 Constrained versus Unconstrained Regression Function ..... 55
2.9 Examples ..... 57
2.9.1 Normal Errors Examples ..... 57
2.9.2 Bernoulli Example ..... 59
3 Bayesian Shape-restricted Regression Splines and Mixed Models ..... 62
3.1 Introduction to Generalized Linear Model ..... 62
3.2 Shape-restricted Splines Generalized Additive Mixed Model ..... 65
3.3 Random Intercept Normal Errors Model ..... 66
3.3.1 Bayesian Model ..... 67
3.3.2 Random Intercept Normal Errors Model with Parametrically Modeled Covariates ..... 70
3.3.3 Inference with Mixed Model ..... 72
3.4 Applications ..... 73
3.4.1 Simulated Data ..... 73
3.4.2 CD4 Data ..... 73
4 Bayesian Shape-restricted Regression with Free-knot Splines ..... 78
4.1 Introduction ..... 78
4.2 Free-knot Spline Model ..... 80
4.2.1 Model and Algorithm ..... 80
4.2.2 Priors on Model Parameters ..... 81
4.3 Implementation ..... 82
4.3.1 Birth Move ..... 82
4.3.1.1 Addition of knot ..... 82
4.3.1.2 Acceptance Probability ..... 83
4.3.2 Death Move ..... 85
4.3.3 Relocation Move ..... 86
4.4 Examples ..... 88
4.4.1 Smooth Sigmoid Function ..... 88
4.4.2 Wiggly Piecewise Sigmoid Function ..... 92
5 Bayesian Shape-restricted Regression Splines Model with Change- points ..... 97
5.1 Introduction and Motivation ..... 97
5.2 Background ..... 98
5.3 Multiple Change-point Shape-restricted Regression Spline Model ..... 101
5.3.1 Shape-restricted Spline Approximation ..... 102
5.3.2 Monotonicity Constraint ..... 103
5.3.3 Normal Errors Model ..... 105
5.4 Multiple Change-point Model with Change-point Locations Fixed ..... 107
5.4.1 Overview of the RJMCMC 1 Algorithm for Fixed Change-points ..... 108
5.4.2 Priors for the RJMCMC 1 Algorithm ..... 109
5.4.3 RJMCMC 1 Algorithm Implementation ..... 111
5.4.3.1 Birth Move ..... 111
5.4.3.2 Death Move ..... 115
5.4.3.3 Relocation Move ..... 118
5.4.4 Function Estimation for the RJMCMC 1 Algorithm ..... 122
5.4.5 Examples for the RJMCMC 1 Algorithm ..... 122
5.5 Single Change-point Model with Change-point Location Unknown ..... 126
5.5.1 Overview of the RJMCMC 2 Algorithm ..... 126
5.5.2 Priors for the RJMCMC 2 Algorithm ..... 128
5.5.3 RJMCMC 2 Algorithm Implementation ..... 128
5.5.3.1 Birth, Death, and Relocation Moves for the RJMCMC 2 Algorithm ..... 129
5.5.3.2 Change-point Relocation ..... 131
5.5.4 Function Estimation for the RJMCMC 2 Algorithm ..... 133
5.5.5 Example for the RJMCMC 2 Algorithm ..... 133
5.6 Existence of Change-point Model ..... 135
5.6.1 Overview of the RJMCMC 3 Algorithm ..... 136
5.6.2 Priors for the RJMCMC 3 Algorithm ..... 137
5.6.3 RJMCMC 3 Algorithm Implementation ..... 138
5.6.3.1 Birth, Death, and Relocation Moves for the RJMCMC 3 Algorithm ..... 138
5.6.3.2 Birth of a Change-point Move ..... 139
5.6.3.3 Death of a Change-point Move ..... 143
5.6.4 Function Estimation for the RJMCMC 3 Algorithm ..... 146
5.6.5 Examples for the RJMCMC 3 Algorithm ..... 146
6 Conclusions and Future work ..... 151
6.1 Conclusion ..... 151
6.2 Future Work ..... 154
Bibliography ..... 156
7 References ..... 156

## LIST OF TABLES

2.1 Estimated Mean Squared Error for Simulation Study ..... 48
2.2 Test Size and Power for the Parallel-curves Models with $r=2$ ..... 50
2.3 Test Size and Power for the Parallel-curves Models with $r=3$ ..... 50
2.4 Test Size and Power for the Interaction Model ..... 52
2.5 Constant versus Increasing Simulation Results ..... 54
2.6 Linear versus Increasing Simulation Results ..... 55
2.7 Monotonically Increasing versus Unconstrained Simulation Results ..... 56
4.1 Estimated Mean Squared Error for Fixed versus Free-knot Splines for Smooth Sigmoid Example ..... 94
5.1 RJMCMC 2 Move Probabilities ..... 128
5.2 RJMCMC 3 Move Probabilities when $h=0$ ..... 137
5.3 RJMCMC 3 Move Probabilities when $h=1$ ..... 138

## LIST OF FIGURES

1.1 $B$-spline Basis Functions ..... 4
1.2 Quadratic $I$-spline Basis Functions ..... 9
1.3 Cubic $C$-spline Basis Functions ..... 12
2.1 Sensitivity Plots for Number of Interior Knots ..... 21
2.2 Sample Data Sets from Simulation Study ..... 47
2.3 Credible and Prediction Interval Coverage Probabilities and Interval Lengths for the Simulation Study ..... 49
2.4 Non-parallel Sigmoids Example ..... 52
2.5 Simulated Data Sets from the Monotonically Increasing versus Uncon- strained Simulation Study ..... 57
2.6 Fertility Rate Example ..... 59
2.7 Onion Data Example ..... 60
2.8 Diabetes Example ..... 61
3.1 Simulated Data Example ..... 74
3.2 Plots for $\alpha$ for Simulated Data Example ..... 75
3.3 CD4 Data Set Example ..... 76
3.4 Partial Trace Plots for CD4 Data Set Example ..... 76
3.5 Plots for $\alpha$ for CD4 Data Set Example ..... 77
4.1 Smooth Sigmoid Example ..... 89
4.2 Partial Trace Plots for Smooth Sigmoid Example ..... 91
4.3 Diagnostic Plots for $\alpha$ and $\tau$ for Smooth Sigmoid Example ..... 92
4.4 Fixed versus Free-knot Splines for Smooth Sigmoid Example ..... 93
4.5 Fixed versus Free-knot Splines for Wiggly Sigmoid Example ..... 95
4.6 Partial Trace Plots for Wiggly Sigmoid Example ..... 95
4.7 Diagnostic Plots for $\alpha$ and $\tau$ for Wiggly Sigmoid Example ..... 96
4.8 Wiggly Sigmoid Example ..... 96
5.1 Single Change-point Example ..... 123
5.2 Diagnostic Plots for Single Change-point Example ..... 124
5.3 Two Change-points Example ..... 125
5.4 Diagnostic Plots for Two Change-points Example ..... 127
5.5 Single Change-point with Location Unknown Example ..... 134
5.6 Diagnostic Plots for the Single Change-point with Location Unknown
Example ..... 136
5.7 Plots for the RJMCMC 3 Algorithm for the Single Change-point Example 1485.8 Plots for the RJMCMC 3 Algorithm for the Single Change-point withLarger Error Variance Example149
5.9 Plots for the RJMCMC 3 Algorithm for the No Change-points Example . 150

## Chapter 1

## INTRODUCTION

### 1.1 Introduction to Regression Splines

Non-parametric regression methods have become quite popular recently to examine the relationship between design variables and response variables as well as to make predictions. These non-parametric regression methods include kernel smoothers (Altman, 1992; Wand and Jones, 1995), smoothing splines (Eubank, 1988, Ch 5), and regression splines (Schumaker, 2007; de Boor, 2001; Ramsay, 1988). Nonparametric methods have the benefit of providing estimates of regression functions without requiring the assumption of a parametric form. In addition to providing greater flexibility, they allow local effects to parameter changes and provide a nice framework for imposing shape restrictions (Ramsay, 1988; Meyer, 2008).

Polynomial regression splines or regression splines have properties that make them particularly useful in the regression setting (Schumaker, 2007, Ch 1). A particularly useful property is that a space spanned by polynomial spines is a finite dimensional linear space with bases that are easy to construct and to manipulate in a computer. For a fixed order polynomial spline, every continuous function on an interval can be approximated arbitrarily well by regression splines given a sufficient number of knot points. Therefore, polynomial splines are flexible enough to approximate any function well. Compared to smoothing splines, estimation using regression splines is often preferred because they involve fewer knots and therefore involves the estimation of fewer parameters than smoothing splines (Ramsay, 1988).

Consider a function of a covariate $x$ where the covariate takes on values in $[L, U]$. Estimation of the function using regression splines involves fitting piecewise
polynomials of degree $d$ over the interval $[L, U]$. Given knot points $\boldsymbol{t}=\left(t_{1}, \ldots, t_{q}\right)$ with $L=t_{1}<\ldots<t_{q}=U$, polynomials of at most degree $d$ are fit within the sub-intervals $\left[t_{i}, t_{i+1}\right)$ for $i=1, \ldots q-1$ and continuity constraints at the interior knot points require the polynomials to join with a given degree of smoothness. The continuity constraints are imposed by requiring the derivative of the splines up to $d-1$ to be equal at the interior knot points. Thus, let $p_{d, i}(x)$ be the polynomial between $t_{i}$ and $t_{i+1}$ for the regression spline function of degree $d$ and let $p_{d, i}^{(l)}(x)$ be the $(l)$ th derivative evaluated at $x$. The regression spline function satisfies

$$
\begin{equation*}
p_{d, i-1}^{(l)}\left(t_{i}\right)=p_{d, i}^{(l)}\left(t_{i}\right) \tag{1.1}
\end{equation*}
$$

for $l=1, \ldots, d-1$ and $i=2, \ldots, q-1$ (Dierckx, 1993, Ch 1). Function estimation using regression splines is examined further throughout this chapter.

Regression splines provide a framework that facilitates shape-restricted regression. A useful property of regression splines which is exploited for shaperestricted regression is that a shape-restriction can be enforced by restricting the coefficients. For example, a spline function estimate created using a linear combination of quadratic $I$-spline basis functions (defined in Section 1.4.1) is monotonically increasing if and only if the coefficients on the quadratic $I$-spline basis functions are positive. Thus, compared with kernel smoothers, regression splines produce a framework that makes imposing shape restrictions easier. Using that the derivatives of a polynomial spline are also polynomial splines and can be computed, one can derive asymptotic properties for polynomial splines when estimating smooth functions which we will use to demonstrate consistency of shape-restricted polynomial splines.

We begin this dissertation with an introduction to regression splines. We explain how to construct regression splines using a linear combination of basis functions. We describe how to construct several different basis functions and estimate the linear coefficients. Next, we introduce shape-restricted regression and review
the existing literature on this topic. We describe basis functions for shape-restricted function estimation. We conclude the first chapter with a literature review on non-parametric and semi-parametric shape-restriction regression estimation in the Bayesian paradigm. This will provide background information for the Bayesian shape-restricted regression spline model proposed in Chapter 2.

### 1.2 Basis Functions

Regression splines can be constructed using several different types of basis functions. Given knot points $\boldsymbol{t}$, regression splines are easily constructed using a linear combination of basis functions that span the space of piecewise polynomials of degree $d$. $B$-Splines (Eubank, 1988, Ch 7) basis functions are commonly used to construct regression splines because they are rather straightforward to construct and recursively defined. Figure 1.1 gives the $B$-spline basis functions for a variable with range $[0,1]$ and four equally spaced interior knots. Degree $1 B$-splines are given in Figure 1.1(a) and degree $3 B$-splines in Figure 1.1(b). For observed data $\left(x_{i}, y_{i}\right)$ for $i=1, \ldots, n$, a $B$-spline of degree $d$ and order $q=d+1$ is found by first dividing the range of the covariate, $[L, U]$, into sub-intervals using knot points $\boldsymbol{t}=\left(t_{1}=U, \ldots, t_{k+2}=L\right)$ where $k$ is the number of user-specified interior knots. Define mesh points $\boldsymbol{\xi}=\left(\xi_{1}, \ldots, \xi_{k+2 d}\right)^{\prime}$ to be the vector of knot points with additional knots created at the endpoints to allow for a recursive formula to construct the $B$-splines. These mesh points are given by $L=\xi_{1}=\ldots=\xi_{d}=t_{1}<\xi_{d+1}=$ $t_{2}<\ldots<\xi_{d+k}=t_{k+1}<\xi_{d+k+1}=\ldots=\xi_{d+2 k}=t_{k+2}=U$. Using these mesh points, $j=1, \ldots, k+q B$-splines of order $q=d+1$, denoted by $B_{j}^{q}$, are recursively found using

$$
\begin{align*}
B_{j}^{1}(x) & = \begin{cases}1 & \xi_{j} \leq x<\xi_{j+1} \\
0 & \text { otherwise }\end{cases} \\
B_{j}^{q}(x) & =\frac{x-\xi_{j}}{\xi_{j+q-1}-\xi_{j}} B_{j}^{q-1}+\frac{\xi_{j+q}-x}{\xi_{j+q}-\xi_{j+1}} B_{j+1}^{q-1} \text { for } q>1 \tag{1.2}
\end{align*}
$$

(Eubank, 1988, Ch 7). Other basis functions include the truncated power function (Eubank, 1988, Ch 5) and $M$-splines (Curry and Schoenberg, 1966).
$M$-splines of order $q$, denoted $M_{j}^{q}$, can be integrated to create basis functions for shape-restricted regression splines (Ramsay, 1988) (discussed further in Section 1.4) and are related to $B$-splines. Consider a covariate with range $[L, U]$ and the knot and mesh points as defined above for the $B$-splines. The $j=1, \ldots, k+q M$-splines of order $q$ are found recursively by

$$
\begin{align*}
& M_{j}^{1}(x)= \begin{cases}\frac{1}{\xi_{j+1}-\xi_{j}} & \xi_{j} \leq x<\xi_{j+1} \\
0 & \text { otherwise }\end{cases} \\
& M_{j}^{q}(x)=\left\{\begin{array}{ll}
\frac{q\left[\left(x-\xi_{j}\right) M_{j}^{q-1}(x)+\left(\xi_{j+q}-x\right) M_{j+1}^{q-1}(x)\right]}{(q-1)\left(\xi_{j+q}-\xi_{j}\right)} & \xi_{j} \leq x<\xi_{j+q} \\
0 & \text { otherwise }
\end{array} \text { for } q>1 .\right. \tag{1.3}
\end{align*}
$$

Note the $M$-splines of order $q$ are related to $B$-splines of order $q$ by the identity

$$
\begin{equation*}
B_{j}^{q}(x)=\frac{\left(\xi_{j+q}-\xi_{j}\right)}{q} M_{j}^{q}(x) \tag{1.4}
\end{equation*}
$$

(Ramsay, 1988).


Figure 1.1: Plots of $B$-spline basis functions. The " X " and vertical dotted lines represent the knot points. (a) Degree $1 B$-splines. (b)Degree $3 B$-splines.

### 1.3 Function Estimation

Once an appropriate set of basis functions has been constructed, we can create a spline function estimate of the regression function, $f$. For a given observed data
set $\left(x_{i}, y_{i}\right)$ for $i=1, \ldots, n$, we use a spline function, denoted by $s(x)$, of degree $d$ to estimate regression function $f$ over the range of the covariate $x,[L, U]$. Given knot points $\boldsymbol{t}=\left(t_{1}, \ldots, t_{k+2}\right)$, the spline function estimate is

$$
s(x)=\sum_{j} \beta b_{j}(x)
$$

where $b_{j}(x)$ is the $(j)$ th basis function. For each $j$, we evaluate $b_{j}$ at each $x_{i}$ to give basis vector $\boldsymbol{\delta}_{j}$. To obtain the spline function estimate, we use the method of least squares and a design matrix constructed from the basis vectors.

Consider the univariate case and observed data set $\left(x_{i}, y_{i}\right)$ with $i=1, \ldots, n$. Suppose

$$
\begin{equation*}
y_{i}=f\left(x_{i}\right)+\sigma \epsilon_{i}, \tag{1.5}
\end{equation*}
$$

where $\epsilon_{i}$ for $i=1, \ldots n$ are independent and identically distributed (i.i.d.) random errors. Using $B$-spline basis functions, the regression spline estimate for $f$ is given by

$$
\begin{equation*}
\hat{f}\left(x_{i}\right)=\sum_{j=1}^{d+k} \hat{\beta}_{j} \boldsymbol{B}_{j}, \tag{1.6}
\end{equation*}
$$

where $\boldsymbol{B}_{j}=\left(B_{j}^{q}\left(x_{1}\right), \ldots, B_{j}^{q}\left(x_{n}\right)\right)^{\prime}$ and $B_{j}^{q}\left(x_{i}\right)$ is as defined in (1.2). An estimate for the regression coefficient vector, $\boldsymbol{\beta}=\left(\beta_{1}, \ldots, \beta_{d+k}\right)^{T}$ is found by minimizing

$$
\begin{equation*}
\sum_{i=1}^{n}\left(y_{i}-\sum_{j=1}^{d+k} \beta_{j} \boldsymbol{B}_{j}\right)^{2} \tag{1.7}
\end{equation*}
$$

Thus, the estimates can be computed by $\hat{\boldsymbol{\beta}}=\left(B^{\prime} B\right)^{-1} B^{\prime} \boldsymbol{y}$ where $B=\left[B_{1} \ldots B_{d+k}\right]$ is the matrix of basis vectors and $B^{\prime}$ denotes the transpose of matrix $B$ (Dierckx, 1993, Ch 4). As we will explain in the following section, this estimation procedure can be extended to shape-restricted regression splines.

### 1.4 Shape-restricted Regression

We will now focus on non-parametric and semi-parametric function estimation under shape restrictions. Shape-restricted functions include functions that are
monotonically increasing, monotonically decreasing, convex, concave, and combinations of these restrictions such as increasing and convex. Shape restrictions allow researchers to incorporate knowledge about the relationship between variables into the estimation process. For example, we would like the regression function of a growth curve to be non-decreasing and want to avoid function estimates that incorrectly show the function as decreasing for some values of the predictor. Shape restrictions can be imposed when estimating functions using non-parametric estimators such as kernel smoothers (Hall and Huang, 2001; Mammen, 1991), smoothing splines (Mammen and Thomas-Agnan, 1999), and regression splines (Meyer, 2008; Ramsay, 1988).

Extensive research has been done on non-parametric function estimation under shape restrictions and several different non-parametric and semi-parametric shaperestricted regression estimators have been proposed. Mammen (1991) combined kernel smoothing and isotonic (monotonically increasing) regression in a two-step procedure, and showed that they have the same convergence rate as for the unconstrained kernel estimator. Delecroix et al. (1995) examined non-parametric function estimation using kernel smoothers and demonstrated through simulations that imposing shape restrictions on top of smoothing leads to substantial reductions of squared error loss for moderately sized samples and many choices of underlying function and error variance (when the shape assumptions are indeed correct). Hall and Huang (2001) provided a method to make a kernel smoother monotone and smooth.

Several other shape-restricted function estimation procedures have been proposed using smoothing splines as opposed to kernel smoothers. Smoothing splines are a penalized least-squares method involving polynomial splines where a penalty on roughness is imposed (Wang and Li, 2008; Mammen and Thomas-Agnan, 1999). Kelly and Rice (1990) used smoothing splines restricted to be monotone to estimate dose-response curves. Mammen and Thomas-Agnan (1999) considered imposing
shape restrictions on smoothing splines and showed that, under certain conditions, shape-restricted smoothing splines have the same convergence rate as unconstrained regression splines. They provided rates of convergence for shape-restricted smoothing splines under various conditions, and proved that the convergence rates of the smoothing splines are optimal for appropriate choices of the smoothing parameter. Mammen et al. (2001) proposed a projection framework for constrained smoothing for which a smoothed fit is projected onto a constraint set. Wang and Li (2008) considered the monotone shape restriction and smoothing splines created using natural cubic splines. The function estimates were found by minimizing a linear objective function over the intersection of an affine set. Turlach (2005) provided an algorithm for constrained spline smoothing for several different shape restrictions including monotonicity and concavity that involves adaptively choosing to replace an infinite number of constraints with a finite number.

Shape-restricted regression splines have also been studied by many researchers and is the focus of this dissertation. Monotone shape-restricted regression splines were discussed by Ramsay (1988). His work was later extended to include other shape restrictions such as convexity and concavity (Meyer, 2008). Shape-restricted regression splines provide flexible as well as smooth function estimates and are more robust to knot selection than unconstrained splines (Meyer, 2008). Huang and Stone (2002) established the consistency and convergence rate for unrestricted polynomial splines and Meyer (2008) showed that the convergence rate for shape-restricted regression splines is at least as good as the convergence rate for the unconstrained polynomial splines. However, several other smooth shape-restricted estimators are found in the literature. Tutz and Leitenstorfer (2007) applied regression spline techniques to the generalized regression model under the monotone shape restriction, using $I$-spline basis functions, "boosting" techniques and a stopping rule based on AIC model selection. Dilleen et al. (2003) proposed a monotone dose-response curve using
least squares estimation procedures and $B$-splines along with a bias term adjustment. Yatchew and Hardle (2006) used a non-parametric regression spline function estimation procedure and constrained least squares for state price density estimation under the monotone or convex shape restriction. Banerjee et al. (2009) proposed a semi-parametric model for binomial data under the monotone shape restriction using regression splines and maximum likelihood estimation along with likelihood ratiobased inferential procedures. Our work focuses on using shape-restricted regression splines (SRRS) to produce flexible estimates of smooth functions. Imposing shape restrictions on regression splines produces more flexible fits than assuming a parametric form and is less computationally intensive than imposing shape restrictions on smoothing splines or kernel smoothers.

### 1.4.1 Monotone Shape-restricted Regression Splines

Monotone shape-restricted function estimates can be found using a linear combination of $B$-spline basis functions defined in (1.2) and requiring the coefficients on the appropriately scaled $m$ basis vectors to be ordered as in Kelly and Rice (1990). One can also estimate monotone functions using a linear combination of integrated splines or $I$-splines of degree 2 introduced by Ramsay (1988) and requiring the coefficients of the linear combination of these basis functions to be non-negative (Ramsay, 1988; Meyer, 2008; Tutz and Leitenstorfer, 2007). The $I$-splines are similar to $B$-splines but are found by integrating the $M$-splines. In particular, $I$-splines of order $q+1$ are found by integrating $M$-splines of order $q$ where $M$-splines are as defined in (1.3). Thus, an $I$-spline of order $q+1$ is given by

$$
\begin{equation*}
I_{j}^{q+1}(x)=\int_{L}^{x} M_{j}^{q}(u) d u \tag{1.8}
\end{equation*}
$$

where $j=1, \ldots, q+k, x \in[L, U], q$ is the order of the $M$-splines, and $k$ is the number of interior knots. The degree 2 (quadratic) $I$-splines have the property that along with the constant function they span the space of piecewise quadratic functions and
any linear combination of the $I$-splines and the constant function is non-decreasing if and only if the coefficients on the $I$-splines are non-negative (Meyer, 2008). This is not the case for degree 3 (cubic) $I$-splines since a linear combination of cubic $I$-splines might be non-decreasing if one or more of the coefficients is negative (Meyer, 2008). For this reason, quadratic $I$-splines will be used in the estimation of monotone regression. Examples and simulations in Chapter 2 suggest that they are sufficiently flexible enough to estimate many types of regression functions well. Figure 1.2 gives the basis functions (solid lines) excluding the one vector for the monotone shape restriction for 30 equally spaced $x$-values over $[0,1]$ using four interior knots placed at equal quantiles with basis vectors denoted by points and knot points denoted by " X ".


Figure 1.2: Quadratic basis functions for monotone shape-restricted regression splines. The "X" and vertical dotted lines represent the knot points. The lines represent the basis functions and the dots represent the basis vectors.

For the majority of the analyses in this dissertation, the basis functions and basis vectors (excluding the one vector) are scaled such that they are orthogonal to the one vector and have a range of one. Consider the univariate data set $\left(x_{i}, y_{i}\right)$ with
$i=1, \ldots, n$. Given the knot locations, the basis vectors for the monotone shape restriction are constructed by first defining a set of vectors $\left\{\boldsymbol{\sigma}_{1}, \ldots, \boldsymbol{\sigma}_{d+k}\right\}$ where $\boldsymbol{\sigma}_{j}=\left(\sigma_{j 1}, \ldots, \sigma_{j n}\right)^{T}$ for $j=1, \ldots, d+k$ and $\sigma_{j i}=I_{j}^{3}\left(x_{i}\right)$. Next, define $V$ as the linear space contained in the constraint set so for the monotone shape restriction $V=\mathcal{L}(\mathbf{1})$ where $\mathbf{1}$ is a $n \times 1$ vector of ones and $\mathcal{L}$ denotes the "linear space spanned by." The basis vectors $\boldsymbol{\delta}_{j}, j=1, \ldots, d+k$ are then created by $\boldsymbol{\delta}_{j}=\boldsymbol{\sigma}_{j}-P\left(\boldsymbol{\sigma}_{j} \mid V\right)$, where $P$ is the projection operator. For monotone regression splines

$$
P\left(\boldsymbol{\sigma}_{j} \mid V\right)=P\left(\boldsymbol{\sigma}_{j} \mid \mathbf{1}\right)=\frac{1}{n} \sum_{i=1}^{n} \sigma_{j i} .
$$

Therefore, under the monotone shape restriction we compute the basis vectors $\boldsymbol{\delta}_{j}$ with elements $\delta_{j i}$ where

$$
\begin{equation*}
\delta_{j i}=I_{j}^{3}\left(x_{i}\right)-\frac{1}{n} \sum_{i=1}^{n} I_{j}^{3}\left(x_{i}\right) . \tag{1.9}
\end{equation*}
$$

As with unconstrained regression splines, the shape-restriction regression spline estimate for univariate data set $\left(x_{i}, y_{i}\right)$ with $i=1, \ldots, n$ can be found using least squares estimation. Thus, it is found by minimizing

$$
\begin{equation*}
\sum_{i=1}^{n}\left(y_{i}-\eta_{i}\right)^{2} \tag{1.10}
\end{equation*}
$$

with respect to $\boldsymbol{\eta}=\left(\eta_{1}, \ldots, \eta_{n}\right)^{\prime}$ over a constraint set. The constraint set for the monotonically increasing shape restriction is

$$
\begin{equation*}
\mathcal{C}=\left\{\boldsymbol{\eta}: \boldsymbol{\eta}=\sum_{l=1}^{L} \alpha_{l} \boldsymbol{v}_{l}+\sum_{j=1}^{d+k} \beta_{j} \boldsymbol{\delta}_{j}, \text { and } \beta_{j} \geq 0 \text { for } j=1, \ldots, d+k\right\} . \tag{1.11}
\end{equation*}
$$

where $\boldsymbol{v}_{l} \in V$ for $l=1, \ldots, L$. To obtain the constraint set for the monotonically decreasing shape restriction, replace the condition $\beta_{j} \geq 0$ with $\beta_{j} \leq 0$. Ramsay (1988) used the gradient projection algorithm to obtain the estimates of the linear coefficients. Meyer (1999) extended this by noting that the constraint set for the regression coefficients is a closed convex polyhedral cone in $\mathbb{R}^{n}$ and found coefficient estimates using the hinge algorithm. This procedure projects the $\boldsymbol{y}=\left(y_{1}, \ldots, y_{n}\right)^{\prime}$
onto a linear subspaces and exploits the property of convex cones given in Proposition 1 in Meyer (2008).

Here we summarize the estimation procedure. Let $\Omega$ be the set of all possible non-negative linear combinations of $\boldsymbol{\delta}_{j}, j=1, \ldots, m=d+k$ where $\boldsymbol{\delta}_{j}$ are basis vectors. Define a "face" by $\mathcal{F}(J)=\left\{\sum_{j \in J} \beta_{j} \boldsymbol{\delta}_{j}: \beta_{j} \geq 0\right.$ and $\left.j \in J\right\}$ where $J \subseteq\{1,2, \ldots, m\}$ and there are $2^{m}$ faces. The projection onto $\Omega$ that minimizes condition (1.10) can be found by using one of the faces of the constraint cone. Using Proposition 1 from Meyer (2008), we know that the projection onto the cone is equal to the projection onto the linear space coinciding with $\mathcal{F}(J)$ for some subset $J$ where the linear space coinciding with $\mathcal{F}(J)$ is spanned by $\mathcal{D}(J)=\left\{\delta_{j}: j \in J\right\}$. Using the mixed primal dual algorithm (Fraser and Massam, 1989) or the hinge algorithm (Meyer, 1996), we can efficiently determine which face, $\mathcal{F}(J)$, the projection lies and then project onto the linear subspace spanned by $\mathcal{D}(J)$ to obtain the function estimate. Thus, function estimate, $\hat{\boldsymbol{\eta}}$, is found by projecting onto the space spanned by $\mathcal{D}(J)$ and $V$ separately and then adding the projections. For instance, suppose we are estimating a regression spline assuming a monotone increasing shape restriction. Suppose we used the hinge algorithm to determine face indexed by $J^{*}$ is where the minimizer lies. Let $N_{J^{*}}$ be the matrix whose columns are $\boldsymbol{\delta}_{j}$ where $j \in J^{*}$ and let $\boldsymbol{\eta}^{*}$ be the projection onto the linear space spanned by columns of $N_{J^{*}}$. Therefore, $\boldsymbol{\eta}^{*}=N_{J^{*}}\left(N_{J^{*}}^{\prime} N_{J^{*}}\right)^{-1} N_{J^{*}}^{\prime} \boldsymbol{y}$. Let $\tilde{\boldsymbol{\eta}}$ be the linear projection onto $V$. For the monotonically increasing shape restriction, $V$ is spanned by $\boldsymbol{v}_{1}=\mathbf{1}$ so $\tilde{\boldsymbol{\eta}}=\boldsymbol{v}_{1}\left(\boldsymbol{v}_{1}^{\prime} \boldsymbol{v}_{1}\right)^{-1} \boldsymbol{v}_{1}^{\prime} \boldsymbol{y}$. Hence, the estimate for the regression function is then $\hat{\boldsymbol{\eta}}=\boldsymbol{\eta}^{*}+\tilde{\boldsymbol{\eta}}$.

### 1.4.2 Other Shape-restricted Regression Splines

The convex and concave shape restrictions can be imposed in a similar fashion using $C$-spline basis functions (Meyer, 2008) which are integrated $I$-splines. The
$C$-splines of degree $q+1$ and order $q+2$ are constructed by integrating $I$-splines of order $q+1$ and are given by

$$
\begin{equation*}
C_{j}^{q+2}(x)=\int_{L}^{x} I_{j}^{q+1}(u) d u \tag{1.12}
\end{equation*}
$$

for $i=1, \ldots, m=q+k$ and $x \in[L, U]$. Similar to the monotone shape-restricted regression splines, define set of vectors $\left\{\boldsymbol{\sigma}_{1}, \ldots, \boldsymbol{\sigma}_{m}\right\}$ where $\boldsymbol{\sigma}_{j}=\left(\sigma_{j 1}, \ldots, \sigma_{j n}\right)^{T}$ for $j=1, \ldots, m$ and $\sigma_{j i}=C_{j}^{q+2}\left(x_{i}\right)$. Again, let $V$ be the linear space contained in the constraint set. For the convex or concave shape restriction let $V=\mathcal{L}(\mathbf{1}, \boldsymbol{x})$. As with the $I$-splines, the basis vectors are then found by projecting onto $V$ and are given by $\boldsymbol{\delta}_{j}=\boldsymbol{\sigma}_{j}-P\left(\boldsymbol{\sigma}_{j} \mid V\right)$. The basis functions for the concave shape restrictions over a range of $[0,1]$ with four equally spaced interior knots are given in Figure 1.3.


Figure 1.3: Cubic basis functions for convex shape-restricted regression splines. The " X " and vertical dotted lines represent the knot points. The lines represent the basis functions and the dots represent the basis vectors.

Cubic $C$-splines have the property that a linear combination of them, the constant function and the identity function is convex if and only if the coefficients on the cubic $C$-splines are non-negative (Meyer, 2008). A convex regression function
can be estimated using a linear combination of the basis vectors created from cubic $C$-splines where the linear coefficients are restricted to be non-negative plus a linear combination of the one vector and identity vector with unrestricted coefficients. Similarly, a concave regression function estimate is created using a linear combination of the basis vectors created from cubic $C$-splines where the linear coefficients are restricted to be non-positive plus a linear combination of the one vector and identity vector with unrestricted coefficients. Thus, the estimate for the regression spline can be obtained by minimizing (1.10) over constraint set in (1.11) with $\boldsymbol{\delta}_{\boldsymbol{j}}$ equal to the $C$-spline basis vectors and $l=2$ with $\boldsymbol{v}_{1}$ equal to the one vector of length $n$ and $\boldsymbol{v}_{2}$ an identity vector of length $n$ with elements $v_{2 i}=x_{i}$.

Estimates for regression functions under a combination of restrictions can be found similarly. For instance, to impose the restriction of both an increasing and convex function, we estimate the regression function using a linear combination of cubic $C$-spline basis vectors plus the one vector and identity vector and restrict the coefficients of both the cubic $C$-spline basis vectors and the identity vector to be positive. Likewise, a decreasing concave function can be estimated using a linear combination of $C$-spline basis vectors plus the one vector and identity vector and requiring the coefficients of the $C$-spline basis vectors and the coefficient for the identity vector to be negative. The estimates for these functions can then be found using the same minimization techniques as for the monotone and convex or concave shape restrictions.

### 1.5 Bayesian Inference and Shape-restricted Regression

Non-parametric and semi-parametric regression estimation can be modeled using a Bayesian paradigm. Bayesian methods provide a joint posterior distribution for the coefficients and hence allow for inference through various sampling methods. A number of methods for non-parametric and semi-parametric function estimation in a Bayesian framework both with and without shape restrictions have been proposed.

Silverman (1985) discussed a Bayesian approach to function estimation using $B$-splines. Smith and Kohn (1996) used piecewise cubic polynomials and a Bayesian framework to estimate additive regression models with error modeled using a mixture of normal distributions. Zhao (2000) considered function estimation using Bayesian methods and regression splines and found priors that attain the optimal minimax convergence rate. Berry et al. (2002) provided a Bayesian model for function estimation using smoothing splines and penalized regression splines known as $P$-splines. Lang and Brezger (2004) proposed a model that also uses $P$-splines and allowed the smoothing parameters to be locally adaptive. They imposed normal priors on the basis function coefficients, a gamma prior on the precision and used the penalized likelihood. Muller and Quintana (2004) reviewed many non-parametric Bayesian inference models including regression function estimation models.

Non-parametric and semi-parametric Bayesian shape-restricted regression models, especially under the monotone shape restriction, have been studied by several researchers. Lavine and Mockus (1995) introduced a Bayesian monotone regression approach using Dirichlet process priors. Ramgopal et al. (1993) estimated a monotone dose-response curve, also using a Dirichlet process prior. Perron and Mengersen (2001) proposed a mixture of triangular distributions where the dimension is estimated as part of the Bayesian analysis. Holmes and Heard (2003) used a piecewise constant model with the number and location of the jump points are modeled as random. Neelon and Dunson (2004) proposed a piecewise linear model where the monotonicity is enforced via prior distributions on the model parameter. Their model allowed for flat spots in the regression function by using a prior that is mixture of a continuous distribution and point mass at the origin. Dunson (2005) considered modeling count data using a Bayesian semi-parametric model. He modeled the mean of the process using isotonic function $f$ assigning independent gamma densities to increments on $f$ according to a gamma process. Wang and Li (2008) estimated monotone functions using cubic splines by minimizing a linear
objective function over the intersection of an affine set. Wang (2008) extended the method to a Bayesian framework in which the knot locations are free parameters (free-knot splines). Bornkamp and Ickstadt (2009) applied a Bayesian monotonic regression model to dose-response curves. Their regression function was a mixture of parametric probability distributions with a general discrete random probability measure as the prior for the mixing distribution. Johnson (2007) estimated item response functions with free-knot splines restricted to be monotone by requiring spline coefficients to monotonically increasing and using truncated normal priors. Cai and Dunson (2007) proposed a Bayesian model for multi-site tumor data using multivariate smoothing splines and a hierarchical Markov random field prior for the spline coefficients. Brezger and Steiner (2008) proposed a monotone regression model that used linear inequality constraints along with truncated normal priors on the basis function coefficients to ensure monotonicity. Shively et al. (2009) proposed two Bayesian approaches to non-parametric monotone function estimation with one involving piecewise linear approximation and a Weiner process prior and the other involving regression spline estimation and a prior that is a mixture distribution of constrained normal distributions for the regression coefficients. They discussed the asymptotic properties of shape-restricted regression splines under the monotone shape restriction for a class of Bayesian estimators. Curtis and Ghosh (2011) proposed a Bayesian model for function estimation under the monotone, convex, or concave shape restriction using Bernstein polynomials. Shively et al. (2011) used fixed and free-knot quadratic regression splines in a Bayesian context for non-parametric function estimation subject to shape restrictions. However, the shape-restrictions are imposed by requiring sums of spline coefficients to be positive as opposed to just requiring the coefficients themselves to be positive as with shape-restricted estimation using quadratic $I$-splines or cubic $C$-splines. In the following chapter, we extend the previous work on shape-restricted regression splines
and introduce a new Bayesian model for shape-restricted regression using $I$-splines and $C$-splines .

## Chapter 2

## BAYESIAN SHAPE-RESTRICTED REGRESSION SPLINES MODEL

### 2.1 Motivation for Model

In this chapter, we propose models that use the shape-restricted regression splines of Meyer (2008) within a Bayesian framework for function estimation for generalized linear models. The Bayesian paradigm allows for a wider range of inferences and thus expands the usefulness of the models. By using $I$-splines or $C$ splines, the shape-restrictions are imposed simply by requiring the coefficients on the spline basis functions to be non-negative as opposed to ordered as in Johnson (2007) and Brezger and Steiner (2008). In the Bayesian setting, we adopt gamma prior distributions for coefficients that have support on the positive reals with hyperparameters chosen such that the variance is large and the mean is relatively small. The basic model allows for both monotonic and convex shape restrictions with other extensions such as increasing and convex. Furthermore, the diffuse prior provides estimates for the coefficient of the spline basis functions that perform well when estimating functions with both flat and steep spots as evidenced by a simulation study. In addition to enabling implementation of several different types of shape restrictions, the generalized linear model framework can be used to model several different types of data such as normal, Poisson, and Bernoulli response variables. A simulation study shows that the proposed method of function estimation performs similarly, if not better, in terms of mean squared error when compared to the shaperestricted maximum likelihood estimate (ML SRRS) (Meyer, 2008) and better than the constrained monotone estimation of Brezger and Steiner (2008) for functions that involve flat regions.

We begin this chapter by proposing a Bayesian shape-restricted regression spline (Bayes SRRS) model in a generalized linear model framework. In Section 2.3, we briefly discuss knot selection for this model which will be further addressed in Chapter 4. Specific applications to data with independent normal errors are discussed in Section 2.4. Several different types of inference for the normal errors model are discussed in Section 2.6. Bernoulli and Poisson models are discussed in Section 2.5. In Section 2.7, we establish consistency of Bayes SRRS estimate for the normal errors model under the monotone shape restriction. In Section 2.8, we examine the small sample behavior of this model via a simulation study. We give results from simulation studies examining the inference methods proposed in Section 2.6 for the normal errors model in Sections 2.8.2-2.8.4.3. We conclude Chapter 2 with the application of the Bayes SRRS estimation procedure to two data sets.

### 2.2 Shape-restricted Regression Spline Model

Consider independent observations $y_{1}, \ldots, y_{n}$ generated from the distribution

$$
\begin{equation*}
p\left(y_{i} ; \theta, \phi\right)=\exp \left\{\left[y_{i} \theta_{i}-b\left(\theta_{i}\right)\right] / \phi-c\left(y_{i}, \phi\right)\right\} \tag{2.1}
\end{equation*}
$$

where the specifications of $b$ and $c$ determine the sub-family of models. Common examples are $b(\theta)=\log \left(1+e^{\theta}\right)$ for the Bernoulli, $b(\theta)=\exp (\theta)$ for the Poisson model, and $b(\theta)=\theta^{2} / 2$ with $c\left(y_{i}, \phi\right)=y^{2} \phi / 2$ for the normal errors regression model. The mean vector $\boldsymbol{\mu}=E(\boldsymbol{y})$ has values $\mu_{i}=b^{\prime}\left(\theta_{i}\right)$, and is related to a design matrix of predictor variables through a link function $g\left(\mu_{i}\right)=\eta_{i}, i=1, \ldots, n$. We consider the additive model

$$
\begin{equation*}
\eta_{i}=f_{1}\left(x_{1 i}\right)+\cdots+f_{L}\left(x_{L i}\right)+\boldsymbol{z}_{i}^{\prime} \boldsymbol{\gamma} \tag{2.2}
\end{equation*}
$$

where $\boldsymbol{\gamma}$ is a parameter vector and $\boldsymbol{z}_{i}$ is a vector of variables to be modeled parametrically. The functions $f_{l}$ of the continuous predictors $x_{l}$ are assumed to be smooth;
shape restrictions such as monotonicity or convexity may be imposed on the functions $f_{l}$. For the general case, the vector $\boldsymbol{\eta}$ with elements $\eta_{i}$ for $i=1, \ldots, n$ is approximated by

$$
\begin{equation*}
\sum_{j=1}^{m_{1}} \beta_{1 j} \boldsymbol{\delta}_{1 j}+\cdots+\sum_{j=1}^{m_{L}} \beta_{L j} \boldsymbol{\delta}_{L j}+\sum_{j=1}^{p} \alpha_{j} \boldsymbol{v}_{j}, \text { where } \beta_{l j} \geq 0, \quad \text { for all } l, j \text {. } \tag{2.3}
\end{equation*}
$$

The $\boldsymbol{\delta}_{l j}$ for $j=1, \ldots, m_{l}$ are basis vectors corresponding to the appropriate shaperestricted basis functions for $f_{l}$. For example, suppose $L=2$ and we assume that $f_{1}$ is monotone increasing and $f_{2}$ is convex. Two sets of knots are chosen to span the ranges of the two predictors, and two sets of basis vectors $\boldsymbol{\delta}_{l j}, j=1, \ldots, m_{l}, l=1,2$ are computed, where the first set uses monotone $I$-splines and the second are cubic $C$-splines. The $\boldsymbol{v}_{j}$ consist of the one vector and the vectors of the observed values of covariates to be modeled parametrically. Further, when $f_{l}$ is assumed to be convex, the $\boldsymbol{x}_{l}$ vector is included as one of the $\boldsymbol{v}_{j}$. Note that the $\boldsymbol{\gamma}$ in (2.2) is contained within $\boldsymbol{\alpha}=\left(\alpha_{1}, \ldots, \alpha_{p}\right)^{\prime}$. We adopt a Bayesian method for inference, defining prior densities for the $\alpha$ and $\beta$ coefficients. To impose shape restrictions, we restrict the prior for the $\beta$ coefficients to the positive reals.

### 2.3 Knot Selection

The basis functions in (2.3) depend on number and location of knots points. A key advantage of shape-restricted function estimation is that this method is generally robust to the choice of knot number and placement, a feature not typically shared by unrestricted smoothing methods (Meyer, 2008). Because the monotonicity or convexity obviates the wiggling associated with over-fitting, the number of knots must simply be large enough to capture the overall curvature in the data. One could increase the number of knots until the fit appears to converge.

Huang and Stone (2002) show that for unrestricted regression splines, the optimal number of knots to attain the pointwise convergence rate of $O_{r}\left(n^{-r /(2 r+1)}\right)$ is $n^{1 /(2 r+1)}$ where $r$ is the order of the regressions splines. This rounds to two or
three interior knots for $n$ as large as 500. The default placement should be at equal $x$-quantiles, although more knots may be placed where the function is thought to change slope more quickly. Meyer (2008) found via a simulation study that for functions with steep increases and sample sizes less than 200, estimates using more than 2 interior knots produce better fits in terms of means squared error. The simulation study in Section 2.8 uses 2 or 3 equally spaced interior knots for sample sizes of 20 and 50, respectively. For a more rigorous discussion of knot choice see Meyer (2008).

Robustness of shape-restricted regression splines to number of knots for the normal errors model is demonstrated in Figure 2.1 which gives function estimates using the Bayes SRRS model and priors as given in Section 2.4 with $k$ equally spaced interior knots over the range of covariate $x$. In Figure 2.1 (a), it appears that the monotone fits for the various knot choices are close together. The convex fits in Figure 2.1(b) are almost indistinguishable for $k=2$ to 5 interior knots.

Though the Bayes SRRS are robust to the number and location of knot points, the use of a Bayesian paradigm enables the joint estimation of regression parameters, number of knots, and knot locations via reversible jump Markov chain Monte Carlo (RJMCMC) (Green, 1995). A RJMCMC model will allow one to simultaneously estimate the regression parameters (regression coefficients and variance parameters) in addition to the location and number of interior knots. The extension of this model to free-knot splines is given in Chapter 4 . We believe that fixing $k$ interior knots is sufficient for the majority cases as evidenced by the simulation study below while at the same time being less computationally intensive. However, if one is uncomfortable fixing the number and location of interior knots, the RJMCMC model in Chapter 4 can be implemented.


Figure 2.1: (a) Estimated regression functions for $n=50$ data points simulated from sigmoid function $f(x)=5 \exp (10 x-5) /(1+\exp (10 x-5)$ plus standard normal error with $k=2,3$, 4 , and 5 equally spaced interior knot points. (b) Estimated regression functions for $n=50$ data points simulated from cubic function $f(x)=5 x^{3}$ plus standard normal error with $k=2,3,4$, and 5 equally spaced interior knot points.

### 2.4 Normal Errors Model

### 2.4.1 Bayesian Model

Suppose $\boldsymbol{y}=\boldsymbol{\eta}+\boldsymbol{\epsilon}$, where $\boldsymbol{\eta}=\left(\eta_{1}, \ldots, \eta_{n}\right)^{\prime}$ is modeled as in (2.3) and $\boldsymbol{\epsilon}$ is normally distributed with mean zero and covariance matrix $\tau^{-1} I$ where $I$ denotes the identity matrix. To simplify notation, let $\boldsymbol{\beta}=\left(\beta_{11}, \beta_{12}, \ldots, \beta_{1 m_{1}}, \ldots, \beta_{L 1}, \ldots, \beta_{L m_{L}}\right)^{\prime}$, and $\boldsymbol{\alpha}=\left(\alpha_{1}, \ldots, \alpha_{p}\right)^{\prime}$. We assume a gamma prior for $\tau$ with shape $d_{1}$ and rate $d_{2}$ so its density is

$$
\tau^{d_{1}} \exp \left\{-d_{2} \tau\right\} \mathrm{I}\{0<\tau<\infty\}
$$

and we denote it by $\operatorname{Gamma}\left(d_{1}, d_{2}\right)$. The prior parameters are chosen so that the mean of the prior density, $d_{1} / d_{2}$, is the inverse of a guess for the model variance. For a vague prior, let the prior variance $d_{1} / d_{2}^{2}$ be some multiple of the mean, say ten times the mean. The priors for the $\beta_{l j}$ coefficients should have support on the positive reals. For each $\beta_{l j}$, we choose a gamma prior with shape $c_{l 1}$ and rate $c_{l 2}$
so $\beta_{l j} \sim \operatorname{Gamma}\left(c_{l 1}, c_{l 2}\right)$. Using this prior, the mean $c_{l 1} / c_{l 2}$ can be chosen to be a prior guess for $R_{l}$ divided by $m_{l}$. Note that for monotone $f_{l}$, if the range of each basis function is one, then $\sum_{j=1}^{m_{l}} \beta_{l j}$ coincides with the range of $f_{l}$ given by $R_{l}=f_{l}\left(\max \left(\boldsymbol{x}_{l}\right)\right)-f_{l}\left(\min \left(\boldsymbol{x}_{l}\right)\right)$. To see this, note that $\left(\delta_{l j n}-\delta_{l j 1}\right)=1$ for all $j$ and $l$ so

$$
\begin{aligned}
R_{l} & =f_{l}\left(\max \left(\boldsymbol{x}_{l}\right)\right)-f_{l}\left(\min \left(\boldsymbol{x}_{l}\right)\right) . \\
& =\sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j n}+\alpha-\left(\sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j 1}+\alpha\right) \\
& =\sum_{j=1}^{m_{l}} \beta_{j}\left(\delta_{l j n}-\delta_{l j 1}\right)=\sum_{j=1}^{m_{l}} \beta_{l j} .
\end{aligned}
$$

Here $m_{l}=k_{l}+2$ where $k_{l}$ is the number of interior knots used for function $f_{l}$ and $\delta_{l j i}$ is the value for the $(l j)$ th basis vector found using knot locations $\boldsymbol{t}_{l}=$ $\left(t_{l, 1}=\min \left(\boldsymbol{x}_{l}\right), \ldots, t_{l, k_{l}+2}=\max \left(\boldsymbol{x}_{l}\right)\right)$ evaluated at $x_{i}$. For a diffuse prior that leads to flexible fits, the prior variance is some multiple of the mean. For all examples in Section 2.9 and simulations in Section 2.8, that multiple is ten. A truncated normal prior distribution might be considered, as used in Brezger and Steiner (2008) and Johnson (2007). However, it is easily seen that whatever the mean and variance of the original density, when a normal density is truncated above the mean, the standard deviation is never larger than the mean (Barr and Sherrill, 1999). This leads to non-diffuse priors and a difficulty in modeling flat spots. We assume $\boldsymbol{\alpha} \sim$ $\mathrm{N}(\mathbf{0}, M I)$ where $M$ is chosen by the user; for all the simulations and examples presented in Sections 2.8 and 2.9, we used $M=1000$. Informative priors for the regression function may also be imposed.

The likelihood for the normal errors model is proportional to

$$
\begin{equation*}
L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau ; \boldsymbol{Y}) \propto \tau^{n / 2} \exp \left\{-\frac{\tau}{2} \sum_{i=1}^{n}\left(y_{i}-\sum_{j=1}^{p} \alpha_{j} v_{j i}-\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j i}\right)^{2}\right\} \tag{2.4}
\end{equation*}
$$

and the joint prior density is proportional to

$$
\begin{align*}
p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau) \propto & {\left[\prod_{l=1}^{L} \prod_{j=1}^{m_{l}} \mathrm{I}_{(0, \infty)}\left(\beta_{l j}\right)\right]\left[\prod_{l=1}^{L} \prod_{j=1}^{m_{l}} \beta_{l j}^{c_{l 1}-1} \exp \left(-\beta_{l j} c_{l 2}\right)\right] \times } \\
& {\left[\exp \left\{-\sum_{j=1}^{p} \frac{1}{2 M} \alpha_{j}^{2}\right\} \tau^{d_{1}-1} \exp \left\{-d_{2} \tau\right\} \mathrm{I}_{(0, \infty)}(\tau)\right] . } \tag{2.5}
\end{align*}
$$

The posterior distribution is proportional to the product of the likelihood and the prior. It is proper but analytically intractable, so Markov chain Monte Carlo methods (Givens and Hoeting, 2005, Ch 7) are used to obtain samples from the posterior distribution.

In particular, we use a Gibbs sampler to sample from the posterior distributions and the conditional distributions used in this sampler are given below. Let $\boldsymbol{\beta}_{\left(-j_{0}\right)}$ be the $\boldsymbol{\beta}$ vector with $\beta_{j_{0}}$ removed and similarly let $\boldsymbol{\alpha}_{\left(-j_{0}\right)}$ be the $\boldsymbol{\alpha}$ vector with $\alpha_{j_{0}}$ removed. The conditional distribution for $\alpha_{j_{0}}$, given the data, $\tau, \boldsymbol{\beta}$, and $\alpha_{j}, j \neq j_{0}$, $p\left(\alpha_{j_{0}} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}_{\left(-j_{0}\right)}, \tau, \boldsymbol{y}\right)$, is distributed as

$$
\begin{equation*}
\mathrm{N}\left(\left[\tau \sum_{i=1}^{n} r_{i} v_{j i}\right] /\left[1 / M+\tau \sum_{i=1}^{n}\left(v_{j 0 i}\right)^{2}\right],\left[1 / M+\tau \sum_{i=1}^{n}\left(v_{j 0}\right)^{2}\right]^{-1}\right) \tag{2.6}
\end{equation*}
$$

where, for each $j_{0}=1, \ldots, p, r_{i}=y_{i}-\sum_{j \neq j_{0}} \alpha_{j} v_{j i}-\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j i}$ and $\mathrm{N}\left(\mu, \sigma^{2}\right)$ denotes a normal distribution with mean $\mu$ and variance $\sigma^{2}$. Note that the sum over $j \neq j_{0}$ is the sum from $j=1, \ldots, p$ minus $\alpha_{j_{0}} v_{j_{0} i}$. The conditional posterior density for $\tau$ given the data, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ coefficients, is

$$
\begin{equation*}
p(\tau \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{y}) \sim \operatorname{Gamma}\left(d_{1}+n / 2, d_{2}+S S E / 2\right), \tag{2.7}
\end{equation*}
$$

where $\mathrm{SSE}=\sum_{i=1}^{n}\left(y_{i}-r_{i}\right)^{2}$ with $r_{i}=\sum_{j=1}^{p} \alpha_{j} v_{j i}+\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j i}$ (the sum of squared residuals given the coefficients). The conditional posterior density for $\beta_{l_{0} j_{0}}$, given the data, $\tau, \boldsymbol{\alpha}$, and $\boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}$ is proportional to

$$
\begin{equation*}
\beta_{l_{0} j_{0}}^{c_{l_{0} 1}-1} \exp \left\{-\frac{s_{l_{0} j_{0}} \tau}{2}\left[\beta_{l_{0} j_{0}}-\left(\sum_{i=1}^{n} r_{i} \delta_{l_{0} j_{0} i} / s_{l_{0} j_{0}}-c_{l_{0} 2} /\left(s_{l_{0} j_{0}} \tau\right)\right)^{2}\right\} I_{(0, \infty)}\left(\beta_{l_{0} j_{0}}\right)\right. \tag{2.8}
\end{equation*}
$$

where $r_{i}=y_{i}-\sum_{j=1}^{p} \alpha_{j} v_{j i}-\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j i}+\beta_{l_{0} j_{0}} \delta_{l_{0} j_{0} i}$ and $s_{l_{0} j_{0}}=\sum_{i=1}^{n}\left(\delta_{l_{0} j_{0} i}\right)^{2}$. The density is of the form $f(x) \propto x^{a} \exp \left\{-b(x-c)^{2}\right\} \mathrm{I}\{x>0\}$ where $x=\beta_{l_{0} j_{0}}, a=$ $c_{l_{0} 1}-1, b=s_{l_{0} j_{0}} \tau / 2>0, c=\sum_{i=1}^{n} r_{i} \delta_{l_{0} j_{0} i} / s_{l_{0} j_{0}}-c_{l_{0} 2} /\left(s_{l_{0} j_{0}} \tau\right)$, and I is the indicator function. This can be sampled from using the Metropolis-Hastings algorithm or an auxiliary variable Markov chain Monte Carlo technique (Meyer and Laud, 2002; Givens and Hoeting, 2005, Ch 8.1).

The auxiliary variables method in the normal errors model involves the auxiliary variable $u$ and the joint density

$$
\begin{equation*}
f(x, u) \propto x^{a} \mathrm{I}\left\{0<u<\exp \left(-b(x-c)^{2}\right)\right\} \mathrm{I}\{x>0\} . \tag{2.9}
\end{equation*}
$$

Note that if you integrate $f(x, u)$ over $u$, you get a density proportional to $p\left(\beta_{l_{0} j_{0}} \mid \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \boldsymbol{\alpha}, \tau, \boldsymbol{y}\right)$. By alternating between random draws from the conditional distributions, where

$$
\begin{equation*}
f(u \mid x) \propto \mathrm{I}\left\{0<u<\exp \left(-b(x-c)^{2}\right)\right\} \tag{2.10}
\end{equation*}
$$

and

$$
\begin{equation*}
f(x \mid u) \propto x^{a} \mathrm{I}\left\{\max \left\{0, c-\sqrt{\frac{-\ln (u)}{b}}\right\}<x<c+\sqrt{\frac{-\ln (u)}{b}}\right\} \tag{2.11}
\end{equation*}
$$

and discarding the draws from the conditional distribution of $u$, we can generate a chain whose stationary distribution is $p\left(\beta_{l_{0} j_{0}} \mid \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \boldsymbol{\alpha}, \tau, \boldsymbol{y}\right)$.

Random draws from the conditional distribution for $u$ are made by randomly sampling from a Uniform $\left(0, \exp \left(-b(x-c)^{2}\right)\right)$ where Uniform $(a, b)$ denotes a continuous uniform distribution between $a$ and $b$. To obtain random draws from the conditional distribution of $x$ given $u$, we randomly sample $d$ from a $\operatorname{Uniform}(0,1)$, then find $x^{(t)}=F^{-1}(d)$ where $x^{(t)}$ is the $(t)$ th random draw for the variable with probability distribution function (pdf) $f(x \mid u)$ with conditional distribution (cdf) $F . \quad F^{-1}$ is the inverse of the cdf. The normalizing constant for the conditional distribution $f(x \mid u)$ in (2.11) is given by

$$
\begin{equation*}
\int_{l_{1}}^{l_{2}} x^{a} d x=\frac{l_{2}^{(a+1)}-l_{1}^{(a+1)}}{a+1} \tag{2.12}
\end{equation*}
$$

where $l_{1}=\max \left\{0, c-\sqrt{\frac{-\ln (u)}{b}}\right\}$ and $l_{2}=c+\sqrt{\frac{-\ln (u)}{b}}$. Thus, the conditional distribution for $x$ given $u$ is

$$
\begin{equation*}
f(x \mid u)=\frac{a+1}{l_{2}^{(a+1)}-l_{1}^{(a+1)}} x^{a} I\left\{l_{1}<x<l_{2}\right\} \tag{2.13}
\end{equation*}
$$

and integrating to obtain the cdf of the conditional distribution, we obtain

$$
F(x \mid u)= \begin{cases}\int_{l_{1}}^{x} f(t \mid u) d t & l_{1}<x<l_{2}  \tag{2.14}\\ 1 & x \geq l_{2} \\ 0 & x \leq l_{1}\end{cases}
$$

with

$$
\begin{equation*}
\int_{l_{1}}^{x} f(t \mid u) d t=\frac{a+1}{l_{2}^{(a+1)}-l_{1}^{(a+1)}} \int_{l_{1}}^{x} t^{a} d t=\frac{x^{(a+1)}-l_{1}^{(a+1)}}{l_{2}^{(a+1)}-l_{1}^{(a+1)}} . \tag{2.15}
\end{equation*}
$$

Therefore, the inverse of the cdf for $l_{1}<x<l_{2}$ is given by

$$
\begin{equation*}
F^{-1}(d)=\left[l_{1}^{(a+1)}+l_{2}^{(a+1)} d-l_{1}^{(a+1)} d\right]^{\left(\frac{1}{a+1}\right)}=\left[l_{2}^{(a+1)} d+(1-d) l_{1}^{(a+1)}\right]^{\left(\frac{1}{a+1}\right)} . \tag{2.16}
\end{equation*}
$$

At each iteration $t$, let the random draws from the full conditional distributions for the parameters in the Gibbs Sampler (MCMC realizations) be denoted by a subscript of $(t)$. Thus, $\beta_{l_{0} j_{0}}^{(t)}$ is the $(t)$ th sampled value for $\beta_{l_{0} j_{0}}$ and $\alpha_{j_{0}}^{(t)}$ is the $(t)$ th sampled value for $\alpha_{j_{0}}$. Let

$$
\begin{equation*}
\eta_{i}^{(t)}=\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j}^{(t)} \delta_{l j i}+\sum_{j=1}^{p} \alpha_{j}^{(t)} v_{j i} . \tag{2.17}
\end{equation*}
$$

The function estimate at $x_{i}$ is found by taking the mean of these $\eta_{i}^{(t)}$ values after removing burn-in and is given by

$$
\begin{equation*}
\hat{f}\left(x_{i}\right)=\frac{1}{N-B} \sum_{t=B+1}^{N} \eta_{i}^{(t)} \tag{2.18}
\end{equation*}
$$

where $N$ is the total number of iterations in the MCMC algorithm, and $B$ is the burn-in.

### 2.4.2 Additive Regression Model

The normal errors model can be applied to the additive regression model where there are multiple parallel curves. Consider the case where there is one continuous variable and a categorical predictor variable with $r$ levels. Let $\boldsymbol{v}_{q}=\left(v_{q 1}, \ldots, v_{q n}\right)$ for $q=g, \ldots, p$ be $r-1$ dummy variables for all but one of the levels of the categorical variable and let $\boldsymbol{v}_{q}$ for $q=1, \ldots, g-1$ be the other variables with unconstrained regression parameters. For the monotonically increasing or decreasing shape restriction with one continuous variable and a categorical predictor variable with $r$ levels, we would let $p=r$ and let $\boldsymbol{v}_{1}$ be the one vector. Likewise, for the convex or concave restriction with one continuous variable and a categorical predictor variable with $r$ levels, we would let $p=r+1$ and let $\boldsymbol{v}_{1}$ be the one vector and $\boldsymbol{v}_{2}$ be the vector of the observed values of the continuous covariates, $\boldsymbol{x}=\left(x_{1}, \ldots, x_{n}\right)^{\prime}$. The parallel curves model is given by

$$
\begin{equation*}
y_{i}=\sum_{j=1}^{m} \beta_{j} \delta_{j i}+\sum_{j=1}^{p} \alpha_{j} v_{j i}+\epsilon_{i} \tag{2.19}
\end{equation*}
$$

with independent normal error $\epsilon_{i} \sim N\left(0, \tau^{-1}\right)$. For the monotone shape restriction with one continuous variable and a categorical predictor variable with $r$ levels, an alternative parametrization for the parallel curves model is to let $p=r$ and each $\boldsymbol{v}_{j}$ for $j=1, \ldots, p$ be indicator variables for level $j$ of the categorical covariate. Using either parametrization, this model along with the priors in the previous Section 2.4.1 can be used to estimate parallel curves under shape restrictions.

### 2.4.3 Non-Additive Regression Model

An interaction effect between a categorical and a continuous variable may be included in the normal errors model. Consider a model with one continuous variable, $x$, and one categorical variable with $r$ levels. Let $y_{i}=f_{1}\left(x_{i}\right) d_{1 i}+\cdots+f_{r 1}\left(x_{i}\right) d_{r i}+$ $\epsilon_{i}$, where $d_{l}, l=1, \ldots, r$ are dummy variables for the $r$ levels of the categorical predictor. A monotone shape restriction is imposed on all $f_{l}$ by defining $m I$-spline
basis functions defined for a given set of knots on the range of the $x$-values (convex assumptions may be similarly imposed).

Let $\boldsymbol{x}_{l}=\left(x_{l 1}, \ldots, x_{l_{n}}\right)^{\prime}$ be the values of covariate $x$ observed at level $l$ of the categorical variable where $n_{l}$ is the number of observations at level $l$ of the categorical variable and let $\boldsymbol{x}$ be a vector of unique values created by combining and sorting the vector $\boldsymbol{x}_{l}$ for $l=1, \ldots, r$. Let $\boldsymbol{\delta}_{j}$ for $j=1, \ldots, m=k+2$ be the $I$-spline basis vectors create using vector $\boldsymbol{x}$ and $k$ equally spaced interior knots over the range of $\boldsymbol{x}$. Create new vectors $\boldsymbol{\delta}_{l j}$ for $l=1, \ldots, r$ and $j=1, \ldots, m$ of length $n=\sum_{l=1}^{r} n_{l}$ whose elements are defined by

$$
\boldsymbol{\delta}_{l j}= \begin{cases}\boldsymbol{\delta}_{j}\left(x_{i}\right) & d_{l i}=1 \\ 0 & d_{l i}=0\end{cases}
$$

where $\boldsymbol{\delta}_{j}\left(x_{i}\right)$ denotes the element in $\boldsymbol{\delta}_{j}$ that corresponds $x_{i}$. Let $\boldsymbol{v}_{1}$ be the one vector of length $n$ and $\boldsymbol{v}_{l}$ for $l=2, \ldots, r$ be defined by

$$
v_{l i}= \begin{cases}1 & d_{l, i}=1 \\ 0 & d_{l, i}=0\end{cases}
$$

Thus, we approximate the mean of $\boldsymbol{y}$ by

$$
\begin{equation*}
\sum_{l=1}^{r} \sum_{j=1}^{m} \beta_{l j} \boldsymbol{\delta}_{l j}+\sum_{l=1}^{r} \alpha_{l} v_{l i} \tag{2.20}
\end{equation*}
$$

To impose the shape restrictions, we again use diffuse gamma priors on each $\beta$ coefficient. As before, we can assume a $\operatorname{Gamma}\left(d_{1}, d_{2}\right)$ prior for the precision $\tau$ and vague normal priors for the unconstrained $\alpha$ coefficients.

### 2.5 Other Generalized Linear Models

It is often natural to impose monotonicity restrictions on the mean function in generalized linear regression models. If the link function is one-to-one, as for Bernoulli or Poisson, this is equivalent to constraining the function components, $f_{l}$, in (2.3). In the following subsections, we propose Bayesian regression models for the Bernoulli and Poisson models under the monotone shape restriction.

### 2.5.1 Bernoulli Model

Assume $y_{i}$ has a Bernoulli distribution with probability $p_{i}$ for $i=1, \ldots, n$ and that $p_{i}$ increases as the values of some covariates increase. We can model this relationship using the logit link function and letting $\operatorname{logit}\left(p_{i}\right)=\eta_{i}$ with $\eta_{i}$ the elements of the $\boldsymbol{\eta}$ as defined as in (2.3). Since for this link function, $p_{i}$ increases as covariates increase if and only if $\eta$ increases as covariates increase, we can impose the monotone shape restrictions by requiring $\eta$ to be monotonically increasing. This is done by using $I$-splines and restricting each $\beta_{l j}$ to be positive. The likelihood is given by

$$
L(\boldsymbol{\alpha}, \boldsymbol{\beta} ; \boldsymbol{Y}) \propto \prod_{i=1}^{n} \frac{\exp \left(\eta_{i} y_{i}\right)}{1+\exp \left(\eta_{i}\right)}
$$

For the Bayesian model, we assume the restricted regression coefficients ( $\beta_{l j}$ for $j=1, \ldots, m_{l}$ and $\left.l=1, \ldots, L\right)$ are independent with a $\operatorname{Gamma}\left(c_{l 1}, c_{l 2}\right)$ prior for each $\beta_{l j}$. As with the normal error model, we could have assumed a normal prior truncated on $(0, \infty)$ but this would not allow as much flexibility as the Gamma prior since it does not allow the mean of each $\beta_{l j}$ to be smaller than the variance. For the unrestricted regression coefficients ( $\alpha_{j}$ for $j=1, \ldots, p$ ), we assume that they are independent using a vague normal prior with mean of zero and large variance $M$, say 1000, for each $\alpha_{j}$. For the Bayesian SRRS Bernoulli model, the joint likelihood is proportional to

$$
\begin{align*}
L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{Y}) \propto & \prod_{i=1}^{n} \frac{\exp \left(\eta_{i} y_{i}\right)}{1+\exp \left(\eta_{i}\right)}\left[\exp \left\{-\sum_{j=1}^{p} \frac{1}{2 M} \alpha_{j}^{2}\right\}\right] \times \\
& {\left[\prod_{l=1}^{L} \prod_{j=1}^{m_{l}}\left(\beta_{l j}^{c_{l 1}-1} \mathrm{I}\left\{0<\beta_{l j}<\infty\right\}\right) \exp \left(-\sum_{l=1}^{L} c_{l 2} \sum_{j=1}^{m_{l}} \beta_{l j}\right)\right] } \tag{2.21}
\end{align*}
$$

where $\eta_{i}=\sum_{j=1}^{p} \alpha_{j} v_{j i}+\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j i}$.
As with the normal errors, the joint posterior distribution of the parameters is proper but analytically intractable so we use a Gibbs sampler to estimate the regression function. To sample from the full conditional distributions for each $\alpha_{j_{0}}$ and
each $\beta_{l_{0} j_{0}}$ we again use the auxiliary variable Markov chain Monte Carlo technique (Meyer and Laud, 2002; Givens and Hoeting, 2005, Ch 8.1) as in Section 2.4.1 and introduce auxiliary variables $\boldsymbol{u}=\left(u_{1}, \ldots, u_{n}\right)^{\prime}$ so

$$
\begin{align*}
& p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{Y}, \boldsymbol{u}) \\
& \propto \prod_{i=1}^{n} I\left\{0<u_{i}<\left[1+\exp \left(\eta_{i}\right)\right]^{-1}\right\} \exp \left(\eta_{i} y_{i}\right)\left[\exp \left\{-\sum_{j=1}^{p} \frac{1}{2 M} \alpha_{j}^{2}\right\}\right] \\
& {\left[\prod_{l=1}^{L} \prod_{j=1}^{m_{l}}\left(\beta_{l j}^{c_{l 1}-1} \mathrm{I}\left\{0<\beta_{l j}<\infty\right\}\right)\right] \exp \left(-\sum_{l=1}^{L} c_{l 2} \sum_{j=1}^{m_{l}} \beta_{l j}\right) } \tag{2.22}
\end{align*}
$$

Integrating out the $\boldsymbol{u}$ in (2.22), we get a function that is proportional to the likelihood for the Bernoulli Bayes SRRS model. The probability density for $u_{i}$ given all other values of the parameters and data, denoted $p\left(u_{i} \mid \cdot\right)$, follows a $\operatorname{Uniform}\left(0,\left[1+\exp \left(\eta_{i}\right)\right]^{-1}\right)$. For each $\alpha_{j_{0}}$, the density given all other parameters, auxiliary variables, and data is given by

$$
\begin{aligned}
& p\left(\alpha_{j_{0}} \mid \boldsymbol{\alpha}_{\left(-j_{0}\right)}, \boldsymbol{\beta}, \tau, \boldsymbol{y}, \boldsymbol{u}\right) \\
& \quad \propto\left[\prod_{i=1}^{n} \mathrm{I}\left\{0<u_{i}<\left[1+\exp \left(\eta_{i}\right)\right]^{-1}\right\}\right] \exp \left\{-\frac{1}{2 M}\left(\alpha_{j_{0}}-M \sum_{i=1}^{n} y_{i} v_{j_{0} i}\right)^{2}\right\}
\end{aligned}
$$

and is distributed as a normal distribution truncated on $(l b, u b)$. To find the lower truncation limit, $l b$, let $A_{j_{0}}=\left\{i: v_{j_{0} i}<0\right\}$ and

$$
l b=\left\{\begin{array}{ll}
-\infty & \text { if } A_{j_{0}}=\emptyset \\
\max _{i \in A_{j_{0}}}\left\{v_{j_{0} i}^{-1}\left[\log \left(u_{i}^{-1}-1\right)-\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j i}-\sum_{j \neq j_{0}} \alpha_{j} v_{j i}\right]\right\} & \text { if } A_{j_{0}} \neq \emptyset
\end{array} .\right.
$$

To find upper truncation limit, $u b$, let $B_{j_{0}}=\left\{i: v_{j_{0} i}>0\right\}$ which will not be empty and define

$$
u b=\min _{i \in B_{j_{0}}}\left\{v_{j_{0} i}^{-1}\left[\log \left(u_{i}^{-1}-1\right)-\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j i}-\sum_{j \neq j_{0}} \alpha_{j} v_{j i}\right]\right\} .
$$

The density for each $\beta_{l_{0} j_{0}}$ given all other parameters, auxiliary variables, and data is given by

$$
\begin{aligned}
p( & \left.\beta_{l_{0} j_{0}} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right) \\
\propto & \beta_{l_{0} j_{0}}^{c_{l_{1}-1}-1} \exp \left\{-\beta_{l_{0} j_{0}}\left[c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}\right]\right\} \mathrm{I}\left\{0<\beta_{l_{0} j_{0}}<\infty\right\} \times \\
& \prod_{i=1}^{n} \mathrm{I}\left\{0<u_{i}<\left[1+\exp \left(\eta_{i}\right)\right]^{-1}\right\} \\
\propto & \beta_{l_{0} j_{0}}^{c_{0_{0} 1}-1} \exp \left\{-\beta_{l_{0} j_{0}}\left[c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}\right]\right\} \mathrm{I}\left\{l b<\beta_{l_{0} j_{0}}<u b\right\}
\end{aligned}
$$

with $l b$ defined by letting $A_{l_{0} j_{0}}=\left\{i: \delta_{l_{0} j_{0} i}<0\right\}$ (which will not be empty if basis functions are scaled such that they are orthogonal to $\boldsymbol{v}_{1}$ ) and

$$
l b=\max \left\{0, \max _{i \in A_{l_{0} j_{0}}}\left\{\delta_{l_{0} j_{0} i}^{-1}\left[\log \left(u_{i}^{-1}-1\right)-\sum_{l \neq l_{0}} \sum_{j \neq j_{0}} \beta_{l j} \delta_{l j i}-\sum_{j=1}^{p} \alpha_{j} v_{j i}\right]\right\}\right\} .
$$

Similarly, let $B_{l_{0} j_{0}}=\left\{i: \delta_{l_{o} j_{0} i}>0\right\}$ (which will not be empty if basis functions are scaled such that they are orthogonal to $\boldsymbol{v}_{1}$ ) and

$$
u b=\min _{i \in B_{l_{0} j_{0}}}\left\{\delta_{l_{0} j_{0} i}^{-1}\left[\log \left(u_{i}^{-1}-1\right)-\sum_{l \neq l_{0}} \sum_{j \neq j_{0}} \beta_{l j} \delta_{l j i}-\sum_{j=1}^{p} \alpha_{j} v_{j i}\right]\right\} .
$$

Thus, $p\left(\beta_{l_{0} j_{0}} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right)$ is of a form of a $\operatorname{Gamma}\left(c_{l_{0} 1},\left[c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0}}\right]\right)$ truncated on $(l b, u b)$ except that $\left[c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0}}\right]$ can and often will be negative. Thus, to sample from this distribution we can use another auxiliary variable. Consider the auxiliary variable $w$ and let

$$
\begin{aligned}
& p\left(\beta_{l_{0} j_{0}}, w \mid \boldsymbol{\alpha}, \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right) \\
& \quad \propto \beta_{l_{0} j_{0}}^{c_{l_{0} 1}-1} \mathrm{I}\left\{l b<\beta_{l_{0} j_{0}}<u b\right\} \mathrm{I}\left\{0<w<\exp \left(-\beta_{l_{0} j_{0}}\left[c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}\right]\right)\right\} .
\end{aligned}
$$

Note that if we integrate out $w$, we get $p\left(\beta_{l_{0} j_{0}} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right)$. We alternate sampling from $p\left(\beta_{l_{0} j_{0}} \mid w, \boldsymbol{\alpha}, \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right)$ and $p(w \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \tau, \boldsymbol{y}, \boldsymbol{u})$. Under this specification, $p(w \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \tau, \boldsymbol{y}, \boldsymbol{u})$ has a uniform distribution on

$$
\left(0, \exp \left(-\beta_{l_{0} j_{0}}\left[c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}\right]\right)\right)
$$

and

$$
p\left(\beta_{l_{0} j_{0}} \mid w, \boldsymbol{\alpha}, \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right) \propto \beta_{l_{0} j_{0}}^{c_{l_{0}}-1} \mathrm{I}\left\{l b^{*}<\beta_{l_{0} j_{0}}<u b^{*}\right\}
$$

where

$$
l b^{*}= \begin{cases}\max \left\{l b, \frac{-\log (w)}{c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}}\right\} & \text { if } c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}<0 \\ l b & \text { if } c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}>0\end{cases}
$$

and

$$
u b^{*}=\left\{\begin{array}{ll}
u b & \text { if } c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}<0 \\
\min \left\{u b, \frac{-\log (w)}{c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l j i}}\right\} & \text { if } c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}>0
\end{array} .\right.
$$

To obtain random draws from this density, we invert the cdf as we did for $\beta_{l_{0} j_{0}}$ in the normal errors model. We randomly sample $d$ from a $\operatorname{Uniform}(0,1)$, then find $\beta_{l_{0} j_{0}}^{(t)}=$ $F^{-1}(d)$ where $F$ is the cdf with corresponding pdf, $p\left(\beta_{l_{0} j_{0}} \mid w, \boldsymbol{\alpha}, \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right)$, and $\beta_{l_{0} j_{0}}^{(t)}$ represents the value for $\beta_{l_{0} j_{0}}$ at the $(t)$ th random draw of the Gibbs sampler. Note if $c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}=0$ then can sample from $p\left(\beta_{l_{0} j_{0}} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right)$ by inverting cdf and do not need auxiliary variable $w$.

### 2.5.2 Poisson Model

For the Poisson model, suppose you have independent observations $y_{i}$ for $i=$ $1, \ldots, n$ assumed to follow a Poisson distribution with mean $\lambda_{i}$ and that the mean is increasing with covariate(s). The shape restrictions can be imposed on the mean in a similar fashion as with the Bernoulli model by assuming the mean of the $\boldsymbol{y}=$ $\left(y_{1}, \ldots, y_{n}\right)^{\prime}$ is approximated by $\boldsymbol{\eta}$ in (2.3) and is related to design variables through link function $\log (\cdot)$. Note that $\eta$ is increasing if and only if $\log (\eta)$ is increasing. As with the normal errors model and Bernoulli model under the monotone assumption, we can use $I$-spline basis functions and restrict each $\beta_{l j}$ to be positive. Thus, we can again assume a $\operatorname{Gamma}\left(c_{l 1}, c_{l 2}\right)$ for each $\beta_{l j}$, a $\operatorname{Normal}(0, M)$ prior for each $\alpha_{j}$,
and assume the parameters are independent. The joint likelihood is given by

$$
\begin{aligned}
& L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{Y}) \\
& \propto \quad \exp \left\{\sum_{i=1}^{n} y_{i} \eta_{i}\right\} \prod_{i=1}^{n}\left\{\exp \left[-\exp \left(\eta_{i}\right)\right]\right\} \prod_{i=1}^{n} \mathrm{I}\left\{y_{i} \in 0 \cup \mathbb{Z}^{+}\right\} \times \\
& \\
& \quad \prod_{l=1}^{L} \prod_{j=1}^{m}\left(\beta_{l j}^{c_{l j}-1} \mathrm{I}\left\{0<\beta_{l j}<\infty\right\}\right) \exp \left\{-c_{l 2} \sum_{l=1}^{L} \sum_{j=1}^{m} \beta_{l j}\right\} \exp \left\{-\frac{1}{2 M} \sum_{j=1}^{p} \alpha_{j}^{2}\right\} .
\end{aligned}
$$

We sample from the posterior distribution using Gibbs sampler and as with the Bernoulli model, we use auxiliary variables $\boldsymbol{u}=\left(u_{1}, \ldots, u_{n}\right)^{\prime}$. Consider the density for all model parameters and the auxiliary variable vector,

$$
\begin{aligned}
& p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{y}, \boldsymbol{u}) \\
& \propto \exp \left\{\sum_{i=1}^{n} y_{i} \eta_{i}\right\} \prod_{i=1}^{n} \mathrm{I}\left\{0<u_{i}<\exp \left[-\exp \left(\eta_{i}\right)\right]\right\} \times \\
& \quad \prod_{l=1}^{L} \prod_{j=1}^{m}\left(\beta_{l j}^{c_{l j}-1} \mathrm{I}\left\{0<\beta_{l j}<\infty\right\}\right) \exp \left\{-c_{l 2} \sum_{l=1}^{L} \sum_{j=1}^{m} \beta_{l j}\right\} \exp \left\{-\frac{1}{2 M} \sum_{j=1}^{p} \alpha_{j}^{2}\right\} .
\end{aligned}
$$

The density for each $u_{i}$ given all other parameters, auxiliary variables, and the data follows a uniform distribution with a lower limit of zero and a upper limit of $\exp \left\{-\exp \left(\eta_{i}\right)\right\}$. For each $\alpha_{j_{0}}, p\left(\alpha_{j_{0}} \mid \boldsymbol{\beta}, \boldsymbol{\alpha}_{\left(-j_{0}\right)}, \tau, \boldsymbol{y}, \boldsymbol{u}\right)$ is a normal distribution with mean of $M \sum_{i=1}^{n} y_{i} v_{j_{0} i}$ and variance of $M$ truncated on $(l b, u b)$. For $l b$, let $A_{j_{0}}=\left\{i: v_{j_{0} i}>0\right\}$ and define

$$
l b= \begin{cases}-\infty & \text { if } A_{j_{0}}=\emptyset \\ \max _{i \in A_{j_{0}}}\left\{v_{j_{0} i}^{-1}\left[\log \left\{-\log \left(u_{i}\right)\right\}-r_{i}\right]\right\} & \text { if } A_{j_{0}} \neq \emptyset\end{cases}
$$

where $r_{i}=\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l} \delta_{l j i}+\sum_{j \neq j_{0}} \alpha_{j} v_{j i}$ and sum over $j \neq j_{0}$ means take the sum of $j$ over $j=1, \ldots, p$ excluding $j=j_{0}$. Similarly, let $B=\left\{i: v_{j_{0} i}<0\right\}$ and define

$$
u b= \begin{cases}\infty & \text { if } B_{j_{0}}=\emptyset \\ \min _{i \in B_{j_{0}}}\left\{v_{j_{0} i}^{-1}\left[\log \left\{-\log \left(u_{i}\right)\right\}-r_{i}\right]\right\} & \text { if } B_{j_{0}} \neq \emptyset\end{cases}
$$

For each $\beta_{l_{0} j_{0}}$, the conditional distribution for the Gibbs sampler is given by

$$
\begin{aligned}
& p\left(\beta_{l_{0} j_{0}} \mid \boldsymbol{\beta}_{\left(-l_{0} j_{0}\right)}, \boldsymbol{\alpha}, \tau, \boldsymbol{y}, \boldsymbol{u}\right) \\
& \propto \quad \prod_{i=1}^{n} \mathrm{I}\left\{0<u_{i}<\exp \left[-\exp \left(\eta_{i}\right)\right]\right\} \beta_{l_{0} j_{0}}^{c_{0_{0}}-1} \mathrm{I}\left\{0<\beta_{l_{0} j_{0}}<\infty\right\} \times \\
& \quad \exp \left\{-\beta_{l_{0} j_{0}}\left[c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}\right]\right\},
\end{aligned}
$$

and is the form of gamma distribution with shape and scale parameters $c_{l_{0} 1}$ and $c_{l_{0} 2}-\sum_{i=1}^{n} y_{i} \delta_{l_{0} j_{0} i}$ that is truncated except that the scale parameter may and often will be negative. This distribution is truncated on $(l b, u b)$ with $l b$ defined by letting $A_{l_{0} j_{0}}=\left\{i: \delta_{l_{0} j_{0} i}<0\right\}$ (which will not be empty if scale basis vectors to have range of one) and

$$
l b=\max \left\{0, \max _{i \in A_{l_{0} j_{0}}}\left\{\delta_{l_{0} j_{0} i}^{-1}\left[\log \left\{-\log \left(u_{i}\right)\right\}-\sum_{l \neq l_{0}} \sum_{j \neq j_{0}} \beta_{l j} \delta_{l j i}-\sum_{j=1}^{p} \alpha_{j} v_{j i}\right]\right\}\right\} .
$$

Similarly, let $B_{l_{0} j_{0}}=\left\{i: \delta_{l_{0} j_{0} i}>0\right\}$ (which will not be empty) and

$$
u b=\min _{i \in B_{l_{0} j_{0}}}\left\{\delta_{l_{0} j_{0} i}^{-1}\left[\log \left\{-\log \left(u_{i}\right)\right\}-\sum_{l \neq l_{0}} \sum_{j \neq j_{0}} \beta_{l j} \delta_{l j i}-\sum_{j=1}^{p} \alpha_{j} v_{j i}\right]\right\}
$$

To sample from this distribution we can use another auxiliary variable as we did for the Bernoulli model described in Section 2.5.1.

### 2.6 Inference

The use of a Bayesian framework along with the Gibbs sampler aids in several different types of inference. We perform inference using random draws from the posterior distribution This is helpful since the variance of the restricted regression coefficient vector is not easy to obtain in a frequentist setting. Furthermore, the Bayesian framework provides several different model selection tools that aid in determining whether a particular shape restriction appears reasonable for a given data set while hypothesis tests to determine this are much more difficult to perform (or not possible) in a frequentist setting. Several examples of inference using the proposed Bayesian shape-restricted regression spline model are provided in the following subsections.

### 2.6.1 Credible and Prediction Intervals

Pointwise confidence bands for the regression function are easily produced from the same posterior draws from which the function estimate is calculated. For each set of $\beta$ and $\alpha$ parameters produced by an iteration of the Gibbs sampler (after burnin), we compute the function estimate at each $x$-value and get $\eta_{i}^{(t)}$ as in (2.17). At each $x$-value, we estimate the $(0.95 \cdot 100) \%$ highest posterior density (HPD) using these $\eta_{i}^{(t)}$ values and the Chen-Shao estimation algorithm (Chen et al., 2000, Ch 7). For a $(q \cdot 100 \%)$ pointwise HPD for $f\left(x_{i}\right)$, the algorithm involves first ordering $\eta_{i}^{(t)}$ from least to greatest where $t=B+1, \ldots, N, B$ is the burn-in, and $N$ is the total number of iterations of the MCMC algorithm. Denote these ordered realizations by $\theta_{(i)}$ with $i=1, \ldots, N-B$ and let $K=\lceil q \cdot(N-B)\rceil$ where $\lceil\cdot\rceil$ denotes the ceiling. Compute the lengths $l_{i}=\theta_{(i+K)}-\theta_{(i)}$ for $i=1, \ldots, N-B-K$ and select the $i^{*}=\underset{i}{\operatorname{argmin}}\left(l_{i}\right)$. The $(q \cdot 100 \%)$ pointwise $\operatorname{HPD}$ is then $\left(\theta_{\left(i^{*}\right)}, \theta_{\left(i^{*}+K\right)}\right)$.

For the normal errors model in Section 2.4, we can also construct prediction intervals by generating $\tilde{\eta}_{i}^{(t)}$ for each iteration $t$ from a normal distribution with mean equal to $\eta_{i}^{(t)}$ and variance equal to the $(t)$ th realization of the variance parameter from the MCMC algorithm, $\left(\tau^{(t)}\right)^{-1}$. We use $\tilde{\eta}_{i}^{(t)}$ values and the Chen-Shao algorithm to find the $95 \%$ pointwise HPD prediction interval.

### 2.6.2 Inference using the Posterior Distribution

For the normal errors model in Section 2.4, the Bayesian model provides a joint posterior distribution for the $\boldsymbol{\alpha}$ vector that is approximately normal so we can perform many types of inference on the parameters that are modeled parametrically. We can create confidence intervals or compare mean values using the joint normal distribution with estimates of mean vectors and variance easily obtained from the MCMC output. Let $J=\{1, \ldots, p\}$ be the induces for the $p$ parametrically modeled $\alpha_{j}$ parameters and $J^{\prime}=\left\{j_{1}, \ldots, j_{q}\right\} \subseteq J$. Suppose $\alpha_{j_{i}}^{(t)}$ is the $(t)$ th MCMC replicate
for $\alpha_{j_{i}}$ with $j_{i} \in J^{\prime}$. Define $\overline{\boldsymbol{\alpha}}=\left(\bar{\alpha}_{j_{1}}, \ldots, \bar{\alpha}_{j_{q}}\right)$ and the sample covariance matrix $R$ by

$$
\begin{equation*}
\bar{\alpha}_{j_{i}}=\frac{1}{T} \sum_{t=1}^{T} \alpha_{j_{i}}^{(t)} \text { and } R_{i l}=\frac{1}{T-1} \sum_{t=1}^{T}\left(\alpha_{j_{i}}^{(t)}-\bar{\alpha}_{j_{i}}\right)\left(\alpha_{j_{l}}^{(t)}-\bar{\alpha}_{j_{l}}\right), \tag{2.23}
\end{equation*}
$$

where $T$ is the number of post-burn in posterior replicates.
In particular, consider the additive regression model as in Section 2.4.2 with one continuous and one categorical predictor variable with $r$ levels. Let $d_{j}, j=$ $1, \ldots, r-1$ be dummy variables for all but one of the levels of the categorical variable and model the data as

$$
\begin{equation*}
y_{i}=\alpha_{0}+f\left(x_{i}\right)+\sum_{j=1}^{r-1} \alpha_{j} d_{j i}+\epsilon_{i}, \text { for } i=1, \ldots, n \tag{2.24}
\end{equation*}
$$

where $f$ is assumed to be smooth with a shape restriction and $f(0)=0$ for identifiably. If $r=2$, to determine whether a parallel curves model is appropriate for the data, we can compute the $95 \%$ credible interval for $\alpha_{1}$ and determine whether it includes zero. For $r>2$, we utilize the approximate normality of the joint posterior and perform a chi-squared test with $r-1$ degrees of freedom. The test statistic $X=\overline{\boldsymbol{\alpha}}^{\prime} R^{-1} \overline{\boldsymbol{\alpha}}$, where $\overline{\boldsymbol{\alpha}}$ and $R$ are defined in (2.23), is compared with a $\chi^{2}(r-1)$ distribution and the hypothesis of a single regression curve is rejected if $X$ is large.

### 2.6.3 Model Selection

Model comparison between the many different possible model types is an important component to any regression modeling. There are several options for model comparisons which include AIC (Akaike, 1974), AICC (Hurvich and Tsai, 1989), BIC (Schwarz, 1978), DIC (Spiegelhalter et al., 2002), and Bayes Factors (Kass and Raftery, 1995). We use Bayes factors to select models. For model $M_{1}$ with parameters $\boldsymbol{\theta}_{M_{1}}$, the Bayes factor in comparison to model $M_{2}$ assuming equal model probabilities is given by $\mathrm{BF}=p\left(\boldsymbol{y} \mid M_{1}\right) / p\left(\boldsymbol{y} \mid M_{2}\right)$ where

$$
p\left(\boldsymbol{y} \mid \mathrm{M}_{1}\right)=\int p\left(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_{1}}, M_{1}\right) p\left(\boldsymbol{\theta}_{M_{1}} \mid M_{1}\right) d \boldsymbol{\theta}_{M_{1}}
$$

is the integrated likelihood for $M_{1}$. For our models the integrated likelihood is generally intractable, so the BF must be estimated using numerical methods.

We approximate the BF using the harmonic mean estimate (Newton and Raftery, 1994; Raftery et al., 2007) given by

$$
\begin{equation*}
\hat{p}\left(\boldsymbol{y} \mid M_{1}\right)=\left[\frac{1}{T} \sum_{t=1}^{T} \frac{1}{p\left(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_{1}}^{(t)}, \mathrm{M}_{1}\right)}\right]^{-1} \tag{2.25}
\end{equation*}
$$

where $\boldsymbol{\theta}_{M_{1}}^{(t)}$ is the estimate of the parameters for $M_{1}$ from iteration $t$ of the MCMC algorithm and $T$ is the number of iterations after burn-in. This estimate is based on the harmonic mean identity given by

$$
\frac{1}{f(\boldsymbol{y})}=\int[f(\boldsymbol{\theta} \mid \boldsymbol{y}) / f(\boldsymbol{y} \mid \boldsymbol{\theta})] d \boldsymbol{\theta}=\mathrm{E}\left[\left.\frac{1}{f(\boldsymbol{y} \mid \boldsymbol{\theta})} \right\rvert\, \boldsymbol{y}\right]
$$

For some cases, we found this estimator was unsatisfactory with some extremely small values of $p\left(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_{i}}^{(t)}, M_{i}\right)$. We used importance sampling ideas (Givens and Hoeting, 2005, Ch 6) to stabilize the estimate by discarding the realizations $\boldsymbol{\theta}_{M_{i}}^{(t)}$ from the MCMC realizations where $\left[p\left(\boldsymbol{y} \mid \boldsymbol{\theta}_{M_{i}}^{(t)}, M_{i}\right)\right]^{-1}$ was above the 95 th quantile of these estimates. For simulations where we know the model that generated the data, this approach lead to reasonable model selection results and generally lead to BF estimates that converged to reasonable values. Below we refer to this approach as the approximate BF. For the simulations in Section 2.8 we used a cutoff of $B F=3$ as suggested by Kass and Raftery (1995). Alternatively, the Bayes factor cutoff may be calibrated to produce a desired test size.

### 2.7 Consistency

### 2.7.1 Background

We now focus on the asymptotic behavior of the Bayes SRRS estimate for the normal errors model in Section 2.4.1 assuming the monotonicity constraint holds and the interior knot locations are fixed. To demonstrate that the regression function and precision estimates under the strictly monotone assumption are consistent,
we show convergence of Bayes SRRS parameter estimates to their corresponding unconstrained maximum likelihood regression spline (ML RS) estimate which are known to be consistent (Huang and Stone, 2002).

In this section, we consider the consistency of non-parametric strictly monotone quadratic regression splines estimates for data assumed to be from the normal errors model in Section 2.4.1 with unknown variance, fixed number and locations of interior knot points, and assuming the constraint is valid. Recall that when fixing the number and locations of interior knots, the Bayes SRRS estimate for the normal errors model at $x_{i}$ is given by (2.18). Let the Bayes SRRS estimates for each $\beta_{l j}$ for $j=1, \ldots, m_{l}$ and $l=1, \ldots, L$, be denoted by $\hat{\beta}_{l j}$, the Bayes SRRS estimates for each $\alpha_{j}$ for $j=1, \ldots, p$, be denoted by $\hat{\alpha}_{j}$, and the Bayes SRRS estimate for $\tau$ be denoted by $\hat{\tau}$. In the following subsections, we show that the Bayes SRRS estimate for these parameters converge to their corresponding ML RS parameter estimates and therefore converge to their corresponding parameters.

### 2.7.2 Consistency of the Unrestricted Regression Coefficients

We define each $\boldsymbol{v}_{j}=\left(v_{j 1}, \ldots, v_{j n}\right)^{\prime}$ in (2.3) such that they form an orthogonal set and thus $\boldsymbol{v}_{j}$ is orthogonal to $\boldsymbol{v}_{k}$ for all $j, k \in\{1, \ldots, p\}$ with $j \neq k$. We define $\boldsymbol{\delta}_{l j}$ for $j=1, \ldots, m_{l}$ and $l=1, \ldots, L$ in (2.3) such that they are orthogonal to each $\boldsymbol{v}_{j}$ for $j=1, \ldots, p$ using

$$
\begin{equation*}
\boldsymbol{\delta}_{l j}=\boldsymbol{\delta}_{l j}^{*}-\frac{\boldsymbol{v}_{1}^{\prime} \boldsymbol{\delta}_{l j}^{*}}{\boldsymbol{v}_{1}^{\prime} \boldsymbol{v}_{1}} \boldsymbol{v}_{1}-\ldots-\frac{\boldsymbol{v}_{p}^{\prime} \boldsymbol{\delta}_{l j}^{*}}{\boldsymbol{v}_{p}^{\prime} \boldsymbol{v}_{p}} \boldsymbol{v}_{p} \tag{2.26}
\end{equation*}
$$

where $\delta_{l j}^{*}$ is the $(l j)$ th $I$-spline prior to rescaling. To obtain the unconstrained maximum likelihood estimate of the parameters in (2.3), we let $q=\sum_{l=1}^{L} m_{l}$ and $X$ be a $n \times(p+q)$ matrix consisting of $\boldsymbol{v}_{j}$ for $j=1, \ldots, p$ followed by $\boldsymbol{\delta}_{l j}$ for $l=1, \ldots, L$ and $j=1, \ldots, m_{l}$ as columns. Thus, $X=\left[\boldsymbol{x}_{1}, \ldots, \boldsymbol{x}_{p+q}\right]=\left[\boldsymbol{v}_{1} \ldots \boldsymbol{v}_{p} \boldsymbol{\delta}_{11} \ldots \boldsymbol{\delta}_{L m_{L}}\right]$. The ML RS estimates for the parameter vector

$$
\boldsymbol{\theta}=\left(\theta_{1} \ldots, \theta_{p+q}\right)^{\prime}=\left(\alpha_{1}, \ldots, \alpha_{p}, \beta_{11}, \ldots, \beta_{L m_{L}}\right)^{\prime}
$$

denoted by

$$
\tilde{\boldsymbol{\theta}}=\left(\tilde{\theta}_{1}, \ldots, \tilde{\theta}_{p+q}\right)^{\prime}=\left(\tilde{\alpha}_{1}, \ldots, \tilde{\alpha}_{p}, \tilde{\beta}_{11}, \ldots, \tilde{\beta}_{L m_{L}}\right)^{\prime}
$$

is given by

$$
\begin{equation*}
\tilde{\boldsymbol{\theta}}=X\left(X^{\prime} X\right)^{-1} X^{\prime} \boldsymbol{y} \tag{2.27}
\end{equation*}
$$

This implies that the ML SR estimate for the unconstrained regression parameters are found by

$$
\tilde{\alpha}_{j}=\tilde{\theta}_{j}=\frac{\boldsymbol{x}_{j}^{\prime}\left(\boldsymbol{y}-P_{X_{(-j)}} \boldsymbol{y}\right)}{\boldsymbol{x}_{j}^{\prime} \boldsymbol{x}_{j}}
$$

for $j=1, \ldots, p$ where $\boldsymbol{x}_{j}$ is the $(j)$ th column of $X, X_{(-j)}$ is the $X$ matrix removing $\boldsymbol{x}_{j}$, and $P_{X_{(-j)}}=X_{(-j)}\left(X_{(-j)}^{\prime} X_{(-j)}\right)^{-1} X_{(-j)}^{\prime}$ (Takeuchi et al., 1982, Ch 3). Now $\boldsymbol{x}_{j}^{\prime} P_{X_{(-j)}}=\mathbf{0}$ since $\boldsymbol{x}_{j}$ is orthogonal to all columns of $X_{(-j)}$ so the ML RS is

$$
\tilde{\alpha}_{j}=\frac{\boldsymbol{x}_{j}^{\prime} \boldsymbol{y}}{\boldsymbol{x}_{j}^{\prime} \boldsymbol{x}_{j}}=\frac{\sum_{i=1}^{n} v_{j i} y_{i}}{\sum_{i=1}^{n} v_{j i}^{2}} .
$$

We now show that as the sample size, $n$, approaches infinity, the Bayes SRRS estimate for $\alpha_{j}$ approaches $\tilde{\alpha}_{j}$. From (2.6), we know that the conditional distribution for $\alpha_{j}$ given all other parameters and the data that is used in the Gibbs sampler is a normal distribution with mean

$$
\begin{align*}
\mathrm{E}\left(\alpha_{j} \mid \boldsymbol{\alpha}_{(-j)}, \boldsymbol{\beta}, \tau, \boldsymbol{y}\right) & =\frac{\tau}{\frac{1}{M}+\tau \sum_{i=1}^{n} v_{j i}^{2}} \sum_{i=1}^{n} v_{j i}\left(y_{i}-\sum_{l=1}^{L} \sum_{k=1}^{m_{l}} \beta_{l k} \delta_{l k i}-\sum_{k \neq j} \alpha_{k} v_{k i}\right) \\
& =\frac{\tau}{\frac{1}{M}+\tau \sum_{i=1}^{n} v_{j i}^{2}} \sum_{i=1}^{n} v_{j i} y_{i}=\frac{\tau M \sum_{i=1}^{n} v_{j i} y_{i}}{1+\tau M \sum_{i=1}^{n} v_{j i}^{2}} \tag{2.28}
\end{align*}
$$

where summing over $k \neq j$ means summing over $k=1, \ldots, p$ excluding $k=j$. The second line is obtained by noting that $\boldsymbol{v}_{j}$ is orthogonal to other $\boldsymbol{v}_{k}$ when $k \neq j$ and orthogonal to the $I$-spline basis vectors. We assume

$$
\tau \sum v_{j i}^{2} \rightarrow \infty
$$

as $n \rightarrow \infty$. So as $n \rightarrow \infty, \tau M \sum_{i=1}^{n} v_{j i}^{2}$ dominates the one in the denominator in 2.28. Thus, as $n \rightarrow \infty$, the mean of full conditional distribution used in the Gibbs sampler approaches

$$
\frac{\sum_{i=1}^{n} v_{j i} y_{i}}{\sum_{i=1}^{n} v_{j i}^{2}}
$$

which is the ML SR estimate.
Using (2.6), we note that the variance of the conditional distribution for $\alpha_{j}$ given all other parameters and the data that is used in the Gibbs sampler $\alpha_{j}$ is given by

$$
\left[1 / M+\tau \sum_{i=1}^{n} v_{j i}^{2}\right]^{-1}
$$

Using the assumption that $\tau \sum v_{j i}^{2}$ goes to $\infty$ as $n \rightarrow \infty$, we have that this variance approaches zero as $n \rightarrow \infty$. Thus, the marginal distribution for each $\alpha_{j}$ approaches $\tilde{\alpha}_{j}$ as $n \rightarrow \infty$. By Geman and Geman (1984), we know that the marginalization of the target distribution is the limiting marginal distribution. Thus, the Bayes SRRS estimate for $\alpha_{j}$ converges to the ML SR as $n \rightarrow \infty$ and is therefore consistent.

### 2.7.3 Consistency of the Restricted Regression Coefficients

In this section, we consider the consistency of the restricted regression coefficients, $\beta_{l j}$ for $j=1, \ldots, m_{l}$ and $l=1, \ldots, L$. To simplify the notation, we let $\boldsymbol{\beta}=\left(\beta_{11}, \ldots, \beta_{L m_{L}}\right)^{\prime}=\left(\beta_{1}, \ldots, \beta_{q}\right)^{\prime}$ where $q=\sum_{l=1}^{L} m_{l}$. To show consistency, we consider the conditional distribution of $\boldsymbol{\beta}$ given all other parameters and the data that is sampled from in the Gibbs sampler, denoted $f(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \tau, \boldsymbol{y})$. We show that as $n \rightarrow \infty, f(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \tau, \boldsymbol{y})$ is a distribution with a mass point at the ML RS estimate for $\boldsymbol{\beta}, \tilde{\boldsymbol{\beta}}=\left(\tilde{\beta}_{11}, \ldots, \tilde{\beta}_{L m_{L}}\right)^{\prime}=\left(\tilde{\beta}_{1}, \ldots, \tilde{\beta}_{q}\right)^{\prime}=\left[\Delta^{\prime} \Delta\right]^{-1} \Delta^{\prime} \boldsymbol{y}$ where $\Delta$ is a $n$ by $q$ matrix with columns of the basis vectors as in (2.26). Thus the Bayes SRRS estimate for $\boldsymbol{\beta}$ approaches the ML SR as $n \rightarrow \infty$.

To determine the limiting behavior for $f(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \tau, \boldsymbol{y})$, we note that

$$
\begin{align*}
f(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \tau, \boldsymbol{y}) \propto & \prod_{j=1}^{q}\left[\beta_{j}^{c_{j}^{*}-1} I\left\{0<\beta_{j}<\infty\right\}\right] \\
& \exp \left\{\frac{-\tau}{2}(\boldsymbol{y}-\Delta \boldsymbol{\beta}-V \boldsymbol{\alpha})^{\prime}(\boldsymbol{y}-\Delta \boldsymbol{\beta}-V \boldsymbol{\alpha})-\boldsymbol{c}_{2}^{\prime} \boldsymbol{\beta}\right\} \tag{2.29}
\end{align*}
$$

where $V$ is a matrix with columns from the orthogonal set of vectors $\boldsymbol{v}_{j}$ for $j=$ $1, \ldots, p, c_{j}^{*}$ denotes the $c_{l 1}$ corresponding to $\beta_{j}$, and $\boldsymbol{c}_{2}=\left(\boldsymbol{c}_{21}, \ldots, \boldsymbol{c}_{2 L}\right)^{\prime}$ is a vector
of length $q$ created using vectors $\boldsymbol{c}_{2 l}$ for $l=1, \ldots, L$ where $\boldsymbol{c}_{2 l}$ is a vector of length $m_{l}$ with elements $c_{l 2}$. Since the columns of $\Delta$ are orthogonal to the columns of $V$,

$$
\begin{align*}
& \exp \left\{-\frac{\tau}{2}(\boldsymbol{y}-\Delta \boldsymbol{\beta}-V \boldsymbol{\alpha})^{\prime}(\boldsymbol{y}-\Delta \boldsymbol{\beta}-V \boldsymbol{\alpha})-\boldsymbol{c}_{2}^{\prime} \boldsymbol{\beta}\right\} \\
& \propto \exp \left\{-\frac{\tau}{2}(\boldsymbol{y}-\Delta \boldsymbol{\beta})^{\prime}(\boldsymbol{y}-\Delta \boldsymbol{\beta})-\boldsymbol{c}_{2}^{\prime} \boldsymbol{\beta}\right\} \\
& \propto \exp \left\{-\frac{\tau}{2}\left[\boldsymbol{\beta}^{\prime} \Delta^{\prime} \Delta \boldsymbol{\beta}-2 \boldsymbol{\beta}^{\prime}\left(\boldsymbol{y}^{\prime} \Delta-\frac{1}{\tau} \boldsymbol{c}_{2}\right)\right]\right\} \\
& \propto \exp \left\{-\frac{\tau}{2}\left[\boldsymbol{\beta}^{\prime} \Delta^{\prime} \Delta \boldsymbol{\beta}-2 \boldsymbol{\beta}^{\prime} \boldsymbol{a}\right]\right\} \\
& \propto \exp \left\{-\frac{\tau}{2}\left(\boldsymbol{\beta}-\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}\right)^{\prime} \Delta^{\prime} \Delta\left(\boldsymbol{\beta}-\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}\right)\right\} \\
& \propto \exp \left\{-\frac{1}{2}\left(\boldsymbol{\beta}-\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}\right)^{\prime} \Delta^{\prime} \Delta\left(\boldsymbol{\beta}-\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}\right)\right\} \tag{2.30}
\end{align*}
$$

where $\boldsymbol{a}=\Delta^{\prime} \boldsymbol{y}-\frac{1}{\tau} \boldsymbol{c}_{\mathbf{2}}$. The second to last line of (2.30) is obtained by noting that

$$
\begin{aligned}
& \left(\boldsymbol{\beta}-\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}\right)^{\prime} \Delta^{\prime} \Delta\left(\boldsymbol{\beta}-\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}\right) \\
& =\left(\boldsymbol{\beta}^{\prime}-\boldsymbol{a}^{\prime}\left[\Delta^{\prime} \Delta\right]^{-1}\right) \Delta^{\prime} \Delta\left(\boldsymbol{\beta}-\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}\right) \\
& =\left(\boldsymbol{\beta}^{\prime} \Delta^{\prime} \Delta-\boldsymbol{a}^{\prime}\left[\Delta^{\prime} \Delta\right]^{-1} \Delta^{\prime} \Delta\right)\left(\boldsymbol{\beta}-\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}\right) \\
& =\boldsymbol{\beta}^{\prime} \Delta^{\prime} \Delta \boldsymbol{\beta}-\boldsymbol{\beta}^{\prime} \Delta^{\prime} \Delta\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}-\boldsymbol{a}^{\prime} \boldsymbol{\beta}+\boldsymbol{a}^{\prime}\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a} \\
& =\boldsymbol{\beta}^{\prime} \Delta^{\prime} \Delta \boldsymbol{\beta}-2 \boldsymbol{\beta}^{\prime} \boldsymbol{a}+\boldsymbol{a}^{\prime}\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a} \\
& \propto \boldsymbol{\beta}^{\prime} \Delta^{\prime} \Delta \boldsymbol{\beta}-2 \boldsymbol{\beta}^{\prime} \boldsymbol{a} .
\end{aligned}
$$

We use (2.30) and that the diagonal elements of $\left[\Delta^{\prime} \Delta\right]^{-1}$ approach zero (shown in next two paragraphs) to show that $f(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \tau, \boldsymbol{y})$ has a mass point at $\tilde{\boldsymbol{\beta}}$ as $n \rightarrow \infty$.

Let $A=\Delta^{\prime} \Delta$ and note that $A$ is symmetric as well as positive definite so $A^{-1}$ is positive definite. From page 338 of Horn and Johnson (1991), we have the following corollary

Corollary 1. Let $\operatorname{diag}(A)=\left(a_{11}, \ldots, a_{n n}\right)^{\prime}$ denote the diagonal entries of the $n \times n$ square matrix $A, \operatorname{tr}(A)$ denote the trace of matrix $A$, and $\lambda_{1}(A) \leq \ldots \leq \lambda_{n}(A)$ be the eigenvalues of $A$ then

$$
\operatorname{tr}(A)=\sum_{i=1}^{n} a_{i i}=\sum_{i=1}^{n} \lambda_{i}(A)
$$

Using Corollary 1, we know that the sum of the eigenvalues of $A^{-1}=\left[\Delta^{\prime} \Delta\right]^{-1}$ are equal to the trace so $\operatorname{tr}\left(A^{-1}\right)=\sum_{i=1}^{q} \lambda_{i}\left(A^{-1}\right)$ where we denote the ordered eigenvalues of $A^{-1}$ by $\lambda_{i}\left(A^{-1}\right)$ for $i=1, \ldots, q$ with $\lambda_{1}\left(A^{-1}\right) \leq \lambda_{2}\left(A^{-1}\right) \leq \cdots \leq$ $\lambda_{q}\left(A^{-1}\right)$. Using that $\lambda_{i}\left(A^{-1}\right)=\left[\lambda_{i}(A)\right]^{-1}$ (Johnson et al., 2002, pg 302), we can also conclude $\operatorname{tr}\left(A^{-1}\right)=\sum_{i=1}^{q} \lambda_{i}\left(A^{-1}\right)=\sum_{i=1}^{q}\left[\lambda_{i}(A)\right]^{-1}$. Since $A$ is positive definite, $\lambda_{i}(A) \geq 0$ for all $i \in\{1, \ldots, q\}$ and $\sum_{i=1}^{q}\left[\lambda_{i}(A)\right]^{-1} \leq \sum_{i=1}^{q}\left[\lambda_{1}(A)\right]^{-1}=$ $q \cdot\left[\lambda_{1}(A)\right]^{-1}$. Thus,

$$
\begin{equation*}
\operatorname{tr}\left(A^{-1}\right) \leq q \cdot\left[\lambda_{1}(A)\right]^{-1} \tag{2.31}
\end{equation*}
$$

Consider Theorem 5.3.4 on page 316 of Horn and Johnson (1991) which states
Theorem 1. Let $A, B$ be $n \times n$ Hermitian matrices and let $A=\left[a_{i j}\right]$ be positive semi-definite. Any eigenvalues $\lambda(A \cdot B)$ of $A \cdot B$ satisfies

$$
\begin{aligned}
\lambda_{1}(A) \lambda_{1}(B) & \leq \min _{1 \leq i \leq n}\left\{a_{i i}\right\} \lambda_{1}(B) \\
& \leq \lambda(A \cdot B) \\
& \leq \max _{1 \leq i \leq n}\left\{a_{i i}\right\} \lambda_{n}(B) \leq \lambda_{n}(A) \lambda_{n}(B)
\end{aligned}
$$

where $A \cdot B$ is the dot product or Hadamard product of matrices $A$ and $B$.
Recall that a Hermitian matrix is a matrix whose conjugate transpose is that matrix. To apply this theorem to show that the diagonal entries of $A=\Delta^{\prime} \Delta$ approach $\infty$ as $n \rightarrow \infty$, we let $B$ be a $q \times q$ matrix with all off diagonal elements zero and all diagonal elements equal to -1 . This diagonal matrix is Hermitian with $\lambda_{1}(B)=-1$ and $\lambda_{q}(B)=-1 . A=\Delta^{\prime} \Delta$ is the form

$$
\left[\begin{array}{cccc}
\boldsymbol{\delta}_{1}^{\prime} \boldsymbol{\delta}_{1} & \boldsymbol{\delta}_{1}^{\prime} \boldsymbol{\delta}_{2} & \cdots & \boldsymbol{\delta}_{1}^{\prime} \boldsymbol{\delta}_{q}  \tag{2.32}\\
\boldsymbol{\delta}_{1}^{\prime} \boldsymbol{\delta}_{2} & \boldsymbol{\delta}_{2}^{\prime} \boldsymbol{\delta}_{2} & \cdots & \boldsymbol{\delta}_{2}^{\prime} \boldsymbol{\delta}_{q} \\
\vdots & \vdots & \ddots & \vdots \\
\boldsymbol{\delta}_{1}^{\prime} \boldsymbol{\delta}_{q} & \boldsymbol{\delta}_{2}^{\prime} \boldsymbol{\delta}_{q} & \cdots & \boldsymbol{\delta}_{q}^{\prime} \boldsymbol{\delta}_{q}
\end{array}\right]
$$

so $\max _{1 \leq i \leq q}\left\{a_{i i}\right\}=\boldsymbol{\delta}_{k}^{\prime} \boldsymbol{\delta}_{k}$ for some $k \in\{1, \ldots, q\}$. Using inequality $\lambda_{1}(A) \lambda_{1}(B) \leq$ $\max _{1 \leq i \leq n}\left\{a_{i i}\right\} \lambda_{n}(B)$ in Theorem 1, we can conclude $-\lambda_{1}(A) \leq-\boldsymbol{\delta}_{k}^{\prime} \boldsymbol{\delta}_{k}$ and therefore

$$
\begin{equation*}
\lambda_{1}(A) \geq \boldsymbol{\delta}_{k}^{\prime} \boldsymbol{\delta}_{k} \tag{2.33}
\end{equation*}
$$

If we assume $\boldsymbol{\delta}_{k}^{\prime} \boldsymbol{\delta}_{k} \rightarrow \infty$ as $n \rightarrow \infty$ for some $k \in\{1, \ldots, q\}$, then we can conclude $\lim _{n \rightarrow \infty} \lambda_{1}(A)=\infty$ and thus all diagonal entries of $\Delta^{\prime} \Delta$ approach $\infty$ as $n \rightarrow \infty$. Using $\lim _{n \rightarrow \infty} \lambda_{1}(A)=\infty, \lim _{n \rightarrow \infty}\left[\lambda_{1}(A)\right]^{-1}=0$. From (2.31), we have that $\operatorname{tr}\left(A^{-1}\right) \leq 0$ as $n \rightarrow \infty$ and since $A^{-1}$ is positive definite then the trace must go to zero as $n \rightarrow \infty$. Therefore, we have shown that the diagonal entries of $\left[\Delta^{\prime} \Delta\right]^{-1}$ are zero.

Now we show that $f(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \tau, \boldsymbol{y})$ will have a mass point at $\tilde{\boldsymbol{\beta}}$ as $n \rightarrow \infty$. First, we note that the trace of a positive semi-definite matrix is zero if and only if the matrix is a zero matrix (Horn, 1990, pg 80). Thus, since, as $n \rightarrow \infty,\left[\Delta^{\prime} \Delta\right]^{-1}$ is a positive semi-definite matrix with a trace of zero, we have that that $\left[\Delta^{\prime} \Delta\right]^{-1}$ approaches the zero matrix as $n \rightarrow \infty$. We then note that the exponential term in (2.30) is proportional to the pdf of a multivariate normal distribution with a mean

$$
\begin{equation*}
\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{a}=\left[\Delta^{\prime} \Delta\right]^{-1} \Delta^{\prime} \boldsymbol{y}-\frac{1}{\tau}\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{c}_{2} \tag{2.34}
\end{equation*}
$$

and variance-covariance matrix $\left[\Delta^{\prime} \Delta\right]^{-1}$ which we have shown approaches zero as $n \rightarrow \infty$. The mean in (2.34) approaches $\left[\Delta^{\prime} \Delta\right]^{-1} \Delta^{\prime} \boldsymbol{y}$ since $\frac{1}{\tau}\left[\Delta^{\prime} \Delta\right]^{-1} \boldsymbol{c}_{2}$ approaches zero vector. Since $\left[\Delta^{\prime} \Delta\right]^{-1} \Delta^{\prime} \boldsymbol{y}$ is the ML SR estimate, we have that the term in (2.30) is proportional to a multivariate normal pdf that has a mass point at $\tilde{\boldsymbol{\beta}}$ as $n \rightarrow \infty$. Since

$$
\begin{equation*}
\prod_{j=1}^{q}\left[\beta_{j}^{c_{j}^{*}-1} I\left\{0<\beta_{j}<\infty\right\}\right] \tag{2.35}
\end{equation*}
$$

from (2.29) does not depend on $n$, the exponential term in (2.30) denominates $f(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \tau, \boldsymbol{y})$. Thus, as $n \rightarrow \infty, f(\boldsymbol{\beta} \mid \boldsymbol{\alpha}, \tau, \boldsymbol{y})$ will have a mass point at $\tilde{\boldsymbol{\beta}}$. Again using Geman and Geman (1984), we know that the Bayes SRRS estimate converges to the unrestricted ML SR estimate as $n \rightarrow \infty$ so we have consistency of the Bayes SRRS estimates for the restricted coefficients.

### 2.7.4 Consistency of Precision

We conclude this section with the consistency for the precision estimate $\tau$. The normal errors model in Section 2.4 is a linear regression model with $\boldsymbol{Y}=X \boldsymbol{\theta}+\boldsymbol{\epsilon}$
with $\boldsymbol{\theta}$ and design matrix $X$ as defined in Section 2.7.2 and $\epsilon$ a $n \times 1$ vector of the independent identically distributed normal errors with mean zero and variance $\tau^{-1}$. For the unrestricted regression model where the parameters can take on any real value, the maximum likelihood estimate for the variance of the spline regression model is

$$
\tilde{\sigma}^{2}=\frac{\mathrm{S} \tilde{\mathrm{~S}} \mathrm{E}}{n}
$$

where $\mathrm{S} \tilde{\mathrm{S}} \mathrm{E}=(y-X \tilde{\boldsymbol{\theta}})^{\prime}(y-X \tilde{\boldsymbol{\theta}})$ with $\tilde{\boldsymbol{\theta}}$ the ML SR estimate as defined in Section 2.7.2. Using that $\tilde{\sigma}^{2}$ is a consistent estimator for $\sigma^{2}$ and that under certain regularity conditions, a continuous function of consistent MLE estimator is also consistent (Casella and Berger, 2002, p 470), we have that $\tilde{\tau}=n / \mathrm{SSE}=\left(\tilde{\sigma}^{2}\right)^{-1}$ is a consistent estimator for $\tau$.

From (2.7), $f(\tau \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{y})$ follows a $\operatorname{Gamma}\left(d_{1}+n / 2, d_{2}+\mathrm{SSE} / 2\right)$ with mean $\mathrm{E}(\tau \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{y})$ approaches

$$
\mathrm{E}(\tau \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{y})=\frac{d_{1}+n / 2}{d_{2}+\mathrm{SSE} / 2}
$$

where SSE is as in (2.7). Using that the ML SR estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ converge to their corresponding parameters, we have that S $\mathrm{S} E$ converges to SSE as $n \rightarrow \infty$. Thus, $\mathrm{E}(\tau \mid \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{y})$ approaches $\tilde{\tau}$ as $n \rightarrow \infty$. Further note that the variance for the distribution of $\tau$ given $\boldsymbol{\beta}, \boldsymbol{\alpha}$, and $\tau$ from (2.7) is given by

$$
\begin{equation*}
\frac{d_{1}+n / 2}{\left(d_{2}+\mathrm{SSE} / 2\right)^{2}} \tag{2.36}
\end{equation*}
$$

For least squares linear regression, the SSE for the maximum likelihood estimation is of order $n$ and is the minimum variance estimator. Thus, the SSE in (2.36) will be at least of order $n$. This implies that as $n \rightarrow \infty$, the variance of the conditional distribution for $\tau$ in the Gibbs sampler will be of order no larger than $n^{-1}$ and thus goes to zero. So as $n \rightarrow \infty$, the Bayes SRRS estimate for $\tau$ approaches the ML SR estimate which is known to be consistent. Therefore, the Bayes SRRS spline estimator is a consistent estimator for $\tau$.

### 2.8 Simulation Study

In this section, we perform a number of simulations to show that the proposed Bayes SRRS estimates have desirable properties and performance. To obtain the Bayes SRRS estimate, we fix the interior knot locations at $k$ equally spaced $x$ quantiles and use 5000 iterations for each run of the MCMC algorithms, with a burn-in of 500 and the priors for the model parameters described in Section 2.4.1. At each iteration, we compute an estimate for the regression function using the realizations of the parameters for that iteration and the function estimate is found by taking the mean of these 4500 estimates, $\eta_{i}^{(t)}$ (after removing burn-in) as in (2.18). The algorithms generally converged quickly unless the sample sizes were quite small or the error variance so large that the signal to noise ratio was very low. We performed a sensitivity analysis for the prior distributions about the parameters and found that the estimates are insensitive to a wide range of hyperparameter values.

### 2.8.1 Method Comparisons

For the normal errors model as in Section 2.4 with $L=1$, we compare the Bayes SRRS function estimates with maximum likelihood (ML) SRRS and the monotone shape-restricted P-spline method of Brezger and Steiner (2008) using their default priors and knot locations with a correction to their equation (19). The P-spline procedure estimates the regression function using penalized B-splines and imposes the monotone shape restriction by using truncated normal priors for the regression coefficients. For the P-spline estimation procedure, we used the recommended inverse Gamma priors for the smoothing parameter of the regression coefficients and the variance parameter for the independent normal errors with default hyperparameters of $a=b=0.001$. Following Brezger and Steiner (2008), we used 20 interior knots and ran the MCMC algorithm for 11,000 iterations throwing out the first

1000 as burn-in and taking every 10th iteration. To decrease the required computational time, for the single-move Gibbs sampler for the regression coefficients given in Brezger and Steiner (2008), we let $T=1$ and found similar convergence results as compared to using a larger values such as $T=10$.

The ML SRRS, $\tilde{\boldsymbol{\eta}}$, is found by minimizing $\sum_{i=1}^{n}\left(y_{i}-\eta_{i}\right)^{2}$ over the constraint set

$$
\begin{equation*}
\mathcal{C}=\left\{\boldsymbol{\eta}: \boldsymbol{\eta}=\sum_{l=1}^{L} \alpha_{l} \boldsymbol{v}_{l}+\sum_{j=1}^{m} \beta_{j} \boldsymbol{\delta}_{j}, \text { where } \beta_{j} \geq 0, j=1, \ldots, m\right\} \tag{2.37}
\end{equation*}
$$

The set of all possible linear combinations of the basis vectors such that the shape restriction is met make up a convex cone and the estimate for $\boldsymbol{\eta}$ can be found via a projection onto a linear subspaces by exploiting the properties of convex cones. In our analysis, an estimate was found using the hinge algorithm (Meyer, 2008).

For each method,

$$
\begin{equation*}
\mathrm{SMSE}=\sqrt{\frac{1}{\mathrm{~N} \cdot n} \sum_{j=1}^{\mathrm{N}} \sum_{i=1}^{n}\left\{f\left(x_{i}\right)-\hat{f}_{j}\left(x_{i}\right)\right\}^{2}} \tag{2.38}
\end{equation*}
$$

is used to compare function estimates to the function that generated the data, $f$, where $N$ is the number of data sets generated in the simulation study and $\hat{f}_{j}$ is the estimate for the regression function for the $(j)$ th simulated data set.

Data were simulated under eight simulations scenarios, with results given in Table 2.1. The regression functions in the study were the constant $(f(x)=0)$, slope 1 line $(f(x)=x)$, slope 5 line $(f(x)=5 x)$, sigmoid $(f(x)=5 \exp (10 x-5) /[1+$ $\exp (10 x-5)])$, truncated cubic $\left(f(x)=0\right.$ for $x \leq 0.6$ and $f(x)=(5 x-3)^{3}$ for $x>0.6$ ), and the truncated cubic function multiplied by 3 denoted 3 (trun. cubic). To create data sets, functions were evaluated at $n$ equally spaced $x$ values in $[0,1]$ and independent standard normal errors with a standard deviation of 1.5 were added. $N=1000$ data sets were simulated for each scenario. Regression functions for four of the scenarios are shown in Figure 2.2, along with typical data sets, to illustrate
the differences in the methods under comparison. The Bayes and ML SRRS use $k=2$ interior knots for $n=20$ data points, and $k=3$ for $n=50$. However, we find the Bayes SRRS function estimates to be robust to the number of interior knots. For example, the SMSE as in (2.38) is 0.20 for the Bayes SRRS estimate using $k=2$ equally spaced interior knots and the SMSE is 0.25 for Bayes SRRS using $k=4$ equally spaced interior knots for 500 data sets simulated from the slope 5 line with $n=50$ observations and normal errors with standard deviation of 1.5.

Figure 2.2 gives estimates using all three methods along with $95 \%$ pointwise HPD credible intervals for Bayes SRRS method constructed as in Section 2.6.1. The constant function, shown in Figure 2.2(a) as the dotted line, is challenging for the P-spline (dashed curve) because the truncated normal prior is too stiff to allow flat spots to be estimated well. The ML SRRS (dot-dash) is flat in the middle but will dip at the edges if pulled by the data. Because of the lack of penalization or prior information, the ML SRRS will rise sharply at either end if there are positive errors at the right or negative errors at the left. The Bayes SRRS (solid) fits flat or steep spots equally well due to its vague prior, but resists pulling at the edges. In Table 2.1, we see that the SMSE is over $30 \%$ smaller for the Bayes SRRS than for the other two methods.

Two lines were considered, with slopes 1 and 5 . The P-spline does well for these (Table 2.1, cases 2 and 3), because the prior pulls all fits towards a line. For the sigmoid function, $f(x)=5 \exp (10 x-5) /[1+\exp (10 x-5)]$, shown in Figure 2.2(c) and case 4 in Table 2.1, the P-spline is seen to be less flexible, while the ML and Bayes SRRS fits are quite close except at the right end where the ML SRRS is pulled upward. The SMSE for the Bayes SRRS is about $30 \%$ smaller than for the P-spline due to its greater flexibility, and about $15 \%$ smaller than that for the ML SRRS because of its resistance to edge effects. The truncated cubic $f(x)=(5 x-3)_{+}^{3}$ is flat over most of the range, then quite steep (Figure 2.2(d)). The P-spline estimate is outside the pointwise credible interval for the Bayes SRRS (shown as the shaded
area). The Bayes and ML SRRS both have sufficient flexibility to estimate this function well, and have considerably smaller SMSE than the P-spline for different choices of $n$ and signal to noise ratio.


Figure 2.2: For each plot, the solid line is the mean Bayes posterior estimate for the Bayes SRRS method and the shading represents the corresponding pointwise $95 \%$ credible band. The dot-dash curve is the ML SRRS estimate using the same basis functions. The dashed curve is the monotone $P$-spline with the the suggested amount of knots. The true function $f$ is shown as the dotted curve.

Using the procedure discussed in Section 2.6.1, pointwise credible intervals and pointwise prediction intervals for the P-spline and Bayes SRRS methods were constructed for the eight scenarios in the simulation study. The means of the interval lengths of $95 \%$ credible intervals and the $95 \%$ prediction intervals as well as the percent coverage at $x_{n / 2}$ are show in Figure 2.3(a) and (b), respectively. The coverage probabilities for the the Bayes SRRS are very close to target for all scenarios. Both

Table 2.1: Estimated SMSE using three different estimation procedures: Bayes SRRS, ML SRRS, and P-spline method.

| Case | Function | $n$ | ML SRRS | Bayes SRRS | P-spline |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | constant | 50 | 0.32 | 0.23 | 0.34 |
| 2 | slope 1 line | 50 | 0.35 | 0.30 | 0.27 |
| 3 | slope 5 line | 50 | 0.44 | 0.47 | 0.32 |
| 4 | sigmoid | 50 | 0.41 | 0.35 | 0.50 |
| 5 | 3(trun. cubic) | 20 | 0.94 | 0.95 | 3.87 |
| 6 | 3(trun. cubic) | 50 | 0.47 | 0.51 | 1.14 |
| 7 | trun. cubic | 20 | 0.62 | 0.66 | 1.51 |
| 8 | trun. cubic | 50 | 0.40 | 0.38 | 1.17 |

methods appear to perform well for the steeper line and the sigmoid function but the Bayes SRRS method appears to perform better in other cases, especially when considering the truncated cubic functions where the prediction and credible interval lengths for the P-spline method are larger than the Bayes SRRS method. Even with the larger credible interval length for the truncated cubic function, the percent coverage for P-spline method is much smaller than that of Bayes SRRS and the P -spline coverage is well below $95 \%$. The percent coverage for the credible interval for the P-spline method is also relatively small for the constant function and less than $90 \%$. For the prediction intervals, the percent coverage for both methods is similar and close to the target levels but interval lengths for the P-spline method are larger when considering the truncated cubic functions. Hence, the simulation study suggests that the prediction and credible intervals for Bayes SRRS method work well for the $x$-values in the middle and are superior to the $P$-spline method for many cases considered in the simulation study.

### 2.8.2 Parallel-curves Regression Model

Consider the regression model with one continuous and one categorical predictor variable with $r$ levels as in Section 2.4.2 where interest lies in the significance of the categorical predictor. If $r=2$, the test for significance of the categorical predictor may be accomplished through credible interval calculation. For $r>2$, we utilize the


Figure 2.3: Mean lengths (left y-axis, solid points) for Bayes SRRS method (circles) and P-spline method (triangles) as well as percent coverage (right y-axis, hollow points) for the the estimate of the regression function evaluated at $x_{n / 2}$ for each of the 8 cases in the simulation study. Horizontal dashed line denotes $95 \%$ coverage.
approximate normality of the joint posterior for the several $\alpha$ and perform a chisquared test with $r-1$ degrees of freedom as in Section 2.6.2 with null hypothesis that each $\alpha_{j}=0$. Table 2.2 gives rejection rates for tests under Bayes SRRS model with $r=2$ and Table 2.3 gives rejection rates for $r=3$. We generated 10,000 data sets using the slope 5 line and sigmoid functions given in Section 2.8.1 and shown in Figure 2.2(b) and (c), with independent standard normal errors and equally spaced $x$ values in $[0,1]$. For each combination of choices of regression function, sample size, and $\alpha_{j}$ values, we generated data where the $(i)$ th observation was assigned to levels of the categorical variable in rotation. For data sets of size less than $n=100$, we used $k=2$ interior knots; otherwise we used $k=3$.

In Table 2.2 and 2.3, we compare rejection rates for inference using Bayes SRRS to those of the standard $F$-test for the significance of the categorical predictor, which assumes that the relationship between the response and the continuous predictor is linear. The alternative hypothesis value(s) for $\alpha_{j}$ were selected such that when $f$ is linear, the $F$-test will have power of $50 \%$. The results show that when the true
regression function is linear, the Bayes method performs just as well as the as the $F$-test. Otherwise, the Bayes method correctly rejects the one-curve model in favor of the parallel-curves model more often than the $F$-test. For the sigmoid function, the scatter plot often "looks linear" (Figure 2.2(c)) so that a practitioner might feel confident using the $F$-test but as this simulation results suggest, this may result in loss in power.

Table 2.2: Proportion of data sets for which the null hypothesis of a single curve is rejected when using Bayes SRRS model and a credible interval for $\alpha_{1}$ (Bayes) compared to the $F$-test.

| $f$ | $n$ | $\alpha_{2}$ | $F$-test | $B a y e s$ |
| :---: | :---: | :---: | :---: | :---: |
| slope 5 line | 40 | 0 | 0.051 | 0.050 |
| slope 5 line | 100 | 0 | 0.051 | 0.050 |
| sigmoid | 40 | 0 | 0.031 | 0.040 |
| sigmoid | 100 | 0 | 0.032 | 0.050 |
| slope 5 line | 40 | 0.637 | 0.497 | 0.506 |
| slope 5 line | 100 | 0.395 | 0.498 | 0.500 |
| sigmoid | 40 | 0.637 | 0.368 | 0.447 |
| sigmoid | 100 | 0.395 | 0.391 | 0.498 |

Table 2.3: Proportion of data sets for which the null hypothesis of a single curve is rejected when using Bayes SRRS model and the chi-square test statistic given in Section 2.6.2 (Bayes) compared to the $F$-test.

| $f$ | $n$ | $\alpha_{2}$ | $\alpha_{3}$ | $F$-test | $B a y e s$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| slope 5 line | 60 | 0 | 0 | 0.048 | 0.050 |
| slope 5 line | 150 | 0 | 0 | 0.051 | 0.053 |
| sigmoid | 60 | 0 | 0 | 0.040 | 0.049 |
| sigmoid | 150 | 0 | 0 | 0.041 | 0.054 |
| slope 5 line | 60 | 0.543 | 1.085 | 0.500 | 0.518 |
| slope 5 line | 150 | 0.337 | 0.675 | 0.500 | 0.509 |
| sigmoid | 60 | 0.543 | 1.085 | 0.403 | 0.462 |
| sigmoid | 150 | 0.337 | 0.675 | 0.427 | 0.500 |

### 2.8.3 Non-parallel Curves Regression Model

Consider the non-additive normal error regression model given in Section 2.4.3 and the simpler additive model given in Section 2.4.2. To determine whether the
non-additive normal error regression model explains significantly more of the variation in the response for given data sets, we employ the model selection technique using Bayes factors given in Section 2.6.3. We performed a simulation study to illustrate the performance of the method in this context where we have one continuous variable and a categorical variable with $r=2$ levels. We chose four sets of regression functions; for the first two, the correct choice is the additive model (Section 2.4.2), and for the second two, the correct choice is the model with interaction (Section 2.4.3). The "parallel lines" regression curves are $f(x)=3 x$ and $f(x)=3 x+0.75$, and the "parallel sigmoids" models are $f(x)=3 \exp (10 x-5) /(1+\exp (10 x-5))$ and $f(x)=3 \exp (10 x-5) /(1+\exp (10 x-5))+0.75$. The "non-parallel lines" models are $f(x)=3 x$ and $f(x)=x+0.75$, and the "non-parallel sigmoids" models are $f(x)=$ $3 \exp (10 x-5) /(1+\exp (10 x-5))$ and $f(x)=5 \exp (10 x-8) /(1+\exp (10 x-8))+0.75$. We generated 10,000 data sets with equally spaced $x$ values and alternating assignments of the categorical variable values. Table 2.4 gives percent of runs for which the correct model is selected using the $F$-test compared with the Bayes SRRS model and approximate BF discussed in Section 2.6.3. Again, the $F$-test assumes a linear relationship between the response and the continuous covariate.

When the data are generated from the "parallel line" or "parallel sigmoids" cases, the proportion of data sets for which the correct model is chosen is relatively large for both the $F$-test and the Bayes SRRS. For the "non-parallel lines" case, the $F$ test correctly selects the interaction model more often than the Bayes SRRS method and the approximate BF. However, for "non-parallel sigmoids" case, the $F$-test does not recognize the interaction between the sigmoid curves, while the approximate BF method performs very well, especially for the larger sample size. A typical data set generated from "non-parallel sigmoids" case is shown in Figure 2.4(a) for $n=100$. There is no obvious deviation from linearity, so that a practitioner might feel comfortable using the $F$-test. Yet, as shown in Figure 2.4(b), the interaction between the predictors would be missed. This example demonstrates the dangers of
choosing a parametric model when the only valid assumptions are smoothness and monotonicity.


Figure 2.4: A scatter plot for a typical data set simulated from non-parallel sigmoids. (b) The linear (dashed lines), and Bayes SRRS estimate (solid lines) alone with the sigmoid functions used to generate the data (dot-dashed lines) and simulated data.

Table 2.4: Table gives the proportion of correct model selections using both Bayes factors and the $F$-test with a significance level 0.05.

| Case | $n$ | $F$-test | Bayes |
| :---: | :---: | :---: | :---: |
| parallel lines | 40 | 0.949 | 0.950 |
| parallel lines | 100 | 0.951 | 0.914 |
| parallel sigmoids | 40 | 0.963 | 0.924 |
| parallel sigmoids | 100 | 0.958 | 0.925 |
| non-parallel lines | 40 | 0.443 | 0.380 |
| non-parallel lines | 100 | 0.827 | 0.680 |
| non-parallel sigmoids | 40 | 0.031 | 0.267 |
| non-parallel sigmoids | 100 | 0.045 | 0.951 |

### 2.8.4 Model Selection

Practitioners typically prefer to use a simple parametric model, even when the a priori assumptions are more qualitative such as "smooth and increasing." Bayes factors can be used to determine whether a parametric model or the less restrictive
model imposing shape and smoothness is appropriate. In this section, we report the results of a simulation study to examine the small sample performance of the Bayes SRRS model and Bayes factors when a more flexible shape-restricted regression spline model is compared to a parametric function.

Bayes SRRS and approximate Bayes factors can also be used to compare an exponential growth model to a more flexible convex and increasing model. While exact tests for the case of constant versus increasing and linear versus convex exist in the frequentist paradigm for the normal errors model as discussed in Meyer (2008), a more general test has not yet been proposed.

### 2.8.4.1 Constant versus Increasing Regression Function

It is often of interest to examine the effect of a predictor variable, under the assumption of monotonicity. We performed a simulation study to examine the performance of our model selection methodology described in Section 2.8.1 comparing a monotonically increasing normal errors model with one continuous covariate ( $M_{1}$ ) to the constant model $\left(M_{2}\right)$. Note that the constant model is a subset of the monotonically increasing model with the $\beta_{j}$ parameters set to zero. Thus, using method discussed in Section 2.8.1, we compute $\hat{p}\left(\boldsymbol{y} \mid M_{1}\right)$ for the monotonically increasing model and $\hat{p}\left(\boldsymbol{y} \mid M_{2}\right)$ for a constant model with $\beta_{j}=0$ for all $j$, a normal prior for $\alpha$, and a gamma prior for $\tau$. The proportion of correct model selections as measured by the approximate BF for sample sizes of $n=40$ and $n=100$ using $k=2$ and $k=3$ equally spaced interior knots, respectively, is shown in Table 2.5. The functions considered in this analysis are the constant function $f(x)=3$, the linear function $f(x)=1.25 x$, and the exponential function $0.00007 \exp (10 x)$. The latter is flat over most of the range of $(0,1)$, then increases steeply at the right hand side. We generated 10,000 data sets with independent standard normal errors with standard deviation 1 and equally spaced $x$ values in $[0,1]$. For the larger sample sizes, the approximate BF selects the correct model for a relatively large proportion of the data sets but does not do as well for the smaller sample sizes.

Table 2.5: Simulations to examine the performance of inference concerning the constant versus increasing regression function models. The $\%$ correct columns give the percentage of simulated data sets where the correct model was selected based on approximate BFs.

| $f$ | $n$ | $\%$ correct |
| :---: | :---: | :---: |
| constant | 40 | 97.8 |
| constant | 100 | 98.2 |
| linear | 40 | 56.2 |
| linear | 100 | 91.5 |
| exponential | 40 | 38.0 |
| exponential | 100 | 75.1 |

### 2.8.4.2 Linear versus Monotone Regression Function

We also consider the simple example $y_{i}=f\left(x_{i}\right)+\epsilon_{i}$ and the question of whether $f$ can be modeled as a simple line, when it is known a priori that $f$ is increasing and smooth. We compare model selection performance of the parametric linear versus shape-restricted monotone regression function. Thus, we compute $\hat{p}\left(\boldsymbol{y} \mid M_{1}\right)$ for the monotonically increasing model and $\hat{p}\left(\boldsymbol{y} \mid M_{2}\right)$ for linear model and not that if $f$ is truly linear and the basis functions are scaled to have a range of one, then all the coefficients of the basis functions are equal giving $\boldsymbol{\eta}=\beta \sum_{j=1}^{m} \boldsymbol{\delta}_{j}+\alpha$. We adopt the same priors for $\alpha$ and $\tau$ in the linear model as were used in the monotone model discussed in Section 2.4 and the same prior for $\beta$ as was used for each $\beta_{j}$ in monotone model. The two regression functions were the slope-5 line and sigmoid discussed in Section 2.8.1, and the sample sizes are $n=40$ and 100. The proportions of correct choices using the approximate BF as described in Section 2.6.3 are shown in Table 2.6. When the true regression function is linear, the method produces the correct choice over $90 \%$ of the time for both sample sizes. Otherwise, the percent correct choice is larger for larger samples. This is similar to a frequentist testing situation where the linear model constitutes the null hypothesis, and the power grows with the sample size. With the smaller sample size, the model selection method does a poor job determining whether the function is a line or simply monotonically
increasing. However, this is not surprising considering the example in Figure 2.2(c).

Table 2.6: Simulations to examine the performance of inference concerning the linear versus increasing regression function models. The $\%$ correct columns give the percentage of simulated data sets where the correct model was selected based on approximate BFs.

| $f$ | $n$ | $\%$ correct |
| :---: | :---: | :---: |
| slope 5 line | 40 | 92.2 |
| slope 5 line | 100 | 92.0 |
| sigmoid | 40 | 58.1 |
| sigmoid | 100 | 94.0 |

### 2.8.4.3 Constrained versus Unconstrained Regression Function

Often the shape assumptions are included in the a priori information, but occasionally the shape assumptions are part of the research question. The unconstrained regression spline model may be fit by changing the prior distribution for the $\beta_{j}$ coefficients to the normal prior with large variance and zero mean. We compare the mean of the MCMC realizations of $\boldsymbol{\beta}$ using the unrestricted model $\left(\hat{\boldsymbol{\beta}}_{1}\right)$ to the mean of the MCMC realizations using the monotonically increasing shape-restricted model ( $\hat{\boldsymbol{\beta}}_{0}$ ) by constructing $\chi^{*}=\left(\hat{\boldsymbol{\beta}}_{1}-\hat{\boldsymbol{\beta}}_{0}\right)^{\prime} S_{1}\left(\hat{\boldsymbol{\beta}}_{1}-\hat{\boldsymbol{\beta}}_{0}\right)$ where $S_{1}$ is the sample covariance matrix realizations of $\boldsymbol{\beta}_{1}$. Since the joint posterior distribution of $\boldsymbol{\beta}_{1}$ is approximately normal, $\chi^{*}$ is compared with the 95 th percentile of the $\chi^{2}(m)$ distribution (where $m$ is the number of basis vectors) to determine if $\hat{\boldsymbol{\beta}}_{0}$ is within the $95 \%$ credible ellipsoid for $\boldsymbol{\beta}_{1}$.

Table 2.7 gives the proportions of correct model selections using $\chi^{*}$ for four choices of the underlying regression function: the constant function $f(x)=0$ and the functions $f_{a}(x)=10\left[1+x-a \exp \left\{-(x-0.5)^{2} / 0.02\right\}\right]$ with $a=0.15,0.25$, and 0.45 . The functions $f_{a}$ (shown in Figure 2.5) are the functions of Bowman et al. (1998) scaled by 10 to allow for comparison between our and their simulation results (they use $\sigma=0.1$; we use $\sigma=1$ ). Their method for testing monotonicity of
a regression function involves a critical bandwidth. Note that the constant function and $f_{.15}$ are non-decreasing, while $f_{.25}$ has a slight dip in the center of the range, and $f_{.45}$ has a more pronounced dip. We considered $n=50$ and $n=100$ and four interior knots. The method using $\chi^{*}$ is conservative, with small proportions of rejection for the two non-decreasing scenarios and also for the scenario with a small decreasing interval. The method using $\chi^{*}$ has higher percent correct than found for Bowman et al. (1998) for the steeper $f_{.45}$. When the dip in the true function is more pronounced $(a=0.45)$, the chi-square test selects the correct model for a high proportion of the data sets even when $n=50$. Figure 2.5(c) illustrates how the unrestricted Bayes regression estimate is much closer to $f_{0.45}$ than the Bayes SRRS estimate and appears a more appropriate estimate since it captures the dip in data points around 0.5 . The lower percent correct for method using $\chi^{*}$ than for Bowman et al. (1998) for $f_{.25}$ is supported by the simulated data and estimates shown in Figure 2.5(b). It illustrates how both the Bayes SRRS and the unrestricted Bayes regression spline estimate fail to capture the small dip around 0.5 in $f_{.25}$. Figure $2.5(\mathrm{~b})$ also shows how difficult it is to visually detect a dip in data points around 0.5. Thus, supporting that it is hard to conclude that the unrestricted model is more appropriate.

Table 2.7: Simulations to examine the performance of inference using the Bayes SRRS model concerning the monotonically increasing shape assumption. The $\%$ correct columns give the percentage of simulated data sets where the correct model was selected based on $\chi^{*}$.

| $f$ | $n$ | $\chi^{*} \%$ correct | Bowman et al. \% correct |
| :---: | :---: | :---: | :---: |
| constant | 50 | 99.6 | - |
| constant | 100 | 99.4 | - |
| $f_{0.15}$ | 50 | 100.0 | 98.2 |
| $f_{0.15}$ | 100 | 100.0 | 99.2 |
| $f_{0.25}$ | 50 | 2.0 | 10.0 |
| $f_{0.25}$ | 100 | 6.1 | 17.4 |
| $f_{0.45}$ | 50 | 80.9 | 54.4 |
| $f_{0.245}$ | 100 | 99.8 | 87.4 |



Figure 2.5: Scatter plots for typical data sets simulated from $f_{a}(x)=$ $10\left[1+x-a \cdot \exp \left\{-(x-0.5)^{2} / 0.02\right\}\right]$ with $a=0.15$ (plot a), $a=0.25$ (plot b), and $a=0.45$ (plot c) along with the functions used to generate the data (dot-dashed lines), the Bayes SRRS estimate under the monotone assumption (solid lines) and the Bayes unrestricted regression spline model estimate (dashed lines).

### 2.9 Examples

In this section, we apply the Bayes SRRS model to estimate regression functions for three real data sets. We use the normal errors model to model fertility rate in third world countries as a monotonically decreasing function of contraceptive use and use the normal errors model with categorical covariate to model $\log$ of onion yield (grams per plant) as a monotonically decreasing function of areal density of plants (plants per square meter) for onions grown in two locations. We use the Bernoulli model to estimate risk of diabetes as a monotonically increasing function of cholesterol.

### 2.9.1 Normal Errors Examples

We adopt the normal errors model as in Section 2.4 to estimate the mean function when analyzing the "Robey" data set available in the car library in the R program (R Development Core Team, 2011). The data set obtained from Robey et al. (1992) includes total fertility rate (children per woman), percent of married women of childbearing age who use contraceptives, and a region variable that has
levels of Africa, Asia and Pacific, Latin America and Caribbean, and Near East and North Africa. We estimate the mean fertility rate as a function of married women of childbearing age who use contraceptives. We assume a normal errors model and that the mean function is monotonically decreasing. Figure 2.6 gives the mean function estimated using Bayes SRRS (solid black line) and quadratic $I$-splines with each knot point indicated by a red "X" along with $95 \%$ pointwise HPD intervals (the shaded region) computed as in Section 2.6.1. Figure 2.6 also gives the estimated mean function under a simple linear regression model (dashed blue line). Notice that the two estimates are rather similar but differ for contraceptive use between 5 and 40 percent with the Bayes SRRS estimate lower for contraceptive use between 5 and $20 \%$ and higher for contraceptive use between 25 and 55 percent. However, the simple linear regression estimate is within the $95 \%$ HPD for each data point. We use the model selection procedure described in Section 2.6.3 to determine whether the monotonically decreasing model or the simple linear model is most appropriate for the data. The Bayes factor equals 0.19 suggesting that a simple linear regression model is appropriate for this data set.

To illustrate how we can use our model to perform inference on a categorical covariate, we adopt the normal errors model with a categorical covariate (as in Section 2.4.2) to estimate the mean function for the "onions" data set available in the SemiPar library in the R program (R Development Core Team, 2011). The data set measures onion yield (grams per plant) and areal density of plants (plants per square meter) from an experiment studying the production of white Spanish onions at two South Australian locations (Virginia and Purnong Landing). This data set is described in Ratkowsky (1983) and also discussed in Ruppert et al. (2003). It has 84 sets of observations with 42 from Virginia and 42 from Purnong Landing. We model log of onion yield as a monotonically decreasing function of areal density of plants with an indicator function for location that equals 1 if grown in Virginia. Figure 2.7(a) gives the Bayes SRRS estimate for each location (solid

## Fertility Rate Example



Figure 2.6: The Bayes SRRS estimate (solid black line) for the monotonically decreasing model along with the $95 \%$ pointwise HPD intervals (shaded region) where each knot point is indicated by a red "X." The simple linear regression estimate is given by the dashed blue line.
lines) along with $95 \%$ HPD pointwise intervals for each function (shaded regions). The linear regression model estimates are given by the blue dashed lines. Note that the Bayes SRRS estimates for areal density less than 50 estimate lie between the data points for both locations while the linear estimates lie below the data points for both locations. Thus, the Bayes SRRS model appears to model the data better than the linear regression model. The posterior distribution for the coefficient on the indicator for Virginia location, $\alpha_{2}$, is shown in Figure 2.7 (b). The vertical green dashed lines mark the $95 \%$ HPD interval computed as in Section 2.6.1. Note that it does not include zero suggesting that onion yield does depend on location.

### 2.9.2 Bernoulli Example

As an example of implementing the Bernoulli model described in Section 2.5.1, we consider a study looking at cardiovascular risk factors for African Americans in central Virginia (Willems et al., 1997). This study involved measuring 19 factors believed to contribute to heart disease including total cholesterol, stabilized


Figure 2.7: (a) The Bayes SRRS estimate (solid black line) for the monotonically decreasing model for onion data set along with the $95 \%$ pointwise HPD intervals (shaded region) where each knot point is indicated by a red "X." The simple linear regression estimate is given by the dashed blue line. (b) Histogram of posterior draws after discarding burn-in for coefficient on indicator for the Virginia location, $\alpha_{2}$ where the vertical dashed lines mark the $95 \%$ HPD computed as in Section 2.6.1.
glucose, high density lipoprotein (hdl), cholesterol/hdl ratio (total cholesterol over hdl), glycosolated hemoglobin, age, gender, height, weight, and postprandial time when labs were drawn. The data were obtained from the website http://biostat. mc.vanderbilt.edu/twiki/bin/view/Main/DataSets.

We focus on modeling risk of diabetes as a monotonically increasing function of cholesterol/hdl ratio ignoring other covariates to simplify the analysis. A glycosolated hemoglobin greater than 7.0 is an indicator of diabetes so an indicator variable for diabetes was created and subjects with a hemoglobin of 7 or larger were coded as 1. Patients that were missing values for glycosolated hemoglobin or cholesterol/hdl ratio were discarded from the analysis as well as one individual who had a cholesterol/hdl ratio of 19.3 leaving a sample size of 388 . Figure 2.8 gives the estimated mean function using Bayes SRRS (solid line) with each knot point indicated by a red "X" and $95 \%$ pointwise HPD intervals (shaded region) computed as in Section 2.6.1. The risk of diabetes seems to increase at a larger rate for cholesterol/hdl ratios greater than about 7 . However, as indicated by the point wise HPD intervals,
we are less certain about these estimates most likely due to the fact that we have less observations for the higher cholesterol/hdl ratios. Considering the $95 \%$ HPD at cholesterol/hdl ratio of 5.0, this analysis suggests that having a cholesterol/hdl ratio of below 5.0 (vertical dotted line) will keep the risk of diabetes below $20 \%$ (horizontal dotted line) for African Americans in central Virginia.


Figure 2.8: The Bayes SRRS spline estimate for risk of diabetes as a function of cholesterol/hdl ratio along with $95 \%$ HPD computed as in Section 2.6.1 and data points indicated by vertical bars. Each knot point is indicated by a red "X" and the vertical dotted line marks a cholesterol/hdl ratio of 5 . The horizontal dotted line marks a 0.2 probability of diabetes and the estimate using logistic regression is given by blue dashed line.

## Chapter 3

## BAYESIAN SHAPE-RESTRICTED REGRESSION SPLINES AND MIXED MODELS

### 3.1 Introduction to Generalized Linear Model

In this chapter, we extend the Bayes SRRS model to include variance components and introduce a generalized linear mixed model (GLMM). Parameter estimation for GLMMs is challenging because the maximum likelihood estimates are not available in closed form. Generalized linear mixed models have been studied intensively both using frequentist (Breslow and Clayton, 1993; McCulloch and Searle, 2001; Rice and Wu, 2001; Zeger and Diggle, 1994; Verbyla et al., 1999; Lin and Zhang, 1999; Zhang, 2004; McCulloch, 1997) and Bayesian (Tiao and Tan, 1965; Box and Tiao, 1968; Gelman, 2006; Namata et al., 2007; Natarajan and Kass, 2000; Gelfand et al., 1995, 1996) frameworks. Breslow and Clayton (1993) examined the performance of penalized quasi-likelihood (PQL) for function estimation for generalized linear mixed models. Walker (1996) used an EM algorithm (Dempster et al., 1977) for function estimation for a nonlinear random effects model. McCulloch (1997) proposed a Monte Carlo Newton-Raphson algorithm which is a Monte Carlo version of the EM algorithm to calculate maximum likelihood estimates for GLMMS. Zeger and Diggle (1994) proposed a semi-parametric model for longitudinal data that models the trend across time using a locally adaptive kernel estimate. They used a back-fitting algorithm and cross validation to estimate the regression functions. Rice and Wu (2001) proposed a nonparametric method for estimation of mixed models using $B$-splines to model both the fixed and random effects and
estimate the parameters using the EM algorithm. Verbyla et al. (1999) used cubic smoothing splines to fit data for generalized linear models and estimated the parameters using Average Information REML (Gilmour et al., 1995). Namata et al. (2007) used a generalized linear mixed model along with penalized splines and penalized quasi-likelihood estimation (Ruppert et al., 2003) to estimate age-specific rates at which susceptible individuals contract an infection. Zhang (2004) also proposed a semi-parametric model for generalized linear mixed models that used penalized splines and suggested using double penalized quasi-likelihood estimation (Lin and Zhang, 1999) to estimate parameters.

Other researchers have considered analyzing linear and generalized linear mixed models in a Bayesian setting. Gelfand et al. (1995) proposed a Bayesian model to estimate normal linear mixed models using hierarchical centering. Hierarchical centering is discussed further in Section 3.3. Gelfand et al. (1996) extended this Bayesian model to generalized linear mixed models. Vines et al. (1996) proposed a Bayesian random effects model and obtained function estimates using Gibbs sampling. Zhang et al. (1998) proposed a semi-parametric model for mixed model for longitudinal data using penalized natural cubic splines and compared both Bayesian and frequentist inference methods for this model. Thompson and Rosen (2008) used a Bayesian model and modeled curves as a linear combinations of B-splines with random coefficients. The curves are estimated using a MCMC algorithm.

It often can be difficult to estimate the random effect variance components for Bayesian models without using subjective priors (Natarajan and Kass, 2000; Gelman, 2006). Natarajan and Kass (2000) considered the performance of Bayesian estimators for generalized linear mixed models and proposed two different priors for the random effect variance component, that along with uniform priors for the fixed effects, lead to proper posterior distributions with desirable properties under given conditions. Gelman (2006) considered some non-informative prior distributions for
hierarchical variance parameters and examined their performance for several different hierarchical Bayesian models. He also proposed a half- $t$ model and considered models where it might be useful as a weakly-informative prior for variance parameters.

Other researchers have proposed function estimation procedures for generalized linear mixed models under shape restrictions. Ghosh (2007) considered using a semi-parametric regression model and smoothing splines to estimate regression functions for mixed models for binary data under the monotone shape restriction. He estimated the function using a two step procedure. He first used a likelihoodbased algorithm as in Lin and Zhang (1999) and then projected it on monotone function using the PAVA algorithm (Robertson et al., 1988). Brezger and Steiner (2008) proposed a random intercept model for function estimation under the monotone shape restriction that used penalized splines and a Bayesian framework. The monotone shape restriction was imposed by requiring the spline coefficients to be ordered and using truncated normal priors for the spline coefficients. Schipper et al. (2007) used a functional mixed model and a Bayesian framework to model effects of radiation dose on normal tissue complications. They estimated the weight function using regression splines and restricted the weight function in the functional mixed model to be non-decreasing using either linear splines or quadratic $I$-splines and a mixture of point mass at zero and gamma random variables. Our work also uses $I$ splines and gamma random variables to impose the monotone shape restriction but uses a generalized linear mixed model as opposed to a functional mixed model. Our model allows a researcher to impose other shape restrictions besides monotonicity such as convexity. Hazelton and Turlach (2011) considered using MCMC and penalized regression splines to estimate regression functions for mixed models with shape restrictions. They imposed the shape restrictions by using linear constraints on the coefficients. They used either linear or quadratic regression splines (for monotone shape restriction), or cubic splines (for convexity). The linear constraints on spline
coefficients are imposed using a Bayesian framework with truncated multivariate normal priors. We also consider a Bayesian framework to estimate regression functions for mixed models with shape restrictions such as monotonicity or convexity but use the shape-restricted regression splines of Meyer (2008) as opposed to penalized regression splines. This allows the shape restrictions to be imposed by simply requiring the spline coefficients to be positive and does not require the multivariate truncated prior for the spline coefficients, which simplifies the MCMC algorithm.

In Section 3.2, we consider the generalized additive mixed model. We propose a Bayesian framework to estimate generalized additive mixed models with shape restrictions using either $I$-splines or $C$-splines, depending on the shape restriction. We focus on the random intercept model under the monotone shape restriction in Section 3.3 and conclude the chapter with some examples in Section 3.4.

### 3.2 Shape-restricted Splines Generalized Additive Mixed Model

We consider the generalized additive shape-restricted regression spline model as defined in Section 2.2 and extend this to a generalized linear additive mixed model by adding random effect terms. In particular, suppose we have conditionally independent observations $y_{1}, \ldots y_{n}$ such that, conditional on a random effects vector $\boldsymbol{a}=\left(a_{1}, \ldots, a_{q}\right)$, the conditional distribution of $y_{i}$ given $\boldsymbol{a}$ is a member of the exponential family with

$$
f\left(y_{i} \mid \boldsymbol{a}\right)=\exp \left\{\left[y_{i} \theta_{i}-b\left(\theta_{i}\right)\right] / h_{i}(\phi)+c_{i}\left(y_{i}, \phi\right)\right\}
$$

where $b(\cdot), h_{i}(\cdot)$, and $c_{i}(\cdot, \cdot)$ are known functions and $\phi$ is a dispersion parameter (Jiang, 2007). As in (2.1), $b^{\prime}\left(\theta_{i}\right)=\mu_{i}$ and $g\left(\mu_{i}\right)=\eta_{i}$ where $g(\cdot)$ is a given link function and $\mu_{i}$ is the expected value of $Y$ for $x_{i}$. For the mixed model, we redefine $\eta_{i}$ to include the random effects and let

$$
\begin{equation*}
\eta_{i}=f_{1}\left(x_{1 i}\right)+\cdots+f_{L}\left(x_{L i}\right)+\boldsymbol{z}_{i}^{\prime} \boldsymbol{\gamma}+\boldsymbol{w}_{i}^{\prime} \boldsymbol{a} \tag{3.1}
\end{equation*}
$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{z}_{i}$ are defined as is (2.2) and $\boldsymbol{w}_{i}$ is the vector of known values corresponding to the random effects, $\boldsymbol{a}$. Each $f_{l}$ for $l=1, \ldots, L$ is assumed to be a continuous and smooth function of a predictor $x_{l i}$ for $i=1, \ldots, n$ where $n$ is the number of observations for predictor $x_{l}$.

As with the generalized linear model in Chapter 2 , we assume $f_{l}$ follows some shape restriction and model each shape-restricted function using the appropriate shape-restricted regression splines. For a given covariate vector of $n$ observations given by $\boldsymbol{x}_{l}=\left(x_{l 1}, \ldots, x_{l n}\right)^{\prime}$ and set of $m_{l}$ knot points given by $\min \left(\boldsymbol{x}_{l}\right)=t_{0}<\cdots<$ $t_{m_{l}}=\max \left(\boldsymbol{x}_{l}\right)$, we create basis vectors $\boldsymbol{\delta}_{l j}=\left(\delta_{l j 1}, \ldots, \delta_{l j n_{l}}\right)^{\prime}$ with $\delta_{l j i}=s_{l j}\left(x_{l i}\right)$. Note the basis vectors will depend on the shape restriction. For instance, we would use basis vectors created using quadratic $I$-spline for the monotone shape restriction. Given these basis vectors, we model $\eta_{i}$ in (3.1) as

$$
\begin{equation*}
\sum_{l=1}^{L} \sum_{j=1}^{m_{l}} \beta_{l j} \delta_{l j i}+\sum_{j=1}^{p} \alpha_{j} v_{j i}+\sum_{j=1}^{q} a_{j} w_{j i} \tag{3.2}
\end{equation*}
$$

with $\boldsymbol{v}_{j}=\left(v_{j 1}, \ldots, v_{j n}\right)^{\prime}$ as in (2.3) and $\boldsymbol{w}_{j}=\left(w_{j 1}, \ldots, w_{j n}\right)^{\prime}$ the vector of known values corresponding to random effect $a_{j}$.

As in Chapter 2, we estimate the regression functions using a Bayesian framework. Since we again restrict the $\beta$ parameters to be positive, we will use independent gamma priors with a shape parameter equal to $c_{1 l}$ and rate parameter equal to $c_{2 l}$ so $\beta_{l j} \sim \operatorname{Gamma}\left(c_{1 l}, c_{2 l}\right)$ for $l=1, \ldots L$ and $j=1, \ldots, m_{l}$. Since each $\alpha_{j}$ can take on any real value, we again use independent vague normal priors with mean zero and large variance. For the random effect parameters, $a_{j}$, we assume independent normal priors with mean $\mu_{a}$ and variance $\sigma_{a}^{2}$ where depending on the model, these parameters may be known or unknown.

### 3.3 Random Intercept Normal Errors Model

For the remainder of the chapter, we focus on a random intercept normal errors model and the monotone shape restriction. Extensions to the convex/concave shape
restriction can be made by replacing $I$-splines with $C$-splines. Monotone random intercept models are useful tools to model regression relationships for several different types of data sets. For example, the monotone shape restriction could be assumed if we are modeling growth of trees exposed to different growing conditions across time. We would expect that true relationship to be monotone increasing across time so we would want to avoid models which allow the function to dip. We want to use random intercepts because the initial level of tree growth is likely to be different across trees. Additionally, a monotonically decreasing random intercept model may be used to model tumor growth for subjects in a cancer treatment study. We consider the application of a monotone random intercept model to data set measuring CD4 cell numbers for HIV patients in Section 3.4.2.

### 3.3.1 Bayesian Model

Let us first consider the random intercept normal errors model with one covariate and assume we have $j=1, \ldots, n_{i}$ observations from group or individual $i$ and $i=1, \ldots, q$. We denote the observations by $\left(x_{i j}, y_{i j}\right)$ and suppose

$$
\begin{equation*}
y_{i j}=a_{i}+f\left(x_{i j}\right)+\epsilon_{i j} \tag{3.3}
\end{equation*}
$$

with independent random errors, $\epsilon_{i j} \sim \mathrm{~N}\left(0, \sigma^{2}\right)$, and random effects vector $\boldsymbol{a}=$ $\left(a_{1}, \ldots, a_{q}\right)$. We assume $f$ is a smooth and monotonically increasing function. As in Section 3.3, we assume each $a_{j}$ follows a normal distribution with mean $\mu_{a}$ and variance $\sigma_{a}^{2}$ and further assume that they are independent of each other. We also assume that the normal errors are independent of the random effects.

Let $\boldsymbol{x}=\left(x_{11}, x_{12}, \ldots, x_{1 n_{1}}, \ldots, x_{q 1}, x_{q 2}, \ldots, x_{q n_{q}}\right)^{\prime}$ be a vector of length $n=$ $\sum_{i=1}^{q} n_{i}$ of observed values of the covariate and let $\boldsymbol{y}$ be the vector of corresponding observed response values, $\left(y_{11}, y_{12}, \ldots, y_{1 n_{1}}, \ldots, y_{q 1}, y_{q 2}, \ldots, y_{q n_{q}}\right)^{\prime}$. To create shaperestricted basis vectors to approximate $f\left(x_{i j}\right)$, we fix $k$ interior knots and compute $s_{l}\left(x_{i j}\right)=\delta_{l i j}$ for $i=1, \ldots, q, j=1, \ldots, n_{i}$ and $l=1, \ldots, k+2$ where $s_{l}$ for
$l=1, \ldots, k+2$ are the quadratic $I$-spline basis functions created using the unique values of $\boldsymbol{x}$ and fixed knot points. We define $f$ to have an intercept of zero to avoid confounding with the random effects and let $f\left(x_{i j}\right)$ be approximated by $\sum_{l=1}^{m} \beta_{l} \delta_{l i j}$.

To estimate the shape-restricted regression function, we propose a hierarchical Bayesian model (Gelfand et al., 1996) where the likelihood for $\boldsymbol{y}$ is conditional on the random effects. We assume $y_{i j} \mid a_{i}, \sigma^{2}, \boldsymbol{\beta} \sim \mathrm{~N}\left(a_{i}+\sum_{l=1}^{m} \beta_{l} \delta_{l i j}, \sigma^{2}\right)$ for $i=1, \ldots, q$ and $j=1, \ldots, n_{i}$. We then assume the random intercepts are independent and normally distributed with $a_{i} \mid \mu_{a}, \sigma_{a}^{2} \sim \mathrm{~N}\left(\mu_{a}, \sigma_{a}^{2}\right), \mu_{a} \sim \mathrm{~N}(0, M)$, and $\tau_{a}=1 / \sigma_{a}^{2} \sim$ Gamma $\left(d_{a 1}, d_{a 2}\right)$. The hyperparameters for $\mu_{a}$ and $\tau_{a}$ are chosen such that the priors for these parameters are weakly informative. We use a hierarchical model to help improve mixing. The posterior surface of high dimensional models such as those for random intercept models with vague priors can be difficult to "explore" due to weak identifiability between the parameters. Using hierarchical centering, we can reparameterize the model in such a way that the resulting posterior surface is "better" behaved and easier to explore (Gelfand et al., 1996).

Letting $\boldsymbol{\beta}=\left(\beta_{1}, \ldots, \beta_{m}\right)^{\prime}$ where $m=k+2$, the hierarchical Bayesian model using shape-restricted regression splines is given by

$$
\begin{align*}
y_{i j} \mid a_{i}, \sigma^{2}, \boldsymbol{\beta} & \sim \mathrm{~N}\left(a_{i}+\sum_{l=1}^{m} \beta_{l} \delta_{l i j}, \sigma^{2}\right) \text { for } i=1, \ldots q \text { and } j=1, \ldots, n_{i} \\
a_{i} \mid \mu_{a}, \sigma_{a}^{2} & \sim \mathrm{~N}\left(\mu_{a}, \sigma_{a}^{2}\right) \text { for } i=1, \ldots q \\
\tau=1 / \sigma^{2} & \sim \operatorname{Gamma}\left(d_{1}, d_{2}\right)  \tag{3.4}\\
\tau_{a}=1 / \sigma_{a}^{2} & \sim \operatorname{Gamma}\left(d_{a 1}, d_{a 2}\right) \\
\mu_{a} & \sim \mathrm{~N}(0, M) \\
\beta_{l} & \sim \operatorname{Gamma}\left(c_{1}, c_{2}\right) \text { for } l=1, \ldots m
\end{align*}
$$

where independence is assumed among all parameters. Note that a vague uniform prior may also be used for the variance parameters as in Gelman (2006), especially if the value for $\sigma_{a}^{2}$ is believed to be very small.

The posterior distribution,

$$
\begin{aligned}
& p\left(\boldsymbol{\beta}, \boldsymbol{a}, \mu_{a}, \tau_{a}, \tau \mid \boldsymbol{y}\right) \\
& \propto \quad \tau^{n / 2} \exp \left\{\frac{-\tau}{2} \sum_{i=1}^{q} \sum_{j=1}^{n_{i}}\left(y_{i j}-a_{i}-\sum_{l=1}^{m} \beta_{l} \delta_{l i j}\right)^{2}\right\} \tau_{a}^{q / 2} \exp \left\{\frac{-\tau_{a}}{2} \sum_{i=1}^{q}\left(a_{i}-\mu_{a}\right)^{2}\right\} \times \\
& \quad \exp \left\{-\frac{\mu_{a}^{2}}{2 M}\right\} \tau_{a}^{d_{a 1}-1} \exp \left\{-d_{a 2} \tau_{a}\right\} \tau^{d_{1}-1} \exp \left\{-d_{2} \tau\right\} \times \\
& \\
& \quad \prod_{l=1}^{m}\left[\beta_{l}^{c_{1}-1} \mathrm{I}\left\{0<\beta_{l}<\infty\right\}\right] \exp \left\{-c_{2} \sum_{l=1}^{m} \beta_{l}\right\}
\end{aligned}
$$

is proper but analytically intractable so we use a Gibbs sampler to sample from the posterior distribution of the parameters. The conditional distribution used in the Gibbs sampler for $a_{i_{0}}$ given the data, $\boldsymbol{\beta}, \tau, \tau_{a}, \mu_{a}$, and all other random intercepts, $\boldsymbol{a}_{\left(-i_{o}\right)}=\left(a_{1}, \ldots, a_{i_{o}-1}, a_{i_{0}+1}, \ldots, a_{q}\right)$, is
$p\left(a_{i_{0}} \mid \boldsymbol{a}_{\left(-i_{o}\right)}, \boldsymbol{\beta}, \tau, \tau_{a}, \mu_{a}, \boldsymbol{y}\right) \sim \mathrm{N}\left(\left(\tau_{a_{i}}^{*}\right)^{-1}\left[\tau_{a} \mu_{a}+\tau \sum_{j=1}^{n_{i_{0}}}\left(y_{i_{0} j}-\sum_{l=1}^{m} \beta_{l} \delta_{l i_{0} j}\right)\right],\left(\tau_{a_{i}}^{*}\right)^{-1}\right)$
where $\tau_{a_{i}}^{*}=\tau n_{i_{0}}+\tau_{a}$. The conditional distribution for $\mu_{a}$ given all other parameters and the data is given by

$$
p\left(\mu_{a} \mid \boldsymbol{a}, \boldsymbol{\beta}, \tau, \tau_{a}, \boldsymbol{y}\right) \sim \mathrm{N}\left(\left(\tau_{\mu_{a}}^{*}\right)^{-1}\left[\tau_{a} \sum_{i=1}^{q} a_{i}\right],\left(\tau_{\mu_{a}}^{*}\right)^{-1}\right)
$$

where $\tau_{\mu_{a}}^{*}=q \tau_{a}+1 / M$. The conditional distribution for $\tau_{a}$ given all other parameters and the data is given by

$$
p\left(\tau_{a} \mid \boldsymbol{a}, \boldsymbol{\beta}, \tau, \mu_{a}, \boldsymbol{y}\right) \sim \operatorname{Gamma}\left(q / 2+d_{a 1}, d_{a 2}+\frac{1}{2} \sum_{i=1}^{q}\left(a_{i}-\mu_{a}\right)^{2}\right)
$$

The conditional distribution for $\tau$ given all other parameters and the data is given by
$p\left(\tau \mid \boldsymbol{a}, \boldsymbol{\beta}, \tau_{a}, \mu_{a}, \boldsymbol{y}\right) \sim \operatorname{Gamma}\left(n / 2+d_{1}, d_{2}+\frac{1}{2} \sum_{i=1}^{q} \sum_{j=1}^{n_{i}}\left(y_{i j}-a_{i}-\sum_{l=1}^{m} \beta_{l} \delta_{l i j}\right)^{2}\right)$.
Lastly, the conditional distribution for $\beta_{l_{0}}$ given the data, $\boldsymbol{a}, \tau, \tau_{a}, \mu_{a}$, and all other $\beta$ values denoted by $\boldsymbol{\beta}_{\left(-l_{o}\right)}=\left(\beta_{1}, \ldots, \beta_{l_{o}-1}, \beta_{l_{0}+1}, \ldots, \beta_{m}\right)$ is given by

$$
\begin{aligned}
p\left(\beta_{l_{0}} \mid \boldsymbol{a}, \boldsymbol{\beta}_{\left(-l_{0}\right)}, \tau, \tau_{a}, \mu_{a}, \boldsymbol{y}\right) \propto & \exp \left\{-\frac{s_{l_{0}} \tau}{2}\left[\beta_{l_{0} j_{0}}-\left(\sum_{i=1}^{q} \sum_{j=1}^{n_{i}} \frac{r_{i j} \delta_{l_{0} i j}}{s_{l_{0}}}-\frac{c_{2}}{s_{l_{0}} \tau}\right)\right]^{2}\right\} \times \\
& \beta_{l_{0}}^{c_{1}-1} \mathrm{I}\left\{0<\beta_{l_{0}}<\infty\right\}
\end{aligned}
$$

where $s_{l_{o}}=\sum_{i=1}^{q} \sum_{j=1}^{n_{i}} \delta_{l_{0} i j}^{2}, r_{i j}=y_{i j}-a_{i}-\sum_{l \neq l_{o}} \beta_{l} \delta_{l i j}$, and the sum over $l \neq l_{0}$ means take the sum over $l=1, \ldots, m$ subtracting the value when $l=l_{0}$. Note that $p\left(\beta_{l_{0}} \mid \boldsymbol{a}, \boldsymbol{\beta}_{-\left(l_{0}\right)}, \tau, \tau_{a}, \mu_{a}, \boldsymbol{y}\right)$ is of the same form as the conditional distribution in the Gibbs sampler for $\beta_{l_{0} j_{o}}$ in Section 2.4 but with $b=s_{l_{0}} \tau / 2>0$, and $c=$ $\sum_{i=1}^{q} \sum_{j=1}^{n_{i}} r_{i j} \delta_{l_{0} i j} / s_{l_{0}}-c_{2} /\left(s_{l_{0}} \tau\right)$. Therefore, we can sample from this distribution in the same way as described for $\beta_{l_{0} j_{0}}$ in Section 2.4.

The estimate for the regression function for the ( $i j$ )th observation is computed in the same manner as the function estimate in the generalized linear model in Chapter 2. It is the mean of the function estimates at each iteration $t$ of the Gibbs sampler for the (ij)th observation excluding burn-in where the estimate at iteration $t$ is given by

$$
\begin{equation*}
\eta_{i j}^{(t)}=a_{i}^{(t)}+\sum_{l=1}^{m} \beta_{l}^{(t)} \delta_{l j i} \tag{3.5}
\end{equation*}
$$

and the superscript $(t)$ on a parameter indicates the value for that parameter at the $(t)$ th iteration of the Gibbs sampler.

### 3.3.2 Random Intercept Normal Errors Model with Parametrically Modeled Covariates

Let us extend the random intercept model in Section 3.3.1 to include parametrically modeled covariates. First consider the case where there is one continuous variable and a categorical predictor variable with $r$ levels. Let $\boldsymbol{v}_{l}=\left(v_{l 11}, \ldots, v_{l q n_{q}}\right)$ for $l=1, \ldots, r-1$ be $r-1$ dummy variables for all but one of the levels of the categorical variable. Thus, we model data using

$$
\begin{equation*}
y_{i j}=a_{i}+f\left(x_{i j}\right)+\sum_{l=1}^{r-1} \alpha_{l} v_{l i}+\epsilon_{i j} \tag{3.6}
\end{equation*}
$$

where $f\left(x_{i j}\right), a_{i}$, and $\epsilon_{i j}$ are as in (3.3) and $r-1$ dummy variables are used to avoid over-parameterizing the model. Again, assume $f\left(x_{i j}\right)$ is a smooth monotonically increasing function with intercept of zero and approximate the mean at each $y_{i j}$ by $a_{i}+\sum_{l=1}^{m} \beta_{l} \delta_{l i j}+\sum_{l=1}^{r-1} \alpha_{l} v_{l i j}$.

We can further extend this model to include $g$ parametrically modeled covariates by simply including them in the $\sum_{l=1}^{r-1} \alpha_{l} v_{l i}$ term in (3.6) and obtaining

$$
y_{i j}=a_{i}+f\left(x_{i j}\right)+\sum_{l=1}^{r-1+g} \alpha_{l} v_{l i}+\epsilon_{i j} .
$$

Thus, the mean at each mean $y_{i j}$ is approximated using $a_{i}+\sum_{l=1}^{m} \beta_{l} \delta_{l i j}+\sum_{l=1}^{r-1+g} \alpha_{l} v_{l i}$.
The hierarchical Bayesian model with a categorical variable with $r$ levels and $g$ parametrically modeled covariates is

$$
\begin{align*}
y_{i j} \mid \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \mu_{a}, \tau_{a} & \sim \mathrm{~N}\left(a_{i}+\sum_{l=1}^{m} \beta_{l} \delta_{l i j}+\sum_{l=1}^{r-1+g} \alpha_{l} v_{l i j}, \sigma^{2}\right) \\
\boldsymbol{\alpha} & \sim \mathrm{N}\left(\mathbf{0}_{r-1}, M_{\alpha} I\right) \tag{3.7}
\end{align*}
$$

where $\mathrm{N}(\boldsymbol{\theta}, \Sigma)$ denotes a multivariate normal distribution with mean $\boldsymbol{\theta}$ and variancecovariance matrix $\Sigma, \mathbf{0}_{r-1}$ is a $(r-1)$ vector of zeros, $I$ is a $(r-1) \times(r-1)$ identity matrix. We use the same priors for $\boldsymbol{a}, \mu_{a}, \tau_{a}, \tau$, and $\boldsymbol{\beta}$ as in (3.4). We again use a Gibbs sampling algorithm to sample from the marginal distributions of the parameters. The conditional distributions used in the Gibbs sampler for $\boldsymbol{a}, \mu_{a}, \tau_{a}$, $\tau$, and $\boldsymbol{\beta}$ are very similar to the random intercepts model without a categorical variable. The conditional distribution for $a_{i_{0}}$ given all other parameters is given by

$$
p\left(a_{i_{0}} \mid \boldsymbol{a}_{\left(-i_{o}\right)}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau, \tau_{a}, \mu_{a}, \boldsymbol{y}\right) \sim \mathrm{N}\left(\left(\tau_{a_{i}}^{*}\right)^{-1}\left[\tau_{a} \mu_{a}+\tau \sum_{j=1}^{n_{i_{0}}}\left(r_{i_{0} j}\right)\right],\left(\tau_{a_{i}}^{*}\right)^{-1}\right)
$$

where $r_{i_{0} j}=y_{i_{0} j}-\sum_{l=1}^{m} \beta_{l} \delta_{l i_{0} j}-\sum_{l=1}^{r-1+g} \alpha_{l} v_{l i_{0} j}$. The conditional distribution for $\mu_{a}$ given all other parameters is given by

$$
p\left(\mu_{a} \mid \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau, \tau_{a}, \boldsymbol{y}\right) \sim \mathrm{N}\left(\left(\tau_{\mu_{a}}^{*}\right)^{-1}\left[\tau_{a} \sum_{i=1}^{q} a_{i}\right],\left(\tau_{\mu_{a}}^{*}\right)^{-1}\right) .
$$

The conditional distribution for $\tau_{a}$ given all other parameters is given by

$$
p\left(\tau_{a} \mid \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau, \mu_{a}, \boldsymbol{y}\right) \sim \operatorname{Gamma}\left(q / 2+d_{a 1}, d_{a 2}+\frac{1}{2} \sum_{i=1}^{q}\left(a_{i}-\mu_{a}\right)^{2}\right) .
$$

The conditional distribution for $\tau$ given all other parameters is given by

$$
p\left(\tau \mid \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau_{a}, \mu_{a}, \boldsymbol{y}\right) \sim \operatorname{Gamma}\left(n / 2+d_{1}, d_{2}+\sum_{i=1}^{q} \sum_{j=1}^{n_{i}} r_{i j}^{2}\right)
$$

where $r_{i j}=y_{i j}-a_{i}-\sum_{l=1}^{m} \beta_{l} \delta_{l i j}-\sum_{l=1}^{r-1+g} \alpha_{l} v_{l i j}$. The conditional distribution for $\beta_{l_{0}}$ given all other parameters is given by

$$
p\left(\beta_{l_{0}} \mid \boldsymbol{a}, \boldsymbol{\beta}_{-\left(l_{0}\right)}, \boldsymbol{\alpha}, \tau, \tau_{a}, \mu_{a}, \boldsymbol{y}\right) \propto \beta_{l_{0}}^{c_{1}-1} \exp \left\{-\frac{s_{l_{0}} \tau}{2}\left(\beta_{l_{0} j_{0}}-c\right)^{2}\right\} \mathrm{I}\left\{0<\beta_{l_{0}}<\infty\right\}
$$

with $c=\sum_{i=1}^{q} \sum_{j=1}^{n_{i}} r_{i j} \delta_{l_{0} i j} / s_{l_{0}}-c_{2} /\left(s_{l_{0}} \tau\right)$ and $r_{i j}=y_{i j}-a_{i}-\sum_{l \neq l_{o}} \beta_{l} \delta_{l i j}-$ $\sum_{l=1}^{r-1+g} \alpha_{l} v_{l i j}$. The conditional distribution in the Gibbs sampler for $\alpha_{l_{0}}$ given all other parameters and the data is

$$
p\left(\alpha_{l_{0}} \mid \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{\alpha}_{\left(-l_{0}\right)}, \tau, \tau_{a}, \mu_{a}, \boldsymbol{y}\right) \sim \mathrm{N}\left(\frac{\tau}{\tau_{\alpha}^{*}} \sum_{i=1}^{q} \sum_{j=1}^{n_{i}} v_{l_{0} i j} r_{i j},\left(\tau_{\alpha}^{*}\right)^{-1}\right)
$$

where $r_{i j}=y_{i j}-a_{i}-\sum_{l=1}^{m} \beta_{l} \delta_{l i j}-\sum_{l_{0} \neq l} \alpha_{l} v_{l i j}, \tau_{\alpha}^{*}=\tau \sum_{i=1}^{q} \sum_{j=1}^{n_{i}} v_{l_{0} i j}^{2}+1 / M_{\alpha}$, and the sum over $l \neq l_{0}$ means take the sum over $l=1, \ldots, r-1+g$ subtracting the value when $l=l_{0}$. The estimate of the regression function for the $(i j)$ th observation is the mean of the function estimates at each iteration $t$,

$$
\begin{equation*}
\eta_{i j}^{(t)}=a_{i}^{(t)}+\sum_{l=1}^{m} \beta_{l}^{(t)} \delta_{l i j}+\sum_{l=1}^{r-1} \alpha_{l}^{(t)} v_{l i j} \tag{3.8}
\end{equation*}
$$

excluding burn-in.

### 3.3.3 Inference with Mixed Model

As with the generalized linear model, the Bayes SRRS generalized linear mixed model lends itself to several types of inference. Credible intervals can be easily constructed by replacing $\eta_{i}^{(t)}$ in Section 2.6.1 with $\eta_{i j}^{(t)}$ in (3.5) or (3.8). Bayesian model selection tools such as Bayes factors can be used to select the appropriate model given the data as in Section 2.6.3. The approximate joint normality of the posterior distribution of the fixed effects $\boldsymbol{\alpha}$ in (3.2) can be used to determine the significance of these effects as in Section 2.6.2. In particular, considering the random intercept model with a categorical variable as in Section 3.3, we can determine whether the function depends on the level of the categorical variable by considering the posterior distribution of $\boldsymbol{\alpha}$. If $r=2$, we can construct a credible interval for $\alpha$ and see if this interval includes zero. If $r>2$, we can perform inference using (2.23) to determine whether the $\alpha_{1}=\alpha_{2}=\ldots=\alpha_{r-1}=0$.

### 3.4 Applications

To examine the performance of the Bayes SRRS method for generalized linear mixed models under the monotone shape restriction we consider applying our Bayes SRRS estimation procedure to simulated data and a subset of a data set from the Multicenter AIDS Cohort Study, "MACS", discussed in Kaslow et al. (1987) and analyzed by Zeger and Diggle (1994). We refer to this data set as the CD4 data set.

### 3.4.1 Simulated Data

For the simulated data set, we simulate data under the random intercept model with a categorical covariate with $r=2$ levels and let $y_{i j}=a_{i}+f\left(x_{i j}\right)+\alpha v_{i}+\epsilon_{i j}$ with $i=1, \ldots, 30$. For each $i$ we compute $f$ at 20 equally spaced values between 0 and 1. We randomly generate $n_{l}=20 a_{i}$ values from $\mathrm{N}(2,0.5)$ and randomly generate 600 normal errors $\epsilon_{i j}$ values from $\mathrm{N}(0,1)$ distribution. We let $\boldsymbol{v}_{l}=\left(v_{l 1}, \ldots, v_{l 20}\right)^{\prime}$, $\boldsymbol{v}=\left(v_{11}, \ldots, v_{1,20}, \ldots, v_{30,1}, \ldots, v_{30,20}\right)^{\prime}=\left(\boldsymbol{v}_{1}^{\prime} \ldots \boldsymbol{v}_{30}^{\prime}\right)^{\prime}$, and let $\boldsymbol{v}_{l}$ be a vector of zeros if $l$ is odd and a vector of ones if $l$ is even. We set $\alpha=1.5$ and let $f(x)=$ $5 \exp (10 x-5) /(1+\exp (10 x-5))$. To estimate $f$, we use $k=3$ equally spaced interior knots fixed at $0.25,0.5$, and 0.75 , run the algorithm for 50,000 iterations and discard the first 10,000 as burn-in. The Bayes SRRS estimate (solid red line) along with the $95 \%$ HPD (grey shaded region) are given in Figure 3.1. Note that the HPD intervals include the true values for all $x_{i j}$. The partial trace plot and histogram of posterior draws for $\alpha$ is given in Figure 3.2. The posterior distribution is centered close to the value of $\alpha$ used to simulate the data. The $95 \%$ HPD interval for $\alpha$ is $(1.20,2.07)$ which includes the true value and supports that $\alpha$ is different from zero.

### 3.4.2 CD4 Data

To further examine the performance of our model, we use our model to estimate regression curves for a subset of the data set from the Multicenter AIDS Cohort


Figure 3.1: Data set generated from random intercept model with $i=3$ (a), $i=$ $14(\mathrm{~b})$, and $i=21$ (c) along with $f\left(x_{i j}\right)+a_{i}$ used to simulate the data (grey dashed line), the estimate found using the Bayes SRRS model estimate (solid red line), and $95 \%$ HPD interval found using Bayes SRRS model and method from Section 2.6.1.

Study (MACS) analyzed by Zeger and Diggle (1994). The MACS followed nearly 5,000 gay and bisexual men from Baltimore, Pittsburgh, Chicago, and Los Angeles with approximately $37 \%$ of participants infected with HIV when the study began in 1984 and about 7\% more seroconverted (HIV positive) during the follow-up. HIV destroys CD4 cells and CD4 cell loss can be used to study disease progression. We model the CD4 cell levels as a function of time (in years) since seroconversion (HIV antibody positive). The data set includes several other covariates such as age (relative to an arbitrary origin), packs of cigarettes smoked per day, an indicator for recreational drug use, number of sexual partners (centered), and mental illness score (depression symptoms as measured by CSED scale where larger values indicate increased depressive symptoms). We focus on modeling CD4 levels (per 1000 cells) for individual $i$ as a monotonically decreasing function of years since seroconversion, $x_{i j}$. We consider a subset of the data and include individuals who have at least 7 observations after seroconversion and smoked less than one pack a day. This results in observations on 45 individuals who were HIV positive at some point. We omit all other covariates except time since seroconversion to simplify the analysis and use the random intercept normal errors model discussed in Section 3.3.1 to model the regression relationship assuming a monotonically decreasing regression function.


Figure 3.2: (a) Partial trace plots for $\alpha$ corresponding to the second level of the categorical variable where the value of $\alpha$ used to generate the data is given by the dashed red line. (b) Histogram of posterior draws for $\alpha$ after discarding burn-in where the value of $\alpha$ used to generate the data is given by the dashed red line.

Thus, we model the mean for the $(j)$ th observation for individual $i, \eta_{i j}$, by

$$
a_{i}+\sum_{l=1}^{m} \beta_{l} \cdot\left(-\delta_{l i j}\right)
$$

where $i=1, \ldots, 45, m=6$, and $a_{i}$ is random intercept for individual $i . \delta_{l i j}$ is the value of the $(l)$ th basis function evaluated at $x_{i j}$ where $x_{i j}$ is the time since seroconversion for the $(j)$ th observation of the $(i)$ th individual. $\beta_{l}$ for $l=1, \ldots, m$ are the regression coefficients constrained to be positive. The $I$-spline basis vectors are created using the unique observed $x$-values and knot points $(0.211,1.254,2.297,3.340,4.383,5.426)^{\prime}$.

The Bayes SRRS estimate (solid red line) found by running the MCMC algorithm for 50,000 iterations and discarding the first 10,000 as burn-in along with the 95\% HPD (grey shaded region) for three of the individuals are shown in Figure 3.3. The function estimates for these individuals suggest that there is a gradual decrease in CD4 cells after seroconversion. The partial trace plot for $\tau, \tau_{a}$, and $\mu_{a}$ are given in Figure 3.4 suggesting the algorithm is mixing well.


Figure 3.3: The Bayes SRRS model estimates (solid black line) and $95 \%$ HPD interval found using Bayes $S R R S$ model for three individuals from the C f 4 data set.


Figure 3.4: (a) Partial trace plot for $\tau$. (b) Partial trace plot for $\tau_{a}$. (c) Partial trace plot for $\mu_{a}$.

We could also consider a random intercept model including the recreational drug use covariate and use the Bayes SRRS spline model with a categorical covariate as discussed in Section 3.3 to determine if drug use affects the CD4 levels. Thus, we estimate the mean of $y_{i j}, \eta_{i j}$, by $a_{i}+\sum_{l=1}^{m} \beta_{l}\left(-\delta_{l i j}\right)+\alpha v_{i j}$ where $v_{i j}$ is an indicator whether the $(i j)$ th observation was taken from an individual who is presently participating in recreational drug use. The partial trace plot and histogram for $\alpha$ is given in Figure 3.5. The $95 \%$ HPD for $\alpha$ is $(-0.025,0.107)$ and suggests that recreation drug use to does not drastically affect CD4 levels when assuming the random intercept normal errors model.


Figure 3.5: (a) Partial trace plot for $\alpha$. (b) Histogram of posterior draws for $\alpha$.

## Chapter 4

## BAYESIAN SHAPE-RESTRICTED REGRESSION WITH FREE-KNOT SPLINES

### 4.1 Introduction

A key advantage of shape-restricted function estimation is that it is generally robust to the choice of knot number and placement compared to unrestricted spline models (Meyer, 2008). However, the Bayesian framework allows one to use reversible-jump Markov chain Monte Carlo (Richardson and Green, 1997; Green, 1995) to integrate over the number and location of interior knots. RJMCMC is a generalization of the Metropolis Hastings sampler (Tierney, 1994) that allows moves that increase or decrease the dimension of the parameter vector.

Estimation of regression splines where the knots for the $B$-splines are free parameters and the function estimates are found using reversible-jump Markov chain Monte Carlo (RJMCMC) has been studied by many other researchers (Holmes and Mallick (2003); DiMatteo et al. (2001); Biller (2000); Lindstrom (2002); Zhou and Shen (2001); Denison et al. (1998)). In particular, Denison et al. (1998) used RJMCMC to determine the number and placement of knots for piecewise polynomials. They used least squares to update and estimate the coefficients so this approach is not fully Bayesian. Biller (2000) and DiMatteo et al. (2001) used a fully Bayesian approach involving a RJMCMC algorithm which estimate functions using cubic $B$ splines. DiMatteo et al. (2001) assumed normal errors and used normal priors for the regression coefficients which simplifies the algorithm by allowing them to marginalize out the regression coefficients. Thus, they used RJMCMC to simulate a chain
for the number of knots and knot locations only and obtained regression coefficient estimates from the marginal distribution separately. Biller (2000) suggested normal priors for the regression coefficients and used RJMCMC to simulate the number and location of interior knots as well as the regression coefficients. He assumed a grid of possible knot locations over the range of the data as opposed to DiMatteo et al. (2001) who allowed the knot locations to fall anywhere in the range of the data. Both analyzes did not involve shape restrictions and imposing shape restrictions make the choice of normal priors for the regression coefficients inappropriate. Wang (2008) considered using RJMCMC along with cubic $B$-splines to estimate monotonically increasing functions. His method involved using normal priors for the unconstrained regression coefficients and then projecting the unconstrained distribution on the constrained space. He used second-order cone programming (Alizadeh and Goldfarb, 2003) and maximum likelihood estimation to obtain estimates for the regression coefficients, but this procedure can be difficult to implement and may involve problems with optimization. Using $B$-splines and RJMCMC, Johnson (2007) considered modeling dichotomous item response theory where the mean response function is non-decreasing. The monotone shape restriction was imposed by truncated normal priors that required the regression coefficients on the $B$-splines to be ordered. Shively et al. (2011) proposed a method for shape-restricted function estimation using freeknot quadratic polynomial regression splines in a Bayesian context. They imposed shape restrictions using linear constraints on the spline coefficients and multivariate a normal prior for them constrained to the multi-dimensional generalization of the first quadrant.

The procedure described here considers estimating functions using I-splines and RJMCMC where number and location of interior knots are unknown. Our use of quadratic $I$-splines allows monotonicity to be imposed by requiring the spline coefficients to be positive, which simplifies the RJMCMC algorithm. Our algorithm can also be generalized to other shape restrictions by selecting the appropriate basis
functions and restricting the coefficients to be positive. Changes to the coefficients updates in the RJMCMC algorithm based on the properties of the basis vectors may also be desired. In Section 4.2, we present the Bayesian model for function estimation using free-knot splines under the monotonically increasing shape restriction with the model and outline of the RJMCMC algorithm given in Section 4.2.1 and the priors used in our analysis given in Section 4.2.2. In Section 4.3, we give the specifics of the algorithm and describe each move in detail. This work can be found in Meyer et al. (2011).

### 4.2 Free-knot Spline Model

### 4.2.1 Model and Algorithm

Consider the normal errors model described in Section 2.4 with $L=1$ and the monotonically increasing shape restriction. Similar to Biller (2000), we consider three possible moves: "birth," "death," or "relocation," of knots (described further in Sections 4.3.1, 4.3.2, and 4.3.3, respectively) with probabilities $b_{k}, d_{k}$, and $r_{k}$, respectively. Let $k$ be the current number of interior knots and as in DiMatteo et al. (2001), we let $b_{k}=c \cdot \min \left\{1, f_{k}(k+1) / f_{k}(k)\right\}, d_{k}=c \cdot \min \left\{1, f_{k}(k-1) / f_{k}(k)\right\}$, $r_{k}=1-b_{k}-d_{k}$, and $c=0.4$ where $f_{k}(\cdot)$ is the density of the prior on the number of interior knots (discussed in Section 4.2.2). We place bounds on the possible values for $k$ with lower bound denoted by $k_{\min }$ and upper bound denoted $k_{\max }$. To avoid proposing moves with $k$ outside of $\left[k_{\text {min }}, k_{\text {max }}\right]$, we let $b_{k}=1, d_{k}=0$, and $r_{k}=0$ if $k=k_{\min }$ knots and $b_{k}=0, d_{k}=1$, and $r_{k}=0$ if $k=k_{\max }$ knots. The parameters proposed in each move are accepted with probability $\min \{1, A\}$ with

$$
\begin{equation*}
A=(\text { likelihood ratio }) \times(\text { prior ratio }) \times(\text { proposal ratio }) \times \mid \operatorname{det}(\text { Jacobian }) \mid \tag{4.1}
\end{equation*}
$$

where the Jacobian is found by first defining $\boldsymbol{\theta}_{1}$ as the vector of all the parameter values in the current model $M_{1}$ and $\boldsymbol{\theta}_{2}$ the parameter vector for the proposed model, $M_{2}$. We create auxiliary variable vectors $\boldsymbol{u}_{1}$ and $\boldsymbol{u}_{2}$ such that length $\left(\boldsymbol{\theta}_{1}\right)+$ length $\left(\boldsymbol{u}_{1}\right)=$
length $\left(\boldsymbol{\theta}_{2}\right)+$ length $\left(\boldsymbol{u}_{2}\right)$ and define $g$ as a function that maps $\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}$ to $\left(\boldsymbol{\theta}_{2}, \boldsymbol{u}_{2}\right)^{\prime}$. The Jacobian is

$$
\begin{equation*}
\frac{\partial g\left(\left(\boldsymbol{\theta}, \boldsymbol{u}_{1}\right)^{\prime}\right)}{\partial\left(\boldsymbol{\theta}, \boldsymbol{u}_{1}\right)^{\prime}} \tag{4.2}
\end{equation*}
$$

The proposal ratio is the probability of proposing $\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}$ given $\left(\boldsymbol{\theta}_{2}, \boldsymbol{u}_{2}\right)^{\prime}$ over the probability of proposing $\left(\boldsymbol{\theta}_{2}, \boldsymbol{u}_{2}\right)^{\prime}$ given $\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}$ and it depends on the move type. The acceptance probability for the birth, death, and relocation moves are given in Sections 4.3.1, 4.3.2, and 4.3.3, respectively.

After each birth, death, or relocation move, we perform a "coefficient update" step, updating $\boldsymbol{\beta}=\left(\beta_{1}, \ldots, \beta_{m}\right)^{\prime}, \boldsymbol{\alpha}=\left(\alpha_{1}, \ldots, \alpha_{p}\right)^{\prime}$, and $\tau$ using a Gibbs sampler and the conditional distributions given in Section 2.4.1. An estimate of the mean regression function for the $(t)$ th iteration is found by

$$
\begin{equation*}
\boldsymbol{\eta}^{(t)}=\sum_{j=1}^{m^{(t)}} \beta_{j}^{(t)} \boldsymbol{\delta}_{j}^{(t)}+\sum_{i=1}^{p} \alpha_{i}^{(t)} \boldsymbol{v}_{j i} \tag{4.3}
\end{equation*}
$$

where $\beta_{j}^{(t)}$ for $j=1, \ldots, m^{(t)}$ and $\boldsymbol{\alpha}^{(t)}=\left(\alpha_{1}^{(t)}, \ldots, \alpha_{p}^{(t)}\right)^{\prime}$ are the parameter values for the $(t)$ th iteration after the Gibbs sampler coefficient update step, $m^{(t)}=k^{(t)}+2$, $k^{(t)}$ is the number of interior knots for the $(t)$ th iteration, $\boldsymbol{\delta}_{j}^{(t)}$ is the $(j)$ th quadratic $I$-spline basis vector created using the $k^{(t)}$ interior knot locations for iteration $t$, and $\boldsymbol{v}_{j i}$ is as in (2.3). The regression spline estimate is then found by averaging $\boldsymbol{\eta}^{(t)}$ after discarding burn-in.

### 4.2.2 Priors on Model Parameters

As in Green (1995), Biller (2000), Denison et al. (1998) DiMatteo et al. (2001), and Johnson (2007) we use a Poisson $(\lambda)$ prior for the number of interior knots $k$. We let the prior for $k$ be truncated on $k_{\min }=1$ and $k_{\max }=100$ so the density is given by

$$
f_{k}(k)=\left(\sum_{i=k_{\min }}^{k_{\max }} \frac{\lambda^{i}}{i!}\right)^{-1} \frac{\lambda^{k}}{k!} .
$$

The hyperparameter $\lambda$ can be adjusted to help improve mixing. Given $k$, we assume a prior for the $k$ interior knot locations as in the step function model in Green (1995) and assume the knot locations are distributed as the even-numbered order statistics from $2 k+1$ locations uniformly distributed on $[\min (\boldsymbol{x})$, $\max (\boldsymbol{x})]$. Let $\boldsymbol{t}=\left(t_{1}, \ldots, t_{k+2}\right)^{\prime}$ be the ordered knot locations for model with $k$ interior knots (set $t_{1}=\min (\boldsymbol{x})$ and $\left.t_{k+2}=\max (\boldsymbol{x})\right)$. The density of the prior for the $k$ interior knot locations is given by

$$
\frac{(2 k+1)!\cdot t_{2}\left(t_{3}-t_{2}\right) \cdot \ldots \cdot\left(t_{k+1}-t_{k}\right)\left(\max (\boldsymbol{x})-t_{k+1}\right)}{[\max (\boldsymbol{x})-\min (\boldsymbol{x})]^{2 k+1}} .
$$

As with the fixed-knot model in Chapter 2, we use independent Gamma $\left(c_{1}, c_{2}\right)$ priors for $\beta_{j}$, independent $\mathrm{N}(0, M)$ priors for the $\alpha_{j}$ 's, and a Gamma $\left(d_{1}, d_{2}\right)$ prior for $\tau$.

### 4.3 Implementation

In this section, we describe how we propose the new knot locations for each move type as well as how we propose new values for the restricted regression coefficients. We also derive the acceptance probability for each move type.

### 4.3.1 Birth Move

### 4.3.1.1 Addition of knot

A birth move from a model with $k$ interior knots to a model with $k+1$ interior knots adds a new interior knot location by selecting a new knot location, $t^{*}$, from a Uniform $(\min (\boldsymbol{x}), \max (\boldsymbol{x}))$. Now, $t^{*}$ will fall between $t_{j}$ and $t_{j+1}$ where $t_{j}$ is the $(j)$ th largest knot location in the current model with $k$ interior knots and $j=$ $1, \ldots, k+1$. By exploiting the fact that quadratic $I$-spline basis functions can be generated such that they have a range of one and that each $I$-spline has a positive slope at one and only one knot location, we need only to update coefficients in close proximity of the new knot location. Let $\boldsymbol{\beta}=\left(\beta_{1}, \ldots, \beta_{k+2}\right)^{\prime}$ be the current $I$-spline basis function coefficient values for a model with $k$ interior knots where $\beta_{j}$
is the coefficient for the basis function corresponding to the $I$-spline with a positive slope at the $(j)$ th largest knot location. Given the proposed knot locations $\tilde{\boldsymbol{t}}=$ $\left(t_{1}, \ldots, t_{j}, t^{*}, t_{j+1}, \ldots, t_{k+2}\right)^{\prime}=\left(\tilde{t}_{1}, \ldots, \tilde{t}_{k+3}\right)^{\prime}$, we propose a new coefficient vector $\tilde{\boldsymbol{\beta}}=\left(\tilde{\beta}_{1}, \ldots, \tilde{\beta}_{k+3}\right)^{\prime}$ using

$$
\tilde{\beta}_{i}=\left\{\begin{array}{ll}
\beta_{i} & i=1, \ldots, j-1  \tag{4.4}\\
(1-u) \beta_{i} & i=j \\
u\left(\beta_{i-1}+\beta_{i}\right) & i=j+1 \\
(1-u) \beta_{i-1} & i=j+2 \\
\beta_{i-1} & i=j+3, \ldots, k+3
\end{array} .\right.
$$

where $u \sim$ Uniform $(0,1)$. Note that this coefficient update ensures that the proposed coefficients are positive to preserve the shape assumption and ensures that $\sum_{i=j}^{j+1} \beta_{i}=\sum_{i=j}^{j+2} \tilde{\beta}_{i}$ to preserve the range of the estimate which is equal to the sum of the regression coefficients. Note that both $\boldsymbol{\alpha}$ and $\tau$ are not updated until the coefficient update step in the Gibbs sampler algorithm. If we denote the current values of these parameters by $\boldsymbol{\alpha}$ and $\tau$, the proposed values are $\tilde{\boldsymbol{\alpha}}=\boldsymbol{\alpha}$ and $\tilde{\tau}=\tau$.

### 4.3.1.2 Acceptance Probability

Following (4.1), we derive the acceptance probability for a move from a model with $k$ interior knots to $k+1$ interior knots where $t^{*}, t_{j}, t_{j+1}, \boldsymbol{t}, \boldsymbol{\beta}, \tilde{\boldsymbol{t}}$ and $\tilde{\boldsymbol{\beta}}$ are defined as in Section 4.3.1.1. The likelihood ratio is given by

$$
\begin{equation*}
\frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k+1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \tau\right)}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right)} \tag{4.5}
\end{equation*}
$$

where $f_{y}\left(y_{i} \mid k+1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+3} \tilde{\beta}_{j} \tilde{\delta}_{j i}+$ $\sum_{j=1}^{p} \alpha_{j} v_{j i}$ and standard deviation of $\tau^{-1 / 2}, f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\eta_{i}=\sum_{j=1}^{k+2} \beta_{j} \delta_{j i}+\sum_{j=1}^{p} \alpha_{j} v_{j i}$ and standard deviation of $\tau^{-1 / 2}, \tilde{\delta}_{j i}$ is the $(i)$ th value of the $(j)$ th $I$-spline basis vector created using proposed knot locations $\tilde{\boldsymbol{t}}$, and $\delta_{j i}$ is the $(i)$ th value of the $(j)$ th $I$-spline basis vector created using current knot locations $\boldsymbol{t}$.

With the priors as in Section 4.2.2, the prior ratio is

$$
\begin{align*}
& \frac{p(k+1) p(\tilde{\boldsymbol{t}} \mid k+1) p(\tilde{\boldsymbol{\beta}} \mid k+1) p(\tilde{\boldsymbol{\alpha}}) p(\tilde{\tau})}{p(k) p(\boldsymbol{t} \mid k) p(\boldsymbol{\beta} \mid k) p(\boldsymbol{\alpha}) p(\tau)} \\
& =\frac{\frac{\lambda^{k+1}}{(k+1)!} \cdot \frac{(k)!}{\lambda^{k}} \cdot \frac{(2 k+3)![\max (\boldsymbol{x})-\min (\boldsymbol{x})]^{2 k+1}}{(2 k+1)![\max (\boldsymbol{x})-\min (\boldsymbol{x})]^{2 k+3}} \times}{} \quad \frac{t_{2}\left(t_{3}-t_{2}\right) \cdot \ldots \cdot\left(t_{j}-t_{j-1}\right)\left(t^{*}-t_{j}\right)\left(t_{j+1}-t^{*}\right) \cdot \ldots \cdot\left(\max (\boldsymbol{x})-\min (\boldsymbol{x})-t_{k+1}\right)}{t_{2}\left(t_{3}-t_{2}\right) \cdot \ldots \cdot\left(t_{j}-t_{j-1}\right)\left(t_{j+1}-t_{j}\right) \cdot \ldots \cdot\left(\max (\boldsymbol{x})-\min (\boldsymbol{x})-t_{k+1}\right)} \times \\
& \\
& \quad \frac{\prod_{j=1}^{k+3}\left[\frac{c_{2}^{c_{1}}}{\Gamma\left(c_{1}\right.} \tilde{\beta}_{j}^{c_{1}-1} \exp \left\{-c_{2} \tilde{\beta}_{j}\right\} I\left\{0<\tilde{\beta}_{j}<\infty\right\}\right] f(\tilde{\alpha}) f(\tilde{\tau})}{\prod_{j=1}^{k+2}\left[\frac{c_{2}^{c_{1}}}{\Gamma\left(c_{1}\right)} \beta_{j}^{c_{1}-1} \exp \left\{-c_{2} \beta_{j}\right\} \mathrm{I}\left\{0<\beta_{j}<\infty\right\}\right] f(\alpha) f(\tau)} \\
& = \\
& \quad \frac{\lambda(2 k+3)(2 k+2)\left(t^{*}-t_{j}\right)\left(t_{j+1}-t^{*}\right) c_{2}^{c_{1}}}{(k+1)[\max (\boldsymbol{x})-\min (\boldsymbol{x})]^{2}\left(t_{j+1}-t_{j}\right) \Gamma\left(c_{1}\right)}\left(\frac{\tilde{\beta}_{j} \tilde{\beta}_{j+1} \tilde{\beta}_{j+2}}{\beta_{j} \beta_{j+1}}\right)^{c_{1}-1} \times  \tag{4.6}\\
& \\
& \quad \exp \left\{-c_{2}\left[\sum_{j=1}^{k+3} \tilde{\beta}_{j}-\sum_{j=1}^{k+2} \beta_{j}\right]\right\}
\end{align*}
$$

Note that $f(\theta)$ is used to denote the density of the prior for a given parameter $\theta$, $c_{1}$ and $c_{2}$ are the hyperparameters for the priors for the $\beta$ coefficients, and $\lambda$ is the hyperparameter for Poisson prior for the number of interior knots.

Performing the birth move as in Section 4.3.1.1, the proposal ratio is given by

$$
\begin{align*}
\frac{\operatorname{Pr}(\text { death in } k+1 \text { model }) \operatorname{Pr}\left(\text { delete } t^{*}\right)}{\operatorname{Pr}(\text { birth in } k \text { model }) \operatorname{Pr}\left(t^{*}\right) \operatorname{Pr}(u)} & =\frac{d_{k+1} \cdot \frac{1}{k+1}}{b_{k} \cdot \frac{1}{[\max (\boldsymbol{x})-\min (\boldsymbol{x})]} \cdot \frac{1}{1-0}} \\
& =\frac{d_{k+1}[\max (\boldsymbol{x})-\min (\boldsymbol{x})]}{b_{k}(k+1)} \tag{4.7}
\end{align*}
$$

where $\operatorname{Pr}$ (death in $k+1$ model) means probability of performing a death move from a model with $k+1$ interior knots to a model with $k$ interior knots and $\operatorname{Pr}($ birth in $k$ model) means probability of performing a birth move from a model with $k$ interior knots to a model with $k+1$ interior knots. Recall that $b_{k}$ and $d_{k+1}$ are defined as in Section 4.2.1.

The Jacobian in (4.1) is found by letting $\boldsymbol{u}_{1}=\left(t^{*}, u\right),\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}=\left(k, \boldsymbol{t}, \boldsymbol{\beta}, \alpha, \tau, t^{*}, u\right)^{\prime}$, and $g(\cdot)$ be the function that maps $\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}$ to the proposed $\boldsymbol{\theta}_{2}^{\prime}=(\tilde{k}=k+1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\beta}}, \tilde{\alpha}, \tilde{\tau})$ given in (4.4). The Jacobian is $\partial\left[g\left(\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}\right)\right] / \partial\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}$. Hence, the determinant
of the Jacobian is equal to the determinant of

$$
\left[\begin{array}{ccc}
1-u & u & 0 \\
0 & u & 1-u \\
-\beta_{j} & \beta_{j}+\beta_{j+1} & -\beta_{j+1}
\end{array}\right]
$$

so the absolute value of the determinant is given by

$$
\begin{equation*}
\left|(u-1)\left(\beta_{j}+\beta_{j+1}\right)\right|=\left|\beta_{j}+\beta_{j+1}-\tilde{\beta}_{j+1}\right| \tag{4.8}
\end{equation*}
$$

To obtain the acceptance probability for birth move, substitute (4.5), (4.6), (4.7), and (4.8) into (4.1) and obtain

$$
\begin{align*}
A= & \frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k+1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \tau\right) \lambda(2 k+3)(2 k+2)\left(t^{*}-t_{j}\right)\left(t_{j+1}-t^{*}\right) c_{2}^{c_{1}}}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right)(k+1)[\max (\boldsymbol{x})-\min (\boldsymbol{x})]^{2}\left(t_{j+1}-t_{j}\right) \Gamma\left(c_{1}\right)} \times \\
& \left(\frac{\tilde{\beta}_{j} \tilde{\beta}_{j+1} \tilde{\beta}_{j+2}}{\beta_{j} \beta_{j+1}}\right)^{c_{1}-1} \times \exp \left\{-c_{2}\left[\sum_{j=1}^{k+3} \tilde{\beta}_{j}-\sum_{j=1}^{k+2} \beta_{j}\right]\right\} \times \\
& \frac{\prod_{j=1}^{k+3} \mathrm{I}\left\{0<\tilde{\beta}_{j}<\infty\right\}}{\prod_{j=1}^{k+2} \mathrm{I}\left\{0<\beta_{j}<\infty\right\}} \times \frac{d_{k+1}[\max (\boldsymbol{x})-\min (\boldsymbol{x})]}{b_{k}(k+1)} \times \\
& \left|\beta_{j}+\beta_{j+1}-\tilde{\beta}_{j+1}\right| \tag{4.9}
\end{align*}
$$

### 4.3.2 Death Move

The death move from a model with $k$ interior knots to a model with $k-1$ interior knots deletes one of the existing interior knots and updates the coefficients in a deterministic fashion. From the existing ordered knot locations, one interior knot location, say $t_{j+1}$, is randomly selected to be removed giving proposed knot locations $\tilde{\boldsymbol{t}}=\left(t_{1}, \ldots, t_{j}, t_{j+2}, \ldots, t_{k+2}\right)^{\prime}=\left(\tilde{t}_{1}, \ldots, \tilde{t}_{j}, \tilde{t}_{j+1}, \ldots, \tilde{t}_{k+1}\right)^{\prime}$. The death move is the inverse of a birth move from $k-1$ interior knots to $k$ interior knots with new knot location $t^{*}=t_{j+1}$ and the proposed values for the $I$-spline coefficients can be found by inverting the function given in (4.4). Thus, we propose a new coefficient vector $\tilde{\boldsymbol{\beta}}=\left(\tilde{\beta}_{1}, \ldots, \tilde{\beta}_{k+1}\right)^{\prime}$ by letting

$$
\tilde{\beta}_{i}=\left\{\begin{array}{ll}
\beta_{i} & i=1, \ldots, j-1  \tag{4.10}\\
\frac{\beta_{i}\left(\beta_{i}+\beta_{i+1}+\beta_{i+2}\right)}{\left.\beta_{i}+\beta_{i+2}+\beta_{i}\right)} & i=j \\
\frac{\beta_{i+1}\left(\beta_{i-1}+\beta_{i}+\beta_{i+1}\right)}{\beta_{i-1}+\beta_{i+1}} & i=j+1 \\
\beta_{i+1} & i=j+2, \ldots, k+1
\end{array} .\right.
$$

The acceptance probability is found by inverting the acceptance probability, $A_{k-1}$, for a birth from $k-1$ to $k$ interior knots with $t^{*}=t_{j+1}$. The acceptance probability is equal to $\min \left\{1, A_{k-1}^{-1}\right\}$ where

$$
\begin{aligned}
A_{k-1}= & \frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right) \lambda(2 k+1)(2 k)\left(t_{j+1}-t_{j}\right)\left(t_{j+2}-t_{j+1}\right) c_{2}^{c_{1}}}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k-1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \tau\right)(k)[\max (\boldsymbol{x})-\min (\boldsymbol{x})]^{2}\left(t_{j+2}-t_{j}\right) \Gamma\left(c_{1}\right)} \times \\
& \left(\frac{\beta_{j} \beta_{j+1} \beta_{j+2}}{\tilde{\beta}_{j} \tilde{\beta}_{j+1}}\right)^{c_{1}-1} \times \exp \left\{-c_{2}\left[\sum_{j=1}^{k+2} \beta_{j}-\sum_{j=1}^{k+1} \tilde{\beta}_{j}\right]\right\} \times \\
& \frac{\prod_{j=1}^{k+2} \mathrm{I}\left\{0<\beta_{j}<\infty\right\}}{\prod_{j=1}^{k+1} \mathrm{I}\left\{0<\tilde{\beta}_{j}<\infty\right\}} \times \frac{d_{k}[\max (\boldsymbol{x})-\min (\boldsymbol{x})]}{b_{k-1}(k)} \times\left|\tilde{\beta}_{j}+\tilde{\beta}_{j+1}-\beta_{j+1}\right| .
\end{aligned}
$$

Note that $f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\eta_{i}=\sum_{j=1}^{k+2} \beta_{j} \delta_{j i}+$ $\sum_{j=1}^{p} \alpha_{j} v_{j i}$ and standard deviation of $\tau^{-1 / 2}$ and $f_{y}\left(y_{i} \mid k-1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+1} \tilde{\beta}_{j} \tilde{\delta}_{j i}+\sum_{j=1}^{p} \alpha_{j} v_{j i}$ and standard deviation of $\tau^{-1 / 2}$. Also, $\tilde{\delta}_{j i}$ is the $(i)$ th value of the $(j)$ th $I$-spline basis vector created using proposed knot locations after deleting knot $t_{j+1}, \tilde{\boldsymbol{t}}$, and $\delta_{j i}$ is the $(i)$ th value of the $(j)$ th $I$-spline basis vector created using knot locations before deleting $t_{j+1}, \boldsymbol{t}$.

### 4.3.3 Relocation Move

A relocation move keeps the same number of interior knots, $k$, but the location of one knot is moved and the coefficients are updated deterministically based on this location change. From the set of existing ordered interior knots, a knot is randomly selected to be moved and denoted $t_{j+1}$. The new knot location $t^{*}$ is proposed from Uniform $\left(t_{j}, t_{j+2}\right)$ and the proposed knots for the relocation move are given by $\tilde{\boldsymbol{t}}=\left(\tilde{t}_{1}, \ldots, \tilde{t}_{k+2}\right)^{\prime}=\left(t_{1}, \ldots, t_{j}, t^{*}, t_{j+2}, \ldots, t_{k+2}\right)^{\prime}$. To obtain updated coefficients, $\tilde{\boldsymbol{\beta}}=\left(\tilde{\beta}_{1}, \ldots, \tilde{\beta}_{k+2}\right)^{\prime}$, note that if $t^{*} \geq t_{j+1}$ this will cause $\tilde{\beta}_{j}$ to be greater than $\beta_{j}$ as well as $\tilde{\beta}_{j+1}<\beta_{j+1}$ and $\tilde{\beta}_{j+2}<\beta_{j+2}$. Since this change is influenced by the magnitude of the difference between $t^{*}$ and $t_{j+1}$, when $t^{*} \geq t_{j+1}$, the coefficients are
updated according to

$$
\tilde{\beta}_{i}= \begin{cases}\beta_{i} & i=1, \ldots, j-1  \tag{4.11}\\ \beta_{i}+\frac{0.5 r_{2}}{r_{1}+r_{2}+r_{3}} \beta_{i+1} & i=j \\ \frac{r_{1}+0 . r_{3}}{r_{1}+r_{2}+r_{3}} \beta_{i}+\frac{r_{2}}{r_{2}+r_{3}} \beta_{i+1} & i=j+1 \\ \frac{r_{3}}{r_{2}+r_{3}} \beta_{i}+\frac{0.5\left(r_{2}+r_{3}\right)}{r_{1}+r_{2}+r_{3}} \beta_{i-1} & i=j+2 \\ \beta_{i} & i=j+3, \ldots, k+2\end{cases}
$$

where $r_{1}=t_{j+1}-t_{j}, r_{2}=t^{*}-t_{j+1}$, and $r_{3}=t_{j+2}-t^{*}$. Likewise, when $t^{*}<t_{j+1}$, the coefficients are updated according to

$$
\tilde{\beta}_{i}= \begin{cases}\beta_{i} & i=1, \ldots, j-1  \tag{4.12}\\ \frac{r_{1}}{r_{1}+r_{2}} \beta_{i}+\frac{0.5\left(r_{1}+r_{2}\right)}{r_{1}+r_{2}+r_{3}} \beta_{i+1} & i=j \\ \frac{0.5 r_{1}+r_{3}}{r_{1}+r_{2}+r_{3}} i_{i}+\frac{r_{2}}{r_{1}+r_{2}} \beta_{i-1} & i=j+1 \\ \beta_{i}+\frac{0.5 r_{2}}{r_{1}+r_{2}+r_{3}} \beta_{i-1} & i=j+2 \\ \beta_{i} & i=j+3, \ldots, k+2\end{cases}
$$

where $r_{1}=t^{*}-t_{j}, r_{2}=t_{j+1}-t^{*}$, and $r_{3}=t_{j+2}-t_{j+1}$.
Since the relocation move does not involve a change in dimensions, the acceptance probability is $\min \left\{1, A_{M H}\right\}$ where

$$
\begin{aligned}
A_{M H}= & (\text { likelihood ratio }) \times(\text { prior ratio }) \times(\text { proposal ratio }) \\
= & \frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right)\left(t^{*}-t_{j}\right)\left(t_{j+2}-t^{*}\right)}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k-1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \tau\right)\left(t_{j+1}-t_{j}\right)\left(t_{j+2}-t_{j+1}\right)}\left(\frac{\tilde{\beta}_{j} \tilde{\beta}_{j+1} \tilde{\beta}_{j+2}}{\beta_{j} \beta_{j+1} \beta_{j+2}}\right)^{c_{1}-1} \times \\
& \exp \left\{-c_{2}\left[\sum_{j=1}^{k+2} \tilde{\beta}_{j}-\sum_{j=1}^{k+2} \beta_{j}\right]\right\} \times \frac{\prod_{j=1}^{k+2} \mathrm{I}\left\{0<\tilde{\beta}_{j}<\infty\right\}}{\prod_{j=1}^{k+2} \mathrm{I}\left\{0<\beta_{j}<\infty\right\}} \times 1
\end{aligned}
$$

Note that $f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\eta_{i}=\sum_{j=1}^{k+2} \beta_{j} \delta_{j i}+$ $\sum_{j=1}^{p} \alpha_{j} v_{j i}$ and standard deviation of $\tau^{-1 / 2}$ and $f_{y}\left(y_{i} \mid k-1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+2} \tilde{\beta}_{j} \tilde{\delta}_{j i}+\sum_{j=1}^{p} \alpha_{j} v_{j i}$ and standard deviation of $\tau^{-1 / 2}$. Also, $\tilde{\delta}_{j i}$ is the $(i)$ th value of the $(j)$ th $I$-spline basis vector created using proposed knot locations, $\tilde{\boldsymbol{t}}$, and $\delta_{j i}$ is the $(i)$ th value of the $(j)$ th $I$-spline basis vector created using knot locations before moving knot, $\boldsymbol{t}$.

### 4.4 Examples

To examine the performance of the proposed RJMCMC algorithm, we considered applying the algorithm to two simulated data sets with the first data set simulated using a smooth regression sigmoid regression function,

$$
\begin{equation*}
f(x)=5 \cdot \exp (10 x-5) /[1+\exp (10 x-5)] \tag{4.13}
\end{equation*}
$$

(Figure 4.1) and the second data set simulated using a more wiggly piecewise sigmoid function

$$
\begin{align*}
f(x)= & f_{1}(x) I_{[0,0.3]}(x)+\left[f_{1}(0.3)+f_{2}(x)\right] I_{(0.3,0.65]}(x) \\
& +\left[f_{2}(0.65)+f_{3}(x)\right] I_{(0.65,1]}(x) \tag{4.14}
\end{align*}
$$

with

$$
\begin{aligned}
& f_{1}(x)=8 \cdot \exp (40 x-4) /(1+\exp (40 x-4)) \\
& f_{2}(x)=8 \cdot \exp (40 x-20) /(1+\exp (40 x-20)) \\
& f_{3}(x)=8 \cdot \exp (40 x-36) /(1+\exp (40 x-36))
\end{aligned}
$$

(Figure 4.5). The data sets are simulated under the normal errors model with $y_{i}=$ $f\left(x_{i}\right)+\epsilon_{i}$ where $x_{i} i=1, \ldots, n$ are $n=50$ values at equally spaced quantiles over the interval $[0,1]$ and $\epsilon_{i}$ are the random errors generated from a normal distribution with mean zero and a standard deviation of 1.5. The algorithm is run 110,000 iterations with the first 10,000 discarded as burn-in.

### 4.4.1 Smooth Sigmoid Function

Figure 4.1 shows the data set generated from the sigmoid function (dottedline) given in (4.13) and the regression function estimate found using the RJMCMC algorithm discussed in Section 4.3 with $\lambda=5$ (red solid line) along with the $95 \%$ pointwise highest posterior density (HPD) interval for the $\boldsymbol{\eta}=\sum_{j=1}^{m} \beta_{j} \boldsymbol{\delta}_{j}+\alpha \mathbf{1}$
found using the $\hat{\boldsymbol{\eta}}^{(i)}$ values and Chen-Shao estimation algorithm (Chen et al., 2000, Ch 7). Figure 4.1 also shows the constrained maximum likelihood estimator using shape-restricted regression splines, denoted ML SRRS, (dashed line) found using $k=3$ interior knots at the $1 / 4,1 / 2$, and $3 / 4 x$-quantiles marked by "X." Note that both the ML SRRS estimate and the estimate found using RJMCMC, which we will denote RJMCMC, are close to the sigmoid function. This is further supported by considering the square root of the estimate for MSE given by

$$
\begin{equation*}
\widehat{\mathrm{SMSE}}=\sqrt{\frac{1}{n} \sum_{i}^{n}\left(\hat{f}\left(x_{i}\right)-f\left(x_{i}\right)\right)^{2}} \tag{4.15}
\end{equation*}
$$

where $\hat{f}\left(x_{i}\right)$ is the estimate for the function at $x_{i}$ for the given estimation procedure and $f\left(x_{i}\right)$ is the true value of the regression function at $x_{i}$. For ML SRRS, the $\widehat{\text { SMSE }}$ equals 0.40 and for RJMCMC, it equals 0.35 . The $95 \%$ HPD found from the RJMCMC algorithm output contains the true function for all $x_{i}$.


Figure 4.1: Data set generated from sigmoid function (grey dotted line) along with RJMCMC estimate (red solid line), constrained maximum likelihood estimate using shape-restricted regression splines (blue dashed line) fit using three fixed knot points (marked by X's) and pointwise 95\% highest posterior density interval for RJMCMC estimate (shaded region).

To further assess the performance of the RJMCMC algorithm, we examine the mixing of $k$ and the behavior of $\alpha$ and $\tau$ since these two parameters do not change meaning as the value of $k$ changes. Sisson (2005) discussed the current approaches to assessing convergence of RJMCMC algorithms such as using a marked point-process and difficulties in assessing convergence for trans-dimensional Markov chains. Point processes are not easy to construct and are not the focus of this work so we will focus on examining the mixing of $k$ and behavior of $\alpha$ and $\tau$ through some common diagnostics. Figure $4.2(\mathrm{a})$ shows good mixing for $k$ and demonstrates that the algorithm explores the parameter space. The convergence of the RJMCMC is further supported by considering partial trace plots for $\alpha$ and $\tau$ (Figure 4.2(b) and (c)). Likewise, the plots of the auto-correlation function (ACF) for $\alpha$ and $\tau$ (Figure 4.3(c) and (d)) suggest that the algorithm is performing well. The histogram of the posterior for $\tau$ found using RJMCMC has the mode very close to the true value used to simulate the data (Figure 4.3(b)).

Another measure to consider when assessing the performance of our RJMCMC algorithm is the Monte Carlo standard error (MCSE). We use MCSE to help determine if the chain has run long enough. Using the batch method (Givens and Hoeting, 2005, Ch 7) and considering realizations for a given parameter $\theta^{(i)}$ from the MCMC output, the MCSE is found by first separating the iterations after discarding burn-in into $L$ batches ( $L=2000$ for our analysis) of $m$ ( $m=50$ in our analysis) consecutive iterations. The MCSE is found by

$$
M C S E=\frac{1}{\sqrt{L}} \cdot \sqrt{\frac{1}{L-1} \sum_{l=1}^{L}\left(\bar{\theta}_{l}-\bar{\theta}\right)^{2}}
$$

where $\bar{\theta}_{l}$ is the mean of batch $l$ and $\bar{\theta}$ is the mean of the $L$ batch means. As a rule of thumb, if the MCSE for parameter $\theta$ is less than $5 \%$ of the sample standard deviation of the MCMC realizations (minus burn-in) of $\theta$, the chain is thought to have run long enough. The MCSE for $\alpha$ equal to 0.0007 which is less than $5 \%$ of the sample
standard deviation for $\alpha$ (0.011) and the MCSE for $\tau$ equal to 0.0004 which is less than $5 \%$ of the sample standard deviation for $\tau$ (0.005). Hence, when applying the RJMCMC algorithm to this data set under the monotone shape restriction, we find no evidence that our algorithm does not converge to the stationary distribution.
(a) Trace Plot for $k$

(b) Trace Plot for $\alpha$

(c) Trace Plot for $\tau$


Figure 4.2: (a) Partial trace plot for the number of interior knots, $k$ for the sigmoid example. (b) Partial trace plot for $\alpha$ for the sigmoid example. (c) Partial trace plot for the precision for the normal errors, $\tau$, for the sigmoid example.

We also compare the estimates for the regression function when using RJMCMC (free-knot splines) to the estimate found when fixing the number and knot locations by comparing the function estimates for this data set found using the RJMCMC algorithm to the function estimates found fixing knots at minimum of $\boldsymbol{x}=\left(x_{1}, \ldots, x_{n}\right)^{\prime}$, maximum of $\boldsymbol{x}$, and $k$ equally spaced quantiles in the interior of the range of $\boldsymbol{x}$. The function estimates using RJMCMC along with estimates with $k$ equal to 2 (blue dashed), 3 (dark green dotted), 4 (purple dot-dash) and 5 (orange short-long dashed line) are shown in Figure 4.4. We find that having free versus fixed knots gives similar estimates as evidenced by Figure 4.4 and the values $\widehat{\text { SMSE }}$ in Table 4.1.


Figure 4.3: (a) and (b) Histogram of posterior draws after discarding burn-in for the sigmoid example for $\alpha$ and $\tau$, respectively. The vertical red dashed line represents the true value of $\tau$ used to generate the data. (c) and (d) Auto-correlation functions for $\alpha$ and $\tau$, respectively.

For this smooth function, the estimate with fixed knots may be preferred since the computational demand is significantly less and the estimates with and without fixing the knots are very similar. The run time on a Dell ${ }^{\text {TM }}$ Inspiron ${ }^{\text {TM }}$ laptop with an intel ${ }^{\circledR}$ Core ${ }^{\mathrm{TM}} \mathrm{i} 5$ processor and Windows ${ }^{\circledR} 7$ Professional using the 64 bit 2.11.1 version of $R$ is 1.515 minutes for 110,000 iterations of the MCMC algorithm to estimate regression function for the data set in Figure 4.1(a) with fixed knots and $k=3$. The run time for RJMCMC estimate for the same data set on the same computer for the same number of iterations is 11.734 minutes.

### 4.4.2 Wiggly Piecewise Sigmoid Function

To further examine the performance of RJMCMC, we examine the RJMCMC estimate with $\lambda=10$ for the wiggly function (Figure 4.5 ). Figure 4.5 gives the RJMCMC (red solid line) and ML SRRS (blue dashed line) estimates with $k=$


Figure 4.4: Data set generated from sigmoid function along with RJMCMC estimate (solid line) and Bayesian shape-restricted regression spline estimates with fixed interior knots at $k$ equal quantiles over $(0,1)$ with $k$ equal to 2 (blue dashed line), 3 (dark green dotted line), 4 (purple dot-dash line), and 5 (orange short-long dashed line).

9 interior knots (knots marked by "X") fixed at equal quantiles on $(0,1)$. Both are close to the true function with $\widehat{\text { SMSE }}=0.34$ for the RJMCMC estimate and $\widehat{\text { SMSE }}=0.30$ for ML SRRS estimate. The $95 \%$ pointwise HPD intervals found using RJMCMC (Figure 4.5, shaded region) also contain the true function for most $x_{i}$. The partial trace plot for $k$ shows good mixing and the partial trace plots for $\alpha$ and $\tau$ support that the algorithm converged (Figure 4.6). The histograms of the posterior distributions and ACF plots for $\alpha$ and $\tau$ (Figure 4.7) also suggest the algorithm has converged. The MCSE of $\alpha$ and $\tau$ suggests that the algorithm has run long enough. The MCSE for $\alpha$ equal to 0.0007 which is less than $5 \%$ of the sample standard deviation for $\alpha$ (0.011), and the MCSE for $\tau$ equal to 0.0006 which is less than $5 \%$ of the sample standard deviation for $\tau$ (0.005). Thus, the RJMCMC algorithm provides reasonable estimates for the the piecewise sigmoid function without having to specify the number and location of interior knots even for functions that are not very smooth.

Table 4.1: Table of $\widehat{\text { SMSE }}$ values as in (4.15) comparing the Bayes SRRS with $k$ equally spaced interior knots to the function used to generate the data as well as the $\widehat{\text { SMSE }}$ comparing the RJMCMC to the function used to generate the data.

| $k$ | $\widehat{\text { SMSE }}$ |
| :---: | :---: |
| 2 | 0.44 |
| 3 | 0.26 |
| 4 | 0.41 |
| 5 | 0.27 |
| RJMCMC | 0.35 |

As with the smooth sigmoid function in Section 4.4.1, we compare the estimates for the regression function when using RJMCMC to the estimate found when fixing the number and knot locations. The function estimates using RJMCMC along with estimates with $k$ equal to 7 (blue dashed), 8 (dark green dotted), 9 (purple dot dash) and 10 (orange short-long dashed line) are shown in Figure 4.8. Like the smooth sigmoid function, the fixed and free-knot spline estimates are similar.

It is important to note that we found some sensitivity to the choice of the hyperparameter $\lambda$ for the RJMCMC algorithm and found that if $\lambda$ is too small, $\tau$ has slow convergence. Therefore, for functions that are wiggly, such as in Figure 4.5, we recommend higher values of $\lambda$. For a simpler and computationally faster result, one can fix the number and location of the interior knots. However, if one feels uncomfortable fixing the knot locations, RJMCMC along with $I$-splines provide reasonable estimates and is relatively easy to implement. It allows one to average over several different possible knot locations and estimate functions without having to make assumptions about the number and location of knot points.


Figure 4.5: Data set generated from the piecewise sigmoid function (grey dotted line) along with RJMCMC estimate (red solid line), constrained maximum likelihood estimate using shape-restricted regression splines (blue dashed line) fit using 9 fixed knot points (marked by X's), and point wise $95 \%$ highest posterior density interval (shaded region).

(b) Trace Plot for $\alpha$

(c) Trace Plot for $\tau$


Figure 4.6: Partial trace plots for the number of interior knots, $k$ (a), the y-intercept, $\alpha$ (b), and the precision for the normal errors, $\tau$ (c) when using RJMCMC to find regression function estimate for data simulated from the piecewise sigmoid function.


Figure 4.7: (a) and (b) Histogram of parameter realizations in the RJMCMC algorithm after discarding burn-in for $\alpha$ and $\tau$, respectively. The vertical red dashed line represents the true value of $\tau$ used to generate the data. (c) and (d) Auto-correlation functions for $\alpha$ and $\tau$, respectively.


Figure 4.8: Data set generated from the piecewise sigmoid function along with RJMCMC estimate (solid line) and Bayesian shape-restricted regression spline estimates with fixed interior knots at $k$ equal quantiles over $(0,1)$ with $k$ equal to 7 (blue dashed line), 8 (dark green dotted line), 9 (purple dot-dash line), and 10 (orange short-long dashed line).

## Chapter 5

## BAYESIAN SHAPE-RESTRICTED REGRESSION SPLINES MODEL WITH CHANGE-POINTS

### 5.1 Introduction and Motivation

Consider an extension to the Bayes SRRS model discussed in Chapter 2 where there exists an event or events that alter the relationship between covariates and response variable that may violate the shape-restrictions. For instance, consider measuring growth of vegetation in a given region across time but one suspects that there were instances of major shock to the system (such as droughts, fires, floods, disease, or pollution) that may disrupt the monotonically increasing relationship between vegetation growth and time. The number of major shocks and the time when these events occur may be known or unknown. monotonically increasing regression splines can be used to model the relationship between time and vegetation growth between major events with more flexibility than assuming a parametric form such as a line.

We consider a Bayesian model with shape-restricted regression splines and change-points which imposes shape restrictions on the relationship between covariates and a response variable between major events but accounts for events that may violate the shape restrictions with change-points. We define change-points as covariate values for which violations of smoothness or the shape restriction occur in the regression function. To obtain a multiple change-point model, we introduce additional basis functions and combine them with the basis function for the Bayes SRRS model in Chapter 2.

We begin with some background on multiple change-point analysis in Section 5.2. We propose a multiple change-point model with shape-restricted regression splines in a Bayesian framework in Section 5.3. In Section 5.4, we propose a RJMCMC algorithm to estimate regression functions with $h$ change-points with the number of change-points and their locations known. In Section 5.5, we propose a RJMCMC algorithm for function estimation for a model with a single changepoint whose location is unknown where we allow the number and location of knots points between the minimum covariate value and the change-point be random as well as allow the number and location of knots points between the change-point and the maximum covariate to be random. We conclude the chapter by proposing a RJMCMC algorithm to determine whether to use a Bayes SRRS model without change-points or use a model with a single change-point. This model can be used to determine the existence of a change-point for a given data set when it is assumed that the relationship should be monotonically increasing between covariate and response variable.

### 5.2 Background

Change-point models have been studied extensively and are currently a hot topic in statistics literature. Non-parametric tests for the existence of a changepoint have been around for quite awhile. Wolfe and Schechtman (1984) performed a simulation study to compare the performance of several non-parametric tests used to determine whether or not a change-point is appropriate for a given data set. In the context of regression analysis, Muller (1992) provided a non-parametric changepoint model to detect a single jump point or discontinuity in slope using kernel smoothers. The change-point location was found by maximizing the difference in function estimates using one-sided kernels. He provided asymptotic distributions and rates of convergence for a Gaussian process. Braun and Muller (1998) extended
this idea and examined the detection of multiple change-points using local polynomial estimation with applications to the deoxyribonucleic acid (DNA) sequencing. They assumed smooth polynomial functions between jump-points and suggested using cross validation or a threshold for estimated jump-heights to determine the number of jump points. Once the number of jump-points was found along with an appropriate bandwidth, local polynomials were fit to segments of data between estimated jump-points using an appropriate kernel. Other models that involve estimating discontinuous regression functions using non-parametric smoothers between change-points can be found in Wu and Chu (1993), Eubank and Speckman (1994), Muller and Song (1997), Qiu and Yandell (1998), Gijbels et al. (1999), and Gijbels and Goderniaux (2004), among others. The majority of these articles allowed discontinuities in slope in the regression function as well as jumps. Koo (1997) estimated regression functions with discontinuities using piecewise linear splines between discontinuities in a frequentist framework. The placement of knot and change-points were determined by an algorithm that involves stepwise knot addition, stepwise knot deletion, as well as deletion of basis vectors and what they termed "knot merging." They used knot merging to determine whether a continuous or discontinuous linear spline was appropriate at a given candidate location for a change-point. Gijbels et al. (2007) proposed a method for estimating regression curves with unknown discontinuities using local linear kernel smoothing and an algorithm which, for a given value of $x$, decides between a local linear estimate based on data points to left of $x$, data points to the right of $x$, or data points to the left and right of $x$.

Change-point models in a Bayesian framework have also been extensively studied. Carlin et al. (1992) used a hierarchical Bayesian model along with a Gibbs sampler to model data with at most one change-point for a Poisson process model and simple linear regression. Raftery (1993) produced a fully Bayesian framework to model data where at most one change-point is assumed to occur and considered
a linear regression model and a Poisson process. He extended his model to two dimensions and considered change curves. Chib (1998) provided a general framework to estimate multiple change-point models using a Bayesian approach with a latent discrete state variable used to determine the number of change-points. This method involved a Markov chain Monte Carlo sampling algorithm and was applied to binary data and a Poisson process. However, it was not applied to non-parametric models or regression splines. Rotondi (2002) also considered a change-point model with an unknown number of change-points and change-point locations. He used a RJMCMC algorithm to model a Poisson process and applied his method to earthquake data. Fearnhead (2006) provided an algorithm to estimate the number and location of multiple change-points in a Bayesian setting that is based on recursions as opposed to the RJMCMC algorithm and applied this algorithm to several data sets including one that was fit using a piecewise constant model. However, he did not consider any higher order regression splines or shape restrictions. He considered two priors on the number and location of change-points with one prior similar to the one used to in the RJMCMC algorithm to fit piecewise quadratic $I$-splines in Chapter 4 and the other based on a point process on positive and negative integers. His procedure was able to generate draws for the number of change-points separate from the other parameters which can not be duplicated when there are constrained regression splines. Giordani and Kohn (2008) approached change-points from a different perspective and modeled a time series with unknown magnitudes and number of changes to parameters using a state-space representation. They modeled changepoints using mixture distributions for state innovations. Thomson et al. (2010) used linear regression splines along with change-points to model changes in the relationship between abundance of pelagic fish species and time. They considered a model that identifies change-points common to multiple species. They essentially fit multiple change-point models with piecewise linear splines fit between jump-points. The knots for the piecewise linear splines as well as the jump points were treated as
random and RJMCMC was used to estimate these values. We extend this idea by considering piecewise quadratic splines and impose shape restrictions on the spline functions. Since we are dealing with piecewise splines of a degree higher than one, we also allow changes in slope as well as jumps at change-points.

Multiple change-points models with shape-restrictions have also been studied. Holmes and Heard (2003) used piecewise constant functions to model monotonic regression functions where the location and the number of steps were random. They used a RJMCMC algorithm with priors for the unconstrained step function and discarded draws where the monotone constraint was not met in order to estimate a monotonically increasing function. For normal errors data, Alvarez and Dey (2009) proposed a fully Bayesian parametric model for function estimation for several different shape restrictions including monotonicity that allowed for change-points. However, this model did not allow one to directly estimate the location and the number of change-points and required the specification of a "tuning" parameter which determined the smoothness of the fit. Ma and Yang (2010) provided a model for detecting existence of jump points by using either piecewise constant or piecewise linear $B$ splines. If change-points were found to exist, they used BIC as well as a hypothesis test to determine the location and magnitude of the jump points. We extend previous work by proposing Bayesian models using quadratic $I$-splines which allow us to imposed shape-restrictions on the regression functions between the change-points simply by restricting some of the regression coefficients to be positive.

### 5.3 Multiple Change-point Shape-restricted Regression Spline Model

The proposed multiple change-point model uses shape-restricted regression splines. We begin by considering the shape-restricted regression spline model in Section 2.2 with $\eta_{i}$ as defined in (2.2). We generalize the Bayes SRRS model to a multiple change-point model under the monotone shape restriction and redefine the spline approximation of $\boldsymbol{\eta}$ in (2.3) to include change-points. When we say the function
$f(x)$ has a change-point at $x=\xi_{i}$, we mean the function $f$ has a discontinuity, i.e., unequal left and right first derivatives, or a violation of the shape restriction at $\xi_{i}$ by a jump down at a change-point. This model allows for the estimation of regression functions with a finite number of violations of the shape restriction. To simplify notation and the explanation of the multiple change-point shape-restricted spline model, we describe a model for a single continuous covariate and assume we observe data $\left(x_{i}, y_{i}\right)$ for $i=1, \ldots, n$ with $\boldsymbol{x}=\left(x_{1}, \ldots, x_{n}\right)^{\prime}$ and $\boldsymbol{y}=\left(y_{1}, \ldots, y_{n}\right)^{\prime}$. However, this model can be extended to include additional covariates.

### 5.3.1 Shape-restricted Spline Approximation

The shape-restricted spline approximation is created by first defining a new set of basis functions and then taking a linear combination of these basis functions to create our function estimate. For a model with $h$ change-points, we assume $k$ interior knot points with which to create the basis functions. We assume that the change-points occur at the interior knot points so $k \geq h$. We define a knot location vector $\boldsymbol{t}=\left(t_{1}, \ldots, t_{k+2}\right)^{\prime}$ with $\min (\boldsymbol{x})=t_{1}<t_{2}<\ldots<t_{k+1}<t_{k+2}=$ $\max (\boldsymbol{x})$. Let the ordered change-point location vector be $\boldsymbol{\xi}=\left(\xi_{0}, \xi_{1}, \ldots, \xi_{h}, \xi_{h+1}\right)^{\prime}$ with $\min (\boldsymbol{x})=\xi_{0}<\xi_{1}<\ldots<\xi_{h}<\xi_{h+1}=\max (\boldsymbol{x})$ with the $h$ change-points, $\xi_{q}$ for $q=1, \ldots, h$. Since the change-points are required to occur at interior knot points, $\left(\xi_{1}, \ldots, \xi_{h}\right)=\left(t_{j_{1}}, \ldots, t_{j_{h}}\right)$ for $j_{i} \in\{2, \ldots, k+1\}$ and $i=1, \ldots, h$. Let $J=\left\{j_{1}, \ldots, j_{h}\right\}$. We define basis functions $\delta_{j}(x)$ for $j=1, \ldots, m=k+2$ to be the $(j)$ th quadratic $I$-spline basis functions as defined in (1.8) created using $\boldsymbol{x}$ and $\boldsymbol{t}$ where $\delta_{j}$ has a positive slope at the $(j)$ th largest knot location. We define basis functions $\delta_{m+q}(x)$ for $q=1, \ldots, h$ by

$$
\begin{equation*}
\delta_{m+q}(x)=\left(x-\xi_{q}\right) \mathrm{I}\left\{x>\xi_{q}\right\} \tag{5.1}
\end{equation*}
$$

where I $\{\cdot\}$ denotes the indicator function and note these basis functions allow for a difference in the left and right derivatives at the change-points. Define $v_{1}(x)=$
$\mathrm{I}\left\{x \leq \xi_{1}\right\}, v_{h+1}=\mathrm{I}\left\{x>\xi_{h}\right\}$, and $v_{q}(x)$ for $q=2, \ldots, h$ by

$$
\begin{equation*}
v_{q}(x)=\mathrm{I}\left\{\xi_{q-1}<x \leq \xi_{q}\right\} \tag{5.2}
\end{equation*}
$$

and note these functions allow for discontinuities in $f$ at the change-points. Using these basis functions, we approximate $f\left(x_{i}\right)$ by

$$
\begin{equation*}
\sum_{j=1}^{m+h} \beta_{j} \delta_{j}\left(x_{i}\right)+\sum_{j=1}^{h+1} \alpha_{j} v_{j}\left(x_{i}\right) \tag{5.3}
\end{equation*}
$$

where the shape restriction is imposed between change-points by constraining the $\boldsymbol{\beta}=\left(\beta_{1}, \ldots, \beta_{m+h}\right)$ parameter vector. The constraints on $\boldsymbol{\beta}$ are described in Section 5.3.2. Note that if $h=0$ then we have the shape-restricted spline model as in Chapter 2.

### 5.3.2 Monotonicity Constraint

To impose the monotonically increasing shape restriction between change-points, we require $\eta^{\prime}(x)>0$ (or $\eta^{\prime}(x)<0$ for monotonically decreasing) between changepoints. Note that since we estimate $\eta\left(x_{i}\right)$ with a piecewise quadratic polynomial spline, the derivative of our estimate will be a piecewise linear polynomial spline. Therefore, the derivative, given by

$$
\sum_{j=1}^{m} \beta_{j} \delta_{j}^{\prime}(x)+\sum_{j=1}^{h} \beta_{m+j} \mathrm{I}\left\{x>\xi_{j}\right\}
$$

will be positive (negative) if it is positive (negative) at each knot point

$$
\boldsymbol{t}=\left(\min (x)=t_{1, \ldots} \ldots, t_{m}=\max (x)\right)
$$

and the left and right derivatives evaluated at each $\xi_{q}$ are positive (negative). For quadratic $I$-splines, $\delta_{j}^{\prime}(x)$ is a degree one $M$-spline and $\delta_{j}^{\prime}\left(t_{l}\right)$ is nonzero for $j=l$ only. Furthermore, each basis function can be scaled such that

$$
\begin{aligned}
\delta_{j}^{\prime}\left(t_{j}\right) & =1 \\
\delta_{j}^{\prime}\left(t_{i}\right) & =0 \text { for } i \neq j
\end{aligned}
$$

Therefore, we can enforce monotonicity by requiring $\beta_{j} \geq 0$ for $j=1, \ldots, j_{1}, \beta_{j}+$ $\sum_{q=1}^{i} \beta_{m+q} \geq 0$ for $j_{i} \leq j \leq j_{i+1}$ and $i=1, \ldots, h-1$, and $\beta_{j}+\sum_{q=1}^{h} \beta_{m+q} \geq 0$ for $j_{h} \leq j \leq m$.

We can reparameterize the model such that the constraints can be imposed simply by requiring coefficients of some of the transformed basis functions to be positive. This is best illustrated by using matrix notation. Define $\Delta$ as a matrix with columns $\boldsymbol{\delta}_{j}=\left(\delta_{j 1}, \ldots, \delta_{j n}\right)^{\prime}$ for $j=1, \ldots m+h$ where $\delta_{j i}=\delta_{j}\left(x_{i}\right)$ and $V$ as a matrix with columns $\boldsymbol{v}_{j}=\left(v_{j 1}, \ldots, v_{j n}\right)^{\prime}$ for $j=1, \ldots, h+1$. The vectors $\boldsymbol{\delta}_{j}$ are scaled such that $\boldsymbol{\delta}_{j}^{\prime} \boldsymbol{v}_{i}=0$ for $i=1, \ldots, h+1$. Let $\boldsymbol{\beta}=\left(\beta_{1}, \ldots \beta_{m+h}\right)^{\prime}$ and $\boldsymbol{\alpha}=\left(\alpha_{1}, \ldots, \alpha_{h+1}\right)^{\prime}$. Thus, using matrix notation, the shape-restricted regression spline estimate is

$$
\begin{equation*}
\Delta \boldsymbol{\beta}+V \boldsymbol{\alpha} \tag{5.4}
\end{equation*}
$$

Using this matrix notation, the constraints can be imposed by letting

$$
S=\left[\begin{array}{ccccc}
I_{m} & \boldsymbol{J}_{j_{1+1}} & \boldsymbol{J}_{j_{2}+1} & \cdots & \boldsymbol{J}_{j_{h}+1} \\
\boldsymbol{e}_{j_{1}} & 1 & 0 & \cdots & 0 \\
\boldsymbol{e}_{j_{2}} & 1 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
\boldsymbol{e}_{j_{h}} & 1 & 1 & \cdots & 1
\end{array}\right]
$$

where $I_{m}$ is an $m \times m$ identity matrix, $\boldsymbol{e}_{j_{i}}$ is a $1 \times(m+h)$ vector with elements of zero except for the $j_{i}$ element that is equal to one, $\boldsymbol{J}_{j_{i}+1}$ is a $(m+h) \times 1$ vector with elements of zero for rows 1 to $j_{i}$ and elements of 1 from rows $j_{i}+1$ to $m+h$. The shape restriction is met by requiring $S \boldsymbol{\beta} \geq 0$ ( $S \boldsymbol{\beta} \leq 0$ for monotonically decreasing), where $A \geq 0$ implies each element in $A$ is greater than or equal to zero. Let $\boldsymbol{b}=\left(b_{1}, \ldots, b_{m+h}\right)^{\prime}=S \beta$ and thus

$$
\begin{equation*}
\boldsymbol{\beta}=S^{-1} \boldsymbol{b} \tag{5.5}
\end{equation*}
$$

Using the reparameterization in (5.5) and substituting in (5.4), our estimate is

$$
\begin{equation*}
\Delta S^{-1} \boldsymbol{b}+V \boldsymbol{\alpha}=\Delta^{*} \boldsymbol{b}+V \boldsymbol{\alpha} \tag{5.6}
\end{equation*}
$$

Thus, we can approximate $f\left(x_{i}\right)$ by

$$
\begin{equation*}
\sum_{j=1}^{m+h} b_{j} \delta_{j i}^{*}+\sum_{j=1}^{p+h} \alpha_{j} v_{j i} \tag{5.7}
\end{equation*}
$$

where $\boldsymbol{\delta}_{j}^{*}=\left(\delta_{j 1}^{*}, \ldots, \delta_{j n}^{*}\right)^{\prime}$ is the $(j)$ th column of $\Delta^{*}=\Delta S^{-1}$. A monotone spline estimate is obtained by requiring $b_{j} \geq 0$ for all $j=1, \ldots, m+h$. It is important to note that $\Delta^{*}$ will have a column equal to a zero vector if there are no data points between a change-point and any of the interior knot points adjacent to that change-point. When performing analysis using any of the three RJMCMC algorithms discussed in Sections 5.4, 5.5, and 5.6, this will result in a division by zero. Thus, we want to avoid placing knot points that are too close together.

For the Bayesian model, we assume the prior for $\boldsymbol{\alpha}=\left(\alpha_{1}, \ldots, \alpha_{p+h}\right)^{\prime}$ is multivariate normal with mean equal to a zero vector of length $h+1$ and variance equal to $M I_{h+1}$ where $M$ is a constant chosen by the user and $I_{h+1}$ is a $(h+1) \times(h+1)$ identity matrix. For all the examples, we used $M=1000$. We assume the $b$ parameters are independent and note that the prior for each $b_{j}$ coefficient should have support on the positive reals and choose a gamma prior with a shape value of $c_{1}$ and a rate value of $c_{2}$ for each $b_{j}$. The values for $c_{1}$ and $c_{2}$ can again be chosen such that the mean of the gamma prior is large than the variance is relatively small to aid in the estimation of functions with "flat" spots.

### 5.3.3 Normal Errors Model

We now focus on estimating the regression function under the normal errors model. Suppose we observe data $\left(x_{i}, y_{i}\right)$ with $i=1, \ldots, n$ and $y_{i}=f\left(x_{i}\right)+\epsilon_{i}$ where $\epsilon_{i}$ is a random normal error with variance $\sigma^{2}$ and precision $\tau=1 / \sigma^{2}$. We assume that $f$ is monotonically increasing and smooth except at $h$ change-points. To estimate $f$ we use the multiple change-point shape-restricted spline regression model discussed in Section 5.3.2. We propose a Bayesian model using the priors for $\boldsymbol{\alpha}$ and $\boldsymbol{b}$ as given in Section 5.3.2 and assume $\tau \sim \operatorname{Gamma}\left(d_{1}, d_{2}\right)$, where the prior
parameters are chosen so that the mean of the prior density, $d_{1} / d_{2}$, is the inverse of a guess for the model variance.

The likelihood for the normal errors model is proportional to

$$
\begin{equation*}
L(\boldsymbol{\alpha}, \boldsymbol{b}, \tau ; \boldsymbol{Y}) \propto \tau^{n / 2} \exp \left\{-\frac{\tau}{2} \sum_{i=1}^{n}\left(y_{i}-\sum_{j=1}^{m+h} b_{j} \delta_{j i}^{*}-\sum_{j=1}^{h+1} \alpha_{j} v_{j i}\right)^{2}\right\} \tag{5.8}
\end{equation*}
$$

and the joint prior density is

$$
\begin{align*}
p(\boldsymbol{\alpha}, \boldsymbol{b}, \tau) \propto & {\left.\left[\prod_{j=1}^{m+h} \mathrm{I}\left\{0<b_{j}<\infty\right\}\right)\right]\left[\prod_{j=1}^{m+h}\left(b_{j}^{c_{1}-1} \exp \left\{-b_{j} c_{2}\right\}\right)\right] \times } \\
& {\left[\exp \left\{-\sum_{j=1}^{h+1} \frac{1}{2 M} \alpha_{j}^{2}\right\} \tau^{d_{1}-1} \exp \left\{-d_{2} \tau\right\} \mathrm{I}\{0<\tau<\infty\}\right] . } \tag{5.9}
\end{align*}
$$

The posterior distribution is proper but analytically intractable, so Markov chain Monte Carlo methods (Givens and Hoeting, 2005, Ch 7) are used to obtain samples from the posterior distribution as with normal errors model in Chapter 2.

In particular, we use a Gibbs sampler to sample from the posterior distributions and the conditional distributions used in this sampler are given below. Let $\boldsymbol{b}_{\left(-j_{0}\right)}$ be the $\boldsymbol{b}$ vector with $b_{j_{0}}$ removed and similarly let $\boldsymbol{\alpha}_{\left(-j_{0}\right)}$ be the $\boldsymbol{\alpha}$ vector with $\alpha_{j_{0}}$ removed. The conditional posterior density for $\alpha_{j_{0}}$, given the data, $\tau, \boldsymbol{b}$, and $\boldsymbol{\alpha}_{\left(-j_{0}\right)}$ is given by

$$
\begin{equation*}
p\left(\alpha_{j_{0}} \mid \boldsymbol{b}, \boldsymbol{\alpha}_{\left(-j_{0}\right)}, \tau, \boldsymbol{y}\right) \sim \mathrm{N}\left(\frac{\tau}{\tau_{\alpha_{0}}} \sum_{i=1}^{n} r_{i} v_{j i},\left(\tau_{\alpha_{j_{0}}}\right)^{-1}\right), \tag{5.10}
\end{equation*}
$$

where $\tau_{\alpha_{j_{0}}}=1 / M+\tau \sum_{i=1}^{n}\left(v_{j_{0} i}\right)^{2}, r_{i}=y_{i}-\sum_{j=1}^{m+h} b_{j} \delta_{j i}^{*}-\sum_{j \neq j_{0}} \alpha_{j} v_{j i}$, and the sum over $j \neq j_{0}$ denotes the sum over $j=1, \ldots, h+1$ excluding $j=j_{0}$. The conditional posterior density for $\tau$ given the data, $\boldsymbol{b}$ and $\boldsymbol{\alpha}$ coefficients, is

$$
\begin{equation*}
p(\tau \mid \boldsymbol{b}, \boldsymbol{\alpha}, \boldsymbol{y}) \sim \operatorname{Gamma}\left(d_{1}+n / 2, d_{2}+\mathrm{SSE} / 2\right), \tag{5.11}
\end{equation*}
$$

where $\mathrm{SSE}=\sum_{i=1}^{n}\left(y_{i}-r_{i}\right)^{2}$ with $r_{i}=\sum_{j=1}^{m+h} b_{j} \delta_{j i}^{*}+\sum_{j=1}^{h+1} \alpha_{j} v_{j i}$ (the sum of squared residuals given the coefficients). The conditional posterior density for $b_{j_{0}}$, given the
data, $\tau, \boldsymbol{\alpha}$, and $\boldsymbol{b}_{\left(-j_{0}\right)}$ is

$$
\begin{gather*}
p\left(b_{j_{0}} \mid \boldsymbol{b}_{\left(-j_{0}\right)}, \boldsymbol{\alpha}, \boldsymbol{y}\right) \propto \\
b_{j_{0}}^{c_{1}-1} \exp \left\{-\frac{s_{j_{0}}^{*} \tau}{2}\left[b_{j_{0}}-\left(\sum_{i=1}^{n} \frac{r_{i} \delta_{j_{0} i}^{*}}{s_{j_{0}}^{*}}-\frac{c_{2}}{s_{j_{0}}^{*} \tau}\right)\right]^{2}\right\} \times  \tag{5.12}\\
\mathrm{I}\left\{0<b_{j_{0}}<\infty\right\}
\end{gather*}
$$

where $r_{i}=y_{i}-\sum_{j \neq j_{0}} b_{j} \delta_{j i}^{*}-\sum_{j=1}^{h+1} \alpha_{j} v_{j i}$, the sum over $j \neq j_{0}$ denotes the sum over $j=1, \ldots, m+h$ excluding $j=j_{0}$, and $s_{j_{0}}^{*}=\sum_{i=1}^{n}\left(\delta_{j_{0} i}^{*}\right)^{2}$. The density is of the form $f(x) \propto x^{a} \exp \left\{-d(x-c)^{2}\right\} \mathrm{I}\{x>0\}$ where $x=b_{j_{0}}, a=c_{1}-1, d=s_{j_{0}}^{*} \tau / 2>0$, $c=\sum_{i=1}^{n} r_{i} \delta_{j_{0} i}^{*} / s_{j_{0}}^{*}-c_{2} /\left(s_{j_{0}}^{*} \tau\right)$, and I $\{\cdot\}$ is the indicator function. This is the same form as the conditional distribution for each $\beta_{l_{0} j_{0}}$ in Section 2.4 and can be sampled from using the Metropolis-Hastings algorithm or the auxiliary variable Markov chain Monte Carlo technique (Meyer and Laud, 2002; Givens and Hoeting, 2005, Ch 8.1) discussed in Section 2.4.

### 5.4 Multiple Change-point Model with Change-point Locations Fixed

In this section, we consider function estimation when change-points are known. Assume the normal errors multiple change-point model discussed in Section 5.3.3. Further, we assume $h$ change-points with locations known and that the function is monotonically increasing between the change-points. We also assume the function is monotonically increasing between the minimum value for the covariate and the smallest change-point as well as monotonically increasing between the largest change-point and the maximum value for the covariate. We refer to the minimum and maximum values of the covariate as the endpoints. We estimate the regression function using free-knot splines with the number and location of interior knots between change-points as well as number and location of interior knots between endpoints and change-points unknown using a RJMCMC algorithm, which we call the RJMCMC 1 algorithm. The RJMCMC 1 algorithm estimates regression function with $h$ change-points where the change-points locations are known and fixed. For
this algorithm, we scale $\boldsymbol{x}$ such that its range is the unit interval with min $(\boldsymbol{x})=0$ and $\max (\boldsymbol{x})=1$ and scale $\boldsymbol{y}$ such that it has a mean of zero and variance of one.

### 5.4.1 Overview of the RJMCMC 1 Algorithm for Fixed Change-points

The RJMCMC 1 algorithm is similar to the RJMCMC algorithm for Bayes SRRS model without change-points proposed in Chapter 4. The RJMCMC 1 algorithm can be summarized as follows:

1. Perform a "birth", "death", or "relocation" move and accept the proposed parameters with a given acceptance probability.
2. Perform a "coefficient update" step using a Gibbs sampler step to update $\boldsymbol{\alpha}$, $\boldsymbol{b}$, and $\tau$.
3. Repeat until convergence.

The birth, death, and relocation moves are performed with probabilities $b_{k}, d_{k}$, and $r_{k}$, respectively, with these probabilities given in the next paragraph. The birth, death, and relocation moves are discussed further in Sections 5.4.3.1, 5.4.3.2, and 5.4.3.3, respectively. Once a birth, death, or relocation move has been performed, we perform the coefficient update step using the values of $k$ (the current number of interior knots including change-points) and the interior knot locations obtained from the birth, death, or relocation move. This is a Gibbs sampler step where the conditional distributions in Section 5.3.3 are used to update the parameter values for $\boldsymbol{b}, \boldsymbol{\alpha}$, and $\tau$.

Let $k$ be the current number of interior knots including change-points and define $b_{k}=c \cdot \min \left\{1, f_{k}(k+1) / f_{k}(k)\right\}, d_{k}=c \cdot \min \left\{1, f_{k}(k-1) / f_{k}(k)\right\}, r_{k}=1-b_{k}-d_{k}$, and $c=0.4$ where $f_{k}(\cdot)$ is the density of the prior on the number of interior knots (discussed in Section 5.4.2). We place bounds on the possible values for $k$ with lower bound denoted by $k_{\min }$ and upper bound denoted by $k_{\max }$. We choose to let
$k_{\text {min }}=2 h+1$ because we require interior knots at each change-point location, at least one interior knot between change-points, at least one interior knot between the minimum value of the covariate and the smallest change-point, and at least one interior knot between the largest change-point and the maximum value of the covariate. We suggest a large value for $k_{\max }$ but suggest that is be much smaller than $n$ to ensure quite a few data points between knot points. For the examples in this chapter, we would like many data points between knot points so use $k_{\max }=n / 4$ or $k_{\max }=n / 5$. To avoid proposing moves with $k$ outside of $\left[k_{\min }, k_{\max }\right]$, we let $b_{k}=1$, $d_{k}=0$, and $r_{k}=0$ if $k=k_{\min }$ knots and $b_{k}=0, d_{k}=1$, and $r_{k}=0$ if $k=k_{\max }$ knots.

As with the RJMCMC algorithm in Chapter 4, for the RJMCMC 1 algorithm, the parameters proposed in each move in the first step of the algorithm are accepted with probability $\min \{1, A\}$ with

$$
\begin{equation*}
A=(\text { likelihood ratio }) \times(\text { prior ratio }) \times(\text { proposal ratio }) \times \mid \operatorname{det}(\text { Jacobian }) \mid \tag{5.13}
\end{equation*}
$$

where the acceptance probability for the birth, death, and relocation moves are given in Sections 5.4.3.1, 5.4.3.2, and 5.4.3.3, respectively.

### 5.4.2 Priors for the RJMCMC 1 Algorithm

Since the number and location of the interior knots that are not change-points are random, we need to assume priors for the number of interior knots, $k$, and the interior knot locations, $\boldsymbol{t}=\left(t_{1}, \ldots, t_{k+2}\right)^{\prime}$. For $k$, we assume a $\operatorname{Poisson}(\lambda)$ prior truncated on $\left(k_{\min }, k_{\max }\right)$ as in Section 4.2.2 with values of $k_{\min }$ and $k_{\max }$ given in Section 5.4.1. As with the RJMCMC algorithm in Chapter 4, the hyperparameter $\lambda$ can be adjusted to help improve mixing. For the examples in this section, we choose a value $\lambda$ that is at least as large as $k_{\text {min }}$.

We need to take two things into account when assuming a prior for $\boldsymbol{t}$. Since the change-point locations are known and they are assumed to be interior knots then
we only have $k-h$ "free" interior knots. We also need to enforce the restriction that there is at least one interior knot between change-points as well as one interior knot between the minimum value of the covariate and the closest change-point and one interior knot between the maximum value of the covariate and the closest changepoint. Thus, for $h+1$ interior knot locations, we have the following prior distribution

$$
\prod_{i=1}^{h+1} \frac{1}{\xi_{i}-\xi_{i-1}}
$$

where $\boldsymbol{\xi}=\left(t_{1}=\xi_{0}, \xi_{1}, \ldots, \xi_{h}, t_{k+2}=\xi_{h+1}\right)^{\prime}$ is as in Section 5.3.1 with $\xi_{0}<\xi_{1}<\cdots<$ $\xi_{h}<\xi_{h+1}$. If $k>(2 h+1)$, the prior for these $k-(2 h+1)$ interior knots are independent uniform distributions over the range of $\boldsymbol{x}$ excluding small intervals around the change-points to avoid having interior knot points too close to the change-points. Let $d_{l i}$ be the second largest $x$ value contained in the interval $\left(t_{j}, t_{j+1}\right)$ where $t_{j+1}=\xi_{i}$. Thus, of all interior knots smaller than the change-point, $t_{j}$ is the interior knot that is closest to $\xi_{i}$. Let $d_{u i}$ be the second smallest $x$ value that is contained in the inter$\operatorname{val}\left(t_{j+1}, t_{j+2}\right)$ where $t_{j+1}=\xi_{i}$. Let $\boldsymbol{d}=\left(t_{1}=d_{u 0}, d_{l 1}, d_{u 1} \ldots, d_{l h}, d_{u h}, d_{l, h+1}=t_{k+2}\right)$. Thus, the prior for these $k-(2 h+1)$ interior knot locations is the product of independent uniform distributions over

$$
\bigcup_{i=1}^{h+1}\left(d_{u, i-1}, d_{l, i}\right)
$$

Therefore, the priors for each of the $k-(2 h+1)$ interior knot locations (if they exist) is a uniform distribution with density

$$
\frac{1}{\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)}
$$

For the remaining parameters in the Bayesian model, we used the priors given in Sections 5.3.2 and 5.3.3. We use independent gamma priors with shape parameter $c_{1}$ and rate parameter $c_{2}$ for the constrained basis spline coefficients (b) and vague normal priors with mean 0 and large variance $M$ for the unconstrained basis splines and coefficients $(\boldsymbol{\alpha})$. We use a gamma prior with shape parameter $d_{1}$ and rate parameter $d_{2}$ for the precision of the random error $\left(\tau=1 / \sigma^{2}\right)$.

### 5.4.3 RJMCMC 1 Algorithm Implementation

In this section, we describe how we propose the new knot locations for each move type of the RJMCMC 1 algorithm where the change-points are fixed. We explain how we propose new values for the restricted regression coefficients and also derive the acceptance probability for each move type.

### 5.4.3.1 Birth Move

In the birth move, we add a new interior knot that is not a change-point to a model. As with the birth move in Section 4.3.1, we move from a model with $k$ interior knot locations to a model with $k+1$ interior knot locations and update the coefficients based on the addition of the interior knot. Let $\boldsymbol{t}=\left(t_{1}, \ldots, t_{k+2}\right)^{\prime}$ be the ordered current knot locations for the model with $k$ interior knots. Let $\xi_{1}, \ldots, \xi_{h}$ be the ordered known change-points with $\left(\xi_{1}, \ldots, \xi_{h}\right)^{\prime}=\left(t_{j_{1}}, \ldots, t_{j_{h}}\right)^{\prime}$ for $j_{i} \in\{2, \ldots, k+1\}$ and $i=1, \ldots, h$. Let $J=\left\{j_{1}, \ldots, j_{h}\right\}$. As in Section 5.3.1, let $\boldsymbol{\xi}=\left(t_{1}=\xi_{0}, \xi_{1}, \ldots, \xi_{h}, t_{k+2}=\xi_{h+1}\right)^{\prime}$. For the change-point model, we require that the new interior knot location is not a change-point and also want to avoid adding an interior knot too close to existing interior knots. To meet these requirements, we first find a set of candidate intervals defined by
$t_{\text {cand }}=\left\{\left(t_{l}, t_{l+1}\right): l=1, \ldots, k+1\right.$ and at least four $x_{i}^{\prime} s$ are contained within $\left.\left(t_{l}, t_{l+1}\right)\right\}$
and denote the cardinality of $t_{\text {cand }}$ by $n_{t}$. From $t_{\text {cand }}$, we randomly select an interval with which to generate the new knot point and denote this interval by $\left(t_{j}, t_{j+1}\right)$. The new knot location, denoted $t^{*}$, is then generated randomly from a continuous uniform distribution on $(l b, u b)$. The bounds on the uniform distribution, $l b$ and $u b$, are found by defining $\mathrm{U}_{j}=\left\{x_{i}: x_{i} \in\left(t_{j}, t_{j+1}\right)\right\}$. Let $l b$ be the second smallest $x_{i}$ in $\mathrm{U}_{j}$ unless $j=1$. If $j=1$, let $l b=\min (\boldsymbol{x})=0$. Let $u b$ be the second largest $x_{i}$ in $\mathrm{U}_{j}$ unless $j+1=k+2$. If $j+1=k+2$, let $u b=\max (\boldsymbol{x})=1$. The proposed knot location vector is $\tilde{\boldsymbol{t}}=\left(\tilde{t}_{1}, \ldots, \tilde{t}_{k+3}\right)^{\prime}=\left(t_{1}, \ldots, t_{j}, t^{*}, t_{j+1}, \ldots, t_{k+2}\right)^{\prime}$.

With the addition of $t^{*}$, we need to propose new values for the $b$ parameters that preserve the shape restriction but also produce a function estimate that preserves the properties of the current function estimate such as a function estimate with the same range. Let $\boldsymbol{b}=\left(b_{1}, \ldots, b_{k+2+h}\right)^{\prime}$ be the current $b$ values for the model with $k$ interior knots and let $\tilde{\boldsymbol{b}}=\left(\tilde{b}_{1}, \ldots, \tilde{b}_{k+3+h}\right)^{\prime}$ be the proposed $b$ values after adding $t^{*}$. Let $u$ be a random variable generated from a uniform distribution over $(0,1)$. To update $\boldsymbol{b}$, we consider two possible cases. The first case occurs when $t_{j}$ is not a change-point and the second case occurs when $t_{j}$ is a change-point. For the first case where $t_{j}$ is not a change-point, let

$$
\tilde{b}_{l}= \begin{cases}b_{l} & \text { for } l=1, \ldots j-1  \tag{5.14}\\ (1-u) b_{l} & \text { for } l=j \\ u\left(b_{l-1}+b_{l}\right) & \text { for } l=j+1 \\ (1-u) b_{l-1} & \text { for } l=j+2 \\ b_{l-1} & \text { for } l=j+3, \ldots, k+3+h\end{cases}
$$

Note that this coefficient update function only changes coefficients of $\boldsymbol{\delta}_{l}^{*}$ that have positive slope between $t_{j}$ and $t_{j+1}$. Thus, it produces a function estimate that is close to the current function estimate. Furthermore, this coefficient update function ensures all elements of $\tilde{\boldsymbol{b}}$ are positive so it also preserves the monotonically increasing shape restriction. For the second case where $t_{j}$ is a change-point, we have that $t_{j}=\xi_{i}$ for some $i \in\{1, \ldots, h\}$. We let

$$
\tilde{b}_{l}= \begin{cases}b_{l} & \text { for } l=1, \ldots j  \tag{5.15}\\ u\left(b_{l-1}+b_{l}\right) & \text { for } l=j+1 \\ (1-u) b_{l-1} & \text { for } l=j+2 \\ b_{l-1} & \text { for } l=j+3, \ldots, k+3+i-1 \quad . k-1 \\ (1-u) b_{l-1} & \text { for } l=k+3+i \\ b_{l-1} & \text { for } l=k+3+i+1, \ldots, k+3+h\end{cases}
$$

Note that this coefficient update function also only changes coefficients of $\boldsymbol{\delta}_{l}^{*}$ that have positive slope between $t_{j}$ and $t_{j+1}$ and ensures all elements of $\tilde{\boldsymbol{b}}$ are positive so it preserves the monotonically increasing shape restriction. We do not update $\boldsymbol{\alpha}$ and $\tau$ and leave this for the coefficient update step.

Next, we derive the acceptance probability for this birth move from a model with $k$ interior knots to a model with $k+1$ interior knots using (5.13). The acceptance probability for the birth move is equal to $\min \left\{1, A_{b}\right\}$ with

$$
\begin{equation*}
A_{b}=\mathrm{LR}_{b} \times \text { Prior }_{b} \times \operatorname{Prop}_{b} \times \mathrm{Jacob}_{b} \tag{5.16}
\end{equation*}
$$

where $\mathrm{LR}_{b}$ is the likelihood ratio, $\mathrm{Prior}_{b}$ is the prior ratio, $\mathrm{Prop}_{b}$ is the proposal ratio and $\mathrm{Jacob}_{b}$ is the absolute value of the Jacobian, |det (Jacobian)| for the birth move from model with $k$ interior knots to model with $k+1$ interior knots. The likelihood ratio in (5.16) for the birth move is given by

$$
\begin{equation*}
\mathrm{LR}_{b}=\frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k+1, \tilde{\boldsymbol{t}}, \boldsymbol{\xi}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\xi}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)} \tag{5.17}
\end{equation*}
$$

Here, $f_{y}\left(y_{i} \mid k+1, \tilde{\boldsymbol{t}}, \boldsymbol{\xi}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+3+h} \tilde{b}_{j} \tilde{\delta}_{j i}^{*}+$ $\sum_{j=1}^{h+1} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{j i}$ is defined as in (5.7). $\tilde{\delta}_{j i}^{*}$ as defined in Section 5.3 .2 is the $(i)$ th element of $\tilde{\boldsymbol{\delta}}_{j}^{*}$ which is the $(j)$ th basis vector found using $\boldsymbol{x}$, the knot location vector $\tilde{\boldsymbol{t}}$, and $\boldsymbol{\xi}$. In (5.17), $f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\xi}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\eta_{i}=\sum_{j=1}^{k+2+h} b_{j} \delta_{j i}^{*}+\sum_{j=1}^{h+1} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{j i}$ is defined as in (5.7). $\delta_{j i}^{*}$ is the $(i)$ th element of $\boldsymbol{\delta}_{j}^{*}$ found using using $\boldsymbol{x}$, the knot location vector $\boldsymbol{t}$, and $\boldsymbol{\xi}$.

The prior ratio in for the birth move, Prior $_{b}$ in (5.16), is
Prior $_{b}$

$$
\begin{align*}
= & \frac{p(k+1) p(\tilde{\boldsymbol{t}} \mid k+1, h, \boldsymbol{\xi}) p(\tilde{\boldsymbol{b}} \mid k+1, h) p(\boldsymbol{\alpha} \mid h) p(\tau)}{p(k) p(\boldsymbol{t} \mid k, h, \boldsymbol{\xi}) p(\boldsymbol{b} \mid k, h) p(\boldsymbol{\alpha} \mid h) p(\tau)} \\
= & \frac{\lambda^{k+1}}{(k+1)!} \cdot \frac{(k)!}{\lambda^{k}} \cdot \frac{\left[\prod_{i=1}^{h+1}\left(\xi_{i}-\xi_{i-1}\right)^{-1}\right]\left[\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)\right]^{-(k+1-2 h-1)}}{\left[\prod_{i=1}^{h+1}\left(\xi_{i}-\xi_{i-1}\right)^{-1}\right]\left[\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)\right]^{-(k-2 h-1)}} \times \\
& \frac{\prod_{i=1}^{k+3+h} f_{b}\left(\tilde{b}_{i}\right)}{\prod_{i=1}^{k+2+h} f_{b}\left(b_{i}\right)} \\
= & \frac{\lambda}{k+1}\left[\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)\right]^{-1} \frac{\prod_{i=1}^{k+3+h} f_{b}\left(\tilde{b}_{i}\right)}{\prod_{i=1}^{k+2+h} f_{b}\left(b_{i}\right)} \tag{5.18}
\end{align*}
$$

Note that $p(\theta)$ is used to denote the density of the prior for a given parameter $\theta, \lambda$ is the hyperparameter for Poisson prior for the number of interior knots, and $f_{b}(z)$ is used to denote the density for the gamma prior for each $b_{i}$ evaluated at $z$. If we have case (1) where $t_{j}$ is not a change-point, then the last line of (5.18) simplifies to

$$
\begin{equation*}
\frac{c_{2}^{c_{1}}\left(\tilde{b}_{j} \tilde{b}_{j+1} \tilde{b}_{j+2}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(\tilde{b}_{j}+\tilde{b}_{j+1}+\tilde{b}_{j+2}\right)\right\}}{\Gamma\left(c_{1}\right)\left(b_{j} b_{j+1}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(b_{j}+b_{j+1}\right)\right\}} \tag{5.19}
\end{equation*}
$$

If we have case (2) where $t_{j}$ is a change-point and $t_{j}=\xi_{i}$, then the last line of (5.18) simplifies to

$$
\begin{equation*}
\frac{c_{2}^{c_{1}}\left(\tilde{b}_{j+1} \tilde{b}_{j+2} \tilde{b}_{k+3+i}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(\tilde{b}_{j}+\tilde{b}_{j+1}+\tilde{b}_{k+3+i}\right)\right\}}{\Gamma\left(c_{1}\right)\left(b_{j+1} b_{k+2+i}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(b_{j+1}+b_{k+2+i}\right)\right\}} \tag{5.20}
\end{equation*}
$$

The proposal ratio for the birth move is found by first noting that the probability of selecting the interval $\left(t_{j}, t_{j+1}\right)$ from $t_{\text {cand }}$ is one over the cardinality of $t_{\text {cand }}, 1 / n_{t}$, and the probability of $t^{*}$ given we have selected interval $\left(t_{j}, t_{j+1}\right)$ is $1 /(u b-l b)$. Using this, we compute the proposal ratio and find it equal to

$$
\begin{align*}
\operatorname{Prop}_{b} & =\frac{\operatorname{Pr}(\text { death in } k+1 \text { model }) \operatorname{Pr}\left(\text { delete } t^{*}\right)}{\operatorname{Pr}(\text { birth in } k \text { model }) \operatorname{Pr}\left(\text { interval }\left(t_{j}, t_{j+1}\right)\right) \operatorname{Pr}\left(t^{*}\right) \operatorname{Pr}(u)} \\
& =\frac{d_{k+1} \cdot \frac{1}{n_{\text {del }}}}{b_{k} \cdot \frac{1}{n_{t}} \cdot \frac{1}{u b-l b} \cdot \frac{1}{1-0}} \\
& =\frac{d_{k+1} \cdot n_{t} \cdot(u b-l b)}{b_{k} \cdot n_{\text {del }}} \tag{5.21}
\end{align*}
$$

where $\operatorname{Pr}$ (death in $k+1$ model) means the probability of a death move from a model with $k+1$ interior knots to a model with $k$ interior knots and $\operatorname{Pr}$ (birth in $k$ model) means the probability of a birth move from a model with $k$ interior knots to a model with $k+1$ interior knots. In (5.21), $n_{\text {del }}$ is the number of interior knots that are eligible to be deleted in the model with interior knot location $\tilde{\boldsymbol{t}}$ and discussed further in Section 5.4.3.2. The values for $b_{k}$ and $d_{k+1}$ are defined as in Section 5.4.1.

The absolute value of the determinant for the Jacobian for the birth move, $\mathrm{Jacob}_{b}$ in (5.16), is found by letting $\boldsymbol{u}_{1}=\left(t^{*}, u\right),\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}=\left(k, \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau, t^{*}, u\right)^{\prime}$,
and letting $g(\cdot)$ be the function that maps $\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}$ to the proposed $\boldsymbol{\theta}_{2}^{\prime}=(\tilde{k}=k+1, \tilde{\boldsymbol{t}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{\alpha}}, \tilde{\tau})$. The Jacobian is $\partial\left[g\left(\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}\right)\right] / \partial\left(\boldsymbol{\theta}_{1}, \boldsymbol{u}_{1}\right)^{\prime}$. Hence, when $t_{j}$ is not a change-point, we use the coefficient update function proposed in (5.14) and find that the determinant of the Jacobian is equal to the determinant of

$$
\left[\begin{array}{ccc}
1-u & u & 0 \\
0 & u & 1-u \\
-b_{j} & b_{j}+b_{j+1} & -b_{j+1}
\end{array}\right]
$$

So the $|\operatorname{det}(J a c o b i a n)|$ for the birth move for case (1) is

$$
\begin{equation*}
\mathrm{Jacob}_{b}=\left|(u-1)\left(b_{j}+b_{j+1}\right)\right|=\left|\tilde{b}_{j+1}-b_{j}-b_{j+1}\right| \tag{5.22}
\end{equation*}
$$

Likewise, when $t_{j}$ is equal to change-point $\xi_{i}$, we use the coefficient update function proposed in (5.15) and find that the determinant of the Jacobian is equal to the determinant of

$$
\left[\begin{array}{ccc}
u & 0 & 0 \\
u & 1-u & 1-u \\
b_{j+1}+b_{k+2+i} & -b_{j+1} & -b_{k+2+i}
\end{array}\right] .
$$

So the $|\operatorname{det}(J a c o b i a n)|$ for the birth move for case (2) is

$$
\begin{equation*}
\mathrm{Jacob}_{b}=\left|(1-u)\left(b_{j+1}+b_{k+2+i}\right)\right|=\left|b_{j+1}+b_{k+2+i}-\tilde{b}_{j+1}\right| \tag{5.23}
\end{equation*}
$$

### 5.4.3.2 Death Move

The death move deletes an interior knot that is not a change-point. It is a move from a model with $k$ interior knots to a model with $k-1$ interior knots that deletes one of the existing interior knots that is not a change-point and updates the coefficients in a deterministic fashion. From the existing ordered interior knot locations excluding the change-points, we determine a set of candidate knot locations for deletion and denote the set of them by $t_{d e l}$ and denote the cardinality of $t_{\text {del }}$ by $n_{\text {del }}$. The set $t_{\text {del }}$ is created by first defining the set $L_{0}$ which is the set of all interior knots that are not change-points. Let $L_{i}=\left\{t_{j}: t_{j} \in\left(\xi_{i-1}, \xi_{i}\right)\right\}$ for $i=1, \ldots, h+1$. We define $t_{d e l}^{*}$ to be the set of elements in $L_{0}$ excluding any $L_{i}$ that have cardinality of one. To ensure that the death move is the inverse of a birth move, we need at
least four $x_{i}$ 's between $t_{j}$ and $t_{j+2}$ when we delete knot $t_{j+1}$. Therefore, we remove any $t_{j+1}$ from $t_{\text {del }}^{*}$ for which there are less than four $x_{i}$ values between $t_{j}$ and $t_{j+2}$. This gives the set of candidate knots for deletion, $t_{d e l}$, with cardinality $n_{d e l}$. From $t_{d e l}$, a knot location, say $t_{j+1}$, is randomly selected to be removed giving proposed knot locations $\tilde{\boldsymbol{t}}=\left(t_{1}, \ldots, t_{j}, t_{j+2}, \ldots, t_{k+2}\right)^{\prime}=\left(\tilde{t}_{1}, \ldots, \tilde{t}_{j}, \tilde{t}_{j+1}, \ldots, \tilde{t}_{k+1}\right)^{\prime}$.

Given the new interior knot locations, we can update the coefficients deterministically by defining the death move as the inverse of a birth move from $k-1$ interior knots to $k$ interior knots with new knot location $t^{*}=t_{j+1}$. The new $\boldsymbol{b}$ values are denoted by $\tilde{\boldsymbol{b}}=\left(\tilde{b}_{1}, \ldots, \tilde{b}_{k+1+h}\right)$ and are found by inverting the coefficient update function in (5.14) or (5.15) depending on whether $t_{j}$ is a change-point. Thus, for case (1) where $t_{j}$ is not a change-point, we propose a new coefficient vector by letting

$$
\tilde{b}_{l}=\left\{\begin{array}{ll}
b_{l} & l=1, \ldots, j-1  \tag{5.24}\\
\frac{b_{l}\left(b_{l}+b_{l+1}+b_{l+2}\right)}{b_{l}+b_{+2}} & l=j \\
\frac{b_{l+1}\left(b_{l-l}+b_{l}+b_{l+1}\right)}{b_{l-1}+b_{l+1}} & l=j+1 \\
b_{l+1} & l=j+2, \ldots, k+1+h
\end{array} .\right.
$$

For case (2) where $t_{j}$ is a change-point and equal to $\xi_{i}$, we propose a new coefficient vector by letting

$$
\tilde{b}_{l}=\left\{\begin{array}{ll}
b_{l} & l=1, \ldots, j  \tag{5.25}\\
\frac{b_{l+1}\left(b_{l-1}+b_{l}+b_{k+2+i}\right)}{b_{l-1}+b_{k+2+i}} & l=j+1 \\
b_{l+1} & l=j+2, \ldots, k+i \\
\frac{b_{l+1}\left(b_{j+1}+b_{j+2}+b_{l+1}\right)}{b_{j+2}+b_{l+1}} & l=k+1+i \\
b_{l+1} & l=k+2+i, \ldots, k+1+h
\end{array} .\right.
$$

As with the birth move, we do not update $\boldsymbol{\alpha}$ and $\tau$ until the coefficient update step.
The acceptance probability for a death move is found by inverting the acceptance probability for a birth move from $k-1$ to $k$ interior knots with $t^{*}=t_{j+1}$. Thus, the acceptance probability for the death move is equal to $\min \left\{1, A_{d}^{-1}\right\}$ with

$$
A_{d}=\mathrm{LR}_{d} \times \operatorname{Prior}_{d} \times \operatorname{Prop}_{d} \times \mathrm{Jacob}_{d}
$$

where $\mathrm{LR}_{d}$ is the likelihood ratio, $\mathrm{Prior}_{d}$ is the prior ratio, $\mathrm{Prop}_{d}$ is the proposal ratio and $\mathrm{Jacob}_{d}$ is the absolute value of the determinant of the Jacobian, for a
birth move from $k-1$ to $k$ interior knots with $t^{*}=t_{j+1}$. Thus,

$$
\begin{equation*}
\mathrm{LR}_{d}=\frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\xi}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k-1, \tilde{\boldsymbol{t}}, \boldsymbol{\xi}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)} \tag{5.26}
\end{equation*}
$$

where $f_{y}\left(y_{i} \mid k-1, \tilde{\boldsymbol{t}}, \boldsymbol{\xi}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+1+h} \tilde{b}_{j} \tilde{\delta}_{j i}^{*}+$ $\sum_{j=1}^{h+1} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$. Here, $v_{j i}$ is defined as in (5.7). $\tilde{\delta}_{j i}^{*}$ as defined in Section 5.3 .2 is the $(i)$ th element of $\tilde{\boldsymbol{\delta}}_{j}^{*}$ which is the $(j)$ th basis vector found using $\boldsymbol{x}$, the knot location vector $\tilde{\boldsymbol{t}}$, and $\boldsymbol{\xi}$. $f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\xi}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\eta_{i}=\sum_{j=1}^{k+2+h} b_{j} \delta_{j i}^{*}+\sum_{j=1}^{h+1} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{j i}$ is defined as in (5.7). $\delta_{j i}^{*}$ is the $(i)$ th element of $\boldsymbol{\delta}_{j}^{*}$ found using using $\boldsymbol{x}$, the knot location vector $\boldsymbol{t}$, and $\boldsymbol{\xi}$.

For the death move,

$$
\begin{align*}
& \operatorname{Prior}_{d} \\
&= \frac{p(k) p(\boldsymbol{t} \mid k, h, \boldsymbol{\xi}) p(\boldsymbol{b} \mid k, h) p(\boldsymbol{\alpha} \mid h) p(\tau)}{p(k-1) p(\tilde{\boldsymbol{t}} \mid k-1, h, \boldsymbol{\xi}) p(\tilde{\boldsymbol{b}} \mid k-1, h) p(\boldsymbol{\alpha} \mid h) p(\tau)} \\
&= \frac{\lambda^{k}}{(k)!} \cdot \frac{(k-1)!}{\lambda^{k-1}} \cdot \frac{\left[\prod_{i=1}^{h+1}\left(\xi_{i}-\xi_{i-1}\right)^{-1}\right]\left[\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)\right]^{k-2 h-1}}{\left[\prod_{i=1}^{h+1}\left(\xi_{i}-\xi_{i-1}\right)^{-1}\right]\left[\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)\right]^{k-1-2 h-1}} \times \\
& \frac{\prod_{i=1}^{k+2+h} f_{b}\left(b_{i}\right)}{\prod_{i=1}^{k+1+h} f_{b}\left(\tilde{b}_{i}\right)} \\
&= \frac{\lambda}{k}\left[\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)\right]^{-1} \cdot \frac{\prod_{i=1}^{k+2+h} f_{b}\left(b_{i}\right)}{\prod_{i=1}^{k+1+h} f_{b}\left(\tilde{b}_{i}\right)} \tag{5.27}
\end{align*}
$$

If we have case (1) where $t_{j}$ is not a change-point, then
$\left[\prod_{i=1}^{k+2+h} f_{b}\left(b_{i}\right)\right] /\left[\prod_{i=1}^{k+1+h} f_{b}\left(\tilde{b}_{i}\right)\right]$ in the last line of (5.27) simplifies to

$$
\frac{c_{2}^{c_{1}}\left(b_{j} b_{j+1} b_{j+2}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(b_{j}+b_{j+1}+b_{j+2}\right)\right\}}{\Gamma\left(c_{1}\right)\left(\tilde{b}_{j} \tilde{b}_{j+1}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(\tilde{b}_{j}+\tilde{b}_{j+1}\right)\right\}}
$$

If we have case (2) where $t_{j}$ is a change-point and equal to $\xi_{i}$, then $\left[\prod_{i=1}^{k+2+h} f_{b}\left(b_{i}\right)\right] /\left[\prod_{i=1}^{k+1+h} f_{b}\left(\tilde{b}_{i}\right)\right]$ in the last line of (5.27) simplifies to

$$
\frac{c_{2}^{c_{1}}\left(b_{j+1} b_{j+2} b_{k+2+i}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(b_{j+1}+b_{j+2}+b_{k+2+i}\right)\right\}}{\Gamma\left(c_{1}\right)\left(\tilde{b}_{j+1} \tilde{b}_{k+1+i}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(\tilde{b}_{j+1}+\tilde{b}_{k+1+i}\right)\right\}} .
$$

For the death move,

$$
\begin{align*}
\operatorname{Prop}_{d} & =\frac{\operatorname{Pr}(\text { death in } k \text { model }) \operatorname{Pr}\left(\text { delete } t_{j+1}\right)}{\operatorname{Pr}(\text { birth in } k-1 \text { model }) \operatorname{Pr}\left(\text { interval }\left(t_{j}, t_{j+2}\right)\right) \operatorname{Pr}\left(t_{j+1}\right) \operatorname{Pr}(u)} \\
& =\frac{d_{k} \cdot \frac{1}{n_{d e l}}}{b_{k-1} \cdot \frac{1}{n_{t}} \cdot \frac{1}{u b-l b} \cdot \frac{1}{1-0}} \\
& =\frac{d_{k} n_{\tilde{t}}(u b-l b)}{b_{k-1} n_{d e l}} \tag{5.28}
\end{align*}
$$

where $\operatorname{Pr}($ death in $k$ model $)$ denotes the probability of a death move from a model with $k$ interior knots to a model with $k-1$ interior knots and $\operatorname{Pr}$ (birth in $k-1$ model) denotes the probability of a birth move from a model with $k-1$ interior knots to a model with $k$ interior knots. In (5.28), $n_{\tilde{t}}$ is the cardinality of $t_{\text {cand }}$ found using $\tilde{t}$. The values for $u b$ and $l b$ are found using $x$ values in the interval $\left(t_{j}, t_{j+2}\right)$.

For case (1) where $t_{j}$ is not a change-point,

$$
\begin{equation*}
\operatorname{Jacob}_{d}=\left|(u-1)\left(\tilde{b}_{j}+\tilde{b}_{j+1}\right)\right|=\left|b_{j+1}-\tilde{b}_{j}-\tilde{b}_{j+1}\right| . \tag{5.29}
\end{equation*}
$$

Likewise, for case (2) where $t_{j}$ is equal to change-point $\xi_{i}$,

$$
\begin{equation*}
\operatorname{Jacob}_{d}=\left|(1-u)\left(\tilde{b}_{j+1}+\tilde{b}_{k+1+i}\right)\right|=\left|\tilde{b}_{j+1}+\tilde{b}_{k+1+i}-b_{j+1}\right| . \tag{5.30}
\end{equation*}
$$

### 5.4.3.3 Relocation Move

A relocation move changes the location of an interior knot that is not a changepoint. It keeps the same number of interior knots, $k$, but the location of one knot is moved and the coefficients are updated deterministically based on this location change. We create a candidate set of knots to move that excludes change-points and also ensures that we do not place two interior knot points too close together. We define sets $\mathrm{W}_{j+1}=\left\{x_{i}: x_{i} \in\left(t_{j}, t_{j+2}\right)\right\}$ for $j+1=2, \ldots, k+1$ which are the sets of data points between the two knot points surrounding $t_{j+1}$. The set of candidate knots, $t_{\text {move }}$, is the set of all knots, $t_{j+1}$, for $j=1, \ldots, k$ such that $t_{j+1} \neq \xi_{i}$ for all $i \in\{1, \ldots, h\}$ and $\mathrm{W}_{j+1}$ has cardinality of at least four. From $t_{\text {move }}$, we randomly select an interior knot location and denote it by $t_{j+1}$. Let $n_{m}$
be the cardinality of $t_{\text {move }}$ so $t_{j+1}$ is selected with probability $1 / n_{m}$. We propose a new interior knot location, $t^{*}$, from Uniform $(l b, u b)$ where $l b$ is the second smallest $x_{i}$ in $\mathrm{W}_{j+1}$ unless $j=1$ then $l b=0$ and $u b$ is the second largest $x_{i}$ in $\mathrm{W}_{j+1}$ unless $j=k$ then $u b=1$. The proposed knots for the relocation move are given by $\tilde{\boldsymbol{t}}=\left(\tilde{t}_{1}, \ldots, \tilde{t}_{k+2}\right)^{\prime}=\left(t_{1}, \ldots, t_{j}, t^{*}, t_{j+2}, \ldots, t_{k+2}\right)^{\prime}$.

Given new knot locations, $\tilde{\boldsymbol{t}}$, we update the restricted regression coefficients. We update coefficients for this relocation move similar to what is used for the relocation move in Section 4.3.3 but we update the $\boldsymbol{b}$ values for the change-point model instead of the $\boldsymbol{\beta}$ values. The distance between the $t_{j}$ and $t^{*}$, the distance between $t^{*}$ and $t_{j+1}$ and the distance between $t^{*}$ and $t_{j+2}$ which we denote by $r_{1}, r_{2}$, and $r_{3}$, respectively, will affect the magnitude of the change in the $\boldsymbol{b}$ values. Therefore, we use these to update the restricted regression coefficients. For case (1) where $t_{j}$ is not a changepoint and $t^{*} \geq t_{j+1}$, the coefficients are updated according to

$$
\tilde{b}_{l}= \begin{cases}b_{l} & l=1, \ldots, j-1  \tag{5.31}\\ b_{l}+\frac{0.5 r_{2}}{r_{1}+r_{2}+r_{3}} b_{l+1} & l=j \\ \frac{r_{1}+0.5 r_{3}}{r_{1}+r_{2}+r_{3}} b_{l}+\frac{r_{2}}{r_{2}+r_{3}} b_{l+1} & l=j+1 \\ \frac{r_{3}}{r_{2}+r_{3}} b_{l}+\frac{0.5\left(r_{2}+r_{3}\right)}{r_{1}+r_{2}+r_{3}} b_{l-1} & l=j+2 \\ b_{l} & l=j+3, \ldots, k+2\end{cases}
$$

where $r_{1}=t_{j+1}-t_{j}, r_{2}=t^{*}-t_{j+1}$, and $r_{3}=t_{j+2}-t^{*}$. For case (1) where $t_{j}$ is not a change-point and $t^{*}<t_{j+1}$, the coefficients are updated according to

$$
\tilde{b}_{l}= \begin{cases}b_{l} & l=1, \ldots, j-1  \tag{5.32}\\ \frac{r_{1}}{r_{1}+r_{2}} b_{l}+\frac{0.5\left(r_{1}+r_{2}\right)}{r_{1}+r_{2}+r_{3}} b_{l+1} & l=j \\ \frac{0.5 r_{1}+r_{3}}{r_{1}+r_{2}+r_{3}} b_{l}+\frac{r_{2}}{r_{1}+r_{2}} b_{l-1} & l=j+1 \\ b_{l}+\frac{0.5 r_{2}}{r_{1}+r_{2}+r_{3}} b_{l-1} & l=j+2 \\ b_{l} & l=j+3, \ldots, k+2\end{cases}
$$

where $r_{1}=t^{*}-t_{j}, r_{2}=t_{j+1}-t^{*}$, and $r_{3}=t_{j+2}-t_{j+1}$. For case (2) where $t_{j}$ is a change-point and $t^{*} \geq t_{j+1}$, the coefficients are updated according to

$$
\tilde{b}_{l}= \begin{cases}b_{l} & l=1, \ldots, j  \tag{5.33}\\ \frac{0.5 r_{3}+r_{1}}{r_{1}+r_{1}+r_{3}} b_{l}+\frac{r_{2}}{r_{2}+r_{3}} b_{l+1} & l=j+1 \\ \frac{0.5\left(r_{2}+r_{3}\right)}{r_{1}+r_{2}+r_{3}} b_{l-1}+\frac{r_{3}}{r_{2}+r_{3}} b_{l} & l=j+2 \\ b_{l} & l=j+3, \ldots, k+1+i \\ \frac{0.5 r_{2}}{r_{1}+r_{2}+r_{3}} b_{j+1}+b_{l} & l=k+2+i \\ b_{i} & l=k+3+i, \ldots, k+2+h\end{cases}
$$

where $r_{1}=t_{j+1}-t_{j}, r_{2}=t^{*}-t_{j+1}$, and $r_{3}=t_{j+2}-t^{*}$. For case (2) where $t_{j}$ is a change-point and $t^{*}<t_{j+1}$, the coefficients are updated according to

$$
\tilde{b}_{l}= \begin{cases}b_{l} & l=1, \ldots, j  \tag{5.34}\\ \frac{0.5 r_{1}+r_{3}}{r_{1}+r_{2}+r_{3}} b_{l}+\frac{r_{2}}{r_{1}+r_{2}} b_{k+2+i} & l=j+1 \\ b_{l}+\frac{0.5 r_{2}}{r_{1}+r_{2}+r_{3}} b_{l-1} & l=j+2 \\ b_{l} & l=j+3, \ldots, k+1+i \\ \frac{r_{1}}{r_{1}+r_{2}} b_{l}+\frac{0.5\left(r_{1}+r_{2}\right)}{r_{1}+r_{2}+r_{3}} b_{j+1} & l=k+2+i \\ b_{i} & l=k+3+i, \ldots, k+2+h\end{cases}
$$

where $r_{1}=t^{*}-t_{j}, r_{2}=t_{j+1}-t^{*}$, and $r_{3}=t_{j+2}-t_{j+1}$. As with the birth and death moves, we do not update $\boldsymbol{\alpha}$ and $\tau$ until the coefficient update step.

The acceptance probability for the relocation move is $\min \left\{1, A_{r}\right\}$ where

$$
A_{r}=\mathrm{LR}_{r} \times \text { Prior }_{r} \times \text { Prop }_{r} .
$$

Note that we do not have |det (Jacobian)| term in the acceptance probability for the relocation move because the relocation move does not involve a change in dimensions. The likelihood ratio for the relocation move is given by

$$
\begin{equation*}
\mathrm{LR}_{r}=\frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \tilde{\boldsymbol{t}}, \boldsymbol{\xi}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\xi}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)} \tag{5.35}
\end{equation*}
$$

where $f_{y}\left(y_{i} \mid k, \tilde{\boldsymbol{t}}, \boldsymbol{\xi}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+2+h} \tilde{b}_{j} \tilde{\delta}_{j i}^{*}+$ $\sum_{j=1}^{h+1} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{j i}$ is defined as in (5.7). $\tilde{\boldsymbol{\delta}}_{j}^{*}$ is the $(j)$ th basis vector with elements $\tilde{\delta}_{j i}^{*}$ as defined in Section 5.3.2 and is found using knot location vector $\tilde{\boldsymbol{t}}$ and $\boldsymbol{\xi} . f_{y}\left(y_{i} \mid k, \boldsymbol{t}, \boldsymbol{\xi}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean
$\eta_{i}=\sum_{j=1}^{k+2+h} b_{j} \delta_{j i}^{*}+\sum_{j=1}^{h+1} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{j i}$ is defined as in (5.7). $\boldsymbol{\delta}_{j}^{*}$ is the $(j)$ th basis vector with elements $\delta_{j i}^{*}$ as defined in Section 5.3.2 and is found using knot location vector $\boldsymbol{t}$ and $\boldsymbol{\xi}$.

The prior ratio for the relocation move is
Prior $_{r}$

$$
\begin{align*}
& =\frac{p(k) p(\tilde{\boldsymbol{t}} \mid k, h, \boldsymbol{\xi}) p(\tilde{\boldsymbol{b}} \mid k, h) p(\boldsymbol{\alpha} \mid h) p(\tau)}{p(k) p(\boldsymbol{t} \mid k, h, \boldsymbol{\xi}) p(\boldsymbol{b} \mid k, h) p(\boldsymbol{\alpha} \mid h) p(\tau)} \\
& =\frac{\left[\prod_{i=1}^{h+1}\left(\xi_{i}-\xi_{i-1}\right)^{-1}\right]\left[\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)\right]^{k-2 h-1}\left[\prod_{i=1}^{k+2+h} f_{b}\left(\tilde{b}_{i}\right)\right]}{\left[\prod_{i=1}^{h+1}\left(\xi_{i}-\xi_{i-1}\right)^{-1}\right]\left[\sum_{i=1}^{h+1}\left(d_{l i}-d_{u, i-1}\right)\right]^{k-2 h-1}\left[\prod_{i=1}^{k+2+h} f_{b}\left(b_{i}\right)\right]} \\
& =\frac{\prod_{i=1}^{k+2+h} f_{b}\left(\tilde{b}_{i}\right)}{\prod_{i=1}^{k+2+h} f_{b}\left(b_{i}\right)} . \tag{5.36}
\end{align*}
$$

We use the same notation in (5.36) as we did for the prior ratio for the birth move in (5.18). If we have case (1) where $t_{j}$ is not a change-point, then the last line of (5.36) simplifies to

$$
\begin{equation*}
\frac{\left(\tilde{b}_{j} \tilde{b}_{j+1} \tilde{b}_{j+2}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(\tilde{b}_{j}+\tilde{b}_{j+1}+\tilde{b}_{j+2}\right)\right\}}{\left(b_{j} b_{j+1} b_{j+2}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(b_{j}+b_{j+1}+b_{j+2}\right)\right\}} \tag{5.37}
\end{equation*}
$$

If we have case (2) where $t_{j}$ is a change-point and equal to $\xi_{i}$, then the last line of (5.36) simplifies to

$$
\begin{equation*}
\frac{\left(\tilde{b}_{j+1} \tilde{b}_{j+2} \tilde{b}_{k+2+i}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(\tilde{b}_{j+1}+\tilde{b}_{j+2}+\tilde{b}_{k+2+i}\right)\right\}}{\left(b_{j+1} b_{j+2} b_{k+2+i}\right)^{c_{1}-1} \exp \left\{-c_{2}\left(b_{j+1}+b_{j+2}+b_{k+2+i}\right)\right\}} \tag{5.38}
\end{equation*}
$$

The proposal ratio for the relocation move is given by

$$
\begin{align*}
\operatorname{Prop}_{r} & =\frac{\operatorname{Pr}(\text { relocate in } k \text { model }) \operatorname{Pr}\left(\text { move } t^{*}\right) \operatorname{Pr}\left(t_{j+1}\right)}{\operatorname{Pr}(\text { relocate in } k \text { model }) \operatorname{Pr}\left(\text { move } t_{j+1}\right) \operatorname{Pr}\left(t^{*}\right)} \\
& =\frac{r_{k} \cdot \frac{1}{n_{m}} \cdot \frac{1}{u b-l b}}{r_{k} \cdot \frac{1}{n_{m}} \cdot \frac{1}{u b-l b}} \\
& =1 \tag{5.39}
\end{align*}
$$

where $\operatorname{Pr}$ (relocate in $k$ model) denotes the probability of selecting a relocation move when current model has $k$ interior knots. The values for $u b$ and $l b$ are found using $x$ values in the interval $\left(t_{j}, t_{j+2}\right)$.

### 5.4.4 Function Estimation for the RJMCMC 1 Algorithm

To estimate the regression function, we first compute the estimate of $\boldsymbol{\eta}$ as in (5.6) at iteration $i$ of the RJMCMC algorithm and denote it by $\boldsymbol{\eta}^{(i)}$. We use the values of $k, \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{\alpha}$, and $\tau$ from the $(i)$ th iteration of RJMCMC after the coefficient update step and denote them by $k^{(i)}, \boldsymbol{t}^{(i)}, \boldsymbol{b}^{(i)}$, and $\boldsymbol{\alpha}^{(i)}$, respectively. The estimate of the mean regression function at iteration $i$ is then given by $\boldsymbol{\eta}^{(i)}=$ $\sum_{j=1}^{k+2+h} b_{j}^{(i)} \boldsymbol{\delta}_{j}^{*}+\sum_{j=1}^{h+1} \alpha_{j}^{(i)} \boldsymbol{v}_{j}$ where $\boldsymbol{\delta}_{j}^{*}$ and $\boldsymbol{v}_{j}$ are the basis vectors created using the vector of current knot locations, $\boldsymbol{t}$, and $\boldsymbol{\xi}$. The regression spline estimate is then found by averaging $\boldsymbol{\eta}^{(i)}$ after discarding burn-in iterations and is given by

$$
\begin{equation*}
\hat{\boldsymbol{\eta}}=\frac{1}{N-B} \sum_{i=B+1}^{N} \boldsymbol{\eta}^{(i)} \tag{5.40}
\end{equation*}
$$

where $N$ is the total iterations in the RJMCMC algorithm and $B$ is the burn-in.

### 5.4.5 Examples for the RJMCMC 1 Algorithm

In this section, we examine the performance of RJMCMC 1 algorithm discussed in Sections 5.4.1, 5.4.2, and 5.4.3 by considering its application to two simulated data sets. The first simulated data set is simulated from a model with one change-point and is denoted the "single change-point" data set. The second simulated data set is simulated from a model with two change-points and is denoted the "two changepoints" data set. For both data sets, we ran the algorithm for 50,000 iterations and threw out the first 10,000 iterations as burn-in. For the single change-point data set, we let $\lambda=5, k_{\min }=3, k_{\max }=25, c_{1}=0.2, c_{2}=0.1, d_{1}=0.2$, and $d_{2}=0.2$. For the two change-points data set, we kept all other hyperparameters the same except we let $\lambda=9, k_{\text {min }}=5, k_{\max }=37$ because more interior knots are required with the additional change-point.

The single change-point data set is generated using $f(x)$ plus random normal errors with a standard deviation of 0.35 and $n=100$ equally spaced $x$-values between 0 and 1 . The regression function for $x$-values less than or equal to the change-point
of 0.5 is $f(x)=3 \exp (10 x-5) /(1+\exp (10 x-5))$ and the regression function for $x$-values greater than change-point is $f(x)=4 \exp (10(x-0.5)-5) /(1+\exp (10(x-$ $0.5)-5))+0.3$. The data are then scaled such that the mean of the $y$ values is 0 and their variance is 1 . The scaled data are shown in Figure 5.1 along with the function estimate found using the RJMCMC 1 algorithm (red dashed line). The true regression function (solid gray line) and the pointwise $95 \%$ credible intervals (shaded region) found using method given in Section 2.6.1 are also shown in Figure 5.1. Note that estimated regression function is close to the true function and the true function is contained within the pointwise credible intervals.


Figure 5.1: Simulated data from a regression function with a single change-point and the estimated regression function (red dashed line) using the RJMCMC 1 algorithm along with pointwise $95 \%$ HPD intervals (shaded region) and the function used to simulate the data (solid gray line). The change-point is indicated by the vertical line.

The partial trace plots for the number of interior knots, $k$, and the precision of the normal errors discussed in Section 5.3.3, $\tau$, are shown in Figures 5.2(a) and (c), respectively. We selected to look at these parameters because they do not change meaning throughout the RJMCMC 1 algorithm. They suggest good mixing for the RJMCMC algorithm. We also look at the MCSE defined in Section 4.4.1 and note that if the MCSE for parameter $\theta$ is less than $5 \%$ of the sample standard deviation of the MCMC realizations (minus burn-in) of $\theta$, the chain is thought to have run
long enough. The MCSE for $k$ equal to 0.0204 which is less than $5 \%$ of the sample standard deviation for $k$ (0.0433) and the MCSE for $\tau$ equal to 0.0044 which is less than $5 \%$ of the sample standard deviation for $\tau$ (0.0264). The histograms of posterior draws for $k$ and $\tau$ are given in Figures 5.2(b) and (d), respectively. We estimate the number of interior knots by averaging the posterior draws for $k$ after discarding burn-in iterations and find $\hat{k}=3.88$ which is just above $k_{\text {min }}$. Given that the function that simulated the data is not very "wiggly," the smaller number of interior knots seems appropriate. We estimate $\tau$ by averaging the posterior draws for $\tau$ after discarding burn-in iterations and find $\hat{\tau}=3.55$. The estimate for $\tau$ is larger than the value of $\tau$ used to generate the data, 3.03 (for the scaled data), but 3.03 is included in the $95 \%$ HPD interval for $\tau,(2.53,4.60)$.


Figure 5.2: (a) Partial trace plot for $k$ for the single change-point data set. (b) Histogram of posterior draws for $k$ for the RJMCMC 1 algorithm for the single change-point data set after discarding burn-in. (c) Partial trace plot for $\tau$ for the single change-point data set. (d) Histogram of posterior draws for $\tau$ for the RJMCMC 1 algorithm for the single change-point data set after discarding burn-in. The grey solid line denotes the value of $\tau$ used to generate the data and the red dashed lines mark the $95 \%$ HPD interval for $\tau$ found using method discussed in Section 2.6.1.

The two change-points data set is generated using $f(x)$ plus random normal errors with a standard deviation of 0.05 and $n=150$ equally spaced $x$-values between

0 and 1. The change-points are at 0.3 and 0.7 . The regression function for $x$ values less than or equal to 0.3 is $f(x)=5 \exp (10 x-5) /(1+\exp (10 x-5))$. The regression function for $x$-values greater than 0.3 and less than or equal to 0.7 is $f(x)=7 \exp (5(x-0.3)-5) /(1+\exp (5(x-0.3)-5))+0.1$. The regression function for $x$-values greater than 0.7 is $f(x)=3 \exp (10(x-0.7)-5) /(1+\exp (10(x-$ $0.7)-5))+.22$. The data are then scaled such that the mean of the $y$ values is 0 and their variance is 1 . The scaled data are shown in Figure 5.3 along with the function estimate found using the RJMCMC 1 algorithm (red dashed line). The true regression function (solid gray line) and the pointwise $95 \%$ credible intervals (shaded region) found using the method given in Section 2.6.1 are also shown in Figure 5.3. Note that the estimate found using RJMCMC is very close to the true function. The estimate is able to capture the sharper increase in the function before the first change-point as well as the slower increase in the function between the change-points. Note that the true regression function is contained with pointwise credible intervals except for a few $x$-values that are close to one.


Figure 5.3: Simulated data from a regression function with two change-points and the estimated regression function (red dashed line) using the RJMCMC 1 algorithm along with pointwise $95 \%$ HPD intervals (shaded region) and the function used to simulate the data (solid gray line). The change-points are indicated by the vertical lines.

The partial trace plots for $k$ and $\tau$ for the two change-points data set are shown in Figures 5.4(a) and (c), respectively. They suggest good mixing for the RJMCMC algorithm. We again look at the MCSE defined in Section 4.4.1. The MCSE for $k$ equal to 0.0276 which is less than $5 \%$ of the sample standard deviation for $k(0.0487)$ and the MCSE for $\tau$ equal to 0.0057 which is less than $5 \%$ of the sample standard deviation for $\tau$ (0.0407). The histograms of posterior draws for $k$ and $\tau$ are given in Figures 5.4(b) and (d), respectively. We estimate the number of interior knots by averaging the posterior draws for $k$ after discarding burn-in iterations and find $\hat{k}=6.05$. This supports that more interior knot points are needed when we increase the number of change-points. We estimate $\tau$ by averaging the posterior draws for $\tau$ after discarding burn-in iterations and find $\hat{\tau}=6.80$. The estimate for $\tau$ is larger than the value of $\tau$ that generated the data, 6.58 (for the scaled data), but 6.58 is included in the $95 \%$ HPD interval for $\tau,(5.26,8.42)$.

### 5.5 Single Change-point Model with Change-point Location Unknown

We now discuss an RJMCMC algorithm to find the location of an unknown change-point in a model with a single change-point and denote this algorithm by the RJMCMC 2 algorithm. We use the Bayesian model discussed in Section 5.3.3 and extend the RJMCMC 1 algorithm discussed in Section 5.4. We assume the normal errors model in Section 5.3.3 with the regression function monotonically increasing between the change-point and the endpoints.

### 5.5.1 Overview of the RJMCMC 2 Algorithm

The algorithm can be summarized as follows:

1. Perform a "birth", "death", "relocation", or "change-point relocation" move and accept the proposed parameters with a given acceptance probability.
2. Perform a "coefficient update" step using a Gibbs sampler step to update $\boldsymbol{\alpha}$, $\boldsymbol{b}$, and $\tau$.


Figure 5.4: (a) Partial trace plot for $k$ for the two change-points data set. (b) Histogram of the posterior draws for $k$ for the RJMCMC 1 algorithm for the two change-points data set after discarding burn-in. (c) Partial trace plot for $\tau$ for the two change-points data set. (d) Histogram of the posterior draws for $\tau$ for the RJMCMC 1 algorithm for the two change-points data set after discarding burn-in. The grey solid line denotes the value of $\tau$ used to generate the data and the red dashed lines mark the $95 \%$ HPD interval for $\tau$ found using method discussed in Section 2.6.1.

## 3. Repeat until convergence.

The probabilities of performing birth, death, relocation, and change-point relocation moves, denoted by $b_{k}, d_{k}, r_{k}$, and $r c_{k}$, respectively, are given in Table 5.1 where $b_{k}$, $d_{k}$, and $r_{k}$ are defined as in Section 5.4.1 and in Section 4.2.1 but with $c=0.2$ and different values for $b_{k}, d_{k}$, and $r_{k}$ when $k=k_{\min }$ and $k=k_{\max }$. The new values for $b_{k}, d_{k}$, and $r_{k}$ when $k=k_{\min }$ and $k=k_{\max }$ are also given in Table 5.1. The four moves in the RJMCMC 2 algorithm and their acceptance probabilities are discussed in Section 5.5.3. The coefficient update step updates parameters $\boldsymbol{\alpha}$, $\boldsymbol{b}$, and $\tau$ using a Gibbs sampler step and the conditional distributions for the model parameters given in Section 5.3.3.

Table 5.1: Probabilities for four moves in the proposed RJMCMC 2 algorithm to determine the location of a change-point.

| Move | $k=k_{\min }$ | $k_{\min }<k<k_{\max }$ | $k=k_{\max }$ |
| :---: | :---: | :---: | :---: |
| birth | $b_{k}=\frac{1}{2}$ | $b_{k}$ | $b_{k}=0$ |
| death | $d_{k}=0$ | $d_{k}$ | $d_{k}=\frac{1}{2}$ |
| relocation | $r_{k}=\frac{1}{2}$ | $\frac{1}{3} r_{k}$ | $r_{k}=\frac{1}{4}$ |
| change-point relocation | $r c_{k}=0$ | $\frac{2}{3} r_{k}$ | $r c_{k}=\frac{1}{4}$ |

### 5.5.2 Priors for the RJMCMC 2 Algorithm

In addition to introducing new move probabilities, we also assume a new set of priors for the RJMCMC 2 algorithm. In particular, we use the same prior for $k$ as for the RJMCMC 1 algorithm in Section 5.4.2 with $k_{\text {min }}=3$. Given $k$, we use the same prior for the interior knot locations as we did for the RJMCMC algorithm without change-points given in Section 4.2.2 and assume the knot locations are distributed as the even-numbered order statistics from $2 k+1$ locations uniformly distributed on $[\min (\boldsymbol{x}), \max (\boldsymbol{x})]$. Since the change-point location is no longer fixed, the changepoint location is an unknown parameter which we denote by $\xi_{1}$. We require the change-point to occur at one of the interior knots points. We also require one interior knot between $t_{1}$ and $\xi_{1}$ and one interior knot between $\xi_{1}$ and $t_{k+2}$. Given these requirements and knot locations $\boldsymbol{t}=\left(\min (\boldsymbol{x})=t_{1}, t_{2}, \ldots, t_{k+1}, t_{k+2}=\max (\boldsymbol{x})\right)^{\prime}$, we create a set of candidate points for $\xi_{1}$ given by $C P_{\text {cand }}=\left\{t_{3}, \ldots, t_{k}\right\}$. The prior for $\xi_{1}$ is then a discrete uniform distribution over $C P_{\text {cand }}$ so $\operatorname{Pr}\left(\xi_{1} \mid \boldsymbol{t}\right)=1 /(k-2)$. As with the RJMCMC 1 algorithm, for the unconstrained coefficients, $\alpha_{j}$, we assume independent normal priors with mean of zero and variance of $M$. We again assume a gamma prior for $\tau$ with shape $d_{1}$ and rate $d_{2}$. For the constrained coefficients, $b_{j}$ for $j=1, \ldots, k+2+h$, we again assume independent gamma priors with shape $c_{1}$ and rate $c_{2}$.

### 5.5.3 RJMCMC 2 Algorithm Implementation

In this section, we discuss each move type and their corresponding acceptance probabilities for the RJMCMC 2 algorithm to estimate regression function with a
single change-point whose location is unknown under the monotone shape restriction.

### 5.5.3.1 Birth, Death, and Relocation Moves for the RJMCMC 2 Algorithm

The birth, death, and relocation moves in the RJMCMC 2 algorithm propose knot locations for interior knots that are not the change-point. Given current value for $\xi_{1}$ and $h=1$, the birth, death, and relocation moves in the RJMCMC 2 algorithm propose knot locations and update coefficients in the same was as the birth, death, and relocation moves in the RJMCMC 1 algorithm discussed in Sections 5.4.3.1, 5.4.3.2, and 5.4.3.3, respectively. However, we have a different set of priors for the RJMCMC 2 algorithm so the acceptance probabilities for these three moves are different for this algorithm. For the birth move for the RJMCMC 2 algorithm, the acceptance probability is given by $\min \left\{1, A_{b}^{*}\right\}$ where

$$
A_{b}^{*}=\mathrm{LR}_{b}^{*} \times \operatorname{Prior}_{b}^{*} \times \operatorname{Prop}_{b}^{*} \times \mathrm{Jacob}_{b}^{*}
$$

The likelihood ratio for the birth move for the RJMCMC 2 algorithm is same as the likelihood ratio for the birth move in the RJMCMC 1 algorithm. Thus, $\mathrm{LR}_{b}^{*}=\mathrm{LR}_{b}$ with $\mathrm{LR}_{b}$ as in (5.17). The prior ratio for the birth move for the RJMCMC 2 algorithm is

$$
\operatorname{Prior}_{b}^{*}=\frac{\lambda(2 k+3)(2 k+2)\left(t^{*}-t_{j}\right)\left(t_{j+1}-t^{*}\right)(k-2) c_{2}^{c 1}\left(\prod_{j=1}^{k+4} f_{b}\left(\tilde{b}_{j}\right)\right)}{(k+1)\left(t_{j+1}-t_{j}\right)(k-1) \Gamma\left(c_{1}\right)\left(\prod_{j=1}^{k+3} f_{b}\left(b_{j}\right)\right)}
$$

where $f_{b}, t_{j}, t_{j+1}, t^{*}, \tilde{b}_{j}$, and $b_{j}$ are defined as in Section 5.4.3.1. If $t_{j} \neq \xi_{1}$, the ratio of $\prod_{j=1}^{k+4} f_{b}\left(\tilde{b}_{j}\right)$ over $\prod_{j=1}^{k+3} f_{b}\left(b_{j}\right)$ reduces to ratio given in (5.19). If $t_{j}=\xi_{1}$, the ratio of $\prod_{j=1}^{k+4} f_{b}\left(\tilde{b}_{j}\right)$ over $\prod_{j=1}^{k+3} f_{b}\left(b_{j}\right)$ reduces to ratio given in (5.20). The proposal ratio for the birth move for the RJMCMC 2 algorithm is the same as the proposal ratio for the birth move in the RJMCMC 1 algorithm so $\operatorname{Prop}_{b}^{*}=\operatorname{Prop}_{b}$ with $\mathrm{Prob}_{b}$ as in (5.21). The absolute value of the determinant of the Jacobian for
the birth move for the RJMCMC 2 algorithm is the same as it is for the birth move in the RJMCMC 1 algorithm. Thus, if $t_{j} \neq \xi_{1}$, $\operatorname{Jacob}_{b}^{*}=\left|\tilde{b}_{j+1}-b_{j}-b_{j+1}\right|$ and if $t_{j}=\xi_{1}, \mathrm{Jacob}_{b}^{*}=\left|b_{j+1}+b_{k+2+i}-\tilde{b}_{j+1}\right|$.

The acceptance probability for the death move is equal to $\min \left\{1,\left(A_{d}^{*}\right)^{-1}\right\}$ where

$$
A_{d}^{*}=\mathrm{LR}_{d}^{*} \times \operatorname{Prior}_{d}^{*} \times \operatorname{Prop}_{d}^{*} \times \mathrm{Jacob}_{d}^{*}
$$

As with the birth move, the likelihood ratio, the proposal ratio, and the absolute value of the determinant of the Jacobian for the death move in the RJMCMC 2 algorithm is the same as they are for the RJMCMC 1 algorithm. Therefore, $\mathrm{LR}_{d}^{*}=$ $\mathrm{LR}_{d}$ with $\mathrm{LR}_{d}$ as in (5.26), $\operatorname{Prop}_{d}^{*}=\operatorname{Prop}_{d}$ with $\mathrm{Prop}_{d}$ as in (5.28), and Jacob ${ }_{d}^{*}=$ $\mathrm{Jacob}_{d}$ with $\mathrm{Jacob}_{d}$ as in (5.29) or (5.30) depending on whether $t_{j}=\xi_{1}$. The prior ratio for the death move in the RJMCMC 2 algorithm is

$$
\operatorname{Prior}_{d}^{*}=\frac{\lambda(2 k+1)(2 k)\left(t_{j-1}-t_{j}\right)\left(t_{j+2}-t_{j+1}\right)(k-3) \prod_{j=1}^{k+3} f_{b}\left(b_{j}\right)}{k\left(t_{j+2}-t_{j}\right)(k-2) \prod_{j=1}^{k+2} f_{b}\left(\tilde{b}_{j}\right)}
$$

where $f_{b}, t_{j}, t_{j+1}, t^{*}, \tilde{b}_{j}$, and $b_{j}$ are defined as in Section 5.4.3.2.
The acceptance probability for the relocation move in the RJMCMC 2 algorithm is equal to $\min \left\{1, A_{r}^{*}\right\}$ where

$$
A_{r}^{*}=\mathrm{LR}_{r}^{*} \times \text { Prior }_{r}^{*} \times \text { Prop }_{r}^{*}
$$

The likelihood ratio and the proposal ratio for relocation move for the RJMCMC 2 algorithm are the same as for the RJMCMC 1 algorithm. Thus, $\mathrm{LR}_{r}^{*}=\mathrm{LR}_{r}$ with $\mathrm{LR}_{r}$ as in (5.35) and $\operatorname{Prop}_{r}^{*}=\operatorname{Prop}_{r}$ with $\operatorname{Prop}_{r}$ as in (5.39). The prior ratio for relocation move for the RJMCMC 2 algorithm is

$$
\operatorname{Prior}_{r}^{*}=\frac{\left(t^{*}-t_{j}\right)\left(t_{j+2}-t^{*}\right) \prod_{i=1}^{k+3} f_{b}\left(\tilde{b}_{i}\right)}{\left(t_{j+1}-t_{j}\right)\left(t_{j+2}-t_{j+1}\right) \prod_{i=1}^{k+3} f_{b}\left(b_{i}\right)}
$$

where $f_{b}, t_{j}, t_{j+1}, t_{j+2}, t^{*}, \tilde{b}_{j}$, and $b_{j}$ are defined as in Section 5.4.3.3.

### 5.5.3.2 Change-point Relocation

The change-point relocation move proposes a new location for the change-point. It keeps the same number of interior knots but moves the change-point to a new interior knot location. We propose a new change-point denoted $\tilde{\xi}_{1}=t_{j_{1}^{*}}$ such that $j_{1}^{*}=j_{1}-1$ or $j_{1}^{*}=j_{1}+1$ where $\xi_{1}=t_{j_{1}}$ is the current change-point at the $\left(j_{1}\right)$ th largest interior knot location. We let $j_{1}^{*}=j_{1}-1$ and $j_{1}^{*}=j_{1}+1$ with equal probability except when $j_{1}=3$ or when $j_{1}=k$. We let the probability that $j_{1}^{*}=j_{1}+1=4$ be one when $j_{1}=3$ to ensure that there is at least one interior knot between $t_{1}$ and the change-point. Likewise, we let the probability that $j_{1}^{*}=j_{1}-1=k-1$ be one when $j_{1}=k$ to ensure that there is at least one interior knot between the change-point and $t_{k+2}$.

Given $\tilde{\xi}_{1}$, we consider proposing new values for $\boldsymbol{\beta}=S^{-1} \boldsymbol{b}$ instead of $\boldsymbol{b}$. We denote the proposed values for $\boldsymbol{\beta}$ after moving $\xi_{1}$ to $\tilde{\xi}_{1}$ by $\tilde{\boldsymbol{\beta}}=\left(\tilde{\beta}_{1}, \ldots, \tilde{\beta}_{k+3}\right)^{\prime}$. Using $\tilde{\boldsymbol{\beta}}$, we obtain new values for $\boldsymbol{b}$ which are denoted $\tilde{\boldsymbol{b}}=\left(\tilde{b}_{1}, \ldots, \tilde{b}_{m+1}\right)$ and $\tilde{\boldsymbol{b}}=\tilde{S}^{-1} \tilde{\boldsymbol{\beta}}$ where $\tilde{S}$ is the constraint matrix created using new change-point location, $\tilde{\xi}_{1}=t_{j_{1}^{*}}$ and basis vectors created using knot locations $\boldsymbol{t}$. To obtain $\tilde{\boldsymbol{\beta}}$, we let $u=0.5$ and

$$
\tilde{\beta}_{i}= \begin{cases}\beta_{i} & i=1, \ldots, j_{1}-2 \\ u \beta_{i-1}+u \beta_{i} & i=j_{1-1} \\ (1-u) \beta_{i-1}+(1-u)\left(\beta_{i}+\beta_{m+1}\right)-u \beta_{m+1} & i=j_{1} \\ \beta_{i}+(1-u) \beta_{m+1} & i=j_{1}+1, \ldots, m \\ u \beta_{i} & i=m+1\end{cases}
$$

when $j_{1}^{*}=j_{1}-1$. If $j_{1}^{*}=j_{1}+1$, we let

$$
\tilde{\beta}_{i}= \begin{cases}\beta_{i} & i=1, \ldots, j_{1}-1 \\ (1-u) \beta_{i}+(1-u)\left(\beta_{j_{1}}+\beta_{m+1}\right) & i=j_{1} \\ u \beta_{i-1}+u\left(\beta_{i}+\beta_{m+1}\right) & i=j_{1}+1 \\ \beta_{i}+(1-u) \beta_{m+1} & i=j_{1}+2, \ldots, m \\ u \beta_{i} & i=m+1\end{cases}
$$

The acceptance probability for the change-point relocation is $\left\{1, A_{r c}\right\}$ where

$$
A_{r c}=\mathrm{LR}_{r c} \times \text { Prior }_{r c} \times \text { Prop }_{r c}
$$

The likelihood ratio for change-point relocation move is

$$
\mathrm{LR}_{r c}=\frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, h, \tilde{\xi}_{1}, \boldsymbol{t}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, h, \xi_{1}, \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)} .
$$

Note that $f_{y}\left(y_{i} \mid k, h, \tilde{\xi}_{1}, \boldsymbol{t}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+3} \tilde{b}_{j} \tilde{\delta}_{j i}^{*}+$ $\sum_{j=1}^{2} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{j i}$ is defined as in (5.7). $\tilde{\delta}_{j i}^{*}$ as defined in Section 5.3 .2 is the $(i)$ th element of $\tilde{\boldsymbol{\delta}}_{j}^{*}$ which is the $(j)$ th basis vector found using $\boldsymbol{x}$, the knot location vector $\boldsymbol{t}$, and proposed change-point $\tilde{\xi}_{1}$. Here $f_{y}\left(y_{i} \mid k, h, \xi_{1}, \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\eta_{i}=\sum_{j=1}^{k+3} b_{j} \delta_{j i}^{*}+\sum_{j=1}^{2} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{j i}$ is defined as in (5.7). $\delta_{j i}^{*}$ is the $(i)$ th element of $\boldsymbol{\delta}_{j}^{*}$ found using using $\boldsymbol{x}$, the knot location vector $\boldsymbol{t}$, and change-point location prior to the move, $\xi_{1}$.

The prior ratio for the change-point relocation move is

$$
\begin{aligned}
\operatorname{Prior}_{r c} & =\frac{p(k) p(\boldsymbol{t} \mid k, h=1) p\left(\tilde{\xi}_{1}\right) p(\tilde{\boldsymbol{b}} \mid k, h=1) p(\boldsymbol{\alpha} \mid h=1) p(\tau)}{p(k) p(\boldsymbol{t} \mid k, h=1) p\left(\xi_{1}\right) p(\boldsymbol{b} \mid k, h=1) p(\boldsymbol{\alpha} \mid h=1) p(\tau)} \\
& =\left(\frac{\prod_{i=1}^{k+3} \tilde{b}_{i}}{\prod_{i=1}^{k+3} b_{i}}\right)^{c_{1}-1} \exp \left\{-c_{2}\left[\sum_{i=1}^{k+3} \tilde{b}_{i}-\sum_{j=1}^{k+3} b_{i}\right]\right\}
\end{aligned}
$$

The proposal ratio for the change-point relocation move, Prop $_{r c}$ depends on the values of $k, j_{1}, j_{1}^{*}, k_{\min }$, and $k_{\max }$. If $j_{1}=k_{\min }$ and $k>k_{\min }+1$ or if $j_{1}=k$ and $k>k_{\text {min }}+1, \operatorname{Prop}_{r c}=2 . \operatorname{Prop}_{r c}=0.5$ under the following cases:

1. If $j_{1}=k_{\text {min }}+1, k>k_{\text {min }}+1$, and $j_{1}^{*}=j_{1}-1$
2. If $j_{1}=k_{\text {min }}+1, k=k_{\text {min }}+2$, and $j_{1}^{*}=j_{1}+1$
3. If $j_{1}=k-1, k>k_{\text {min }}+2$, and $j_{1}^{*}=j_{1}+1$.

In all other cases, $\operatorname{Prop}_{r c}=1$.

### 5.5.4 Function Estimation for the RJMCMC 2 Algorithm

To find the function estimate, we first need to determine the location of the change-point using the RJMCMC 2 algorithm. We then find the function estimate by running the RJMCMC 1 algorithm with the change-point fixed at the estimate for the change-point found from the RJMCMC 2 algorithm. If we do not fix the change-point first and average the MCMC draws from the RJMCMC 2 algorithm, it is possible to get a function estimate that violates the shape restriction. Thus, we first use the posterior draws from the RJMCMC 2 algorithm for the change-point location to estimate the change-point location. Denote the value of $\xi_{1}$ for the $(i)$ th iteration of the RJMCMC 2 algorithm by $\xi_{1}^{(i)}$. Using a mode estimation procedure, we estimate the mode of the RJMCMC 2 algorithm values of $\xi^{(i)}$, after discarding burn-in and denote the mode estimate by $\hat{\xi}_{1}$. For our analysis, we use the mode estimation method of Robertson and Cryer (1974). Given $\hat{\xi}_{1}$, we run the RJMCMC 1 algorithm discussed in Section 5.4 setting $h=1$ and $\xi_{1}=\hat{\xi}_{1}$. We use the values of $k, \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{\alpha}$, and $\tau$ from the $(i)$ th iteration after the Gibbs sampler update of the RJMCMC 1 algorithm with change-point fixed at $\hat{\xi}_{1}$ and denote them by $k^{(i)}, \boldsymbol{t}^{(i)}$, $\boldsymbol{\beta}^{(i)}$, and $\boldsymbol{\alpha}^{(i)}$, respectively. The estimate of the mean regression function at iteration $i$ is then given by $\boldsymbol{\eta}^{(i)}=\sum_{j=1}^{k+3} b_{j}^{(i)} \boldsymbol{\delta}_{j}^{*}+\sum_{j=1}^{2} \alpha_{j}^{(i)} \boldsymbol{v}_{j}$ where $\boldsymbol{\delta}_{j}^{*}$ and $\boldsymbol{v}_{j}$ are the basis vectors created using vector of current knot locations, $\boldsymbol{t}$, and $\hat{\xi}_{1}$. The regression spline estimate is found by averaging $\boldsymbol{\eta}^{(i)}$ after discarding burn-in and is given by

$$
\begin{equation*}
\hat{\boldsymbol{\eta}}=\frac{1}{N-B} \sum_{i=B+1}^{N} \eta^{(i)} \tag{5.41}
\end{equation*}
$$

where $N$ is the total iterations in the RJMCMC algorithm and $B$ is the burn-in.

### 5.5.5 Example for the RJMCMC 2 Algorithm

To examine the performance of the RJMCMC 2 algorithm for single changepoint model with location of change-point unknown, we consider the single changepoint data set in Section 5.4.5 and shown in Figure 5.1. We run the algorithm
for 100,000 iterations and discard the first 20,000 as burn-in. We use the same hyperparameters as in Section 5.4.5. The posterior distribution for the changepoint location is given in Figure 5.5(a) along with the estimated mode, $\hat{\xi}_{1}=0.509$, denoted by the dashed vertical red line and found using the method of Robertson and Cryer (1974). The value for $\xi_{1}$ used to generate the data is 0.5 and denote by the vertical gray line in Figure 5.5(a). The function estimate found after fixing the change-point at the estimate of 0.509 and using the RJMCMC 1 algorithm in Section 5.4 is shown in Figure 5.5(b). The function estimate is similar to the estimate fixed at the true value for $\xi_{1}$ given in Figure 5.1.



Figure 5.5: (a) Histogram of the posterior draws for $\xi_{1}$ for the RJMCMC 2 algorithm for the single change-point data set in Section 5.4.5 after discarding burn-in. The vertical dashed red line is the estimate of $\xi_{1}$ found using mode estimation procedure of Robertson and Cryer (1974) and the vertical gray line is true value for $\xi_{1}$ used to generate the data. (b) Plot of simulated data from a regression function with one change-points and the estimated regression function (red dashed line) using the RJMCMC 2 algorithm along with pointwise $95 \%$ HPD intervals (shaded region) when fixing $\xi_{1}$ at $\hat{\xi}_{1}$ and the function used to simulate the data (solid gray line). The value of the change-point estimate used to simulate the data is indicated by the vertical line.

The partial trace plots for $k$ and $\tau$ for the single change-point data set are shown in Figures 5.6(a) and (c), respectively. Again, we selected to look at these parameters because they do not change meaning throughout the RJMCMC algorithm. The partial trace for $k$ does appear to propose the same value for $k$ for several iterations. However, the RJMCMC 2 algorithm has two moves that do not change the value
of $k$. In particular, many different change-point locations can be explored using the same value for $k$. Therefore, considering the mixing of $k$ may not be the best indicator of mixing for the entire algorithm. The partial trace for $\tau$ suggests good mixing. We again look at the MCSE defined in Section 4.4.1 but only calculate it for $\tau$ since $k$ is thought to be a poor indicator of mixing for the RJMCMC 2 algorithm. The MCSE for $\tau$ equal to 0.0050 which is less than $5 \%$ of the sample standard deviation for $\tau$ (0.0276) suggesting the algorithm has run long enough. The histograms of posterior draws for $k$ and $\tau$ are given in Figures 5.6(b) and (d), respectively. We again find the estimates for $k$ and $\tau$ by averaging the posterior draws for each parameter after discarding burn-in. We find the estimate for $k$ is $\hat{k}=5.80$ and the estimate for $\tau$ is $\hat{\tau}=3.54$. The estimate for $\tau$ is larger than the value of $\tau$ used to generate the data, 3.03 (for the scaled data), but 3.03 is included in the $95 \%$ HPD interval for $\tau,(2.53,4.60)$.

### 5.6 Existence of Change-point Model

In this section, we propose a RJMCMC algorithm that moves between models with the number of change-points equal to zero (no change-points model) or one (single change-point model) and we will denote this algorithm by the RJMCMC 3 algorithm. It can be used to determine if a model with a change-point or no change-points is most appropriate for a given data set. We consider the normal errors multiple change-point model under the monotone shape restriction discussed in Section 5.3.3. In both the no change-points and single change-point model, the number of interior knots and their locations are free parameters. For the model with no change-points, the function is assumed to be monotone increasing or decreasing over the range of the covariate, $x$. For the single change-point model, the function is assumed to monotone increasing or decreasing between the minimum value for the covariate and the change-point as well as monotonically increasing or decreasing between the change-point and the maximum value for the covariate. Also, for


Figure 5.6: (a) Partial trace plot for $k$ for the single change-point data set for the RJMCMC 2 algorithm. (b) Histogram of the posterior draws for $k$ for the RJMCMC 2 algorithm for the single change-point data set after discarding burn-in. (c) Partial trace plot for $\tau$ for the single change-point data set for the RJMCMC 2 algorithm. (d) Histogram of the posterior draws for $\tau$ for the RJMCMC 2 algorithm for the single change-point data set after discarding burn-in. The grey solid line denotes the value of $\tau$ used to generate the data and the red dashed lines mark the $95 \%$ HPD interval for $\tau$ found using method discussed in Section 2.6.1.
the single change-point model, the location of the change-point is assumed to be unknown. The proposed RJMCMC 3 algorithm combines new moves with moves in the RJMCMC 2 algorithm given in Section 5.5 as well as moves in the RJMCMC algorithm for Bayes SRRS without change-points from Chapter 4.

### 5.6.1 Overview of the RJMCMC 3 Algorithm

The RJMCMC 3 algorithm can be summarized as follows:

1. Perform a "birth", "death", "relocation", "birth of a change-point", or "death of a change-point" move and accept the proposed parameters with given acceptance probability.
2. Perform a "coefficient update" step using a Gibbs sampler step to update either $\boldsymbol{\alpha}, \boldsymbol{b}$, and $\tau$ if have the single change-point model or update $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\tau$ if have the no change-points model.
3. Repeat until convergence.

The probabilities of performing a birth, death, relocation, birth of a change-point, or death of a change-point move, denoted by $b_{k}, d_{k}, r_{k}, b c_{k}$, and $d c_{k}$, respectively, are given in Table 5.2 for the no change-points model $(h=0)$ and in Table 5.3 for the single change-point model $(h=1)$ where $b_{k}$, and $d_{k}$ are defined as in Section 5.4.1 and in Section 4.2.1 but with $c=0.2$. The new values for $b_{k}, d_{k}$, and $r_{k}$ when $k=k_{\min }$ and $k=k_{\max }$ are given in Tables 5.2 and 5.3. These five moves and there acceptance probabilities are discussed in Section 5.6.3. The coefficient update step when $h=1$, updates the coefficient using a Gibbs sampler step and the conditional distributions for the model parameters given in Section 5.3.3. For the coefficient update step, if $h=0$ and we have the no change-points model, we update the parameters using the conditional distributions given in Section 2.4.

Table 5.2: Probabilities for five moves in the proposed RJMCMC 3 algorithm to determine the existence of a change-point when $h=0$.

| Move | $k=k_{\min }$ | $k_{\min }<k<k_{\max }$ | $k=k_{\max }$ |
| :---: | :---: | :---: | :---: |
| birth | $b_{k}=\frac{1}{3}$ | $b_{k}$ | $b_{k}=0$ |
| death | $d_{k}=0$ | $d_{k}$ | $d_{k}=\frac{1}{3}$ |
| relocation | $r_{k}=\frac{1}{3}$ | $r_{k}=\frac{1}{3}\left(1-b_{k}-d_{k}\right)$ | $r_{k}=\frac{1}{3}$ |
| birth of a change-point | $b c_{k}=\frac{1}{3}$ | $b c_{k}=\frac{2}{3}\left(1-b_{k}-d_{k}\right)$ | $b c_{k}=\frac{1}{3}$ |
| death of a change-point | $d c_{k}=0$ | $d c_{k}=0$ | $d c_{k}=0$ |

### 5.6.2 Priors for the RJMCMC 3 Algorithm

In addition to introducing new move probabilities for the model when the number of change-points, $h$, is allowed to take on the values of zero or one, we also need to assume a prior for $h$. We assume a discrete uniform distribution over $\{0,1\}$ for $h$ so $\operatorname{Pr}(h=0)=0.5$ and $\operatorname{Pr}(h=1)=0.5$. Given $h$, we assume priors for

Table 5.3: Probabilities for five moves in the proposed RJMCMC 3 algorithm to determine the existence of a change-point when $h=1$.

| Move | $k=k_{\min }$ | $k_{\min }<k<k_{\max }$ | $k=k_{\max }$ |
| :---: | :---: | :---: | :---: |
| birth | $b_{k}=\frac{1}{3}$ | $b_{k}$ | $b_{k}=0$ |
| death | $d_{k}=0$ | $d_{k}$ | $d_{k}=\frac{1}{3}$ |
| relocation | $r_{k}=\frac{1}{3}$ | $r_{k}=\frac{1}{3}\left(1-b_{k}-d_{k}\right)$ | $r_{k}=\frac{1}{3}$ |
| birth of a change-point | $b c_{k}=0$ | $b c_{k}=0$ | $b c_{k}=0$ |
| death of a change-point | $d c_{k}=\frac{1}{3}$ | $d c_{k}=\frac{2}{3}\left(1-b_{k}-d_{k}\right)$ | $d c_{k}=\frac{1}{3}$ |

the other model parameters as was used in the previous RJMCMC algorithms. In particular, we use the same prior for $k$ as for the RJMCMC 1 algorithm in Section 5.4.2 with $k_{\text {min }}=3$. Given $k$, we use the same prior for the interior knot locations as the RJMCMC 2 algorithm given in Section 5.5.2 and the RJMCMC algorithm without change-points given in Section 4.2.2. When $h=1$, we assume the same prior for the change-point, $\xi_{1}$, as in the the RJMCMC 2 algorithm given in Section 5.5.2. For both the no change-points $(h=0)$ and single change-point model ( $h=1$ ), we assume independent normal priors with mean of zero and variance of $M$ for the unconstrained coefficients, $\alpha_{j}$ and a gamma prior for $\tau$ with shape $d_{1}$ and rate $d_{2}$. For the constrained coefficients in both the no change-points model and the single change-point model, which are given by $\beta_{j}$ for $j=1, \ldots k+1$ and $b_{j}$ for $j=1, \ldots, k+2+h$, respectively, we assume independent gamma prior with shape $c_{1}$ and rate $c_{2}$.

### 5.6.3 RJMCMC 3 Algorithm Implementation

In this section, we discuss each move type and the acceptance probabilities for the RJMCMC 3 algorithm.

### 5.6.3.1 Birth, Death, and Relocation Moves for the RJMCMC 3 Algorithm

The birth, death, and relocation move for the RJMCMC 3 algorithm propose new knot locations for interior knots that are not change-points as in the RJMCMC 1 algorithm in Section 5.4. The new knot locations for the birth, death, and relocation
moves in this algorithm are proposed the same way as in the RJMCMC 1 algorithm in Sections 5.4.3.1, 5.4.3.2, and 5.4.3.3, respectively. If $h=0$, the new coefficients for the birth, death, and relocation moves are proposed the same as in the RJMCMC without change-points given in Sections 4.3.1, 4.3.2, and 4.3.3, respectively. Furthermore, when $h=0$, the acceptance probabilities for birth, death, and relocation moves are the same as acceptance probabilities in Sections 4.3.1, 4.3.2, and 4.3.3, respectively. If $h=1$, the birth, death, and relocation moves propose coefficients in the same was as in the RJMCMC 1 algorithm given in Sections 5.4.3.1, 5.4.3.2, and 5.4.3.3, respectively. For single change-point model, the acceptance probabilities for the birth, death, and relocation moves are the same as in the RJMCMC 2 algorithm given in Section 5.5.3.1.

### 5.6.3.2 Birth of a Change-point Move

The birth of a change-point move adds a change-point to the no change-points model. It proposes a move from a model with $h=0$ to a model with $h=1$. For this move, we add a change-point at an existing interior knot location and update the coefficients based on this change-point addition. Let $\boldsymbol{t}=\left(t_{1}, \ldots, t_{k+2}\right)^{\prime}$ be the ordered current knot locations for the model with $k$ interior knots and no change-point. We select the change-point from a set of candidate knot interior knot locations, denoted by $C P_{\text {cand }}$. Since we want at least one interior knot between $t_{1}$ and $\xi_{1}$ and at least one interior knot between $\xi_{1}$ and $t_{k+2}$, we exclude $t_{2}$ and $t_{k+1}$ from the candidate interior knot locations for the change-point. Thus, $C P_{\text {cand }}=\left\{t_{3}, \ldots, t_{k}\right\}$. From $C P_{\text {cand }}$, we randomly select an interior knot location and add a change-point at this location. We denote the proposed change-point by $\xi_{1}^{*}=t_{j}^{*}$. The probability of $\xi_{1}^{*}$ given $\boldsymbol{t}$ is $1 /(k-2)$. Given $\xi_{1}^{*}$, we create basis vectors $\tilde{\boldsymbol{\delta}}_{l}^{*}$ for $l=1, \ldots, k+3$ as defined in Section 5.3.2 using $\xi_{1}^{*}$ and $\boldsymbol{t}$. Let $\tilde{\boldsymbol{v}}_{1}$ and $\tilde{\boldsymbol{v}}_{2}$ be the basis vectors as defined in Section 5.3.1 found after adding $\xi_{1}^{*}$.

Let $\boldsymbol{\delta}_{l}$ for $l=1, \ldots, k+2$ be the quadratic $I$-spline basis functions for the model when $h=0$ found using $\boldsymbol{t}$ as defined in Section 1.4.1. Furthermore, let
$\boldsymbol{\beta}=\left(\beta_{1}, \ldots, \beta_{k+2}\right)$ be the constrained coefficients prior to adding $\xi_{1}^{*}$ with $\beta_{l}$ the coefficient for $\boldsymbol{\delta}_{l}$ where $\boldsymbol{\delta}_{l}$ is the basis vector corresponding to the basis function with a positive slope at the $(l)$ th largest knot location. Let $\alpha_{1}$ be the unconstrained coefficient for model when $h=0$ and the coefficient of the vector $\boldsymbol{v}_{1}$ which is a one vector of length $n$. Let $\tilde{\boldsymbol{b}}=\left(\tilde{b}_{1}, \ldots, \tilde{b}_{k+3}\right)^{\prime}$ be the coefficient vector for the basis vectors $\tilde{\boldsymbol{\delta}}_{l}^{*}$ for $l=1, \ldots, k+3$ after adding change-point $\xi_{1}^{*}$. Likewise, let $\tilde{\boldsymbol{\alpha}}=\left(\tilde{\alpha}_{1}, \tilde{\alpha}_{2}\right)^{\prime}$ be the unconstrained coefficients for the model after adding changepoint $\xi_{1}^{*}$. Note the $\xi_{1}^{*}=t_{j}$ for some $j=3, \ldots k-1$ and will fall between $t_{j-1}$ and $t_{j+1}$. Using this, we propose values for $\tilde{\boldsymbol{b}}$ according to

$$
\tilde{b}_{l}= \begin{cases}\beta_{l} & \text { for } l=1, \ldots j-1  \tag{5.42}\\ (1-u) \beta_{l} & \text { for } l=j \\ (1-u) \beta_{l} & \text { for } l=j+1 \\ \beta_{l} & \text { for } l=j+2, \ldots, k+2 \\ u\left(\beta_{j}+\beta_{j+1}\right) & \text { for } l=k+3\end{cases}
$$

where $u$ is a random variable from a continuous uniform distribution on $(0,1)$. Note that this coefficient update step ensures all elements of $\tilde{\boldsymbol{b}}$ are positive so it preserves the monotonically increasing shape restriction. We propose values for $\tilde{\boldsymbol{\alpha}}$ according to

$$
\begin{equation*}
\tilde{\alpha}_{1}=\frac{\sum_{j=1}^{k+2} \beta_{j} \sum_{i=1}^{n} \delta_{j i} \tilde{v}_{1 i}}{\sum_{i=1}^{n} \tilde{v}_{1 i}^{2}}+\alpha_{1} \frac{\sum_{i=1}^{n} \tilde{v}_{1 i}}{\sum_{i=1}^{n} \tilde{v}_{1 i}^{2}}+e_{1} \tag{5.43}
\end{equation*}
$$

and

$$
\begin{equation*}
\tilde{\alpha}_{2}=\frac{\sum_{j=1}^{k+2} \beta_{j} \sum_{i=1}^{n} \delta_{j i} \tilde{v}_{2 i}}{\sum_{i=1}^{n} \tilde{v}_{2 i}^{2}}+\alpha_{1} \frac{\sum_{i=1}^{n} \tilde{v}_{2 i}}{\sum_{i=1}^{n} \tilde{v}_{2 i}^{2}} \tag{5.44}
\end{equation*}
$$

where $e_{1}$ is a random normal random variable with mean 0 and standard deviation $\sigma_{e}$. For the examples in this chapter, we let $\sigma_{e}=0.1$. Note the unconstrained coefficient update is the estimate for the regression when $h=0$ projected onto the new basis functions plus some random error. These parameter coefficient updates were chosen because they give a regression function estimate after adding $\xi_{1}^{*}$ that is very similar to the function estimate prior to adding $\xi_{1}^{*}$.

Next, we derive the acceptance probability for the birth of a change-point move for the RJMCMC 3 algorithm which is equal to $\min \left\{1, A_{b c}\right\}$ where

$$
A_{b c}=\mathrm{LR}_{b c} \times \text { Prior }_{b c} \times \text { Prop }_{b c} \times \mathrm{Jacob}_{b c}
$$

The likelihood ratio for the birth of a change-point move for the RJMCMC 3 algorithm is

$$
\operatorname{LR}_{b c}=\frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \tilde{h}=1, \xi_{1}^{*}, \boldsymbol{t}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, h=0, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right)}
$$

where $f_{y}\left(y_{i} \mid k, \tilde{h}=1, \xi_{1}^{*}, \boldsymbol{t}, \tilde{\boldsymbol{b}}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+3} \tilde{b}_{j} \tilde{\delta}_{j i}^{*}+$ $\sum_{j=1}^{2} \tilde{\alpha}_{j} \tilde{v}_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $\tilde{v}_{j i}$ is the $(i)$ th element of $\tilde{\boldsymbol{v}}_{j}$ and $\tilde{\delta}_{j i}^{*}$ is the $(i)$ th element of $\tilde{\boldsymbol{\delta}}_{j}^{*}$. Likewise, $f_{y}\left(y_{i} \mid k, h=0, \boldsymbol{t}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\eta_{i}=\sum_{j=1}^{k+2} \beta_{j} \delta_{j i}+\alpha_{1} v_{1 i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{1 i}$ is the $(i)$ th element of $\boldsymbol{v}_{1}$ and $\delta_{j i}$ is the $(i)$ th element of $\boldsymbol{\delta}_{j}$.

The prior ratio for the birth of a change-point move in the RJMCMC 3 algorithm is

$$
\begin{aligned}
\operatorname{Prior}_{b c} & =\frac{p(k) p(\tilde{h}=1) p(\boldsymbol{t} \mid k, h) p\left(\xi_{1}^{*}\right) p(\tilde{\boldsymbol{b}} \mid k, \tilde{h}=1) p(\boldsymbol{\alpha} \mid \tilde{h}=1) p(\tau)}{p(k) p(h=0) p(\boldsymbol{t} \mid k) p(\boldsymbol{\beta} \mid k, h=0) p(\boldsymbol{\alpha} \mid h=0) p(\tau)} \\
& =\frac{\left[\prod_{i=1}^{k+3} f_{b}\left(\tilde{b}_{i}\right)\right] \cdot\left[\prod_{i=1}^{2} f_{\alpha}\left(\tilde{\alpha}_{i}\right)\right]}{(k-2) \cdot\left[\prod_{i=1}^{k+2} f_{\beta}\left(\beta_{i}\right)\right] \cdot f_{\alpha}\left(\alpha_{1}\right)}
\end{aligned}
$$

where $f_{b}(z)$ is the density of the gamma prior for the $b$ parameters with hyperparameters $c_{1}$ and $c_{2}$ evaluated at $z$ and $f_{\beta}(z)$ is the density of the gamma prior for the $\beta$ parameters evaluated at $z$ which is also a gamma prior with hyperparameters $c_{1}$ and $c_{2}$. Also, $f_{\alpha}(z)$ is the density evaluated at $z$ for the normal prior with mean zero and variance $M$ for the unconstrained coefficients in both the no change-points and single change-points model.

The proposal ratio for the birth of a change-point move in the RJMCMC 3 algorithm is

$$
\begin{align*}
\operatorname{Prop}_{b c} & =\frac{\operatorname{Pr}(\text { delete change-point for } k \text { model }) \operatorname{Pr}\left(\text { delete } \xi_{1}^{*}\right)}{\operatorname{Pr}(\text { add change-point for } k \text { model }) \operatorname{Pr}\left(\xi_{1}^{*}\right) \operatorname{Pr}(u) \operatorname{Pr}\left(e_{1}\right)} \\
& =\frac{d c_{k} \cdot 1}{b c_{k} \cdot \frac{1}{k-2} \cdot \frac{1}{1-0} \cdot f_{e}\left(e_{1}\right)} \\
& =\frac{d c_{k}(k-2)}{b c_{k} \cdot f_{e}\left(e_{1}\right)} \tag{5.45}
\end{align*}
$$

where $f_{e}\left(e_{1}\right)$ is the density for the normal distribution for $e_{1}$ with mean zero and standard deviation of $\sigma_{e}$ evaluated at the value of $e_{1}$ proposed in the add a changepoint move. Note that $\operatorname{Pr}$ (delete change-point for $k$ model) is the probability of performing a death of a change-point move with $h=1$ and $k$ interior knots. Likewise, $\operatorname{Pr}($ add change-point for $k$ model $)$ is the probability of performing an birth of a change-point move with $h=0$ and $k$ interior knots.

The absolute value of the determinant of the Jacobian for the birth of a changepoint move in the RJMCMC 3 algorithm, $\mathrm{Jacob}_{b c}$, is found by using the coefficient update functions given in (5.42), (5.43) and (5.44). For this move, the determinant of the Jacobian is found by taking the determinant of a matrix of the form

$$
J=\left[\begin{array}{ll}
J_{1} & J_{2} \\
J_{3} & J_{4}
\end{array}\right]
$$

where $J_{1}, J_{2}, J_{3}$, and $J_{4}$ are found by first letting

$$
\begin{gathered}
\tilde{a}_{1 l}=\frac{\sum_{i=1}^{n} \delta_{l i} \tilde{v}_{1 i}}{\sum_{i=1}^{n} \tilde{v}_{1 i}^{2}}, \\
\tilde{a}_{2 l}=\frac{\sum_{i=1}^{n} \delta_{l i} \tilde{v}_{2 i}}{\sum_{i=1}^{n} \tilde{v}_{2 i}^{2}} \\
\tilde{s}_{v 1}=\frac{\sum_{i=1}^{n} \tilde{v}_{1 i}}{\sum_{i=1}^{n} \tilde{v}_{1 i}^{2}}
\end{gathered}
$$

and

$$
\tilde{s}_{v 2}=\frac{\sum_{i=1}^{n} \tilde{v}_{2 i}}{\sum_{i=1}^{n} \tilde{v}_{2 i}^{2}} .
$$

Then

$$
\begin{gathered}
J_{1}=\left[\begin{array}{cccccc}
1-u & 0 & -\beta_{j} & 0 & 0 \\
0 & 1-u & -\beta_{j+1} & 0 & 0 \\
u & u & \beta_{j}+\beta_{j+1} & 0 & 0 \\
\tilde{a}_{1 j} & \tilde{a}_{1, j+1} & 0 & \tilde{s}_{v 1} & 1 \\
\tilde{a}_{2 j} & \tilde{a}_{2, j+1} & 0 & \tilde{s}_{v 2} & 0
\end{array}\right], \\
J_{2}=\left[\begin{array}{ccccccc}
0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1, j-1} & \tilde{a}_{1, j+2} & \cdots & \tilde{a}_{1, k+2} \\
\tilde{a}_{21} & \tilde{a}_{22} & \cdots & \tilde{a}_{2, j-1} & \tilde{a}_{2, j+2} & \cdots & \tilde{a}_{2, k+2}
\end{array}\right],
\end{gathered}
$$

$J_{3}$ is a $k \times 5$ matrix of zeros, and $J_{4}$ is an $k \times k$ identity matrix. The $\mathrm{Jacob}_{b c}$ is the absolute value of the the determinant of $J$.

### 5.6.3.3 Death of a Change-point Move

The death of a change-point move is the inverse of the add the change-point move in Section 5.6.3.2. The change-point location, $\xi_{1}$, is removed from a model with $h=1, k$ interior knots, knot location vector $\boldsymbol{t}=\left(t_{1}, \ldots, t_{k+2}\right)^{\prime}$, constrained coefficients $\boldsymbol{b}=\left(b_{1}, \ldots, b_{k+3}\right)^{\prime}$, and unconstrained coefficients $\boldsymbol{\alpha}=\left(\alpha_{1}, \alpha_{2}\right)^{\prime}$. Let $\boldsymbol{\delta}_{j}^{*}=\left(\delta_{j 1}^{*}, \ldots, \delta_{j n}^{*}\right)^{\prime}$ for $j=1, \ldots, k+3, \boldsymbol{v}_{1}$, and $\boldsymbol{v}_{2}$ be the basis vectors for model prior to deleting $\xi_{1}$. After removing $\xi_{1}$, we have only an interior knot at $t_{j}$ where $\xi_{i}$ was. Let $\tilde{\boldsymbol{\delta}}_{l}=\left(\tilde{\delta}_{l 1}, \ldots, \tilde{\delta}_{l n}\right)^{\prime}$ for $l=1, \ldots, k+2$ be the quadratic $I$-spline basis functions found using $\boldsymbol{t}$ after deleting $\xi_{1}$ and let $\tilde{\boldsymbol{v}}_{1}$ be the one vector of length $n$ for the no change-points model after deleting $\xi_{1}$. Let $\tilde{\boldsymbol{\beta}}=\left(\tilde{\beta}_{1}, \ldots, \tilde{\beta}_{k+2}\right)^{\prime}$ be the constrained coefficients for the no change-points model after deleting $\xi_{1}$. We propose values for $\tilde{\boldsymbol{\beta}}$ by inverting (5.42). Thus, the constrained coefficients for the no change-points are

$$
\tilde{\beta}_{l}=\left\{\begin{array}{ll}
b_{l} & \text { for } l=1, \ldots, j-1 \\
\frac{b_{l}\left(b_{l}+b_{l+1}+b_{k+3}\right)}{b_{l}+b_{l+1}} & \text { for } l=j \\
\frac{b_{l+1}\left(b_{l}+b_{l+1}+b_{k+3}\right)}{b_{l}+b_{l+1}} & \text { for } l=j+1 \\
b_{l} & \text { for } l=j+2, \ldots, k+2
\end{array} .\right.
$$

The proposed value for the unconstrained coefficient $\tilde{\alpha}_{1}$ for the no change-points model is found by inverting (5.44) and is equal to

$$
\tilde{\alpha}_{1}=\left(\alpha_{2}-\frac{\sum_{l=1}^{k+2} \tilde{\beta}_{j} \sum_{i=1}^{n} \tilde{\delta}_{j i} v_{2 i}}{\sum_{i=1}^{n} v_{2 i}^{2}}\right)\left(\frac{\sum_{i=1}^{n} v_{2 i}^{2}}{\sum_{i=1}^{n} v_{2 i}}\right) .
$$

Since the death of a change-point move is the inverse of the birth of a changepoint move from $h=0$ to $h=1$, the acceptance probability for the death of a change-point move is found by inverting the acceptance probability for the birth of a change-point move. It is equal to $\min \left\{1, A_{d c}^{-1}\right\}$ with

$$
A_{d c}=\mathrm{LR}_{d c} \times \operatorname{Prior}_{d c} \times \operatorname{Prop}_{d c} \times \mathrm{Jacob}_{d c}
$$

$\mathrm{LR}_{d c}$ is the likelihood ratio used in the death of a change-point move and is equal to

$$
\mathrm{LR}_{d c}=\frac{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, h=1, \xi_{1}, \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)}{\prod_{i=1}^{n} f_{y}\left(y_{i} \mid k, \tilde{h}=0, \boldsymbol{t}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}, \tau\right)}
$$

where $f_{y}\left(y_{i} \mid k, h=1, \xi_{1}, \boldsymbol{t}, \boldsymbol{b}, \boldsymbol{\alpha}, \tau\right)$ is the normal density with mean $\eta_{i}=\sum_{j=1}^{k+3} b_{j} \delta_{j i}^{*}+$ $\sum_{j=1}^{2} \alpha_{j} v_{j i}$ and standard deviation $\tau^{-1 / 2}$ where $v_{j i}$ is the $(i)$ th element of $\boldsymbol{v}_{j}$ and $\delta_{j i}^{*}$ is the $(i)$ th element of $\boldsymbol{\delta}_{j}^{*}$. $f_{y}\left(y_{i} \mid k, \tilde{h}=0, \boldsymbol{t}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}}, \tau\right)$ is the normal density with mean $\tilde{\eta}_{i}=\sum_{j=1}^{k+2} \tilde{\beta}_{j} \tilde{\delta}_{j i}+\tilde{\alpha}_{1} \tilde{v}_{1 i}$ and standard deviation of $\tau^{-1 / 2}$ where $\tilde{v}_{1 i}$ is the $(i)$ th element of $\tilde{\boldsymbol{v}}_{1}$ and $\delta_{j i}$ is the $(i)$ th element of $\tilde{\boldsymbol{\delta}}_{j}$. For the death of a change-point move,

$$
\begin{aligned}
\operatorname{Prior}_{d c} & =\frac{p(k) p(h=1) p(\boldsymbol{t} \mid k) p\left(\xi_{1} \mid \boldsymbol{t}\right) p(\boldsymbol{b} \mid k, h=1) p(\boldsymbol{\alpha} \mid h=1) p(\tau)}{p(k) p(\tilde{h}=0) p(\boldsymbol{t} \mid k) p(\tilde{\boldsymbol{\beta}} \mid k, \tilde{h}=0) p(\tilde{\boldsymbol{\alpha}} \mid \tilde{h}=0) p(\tau)} \\
& =\frac{\left[\prod_{i=1}^{k+3} f_{b}\left(b_{i}\right)\right] \cdot\left[\prod_{i=1}^{2} f_{\alpha}\left(\alpha_{i}\right)\right]}{(k-2) \cdot\left[\prod_{i=1}^{k+2} f_{\beta}\left(\beta_{i}\right)\right] \cdot f_{\alpha}\left(\tilde{\alpha}_{1}\right)}
\end{aligned}
$$

and

$$
\begin{align*}
\operatorname{Prop}_{d c} & =\frac{\operatorname{Pr}(\text { delete change-point for } k \text { model }) \operatorname{Pr}\left(\text { delete } \xi_{1}\right)}{\operatorname{Pr}(\text { add change-point for } k \text { model }) \operatorname{Pr}\left(\xi_{1}\right) \operatorname{Pr}(u) \operatorname{Pr}\left(\tilde{e}_{1}\right)} \\
& =\frac{d c_{k} \cdot 1}{b c_{k} \cdot \frac{1}{k-2} \cdot \frac{1}{1-0} \cdot f_{e}\left(\tilde{e}_{1}\right)} \\
& =\frac{d c_{k}(k-2)}{b c_{k} \cdot f_{e}\left(\tilde{e}_{1}\right)} \tag{5.46}
\end{align*}
$$

where $f_{e}\left(\tilde{e}_{1}\right)$ is the density for a normal distribution with mean zero and standard deviation of $\sigma_{e}$ evaluated at the value of $e_{1}$ proposed in the add a change-point move, $\tilde{e}_{1}$. We find the value for $\tilde{e}_{1}$ by

$$
\tilde{e}_{1}=\alpha_{1}-\frac{\sum_{l=1}^{k+2} \tilde{\beta}_{j} \sum_{i=1}^{n} \tilde{\delta}_{j i} v_{1 i}}{\sum_{i=1}^{n} v_{1 i}^{2}}-\frac{\sum_{i=1}^{n} v_{1 i}^{2}}{\sum_{i=1}^{n} v_{1 i}} \tilde{\alpha}_{1} .
$$

$\mathrm{Jacob}_{d c}$ is found by taking the absolute value of the determinant of a matrix of the form

$$
J=\left[\begin{array}{ll}
J_{1} & J_{2} \\
J_{3} & J_{4}
\end{array}\right]
$$

where

$$
\begin{gathered}
J_{1}=\left[\begin{array}{cccccc}
1-\tilde{u} & 0 & -\tilde{\beta}_{j} & 0 & 0 \\
0 & 1-\tilde{u} & -\tilde{\beta}_{j+1} & 0 & 0 \\
\tilde{u} & \tilde{u} & \tilde{\beta}_{j}+\tilde{\beta}_{j+1} & 0 & 0 \\
a_{1 j} & a_{1, j+1} & 0 & s_{v 1} & 1 \\
a_{2 j} & a_{2, j+1} & 0 & s_{v 2} & 0
\end{array}\right], \\
J_{2}=\left[\begin{array}{ccccccc}
0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
a_{11} & a_{12} & \cdots & a_{1, j-1} & a_{1, j+2} & \cdots & a_{1, k+2} \\
a_{21} & a_{22} & \cdots & a_{2, j-1} & a_{2, j+2} & \cdots & a_{2, k+2}
\end{array}\right],
\end{gathered}
$$

$J_{3}$ is a $k \times 5$ matrix of zeros, and $J_{4}$ is an $k \times k$ identity matrix. For the death of a change-point move,

$$
\begin{aligned}
& a_{1 l}=\frac{\sum_{i=1}^{n} \tilde{\delta}_{l i} v_{1 i}}{\sum_{i=1}^{n} v_{1 i}^{2}}, \\
& a_{2 l}=\frac{\sum_{i=1}^{n} \tilde{\delta}_{l i} v_{2 i}}{\sum_{i=1}^{n} v_{2 i}^{2}} \\
& s_{v 1}=\frac{\sum_{i=1}^{n} v_{1 i}}{\sum_{i=1}^{n} v_{1 i}^{2}} \\
& s_{v 2}=\frac{\sum_{i=1}^{n} v_{2 i}}{\sum_{i=1}^{n} v_{2 i}^{2}}
\end{aligned}
$$

and

$$
\tilde{u}=\frac{b_{k+3}}{b_{j}+b_{j+1}+b_{k+3}} .
$$

### 5.6.4 Function Estimation for the RJMCMC 3 Algorithm

To find the function estimate for the RJMCMC 3 algorithm, we first try to determine whether the no change-points or the single change-point model is appropriate. We can do this by looking at the posterior draws for $h$ after discarding burn-in iterations for the RJMCMC 3 algorithm. If the majority of posterior draws for $h$ after discarding burn-in have $h=0$ then this suggests that the no changepoints model is preferred over the single change-point model. If the majority of the posterior draws for $h$ after discarding burn-in have $h=1$ then this suggests that the single change-point model is the most appropriate model. If the number of draws for $h$ are about the same for $h=0$ and $h=1$, we can estimate the function under both the no change-points and the single change-point model. We can give the proportion of posterior draws of $h$ after discarding burn-in for which $h=0$ and the proportion of posterior draws of $h$ after discarding burn-in for which $h=1$ to give an indication of the preferences for both models. To find the function estimate for the single change-point model, we estimate the function as in Section 5.5.4. To find the function estimate for the no change-points model, we estimate the function as in Section 4.2.1 using (4.3).

### 5.6.5 Examples for the RJMCMC 3 Algorithm

We consider three simulated data sets to examine the performance of the RJMCMC 3 algorithm to determine the existence of a change-point. The first simulated data set is the single change-point data set discussed in Sections 5.4.5 and 5.5.5 and shown in Figure 5.7(a). The second simulated data set is simulated from the same single change-point model as the first data set but with the error standard deviation equal to 0.5 as opposed to 0.35 . Like the first data set, it is scaled such that $y$-values have variance of 1 and mean of zero. The second data set is shown in Figure 5.8(a). The third data set is generated from a function without change-points and we refer to this data set as the no change-points data set. It is the data set from Section 4.4.1
with the sample size, $n$, increased to 100 and the data scaled such that the $y$-values have a mean of zero and variance of 1 and is shown in Figure 5.9(a). For all data sets we run the algorithm for 100,000 iterations discarding the first 20,000 runs as burn-in. We let $\lambda=5, k_{\min }=3, k_{\max }=n / 4, c_{1}=0.2, c_{2}=0.1, d_{1}=0.2, d_{2}=0.2$, and $\sigma_{e}=0.1$.

For the first data set, the RJMCMC 3 algorithm has $h=1$ for all posterior draws of $h$ (after discarding burn-in). Thus, the RJMCMC 3 algorithm easily detects the existence of the change-point. The partial trace plot for $\tau$ for the first data set with the error standard deviation 0.35 is given in Figure 5.7(c) and suggests good mixing of the RJMCMC 3 algorithm. We feel that looking at the mixing of the number of interior knots, $k$, is misleading for the RJMCMC 3 algorithm because the number of interior knots will depend on whether $h=0$ or $h=1$. As with the other examples in this chapter, we look at the MCSE as defined in Section 4.4.1. We estimate the MCSE for $\tau$ for the first simulated data set to be 0.0033 which is less than $5 \%$ of the sample standard deviation for $\tau$ (0.0270) suggesting that the algorithm has run long enough. The histogram for the posteriors draws (after discarding burn-in) for $\tau$ is given in Figure 5.7(d). We compute the posterior estimate for $\tau$ by averaging the posterior draws after discarding burn-in and find $\hat{\tau}=3.61$. The value of $\tau$ used to generate the data (for the scaled $y$-values) is 3.03 so we do overestimate $\tau$ in this case. However, the $95 \%$ HPD interval for $\tau$ is $(2.59,4.69)$ which includes 3.03 .

For the second data set where we increased the error standard deviation to 0.5 (data shown in Figure 5.8(a)), the posterior draws (after discarding burn-in) of the RJMCMC 3 algorithm for $h$ is shown in Figure 5.8(b). Even with the larger error standard deviation, the RJMCMC 3 algorithm still correctly selects the single change-point model for the majority of the runs. The partial trace plot for $\tau$ for the second data set with the error standard deviation 0.5 is given in Figure 5.8(c). We again look at the MCSE and find the estimate of the MCSE for $\tau$ for the


Figure 5.7: (a) Plot of the data set found using the single change-point example in Section 5.4.5 with the error standard deviation equal to 0.35 . (b) Histogram of the posterior draws for $h$ from the RJMCMC 3 algorithm to determine existence of change-point for data set in plot (a).(c) Partial trace plot for $\tau$ for the RJMCMC 3 algorithm for the single change-point data set with the error standard deviation equal to 0.35 . The grey solid line denotes the value of $\tau$ used to generated the data (for the scaled data). (d) Histogram of the posterior draws for $\tau$ for the RJMCMC 3 algorithm for the single change-point data set with the error standard deviation equal to 0.35 after discarding burn-in. The grey solid lines denotes the value of $\tau$ used to generate the data (for the scaled data) and the red dashed lines mark the $95 \%$ HPD interval as found in Section 2.6.1.
second simulated data set to be 0.0038 which is less than $5 \%$ of the sample standard deviation for $\tau$ (0.0176) suggesting that the algorithm has run long enough. The histogram for the posteriors draws (after discarding burn-in) for $\tau$ for the second data set is given in Figure 5.8(d) . We compute the posterior estimate for $\tau$ by averaging the posterior draws after discarding burn-in and find $\hat{\tau}=2.19$. The value of $\tau$ used to generate the data (for the scaled $y$-values) is 1.83 so we do overestimate $\tau$ in this case as well. However, the $95 \%$ HPD interval for $\tau$ is $(1.52,2.89)$ which includes 1.83.

Next, we consider how our algorithm performs when the curve is continuous so no change-point exists (Figure 5.9(a)). The histogram of posterior draws for $h$ after burn-in is shown in Figure 5.9(b). This algorithm selects the no change-points model


Figure 5.8: (a) Plot of the second data set found using the single change-point example in Section 5.4.5 with the error standard deviation equal to 0.5. (b) Histogram of the posterior draws for $h$ from the RJMCMC 3 algorithm to determine existence of change-point for data set in plot (a). (c) Partial trace plot for $\tau$ for the RJMCMC 3 algorithm for the single change-point data set with the error standard deviation equal to 0.5 . The grey solid line denotes the value of $\tau$ used to generated the data (for the scaled data). (d) Histogram of the posterior draws for $\tau$ for the RJMCMC 3 algorithm for the single change-point data set with the error standard deviation equal to 0.5 after discarding burn-in. The grey solid lines denotes the value of $\tau$ used to generate the data (for the scaled data) and the red dashed lines mark the $95 \%$ HPD interval as found in Section 2.6.1.
for the majority of the runs. The partial trace plot for $\tau$ for the no change-points data set is given in Figures 5.9(c). It suggests that the algorithm is mixing well. We find the estimate of the MCSE for $\tau$ for the third simulated data set to be 0.0016 which is less than $5 \%$ of the sample standard deviation for $\tau$ (0.0202) suggesting that the algorithm has run long enough. The histogram for the posteriors draws (after discarding burn-in) for $\tau$ for the third data set is given in Figure 5.9(d). We compute the posterior estimate for $\tau$ for the third data set and find $\hat{\tau}=2.78$. The value of $\tau$ used to generate the data (for the scaled $y$-values) is 3.16 so we under estimate $\tau$ in this case. Yet, the $95 \%$ HPD interval for $\tau$ is (2.00, 3.57) which still includes value for $\tau$ that was used to generate the data,3.16.


Figure 5.9: (a) Scatter plot of the no change-points data set. (b) Histogram of posterior draws for $h$ from the RJMCMC 3 algorithm for data set in plot (a). (c) Partial trace plot for $\tau$ for the RJMCMC 3 algorithm for the no change-points data set. The grey solid line denotes the value of $\tau$ used to generated the data (for the scaled data). (d) Histogram of the posterior draws for $\tau$ for the RJMCMC 3 algorithm for the no change-points data set after discarding burn-in. The grey solid lines denotes the value of $\tau$ used to generate the data (for the scaled data) and the red dashed lines mark the $95 \%$ HPD interval as found in Section 2.6.1.

## Chapter 6

## CONCLUSIONS AND FUTURE WORK

### 6.1 Conclusion

In this dissertation, we introduced a Bayesian semi-parametric model for shaperestricted regression which allows one to model the relationship between a response variable and covariate(s) without having to assume a parametric form for the regression function while incorporating knowledge about the shape of the regression function. The proposed Bayesian shape-restricted regression spline (Bayes SRRS) model uses the shape-restricted regressions splines discussed in Meyer (2008) to estimate the regression functions. By using a linear combination of quadratic $I$-splines (monotone functions) or cubic $C$-splines (convex or concave functions), we estimated the regression functions and the shape restriction was imposed simply by requiring the coefficients of the spline functions to be positive. We used a Bayesian framework which allows for several different types of inference. The Bayesian paradigm lent itself to the use of model selection tools such as Bayes factors and also provided joint posterior distributions of the parameters that were used to construct pointwise credible intervals and perform inference on categorical covariates.

A simulation study demonstrated that the proposed Bayes SRRS model performed well, in terms of mean square error, at estimating several different types regression functions. It demonstrated how the Bayes SRRS model estimated both functions with steep slopes and functions with "flat spots" well. The simulation study also showed that the inference on categorical covariates using the Bayes SRRS model performed as well as inference made in the frequentist paradigm for the majority of
cases in the simulation study. If the assumptions of linear regression model were violated, the Bayes SRRS model performed better at model selection than frequentist tests requiring the assumption of a linearity. Lastly, the simulation study demonstrated good coverage probabilities for the credible intervals created using the Bayes SRRS model. We also showed that the Bayes SRRS spline model produces consistent estimates for data generated from the normal errors model under the monotone shape restriction. We illustrated the usefulness of the Bayes SRRS model by using it to analyze two real data sets.

The Bayes SRRS spline model was extended to generalized linear mixed models. The proposed Bayes SRRS for generalized linear mixed models allows for function estimation for several different types of generalized linear mixed models without having to assume a parametric form for the regression function while still requiring the function to follow a given shape restriction. We illustrated how the Bayesian framework again allowed for several types of inference including model selection, credible intervals, and inference on categorical covariates. We provided a Markov chain Monte Carlo (MCMC) algorithm to estimate data generated from a random intercept model that did not require the same number of observations per group or individual and also did not require the observations per group or individual to be at the same values of the covariate(s) (unbalanced design). The application of the proposed Bayesian SRRS model for generalized linear mixed models to a simulated data set showed that the estimated regression functions were close to the function that generated them. We also demonstrated that a credible interval for a categorical covariate included the true value. We concluded with the application of the Bayes SRRS random intercept model to a real data set.

Next, we considered the Bayes SRRS model in Chapter 2 and utilized the Bayesian paradigm to extend it to a free-knot spline model. We proposed a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm to estimate monotone shape-restricted regression functions using shape-restricted regression splines
without fixing the number or locations of the interior knot points used to create the splines. Thus, we used the RJMCMC algorithm to average over several different models with different knot locations. Other RJMCMC algorithms for monotone shape-restricted functions use $B$-splines and require the coefficients to be ordered. The use of quadratic $I$-splines allowed us to enforce the shape restriction simply by requiring the coefficients to be positive and also allowed us to use gamma priors for the spline coefficients which simplified the RJMCMC algorithm. We applied the proposed Bayesian free-knot spline model to simulated examples and demonstrated that the algorithm mixes well as well as produced reasonable function estimates for both smooth and wiggly functions.

Lastly, we introduced a Bayes SRRS model for functions with change-points. We proposed a model that requires the function to be monotonically increasing over the range of the covariate except at change-points. We allowed for discontinuities in first derivative or jumps at the change-points by introducing additional basis functions to the Bayes SRRS spline model without change-points. We reparameterized the model such that we could enforce the monotonicity shape restriction by requiring the coefficients of some of the basis function to be positive. We then proposed three RJMCMC algorithms for function estimation for change-point models. The first RJMCMC algorithm, denoted the RJMCMC 1 algorithm, estimated shaperestricted functions where the number and location of change-points were known. It allowed for the number and location of interior knots between change-points to be random. The application of the RJMCMC 1 algorithm to simulated data sets showed that the RJMCMC 1 algorithm mixed well and provided reasonable function estimates. The second RJMCMC algorithm, denoted the RJMCMC 2 algorithm, found the location of a single change-point. It allowed the number and location of interior knots to be random. The application of the RJMCMC 2 algorithm to a simulated data set showed that it provided an estimate for the unknown change-point location that was very close to the true value. The third RJMCMC algorithm, denoted the

RJMCMC 3 algorithm, was used to determine the existence of a change-point. It moved between a model without change-points and a model with a change-point. The application of the RJMCMC 3 algorithm to simulated data sets showed that it selected the correct model. In all, we proposed a Bayesian shape-restricted regression spline model to model a regression function under shape restrictions that does not assume a parametric form. It produces reasonable function estimates and has several extensions including extensions to generalized linear mixed models and change-point models.

### 6.2 Future Work

There are several extensions to the work proposed in this dissertation. A simulation study examining the performance of the Bayes SRRS generalized linear mixed model would provide valuable information on the small sample behavior of this model and estimation procedures. Performing a simulation study examining the Bayes SRRS generalized linear mixed model applied to data other than the normal errors data such as binomial or Poisson data would also be of practical interest. Additionally, a simulation study comparing the free-knot spline model proposed in Chapter 4 to other free-knot spline models such as the one proposed in Shively et al. (2011) would provide information on the advantages or any disadvantages of using the proposed RJMCMC algorithm to estimate shape-restricted functions. Likewise, simulation studies on all of the proposed RJMCMC algorithms for estimating functions with change-points would provide information on the usefulness of these algorithms under various conditions.

It is also important to study the asymptotic behavior of the Bayes SRRS generalized linear mixed model. In particular, we can determine the conditions under which the regression function and variance component estimates are consistent. In addition to studying the asymptotic behavior of the Bayes SRRS generalized linear mixed model, we could also examine the asymptotic behavior for the free-knot

Bayesian SRRS model proposed in Chapter 4. We could again focus on showing the consistency of the regression function estimate. Allowing the number and locations of the interior knot to be free parameters greatly complicates the proof of consistency and we will likely want to extend the work of Shively et al. (2011) regarding the consistency of free-knot splines for smooth functions.

Furthermore, a RJMCMC algorithm that moves between models where the number of change-points ranges from 0 to $H$ where $H>1$ is an obvious extension to the RJMCMC algorithms for change-point models proposed in this dissertation. When $H>1$, we would want to use a prior for the change-point locations that avoids placing the change-points too close together. If we allow the change-points to be too close together then we can have problems distinguishing between models. For instance, a function with a jump of height $w$ at change-point $\xi_{1}$ can be modeled by both a model with one change-point and a jump of $w$ or two consecutive changepoints with jumps of $w / 2$. Another thing to consider when $H>1$, is how to find the locations of multiple change-points from the posterior distribution of change-point locations. We would need to determine the best way to find the multiple modes of the posterior distribution for the change-point locations.

## References

Akaike, H. (1974). A new look at the statistical model identification. Automatic Control, IEEE Transactions on, 19(6):716-723.

Alizadeh, F. and Goldfarb, D. (2003). Second-order cone programming. Mathematical Programming, 95:3-51.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3):175-185.

Alvarez, E. and Dey, D. (2009). Bayesian isotonic changepoint analysis. Annals of the Institute of Statistical Mathematics, 61:355-370.

Banerjee, M., Mukherjee, D., and Mishra, S. (2009). Semiparametric binary regression models under shape constraints with an application to Indian schooling data. Journal of Econometrics, 149(2):101-117.

Barr, D. R. and Sherrill, E. T. (1999). Mean and variance of truncated normal distributions. The American Statistician, 53(4):357-361.

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. Journal of the American Statistical Association, 97(457):160-169.

Biller, C. (2000). Adaptive Bayesian regression splines in semiparametric generalized linear models. Journal of Computational and Graphical Statistics, 9(1):122140.

Bornkamp, B. and Ickstadt, K. (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose-response analysis. Biometrics, 65(1):198-205.

Bowman, A. W., Jones, M., and Gijbels, I. (1998). Testing monotonicity of regression. Journal of Computational and Graphical Statistics, 7(4):489-500.

Box, G. E. P. and Tiao, G. C. (1968). Bayesian estimation of mean for the random effect model. Journal of the American Statistical Association, 63(321):174-181.

Braun, J. V. and Muller, H.-G. (1998). Statistical methods for DNA sequence segmentation. Statistical Science, 13(2):142-162.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. Journal of the American Statistical Association, 88(421):925.

Brezger, A. and Steiner, W. J. (2008). Monotonic regression based on Bayesian P-splines. Journal of Business and Economic Statistics, 26(1):90-104.

Cai, B. and Dunson, D. B. (2007). Bayesian multivariate isotonic regression splines. Journal of the American Statistical Association, 102(480):1158-1171.

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. Journal of the Royal Statistical Society. Series C, 41(2):389-405.

Casella, G. and Berger, R. L. (2002). Statistical Inference. Duxbury, United States, second edition.

Chen, M.-H., i Man Shao, Q., and Ibrahim, J. G. (2000). Monte Carlo methods in Bayesian Computation. Springer-Verlag, New York.

Chib, S. (1998). Estimation and comparison of multiple change-point models. Journal of Econometrics, 86(2):221-241.

Curry, H. and Schoenberg, I. (1966). On Pólya frequency functions iv: The fundamental spline functions and their limits. Journal d'Analyse Mathématique, 17:71-107.

Curtis, S. M. and Ghosh, S. K. (2011). A variable selection approach to monotonic regression with Bernstein polynomials. Journal of Applied Statistics, 38:961-976.
de Boor, C. (2001). A Practical Guide to Splines. Springer, New York, revised edition.

Delecroix, M., Simioni, M., and Thomas-Agnan, C. (1995). A shape constrained smoother: simulation study. Computational Statistics, 10(2):155-175.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 39(1):1-38.

Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998). Automatic Bayesian curve fitting. Journal of the Royal Statistical Society, Series B, 60(2):333-350.

Dierckx, P. (1993). Curve and Surface Fitting with Splines. Oxford University Press, New York.

Dilleen, M., Heimann, G., and Hirsch, I. (2003). Non-parametric estimators of a monotonic dose-response curve and bootstrap confidence intervals. Statistics in Medicine, 22(6):869-882.

DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. Biometrika, 88(4):1055-1071.

Dunson, D. B. (2005). Bayesian semiparametric isotonic regression for count data. Journal of the American Statistical Association, 100(470):618-627.

Eubank, R. L. (1988). Spline Smoothing and Nonparametric Regression. Dekker, New York.

Eubank, R. L. and Speckman, P. L. (1994). Nonparametric estimation of functions with jump discontinuities. Lecture Notes-Monograph Series, 23:130-144.

Fearnhead, P. (2006). Exact and efficientBayesian inference for multiple changepoint problems. Statistics and Computing, 16:203-213.

Fraser, D. A. S. and Massam, H. (1989). A mixed primal-dual bases algorithm for regression under inequality constraints. Application to concave regression. Scandinavian Journal of Statistics, 16(1):65-74.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parametrisations for normal linear mixed models. Biometrika, 82(3):479-488.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1996). Efficient parameterizations for generalized linear mixed models. Bayesian Statistics, 5:165-180.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, 1(3):515-533.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6):721-741.

Ghosh, D. (2007). Incorporating monotonicity into the evaluation of a biomarker. Biostatistics, 8(2):402-413.

Gijbels, I. and Goderniaux, A.-C. (2004). Bandwidth selection for changepoint estimation in nonparametric regression. Technometrics, 46(1).

Gijbels, I., Hall, P., and Kneip, A. (1999). On the estimation of jump points in smooth curves. Annals of the Institute of Statistical Mathematics, 51:231-251.

Gijbels, I., Lambert, A., and Qiu, P. (2007). Jump-preserving regression and smoothing using local linear fitting: A compromise. Annals of the Institute of Statistical Mathematics, 59:235-272.

Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995). Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics, 51(4):1440-1450.

Giordani, P. and Kohn, R. (2008). Efficient Bayesian inference for multiple change-point and mixture innovation models. Journal of Business $8 \mathcal{S}$ Economic Statistics, 26(1):66-77.

Givens, G. H. and Hoeting, J. A. (2005). Computational Statistics. Wiley, Hoboken, New Jersey, USA.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 82(4):711-732.

Hall, P. and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. The Annals of Statistics, 29(3):624-647.

Hazelton, M. L. and Turlach, B. A. (2011). Semiparametric regression with shape-constrained penalized splines. Computational Statistics and Data Analysis, 55:2871-2879.

Holmes, C. C. and Heard, N. A. (2003). Generalized monotonic regression using random change points. Statistics in Medicine, 22(4):623-638.

Holmes, C. C. and Mallick, B. K. (2003). Generalized nonlinear modeling with multivariate free-knot regression splines. Journal of the American Statistical Association, 98(462):352-368.

Horn, R. A. (1990). The Hadamard product. In Johnson, C. R., editor, Matrix Theory and Applications. American Mathematical Society.

Horn, R. A. and Johnson, C. S. (1991). Topics in Matrix Analysis. Cambridge University Press, New York, USA.

Huang, J. Z. and Stone, C. J. (2002). Extended linear modeling with splines. In Dension, D. D., Hansen, M. H., Holmes, C. C., Malick, B., and Yu, B., editors, Nonlinear Estimation and Classification, pages 213-234. Springer, New York.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. Biometrika, 76(2):297-307.

Jiang, J. (2007). Linear and Generalized Linear Mixed Models and Their Applications. Springer, New York.

Johnson, L. W., Riess, R. D., and Arnold, J. T. (2002). Introduction to Linear Algebra. Addison Wesley, San Francisco,CA, fifth edition.

Johnson, M. S. (2007). Modeling dichotomous item responses with free-knot splines. Computational Statistics \& Data Analysis, 51:4178-4192.

Kaslow, R. A., G.Ostrow, D., Detels, R., Phair, J. P., Polk, B. F., and Jr., C. R. R. (1987). The multicenter AIDS cohort study: Rationale, organization, and selected characteristics of the participants. American Journal of Epidemiology, 126(2):310-318.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. Journal of American Statistical Association, 90(430):773-795.

Kelly, C. and Rice, J. (1990). Monotone smoothing with application to doseresponse curves and the assessment of synergism. Biometrics, 46(4).

Koo, J.-Y. (1997). Spline estimation of discontinuous regression functions. Journal of Computational and Graphical Statistics, 6(3):266-284.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. Journal of Computational and Graphical Statistics, 13(1):183-212.

Lavine, M. and Mockus, A. (1995). A nonparametric Bayes method for isotonic regression. Journal of Statistical Planning and Inference, 46(2):235-248.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 61(2):381-400.

Lindstrom, M. J. (2002). Bayesian estimation of free-knot splines using reversible jumps. Computational Statistics \& Data Analysis, 41(2):255-269.

Ma, S. and Yang, L. (2010). A jump-detecting procedure based on spline estimation. Journal of Nonparametric Statitics. In press.

Mammen, E. (1991). Estimating a smooth monotone regression function. The Annals of Statistics, 19(2):724-740.

Mammen, E., Marron, J. S., Turlach, B. A., and Wand, M. P. (2001). A general projection framework for constrained smoothing. Statistical Science, 16(3):232248.

Mammen, E. and Thomas-Agnan, C. (1999). Smoothing splines and shape restrictions. Scandinavian Journal of Statistics, 26(2):239-252.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. Journal of the American Statistical Association, 92(437):162-170.

McCulloch, C. E. and Searle, S. R. (2001). Generalized, Linear, and Mixed Models. John Wiley \& Sons, Inc., New York, NY, first edition.

Meyer, M. C. (1996). Shape restriced inference with applications to nonparametric regression, smooth nonparametric function estimation, and density estimation. PhD thesis, University of Michigan.

Meyer, M. C. (1999). An extension of the mixed primal-dual bases algorithm to the case of more constraints than dimensions. Journal of Statisical Planning and Inference, (81):13-31.

Meyer, M. C. (2008). Inference using shape-restricted regression splines. Annals of Applied Statistics, 2(3):1013-1033.

Meyer, M. C., Hackstadt, A. J., and Hoeting, J. A. (2011). Bayesian estimation and infereence for generalised partial linear models using shape-restricted splines. Journal of Nonparametric Statisics. (to appear).

Meyer, M. C. and Laud, P. W. (2002). Predictive variable selection in generalized linear models. Journal of the American Statistical Association, 97(459):859-871.

Muller, H.-G. (1992). Change-points in nonparametric regression analysis. The Annals of Statistics, 20(2):737-761.

Muller, H.-G. and Song, K.-S. (1997). Two-stage change-point estimators in smooth regression models. Statistics \& Probability Letters, 34(4):323-335.

Muller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. Statistical Science, 19(1):95-110.

Namata, H., Shkedy, Z., Faes, C., Aerts, M., Molenberghs, G., Theeten, H., Van Damme, P., and Beutels, P. (2007). Estimation of the force of infection from current status data using generalized linear mixed models. Journal of Applied Statistics, 34(8):923-939.

Natarajan, R. and Kass, R. (2000). Reference Bayesian methods for generalized linear mixed models. Journal of the American Statistical Association, 95(449):227-237.

Neelon, B. and Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. Biometrics, 60(2):398-406.

Newton, M. A. and Raftery, A. E. (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. Journal of Royal Statistical Society, Serie B, 56(1):3-48.

Perron, F. and Mengersen, K. (2001). Bayesian nonparametric modeling using mixtures of triangular distributions. Biometrics, 57(2):518-528.

Qiu, P. and Yandell, B. (1998). A local polynomial jump-detection algorithm in nonparametric regression. Technometrics, 40(2):141-152.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raftery, A. E. (1993). Change point and change curve modeling in stochastic processes and spatial statistics. Journal of Applied Statistical Science, 1(4):403423.

Raftery, A. E., Newton, M. A., Satagopan, J. M., and Krivitsky, P. N. (2007). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. Bayesian Statistics, 8:1-45.

Ramgopal, P., Laud, P. W., and Smith, A. F. M. (1993). Nonparametric Bayesian bioassay with prior constraints on the shape of the potency curve. Biometrika, 80(3):489-498.

Ramsay, J. O. (1988). Monotone regression splines in action. Statistical Science, $3(4): 425-461$.

Ratkowsky, D. A. (1983). Nonlinear Regression Modeling: A Unified Practical Approach. Marcel Dekker, New York.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics, 57(1):253-259.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60(2):731-792.

Robertson, T. and Cryer, J. D. (1974). An iterative procedure for estimating the mode. Journal of the American Statistical Association, 69(348):1012-1016.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). Order Restricted Statistical Inference. Wiley, New York.

Robey, B., Shea, M. A., Rutstein, O., and Morris, L. (1992). The reproductive revolution: New survey findings. Technical Report M-11, Population Reports.

Rotondi, R. (2002). On the influence of the proposal distributions on a reversible jump MCMC algorithm applied to the detection of multiple change-points. Computational Statistics 83 Data Analysis, 40(3):633-653.

Ruppert, D., Wand, M. P., and Carrol, R. J. (2003). Semiparametric Regression. Cambridge University Press, New York.

Schipper, M., Taylor, J. M. G., and Lin, X. (2007). Bayesian generalized monotonic functional mixed models for the effects of radiation dose histograms on normal tissue complications. Statisics in Medicine, 26:4643-4656.

Schumaker, L. L. (2007). Spline Functions: Basic Theory. Cambridge University Press, New York, third edition.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2):461-464.

Shively, T. S., Sager, T. W., and Walker, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71(1):159-175.

Shively, T. S., Walker, S. G., and Damien, P. (2011). Nonparametric function estimation subject to monotonicity, convexity and other shape constraints. Journal of Econometrics, 161(2):166-181.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting. Journal of the Royal Statistical Society. Series B (Methodological), 47(1):1-52.

Sisson, S. A. (2005). Transdimensional Markov chains: A decade of progress and future perspectives. Journal of American Statistical Association, 100(471):10771089.

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. Journal of Econometrics, 75(2):317-343.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Linde, A. v. d. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 64(4):583-639.

Takeuchi, K., Yanai, H., and Mukherjee, B. N. (1982). The Foundations of Multivariate Analysis. John Wiley \& Sons, Inc., New York.

Thompson, W. K. and Rosen, O. (2008). A Bayesian model for sparse functional data. Biometrics, 64(1):54-63.

Thomson, J. R., Kimmerer, W. J., Brown, L. R., Newman, K. B., Nally, R. M., Bennett, W. A., Feyrer, F., and Fleishman, E. (2010). Bayesian change point analysis of abundance trends for pelagic fishes in the upper San Francisco Estuary. Ecological Applications, 20(5):1431-1448.

Tiao, G. C. and Tan, W. Y. (1965). Bayesian analysis of random-effect models in the analysis of variance. I. Posterior distribution of variance-components. Biometrika, 52(1/2):37-53.

Tierney, L. (1994). Markov chains for exploring posterior distributions. The Annals of Statistics, 22(4).

Turlach, B. (2005). Shape constrained smoothing using smoothing splines. Computational Statistics, 20:81-104.

Tutz, G. and Leitenstorfer, F. (2007). Generalized smooth monotonic regression in additive modeling. Journal of Computational and Graphical Statistics, 16(1):165188.

Verbyla, A. P., Cullis, B. R., Kenward, M. G., and Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. Journal of the Royal Statistical Society. Series C (Applied Statistics), 48(3):269311.

Vines, S. K., Gilks, W. R., and Wild, P. (1996). Fitting Bayesian multiple random effects models. Statistics and Computing, 6:337-346.

Walker, S. (1996). An EM algorithm for nonlinear random effects models. Biometrics, 52(3):934-944.

Wand, M. P. and Jones, M. C. (1995). Kernel smoothing. Chapman \& Hall, New York.

Wang, X. (2008). Bayesian free-knot monotone cubic spline regression. Journal of Computational and Graphical Statistics, 17(2):373-387.

Wang, X. and Li, F. (2008). Isotonic smoothing spline regression. Journal of Computational and Graphical Statistics, 17(1):21-37.

Willems, J. P., Saunders, J. T., Hunt, D. E., and Schorling, J. (1997). Prevalence of coronary heart disease risk factors among rural blacks: A community-based study. Southern Medical Journal, 90:814-820.

Wolfe, D. A. and Schechtman, E. (1984). Nonparametric statistical procedures for the changepoint problem. Journal of Statistical Planning and Inference, 9(3):389 - 396.

Wu, J. S. and Chu, C. K. (1993). Kernel-type estimators of jump points and values of a regression function. The Annals of Statistics, 21(3):1545-1566.

Yatchew, A. and Hardle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. Journal of Econometrics, 133(2):579-599.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics, 50(3):689-699.

Zhang, D. (2004). Generalized linear mixed models with varying coefficients for longitudinal data. Biometrics, 60(1):8-15.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. Journal of the American Statistical Association, 93(442):710-719.

Zhao, L. H. (2000). Bayesian aspects of some nonparametric problems. The Annals of Statistics, 28(2):532-552.

Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. Journal of the American Statistical Association, 96(453):247-259.

