DISSERTATION

CHANGE-POINT ESTIMATION USING SHAPE-RESTRICTED REGRESSION SPLINES

Submitted by

Xiyue Liao

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2016

Doctoral Committee:

Advisor: Mary C. Meyer

F. Jay Breidt
Darren Homrighausen
Belfiori Elisa

ABSTRACT

CHANGE-POINT ESTIMATION USING SHAPE-RESTRICTED REGRESSION SPLINES

Change-Point estimation is in need in fields like climate change, signal processing, economics, dose-response analysis etc, but it has not yet been fully discussed. We consider estimating a regression function f_m and a change-point m, where m is a mode, an inflection point, or a jump point. Linear inequality constraints are used with spline regression functions to estimate m and f_m simultaneously using profile methods. For a given m, the maximum-likelihood estimate of f_m is found using constrained regression methods, then the set of possible change-points is searched to find the $\hat{m}$ that maximizes the likelihood. Convergence rates are obtained for each type of change-point estimator, and we show an oracle property, that the convergence rate of the regression function estimator is as if m were known. Parametrically modeled covariates are easily incorporated in the model. Simulations show that for small and moderate sample sizes, these methods compare well to existing methods. The scenario when the random error is from a stationary autoregressive process is also presented. Under such a scenario, the change-point and parameters of the stationary autoregressive process, such as autoregressive coefficients and the model variance, are estimated together via Cochran-Orcutt-type iterations. Simulations are conducted and it is shown that the change-point estimator performs well in terms of choosing the right order of the autoregressive process. Penalized spline-based regression is also discussed as an extension. Given a large number of knots and a penalty parameter which controls the effective degrees of freedom of a shape-restricted model, penalized methods give smoother fits while balance between under- and over-fitting. A bootstrap confidence interval for a change-point is established. By generating random change-points from a curve on the unit interval, we compute the coverage rate of the bootstrap confidence interval using penalized estimators, which shows advantages such as robustness over competitors. The methods are available in the `R` package `ShapeChange` on the Comprehensive R Archival Network (CRAN).

Moreover, we discuss the shape selection problem when there are more than one possible shapes for a given data set. A project with the Forest Inventory & Analysis (FIA) scientists is included as an example. In this project, we apply shape-restricted spline-based estimators, among which the one-jump and double-jump estimators are emphasized, to time-series Landsat imagery for the purpose of modeling, mapping, and monitoring annual forest disturbance dynamics. For each pixel and spectral band or index of choice in temporal Landsat data, our method delivers a smoothed rendition of the trajectory constrained to behave in an ecologically sensible manner, reflecting one of seven possible "shapes". Routines to realize the methodology are built in the `R` package `ShapeSelectForest` on CRAN, and techniques in this package are being applied for forest disturbance and attribute mapping across the conterminous U.S.. The Landsat community will implement techniques in this package on the Google Earth Engine in 2016.

Finally, we consider the change-point estimation with generalized linear models. Such work can be applied to dose-response analysis, when the effect of a drug increases as the dose increases to a saturation point, after which the effect starts decreasing.

DEDICATION

To the cutest, Francis, Nilus and Tanz

.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1. Cone Projection Review

The projection of $\boldsymbol{Y} \in \mathbb{R}^n$ onto a set $C \subseteq \mathbb{R}^n$ is defined as the point $\hat{\boldsymbol{\theta}} \in C$ that minimizes the Euclidean distance

$$\| \boldsymbol{Y} - \boldsymbol{\theta} \|^2 = \sum_{i=1}^{n} (Y_i - \theta_i)^2.$$

A unique minimum exists if C is closed and convex. We are concerned with projecting onto convex polyhedral cones such as

$$C = \{\boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{A\theta} \geq \boldsymbol{0}\}, \tag{1}$$

for an $m \times n$ constraint matrix $\boldsymbol{A}$. The set C is a cone because given $\boldsymbol{\theta} \in C$, we have $\alpha\boldsymbol{\theta} \in C$, for all non-negative real numbers α, and it is straightforward to verify that C is convex. We require that $\boldsymbol{A}$ be "irreducible" as defined by Meyer (1999); the intuitive meaning is "non-redundant." The term "polyhedral" means finitely generated, so that points in the cone can be characterized as linear combinations of a finite set of points (generators) where the coefficients of the linear combination are non-negative.

The `coneproj` package contains routines for cone projection and quadratic programming, plus applications in estimation and inference for shape-restricted regression. For the `coneA` routine, the user specifies $\boldsymbol{Y} \in \mathbb{R}^n$, an $m \times n$ matrix $\boldsymbol{A}$, and an optional weight vector $\boldsymbol{w}$ with positive elements. The routine returns $\hat{\boldsymbol{\theta}}$ to minimize

$$\sum_{i=1}^{n} w_i (Y_i - \theta_i)^2$$

over $\boldsymbol{\theta} \in C$, is returned, where C is defined in (1). The matrix $\boldsymbol{A}$ is required to be irreducible, that is, the rows of $\boldsymbol{A}$ form an irreducible set. A set of vectors is irreducible if none can be written as a positive linear combination of two or more of the others, and the origin is not a positive linear combination of two or more vectors in the set. (The phrase "positive linear combination" means a linear combination with positive coefficients.)

Let V be the null space of $\boldsymbol{A}$; that is, the linear space orthogonal to the space spanned by the rows of $\boldsymbol{A}$. The space V is contained in C. An element in C can be written as the sum of a vector in V and a linear combination of the *edges* or *generators* of C with non-negative coefficients. If $\boldsymbol{A}$ is full row-rank, it is shown in Meyer (2013) that the edges $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_m$ of the cone are the columns of $\boldsymbol{B} = \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}$. If $\boldsymbol{A}$ is not full row-rank, Proposition 1 of Meyer (1999) can be used to obtain the edges $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_M$, where $M \geq m$. The cone (1) can alternatively be written as

$$(2) \qquad C = \left\{ \boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{\theta} = \boldsymbol{v} + \sum_{j=1}^{M} b_j \boldsymbol{\eta}_j, \quad \boldsymbol{v} \in V \ \text{ and } \ b_1, \ldots, b_M \geq 0 \right\},$$

where M is the number of generators and $M = m$ if $\boldsymbol{A}$ is full row rank. The generators $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_M$ are orthogonal to V, so the projection of $\boldsymbol{Y}$ onto C is the sum of the projections onto V and onto the cone

$$\Omega = \left\{ \boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{\theta} = \sum_{j=1}^{M} b_j \boldsymbol{\eta}_j, \quad b_1, \ldots, b_M \geq 0 \right\}.$$

The algorithm of Meyer (2013) provides the projection onto Ω by determining the *face* of the cone on which the projection lands. The faces are used for the inference methods and are indexed by subsets of $\{1,\ldots,M\}$. For such a subset J, the corresponding face is

$$F_J = \left\{ \boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{\theta} = \sum_{j \in J} b_j \boldsymbol{\eta}_j, \;\; b_j > 0 \;\text{ for }\; j \in J \right\}.$$

The faces cover the cone; once the face containing the projection is determined, the projection onto Ω is simply the projection onto the linear space spanned by the edges making up the face. For more details and proofs, see Meyer (1999). The algorithm finds the projection by determining the set J.

The initial guess J_0 can be any subset of $\{1,\ldots,M\}$ for which the corresponding $\boldsymbol{\eta}^j$, $j \in J$, form a linearly independent set. At the kth iteration,

(1) Project $\boldsymbol{Y}$ onto the linear space spanned by $\{\boldsymbol{\eta}^j, j \in J_k\}$, to get $\boldsymbol{\theta}^{(k)} = \sum_{j\in J_k} b_j^{(k)} \boldsymbol{\eta}_j$.

(2) Check to see if all $b_j^{(k)}$ are non-negative:

- If yes, go to step 3.
- If no, choose j for which $b_j^{(k)}$ is minimized, and remove it from J; go to step 1.

(3) Compute $\langle \boldsymbol{Y} - \boldsymbol{\theta}^k, \boldsymbol{\eta}^j \rangle$ for each $j \notin J_k$. If these are all non-positive, then stop. If not, choose j for which this inner product is largest, add it to the set, and go to step 1.

See Meyer (2013) for the proof of convergence.

The *polar cone* is defined as

$$\Omega^o = \{\boldsymbol{\rho} : \langle \boldsymbol{\theta}, \boldsymbol{\rho} \rangle \leq 0, \; \forall \, \boldsymbol{\theta} \in C\},$$

and it can be shown that the projection $\hat{\boldsymbol{\rho}}$ of $\boldsymbol{Y}$ onto Ω^o is $\boldsymbol{Y} - \hat{\boldsymbol{\theta}}$, i.e., the residual of the projection onto C.

The constraint cone edges $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_M$ are not needed if $\boldsymbol{A}$ is provided, because the rows of $-\boldsymbol{A}$ are the edges of the polar cone. See Meyer (1999) for proof. The function `coneA` requires the specification of the constraint matrix (and hence the polar cone edges), while the function `coneB` requires the user to specify the cone edges $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_M$ and a basis for the linear space V that is contained in the cone. When there is no linear space in the cone, the user need only provide $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_M$. In either case, the function returns the projection, the dimension of the face of the cone on which the projection lands (which may be used as a surrogate degrees of freedom of the model), and the number of iterations. It also returns a message concerning convergence when the algorithm does not converge; although theoretically the algorithm must converge, the presence of rounding error in the real world results in a small possibility of non-convergence.

The function `coneA` will return the projection given the vector $\boldsymbol{Y} \in \mathbb{R}^n$ and the $m \times n$ matrix $\boldsymbol{A}$. The function `coneB` will return the projection given $\boldsymbol{Y}$ and the generators of C. In addition, if a positive weight vector $\boldsymbol{w}$ is provided, the functions return the minimizer of $\sum_{i=1}^n w_i(Y_i - \theta_i)^2$ over C.

The `coneA` and `coneB` algorithms have been coded and compiled in `C++`, and are called by `R`, which makes them considerably faster than if coded completely in `R`. They are the core algorithms of the package `coneproj`, which is now available on the Comprehensive R Archive Network (CRAN) at `https://cran.r-project.org/package=coneproj`. A paper about this package was published in the *Journal of Statistical Software* in 2014. The methodology proposed in this dissertation is developed based on `coneA` and `coneB`.

1.2. Regression Spline Review

B-splines are a standard choice for regression basis functions because of their interpolation property and flexibility. Suppose that $\boldsymbol{\eta}_q(\boldsymbol{x})$ is a B-spline basis function of order $q, q > 1$ and $t_j, j = 1, \ldots, k$, are the "knots" such that $t_1 < \ldots < t_k$, $t_1 \leq min(\boldsymbol{x})$, and $t_k \geq \max(\boldsymbol{x})$, then it is piecewise polynomial in $\boldsymbol{x}$, and the expression of the polynomial pieces on $[t_i, t_{i+1}]$ can be derived by de Boor's algorithm recursively as

$$\eta_{i,1}(\boldsymbol{x}) = \begin{cases} 1 \text{ if } t_i \leq \boldsymbol{x} < t_{i+1} \\ \\ 0 \text{ otherwise,} \end{cases}$$

and

$$\eta_{i,q}(\boldsymbol{x}) = \frac{\boldsymbol{x} - t_i}{t_{i+q-1} - t_i}\eta_{i,q-1}(\boldsymbol{x}) + \frac{t_{i+q} - \boldsymbol{x}}{t_{i+q} - t_{i+1}}\eta_{i+1,q-1}(\boldsymbol{x}).$$

The splines are defined so that each basis vector is orthogonal to all but a few of the others. B-spline properties are thoroughly discussed in de Boor (1978).

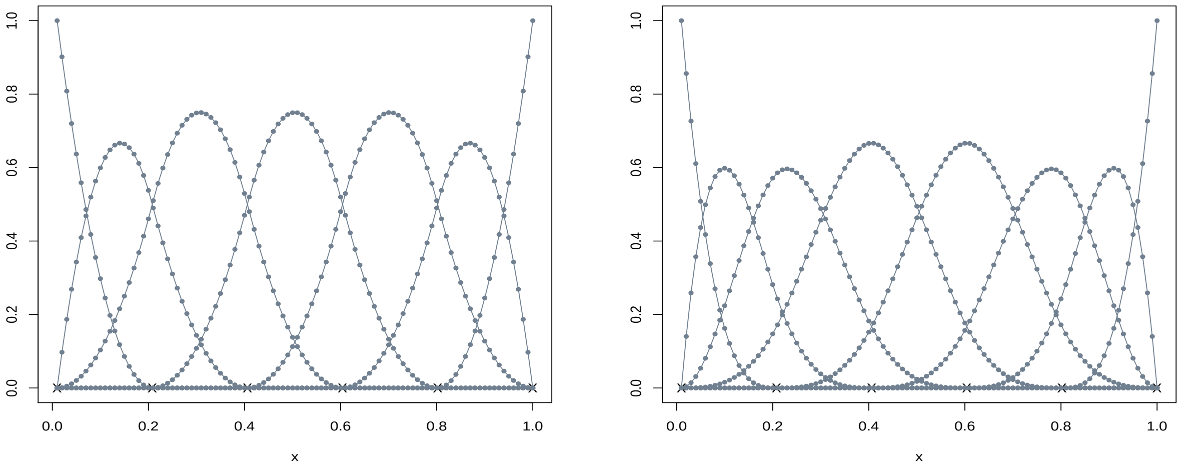

FIGURE 1. B-spline basis functions for a data set with $n = 100$ observations with values marked as dots. Knots are marked as "X". Left: piecewise quadratic. Right: piecewise cubic.

The derivative of a B-spline of order q is a function of B-splines of order $q-1$, i.e.,

$$\frac{d\eta_{i,q}(\boldsymbol{x})}{d\boldsymbol{x}} = (q-1)\Big[\frac{-\eta_{i+1,q-1}(\boldsymbol{x})}{t_{i+q}-t_{i+1}} + \frac{\eta_{i,q-1}(\boldsymbol{x})}{t_{i+q-1}-t_i}\Big].$$

Since the first (second) derivative of quadratic (cubic) B-splines are piecewise linear, it is straight-forward to impose a monotonicity or convexity constraint on an underlying curve using the derivative of B-splines to get a shape-restricted regression.

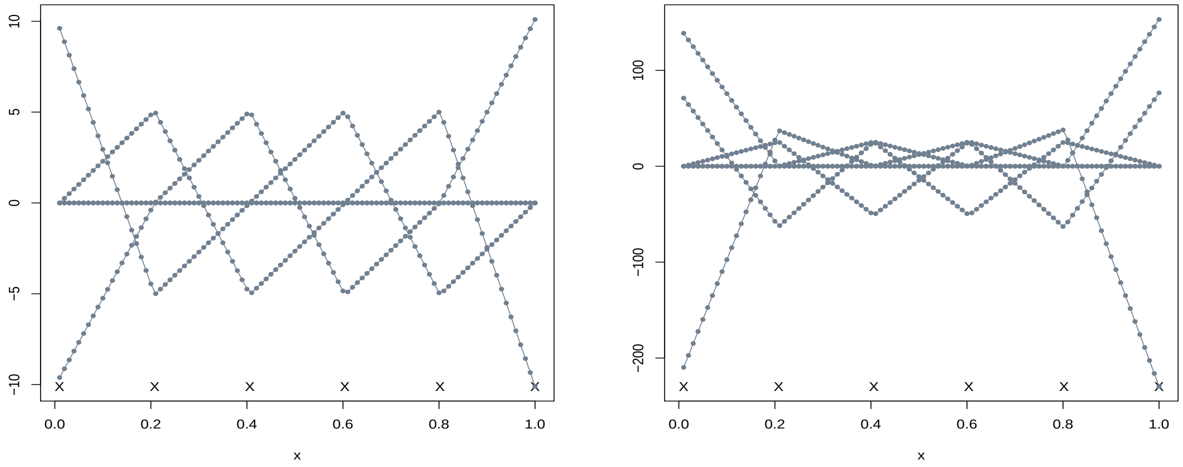

FIGURE 2. Derivatives of B-spline basis functions for a data set with $n = 100$ observations with values marked as dots. Knots are marked as "X". Left: first derivatives of quadratic B-splines. Right: second derivatives of cubic B-splines.

Using quadratic and cubic B-splines, we will develop the change-point estimators with shape constraints in the following chapters.

CHAPTER 2

Change-Point Estimation Methods

Consider the regression model

$$Y_i = f_m(X_i) + \sigma\varepsilon_i, \quad \text{for} \quad i = 1, \ldots, n, \tag{3}$$

where $f_m(x)$ is a function mapping $[0, 1]$ to $\mathbb{R}$ with a change-point m, $0 < m < 1$. Let $\sigma > 0$ and assume (for now) that the errors are independent with mean zero and unit variance. We consider three types of change-points: the mode of an increasing-decreasing (unimodal) regression function, the inflection point of a convex-concave regression function, and a jump point in an otherwise smooth regression function.

2.1. Literature Review

Nonparametric estimation of a unimodal regression function is closely relevant to the estimation of a unimodal density function. A kernel estimator of the mode of a density f is given by Eddy (1980). The estimator is consistent and asymptotically normal. The rate at which the mean squared error of the estimator converges to zero can be decreased from $n^{-4/7}$ to $n^{-1+\varepsilon}$ for any positive ε with sufficient conditions. Meyer (2012b) proposed a smoothed shape-restricted estimator for a unimodal density. The estimator of f is a linear combination of regression quadratic B-splines and it is obtained by a weighted projection onto a convex cone. The least-squares criterion proposed by Greoneboom, Jongbloed, and Wellner (2001) is minimized over the set of linear combinations of basis functions with the coefficients constrained to capture the shape assumption and the convergence rate for $\hat{f}$ and $\hat{m}$ is $n^{-3/7}$ for both the known mode case and unknown mode case.

Shoung and Zhang (2001) provide a nonparametric least squares estimator of the mode m of a unimodal regression function f_m using unsmoothed isotonic regression. When the mode m is known, the estimate of f_m uses the pooled adjacent violators algorithm (PAVA) on each side of m; when m is unknown, the least-squares estimator $\hat{m}$ is that which minimizes the sum of squared residuals for the fitted regression function over $m = x_i$. This function estimator tends to have a "spike" at the largest observation near the true mode, and hence the mode estimator tends to have a larger variance than that for a smoothed function estimator. The convergence rate of $\hat{m}$ satisfies that $\lim_{n\to\infty} sup\{n/(log(n))^{2\gamma}\}^{1/(2s+1)}|\hat{m} - m| < \infty$ *a.s.*. Köllmann, Bornkamp, and Ickstadt (2014) discuss penalized spline methods to achieve unimodal and smooth estimation of a functional relationship in a scenario of dose-response analysis. They use a restricted maximum likelihood approach to choose the tuning parameter and to estimate B-spline coefficients. For choices of m, they choose the tuning parameter and B-spline coefficients that minimize the residual sum of squares; an alternative Bayesian approach is also provided to estimate m instead of searching for m over a grid of possible values. Their methods are in the `R` package `uniReg` Köllmann (2014).

Inflection-Point detection is a challenging problem that is important in fields like signal processing and economics. Kachouie and Schwartzman (2013) propose an estimator using local polynomial regression to detect a single inflection point in an underlying smooth signal curve. To ensure that only one inflection point is detected, a constrained method is proposed for bandwidth selection. Two methods in Christopoulos (2014) are available in the `R` package `inflection` Christopoulos (2013) for identifying the inflection point in a convex-concave curve. They use a generalization of bisection method in root finding without any regression or splines representation.

In the jump-point case, all the well established estimators have the weak convergence property. Müller (1992) provides a kernel estimator for a jump point in an otherwise smooth regression model which is based on maximizing the difference between one-sided kernel smoothers. The estimator is asymptotically normal with a convergence rate exceeding $n^{-1/2}$ in most cases. Under some conditions, the rate can be arbitrarily close to $n^{-1}[log(n)]^{1/2}$. In Loader (1996), a jump-point estimator is the design point which maximizes the difference between the right and left limits at the point based on one-sided nonparametric local polynomial regression of degree p, where $p = 0$ or 1. Grégoire and Hamrouni (2001) provides a local linear regression estimator of a jump point in a regression function which maximizes the jump size. The convergence rate of the estimators in Loader (1996) and Grégoire and Hamrouni (2001) is n^{-1}. Horváth and Kokoszka (2002) provides a test statistic for a discontinuity point in a regression function f or its pth derivative $f^{(p)}$ by fitting local polynomials from the left and right. They also show that the estimator is weakly consistent with a convergence rate of h such that $\lim_{n\to\infty} n^{1/2}h^{p+1/2}(log(1/h))^{1/2} = 0$.

In this chapter, we propose a more computationally straight-forward spline-based nonparametric estimator of the regression function. In Section 2.2, the sets of spline basis functions and constraints are specified for each type of change-point, and the cone-projection algorithm is formulated. In Chapter 3, the convergence rates are established. We also show how to include covariates in the model, and the methods are adapted to the case of correlated errors. In Chapter 4, simulation results show that our estimators perform well when compared to some established estimators, and the methods are demonstrated with “real-data” examples.

Routines to realize the estimation methods discussed in this chapter are built in the `R` package `ShapeChange` Liao and Meyer (2016), and it is open to the public on the Comprehensive R Archival Network (CRAN).

2.2. Regression Spline Estimators

We first consider m to be known, and discuss spline estimation of $f_m(x)$. For each type of change-point, we use a constrained linear combination of B-spline basis functions to estimate the regression function. The basis functions $\boldsymbol{\eta}_1(\boldsymbol{x}), \ldots, \boldsymbol{\eta}_\ell(\boldsymbol{x})$ are determined by the degree of the splines and the knots $0 = t_1 < \cdots < t_k = 1$; for details, see de Boor (2001). For design points $x_1, \ldots, x_n$ in $[0, 1]$, we define $\boldsymbol{\theta} \in \mathbb{R}^n$ as $\theta_i = f_m(x_i)$, $i = 1, \ldots, n$, and similarly define basis vectors $\boldsymbol{\eta}_j \in \mathbb{R}^n$, $j = 1, \ldots, \ell$, where $\eta_{ji} = \eta_j(x_i)$, $i = 1, \ldots, n$. We define the $n \times \ell$ matrix $\boldsymbol{B}$ to have the basis vectors as columns. Then $\boldsymbol{\theta}$ is approximated by $\boldsymbol{Bb}$, where $\boldsymbol{b} \in \mathbb{R}^\ell$ is the coefficient vector. Inequality constraints of the form $\boldsymbol{Sb} \geq \boldsymbol{0}$ will constrain the shape appropriately for each case, and the estimation is accomplished with a quadratic programming routine in `coneproj` Meyer and Liao (2014). For the mode estimation and for the inflection-point estimation, we formulate the solution with the same notation, so that we can derive results that are valid for both cases.

According to Huang (2001), we define $\mathscr{H}$ as a Hilbert space and the norm $\|\cdot\|$ for $\mathscr{H}$ is chosen to be the L_2 norm such that for any $\boldsymbol{h} \in \mathscr{H}$, $\|\boldsymbol{h}\|^2 = \frac{1}{n}\langle \boldsymbol{h}, \boldsymbol{h} \rangle = \sum_{i=1}^{n} h_i^2$. We assume that $\boldsymbol{\theta} \in \mathscr{H}$, and we suppose that $G \subseteq \mathscr{H}$ is a finite-dimensional linear space spanned by bounded functions with the given knots on $[0, 1]$, which we will specify in the following part, and we call G an *approximating* space of $\mathscr{H}$.

In the unimodal case and the inflection-point case, we suppose that G_m is a subspace of G such that an equality constraint with respect to m holds, which will be specified for each case, and we define $\tilde{\boldsymbol{\theta}}_m$ as the orthogonal projection of $\boldsymbol{Y}$ onto G_m and $\bar{\boldsymbol{\theta}}$ as the orthogonal projection of $\boldsymbol{\theta}$ onto G_m; in the jump-point case, we define $\tilde{\boldsymbol{\theta}}_m$ as the orthogonal projection of $\boldsymbol{Y}$ onto G and $\bar{\boldsymbol{\theta}}$ as the orthogonal projection of $\boldsymbol{\theta}$ onto G.

By the definition of $\bar{\boldsymbol{\theta}}$, we know that it is the best approximation to $\boldsymbol{\theta}$ in G. $\tilde{\boldsymbol{\theta}}_m$ is estimating $\bar{\boldsymbol{\theta}}$ and we call it the unconstrained regression estimator of $\boldsymbol{\theta}$. If G is chosen such that $\bar{\boldsymbol{\theta}}$ is close to $\boldsymbol{\theta}$, then $\tilde{\boldsymbol{\theta}}_m$ should be close to $\boldsymbol{\theta}$ as it estimates $\bar{\boldsymbol{\theta}}$. Then we can decompose $\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}$ as

$$\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta} = \bar{\boldsymbol{\theta}} - \boldsymbol{\theta} + \tilde{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}},$$

where $\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}}$ are referred to as the *approximation error* and the *estimation error* respectively. Based on such decomposition, we will establish the convergence rate of each change-point estimator.

2.2.1. Unimodal Case. Given $m \in (0,1)$, the regression function $f_m(x)$ is assumed to be continuous with two continuous derivatives on $[0,1]$. We consider only the increasing-decreasing case, but the decreasing-increasing case is similar. If $f_m'(x) > 0$ on $[0,m)$ and $f_m'(x) < 0$ on $(m,1]$, the smoothness condition implies that $f_m'(m) = 0$. We use quadratic B-splines, with $\ell = k+1$ basis functions spanning the space of piece-wise quadratic spline functions. To enforce zero slope at m, let $\boldsymbol{v} \in \mathbb{R}^\ell$ be defined as $v_j = \eta_j'(m)$, and impose the linear equality constraint $\boldsymbol{v}^\top \boldsymbol{b} = 0$.

Because the first derivative of a quadratic spline function is piece-wise linear, constraining the function to be increasing at two adjacent knots ensures that the function is increasing in the interval between these knots. Suppose that t_p is the knot such that $t_p \leq m < t_{p+1}$,

$p \in \{1, \ldots, k-1\}$. We define a $k \times (k+1)$ matrix $\boldsymbol{S}$ as

$$S_{ij} = \begin{cases} \eta_j'(t_i) & i = 1, \ldots, p \text{ and } j = 1, \ldots, k+1 \\ \\ -\eta_j'(t_i) & i = p+1, \ldots, k \text{ and } j = 1, \ldots, k+1. \end{cases}$$

Then the linear combination $g(\boldsymbol{x}) = \sum_{j=1}^{m} b_j \boldsymbol{\eta}_j(\boldsymbol{x})$ is increasing on $[0, m)$ and decreasing on $(m, 1]$ if and only if the coefficient vector $\boldsymbol{b}$ is in the set $\left\{\boldsymbol{b} \in \mathbb{R}^{k+1} : \boldsymbol{v}^\top \boldsymbol{b} = 0 \text{ and } \boldsymbol{Sb} \geq \boldsymbol{0}\right\}$. These coefficient vectors can be written as $\boldsymbol{b} = \boldsymbol{Wc}$, where the columns of the $\ell \times k$ matrix $\boldsymbol{W}$ are orthogonal to $\boldsymbol{v}$. A vector $\boldsymbol{\theta}$ satisfies the constraints if it is in the set

$$C_m = \{\boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{\theta} = \boldsymbol{BWc}, \text{ where } \boldsymbol{SWc} \geq \boldsymbol{0}\}. \tag{4}$$

Finally, let $G_m \subseteq \mathbb{R}^n$ be the k-dimensional linear subspace containing C_m, defined by

$$G_m = \{\boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{\theta} = \boldsymbol{BWc}, \ \boldsymbol{c} \in \mathbb{R}^k\}. \tag{5}$$

2.2.2. Inflection-Point Case. To estimate $f_m(x)$ with an inflection point at m on $[0, 1]$, we use $\ell = k+2$ cubic B-spline basis functions, spanning the space of piece-wise cubic spline functions with the given knots. To enforce the second derivative to be zero at m, we use a linear equality constraint $\boldsymbol{v}^\top \boldsymbol{b} = 0$; here $v_j = \eta_j''(m)$.

Because the second derivative of a cubic spline function is piecewise linear, constraining the function to be convex at two adjacent knots ensures convexity between the knots. We discuss only the case $f_m''(x) > 0$ on $[0, m)$ and $f_m''(x) < 0$ on $(m, 1]$, but the concave-convex function estimation is similar. If t_p is the knot such that $t_p \leq m < t_{p+1}$, we define a $k \times (k+2)$

matrix $\boldsymbol{S}$ as

$$S_{ij} = \begin{cases} \eta_j^{''}(t_i) & i = 1, \ldots, p \text{ and } j = 1, \ldots, k+2 \\ -\eta_j^{''}(t_i) & i = p+1, \ldots, k \text{ and } j = 1, \ldots, k+2. \end{cases}$$

Then the linear combination of basis functions is convex on $[0, m)$ and concave on $(m, 1]$ if and only if the coefficient vector $\boldsymbol{b}$ satisfies $\boldsymbol{S}\boldsymbol{b} \geq \boldsymbol{0}$ and $\boldsymbol{v}^\top \boldsymbol{b} = 0$, and these coefficient vectors can be written as $\boldsymbol{b} = \boldsymbol{W}\boldsymbol{c}$, where the columns of the $\ell \times (k+1)$ matrix $\boldsymbol{W}$ are orthogonal to $\boldsymbol{v}$. The constraint set C_m and the linear vector space G_m are defined as for the unimodal case using (4) and (5).

The constraint matrix $\boldsymbol{S}$ is readily modified for the case in which the prior information includes a monotonicity assumption. For example, a growth curve might be known to be increasing as well as convex-concave. In this case two rows are added to $\boldsymbol{S}$. The first contains the derivatives of the spline basis functions at the left end-point of the interval, and the second contains the derivatives at the right end-point; that is, $S_{k+1,j} = \eta_j'(0)$, and $S_{k+2,j} = \eta_j'(1)$.

2.2.3. Jump-Point Case. To estimate a function $f_m(x)$ which has a jump point at m but is otherwise continuous with continuous first derivative on $[0, 1]$, we use the $k+1$ quadratic B-spline basis functions, one "jump" basis function and one "ramp" basis function. The first basis function is constant on $[0, m)$ and on $(m, 1]$, for example

$$\eta_{(k+2)}(x_i) = \begin{cases} 0 & \text{if } x_i < m, i = 1, \ldots, n \\ 1 & \text{if } x_i \geq m, i = 1, \ldots, n. \end{cases}$$

The ramp basis function is linear on $[0, m)$ and on $(m, 1]$, and the basis vector is defined as

$$\eta_{(k+3)}(x_i) = \begin{cases} 0 & \text{if } x_i < m, i = 1, \ldots, n \\ x_i - m & \text{if } x_i \geq m, i = 1, \ldots, n. \end{cases}$$

The basis vectors are the basis functions evaluated at the design points, and may be centered and scaled, for numerical stability. The $n \times (k+3)$ matrix $\boldsymbol{B}$ contains the basis vectors, and G is the space of quadratic spline functions with a jump and change of slope at m.

Suppose it is reasonable to assume that the regression function is decreasing, with an upward jump at m for now. This problem was motivated by the need to detect disturbances in forests, caused by fire or logging. The satellite signal is constant or slowly decreasing for a healthy forest, with a jump upward in the signal is caused by mass destruction of trees. The signal is subsequently decreasing as the forest recovers. Later we will relax the monotonicity constraint and the jump direction constraint for a more general problem. To be specific, we define $\gamma = \lim_{x \to m^+} f_m(x) - \lim_{x \to m^-} f_m(x)$ as the jump size of f_m at m. When $\gamma > 0$, there is an upward jump and when $\gamma < 0$, there is a downward jump. We only assume that $f_m(x)$ is decreasing (increasing) on $[t_p, m) \cup (m, t_{p+1}] \subset [0, 1]$, where t_p and t_{p+1} are two consecutive knots. Since $[t_p, t_{p+1}] \to 0$ as $n \to \infty$, we can drop the monotonicity constraint asymptotically.

Let $\boldsymbol{S}$ be a $(k+3) \times (k+3)$ matrix such that

$$S_{ij} = -\eta_j'(t_i) \;\; i = 1, \ldots, k \text{ and } j = 1, \ldots, k+1,$$

$$S_{(k+1)j} = \begin{cases} 1 & j = k+2 \\ 0 & j \neq k+2, \end{cases}$$

$$S_{(k+2)j} = \begin{cases} -\eta_j'(m) & j = 1, \ldots, k+1 \\ \\ 0 & j = k+2 \text{ and } k+3, \end{cases}$$

and finally

$$S_{(k+3)j} = \begin{cases} -\eta_j'(m) & j = 1, \ldots, k+1 \\ \\ 0 & j = k+2 \\ \\ -1 & j = k+3. \end{cases}$$

Then to model a decreasing function with an upward jump at m, we constrain the spline coefficients to be in the set $\left\{\boldsymbol{b} \in \mathbb{R}^{k+3} : \boldsymbol{S}\boldsymbol{b} \geq \boldsymbol{0}\right\}$, and

$$C = \{\boldsymbol{\theta} \in \mathbb{R}^n : \boldsymbol{\theta} = \boldsymbol{B}\boldsymbol{b}, \text{ where } \boldsymbol{S}\boldsymbol{b} \geq \boldsymbol{0}\}. \tag{6}$$

2.2.4. Least Squares Criterion. To minimize sum of squared residuals over appropriate values of $\boldsymbol{b}$, we use the criterion

$$\psi(\boldsymbol{b}) = \boldsymbol{b}^\top \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{b} - 2\boldsymbol{Y}^\top \boldsymbol{B}\boldsymbol{b}, \tag{7}$$

The equality constraint $\boldsymbol{v}^\top \boldsymbol{b} = 0$ is incorporated in the objective function

$$\psi(\boldsymbol{c}) = \boldsymbol{c}^\top \boldsymbol{W}^\top \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{W}\boldsymbol{c} - 2\boldsymbol{Y}^\top \boldsymbol{B}\boldsymbol{W}\boldsymbol{c}. \tag{8}$$

Define $\tilde{\boldsymbol{c}}_m$ as $(\boldsymbol{W}^\top \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{B}^\top \boldsymbol{Y}$ and subsequently $\tilde{\boldsymbol{\theta}}_m = \boldsymbol{B}\boldsymbol{W}\tilde{\boldsymbol{c}}_m$ is the unconstrained least-squares estimate of $\boldsymbol{\theta}$. (Here we use "constrained" and "unconstrained" to reflect the use of the inequality constraints. The unconstrained estimate in the unimodal case, for example, has zero derivative at m, but might not be unimodal.)

To get the constrained estimates, we solve a quadratic programming problem. Let $\boldsymbol{U}^\top \boldsymbol{U}$ be the Cholesky decomposition of $\boldsymbol{W}^\top \boldsymbol{B}^\top \boldsymbol{B}\boldsymbol{W}$. Define $\boldsymbol{\phi} = \boldsymbol{U}\boldsymbol{c}$ and $\boldsymbol{Z} = (\boldsymbol{U}^{-1})^\top \boldsymbol{W}^\top \boldsymbol{B}^\top \boldsymbol{Y}$, so that (7) can be written as $\|\boldsymbol{Z} - \boldsymbol{\phi}\|^2$, and the constraints are $\boldsymbol{A}\boldsymbol{\phi} \geq \boldsymbol{0}$, where $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{W}\boldsymbol{U}^{-1}$. The minimizer $\hat{\boldsymbol{\phi}}_m$ is the projection of $\boldsymbol{Z}$ onto the cone $\tilde{C} = \left\{\boldsymbol{\phi} \in \mathbb{R}^d : \boldsymbol{A}\boldsymbol{\phi} \geq \boldsymbol{0}\right\}$, where $d = k$ for the unimodal case and $d = k+1$ for the inflection-point case. The `coneA` function in the `R` package `coneproj` Meyer and Liao (2014) will provide $\hat{\boldsymbol{\phi}}_m$, and $\hat{\boldsymbol{\theta}}_m = \boldsymbol{B}\boldsymbol{W}\boldsymbol{U}^{-1}\hat{\boldsymbol{\phi}}_m$.

Necessary and sufficient conditions for the projection of $\boldsymbol{Z}$ onto $\tilde{C}$ are: $\hat{\boldsymbol{\phi}}$ minimizes $\|\boldsymbol{Z} - \boldsymbol{\phi}\|^2$ over $\tilde{C}$ if and only if

$$\langle \boldsymbol{Z} - \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\phi}} \rangle = 0 \;\text{ and }\; \langle \boldsymbol{Z} - \hat{\boldsymbol{\phi}}, \boldsymbol{\phi} \rangle \leq 0, \;\text{ for }\; \boldsymbol{\phi} \in \tilde{C},$$

which provides necessary and sufficient conditions for the projection $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{Y}$ onto C_m:

$$\langle \boldsymbol{Y} - \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}} \rangle = 0 \;\text{ and }\; \langle \boldsymbol{Y} - \hat{\boldsymbol{\theta}}, \boldsymbol{\theta} \rangle \leq 0, \;\text{ for }\; \boldsymbol{\theta} \in C_m. \tag{9}$$

These are well-known special forms of the Karush-Kuhn-Tucker conditions; see Section 28 of Rockafellar (1970), or Proposition 3.12.3 of Silvapulle and Sen (2005).

Finally, when m is unknown, we search $[0, 1]$ for the m that minimizes the sum of squared residuals. In this case, we need not utilize the equality constraints imposed with the matrix $\boldsymbol{W}$. For each $j = 1, \ldots, k-1$, we construct the constraint matrix for the coefficients that enforces the shape. For the unimodal case, this is increasing at knots $t_1, \ldots, t_j$, then decreasing at knots $t_{j+1}, \ldots, t_k$. The mode can be inferred from the least-squares fit with these constraints, and the mode estimate is that which minimizes the sum of squared residuals over j. The procedure for estimating the inflection point is similar. In summary, only $k-1$ constrained spline estimators are needed, and $\hat{m}$ is computed from the fit that minimizes the least-squares criterion.

CHAPTER 3

Theoretical Results

3.1. Convergence Rates

When the change-point m is known, established results apply to the unconstrained least-squares spline estimator. References include Stone (1980), Stone (1982), Stone, Hansen, Kooperberg, and Truong (1997), Zhou, Shen, and Wolfe (1998), Huang (1998), Zhou and Wolfe (2000), and Huang (2001). If p is the order of the spline, let the number of knots increase as $n^{1/(2p+1)}$; that is, as $n^{1/7}$ for quadratic splines and as $n^{1/9}$ for cubic splines. Further, the knots must have bounded mesh ratios, that is, ratios of lengths of consecutive knot intervals are bounded away from zero and infinity. Further assume that the design points $X_1, \ldots, X_n$ follow a distribution H with density $h > 0$ on $(0, 1)$, so that the proportion of design points less than $c \in (0, 1)$ approaches $H(c)$ as n increases without bound. These conditions together with the smoothness assumptions for the regression function f_m give the following local and global results. If $\tilde{\boldsymbol{\theta}}_m$ is the projection of $\boldsymbol{Y}$ onto the linear vector space G_m, and $\tilde{f}_m$ is the corresponding spline function, then for $x \in [0, 1]$,

$$|\tilde{f}_m(x) - f_m(x)| = O_p(n^{-p/(2p+1)}),$$

and

$$\frac{1}{n}\|\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}\|^2 = O_p(n^{-2p/(2p+1)}),$$

where $\|\boldsymbol{a}\|^2 = \langle \boldsymbol{a}, \boldsymbol{a} \rangle$ and $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \sum_{i=1}^{n} a_i b_i$. Next, let $\hat{\boldsymbol{\theta}}_m$ be the projection of $\boldsymbol{Y}$ onto C_m and let $\bar{\boldsymbol{\theta}}$ be the projection of $\boldsymbol{\theta}$ onto G_m. Then if $\bar{\boldsymbol{\theta}} \in C_m$ (i.e., the constraints hold), we have

$$\begin{aligned}
\|\tilde{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}}\|^2 &= \|\tilde{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}}_m\|^2 + \|\hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}}\|^2 + 2\langle \tilde{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}}_m, \hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}} \rangle \\
&= \|\tilde{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}}_m\|^2 + \|\hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}}\|^2 + 2\langle \tilde{\boldsymbol{\theta}}_m - \boldsymbol{Y}, \hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}} \rangle + 2\langle \boldsymbol{Y} - \hat{\boldsymbol{\theta}}_m, \hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}} \rangle \\
&= \|\tilde{\boldsymbol{\theta}}_m - \hat{\boldsymbol{\theta}}_m\|^2 + \|\hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}}\|^2 - 2\langle \boldsymbol{Y} - \hat{\boldsymbol{\theta}}_m, \bar{\boldsymbol{\theta}} \rangle \\
&\geq \|\hat{\boldsymbol{\theta}}_m - \bar{\boldsymbol{\theta}}\|^2,
\end{aligned}$$

because by (9), the last inner product is negative. Because $\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}$ is orthogonal to G_m, we have $\|\tilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}\|^2 \geq \|\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}\|^2$, and hence the constrained spline attains the rate for the unconstrained spline.

Next we consider the case when m is unknown, and is estimated by $\hat{m}$, the change-point for which the sum of squared residuals is minimized. By definition, $\|\boldsymbol{Y} - \hat{\boldsymbol{\theta}}_{\hat{m}}\|^2 \leq \|\boldsymbol{Y} - \hat{\boldsymbol{\theta}}_m\|^2$, so

$$\|\boldsymbol{Y} - \boldsymbol{\theta}\|^2 + \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\hat{m}}\|^2 + 2\langle \boldsymbol{Y} - \boldsymbol{\theta}, \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\hat{m}} \rangle \leq \|\boldsymbol{Y} - \boldsymbol{\theta}\|^2 + \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_m\|^2 + 2\langle \boldsymbol{Y} - \boldsymbol{\theta}, \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_m \rangle,$$

or

$$\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_m\|^2 - \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\hat{m}}\|^2 \geq 2\langle \boldsymbol{Y} - \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_m \rangle - 2\langle \boldsymbol{Y} - \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{\hat{m}} \rangle.$$

By (8) of Meyer and Woodroofe (2000), each term on the right is $O_p(d)$, where d is the dimension of G_m, the smallest linear space containing C_m. Therefore with the optimal number of knots, we have

$$\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\hat{m}}\|^2 = O_p(n^{1/(2p+1)}). \tag{10}$$

3.1.1. Convergence of the Change-Point Estimator for the Unimodal Case. To derive the rate of convergence of $\hat{m}$ to m, in addition to $f_m^{'}(m) = 0$, $f_m^{'}(x) > 0$ on $[0, m)$, and $f_m^{'}(x) < 0$ on $(m, 1]$, we assume that $f_m^{''}(x) < -D < 0$ on $[m - \delta, m + \delta]$, $\delta > 0$. The proofs assume that the distribution of x in $[0, 1]$ is uniform, but the results are valid if the density function is bounded away from zero and infinity.

Lemma 1. There is a $c > 0$ that depends only on D, such that for any $\delta > 0$ and any g with $g'(x) \geq 0$ on $[m, m + \delta]$,

$$\int_m^{m+\delta} [f_m(x) - g(x)]^2 dx \geq c\delta^5,$$

and similarly for any g with $g'(x) \leq 0$ on $[m - \delta, m]$,

$$\int_{m-\delta}^{m} [f_m(x) - g(x)]^2 dx \geq c\delta^5.$$

To prove Lemma 1, we start with

Lemma 2. For any function g such that $g'(x) \geq 0$ on $[m, m + \delta]$, we have

$$\int_m^{m+\delta} [f_m(x) - g(x)]^2 dx \geq \int_m^{m+\delta} [f_m(x) - \ell]^2 dx, \text{ where } \ell = \frac{1}{\delta} \int_m^{m+\delta} f_m(x) dx.$$

Proof: Define the constant c as follows. If $g(x) \geq f_m(x)$ over $[m, m + \delta]$, then let $c = g(m)$. If $g(x) \leq f_m(x)$ over $[m, m + \delta]$, then let $c = g(m + \delta)$. If f_m and g intersect at m_0, for $m_0 \in [m, m+\delta]$, then let $c = g(m_0)$. Then $[f_m(x)-g(x)]^2 \geq [f_m(x)-c]^2$ for all $x \in [m, m+\delta]$, and $\int_m^{m+\delta} [f_m(x) - g(x)]^2 dx \geq \int_m^{m+\delta} [f_m(x) - c]^2 dx$. The constant ℓ minimizes the expression $\int_m^{m+\delta} [f_m(x) - c]^2 dx$ over all constants c.

Proof of Lemma 1: Let x_0 be such that $f_m(x_0) = \ell$; it is straightforward to see that $m < x_0 < m + \delta$. By the strict concavity of f_m on $[m - \delta, m + \delta]$, we have for all $x \in [m, m + \delta]$,

$$|f_m(x) - \ell| > [f_m(m) - \ell] \left(\frac{|x_0 - x|}{x_0 - m} \right),$$

so

$$\int_m^{m+\delta} [f_m(x) - \ell]^2 dx \geq \frac{1}{3}[f_m(m) - \ell]^2 \frac{(m + \delta - x_0)^3 + (x_0 - m)^3}{(x_0 - m)^2}.$$

Now, for some ξ_0 between m and x_0, we have by Taylor's expansion of f_m at m:

$$\ell = f_m(m) + \frac{1}{2} f''_m(\xi_0)(x_0 - m)^2,$$

so $f_m(m) - \ell \geq D(x_0 - m)^2/2$, and

$$\int_m^{m+\delta} [f_m(x) - \ell]^2 dx \geq \frac{D^2}{12}(x_0 - m)^2 \left[(m + \delta - x_0)^3 + (x_0 - m)^3\right]. \tag{11}$$

By the strict concavity of f_m at x_0, we can get

$$\int_m^{x_0} \left[f_m(x) - \ell\right] dx < -\frac{1}{2} f'_m(x_0)(x_0 - m)^2,$$

and

$$\int_{x_0}^{m+\delta} \left[\ell - f_m(x)\right] dx > -\frac{1}{2} f'_m(x_0)(m + \delta - x_0)^2.$$

Also note that

$$\int_m^{m+\delta} \left[f_m(x) - \ell\right] dx = 0,$$

we can get

$$\int_{m}^{x_0} \big[f_m(x) - \ell\big]dx = \int_{x_0}^{m+\delta} \big[\ell - f_m(x)\big]dx,$$

so $x_0 - m > m + \delta - x_0$, which implies that

$$x_0 - m > \frac{\delta}{2}. \tag{12}$$

Now we let $x_0 - m = \alpha\delta$ and $m + \delta - x_0 = (1-\alpha)\delta$, $0 < \alpha < 1$. Then we get

$$(m + \delta - x_0)^3 + (x_0 - m)^3 = (3\alpha^2 - 3\alpha + 1)\delta^3.$$

Since we have shown that $\alpha > 1/2$, we can get

$$(m + \delta - x_0)^3 + (x_0 - m)^3 \geq \frac{\delta^3}{4}. \tag{13}$$

Plugging (12) and (13) back in (11), we get

$$\int_{m}^{m+\delta} \big[f_m(x) - \ell\big]^2 dx \geq c\delta^5,$$

where c only depends on D.

Theorem 1. For the unimodal case, $|\hat{m} - m| = O_p(n^{-6/35})$.

Proof: First, we want to show that $|\hat{m} - m| \xrightarrow{p} 0$. Suppose that $|\hat{m} - m| \overset{p}{\nrightarrow} 0$, then there is some $\delta > 0$ such that $|\hat{m} - m| > \delta$ *i.o.*, where *i.o.* stands for infinitely often. Suppose that $\hat{m} > m + \delta$ *i.o.*. When this condition holds, since $[m, \hat{m}] \subset [0, 1]$, we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{\hat{m}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{m+\delta} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx.$$

By Lemma 1, we know that

$$\int_m^{m+\delta} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > c\delta^5.$$

Therefore

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f(x)]^2 dx > c\delta^5, \tag{14}$$

and because

$$\frac{1}{n}\|\hat{\boldsymbol{\theta}}_{\hat{m}} - \boldsymbol{\theta}\|^2 \asymp \int_0^1 \left[\hat{f}_{\hat{m}}(x) - f_m(x)\right]^2 dx, \tag{15}$$

(where $a_n \asymp b_n$ means that $a_n/b_n \overset{p}{\to} 1$ as $n \to \infty$), we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f(x)]^2 dx = O(n^{-6/7}),$$

which implies that there is some $M > 0$ such that

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx \leq n^{-6/7}M \ \text{ as } \ n \to \infty. \tag{16}$$

However

$$c\delta^5 > n^{-6/7}M \ \text{ as } \ n \to \infty,$$

which implies that there is some N, such that

$$c\delta^5 > n^{-6/7}M \ \text{ for all } \ n > N.$$

Then by (14) we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > n^{-6/7} M \ \ i.o.,$$

and this is a contradiction to (16). Therefore, we can conclude that $|\hat{m} - m| \xrightarrow{p} 0$.

Next, we want to show that $|\hat{m} - m| = O_p(n^{-6/35})$. Suppose that for some fixed $\alpha > 0$, $|\hat{m} - m| > n^{-\alpha}$ *i.o.*, and we claim that $\alpha > 6/35$. We can further suppose that $\hat{m} - m > n^{-\alpha}$ *i.o.*. When this condition holds, since $[m, \hat{m}] \subset [0,1]$, we can get

$$(17) \qquad \int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{\hat{m}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{m+n^{-\alpha}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx.$$

By Lemma 1, we know that

$$(18) \qquad \int_m^{m+n^{-\alpha}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > cn^{-5\alpha}.$$

By (17) and (18), we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > cn^{-5\alpha}.$$

Since (16) holds, we can get

$$cn^{-5\alpha} < n^{-6/7} M,$$

which implies

$$n^{-\alpha} < n^{-6/35} (M/c)^{1/5}.$$

This means that for all $\varepsilon > 0$, there is some $\tilde{N}$ such that for all $n > \tilde{N}$ we have

$$\mathcal{P}\big\{|\hat{m} - m| > n^{-6/35} (M/c)^{1/5}\big\} < \varepsilon.$$

Therefore, we can conclude that $|\hat{m} - m| = O_p(n^{-6/35})$.

3.1.2. Convergence of the Change-Point Estimator for the Inflection-Point Case. In addition to $f''_m(m) = 0$, $f''_m(x) > 0$ on $[0, m)$, $f''_m(x) < 0$ on $(m, 1]$, we assume that $f^{(3)}_m(x) < -D < 0$ on $[m - \delta, m + \delta]$, $\delta > 0$, and $f^{(4)}_m(x)$ is continuous on $[m - \delta, m + \delta]$. We again assume that the distribution of x in $[0, 1]$ is uniform, for simplicity of presentation.

Lemma 3. For a function g such that $g''(x) \leq 0$ on $[m - \delta, m]$, there is a $c > 0$ that depends only on D,

$$\int_{m-\delta}^{m} [f_m(x) - g(x)]^2 dx \geq c\delta^7,$$

and similarly for g such that $g''(x) \geq 0$ on $[m, m + \delta]$,

$$\int_{m}^{m+\delta} [f_m(x) - g(x)]^2 dx \geq c\delta^7.$$

To prove Lemma 3, we begin with

Lemma 4. For any function g such that $g''(x) \leq 0$ on $[m - \delta, m]$, we have

$$\int_{m-\delta}^{m} \big[f_m(x) - g(x)\big]^2 dx \geq \int_{m-\delta}^{m} \big[f_m(x) - \ell(x)\big]^2 dx, \quad \text{where} \quad \ell(x)$$

minimizes $\int_{m-\delta}^{m} \big[f_m(x) - g(x)\big]^2 dx$ over the class $L = \{g_l(x) : g_l(x) = c_0 + c_1 x, c_0, c_1 \in \mathbb{R}\}$.

Proof: Define the linear function $g_l(x)$ as follows. If $g(m-\delta) \geq f_m(m-\delta)$ and $g(m) \geq f_m(m)$, then let $g_l(x)$ be the line connecting the points $\big(m-\delta, g(m-\delta)\big)$ and $\big(m, g(m)\big)$. If $g(m-\delta) \geq f_m(m - \delta)$ and $g(x)$ intersects $f_m(x)$ at m_1, for $m_1 \in (m - \delta, m)$, then let $g_l(x)$ be the line connecting the points $\big(m-\delta, g(m-\delta)\big)$ and $\big(m_1, g(m_1)\big)$. If $g(m) \geq f_m(m)$ and $g(x)$ intersects $f_m(x)$ at m_1, for $m_1 \in (m-\delta, m)$, then let $g_l(x)$ be the line connecting the points $\big(m_1, g(m_1)\big)$ and $\big(m, g(m)\big)$. If $g(x)$ intersects $f_m(x)$ at m_1 and m_2, where $[m_1, m_2] \subset (m - \delta, m)$, then let

$g_l(x)$ be the line connecting the points $\big(m_1, g(m_1)\big)$ and $\big(m_2, g(m_2)\big)$. Then $[f_m(x) - g(x)]^2 \geq [f_m(x) - g_l(x)]^2$ on $[m-\delta, m]$, and $\int_{m-\delta}^{m} \big[f_m(x) - g(x)\big]^2 dx \geq \int_{m-\delta}^{m} \big[f_m(x) - g_l(x)\big]^2 dx$.

By taking the first derivative of $\int_{m-\delta}^{m} \big[f_m(x) - g_l(x)\big]^2 dx$ with respect to c_0 and c_1, we can get the minimizer of $\int_{m-\delta}^{m} \big[f_m(x) - g(x)\big]^2 dx$ over the class $L = \{g_l(x) : g_l(x) = c_0 + c_1 x, c_0, c_1 \in \mathbb{R}\}$, and we define it as $\ell(x)$. Let $h_m(x) = f_m(x) - \ell(x)$, then it readily follows that $h_m(x_1) = 0$, $h_m(x_2) = 0$, and $h_m^{(k)}(x) = f_m^{(k)}(x)$, $k \geq 2$; then we bound below

$$\int_{m-\delta}^{m} \big[f_m(x) - \ell(x)\big]^2 dx = \int_{m-\delta}^{m} \big[h_m(x)\big]^2 dx.$$

Lemma 5. We have

$$h_m(m) > \frac{D}{30}\delta^3 + O(\delta^4).$$

Proof: By Taylor's expansion of h_m at m and $h_m''(m) = 0$, we have

$$h_m(x) = h_m(m) + h_m'(m)(x-m) + \frac{1}{6}h_m^{(3)}(m)(x-m)^3 + \frac{1}{24}h_m^{(4)}(\xi_{xm})(x-m)^4,$$

where $\xi_{xm} \in (x, m)$. Moreover, note that $\int_{m-\delta}^{m} h_m(x)dx = 0$ and $\int_{m-\delta}^{m} x h_m(x)dx = 0$, so

$$h_m(m)\delta - h_m'(m)\frac{\delta^2}{2} - h_m^{(3)}(m)\frac{\delta^4}{24} + \frac{1}{24}\left[\int_{m-\delta}^{m} h_m^{(4)}(\xi_{xm})(x-m)^4 dx\right] = 0$$

and

$$h_m(m)\big(m\delta - \frac{\delta^2}{2}\big) + h_m'(m)\big(\frac{\delta^3}{3} - \frac{m\delta^2}{2}\big) + h_m^{(3)}(m)\big(\frac{\delta^5}{30} - \frac{m\delta^4}{24}\big) + \frac{1}{24}\left[\int_{m-\delta}^{m} h_m^{(4)}(\xi_{xm})x(x-m)^4 dx\right] = 0.$$

Solving the two equations for $h_m(m)$, we can get

$$\begin{aligned}
h_m(m) &= -h_m^{(3)}(m)\frac{\delta^3}{30} - \frac{1}{\delta^2}\left[\frac{1}{4}\int_{m-\delta}^{m} h_m^{(4)}(\xi_{xm})(x-m)^5 dx + \frac{\delta}{6}\int_{m-\delta}^{m} h_m^{(4)}(\xi_{xm})(x-m)^4 dx\right] \\
&= -h_m^{(3)}(m)\frac{\delta^3}{30} + O(\delta^4).
\end{aligned}$$

Then observe $h_m^{(3)}(m) < -D < 0$, to get the result.

Analogously to the unimodal case, we connect with a parabola $p(x)$ the points $\big(x_1, h_m(x_1)\big)$, $\big(x_2, h_m(x_2)\big)$ and $\big(m, h_m(m)\big)$. We show that $|h_m(x)| \geq |p(x)|$, and bound below $\int_{m-\delta}^{m} p(x)^2 dx$. To do this, we first show that the curvature of the parabola is sufficiently large.

Lemma 6. The parabola $p(x)$ satisfies

$$p'' > \frac{D}{15}\delta + O(\delta^2),$$

where p'' is the second derivative of $p(x)$, and $|p(x)| < |h_m(x)|$ on $[m-\delta, m]$.

Proof: Note that $h_m(x_1) = 0$ and $h_m(x_2) = 0$, by the definition of $p(x)$, we can get

$$p(x) = h_m(m)\frac{(x-x_1)(x-x_2)}{(m-x_1)(m-x_2)}, \tag{19}$$

which implies that

$$p'' = \frac{2h_m(m)}{(m-x_1)(m-x_2)}. \tag{20}$$

Note that $0 < m - x_2 < \delta$ and $0 < m - x_1 < \delta$, so by Lemma 5,

$$p'' > \frac{D}{15}\delta + O(\delta^2).$$

Then we only need to show that $p'(x_2) < h'(x_2)$ and $p'(x_1) > h'(x_1)$ to get $|p(x)| < |h_m(x)|$ on $[m-\delta, m]$. From (19),

$$p'(x_2) = h_m(m)\frac{x_2 - x_1}{(m-x_1)(m-x_2)}. \tag{21}$$

By Taylor's expansion of $h_m(m)$ at x_2, and $h_m(x_2) = 0$, we have

$$h_m(m) = h'_m(x_2)(m-x_2) + \frac{1}{2}h''_m(x_2)(m-x_2)^2 + \frac{1}{6}h_m^{(3)}(\xi_{2m})(m-x_2)^3,$$

where $\xi_{2m} \in (x_2, m)$. Also note that $h''_m(x_2) = -h_m^{(3)}(\xi_{2m})(m-x_2)$, we can get

$$h_m(m) = h'_m(x_2)(m-x_2) - \frac{1}{3}h_m^{(3)}(\xi_{2m})(m-x_2)^3.$$

Then by (21), we can write $p'(x_2)$ as

$$p'(x_2) = h'_m(x_2)\frac{x_2-x_1}{m-x_1} - \frac{1}{3}h_m^{(3)}(\xi_{2m})\frac{(x_2-x_1)(m-x_2)^2}{m-x_1},$$

which implies that

$$p'(x_2) - h'_m(x_2) = -h'_m(x_2)\frac{m-x_2}{m-x_1} - \frac{1}{3}h_m^{(3)}(\xi_{2m})\frac{(x_2-x_1)(m-x_2)^2}{m-x_1}. \tag{22}$$

By Taylor's expansion of $h_m(x_1)$ at x_2, we have

$$h_m(x_1) = h_m(x_2) + h'_m(x_2)(x_1-x_2) + \tfrac{1}{2}h''_m(x_2)(x_1-x_2)^2 + \tfrac{1}{6}h_m^{(3)}(\xi_{12})(x_1-x_2)^3,$$

where $\xi_{12} \in (x_1, x_2)$. Note that $h_m(x_1) = 0$ and $h_m(x_2) = 0$, we can get

$$h'_m(x_2) = \frac{1}{2}h''_m(x_2)(x_2-x_1) - \frac{1}{6}h_m^{(3)}(\xi_{12})(x_2-x_1)^2. \tag{23}$$

Plugging (23) in the RHS of (22), we can get

$$-\frac{1}{2}h''_m(x_2)\frac{(x_2-x_1)(m-x_2)}{m-x_1}+\frac{1}{6}h^{(3)}_m(\xi_{12})\frac{(x_2-x_1)^2(m-x_2)}{m-x_1}-\frac{1}{3}h^{(3)}_m(\xi_{2m})\frac{(x_2-x_1)(m-x_2)^2}{m-x_1}.$$

Using $h''_m(x_2) = -h^{(3)}_m(\xi_{2m})(m-x_2)$ again, we can further write it as

$$\frac{1}{6}h^{(3)}_m(\xi_{12})\frac{(x_2-x_1)^2(m-x_2)}{m-x_1}+\frac{1}{6}h^{(3)}_m(\xi_{2m})\frac{(x_2-x_1)(m-x_2)^2}{m-x_1}.$$

Since $h^{(3)}_m(\xi_{12}) < 0$ and $h^{(3)}_m(\xi_{2m}) < 0$, we know that the RHS of (22) is negative, which implies that $p'(x_2) < h'(x_2)$. Next, we want to show that $p'(x_1) > h'_m(x_1)$. By (19) and (21), we can get that $p'(x_1) = -p'(x_2)$, so we only need to show that $p'(x_2) < -h'_m(x_1)$. By Taylor's expansion of $h_m(x_2)$ at x_1, we have

$$h_m(x_2) = h_m(x_1) + h'_m(x_1)(x_2-x_1) + \frac{1}{2}h''_m(x_1)(x_2-x_1)^2 + \frac{1}{6}h^{(3)}_m(\xi_{12})(x_2-x_1)^3.$$

Note that $h_m(x_1) = 0$ and $h_m(x_2) = 0$, we further get

$$-h'_m(x_1) = \frac{1}{2}h''_m(x_1)(x_2-x_1) + \frac{1}{6}h^{(3)}_m(\xi_{12})(x_2-x_1)^2. \tag{24}$$

By (23) and $h''_m(x_2) = h''_m(x_1) + h^{(3)}_m(\xi_{12})(x_2-x_1)$, we can get

$$h'_m(x_2) = \frac{1}{2}h''_m(x_1)(x_2-x_1) + \frac{1}{3}h^{(3)}_m(\xi_{12})(x_2-x_1)^2. \tag{25}$$

Since $h^{(3)}_m(\xi_{12}) < -D < 0$, we know that $h'_m(x_2) < -h'_m(x_1)$ by comparing (24) and (25), which implies that $p'(x_2) < -h'_m(x_1)$ and equivalently $p'(x_1) > h'_m(x_1)$. Therefore, we can conclude that $|p(x)| < |h_m(x)|$ on $[m-\delta, m]$ by the previous results.

Proof of Lemma 3: By Lemma 4, we know that for any function g such that $g''(x) \leq 0$ on $[m-\delta, m]$, we have

$$\int_{m-\delta}^{m} [f_m(x) - g(x)]^2 dx \geq \int_{m-\delta}^{m} [h_m(x)]^2 dx.$$

By Lemma 6, we know that the parabola $p(x)$ which connects $\big(x_1, h_m(x_1)\big)$, $\big(x_2, h_m(x_2)\big)$ and $\big(m, h_m(m)\big)$ satisfies that $|p(x)| < |h_m(x)|$ on $[m-\delta, m]$. Therefore

$$\int_s [h_m(x)]^2 dx > \int_s [p(x)]^2 dx$$

holds for s, where $s \in S = \big\{[m-\delta, x_1], [x_1, x_2], [x_2, m]\big\}$. Note that the length of at least one element in S is no less than $\delta/3$. We consider the scenario that $s = [x_1, x_2]$ and $x_2 - x_1 \geq \delta/3$. By Lemma 6, we can get

$$\int_s [p(x)]^2 dx = \frac{1}{4} p''^2 \int_s (x - x_2)^2 (x - x_1)^2 dx = \frac{1}{120} p''^2 (x_2 - x_1)^5 > c\delta^7 + O(\delta^8), \tag{26}$$

where c depends only on D. In another two possible scenarios, i.e., $s = [x_2, m]$ and $m - x_2 \geq \delta/3$, or $s = [m-\delta, x_1]$ and $x_1 - (m - \delta) \geq \delta/3$, we can get (26) similarly. Therefore

$$\int_{m-\delta}^{m} \big[p(x)\big]^2 dx > c\delta^7 + O(\delta^8),$$

which implies that

$$\int_{m-\delta}^{m} [h_m(x)]^2 dx > c\delta^7 + O(\delta^8).$$

As $n \to \infty$, $O(\delta^8)$ is negligible, and this gives the result.

Theorem 2. For the inflection-point case, $|\hat{m} - m| = O_p(n^{-8/63})$.

Proof: First, we want to show that $|\hat{m} - m| \xrightarrow{p} 0$. Suppose that $|\hat{m} - m| \overset{p}{\nrightarrow} 0$, then there is some $\delta > 0$ such that $|\hat{m} - m| > \delta$ *i.o.*, where *i.o.* stands for infinitely often. Suppose that

$\hat{m} > m + \delta$ *i.o.*. When this condition holds, since $[m, \hat{m}] \subset [0, 1]$, we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{\hat{m}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{m+\delta} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx.$$

By Lemma 3, we know that

$$\int_m^{m+\delta} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > c\delta^7.$$

Therefore

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > c\delta^7. \tag{27}$$

With (15), we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx = O(n^{-8/9}),$$

which implies that there is some $M > 0$ such that

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx \leq n^{-8/9} M \text{ as } n \to \infty. \tag{28}$$

However

$$c\delta^7 > n^{-8/9} M \text{ as } n \to \infty,$$

which implies that there is some N, such that

$$c\delta^7 > n^{-8/9} M \text{ for all } n > N.$$

Then by (27) we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > n^{-8/9} M \ \ i.o.,$$

and this is a contradiction to (28). Therefore, we can conclude that $|\hat{m} - m| \xrightarrow{p} 0$.

Next, we want to show that $|\hat{m} - m| = O_p(n^{-8/63})$. Suppose that for some fixed $\alpha > 0$, $|\hat{m} - m| > n^{-\alpha}$ *i.o.*, and we claim that $\alpha > 8/63$. We can further suppose that $\hat{m} - m > n^{-\alpha}$ *i.o.*. When this condition holds, since $[m, \hat{m}] \subset [0, 1]$, we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{\hat{m}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{m+n^{-\alpha}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx. \tag{29}$$

By Lemma 3, we know that

$$\int_m^{m+n^{-\alpha}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > cn^{-7\alpha}. \tag{30}$$

By (29) and (30), we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > cn^{-7\alpha}.$$

Since (28) holds, we can get

$$cn^{-7\alpha} < n^{-8/9} M,$$

which implies

$$n^{-\alpha} < n^{-8/63} (M/c)^{1/7}.$$

This means that for all $\varepsilon > 0$, there is some $\tilde{N}$ such that for all $n > \tilde{N}$ we have

$$\mathcal{P}\big\{|\hat{m} - m| > n^{-8/63} (M/c)^{1/7}\big\} < \varepsilon.$$

Therefore, we can conclude that $|\hat{m} - m| = O_p(n^{-8/63})$.

3.1.3. Convergence of the Change-Point Estimator for the Jump-Point Case.

In addition to $f_m^{'}(x) < 0$ on $[m - \delta, m) \cup (m, m + \delta]$, we define $\gamma = f_m(m+) - f_m(m-)$ as the jump size and we only discuss the case $\gamma > 0$, i.e., $f_m(x)$ has an upward jump at m. The same results hold when $\gamma < 0$.

Lemma 7. For any function g such that $g'(x) \le 0$ on $[m, m+\delta]$, there is a $c > 0$ that depends only on γ, we have

$$\int_m^{m+\delta} [f_m(x) - g(x)]^2 dx \ge c\delta,$$

and similarly for g such that $g'(x) \le 0$ on $[m-\delta, m]$,

$$\int_{m-\delta}^{m} [f_m(x) - g(x)]^2 dx \ge c\delta.$$

Proof of Lemma 7: If $g(m) \le f_m(m-)$, by the monotonicity of f_m and g on $[m, m+\delta]$, we have

$$f_m(x) - g(x) \ge f_m(m+\delta) - g(m) \ge f_m(m+\delta) - f_m(m-) \ge \frac{\gamma}{2}.$$

Similarly, if $g(m) \ge f_m(m+)$, we have

$$g(x) - f_m(x) \ge g(m) - f_m(m-\delta) \ge f_m(m+) - f_m(m-\delta) \ge \frac{\gamma}{2},$$

and if $f_m(m-) < g(m) < f_m(m+)$, we have

$$f_m(x) - g_m(x) \ge f_m(m+\delta) - g_m(m) \ge \frac{(1-\xi)}{2}\gamma,$$

for some $\xi \in (0,1)$. Combining these results we can get

$$\int_m^{m+\delta} [f_m(x) - g(x)]^2 dx \ge c\delta$$

for some c which only depends on γ.

Theorem 3. For the jump-point case, $|\hat{m} - m| = O_p(n^{-6/7})$.

Proof: First, we want to show that $|\hat{m} - m| \xrightarrow{p} 0$. Suppose that $|\hat{m} - m| \overset{p}{\nrightarrow} 0$, then there is some $\delta > 0$ such that $|\hat{m} - m| > \delta$ *i.o.*, where *i.o.* stands for infinitely often. Suppose that $\hat{m} > m + \delta$ *i.o.*. When this condition holds, since $[m, \hat{m}] \subset [0, 1]$, we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{\hat{m}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{m+\delta} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx.$$

By Lemma 7, we know that

$$\int_m^{m+\delta} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > c\delta.$$

Therefore

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > c\delta. \tag{31}$$

With (15), we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx = O(n^{-6/7}),$$

which implies that there is some $M > 0$ such that

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx \leq n^{-6/7} M \text{ as } n \to \infty. \tag{32}$$

However

$$c\delta > n^{-6/7} M \text{ as } n \to \infty,$$

which implies that there is some N, such that

$$c\delta > n^{-6/7}M \text{ for all } n > N.$$

Then by (31) we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > n^{-6/7}M \;\; i.o.,$$

and this is a contradiction to (32). Therefore, we can conclude that $|\hat{m} - m| \xrightarrow{p} 0$.

Next, we want to show that $|\hat{m} - m| = O_p(n^{-6/7})$. Suppose that for some fixed $\alpha > 0$, $|\hat{m} - m| > n^{-\alpha}$ *i.o.*, and we claim that $\alpha > 6/7$. We can further suppose that $\hat{m} - m > n^{-\alpha}$ *i.o.*. When this condition holds, since $[m, \hat{m}] \subset [0, 1]$, we can get

$$(33) \qquad \int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{\hat{m}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > \int_m^{m+n^{-\alpha}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx.$$

By Lemma 7, we know that

$$(34) \qquad \int_m^{m+n^{-\alpha}} [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > cn^{-\alpha}.$$

By (33) and (34), we can get

$$\int_0^1 [\hat{f}_{\hat{m}}(x) - f_m(x)]^2 dx > cn^{-\alpha}.$$

Since (32) holds, we can get

$$cn^{-\alpha} < n^{-6/7}M,$$

which implies

$$n^{-\alpha} < n^{-6/7}M/c.$$

This means that for all $\varepsilon > 0$, there is some $\tilde{N}$ such that for all $n > \tilde{N}$ we have

$$\mathcal{P}\{|\hat{m} - m| > n^{-6/7}M/c\} < \varepsilon.$$

Therefore, we can conclude that $|\hat{m} - m| = O_p(n^{-6/7})$.

3.2. Constrained Penalized Regression Spline Estimators

Penalized spline functions provide more flexibility while avoiding over-fitting. For example, the quadratic splines can have inflection points only at the knots, so having a "small" number of knots might be too limiting. However, providing a large number of knots increases the degrees of freedom for the unpenalized splines. A penalty term can lower the effective degrees of freedom while providing a flexible fit. Hence, penalized spline functions allow for "many" knots while controlling the degrees of freedom with a single tuning parameter λ. For details about unconstrained penalized splines, see Eilers and Marx (1996) and Ruppert, Wand, and Carroll (2003).

To implement a penalty of order q, $q = 1, 2, ...$, we use a penalized sum of squares for the criterion function:

$$\sum_{i=1}^{n} \left[Y_i - \sum_{j=1}^{m} b_j \eta_j(x_i) \right]^2 + \lambda \sum_{j=q+1}^{m} (\Delta^q b_j)^2, \tag{35}$$

where $\Delta^1 b_j = b_j - b_{j-1}$ and $\Delta^q b_j = \Delta^{q-1} b_j$ for $q > 1$. When $q = 1$, the fit gets close to the simple linear regression as λ increases without bound, so the effective degrees of freedom ranges from 2 to the number of basis functions. Similarly, when $q = 2$, the fit converges to the quadratic least-squares curve, etc. The vector form of the criterion is

$$\psi(\boldsymbol{b}) = \boldsymbol{b}^\top(\boldsymbol{B}^\top\boldsymbol{B} + \lambda\boldsymbol{D}^\top\boldsymbol{D})\boldsymbol{b} - 2\boldsymbol{Y}^\top\boldsymbol{B}\boldsymbol{b}, \tag{36}$$

where $\boldsymbol{D}$ is the qth order difference matrix. The unconstrained minimizer of ψ is $\tilde{\boldsymbol{b}}_\lambda = (\boldsymbol{B}^\top\boldsymbol{B} + \lambda\boldsymbol{D}^\top\boldsymbol{D})^{-1}\boldsymbol{B}^\top\boldsymbol{Y}$, and the constrained minimizer $\hat{\boldsymbol{b}}_\lambda$ is found through quadratic programming as described in Section 2.2.

For the unconstrained estimator, the "effective degrees of freedom" (edfu_λ) of the model (see Hastie and Tibshirani (1990), Chapter 5) is the trace of $\boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B} + \lambda\boldsymbol{D}^\top\boldsymbol{D})^{-1}\boldsymbol{B}^\top$. Because the constrained fit is robust to choices of λ, we can simply choose λ so that edfu_λ is a reasonable number such as 8. The constrained estimator then has an effective degrees of freedom $\mathrm{edfc}_\lambda \leq \mathrm{edfu}_\lambda$. See Meyer (2012a) for more details about constrained penalized splines and the edfc. Alternatively, generalized cross validation (GCV) can be used to select a penalty parameter. For a choice of $\lambda \geq 0$, let $\hat{\boldsymbol{\theta}}_{m,\lambda} = \boldsymbol{B}\hat{\boldsymbol{b}}_\lambda$ where $\hat{\boldsymbol{b}}_\lambda$ minimizes (36); then the GCV choice of λ minimizes the criterion:

$$\mathrm{GCV}(\lambda) = \frac{\sum_{i=1}^{n}[Y_i - \hat{\theta}_{m,\lambda,i}]^2}{(1 - \mathrm{edfc}_\lambda/n)^2}. \tag{37}$$

The convergence rates are inherited from the unpenalized version if the penalty term becomes negligible as n grows. The matrix $\boldsymbol{B}^\top\boldsymbol{B}$ is banded with elements on the order of n/k. The elements of $\boldsymbol{D}^\top\boldsymbol{D}$ do not grow with n, so if the number k of knots grows at the same rate as for the unpenalized version, we require that λ grow at a slower rate than $n^{2q/(2q+1)}$.

3.3. Extensions

Two simple extensions greatly increase the utility of these methods in practice.

3.3.1. Extensions to Heteroskedastic and Correlated Error Models. Assume $\mathrm{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$ for some symmetric and positive-definite matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\Sigma}$ is known, as in the case of weighted regression, we can readily transform the model to the *i.i.d.* case, by multiplying the regression equation $\boldsymbol{Y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$ through by $\boldsymbol{\Sigma}^{-1/2}$, to get $\check{\boldsymbol{Y}} = \check{\boldsymbol{\theta}} + \sigma\check{\boldsymbol{\varepsilon}}$, where $\mathrm{cov}(\check{\boldsymbol{\varepsilon}}) = \boldsymbol{I}$. The vector $\check{\boldsymbol{\theta}}$ is estimated by a linear combination of transformed B-spline basis vectors, where $\check{\boldsymbol{\eta}}_j$, $j = 1, \ldots, \ell$, are the columns of $\check{\boldsymbol{B}} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{B}$. Then the results of the previous sections follow.

For time-series data, we can often assume auto-regressive errors of degree p; for example AR(1) errors follow the pattern $\varepsilon_{i+1} = \phi\varepsilon_i + \xi_i$, where ξ_i's are independent. For the case where the covariance parameters are unknown, Wang, Meyer, and Opsomer (2013) proved an oracle property showing that the convergence rate of the spline fit, with estimated auto-regression parameters, is the same as for the known covariance case. Hence the estimates of the auto-regression parameters are themselves consistent. They determine p with a modified AIC criterion, and estimate the parameters and the regression function with Cochran-Orcutt-type iterations. These results apply directly to our change-point regression model with auto-regressive errors.

3.3.2. Change-Point Models with Covariates. For the partial linear model with parameter vector $\boldsymbol{\alpha} \in \mathbb{R}^p$ and covariate vectors $\boldsymbol{z}_i$, $i = 1, \ldots, n$, consider the additive model

$$Y_i = f_m(X_i) + \boldsymbol{z}_i^\top \boldsymbol{\alpha} + \sigma\varepsilon_i, \quad \text{for} \quad i = 1, \ldots, n.$$

If $\boldsymbol{Z}$ is the $n \times p$ matrix whose rows are $\boldsymbol{z}_i$, $i = 1, \ldots, n$, then we can model the expected value of $\boldsymbol{Y}$ as $\boldsymbol{B}\boldsymbol{b} + \boldsymbol{Z}\boldsymbol{\alpha}$. Writing $\tilde{\boldsymbol{B}} = [\boldsymbol{B}|\boldsymbol{Z}]$ and $\boldsymbol{\beta}^\top = [\boldsymbol{b}^\top|\boldsymbol{\alpha}^\top]$, we can define a constraint matrix $\boldsymbol{S}_c = [\boldsymbol{S}|\boldsymbol{0}]$. Assuming that the columns of $\tilde{\boldsymbol{B}}$ are linearly independent, the method of Section 2.2 can be used to minimize $\|\boldsymbol{Y} - \tilde{\boldsymbol{B}}\boldsymbol{\beta}\|^2$ subject to $\boldsymbol{S}_c\boldsymbol{\beta} \geq \mathbf{0}$. In this way the model with covariates is fit with a single step (without back-fitting), and the convergence rate results for $\hat{m}$ and $\hat{f}_{\hat{m}}$ follow.

To add covariates to the penalized spline model, we define the $(m+p) \times (m+p)$ matrix

$$\boldsymbol{D}_c = \begin{bmatrix} \boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Then we minimize $\psi(\boldsymbol{\beta}) = \boldsymbol{\beta}^\top(\tilde{\boldsymbol{B}}^\top\tilde{\boldsymbol{B}} + \lambda\boldsymbol{D}_c^\top\boldsymbol{D}_c)\boldsymbol{\beta} - 2\boldsymbol{Y}^\top\tilde{\boldsymbol{B}}\boldsymbol{\beta}$, subject to $\boldsymbol{S}_c\boldsymbol{\beta} \geq \mathbf{0}$, using the same quadratic programming method.

CHAPTER 4

Simulation and Examples

4.1. Simulation Studies

We compare the performance of the proposed change-point estimator with some other established estimators using the square root of mean squared error (SMSE) criterion:

$$\text{SMSE} = \left[\frac{1}{N}\sum_{i=1}^{N}|\hat{m}_i - m|^2\right]^{1/2},$$

where N is the number of simulations.

4.1.1. Unimodal Case. We compare our estimator $\hat{m}_S$ with the estimator $\hat{m}_P$ of Shoung and Zhang (2001) and the estimator $\hat{m}_U$ of Köllmann et al. (2014). We choose two unimodal functions. The first is $f_m = 6x(1-x)$ with $m = .5$, and the second is $f_m = 30x^4(1-x)$ with $m = .8$. For each function, we use three sample sizes and three standard deviations. X_i's are equally spaced on $[0,1]$. To get $\hat{m}_U$, we use the default choice of knots in the `R` package `uniReg` Köllmann (2014), i.e., 2 exterior knots are placed at the endpoints of the interval and 10 interior knots are equally placed between the 2 knots, and we also use the default choice of penalty in this package, for which the difference penalty of order 2 is used and the tuning parameter is chosen via restricted maximum likelihood; to get $\hat{m}_S$, we use the constrained penalized method with the same knots. We choose $N = 10,000$ to get $\hat{m}_S$ and $\hat{m}_P$, and $N = 1000$ to get $\hat{m}_U$ due to its speed. Results are shown in Figure 3 (Table 1 and Table 2), with examples of the estimators shown in Figure 4. The proposed estimator has a bigger advantage when the function is more peaked or the variance is smaller,

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{P}}$			$\hat{m}_{\mathrm{U}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.069	.103	.163	.104	.139	.182	.077	.088	.157
200	.055	.085	.128	.092	.123	.165	.063	.087	.091
500	.039	.065	.096	.079	.104	.139	.047	.069	.084

TABLE 1. SMSE for the estimators $\hat{m}_{\mathrm{S}}$ (proposed), $\hat{m}_{\mathrm{P}}$ and $\hat{m}_{\mathrm{U}}$ simulated from $f_m(x) = 6x(1-x)$ with $m = .5$.

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{P}}$			$\hat{m}_{\mathrm{U}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.019	.031	.059	.050	.070	.105	.020	.032	.088
200	.016	.023	.042	.044	.061	.087	.017	.028	.051
500	.013	.018	.028	.037	.051	.071	.014	.024	.033

TABLE 2. SMSE for the estimators $\hat{m}_{\mathrm{S}}$ (proposed), $\hat{m}_{\mathrm{P}}$ and $\hat{m}_{\mathrm{U}}$ simulated from $f_m(x) = 30x^4(1-x)$ with $m = .8$.

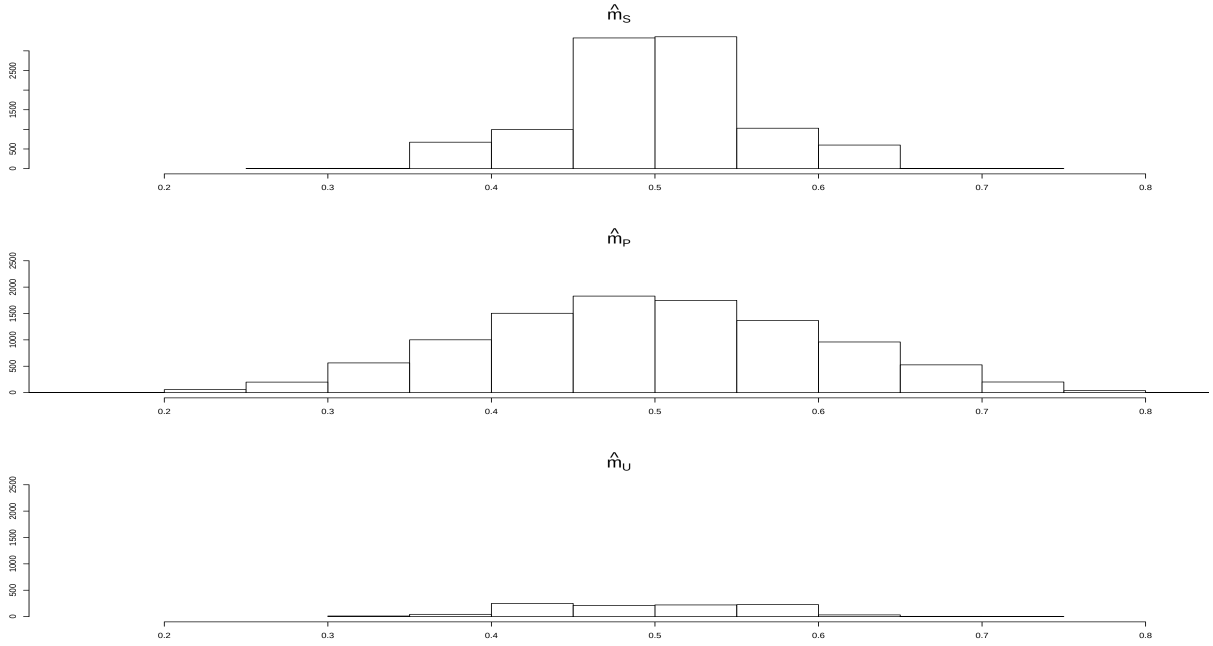

FIGURE 1. $n = 500$, $\sigma = 1$, and $f_m(x) = 6x(1-x)$ with $m = .5$. From top to bottom are the histograms of $\hat{m}_S$, $\hat{m}_P$ and $\hat{m}_U$.

and both spline estimators have considerably smaller SMSE, compared to the mode estimator without smoothing.

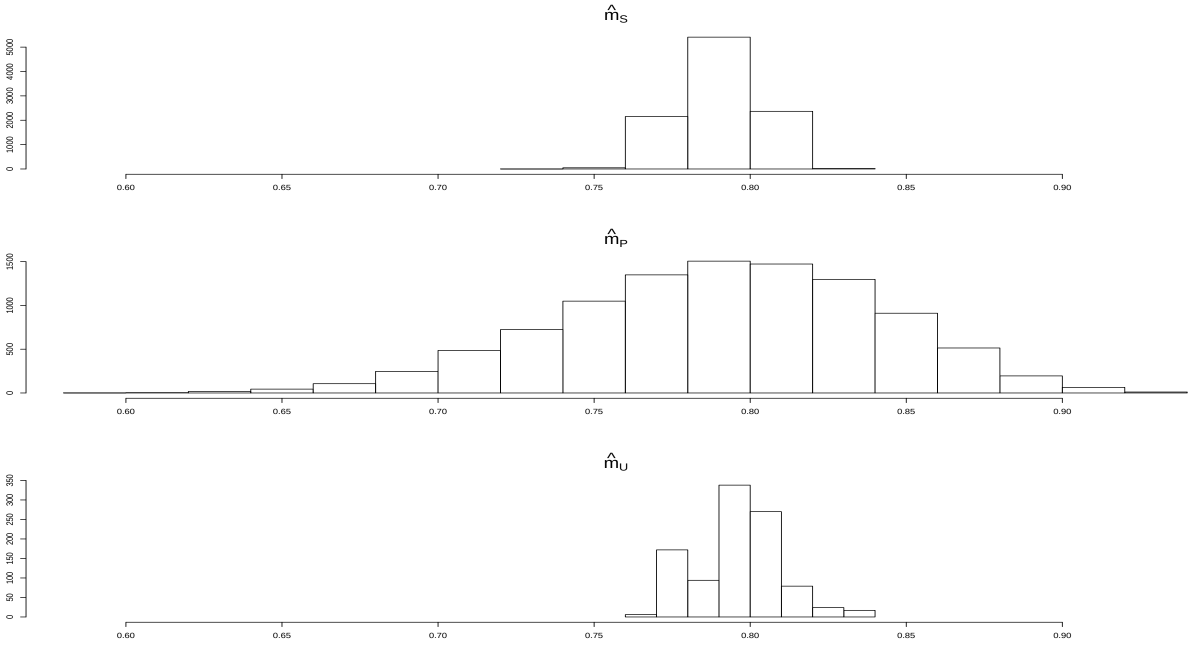

FIGURE 2. $n = 500$, $\sigma = 1$, and $f_m(x) = 30x^4(1 - x)$ with $m = .8$. From top to bottom are the histograms of $\hat{m}_S$, $\hat{m}_P$ and $\hat{m}_U$.

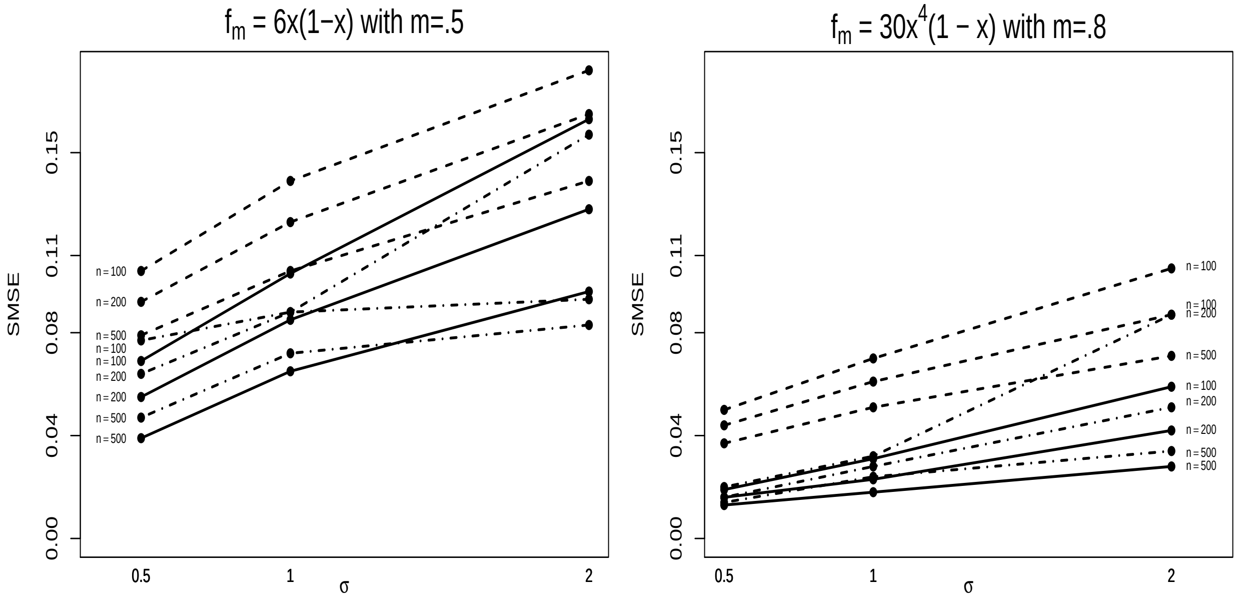

FIGURE 3. Comparisons of $\hat{m}_S$ (solid) with $\hat{m}_P$ (dashed) and $\hat{m}_U$ (dot-dash).

4.1.2. Inflection-Point Case. The method used in Kachouie and Schwartzman (2013) is local polynomial regression. We choose the degree p of the local polynomials to be 3 in the following simulation, which will be applied to construct a confidence interval for

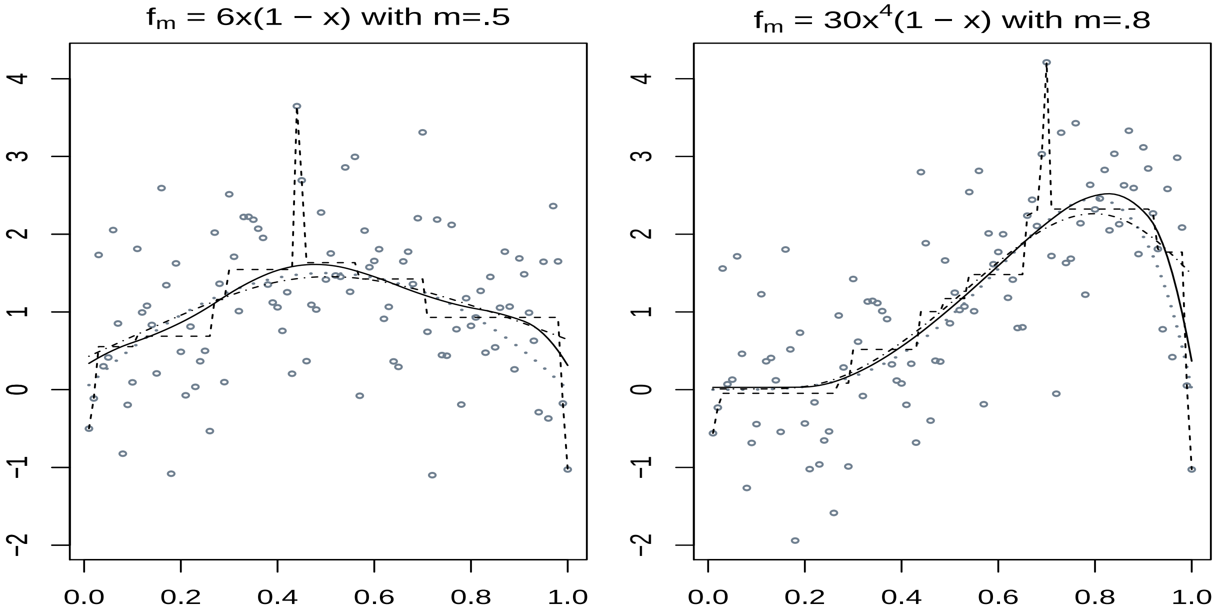

FIGURE 4. Examples of fits for simulated data with $n = 100$. In each case, ε_i's are *i.i.d.* normal with zero mean and unit variance. The dotted line is f_m. Left: $\hat{m}_S = .48$ (solid), $\hat{m}_P = .44$ (dash) and $\hat{m}_U = .51$ (dot-dash). Right: $\hat{m}_S = .83$ (solid), $\hat{m}_P = .7$ (dash) and $\hat{m}_U = .8$ (dot-dash).

the inflection-point in Section 4.3 as discussed in Kachouie and Schwartzman (2013). In Kachouie and Schwartzman (2013), when the authors choose the bandwidth h, they do not use the cross validation score, since this may result in multiple inflection-point estimates. Instead, they choose a h such that the set of the zero down-crossings of the second derivative has only one element to ensure that the smoothed curve has only one inflection point. The authors propose selecting the smallest h within the range that guarantees only one inflection point. To get their proposed h, we try searching for such a h within the range $[.05, .5]$ using an increment of length .02. The smallest h in this range that guarantees that there is only one inflection point is the h to be used. The R package `inflection` Christopoulos (2013) estimates an inflection point by the routine `findiplist`. It can estimate an inflection point by two methods: *extremum surface estimator* (ESE) and *extremum distance estimator* (EDE). We use the ESE method, which is the default method in `inflection`, in the following

n	$\hat{m}_{\mathrm{S}}$ $\sigma = .5$	 $\sigma = 1$	 $\sigma = 2$	$\hat{m}_{\mathrm{K}}$ $\sigma = .5$	 $\sigma = 1$	 $\sigma = 2$	$\hat{m}_{\mathrm{C}}$ $\sigma = .5$	 $\sigma = 1$	 $\sigma = 2$
100	.021 (.022)	.025 (.026)	.040 (.039)	.033	.034	.037	.057	.072	.091
200	.020 (.021)	.022 (.023)	.030 (.031)	.032	.033	.034	.057	.073	.091

Table 3. SMSE for the estimators $\hat{m}_{\mathrm{S}}$ (proposed), $\hat{m}_{\mathrm{K}}$ and $\hat{m}_{\mathrm{C}}$ simulated from $f_m(x) = 10e^{-e^{5(1-2x)}}$ with $m = .5$.

simulation. Note that the `R` package `inflection` does not provide an estimate of f_m and Kachouie and Schwartzman (2013) does.

We choose two convex-concave curves to compare our estimator $\hat{m}_S$ with the estimator $\hat{m}_K$ of Kachouie and Schwartzman (2013) and the estimator $\hat{m}_C$ in the `R` package `inflection` through simulations. The first curve is the Gompertz sigmoid curve $f_m = 10e^{-e^{5(1-2x)}}$, and the second is $f_m = 5\{1+tanh[10(x-.5)]\}$, which is a multiple of the cumulative distribution function of a logistic distribution. Because both functions also satisfy the monotonicity constraint, we also present results for the increasing convex-concave spline estimator. Both curves have inflection-point $m = .5$ and are discussed in , where it is mentioned that the Gompertz sigmoid curve has applications in economics and other disciplines. To compare the estimators, we choose two sample sizes and three standard deviations, with X_i's equally spaced on $[0, 1]$. The constrained penalized method is used for $\hat{m}_S$, with 2 exterior knots placed at the endpoints of the interval and 10 interior knots equally placed between the 2 knots. Simulation results are shown in Figure 7, and some example fits are shown in Figure 8 (Table 3 and Table 4). In the tables below, the SMSE of $\hat{m}_S$ with the monotonicity constraint is included in the bracket.

4.1.3. Jump-Point Case. In this case, we choose the functions $f_m(x) = 4sin(5x) + 3x + I_{[.7,1]}(x)$ and $f_m(x) = x^4 + I_{[.5,1]}(x)$ to compare $\hat{m}_S$ with $\hat{m}_L$ of Loader (1996) and $\hat{m}_G$

	$\hat{m}_{\rm S}$			$\hat{m}_{\rm K}$			$\hat{m}_{\rm C}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.004 (.003)	.008 (.007)	.020 (.018)	.003	.006	.012	.025	.042	.074
200	.002 (.002)	.005 (.004)	.012 (.011)	.002	.004	.008	.024	.040	.074

TABLE 4. SMSE for the estimators $\hat{m}_{\rm S}$ (proposed), $\hat{m}_{\rm K}$ and $\hat{m}_{\rm C}$ simulated from $f_m(x) = 5\{1 + tanh[10(x - .5)]\}$ with $m = .5$.

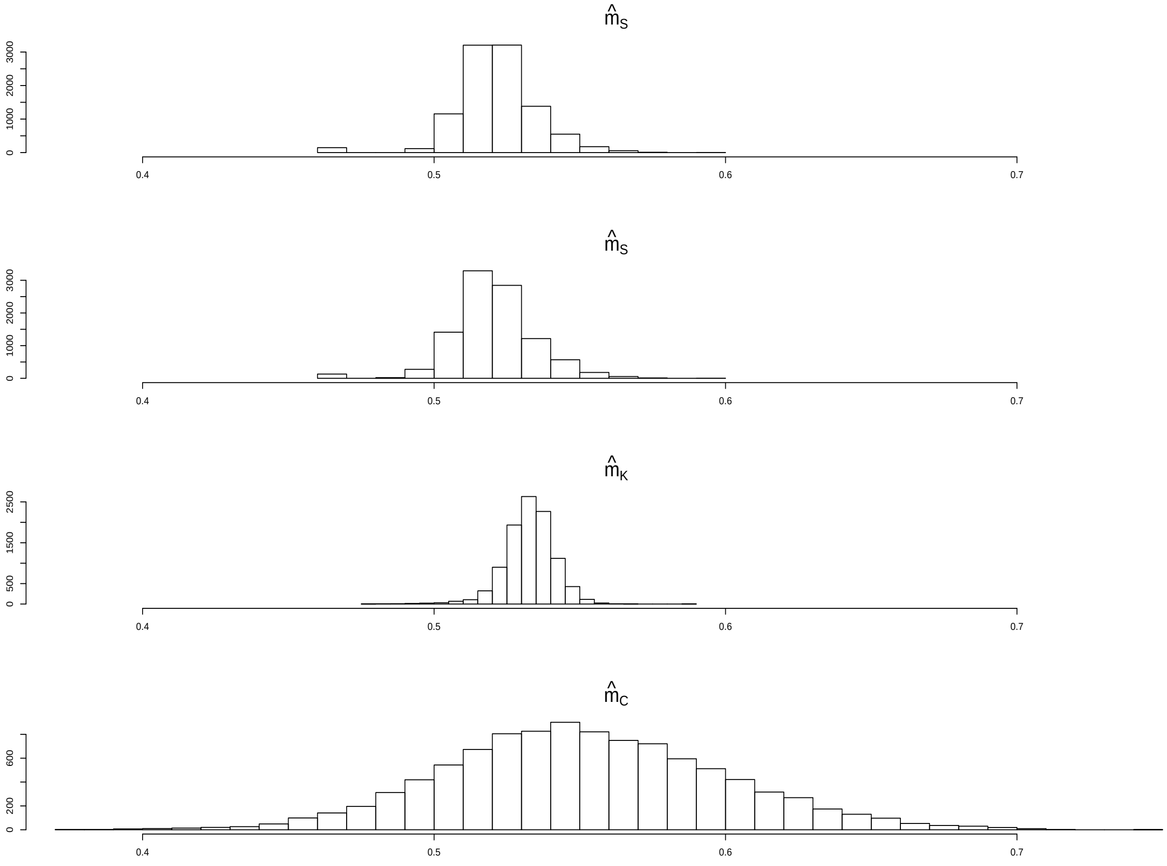

FIGURE 5. $n = 100$, $\sigma = 1$, and $f_m(x) = 10e^{-e^{5(1-2x)}}$ with $m = .5$. From top to bottom are the histograms of $\hat{m}_S$ assuming that f_m is increasing over the whole interval, $\hat{m}_S$ without this assumption, $\hat{m}_K$ and $\hat{m}_C$.

of Grégoire and Hamrouni (2001) by simulation. The first function is discussed in both papers. In Loader (1996), $\hat{m}_L$ is the design point $x_i, i = 1, \ldots, n$, which maximizes the difference between the right and left limits at the point based on one-sided nonparametric local polynomial regression of degree p. The author discusses the estimator when $p = 0$ or 1.

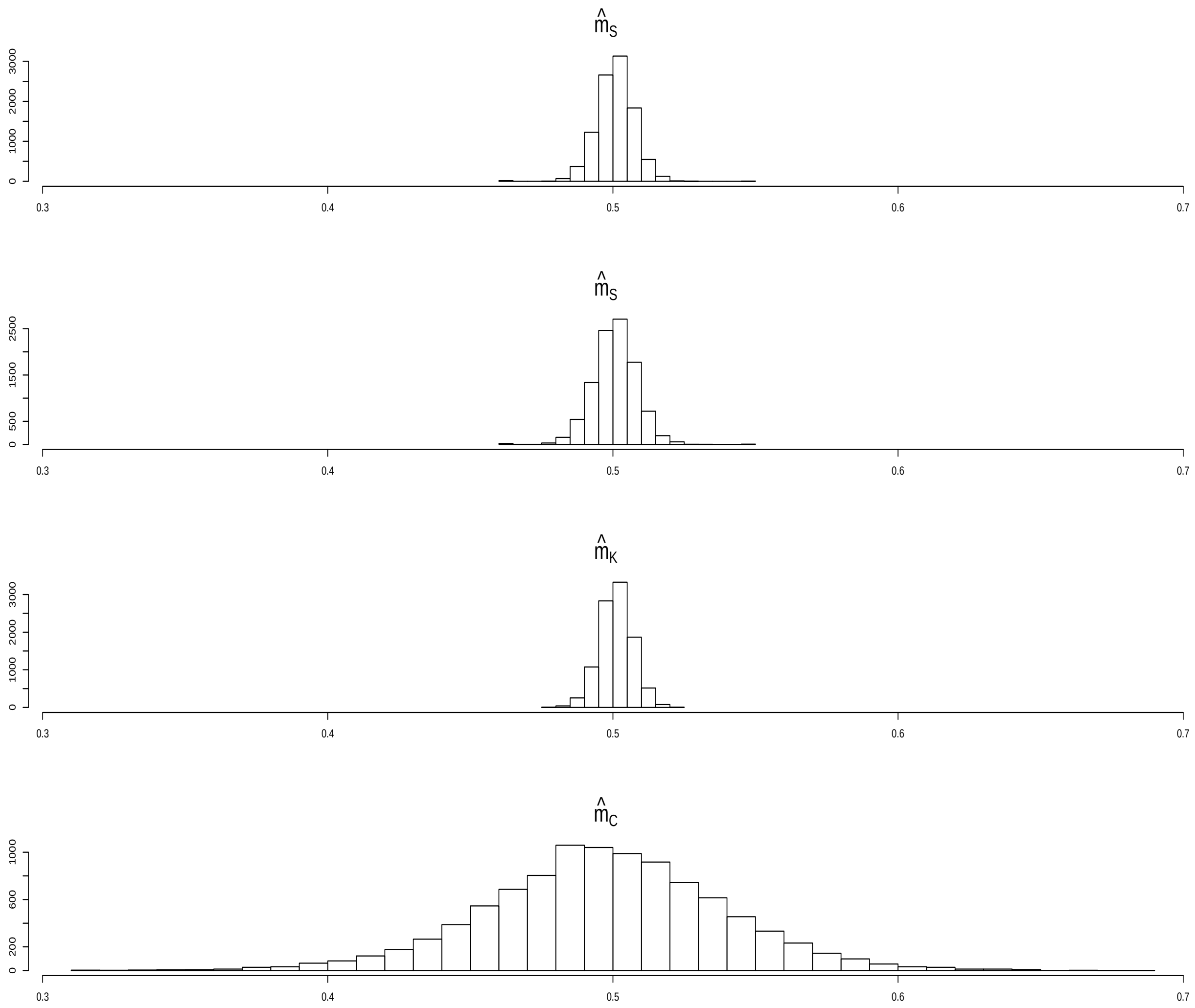

FIGURE 6. $n = 100$, $\sigma = 1$, and $f_m(x) = 5\{1 + tanh[10(x - .5)]$ with $m = .5$. From top to bottom are the histograms of $\hat{m}_S$ assuming that f_m is increasing over the whole interval, $\hat{m}_S$ without this assumption, $\hat{m}_K$ and $\hat{m}_C$.

It is mentioned that the choice of the degree of local polynomials have little impact on the asymptotic results for $\hat{m}_L$, but local linear fitting is suggested to avoid the "boundary problem". So we use local linear fitting to get the following simulation results. In Grégoire and Hamrouni (2001), $\hat{m}_G$ is the design point $x_i, i = 1, \ldots, n$, which maximizes the difference between the right and left limits at the point based on local linear smoothing with a symmetric kernel function defined on $[-1, 1]$. In both papers, the optimal bandwidths include $.15, .17,$

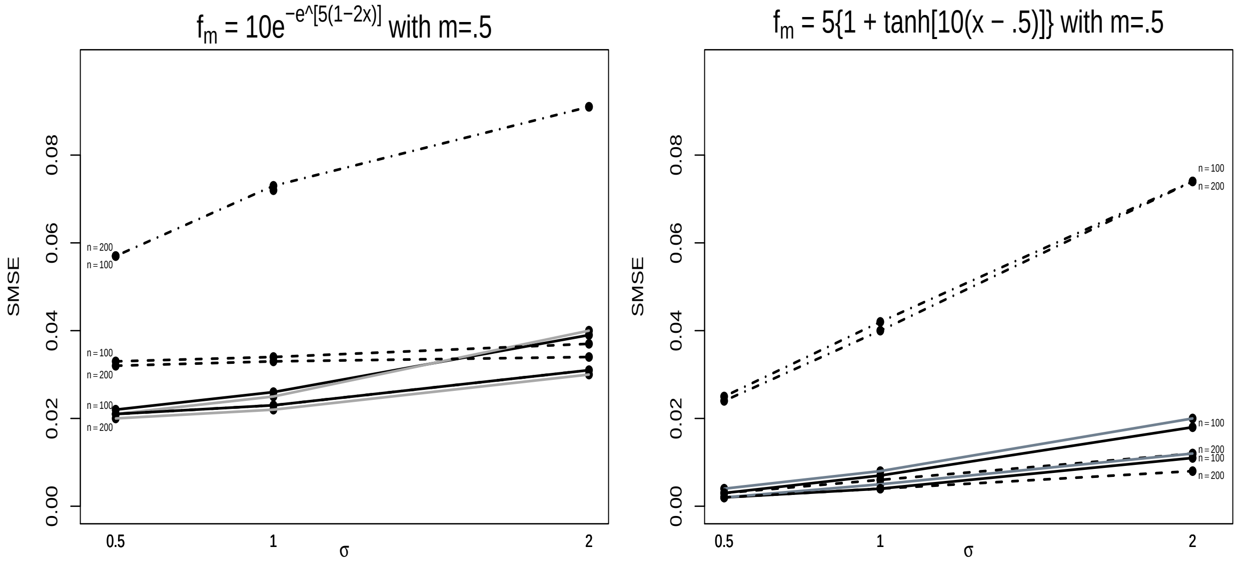

FIGURE 7. Comparisons of $\hat{m}_S$ (solid) with $\hat{m}_K$ (dashed) and $\hat{m}_C$ (dot-dash).

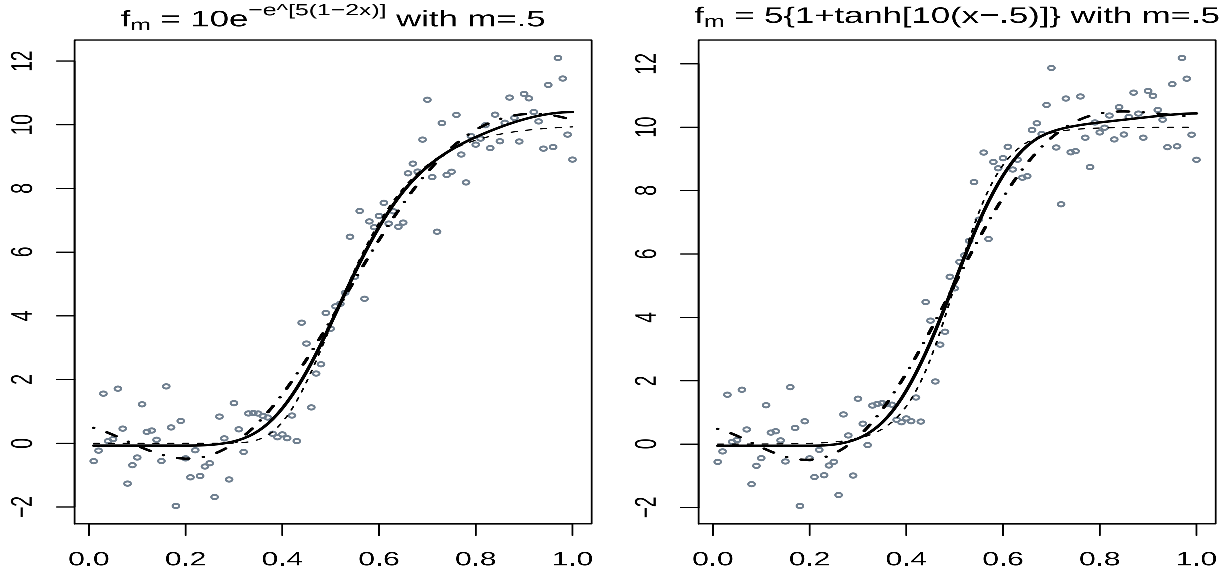

FIGURE 8. Examples of fits for $n = 100$. In each case, ε_i's are *i.i.d.* normal with zero mean and unit variance. Left: $\hat{m}_{\mathrm{S}} = .516$ (solid), $\hat{m}_{\mathrm{K}} = .536$ (dot-dash) and $\hat{m}_C = .565$. Right: $\hat{m}_{\mathrm{S}} = .499$ (solid), $\hat{m}_{\mathrm{K}} = .501$ (dot-dash) and $\hat{m}_C = .48$. There is no estimate of f_m in the R package `inflection`.

and .19, and we choose the bandwidth to be .15 to make the comparison. To compare the estimators, we choose three sample sizes and two standard deviations. To apply our method, we relax the jump direction constraint and we only constrain that f_m is non-increasing in

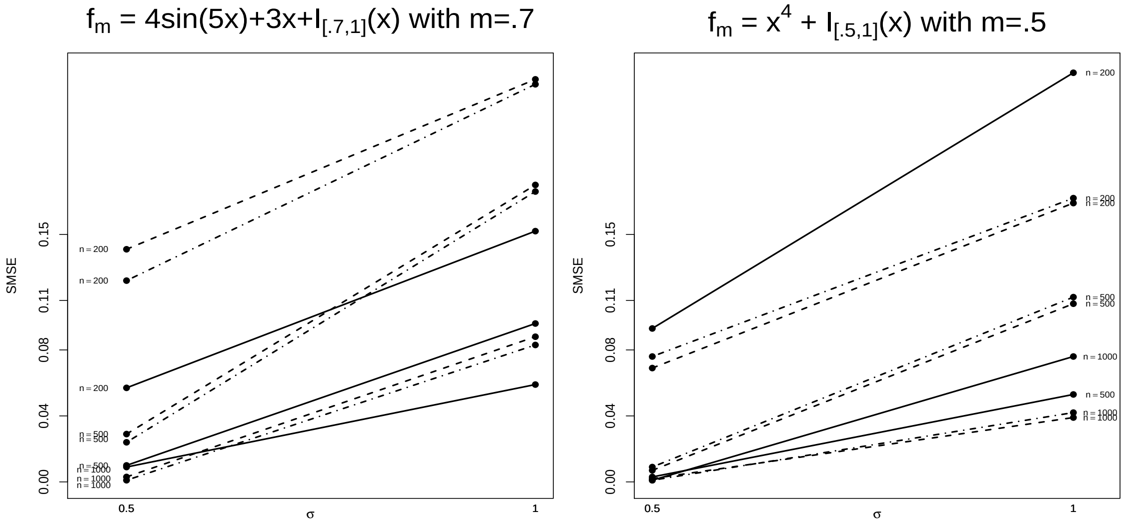

FIGURE 9. Comparisons of $\hat{m}_S$ (solid) with $\hat{m}_L$ (dashed) and $\hat{m}_G$ (dot-dash).

	$\hat{m}_{\mathrm{S}}$		$\hat{m}_{\mathrm{L}}$		$\hat{m}_{\mathrm{G}}$	
n	$\sigma = .5$	$\sigma = 1$	$\sigma = .5$	$\sigma = 1$	$\sigma = .5$	$\sigma = 1$
200	.057	.152	.141	.244	.122	.241
500	.010	.096	.029	.180	.024	.176
1000	.009	.059	.003	.088	.001	.083

TABLE 5. SMSE for the estimators $\hat{m}_{\mathrm{S}}$ (proposed), $\hat{m}_{\mathrm{L}}$ and $\hat{m}_{\mathrm{G}}$ simulated from $f_m(x) = 4sin(5x) + 3x + I_{[.7,1]}(x)$ with $m = .7$.

	$\hat{m}_{\mathrm{S}}$		$\hat{m}_{\mathrm{L}}$		$\hat{m}_{\mathrm{G}}$	
n	$\sigma = .5$	$\sigma = 1$	$\sigma = .5$	$\sigma = 1$	$\sigma = .5$	$\sigma = 1$
200	.093	.248	.069	.169	.076	.172
500	.003	.053	.007	.108	.009	.112
1000	.001	.076	.002	.039	.001	.042

TABLE 6. SMSE for the estimators $\hat{m}_{\mathrm{S}}$ (proposed), $\hat{m}_{\mathrm{L}}$ and $\hat{m}_{\mathrm{G}}$ simulated from $f_m(x) = x^4 + I_{[.5,1]}(x)$ with $m = .5$.

$[t_p, m) \cup (m, t_{p+1}] \subset [0,1]$, where t_p and t_{p+1} are two consecutive knots. Results are in Figure 9 (Table 5 and Table 6). Neither Loader (1996) or Grégoire and Hamrouni (2001) provides an estimate of f_m, and we include two example plots of $\hat{f}_{\hat{m}_S}$ in Figure 10 and Figure 11. It is mentioned in Loader (1996) that for $f_m(x) = 4sin(5x) + 3x + I_{[.7,1]}(x)$ with $n = 1000$ observations and residuals which are *i.i.d.* normal with mean zero and unit variance, "the

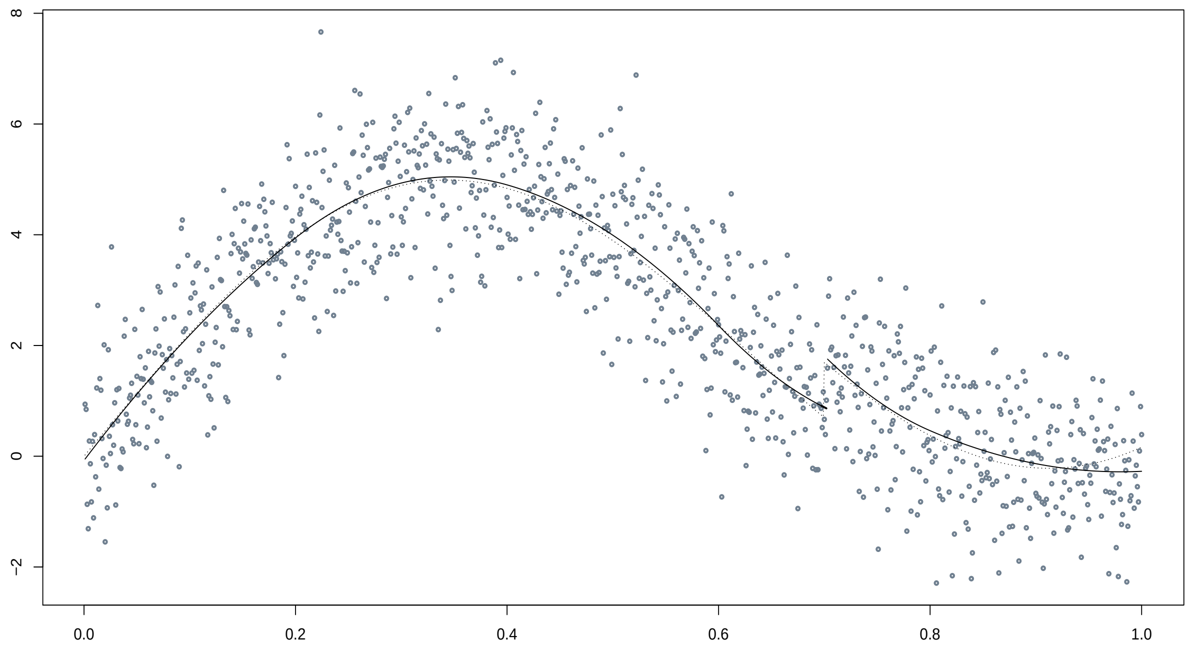

FIGURE 10. Regression estimates for $n = 1000$ independent observations simulated from $f_m(x) = 4sin(5x) + 3x + I_{[.7,1]}(x)$ (dot-dash) with errors which are *i.i.d.* normal with mean zero and unit variance. The solid curve is $\hat{f}_{\hat{m}_S}$ with $\hat{m}_S = .690$.

change is nearly impossible to detect by eye". The following plot is an example showing $\hat{m}_S$ and $\hat{f}_{\hat{m}_S}$ in this scenario.

4.1.4. Choosing p for the AR(p) Model. We generate data from a unimodal regression function with errors from a stationary autoregressive process, to assess the AIC choice of p. The errors are $AR(1)$ with autoregressive coefficient $\phi = .4$; that is, $\varepsilon_i = \phi\varepsilon_{i-1} + \xi_i$, where ξ_i's are *i.i.d.* normal with zero mean and the standard deviation $\sigma = .2$. For $f_m = 6x(1-x)$ and X_i values equally spaced on $[0, 1]$, and five equally spaced knots, three sample sizes are used to calculate the proportion that the estimator chooses the true p, i.e., $p = 1$, for choice $p = 0$, 1, or 2. In Table 7, results from $N = 10,000$ simulated data sets are given.

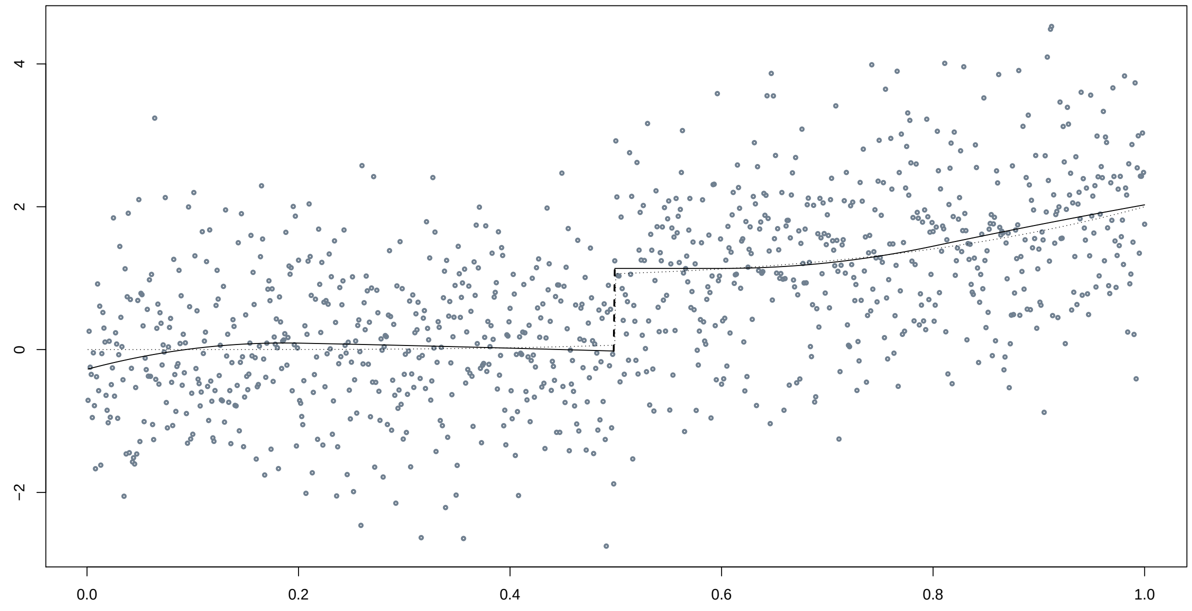

FIGURE 11. Regression estimates for $n = 1000$ independent observations simulated from $f_m(x) = x^4 + I_{[.5,1]}(x)$ (dot-dash) with errors which are *i.i.d.* normal with mean zero and unit variance. The solid curve is $\hat{f}_{\hat{m}_S}$ with $\hat{m}_S = .498$.

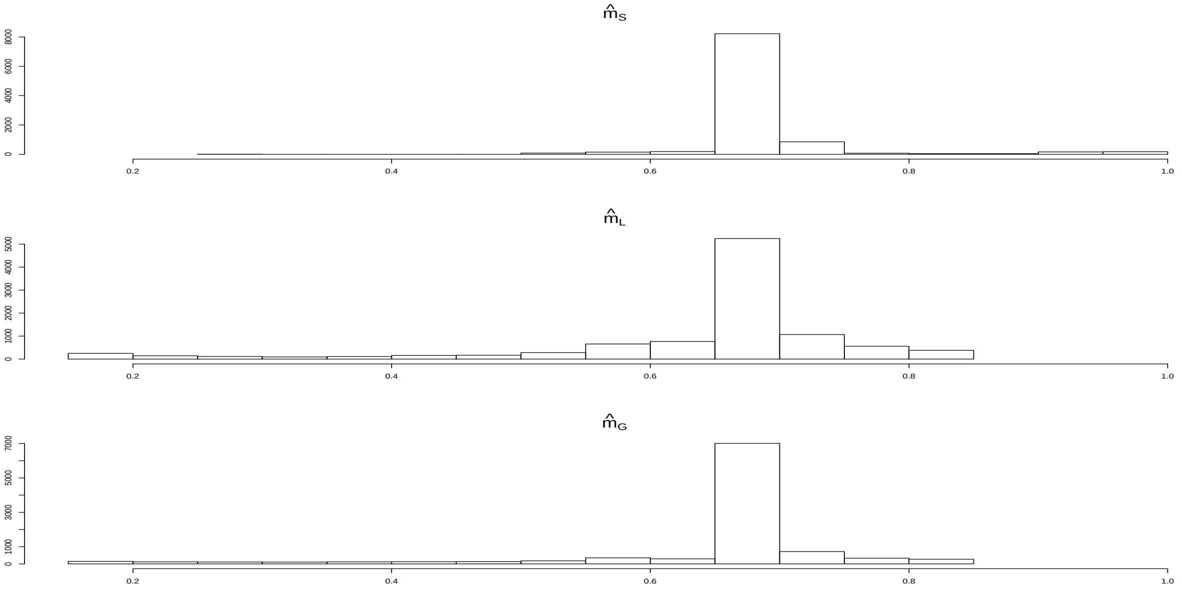

FIGURE 12. $n = 200$, $\sigma = .5$, and $f_m(x) = 4sin(5x) + 3x + I_{[.7,1]}(x)$ with $m = .7$. From top to bottom are the histograms of $\hat{m}_S$, $\hat{m}_L$ and $\hat{m}_G$

.

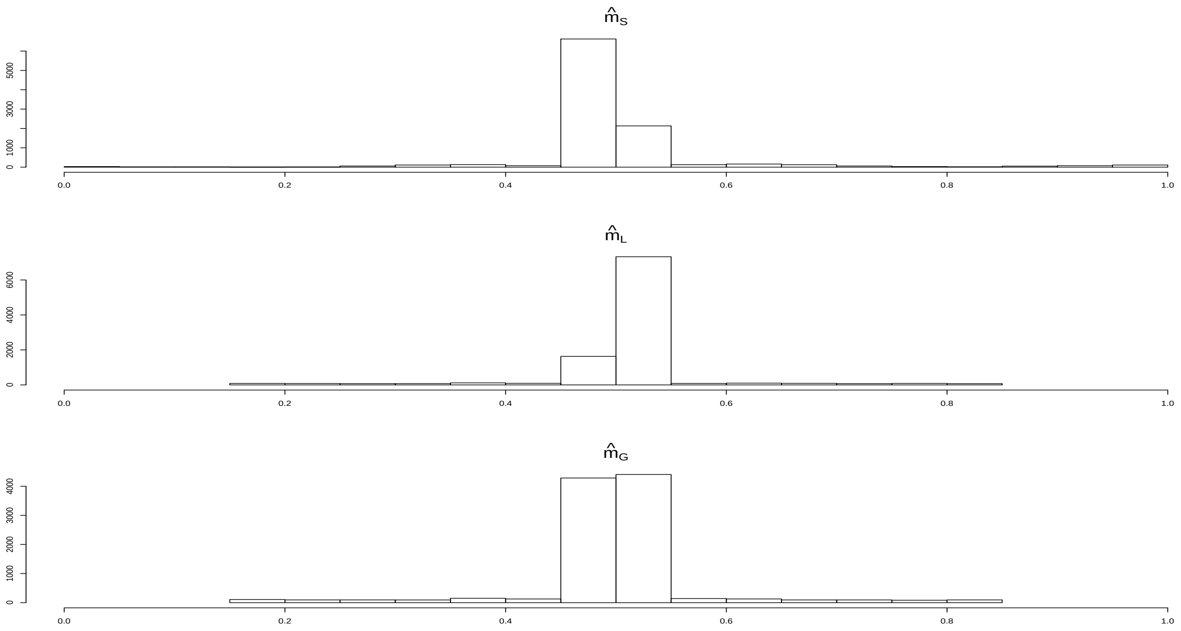

FIGURE 13. $n = 200$, $\sigma = .5$, and $f_m(x) = x^4 + I_{[.5,1]}(x)$ with $m = .5$. From top to bottom are the histograms of $\hat{m}_S$, $\hat{m}_L$ and $\hat{m}_G$

.

	$\hat{p}$			$\hat{\sigma}^2$ ($\hat{p} = 1$)		$\hat{\phi}$ ($\hat{p} = 1$)	
n	0	1	2	mean	SD	mean	SD
100	.000	**.754**	.246	.037	.005	.307	.098
200	.000	**.799**	.201	.039	.004	.353	.067
500	.000	**.828**	.172	.039	.002	.382	.042

TABLE 7. For each of three sample sizes, proportions of datasets for which $\hat{p} = 0, 1$, or 2 are shown, with the mean and standard deviation of the estimated model variance and autocorrelation parameter estimates, when the choice of p is correct.

4.2. Examples

4.2.1. Ethanol Data Set. The **ethanol** data set in the R package SemiPar Wand (2014) contains three variables: the concentration of nitric oxide and nitrogen dioxide in engine exhaust (NOx), the richness of the air-to-ethanol mix (E) and the compression ratio of the engine (C). NOx is formed during combustion of ethanol and there is a relationship between the air-to-ethanol ratio and the completeness of combustion. The observed C has

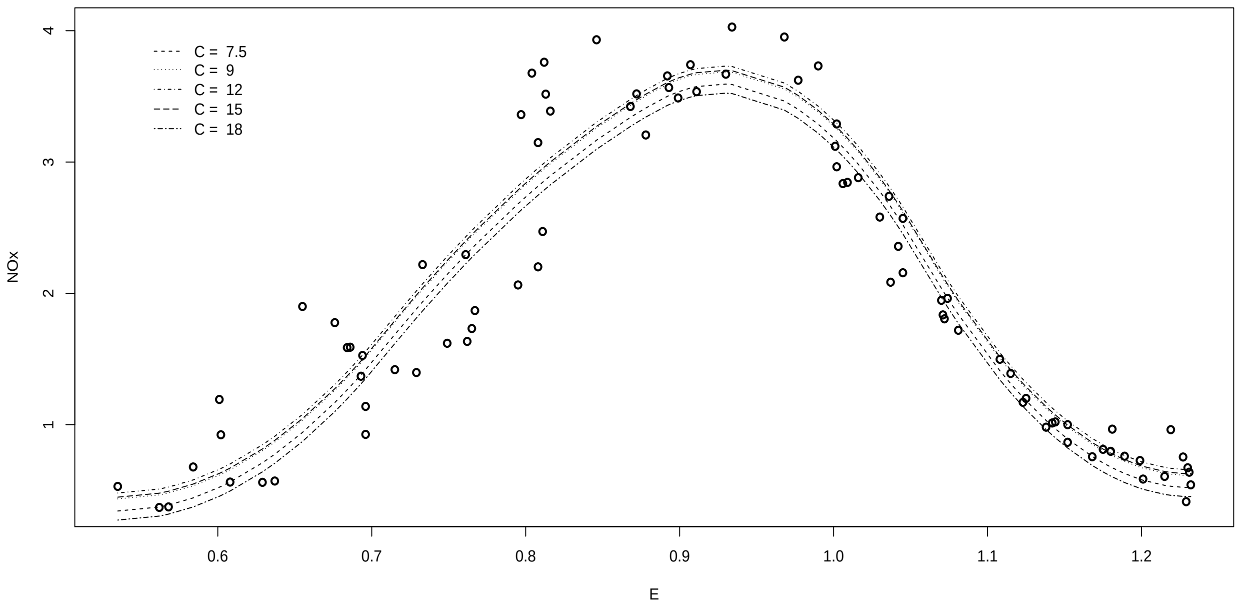

Figure 14. The estimated curve between NOx and E at each level of C is shown in this plot. The estimated mode is at .927.

five levels, and we assume that expected NOx, as a function of E, has mode m at each level of C. In Figure 14, it is revealed that when E is close to 1, combustion is complete and the NOx formed reaches its maximum.

4.2.2. Trade.union Data Set. The `trade.union` data set in the `R` package `SemiPar` Wand (2014) contains data on 534 U.S. workers, which includes workers' age and wage. There are six occupations, and we assume that the expected logarithm of wage, as a function of age, has mode m. We also include gender and race as another two categorical covariates. We find that the logarithm of wage, on average, rises quickly to a peak when a worker is aged around 39, and starts decreasing slowly afterwards. The fits for black male workers are shown in Figure 15, with curves for the six occupations. The curves for Hispanic workers are .125 units lower while those for white workers are .074 units higher; the estimated curves for female workers are .251 units lower. We use the constrained penalized estimator with 12

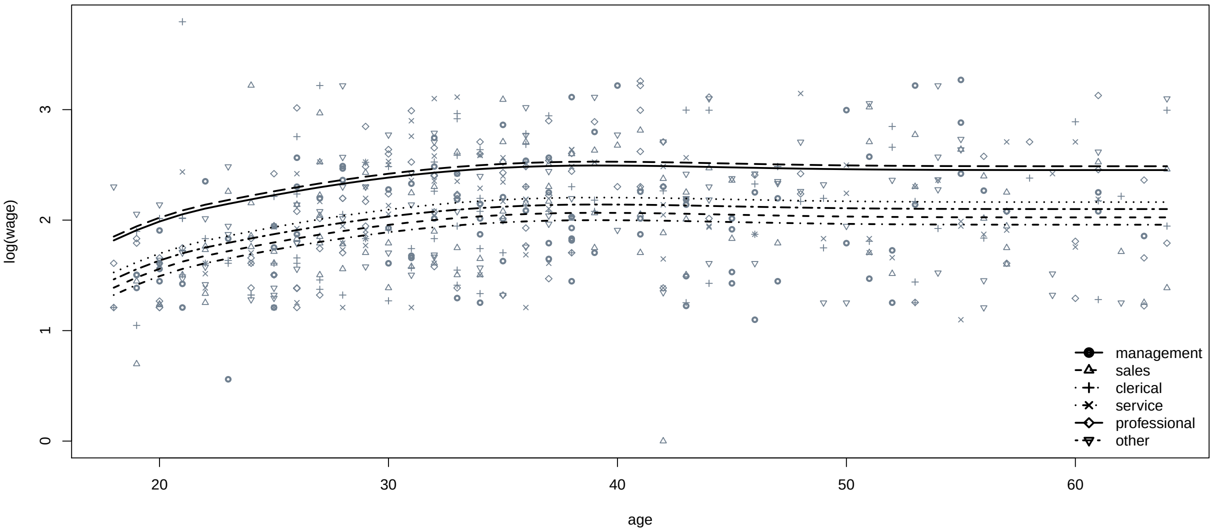

Figure 15. Fitted unimodal curves for log(wage) as a function of age for six occupations, for one of the ethnicity/gender combinations.

equally spaced knots, and choose λ with the GCV method. A range of penalty parameters is used, such that the edfu_λ is an integer from 12 to 21 (including eight degrees of freedom for the covariates), and the GCV choice corresponds to $\text{edfu}_\lambda = 13$.

4.2.3. World Population Growth Rate. We next look at the `world population growth rate` from 1950 to 2015, which is believed to be unimodal. This data set comes from the U.S. Census Bureau at `https://www.census.gov/population/international/data/worldpop/table_population.php`. We assume a stationary autoregressive process with $p = 0$, 1 or 2; the AIC method of Wang et al. (2013) chooses $p = 2$ and $\hat{\boldsymbol{\phi}} = (1.064, -.570)$. The estimated curve is shown in Figure 16, where $\hat{m}$ is around 1964.

4.2.4. $\mathbf{CO_2}$ Emission. We consider the `CO`$_2$ `emissions` (metric tons per capita) in Ghana, Nepal and Hong Kong from 1960 to 2011. This data set comes from the World Bank: Data at `http://data.worldbank.org/indicator/EN.ATM.CO2E.PC`. We assume that the mean curve is convex-concave. The estimated curve for each country is shown in Figure 17.

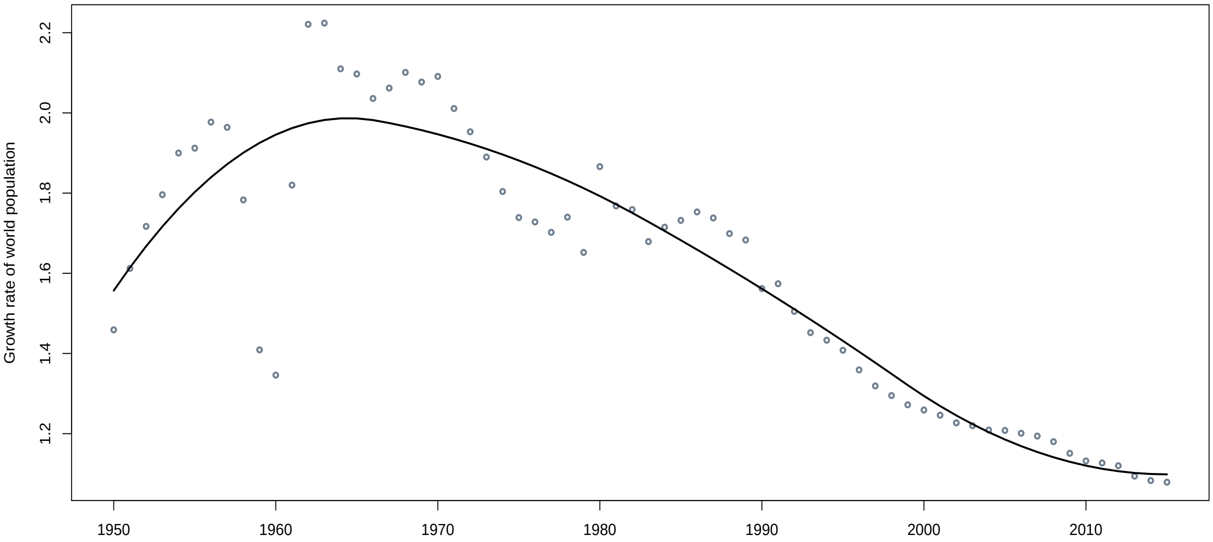

FIGURE 16. The estimated curve of the world population growth rate from 1950 to 2015.

Similarly to the previous example, we have that for Ghana, $p = 1$, $\hat{\boldsymbol{\phi}} = .252$ and $\hat{m}$ is around 1996; for Nepal, $p = 1$, $\hat{\boldsymbol{\phi}} = .491$ and $\hat{m}$ is around 1995; for Hong Kong, $p = 2$, $\hat{\boldsymbol{\phi}} = (.615, -.476)$ and $\hat{m}$ is around 1983. It is suggested that although the CO_2 emissions have been increasing since 1960 in Ghana and Nepal, the increasing rate started decreasing since mid-1990s. Meanwhile, the increasing rate started decreasing since early 1980s in Hong Kong and the CO_2 emissions started decreasing in late 2000s.

4.2.5. Nile River Flow. We consider the Nile river data set, which is about the annual volume of the Nile river from 1871 to 1970. Since the British was constructing the Aswan Low Dam from 1898 to 1902, it is reasonable to believe that there is a jump point in the Nile river flow during this period. We apply our method to this data set to detect a jump point. The estimate is at the year 1898, which is shown in Figure 18. This data set is also discussed in Müller (1992), in which $\hat{m}$ is defined as the maximizer of the difference of right-

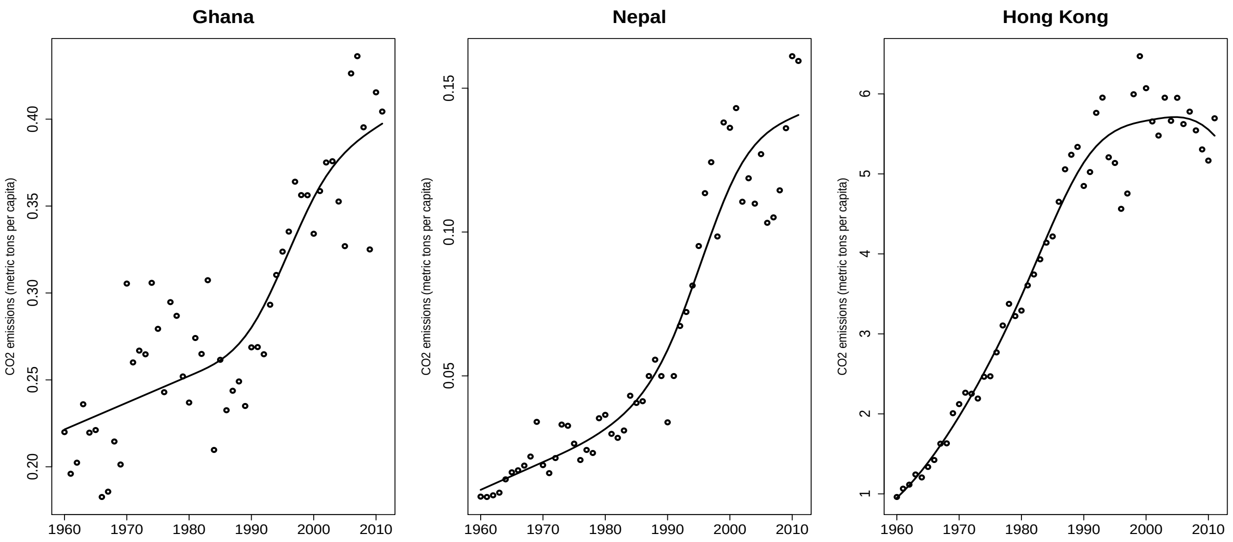

FIGURE 17. The estimated curve of the CO_2 emissions (metric tons per capita) from 1960 to 2011.

and left-sided kernel smoothers, and in Cobb (1978), which applies parametric modelling. Their estimates are the same as our estimate. Again, we use the AIC method of Wang et al. (2013), and it chooses $p = 1$ and $\hat{\boldsymbol{\phi}} = .096$.

4.3. SMSE and Bootstrap Confidence Intervals with Randomly Generated Change-Points

In this section, we discuss the performance of the proposed spline-based estimator $\hat{m}_S$ through simulations using randomly generated change-points on the unit interval. Bootstrap confidence intervals are discussed and we compare the coverage rate of $\hat{m}_S$ and some competitors. Note that none of the jump-point competitors discussed in Section 4.1 estimates f, and we only discuss the unimodal case and the inflection-point case in this section.

It will be shown in the tables below that the B-spline estimator performs better in term of SMSE than its competitors when the change-point is uniformly distributed, which

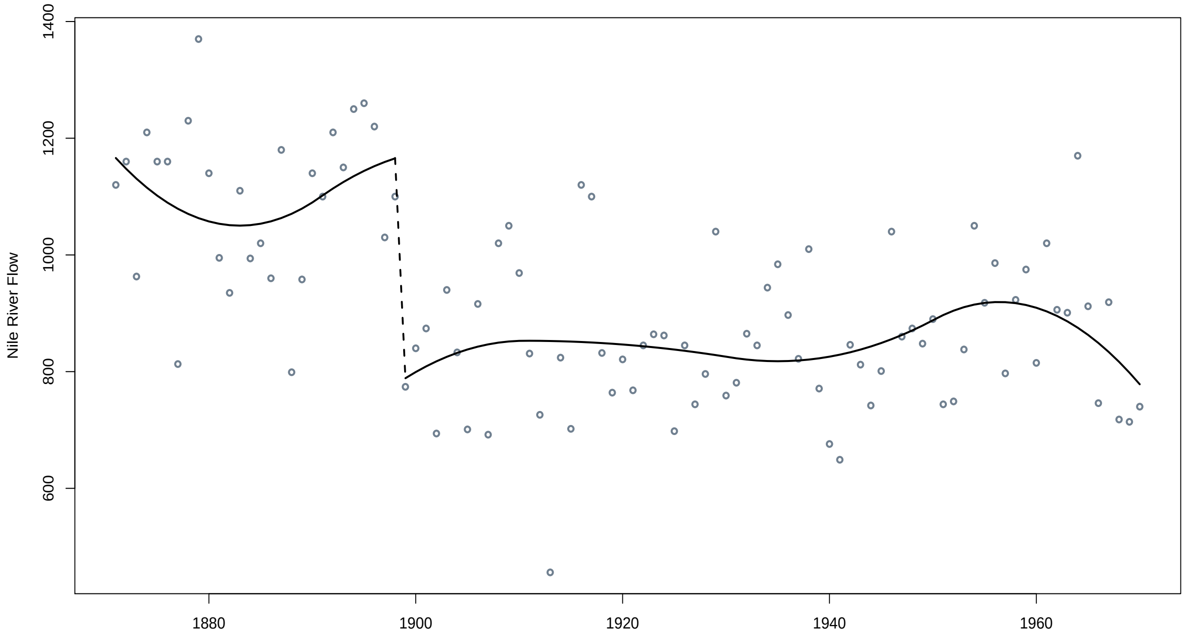

FIGURE 18. Nile river flow from 1871 to 1970. The solid line is the estimated Nile flow curve. The estimated jump point is 1898. The monotonicity constraint is that $f_m(x)$ is increasing on both sides of m and there is no constraint about the jump direction.

indicates its robustness. Besides, in most cases, bootstrap confidence intervals of the B-spline estimator have a much higher coverage rate while the average width of the confidence interval is smaller than its competitors.

4.3.1. Unimodal Case. Based on the two unimodal functions used in Section 4.1, we modify the curves by allowing the mode m to be random on the unit interval to check the robustness of the proposed spline-based estimator $\hat{m}_S$, and we use the penalized version. Now the two curves are $f_m(x) = 6(x+.5-m)(.5+m-x)$ and $f_m(x) = 30(x+.8-m)^4(.2+m-x)$, where m is uniformly distributed on $(0, 1)$. Due to the speed issue of $\hat{m}_U$ of Köllmann et al. (2014), we only compare $\hat{m}_S$ and $\hat{m}_P$ of Shoung and Zhang (2001) in this section. In this section, we compare the estimators based on the SMSE criteria for m and f_m. The SMSE

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{P}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.0061	.0138	.0266	.0109	.0182	.0304
200	.0036	.0094	.0195	.0085	.0146	.0242
500	.0016	.0052	.0124	.0062	.0108	.0181

Table 8. SMSE$_m$ for the estimator $\hat{m}_{\mathrm{S}}$ (proposed) and $\hat{m}_{\mathrm{P}}$ simulated from $f_m(x) = 6(x + .5 - m)(.5 + m - x)$.

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{P}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.0003	.0012	.0047	.0026	.0051	.0105
200	.0037	.0005	.0023	.0020	.0038	.0075
500	9.879e-05	.0002	.0009	.0014	.0026	.0050

Table 9. SMSE$_m$ for the estimators $\hat{m}_{\mathrm{S}}$ (proposed) and $\hat{m}_{\mathrm{P}}$ simulated from $f_m(x) = 30(x + .8 - m)^4(.2 + m - x)$.

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{P}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.0134	.0497	.1820	.0485	.1687	.6296
200	.0069	.0258	.0950	.0304	.1047	.3853
500	.0029	.0108	.0404	.0162	.0547	.1978

Table 10. SMSE$_f$ for the estimators $\hat{m}_{\mathrm{S}}$ (proposed) and $\hat{m}_{\mathrm{P}}$ simulated from $f_m(x) = 6(x + .5 - m)(.5 + m - x)$.

criteria are now defined as

$$\mathrm{SMSE}_m = \left[\frac{1}{N}\sum_{i=1}^{N} |\hat{m}_i - m_i|^2\right]^{1/2}$$

and

$$\mathrm{SMSE}_f = \left[\frac{1}{Nn}\sum_{i=1}^{N}\sum_{j=1}^{n} |\hat{f}_{\hat{m}_i}(x_j) - f_{m_i}(x_j)|^2\right]^{1/2},$$

where $N = 10,000$ and m_i is the random mode for the ith iteration. Three samples sizes and three standard deviations are used. Next, we compare the coverage rate of $\hat{m}_S$ and $\hat{m}_P$

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{P}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.1143	.1537	.3026	.1045	.3337	1.0552
200	.0995	.1155	.1922	.0750	.2231	.6746
500	.0836	.0916	.1231	.0436	.1229	.3613

Table 11. SMSE$_f$ for the estimators $\hat{m}_{\mathrm{S}}$ (proposed) and $\hat{m}_{\mathrm{P}}$ simulated from $f_m(x) = 30(x + .8 - m)^4(.2 + m - x)$.

based on bootstrapping, and the random noise is normal with zero mean and unit variance. For each curve, we construct a 95% bootstrap confidence interval by the following procedure.

1. At the ith iteration, m_i is randomly generated on $(0, 1)$ by the uniform distribution. For a scatterplot (X_j, Y_j), $j = 1, \ldots, n$, which satisfies that $Y_j = f_{m_i}(X_j) + \varepsilon_j, j = 1, \ldots, n$, we get a $\hat{f}_{\hat{m}_i}$ and a residual vector $\boldsymbol{e}$.
2. We sample a new residual vector $\boldsymbol{e}^*$ with replacement from $\boldsymbol{e}$. By adding $\boldsymbol{e}^*$ back to $\hat{f}_{\hat{m}_i}$, we get $\boldsymbol{Y}^*$ and regress it on $\boldsymbol{X}$ to get a bootstrap estimate of m_i.
3. Repeating the previous step for $10,000$ times, we get a bootstrap distribution for each estimator and use the 2.5% quantile and the 97.5% quantile as the 95% bootstrap confidence interval.

Repeating the procedure for 1000 times, we can get a coverage rate from the 1000 bootstrap confidence intervals for each estimator, and we can further compare the coverage rate and the average width of the bootstrap confidence intervals.

4.3.2. Inflection-Point Case. In the inflection-point case, we again generate m uniformly on $(0, 1)$, and the functions in Section 4.1 are changed into $f_m(x) = 10e^{-e^{5(1-\frac{x}{m})}}$ and $f_m(x) = 5\{1 + tanh[10(x - m)]$. In this section, we only consider $\hat{m}_S$ with the monotonicity constraint on f_m. First, we compare $\hat{m}_S$ with $\hat{m}_K$ of Kachouie and Schwartzman (2013) and $\hat{m}_C$ in the `R` package `inflection` by the SMSE criteria defined in the unimodal

	$\hat{m}_{\text{S}}$		$\hat{m}_{\text{P}}$	
n	Coverage Rate	C.I. Width	Coverage Rate	C.I. Width
100	.992	.443	.813	.260
200	.982	.333	.843	.238
500	.984	.266	.828	.199

Table 12. Coverage rate and C.I. width of $\hat{m}_{\text{S}}$ and $\hat{m}_{\text{P}}$ simulated from $f_m(x) = 6(x + .5 - m)(.5 + m - x)$.

	$\hat{m}_{\text{S}}$		$\hat{m}_{\text{P}}$	
n	Coverage Rate	C.I. Width	Coverage Rate	C.I. Width
100	.971	.141	.763	.129
200	.955	.102	.786	.115
500	.929	.060	.831	.099

Table 13. Coverage rate and C.I. width of $\hat{m}_{\text{S}}$ and $\hat{m}_{\text{P}}$ simulated from $f_m(x) = 30(x + .8 - m)^4(.2 + m - x)$.

	$\hat{m}_{\text{S}}$			$\hat{m}_{\text{K}}$			$\hat{m}_{\text{C}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.0014	.0025	.0045	.0666	.0651	.0629	.0031	.0039	.0075
200	.0011	.0019	.0035	.0660	.0638	.0621	.0031	.0037	.0076
500	.0007	.0013	.0023	.0655	.0652	.0645	.0030	.0036	.0081

Table 14. SMSE_m for the estimators $\hat{m}_{\text{S}}$ (proposed), $\hat{m}_{\text{K}}$ and $\hat{m}_{\text{C}}$ simulated from $f_m(x) = 10e^{-e^{5(1-\frac{x}{m})}}$.

case. Note that there is no estimator of f_m in the R package `inflection` , and we don't simulate SMSE_f or bootstrap confidence intervals for their estimator. Three samples sizes and three standard deviations are used. Next, we compare the coverage rate of $\hat{m}_S$ and $\hat{m}_K$. We use the bootstrapping method in the unimodal case to get the coverage rate of $\hat{m}_S$. According to Kachouie and Schwartzman (2013), an approximate 95% confidence interval for m is $[\hat{m}_K - 1.96SE(\hat{m}_K), \hat{m}_K + 1.96SE(\hat{m}_K)]$, where $SE(\hat{m}_K) \approx \dfrac{SE(\hat{f}''_{\hat{m}_K}(\hat{m}_K))}{\hat{f}'''_{\hat{m}_K}(\hat{m}_K)}$,

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{K}}$			$\hat{m}_{\mathrm{C}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.0005	.0009	.0017	.0626	.0615	.0599	.0010	.0018	.0052
200	.0003	.0007	.0012	.0604	.0603	.0583	.0009	.0016	.0052
500	.0002	.0004	.0008	.0618	.0613	.0610	.0009	.0016	.0055

Table 15. SMSE$_m$ for the estimators $\hat{m}_{\mathrm{S}}$ (proposed), $\hat{m}_{\mathrm{K}}$ and $\hat{m}_{\mathrm{C}}$ simulated from $f_m(x) = 5\{1 + tanh[10(x - m)]$.

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{K}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.0181	.0622	.2126	.2986	.3325	.4781
200	.0098	.0334	.1146	.2919	.3033	.3741
500	.0044	.0149	.0510	.2975	.2981	.3217

Table 16. SMSE$_f$ for the estimators $\hat{m}_{\mathrm{S}}$ (proposed) and $\hat{m}_{\mathrm{K}}$ simulated from $f_m(x) = 10e^{-e^{5(1-\frac{x}{m})}}$.

	$\hat{m}_{\mathrm{S}}$			$\hat{m}_{\mathrm{K}}$		
n	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$	$\sigma = .5$	$\sigma = 1$	$\sigma = 2$
100	.0185	.0625	.2104	.3806	.4126	.5543
200	.0100	.0338	.1138	.3772	.3880	.4557
500	.0045	.0152	.0514	.3796	.3799	.4022

Table 17. SMSE$_f$ for the estimators $\hat{m}_{\mathrm{S}}$ (proposed) and $\hat{m}_{\mathrm{K}}$ simulated from $f_m(x) = 5\{1 + tanh[10(x - m)]$.

	$\hat{m}_{\mathrm{S}}$		$\hat{m}_{\mathrm{K}}$	
n	Coverage Rate	C.I. Width	Coverage Rate	C.I. Width
100	.981	.113	.713	.420
200	.983	.097	.729	.311
500	.973	.083	.676	.195

Table 18. Coverage rate and C.I. width of $\hat{m}_{\mathrm{S}}$ and $\hat{m}_{\mathrm{K}}$ simulated from $f_m(x) = e^{-e^{5(1-\frac{x}{m})}}$.

$SE(\hat{f}''_{\hat{m}_K}(\hat{m}_K)) = \sqrt{var(\hat{f}''_{\hat{m}_K}(\hat{m}_K))}$ is the standard error of the estimated second derivative at $\hat{m}_K$, and $\hat{f}'''_{\hat{m}_K}(\hat{m}_K)$ is the estimated third derivative at $\hat{m}_K$.

	$\hat{m}_{\mathrm{S}}$		$\hat{m}_{\mathrm{K}}$	
n	Coverage Rate	C.I. Width	Coverage Rate	C.I. Width
100	.984	.097	.792	.381
200	.987	.084	.797	.277
500	.984	.066	.762	.174

Table 19. Coverage rate and C.I. width of $\hat{m}_{\mathrm{S}}$ and $\hat{m}_{\mathrm{K}}$ simulated from $f_m(x) = 5\{1 + tanh[10(x - m)]$.

CHAPTER 5

Extension to the Generalized Linear Models

In this chapter, we consider the generalized linear model with independent observations from an exponential family, and its mean curve is smooth with a change-point which we have discussed. Such extension can be useful for the estimation of the optimal dosage of a drug which gives the highest survival rate, which is an application of mode estimation, and the estimation of the the inflection-point in a convex-concave dose-response curve.

The generalized linear model with independent observations from an exponential family is of the form

$$p(y_i;\theta,\tau)=exp[\{y_i\theta_i-b(\theta_i)\}\tau-c(y_i,\tau)], i=1,\ldots,n, \tag{38}$$

where the specifications of the functions b and c determine the sub-family of models. The mean vector $\boldsymbol{\mu}=E(\boldsymbol{y})$ has values $\mu_i=b'(\theta_i)$, and is related to a design matrix of predictor variables through a monotonically increasing link function $g(\mu_i)=\eta_i, i=1,\ldots,n$, where $\boldsymbol{\eta}$ is the systematic component and describes the relationship with the predictors. The relationship between $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ is determined by the link function b. For now, we only discuss the Poisson and binomial response.

In the unimodal case and the jump-point case, we use quadratic B-splines and to constrain the monotonicity of $\boldsymbol{\mu}$ is equivalent to constrain the monotonicity of $\boldsymbol{\eta}$. The change-point in $\boldsymbol{\eta}$ is also the change-point in $\boldsymbol{\mu}$. We specify $\boldsymbol{\eta}$ for each observation by $\eta_i=f_m(x_i)$, where f_m satisfies the smoothness and regular conditions defined in Chapter 3. An iteratively re-weighted cone projection algorithm is used to estimate $\hat{m}$, $\hat{\boldsymbol{\eta}}$, and $\hat{\boldsymbol{\mu}}$ is obtained by transforming $\hat{\boldsymbol{\eta}}$ by the inverse of the link function.

To be specific, the algorithm is as following. The negative log-likelihood

$$L(\theta, \tau; y) = \sum_{i=1}^{n} \Big\{ c(y_i, \tau) - \frac{y_i\theta_i - b(\theta_i)}{\tau} \Big\}$$

is written in terms of the systematic component and minimized over $C \subseteq \mathbb{R}^n$. Let $\ell(\boldsymbol{\eta})$ be the negative log likelihood written as a function of $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$. For $\boldsymbol{\eta}_k$ in the constraint set, let

$$\psi_k(\boldsymbol{\eta}) = \ell(\boldsymbol{\eta}_k) + \nabla\ell(\boldsymbol{\eta}_k)^\top(\boldsymbol{\eta} - \boldsymbol{\eta}_k) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}_k)^\top \boldsymbol{Q}_k(\boldsymbol{\eta} - \boldsymbol{\eta}_k), \tag{39}$$

which is equivalent to

$$\frac{1}{2}\boldsymbol{\eta}^\top \boldsymbol{Q}_k \boldsymbol{\eta} + \Big[\nabla\ell(\boldsymbol{\eta}_k) - \boldsymbol{Q}_k\boldsymbol{\eta}_k\Big]^\top \boldsymbol{\eta},$$

where $\nabla\ell(\boldsymbol{\eta}_k)$ is the gradient vector and $\boldsymbol{Q}_k$ is the Hessian matrix for $\ell(\boldsymbol{\eta})$, both evaluated at $\boldsymbol{\eta}_k$. Since we approximate $\boldsymbol{\eta}$ as $\boldsymbol{Bb}$ where the columns of $\boldsymbol{B}$ are quadratic B-splines, we could further write (39) as

$$\psi_k(\boldsymbol{b}) = \frac{1}{2}\boldsymbol{b}^\top \boldsymbol{B}^\top \boldsymbol{Q}_k \boldsymbol{B}\boldsymbol{b} + \Big[\nabla\ell(\boldsymbol{\eta}_k) - \boldsymbol{Q}_k\boldsymbol{B}\boldsymbol{b}_k\Big]^\top \boldsymbol{B}\boldsymbol{b}. \tag{40}$$

Then as we did in Chapter 3, we can include a covariate matrix $\boldsymbol{Z}$ in $\boldsymbol{B}$, and we also have a penalized estimator by adding a penalty term to (40) such that the criterion for penalized regression is

$$\psi_k(\boldsymbol{b}) = \frac{1}{2}\boldsymbol{b}^\top (\boldsymbol{B}^\top \boldsymbol{Q}_k \boldsymbol{B} + \lambda \boldsymbol{D}^\top \boldsymbol{D})\boldsymbol{b} + \Big[\nabla\ell(\boldsymbol{\eta}_k) - \boldsymbol{Q}_k\boldsymbol{B}\boldsymbol{b}_k\Big]^\top \boldsymbol{B}\boldsymbol{b}, \tag{41}$$

where λ and $\boldsymbol{D}$ are defined in Section 3.2 of Chapter 3. To impose a shape constraint on $\boldsymbol{\eta}$, we again use the linear inequality $\boldsymbol{Sb} \geq \mathbf{0}$.

For a fixed m, the iteratively re-weighted algorithm is:

1. Choose a valid starting $\boldsymbol{\eta}_0$, and set $k = 0$.
2. Given $\boldsymbol{\eta}_k$, minimize $\psi_k(\boldsymbol{\eta})$ over C defined by the model. Then $\boldsymbol{\eta}_{k+1}$ minimizes $\ell(\boldsymbol{\eta})$ over the line segment connecting the minimizer of $\psi_k(\boldsymbol{\eta})$ and $\boldsymbol{\eta}_k$.
3. Set $k = k + 1$ and repeat step 2, stopping when a convergence criterion is met.

At step 2, a cone projection algorithm is required, and the `coneA` routine of the `R` package `coneproj` Liao and Meyer (2014) is used. At each iteration of the algorithm, the vector $\boldsymbol{\mu}_k$ is computed where $\mu_{ki} = g^{-1}(\eta_{ki})$. If the Hessian matrix is positive definite for all $\boldsymbol{\eta}$ then the negative log-likelihood function is strictly convex and $\boldsymbol{\mu}_k$ is guaranteed to converge to the MLE $\hat{\boldsymbol{\mu}}_k$.

When m is unknown, we search through knots for $\hat{m}$ along with $\hat{\boldsymbol{\eta}}_{\hat{m}}$ and $\hat{\boldsymbol{\mu}}_{\hat{m}}$. The final set of MLE estimates minimizes the negative log likelihood.

In the inflection-point case, since we use cubic B-splines, restricting the convexity of $\boldsymbol{\mu}$ cannot be obtained by restricting the convexity of $\boldsymbol{\eta}$. Instead, $\boldsymbol{\mu}$ is specified for each observation by $\mu_i = f_m(x_i)$, where f_m satisfies the smoothness and regular conditions in Chapter 3, and is approximated by cubic B-spline basis functions as we discussed for the least-squares model.

We present some examples for each case in the following.

5.1. Unimodal Case

First, we consider the `trade.union` data set in the `R` package `SemiPar` Wand (2014) again. This data set is also discussed in Chapter 11 in Ruppert et al. (2003), in which

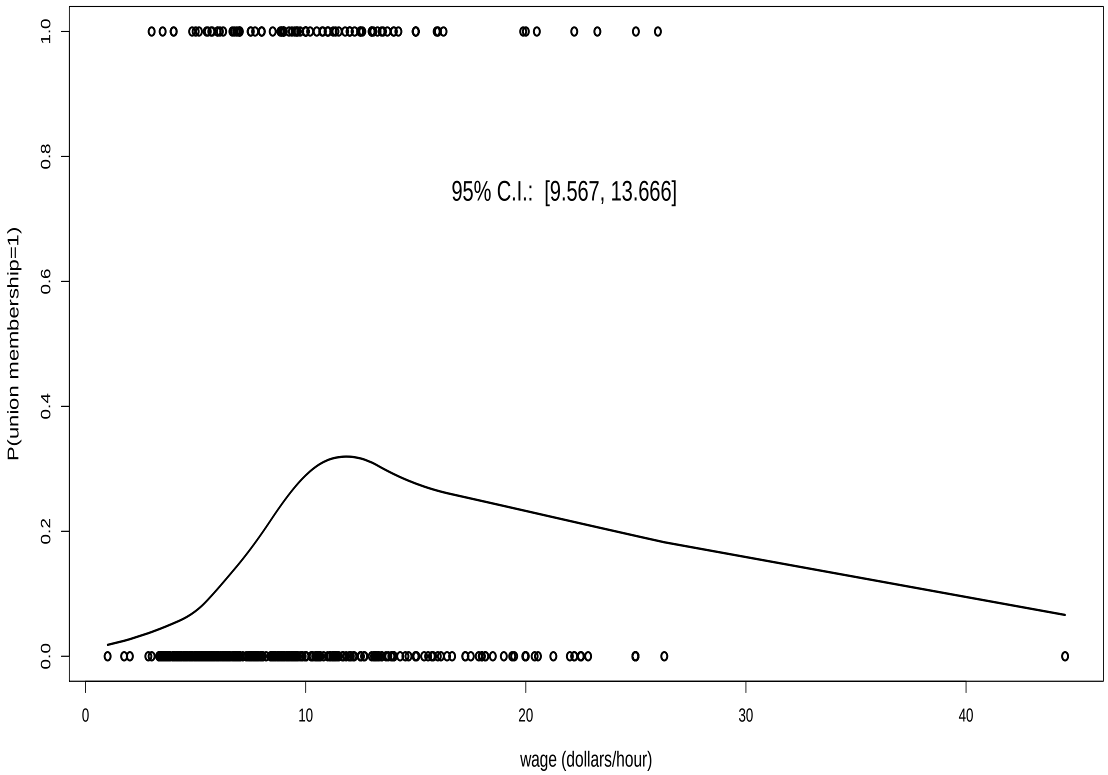

FIGURE 1. Fitted unimodal curve for the probability of union membership = 1 as a function of wages. $\hat{m} = 11.86$.

the binary variable `union membership` (union) is the response and the continuous variable `wages` (wages) is the predictor. As discussed in Chapter 11 in Ruppert et al. (2003), there is strong evidence from the quadratic fit that the linear logistic model is inadequate. It is shown that the probability of `union membership` = 1 increases as `wage` increases up to a point, and then decreases with increasing wages. We could model the relationship between the probability of `union membership` = 1 and `wages` as unimodal in Figure 1. Here we use the penalized estimator and also get a 95% bootstrap confidence interval for m. As a comparison, the log-odds linear and quadratic fits are shown in Figure 2. The P-value of the lack-of-fit test is 3.672e−06 which indicates significant inadequacy of the linear fit. For the unimodal fit, we could further include "occupation" as a categorical covariate as we did in Section 4.2 in Chapter 4. The fit for each occupation is shown in Figure 3.

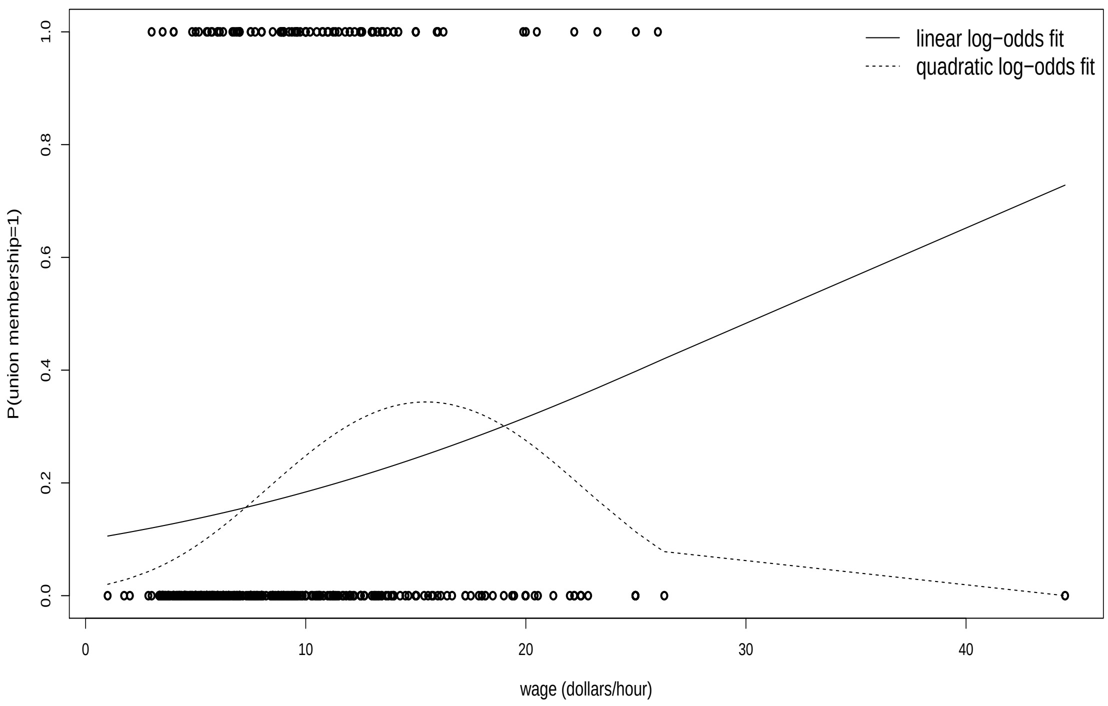

FIGURE 2. Log-odds fits for the probability of union membership = 1 as a function of wages. The solid line is the linear fit. The dash line is the quadratic fit.

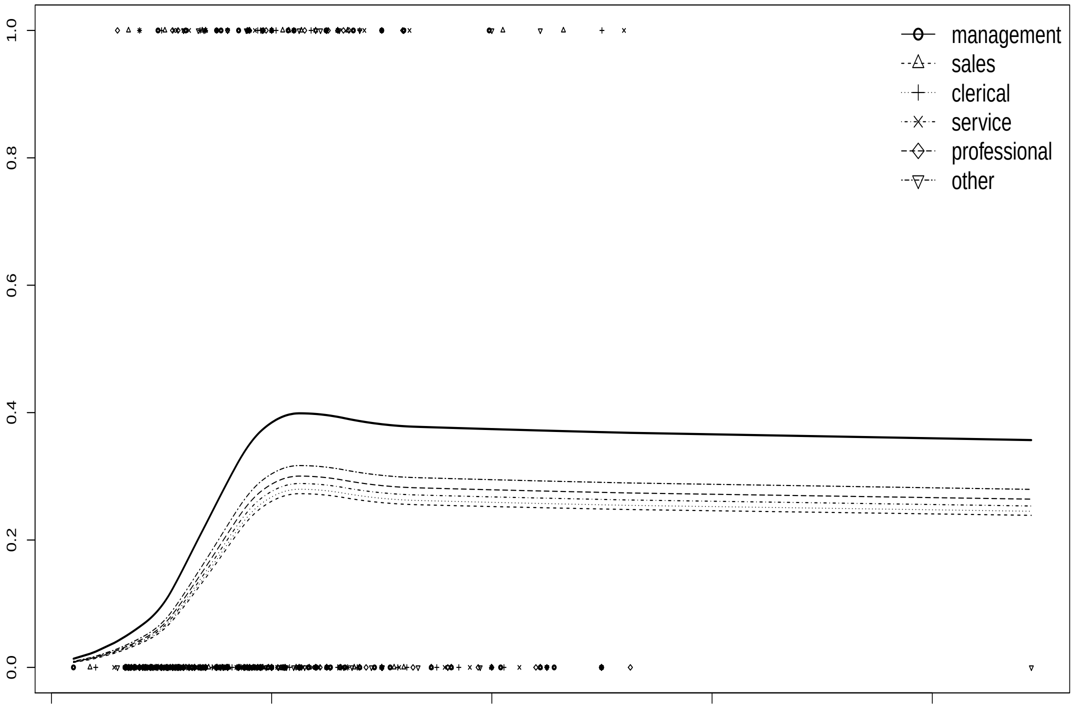

FIGURE 3. Fitted unimodal curve for each occupation for the probability of union membership = 1 as a function of wages. $\hat{m} = 11.25$.

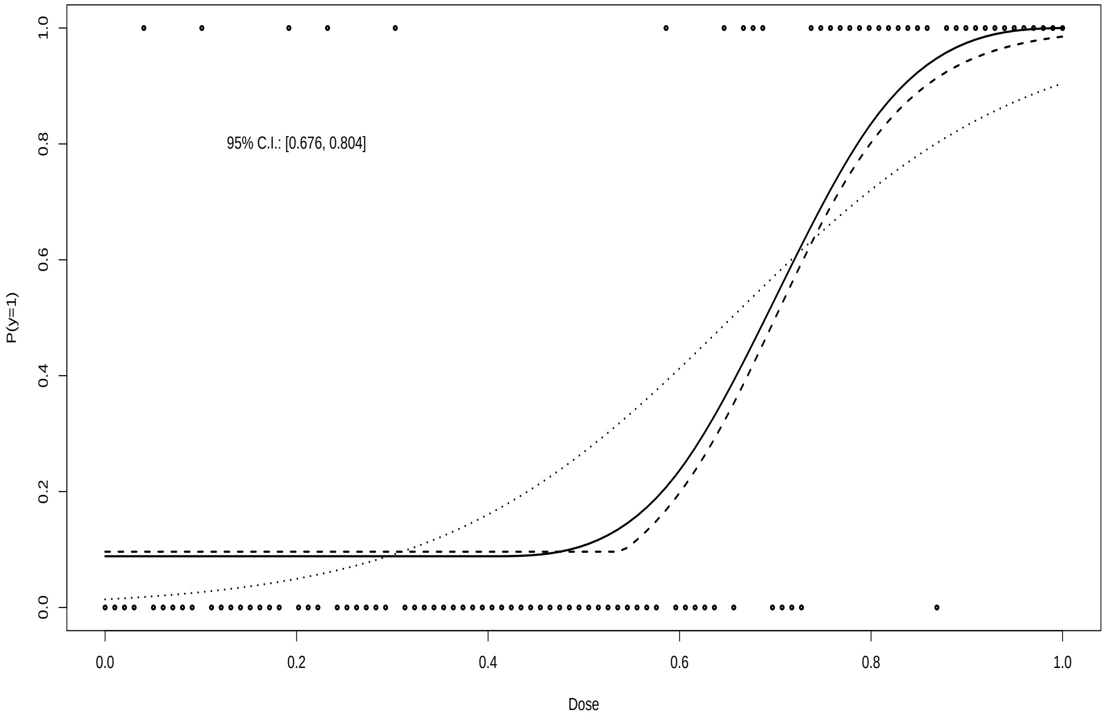

FIGURE 4. The solid curve is $\hat{f}_{\hat{m}}$ for the probability of $y = 1$ as a convex-concave function of dosage with $\hat{m} = .702$, the dot-dash curve is the logistic fit, and the dash curve is the dose-response curve with $m = .7$.

5.2. Inflection-Point Case

We simulate a dose-response data set from the dashed curve. It is shown in Figure 4 that the standard logistic fit cannot capture the true shape, however a smooth curve constrained to be convex-concave is close to the true curve.

5.3. Jump-Point Case

First, we simulate a data set from the curve $f_m(x) = .8sin(5x) + .6x + .4I_{[.7,1]}(x)$ such that f_m has a jump-point $m = .7$, and the response vector is binomial with $n = 100$ observations with the mean vector equal to f_m. The fit is shown in Figure 5. Next, using the curve $f_m(x) = 4sin(5x) + 3x + I_{[.7,1]}(x) + 2$, which has a jump-point at .7, we simulate a Poisson data set with $n = 1000$ observations. The fit is shown in Figure 6.

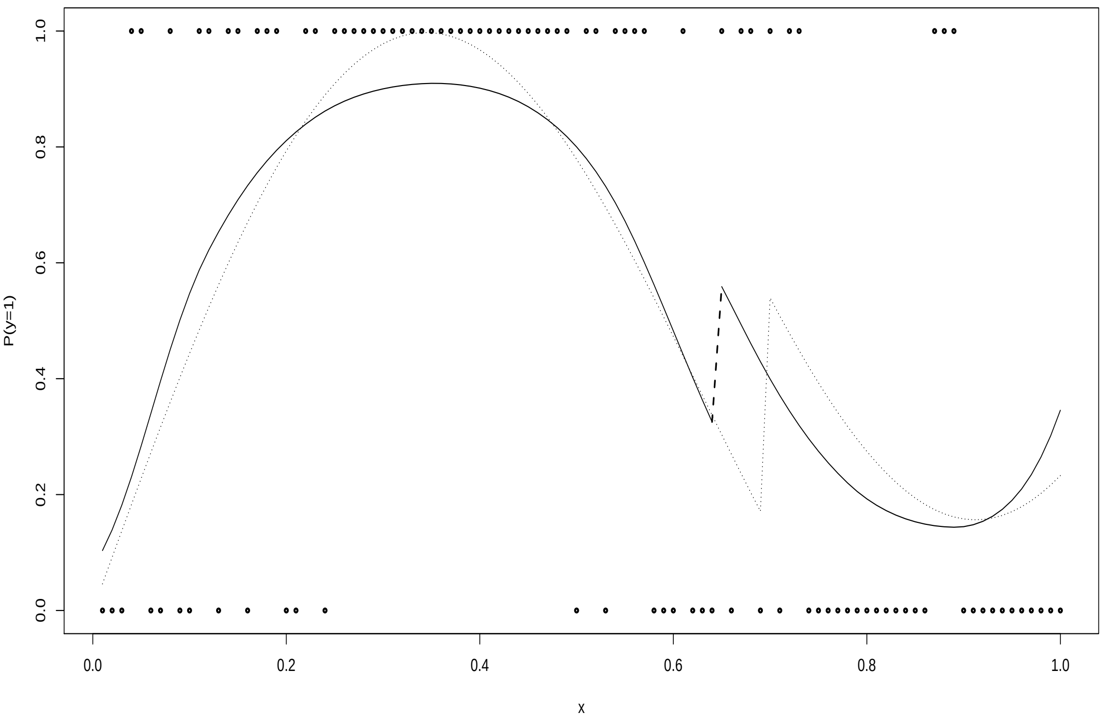

FIGURE 5. The solid curve is $\hat{f}_{\hat{m}}$ for the probability of $y = 1$ as a function of x with $\hat{m} = .64$ and the dot-dash curve is f_m with $m = .7$.

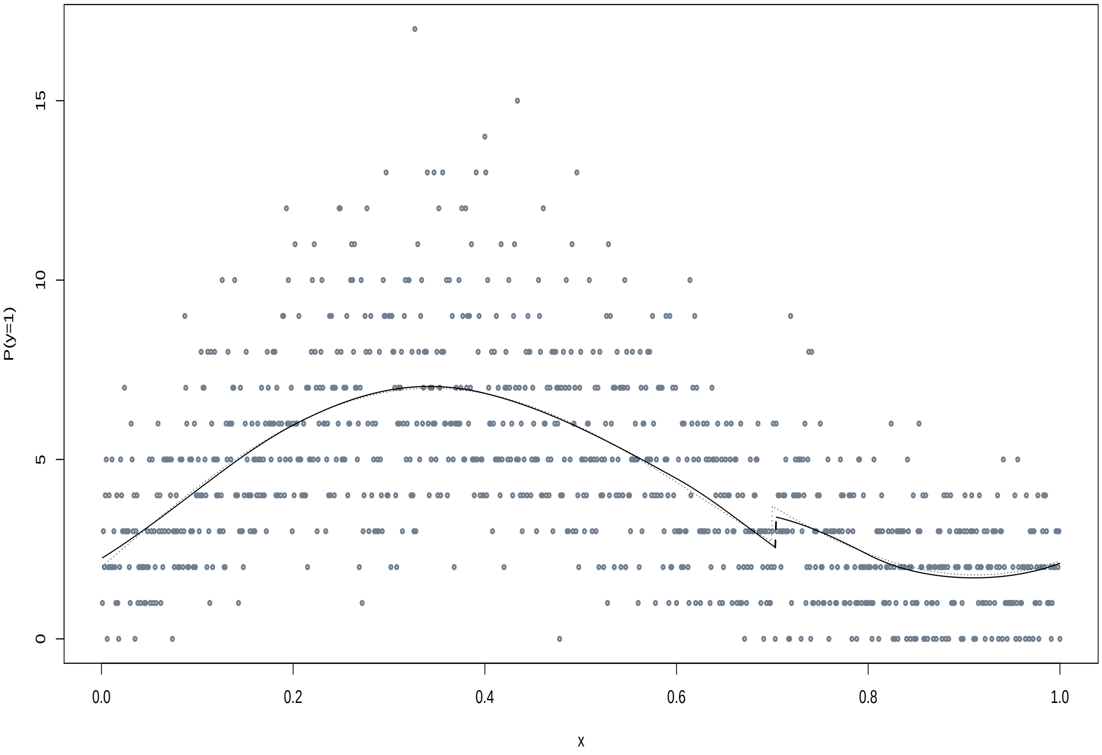

FIGURE 6. The solid curve is $\hat{f}_{\hat{m}}$ with $\hat{m} = .703$ and the dot-dash curve is f_m with $m = .7$.

CHAPTER 6

General Shape Selection

6.1. Overview

In Chapter 2, we propose three change-point estimators in an underlying curve satisfying some smoothness conditions with *a priori* shape constraint. In Chapter 3, convergence rates are developed for the three estimators, and penalized estimation is discussed as an option to avoid over-fitting. We also discuss the model which could include a time-series covariance structure or a categorical covariate.

Now, instead of assuming one possible shape, we suppose that there is a set of possible shapes, either with or without change-point(s), for a given data set, and we want to choose the shape which is closest to the underlying phenomenon. To be specific, we define the basic model as

$$Y_i = f(X_i) + \sigma\varepsilon_i, \quad \text{for} \;\; i = 1, \ldots, n,$$

where f is the underlying trend, X_i is the predictor which could be a time point, and ε_i is the noise. The trend represents several possible phenomena which could be simply monotonic or follow some pattern with a change-point such as an increasing-decreasing (unimodal) curve, a convex-concave curve, or a time series like river flow with jump-point(s).

For all possible shapes, we can assume that each shape satisfies some smoothness conditions, like continuous first derivatives, second derivatives etc, and we use quadratic (cubic) B-splines to approximate the underlying phenomenon. To impose a shape constraint on the fit, we define a $k \times m$ constraint matrix $\boldsymbol{S}$ of slopes or second derivatives at the knots similarly to what we define in Section 2.2 of Chapter 2. For example, if the underlying curve is assumed to be decreasing, then to get a decreasing fit, we find $\boldsymbol{b} \in \mathbb{R}^m$ to minimize the criterion (7) subject to the inequality constraint $\boldsymbol{Sb} \leq \boldsymbol{0}$; if the shape is increasing-decreasing, we apply the inequality constraint defined in the unimodal case in Section 2.2 of Chapter 2; if the shape is decreasing-increasing, we simply swap the sign of the rows of $\boldsymbol{S}$ which are defined at knots greater than the "mode" and the rows defined at knots smaller than the "mode". $\boldsymbol{S}$ can be defined flexibly to catch more general shapes.

A fit will be made for each shape, and for shapes with change-point(s), the estimation methods discussed in previous chapters apply. The best fit can be chosen according to the "Cone Information Criterion" (CIC) Meyer (2013), which is defined as

$$log(SSE) + log\Big\{\frac{2\big[E_0(D) + d_0\big]}{n - d_0 - cE_0(D)} + 1\Big\},$$

where SSE is the sum of squared residuals, d_0 is the dimension of the null space contained in the cone, and $E_0(D)$ is the "null expected degrees of freedom" which is computed by simulating a lot of data sets, usually more than one thousand, as independent and normally distributed with zero mean and taking the average of the used degrees of freedom of the fits. (For the change-point models, one degree of freedom for the change-point is added.). The shape with the smallest CIC is chosen as the best. This criterion is similar to the "Akaike Information Criterion" (AIC) but is specially defined for cone projection problems.

6.1.1. Global Annual Mean Precipitation. We consider estimating the underlying trend and the possible change-point of the global annual mean precipitation from 1901 to 2000, which is available at `http://data.giss.nasa.gov/precip_cru/`. Suppose that the underlying trend has three possible shapes: flat, increasing, or increasing-decreasing. We fit all three shapes and estimate the turn-around point for the increasing-decreasing

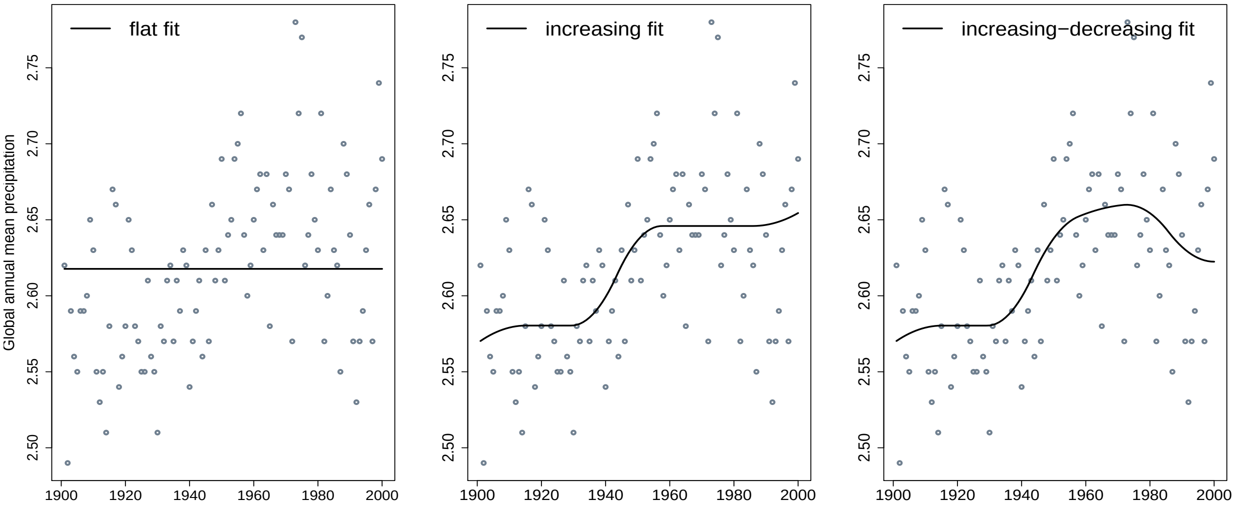

FIGURE 1. Left: flat fit with CIC$= -.415$. Middle: smooth increasing fit with CIC$= -.757$. Right: increasing-decreasing fit with CIC$= -.764$ and $\hat{m} = 1973$.

shape, and the best fit can be chosen according to the CIC criterion, which is a measure of model complexity for cone projection problems. For this time series, we again assume that the random error follows an AR(p) process. The autoregressive process is estimated simultaneously for each shape. The unimodal fit has the smallest CIC value and thus the underlying trend of the global annual mean precipitation is estimated to be slowly increasing with a peak at 1973 and then decreasing. For this fit, the random error is estimated to follow an AR(1) process and the estimated autoregressive coefficient is .146, which implies that the random error of each year has a weakly positive correlation with the random error of the immediately previous year. The three fits are in Figure 1.

6.2. Shape Selection with FIA Data

6.2.1. Background. Understanding forest disturbance is important for carbon assessment and forest management decisions. Forest Inventory & Analysis (FIA) scientists have been working with National Aeronautics and Space Administration (NASA) and university partners on remote sensing-based projects to detect and characterize forest land cover changes over the last three decades. In the North American Dynamics Project (NADP), work has been done to attribute causal agents to the nationwide change maps, making predictions of forest disturbance and cause, as well as fitted spectral trajectories and other useful parameters available at 30m resolution annually over this country. In this project, we use shape-restricted B-splines to fit trajectories of Landsat imagery, which monitors forest conditions, to detect annual forest disturbance dynamics over three decades. (Landsat satellites collect important data about global forest conditions. Documentation about Landsat's role in forest disturbance estimation is available at `http://landsat.gsfc.nasa.gov/?p=9513`.)

In this project, $\boldsymbol{Y}$ is a vector of Landsat band or index measurements, and $\boldsymbol{X}$ is a vector representing years. We assume that $\boldsymbol{Y}$ is trend plus noise, and the trend represents various possible phenomena captured by Landsat trajectories on a single pixel through time. The underlying trend is constrained to behave in an ecologically sensible manner, assuming one of seven possible shapes. An undisturbed forest has a `flat` signal, while the signal is `decreasing` for a forest recovering from a disturbance. An upward `jump` indicates a disturbance in forest canopy or structure, typically caused by a harvest or fire and a `double-jump` signal shows two distinct disturbance events over a short period. A `decreasing-increasing` signal indicates a forest that recovers at first but then encounters a slow disturbance. Conversely, an `increasing-decreasing` signal illustrates the opposite case. Finally, an `increasing` pattern shows gradual decline of a forest caused by a disturbance that might occur very early in the time series of Landsat imagery. An example of each shape is shown in Figure 2.

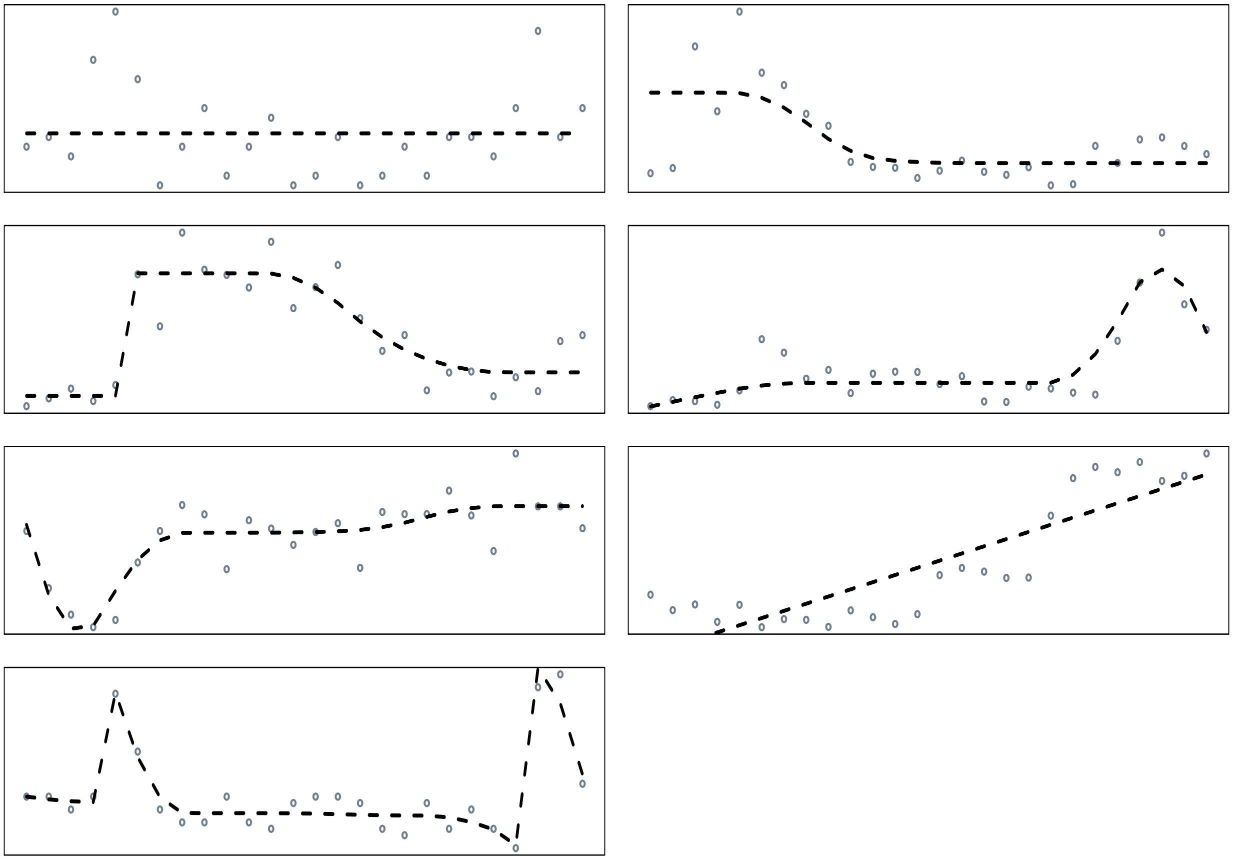

FIGURE 2. Seven possible shapes: flat, decreasing, one-jump (decreasing, jump up, decreasing), increasing-decreasing, decreasing-increasing, linearly increasing, and double-jump (decreasing, jump up, decreasing, jump up, decreasing).

6.2.2. Model Set-Up and Information Criteria. For all shapes except for the flat shape and the linearly increasing shape, we use a constrained linear combination of quadratic B-spline basis functions to estimate the regression function given inequality constraints of the form $\boldsymbol{Sb} \geq \mathbf{0}$. In the one-jump (double-jump) case, we use one (two) "jump" basis function and one (two) "ramp" basis function. This is a constrained quadratic programming problem and it is solved by the `coneA` routine in the `R` package coneproj Meyer and Liao (2014). Suppose that $t_1, \ldots, t_k$ are knots, then

Decreasing: we define the $k \times (k+1)$ constraint matrix $\boldsymbol{S}$ as

$$S_{ij} = -\eta_j'(t_i) \;\; i = 1, \ldots, k \;\text{ and }\; j = 1, \ldots, k+1.$$

One-Jump: we define the $(k+3) \times (k+3)$ constraint matrix $\boldsymbol{S}$ as

$$S_{ij} = -\eta_j'(t_i) \;\; i = 1, \ldots, k \text{ and } j = 1, \ldots, k+1,$$

$$S_{(k+1)j} = \begin{cases} 1 & j = k+2 \\ 0 & j \neq k+2, \end{cases}$$

$$S_{(k+2)j} = \begin{cases} -\eta_j'(m) & j = 1, \ldots, k+1 \\ 0 & j = k+2 \;\text{ and }\; k+3, \end{cases}$$

$$S_{(k+3)j} = \begin{cases} -\eta_j'(m) & j = 1, \ldots, k+1 \\ 0 & j = k+2 \\ -1 & j = k+3. \end{cases}$$

Suppose that t_p is the knot such that $t_p \leq m < t_{p+1}$, $p \in \{1, \ldots, k-1\}$. We define a $k \times (k+1)$ matrix $\boldsymbol{S}$ as

Inverted-Vee:

$$S_{ij} = \begin{cases} \eta_j'(t_i) & i = 1, \ldots, p \;\text{ and }\; j = 1, \ldots, k+1 \\ -\eta_j'(t_i) & i = p+1, \ldots, k \;\text{ and }\; j = 1, \ldots, k+1. \end{cases}$$

Vee:

$$S_{ij} = \begin{cases} -\eta_j'(t_i) & i = 1, \dots, p \text{ and } j = 1, \dots, k+1 \\ \\ \eta_j'(t_i) & i = p+1, \dots, k \text{ and } j = 1, \dots, k+1. \end{cases}$$

Double-Jump: suppose that m is the first jump and m' is the second jump, we define the $(k+6) \times (k+5)$ constraint matrix $\boldsymbol{S}$ as

$$S_{ij} = -\eta_j'(t_i) \;\; i = 1, \dots, k \text{ and } j = 1, \dots, k+1,$$

$$S_{(k+1)j} = \begin{cases} 1 & j = k+2 \\ \\ 0 & j \neq k+2, \end{cases}$$

$$S_{(k+2)j} = \begin{cases} -\eta_j'(m) & j = 1, \dots, k+1 \\ \\ 0 & j \geq k+2, \end{cases}$$

$$S_{(k+3)j} = \begin{cases} -\eta_j'(m) & j = 1, \dots, k+1 \\ \\ 0 & j = k+2, k+4 \text{ and } k+5 \\ \\ -1 & j = k+3. \end{cases}$$

$$S_{(k+4)j} = \begin{cases} 1 & j = k+4 \\ \\ 0 & j \neq k+4, \end{cases}$$

$$S_{(k+5)j} = \begin{cases} -\eta_j'(m') & j = 1, \dots, k+1 \\ \\ 0 & j = k+2, k+4 \text{ and } k+5 \\ \\ -1 & j = k+3, \end{cases}$$

$$S_{(k+6)j} = \begin{cases} -\eta'_j(m') & j = 1, \ldots, k+1 \\ 0 & j = k+2 \text{ and } j = k+4 \\ -1 & j = k+3 \text{ and } j = k+5. \end{cases}$$

For each trajectory with n observations, all seven shapes can be fitted. An information criterion is computed for each fit to choose between the seven fits, and it is defined as the sum of squared residuals (SSE) penalized by adding a measure of model complexity which is a function of the "null effective degrees of freedom" of the model Meyer (2013). The shape with the smallest information criterion is the winner. Two information criteria are considered. One is the "Bayesian Information Criterion" (BIC), which is defined as

$$nlog(SSE) + log(n)E_0(D),$$

and the other is the "Cone Information Criterion" (CIC) Meyer (2013), which is defined as

$$log(SSE) + log\Big\{\frac{2\big[E_0(D)+1\big]}{n-1-1.5E_0(D)} + 1\Big\},$$

where $E_0(D)$ is the null expected dimension of the face of the cone on which the projection lands. For the flat shape and the linearly increasing shape, which are not fitted by the cone projection algorithm, $E_0(D) = 0$ and 1.5.

6.2.3. ShapeSelectForest. The `R` package `ShapeSelectForest` Meyer, Liao, Freeman, and Moisen (2015) package applies constrained regression splines to time series of Landsat imagery for the purpose of modelling, mapping, and monitoring annual forest disturbance dynamics. For each pixel and spectral band or index of choice in temporal Landsat data, the package gives an optimally smoothed fit of the trajectory constrained to behave in an ecologically sensible manner, assuming one of seven possible shapes. It also contains functions for deriving annual predictions of forest disturbance, as well as graphical displays of the shape fits.

In the `ShapeSelectForest` package, for a given set of consecutive years, we compute the `edf0` vector (null effective degrees of freedom) using the routine `getedf0`. Each element of the `edf0` vector is an `edf0` value simulated for one of the seven shapes under the null hypothesis that the response is independent of the predictor. The calculations for the `edf0` vector for a given set of years can be time-consuming, this is accomplished in the subroutine `getedf0`, and the `edf0` vector is an input to the main routine `shape`. The user can choose to simulate the `edf0` value for each single shape, or to simulate the `edf0` vector for all the seven shapes. An `edf0` vector can be used for many response vectors as long as the set of years corresponding to the response vectors are equally spaced and have the same length. We put a matrix `edf0s` in this package. It is a 21×7 matrix. Each row is an `edf0` vector for an equally spaced predictor vector $\boldsymbol{x}$ of n elements. Each vector has seven elements corresponding to the seven shapes. From the first row to the last row, the `edf0` vector is for $\boldsymbol{x}$ of length n which is an integer ranging from 20 to 40. When $\boldsymbol{x}$ is not equally spaced or its number of elements is not between 20 and 40, `getedf0` will be called to get the `edf0` vector.

The main routine in this package is `shape`. Given a predictor vector $\boldsymbol{x}$, e.g., years and a matrix whose columns are response vectors corresponding to $\boldsymbol{x}$. The `shape` routine will select a shape among the seven possible shapes that is the best fit for each response vector according to the BIC criterion or the CIC criterion. The user can choose to simulate an `edf0` vector before calling `shape`, or simulate it inside `shape`. Given a scatter plot, the user can choose any subset of the seven shapes or all the seven shapes to get a shape-restricted fit.

Suppose that we have N scatter plots and for each plot, we have n observations. The main output objects of `shape` are **shape**: a $N \times 1$ vector. Each element is the best shape for each of the N scatter plots; **ic**: a $k \times N$ matrix where the ith column is the vector of BIC or CIC values used to choose the best shape for the ith scatterplot, where k is the number of shapes allowed by the user; **thetab**: a $n \times N$ matrix where the ith column is the $n \times 1$ vector of predicted values for the chosen shape for the ith scatter plot.

Here we present a data set built in this package as an example. The predictor $\boldsymbol{x}$ is a set of consecutive years from 1985 to 2010. For this set of years, there are 36 pixels in South Carolina. For each pixel, there are two spectral bands, B5 and NDVI. It is believed that these pixels have more than one disturbance. We can choose to fit the 36 scatter plots with the seven possible shapes and choose the best shape for each of them according to the CIC criterion. By the graphical routine `plotshape`, we can plot the best shape for each scatter plot along with the CIC value for each candidate shape.

From Figure 3 to Figure 8, each plot on the left panel represents the Landsat signals of a pixel in South Carolina with the fitted trajectory from 1985 to 2010. The fitted trajectory is chosen according to the CIC criterion. On the right panel, each plot shows the CIC values for the seven shapes. The best shape has the smallest CIC value.

Now we are extending the `ShapeSelectForest` package to longer time series and also try to develop the algorithm to allow for new shapes, such as multiple jumps. Also, choosing the best shape of a pixel based on best shapes of spatial neighbors is being considered.

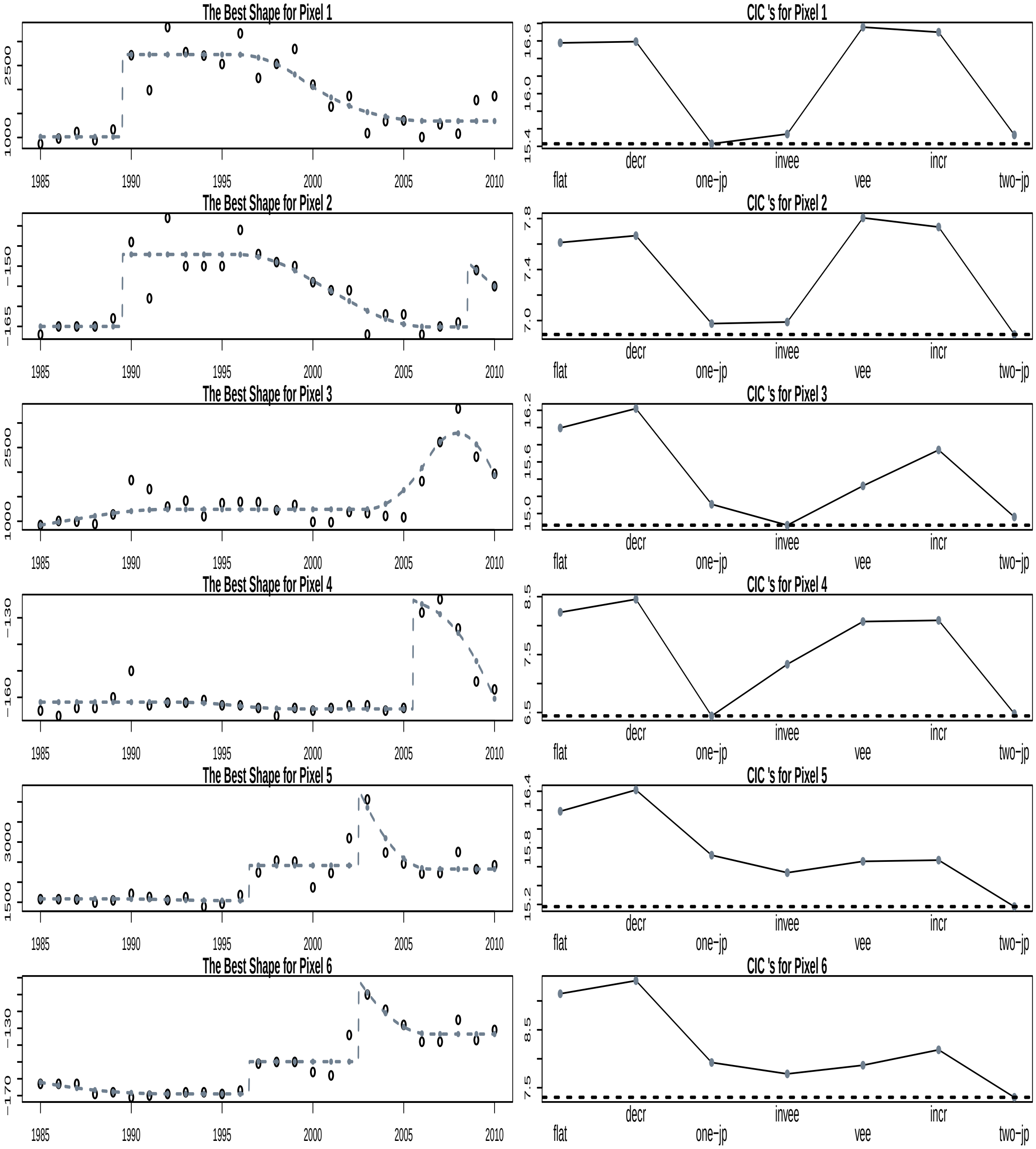

Figure 3. Left: the best shape chosen by the CIC criterion with fitted values marked as dots. Right: CIC values for seven possible shapes. The shape with the smallest CIC value is the winner.

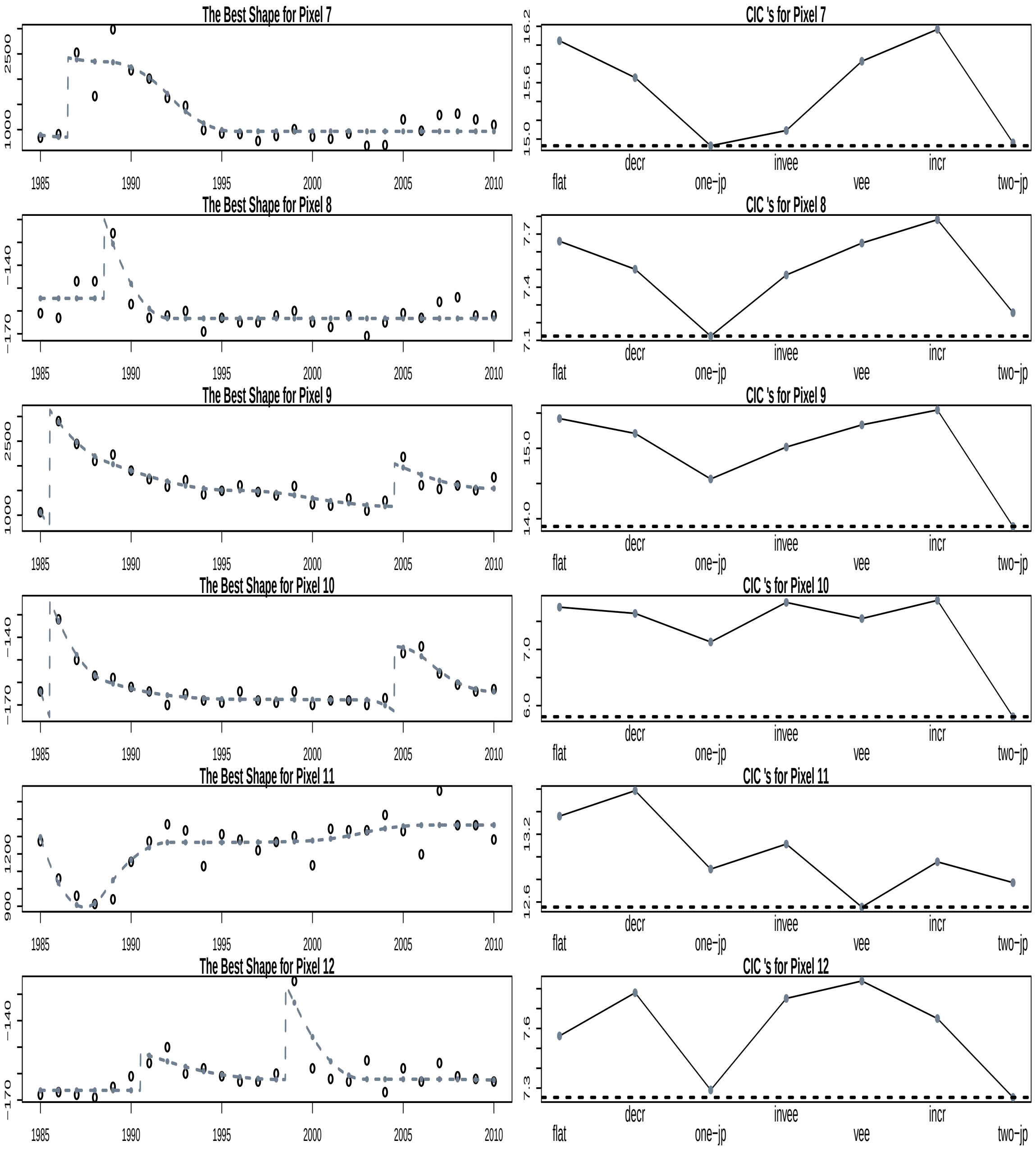

FIGURE 4. Left: the best shape chosen by the CIC criterion with fitted values marked as dots. Right: CIC values for seven possible shapes. The shape with the smallest CIC value is the winner.

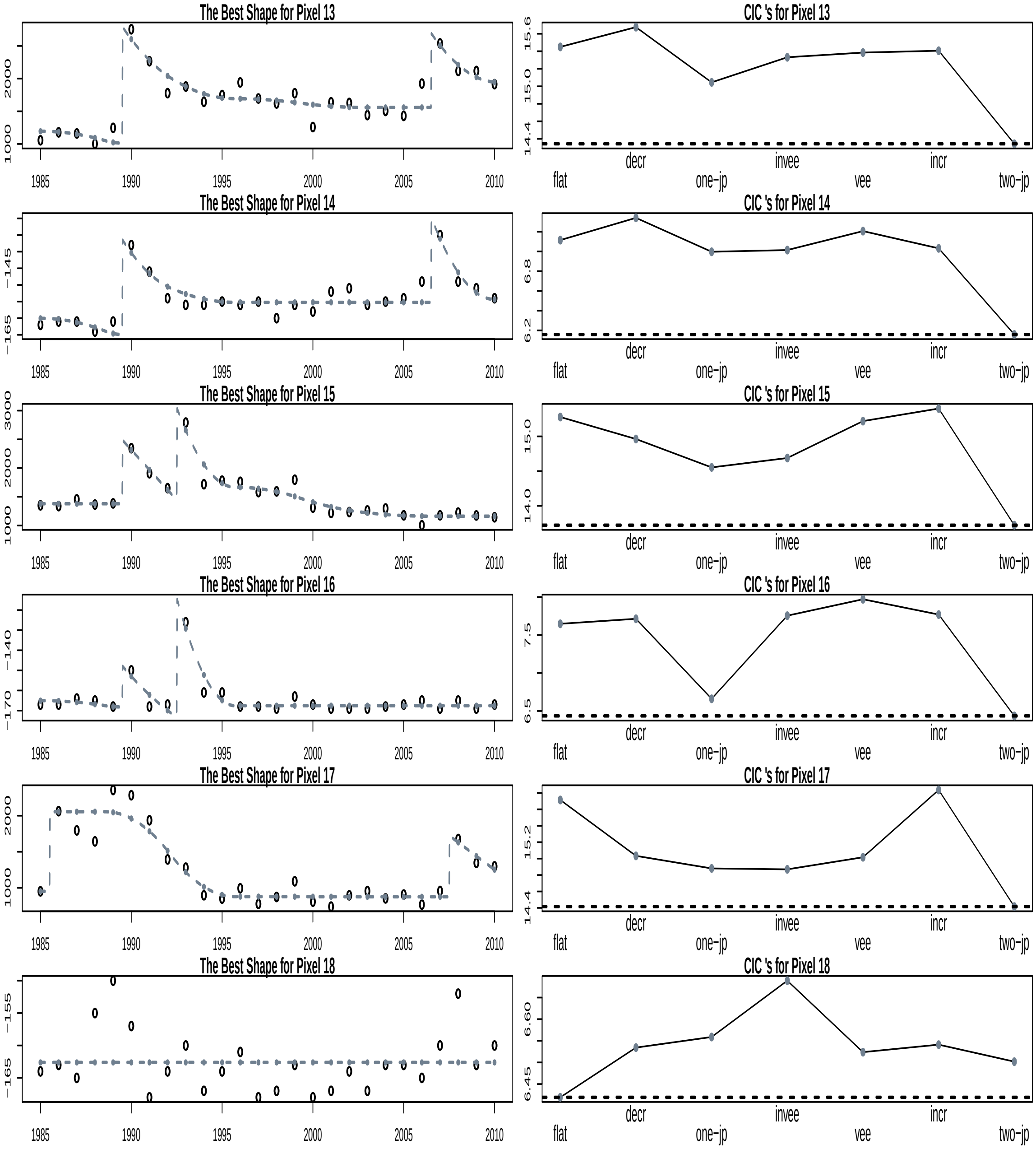

FIGURE 5. Left: the best shape chosen by the CIC criterion with fitted values marked as dots. Right: CIC values for seven possible shapes. The shape with the smallest CIC value is the winner.

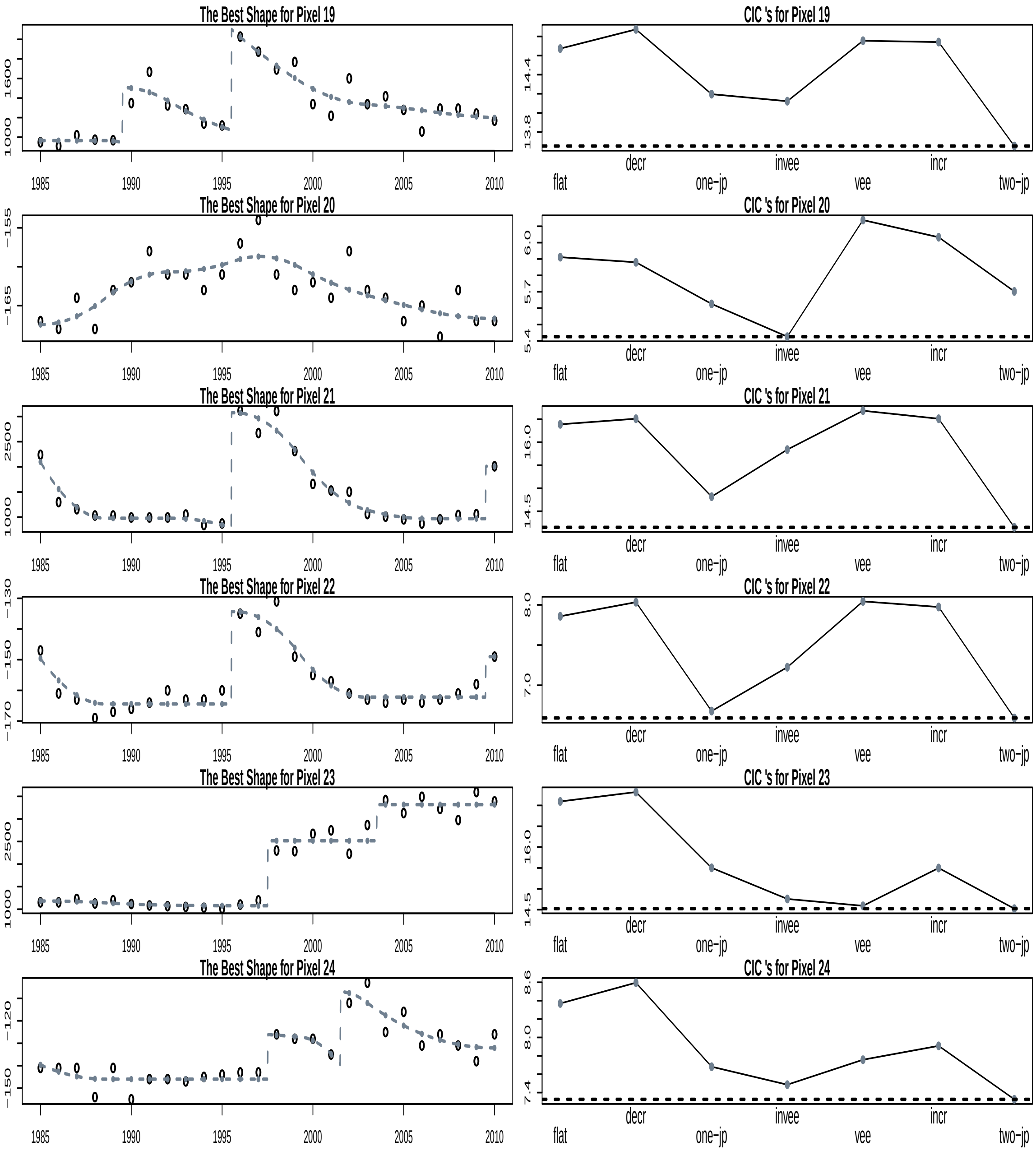

Figure 6. Left: the best shape chosen by the CIC criterion with fitted values marked as dots. Right: CIC values for seven possible shapes. The shape with the smallest CIC value is the winner.

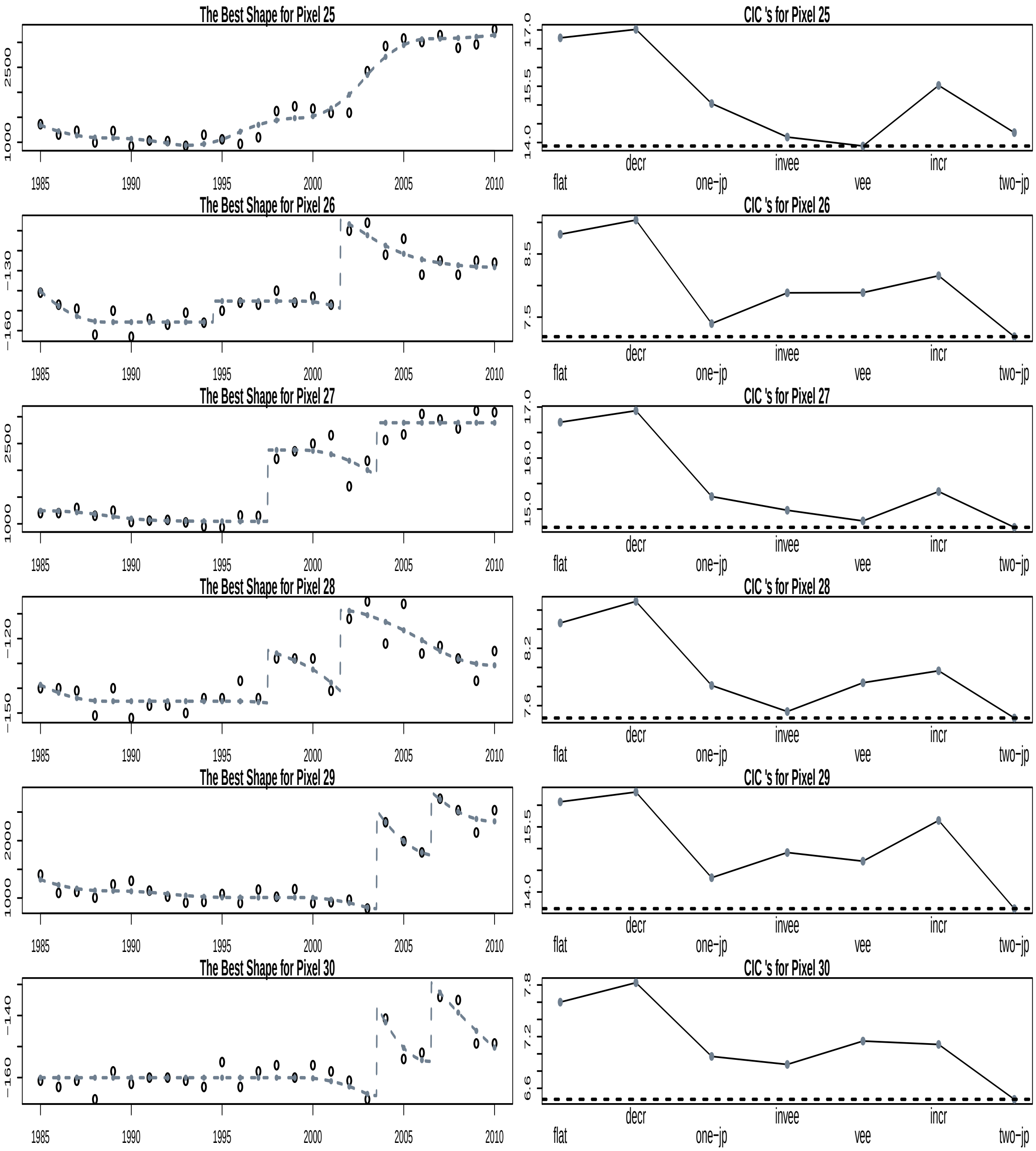

Figure 7. Left: the best shape chosen by the CIC criterion with fitted values marked as dots. Right: CIC values for seven possible shapes. The shape with the smallest CIC value is the winner.

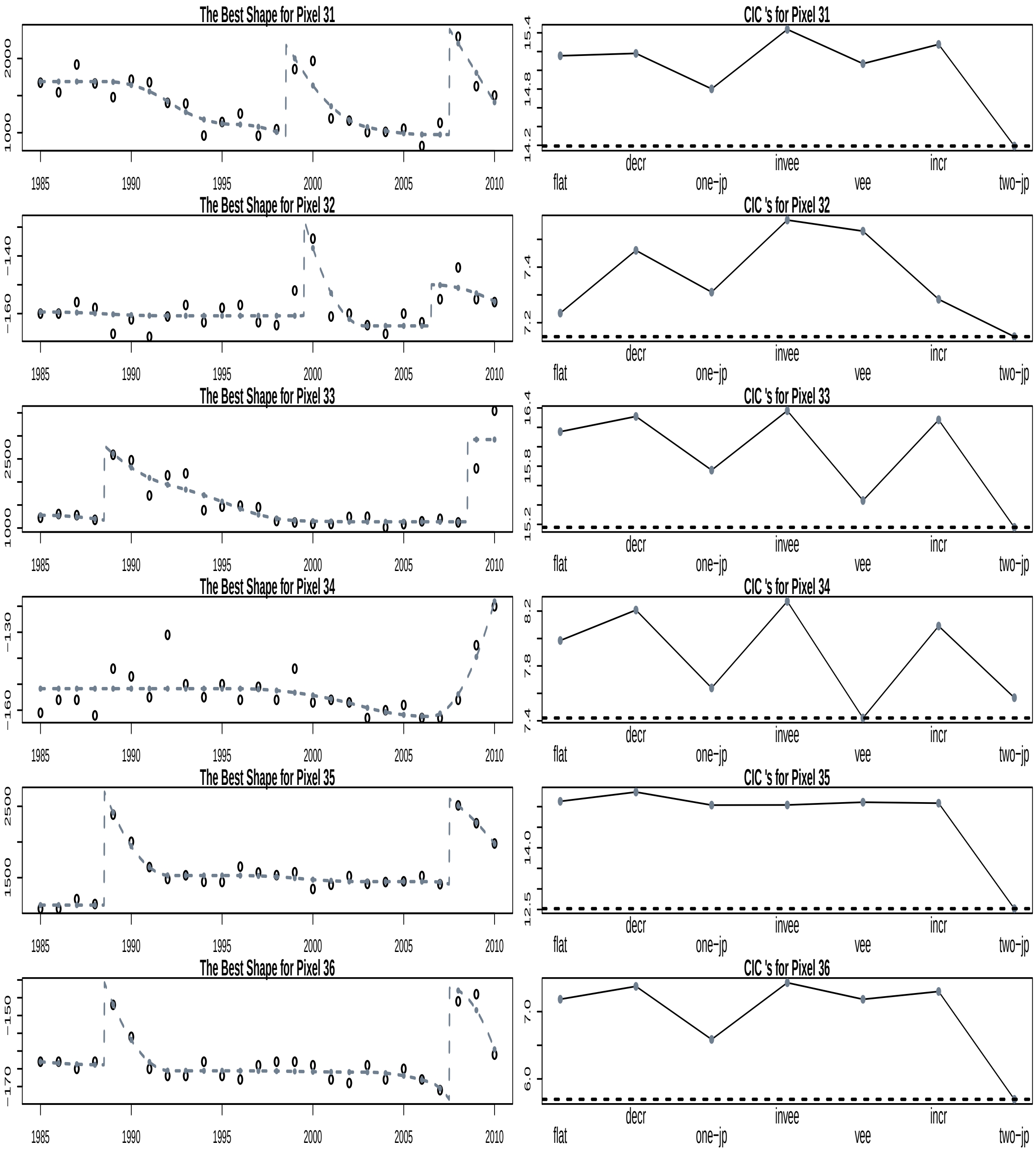

Figure 8. Left: the best shape chosen by the CIC criterion with fitted values marked as dots. Right: CIC values for seven possible shapes. The shape with the smallest CIC value is the winner.

CHAPTER 7

Conclusion and Future Work

7.1. Conclusion

In this dissertation, we propose change-point estimators of a mode, a jump point, and an inflection point, based on constrained splines, with or without penalty. It is shown by simulation that for small or moderate sample sizes, the proposed estimator performs well when compared to some existing estimators. Moreover, convergence rates are established such that the estimator is consistent, while consistency of the other smoothed mode or inflection-point estimators has not been established. The proposed estimators also allow the flexibility to include linear covariates in the model and to incorporate the scenario when the errors follow a stationary autoregressive process with short memory. Change-Point estimation with generalized linear models is also discussed, and the methods are applied to some real and simulated binomial and Poisson data sets.

The proposed methods are implemented in the `R` package `ShapeChange` Liao and Meyer (2016). We compared the speed of our routines with other `R` packages or routines discussed in Chapter 4, using a laptop with a 2.16GHz dual-core Intel(R) Celeron(R) CPU. For example, in the unimodal case, if we simulate data using the curve $f_m = 6x(1-x)$ with 100 observations, it takes roughly 40 milliseconds, 20 milliseconds and 2 minutes per call to get $\hat{m}_S$, $\hat{m}_P$ and $\hat{m}_U$ with 1000 repetitions. In the inflection-point case, we simulate using the curve $f_m = 10e^{-e^{5(1-2x)}}$ and 100 observations; it takes roughly 50 milliseconds, 1.4 seconds and 280 milliseconds per call to get $\hat{m}_S$, $\hat{m}_K$ and $\hat{m}_C$ with 1000 repetitions.

The project with FIA scientists is an application of shape-restricted estimation using B-splines, which emphasize jump-point detection in Landsat time series. It is proven to be an important addition to the Landsat community. A paper about the techniques in the `ShapeSelectForest` package is accepted by *Global Change Biology* as a technical advance paper. The algorithm in the `ShapeSelectForest` package will be implemented on Google Earth Engine in 2016, providing shape parameters and fitted trajectories for use in other disturbance, tree canopy cover, and biomass mapping projects.

7.2. Future Work

Like the analysis we make with the global annual precipitation data set in Chapter 4, we could develop model selection techniques and built them in the `ShapeChange` package as another option. For example, in the unimodal case, we can assume possible shapes of the underlying trend such as flat, increasing, increasing-decreasing etc, and we could choose the best shape according to the CIC criterion. If the best shape is unimodal, a mode estimate will be delivered and the autoregressive parameters will be estimated as well given a time series. Similar ideas can be applied to the jump-point case and the inflection-point case. In the jump-point case, we can assume possible shapes as we did in the `ShapeSelectForest` package but not limit possible data sets to Landsat data; in the inflection-point case, we can assume possible shapes like flat, increasing, convex-concave, or combinations of monotonicity and convexity. Some work has been done in terms of testing the existence of a jump point. We can explore hypothesis testing procedures about the existence of a mode, a jump point, and an inflection point of a regression function.

There are some other interesting change-point estimators which have not been much discussed. For example, the change-point where a regression function, which is flat for a wide range, turns upwards, the change-point where a linear regression function changes its sign, or multiple jump-points in a series.

REFERENCES

Christopoulos, D. T. (2013). *inflection: Finds the inflection point of a curve.*

Christopoulos, D. T. (2014). Developing methods for identifying the inflection point of a convex/concave curve. *arXiv:1206.5478v2*.

Cobb, G. W. (1978). The problem of the nile: Conditional solution to a changepoint problem. *Biometrika 65*(2), 243–251.

de Boor, C. (1978). *A Practical Guide to Splines.* New York: Springer-Verlag.

de Boor, C. (2001). *A Practical Guide to Splines* (Revised ed.). New York: Springer-Verlag.

Eddy, W. F. (1980). Optimum kernel estimators of the mode. *Annals of Statistics 8*, 870–882.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science 11*, 89–121.

Grégoire, G. and Z. Hamrouni (2001). Change point estimation by local linear smoothing. *Journal of Multivariate Analysis 83*, 56–83.

Greoneboom, P., G. Jongbloed, and J. Wellner (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Annals of Statistics 29*, 1653–1698.

Hastie, T. J. and R. J. Tibshirani (1990). *Generalized Additive Models.* Washington, D. C.: Chapman and Hall.

Horváth, L. and P. Kokoszka (2002). Change-point detection with non-parametric regression. *Statistics 36(1), 9-31.*

Huang, J. Z. (1998). Projection estimation in multiple regression with application to functional anova models. *Annals of Statistics 26*(1), 242–272.

Huang, J. Z. (2001). Concave extended linear modeling: A theoretical synthesis. *Statistica Sinica 11*, 173–197.

Kachouie, N. N. and A. Schwartzman (2013). Non-parametric estimation of a single inflection point in noisy observed signal. *Journal of Electrical and Electronic Systems*.

Köllmann, C. (2014). *uniReg: Unimodal penalized spline regression using B-splines.*

Köllmann, C., B. Bornkamp, and K. Ickstadt (2014). Unimodal regression using bernstein-schoenberg splines and penalties. *Biometrics 70*, 783–793.

Liao, X. and M. C. Meyer (2014). coneproj: An R package for the primal or dual cone projections with routines for constrained regression. *Journal of Statistical Software 61*(12), 1–22.

Liao, X. and M. C. Meyer (2016). *ShapeChange: Change-Point Estimation using Shape-Restricted Splines.*

Loader, C. R. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics 24*(4), 1667–1678.

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

Meyer, M. C. (2012a). Constrained penalized splines. *Canadian Journal of Statistics 40*(1), 190–206.

Meyer, M. C. (2012b). Nonparametric estimation of a smooth density with shape regressions. *Statistica Sinica 22*, 681–701.

Meyer, M. C. (2013). A simple new algorithm for quadratic programming with applications in statistics. *Communications in Statistics 42*(5), 1126–1139.

Meyer, M. C. and X. Liao (2014). *coneproj: Primal or Dual Cone Projections with Routines for Constrained Regression.*

Meyer, M. C., X. Liao, E. Freeman, and G. G. Moisen (2015). *ShapeSelectForest: Shape Selection for Landsat Time Series of Forest Dynamics.*

Meyer, M. C. and M. Woodroofe (2000). On the degrees of freedom in shape-restricted regression. *Ann. Statist. 28*, 1083–1104.

Müller, H. G. (1992). Change-points in nonparametric regression analysis. *Annals of Statistics 20*, 737–761.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton, New Jersey: Princeton.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric Regression.* Cambridge, UK: Cambridge University Press.

Shoung, J. and C. Zhang (2001). Least squares estimators of the mode of a unimodal regression function. *The Annals of Statistics 29*(3), 648–665.

Silvapulle, M. J. and P. Sen (2005). *Constrained Statistical Inference.* John Wiley & Sons.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Annals of Statistics 8*, 1348–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics 10*, 1040–1053.

Stone, C. J., M. H. Hansen, C. Kooperberg, and Y. K. Truong (1997). Polynomial splines and their tensor products in extended linear modeling. *Annals of Statistics 25*(4), 1371–1470.

Wand, M. (2014). *SemiPar: Semiparametic Regression.*

Wang, H., M. C. Meyer, and J. D. Opsomer (2013). Constrained spline regression in the presence of ar(p) errors. *Journal of Nonparametic Statistics 25*(4), 809–827.

Zhou, S., X. Shen, and D. A. Wolfe (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics 26*(5), 1760–1782.

Zhou, S. and D. A. Wolfe (2000). On derivative estimation in spline regression. *Statistica Sinica 10*, 93–108.