FUNCTIONAL SINGLE INDEX MODEL AND JENSEN EFFECT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zi Ye

August 2019

FUNCTIONAL SINGLE INDEX MODEL AND JENSEN EFFECT

Zi Ye, Ph.D.

Cornell University 2019

In natural ecosystems, environmental conditions are highly variable over time and space. There is much empirical evidence to show that environmental variability has significant effects on individuals and populations. The nonlinear response is modeled in this thesis in a functional single index model (FSIM), or its generalized version. In this thesis, we will investigate whether the impact is positive or negative. We will develop a nested nonlinear optimization algorithm with local quadratic approximation to estimate the second derivative of the curve in a FSIM. We will show convergence rates and consistency of the estimators. Even though our estimators perform very well theoretically, practical implementations require selections of initial value and bandwidth. Sometimes, our procedures are not able to select them properly, and will negatively impact estimation accuracy. Instead of estimating the curvature, we will estimate the Jensen Effect, or the sign of the Jensen's inequality, which only involves the link function, directly. Inspired by the SiZer method, we will skip the cross-validation step, but evaluate the Jensen Effect over a range of smoothing parameters. We will calculate a t-test statistics based on the Jensen Effect estimates, and perform a hypothesis test with critical values simulating from a Gaussian process. To analyze situations with variance heteroscedasticity in ecology data, we will consider logarithm of response variables, which lead to an exponential single index model. We will investigate the Jensen Effect, and propose a hypothesis test to see the impact of environmental factor. Then we will extend the methodology to generalized single index model for binary response, such as individual survival, or Poisson response, such as number of offspring.

BIOGRAPHICAL SKETCH

Zi Ye was born in Xi'an, China and spent most of her early time in Beijing. In summer 2011, she received her bachelor degree in Mathematics and Applied Mathematics from Shandong University. During her undergraduate study, she was awarded several scholarships, including Weichai Power Scholarship and Shandong University Scholarship for Outstanding Students.

She joined M.S. program in Applied Mathematics at University of Alberta in fall 2011. She was supervised by Dr. Mike Kouritzin and worked on several research projects on stochastic filtering. She received teaching and research assistantship during her master study. Her master dissertation is: *A Discrete-time Particle Filter and Central Limit Theorem*.

In fall 2013, she joined M.S./Ph.D. program in Statistics at Cornell University. During her study, Zi was a teaching assistant for introductory statistics and calculus, linear models and statistical computing courses. She also received research assistantship provided by Dr. Giles Hooker. In summer 2018, she was a biostatistician intern at Takeda Pharmaceutical. Her dissertation is supervised by Dr. Giles Hooker at Cornell University.

Dedicated to my family.

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my Ph.D. advisor, Dr. Giles Hooker. He is an excellent advisor and patiently provides me with his unreserved guidance and help. He always gives me great freedom to pursue independent study and research.

Moreover, I want to thank the rest of my committee members, Dr. Martin Wells and Dr. Stephen P. Ellner for their support, guidance and time during my study at Cornell University. They give me valuable advice to my research and dissertation.

Furthermore, I would like to thank all of my fellow statistics Ph.D. students and friends. They give me continuous and valuable support. I really enjoy the time spending with my friends in Ithaca.

Last but not least, special thanks to my family, especially my parents. They give me support and encouragement when I am facing with difficulties and challenges. Without them, all of these are not possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In natural ecosystems, environmental conditions are highly variable over time and space and many classical questions in ecology and evolution are therefore concerned with the potential consequences of this variation. Ecology theory predicts that nonlinear responses to environmental variability can play an important role in maintaining the diversity of competing species. We would like to develop statistical models and methods for determining whether environmental variability is beneficial or harmful for some component of population growth rate, survival probability or number of offspring, which will depend on whether the nonlinear response curve is concave-up or concave-down over the range of environmental variability.

Considering nonlinearity, we will model the relationship between environmental variability and species growth rates in a functional single index model, and investigate the sign of the Jensen's inequality, which will help us answer the ecology question. To model the count or binary response, we extend the functional single index model to the generalized model, with an additional link function added. We will estimate the effect of environment variability in a Poisson or logistic model.

First, we will investigate the growth model with a single link function. The nonparametric growth model discussed in the thesis is

$$\boldsymbol{G} = g(\boldsymbol{E}) + \epsilon, \tag{1.1}$$

where $\boldsymbol{G}$ and $\boldsymbol{E}$ are the growth (or future size) and environment of a species. We assume that the environment factor $\boldsymbol{E}$ is described by a climate history, such as temperature or precipitation observed at daily resolution. Following Teller et al. (2016), these are thought of as functional covariates, leading to representation of $\boldsymbol{E}$ as a functional linear

term

$$\boldsymbol{E} = \int X(t)\beta(t)\,\mathrm{d}t, \tag{1.2}$$

where $\beta(t)$ is the coefficient function to be estimated, and $X(t)$ is the observed climate history. Combining (1.1) and (1.2), a functional model for observed species growth is

$$Y = g\left(\int X(t)\beta(t)\,\mathrm{d}t\right) + \epsilon. \tag{1.3}$$

This is the Functional Single Index model introduced in Chen et al. (2011) and Ma (2016) with g and β functions unknown.

To assess the impact of variability in environment factor $\boldsymbol{E}$ on the growth rate $\boldsymbol{G}$, we need to compare $g[\mathrm{E}(\boldsymbol{E})]$ (growth under constant conditions) and $\mathrm{E}[g(\boldsymbol{E})]$ (growth under varying conditions). By the Jensen's inequality, we have that $g[\mathrm{E}(\boldsymbol{E})] \leq \mathrm{E}[g(\boldsymbol{E})]$ if g is a convex function, and a varying environment will accelerate the growth. Otherwise, if the link function g is concave, a constant environment at the average of environment factor $\boldsymbol{E}$ would be more beneficial. One way to find the convexity or concavity of a function g is to evaluate the sign of the second derivative g''.

In Chapter 2, to evaluate the nonlinear response of environment factor on the growth rate, we estimate the second derivative of the link function g in a certain range to see if the link function is convex or not. If the coefficient function β is known, we can use local nonparametric methods, such as a Nadaraya-Waston estimator, to get an estimate of g, and then estimate g'' by finite difference method, or use smoothing spline to estimate the curvature g'' directly. However, we have to estimate both the coefficient function and the second derivative of g in a functional single index model. Instead, we iteratively estimate g'' by local quadratic approximation, and β by minimizing the mean square error of the model. Inspired by the Projection Pursuit Regression (PPR) approach introduced in Hardle et al. (1993), our estimation procedure involves a two-step nested nonlinear optimization problem.

We show that the single index value $\int X\beta$ can be estimated root-n consistently. In addition, under some assumptions, we derive the rate of convergence of the estimator as

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\hat{g}''\left(\int X_i\hat{\beta}\right)-g''\left(\int X_i\beta\right)\right]^2=O\left(h_n^4+\frac{1}{nh_n^4}\right), \tag{1.4}$$

where h_n is the bandwidth in the local quadratic approximation and $h_n \to 0$. Even though our estimates of g'' and β have nice theoretical properties over a range of values for h_n, cross-validation step sometimes will fail in selecting the bandwidth to estimate g'' accurately. Moreover, practical implementation of the procedure requires solving a nonlinear optimization problem, and the estimates of the link function and the coefficient function are quite sensitive to the choices of initial values. The nonlinear optimization problem may converge to a local minimum without nicely selecting the bandwidth and initial value. Alternatively, we can generate a large set of initial values and select the best one from them. Considering the computational cost, we need to propose some other methods to evaluate the sign of the Jensen's inequality.

Practically speaking, we need a better estimate of g'' to conclude that in which environment, constant or varying, will accelerate the growth of a species. In Chapter 2, we observe that the second derivative g'' is difficult to estimate non-parametrically, but the estimate of the link function g obtained by our nested procedure is quite accurate. For computational reason, we will estimate the link function g by smoothing spline method, and minimize the penalized mean square error to estimate β. Instead of evaluating the curvature of g in growth model, in Chapter 3, we will examine the "Jensen Effect" directly, and estimate the quantity $\delta \doteq \mathrm{E}\left[g\left(\boldsymbol{E}\right)\right]-g\left[\mathrm{E}\left(\boldsymbol{E}\right)\right]$, or the sign of the Jensen's inequality, which only involves an estimate of g. Similar with local quadratic approximation, to estimate the link function g, we need to select a smoothing parameter λ. The functional single index model for growth has more complex structures, and cross-validation procedure is computationally inefficient. Sometimes, we are not able to select a good λ for estimation purpose.

The SiZer method, was first introduced in Chaudhuri and Marron (1999) for local linear regression, and then extended to smoothing splines settings in Marron and Zhang (2005). The SiZer is a graphical tool to make inference about features on an estimated curve, and detect features of that curve in an innovative way. The SiZer method skips the process of selecting a smoothing parameter or bandwidth by cross-validation step. Instead, it examines the estimate of a curve over a range of smoothing parameters, and observes features of the curve for each combination of observation point and smoothing parameter.

Inspired by the SiZer method, we estimate the "Jensen Effect" δ over a range of smoothing parameters λ, based on estimates of g and β. Then we generate a t-test statistics using the maximum or minimum value of δ over a range of λ, while the critical values are simulated from a Gaussian process. If the hypothesis test is significant and most of δ values are positive, we can conclude that the species will grow better in a varying environment; otherwise a constant environment will contribute to some components of growth success. Our simulation results show that the test is significant for concave and convex function, but not for linear function. In addition, our method also works well for the real North Temperate Lakes LTER data set.

Chapter 3 provides an answer to our ecology question for species growth model. There exists variance heteroscedasticity of the plant growth rates observed in the U.S. Sheep Experiment Station (USSES). To deal with it, we will take the logarithm of plant growth rates. In the previous chapter, we only investigate the nonlinear effect of environmental variability to the logarithm of growth rates. To estimate the Jensen Effect on the original response variables, we take the exponential on both sides and define it in an exponential single index model as

$$Y = \exp\left[g\left(X\beta\right) + \epsilon\right], \tag{1.5}$$

where the composite link function is $\exp(g)$. We know that exponential link will put a null hypothesis on convexity naturally, and we need to develop some hypothesis tests to overcome this assumption. Then, we can compare the growth rates under constant environment $\exp(g)[\mathrm{E}(\boldsymbol{E})]$ with respect to varying environment $\mathrm{E}[\exp(g)(\boldsymbol{E})]$ by analyzing the exponential model.

The exponential model is discussed in Chapter 4. We apply the Delta method to estimate the variance of the difference function δ. Then we obtain a t-test statistics to be compared with the null distribution generated from a Gaussian process. Our simulation results show that our test inspired by the SiZer method also works well for the model with an additional exponential link added.

In addition to analyzing growth rate of individuals or populations, we also want to consider data where the response is binary, such as individual survival, or event counts, such as number of offspring. So we want to figure out the relationship between environmental variability with survival probability or quantity counts of a species. In Chapter 5, it is naturally to extend the single index model to generalized model setting as

$$\eta \doteq g(X\beta), \tag{1.6}$$

$$\mu = h(\eta) = h[g(X\beta)], \tag{1.7}$$

$$Y \sim \text{Distribution}(\mu). \tag{1.8}$$

In (5.24), we define η as a single index model, and add another link function h, which is the generalized linear model (GLM) link. The composite link function observed here is $h \star g$. The response variable is generated by a random distribution with mean μ, determined by which generalized model we use. This is a Generalized Single Index model. If the link function h is exponential, the response variable Y is generated from a Poisson distribution. If the link function is a logistic function, we generate Y from a

Bernoulli distribution.

In Chapter 5, we use the penalized log-likelihood criteria to estimate β and g in the generalized single index model. The Jensen Effect δ and t-test statistics are generated based on our estimates.

CHAPTER 2

LOCAL QUADRATIC ESTIMATION AND CONVERGENCE RATES

2.1 Introduction

2.1.1 Ecological Motivation

Ecological theory predicts that nonlinear responses to environmental variability can play an important role in maintaining the diversity of competing species. For example, species competing for the same resources can nonetheless coexist if their population growth rates are maximized under different environmental conditions, which occur at different times or at different places; see e.g. Hutchinson (1961), Chesson and Warner (1981), Chesson (1994), Chesson (2000b) and Chesson (2000a). For different species, environmental variability can either increase, e.g., Drake (2005); Koons et al. (2009) or decrease population growth rate, e.g., Lewontin and Cohen (1969); Tuljapurkar (2013). The nonlinearity of species' responses to environmental variables also plays an important role in forecasting the effect of projected increases in environmental variability due to climate change.

The data motivating this study come from long-term observations of sagebrush steppe plant communities, see Adler et al. (2010), in which *Artemesia tripartita* (three-tipped sagebrush) is one of the dominant species. Previous models for this community, where *Artemisia* competes with three common grass species and other rarer species, predicted that *Artemisia*'s persistence is dependent on environmental variability – without it, *Artemisia* is gradually out-competed by the dominant grasses, see Adler et al. (2010). However, these models used simple parametric forms for underlying processes such

as plant growth and survival, with year-specific random effects representing between-year variability, and the assumed parametric forms largely determine whether variability is beneficial or harmful, depending on whether the response curve is concave-up or concave-down.

The goal of this paper is to provide improved estimation methods for understanding species' responses to climate variability. Traditional statistical models for individual plant or animal growth, survival and fecundity make parametric assumptions that imply specific forms of nonlinearity, particularly in the presence of high-dimensional covariates. Instead, we would like to use nonparametric models. We also want to model responses to specific covariates such as temperature and rainfall, rather than using random effects to capture between-year variability. The model can then be used to forecast responses under projected novel climatic conditions, different from those experienced during the period of data collection. We therefore consider the nonparametric model

$$\boldsymbol{G} = g(\boldsymbol{E}) + \epsilon, \tag{2.1}$$

where $\boldsymbol{G}$ and $\boldsymbol{E}$ are the growth and environment of a plant, g is a link function to be estimated and ϵ is the random error. To answer the ecological question, "Would the growth be higher if we just gave the plant a constant environment at the average of $\boldsymbol{E}$?", we need to compare $g[\mathrm{E}(\boldsymbol{E})]$ and $\mathrm{E}[g(\boldsymbol{E})]$.

If the link function g is convex, $g[\mathrm{E}(\boldsymbol{E})] \leq \mathrm{E}[g(\boldsymbol{E})]$ by the Jensen's inequality, and the plant grows better in a varying environment. Otherwise, if the link function g is concave, a constant environment is preferred. Assuming a smooth function g, convexity is equivalent to $g''(s) > 0$, for all s in the domain of g. Therefore, in this thesis we consider the problem of estimating the curvature of the link function g.

To complete this model, the environment $\boldsymbol{E}$ is described by the recent history of temperature and rainfall recorded at up to daily resolution. Since plants may be impacted

by climate events over a long period of time (see Dahlgren and Ehrlén, 2011; Clark et al., 2011), in our ecological data example, we will consider the past 22 years environmental histories. Following Teller et al. (2016), these are thought of as functional covariates leading to a representation of $\boldsymbol{E}$ as a functional linear term:

$$\boldsymbol{E} = \int X(t)\beta(t)\,\mathrm{d}t, \tag{2.2}$$

where $\beta(t)$ is the coefficient function to be estimated, and $X(t)$ is the covariate function we observed, typically a measurement of climate history.

The growth model of a plant is now given by

$$Y = g\left(\int X(t)\beta(t)\,\mathrm{d}t\right) + \epsilon. \tag{2.3}$$

This is Functional Single Index model, introduced in Chen et al. (2011) and Ma (2016).

In functional data analysis, a functional linear model (FLM) is defined as

$$Y = \int X(t)\beta(t)\,\mathrm{d}t + \epsilon, \tag{2.4}$$

which is often used in modeling the relationship between a functional covariate and a scalar response. To assess curvature, we need a more flexible model than the FLM. A generalized functional linear model (GFLM) is proposed in Müller and Stadtmüller (2005), James (2002) and Escabias et al. (2007), which has the same form as the functional single index model but with a known link function g. The functional single index model could be considered as an extension to the GFLM, as it is more flexible and could model a variety of real-world data.

Compared to the generalized functional linear model, estimation of the link function g based on a unknown coefficient function β is challenging. Even if β is known, estimating the second derivative of a nonparametric function directly is difficult. In this paper,

we prove a theoretical convergence rate for an estimate of g'' in the functional single index model, even if there is some bias in estimating the coefficient function β.

In this paper, we will use a local quadratic approximation to estimate link function g, and the coefficient function β is represented by a basis expansion whose parameters are obtained by minimizing mean square error of the model. The convergence rates that we derive are based on finding a global solution to a nonlinear optimization problem using a bandwidth that decreases at a known rate with n. However, this requires overcoming several practical issues. First, to find an optimum for β, we rely on nonlinear optimization methods which require an initial value from which to search for a minimum. Our experiments demonstrate that the performance of the estimate can depend critically on this choice of initial condition and natural choices which provide good estimates of g do not necessarily work well for g''. Further, the optimal choice of bandwidth can be quite different between estimation targeting g and that targeting g'' and we provide a heuristic post-cross-validation modification to improve the estimate of bandwidth. We expect similar rates of convergence will hold for alternative nonparametric estimators of g, penalized splines, for example, but that the specifics of smoothing parameter selection and nonlinear optimization can be expected to be quite different. A detailed analysis of the optimization problem is beyond the scope of this thesis.

2.1.2 Previous Results

In this section, we will introduce previous theoretical and empirical results for Single Index model and Functional Single Index model.

Single Index Model

There has been considerable research on the single index model, where the coefficient $\boldsymbol{\beta}$ is finite dimensional. The Single Index Model is defined as

$$Y = g(\boldsymbol{X}\boldsymbol{\beta}) + \epsilon, \tag{2.5}$$

where $\boldsymbol{X}$ is the covariate and $\boldsymbol{\beta}$ is the coefficient vector. Three methods have been used to estimate the link function g and the coefficient vector $\boldsymbol{\beta}$. The Projection Pursuit Regression (PPR) approach introduced in Hardle et al. (1993) is a two-step estimation procedure:

1. Estimate the link function g by the kernel method

$$\hat{g}_i(\boldsymbol{X}_i\boldsymbol{\beta}\,|\boldsymbol{\beta}, h) = \frac{\sum_{j\neq i} Y_j K\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta}-\boldsymbol{X}_j\boldsymbol{\beta}}{h}\right)}{\sum_{j\neq i} K\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta}-\boldsymbol{X}_j\boldsymbol{\beta}}{h}\right)}, \tag{2.6}$$

   where h is the bandwidth.

2. Estimate the coefficient $\boldsymbol{\beta}$ by minimizing the mean squared error

$$\hat{S}(\boldsymbol{\beta}, h) = \sum_{i=1}^{n} \left[Y_i - \hat{g}_i(\boldsymbol{X}_i\boldsymbol{\beta}\,|\boldsymbol{\beta}, h)\right]^2. \tag{2.7}$$

Hardle et al. (1993) proved that the coefficient vector $\boldsymbol{\beta}$ can be estimated root-n consistently. Ichimura (1993) showed the asymptotic normality of the estimator. The other two approaches provide new methods to estimate the coefficient vector. The Average Derivative approach in Hristache et al. (2001) showed that

$$\mathrm{E}\left[\frac{\partial g(\boldsymbol{X}\boldsymbol{\beta})}{\partial \boldsymbol{X}}\right] \overset{z\doteq \boldsymbol{X}\boldsymbol{\beta}}{=} \mathrm{E}\left[\frac{\partial g(z)}{\partial z}\boldsymbol{\beta}\right] = \mathrm{E}\left[\frac{\partial g(z)}{\partial z}\right]\boldsymbol{\beta} \doteq \gamma\boldsymbol{\beta}. \tag{2.8}$$

If we can find a consistent estimator of the average derivative $\mathrm{E}\left[\frac{\partial g(\boldsymbol{X}\boldsymbol{\beta})}{\partial \boldsymbol{X}}\right]$, we can get a consistent estimator of the coefficient $\boldsymbol{\beta}$, up to multiplication by a constant. Normally,

we require the coefficient vector to be norm 1. Stoker (1986) proposed two consistent estimators of the average derivative.

The third method, sliced inverse regression method in Li (1991), considered the estimation of the coefficient vector as a dimension-reduction problem. Any linear combination of the coefficient vector $\boldsymbol{\beta}$ is assumed to be an effective dimension-reduction (EDR) direction. They conduct a principle component analysis on the inverse regression space $\mathrm{E}(\boldsymbol{X}|Y)$, and estimate the coefficient vector $\boldsymbol{\beta}$ by the largest component.

Functional Single Index Model

There are only a few papers in the functional single index model. In Chen et al. (2011), similar to the projection pursuit regression in the single index model, the link function g and the coefficient function β are estimated by a two-step procedure. The coefficient function β is reduced to a finite dimensional coefficient vector by a spline basis. Under some assumptions, Chen et al. (2011) showed that

$$\frac{1}{n}\sum_{j=1}^{n}\left[\hat{g}\left(\int X_j\hat{\beta}\right)-g\left(\int X_j\beta\right)\right]^2 = O\left(n^{-c}\right), \tag{2.9}$$

for $c > 0$. In Ma (2016), two spline bases were used to represent the coefficient function and the link function, respectively, and the MSE was minimized iteratively until convergence. Ma (2016) constructed a asymptotic simultaneous confidence band for the coefficient function β. Our estimates follow Chen et al. (2011) but will examine the properties of $\hat{g}''$. By a clever decomposition of squared error, Chen et al. (2011) were able to avoid the need to directly account for the estimate of β. Unlike that case, to examine g'' we will need to obtain the $\sqrt{n}$ convergence rate for $\int X\beta$ directly, before we can examine our target.

2.2 Estimation Procedure

Suppose that we observe n environment histories and responses $(X_1(t), Y_1), \cdots, (X_n(t), Y_n)$, independent and identically distributed as $(X(t), Y)$, where $t \in \mathcal{I}$, with

$$Y = g\left(\int X(t)\beta(t)\,\mathrm{d}t\right) + \epsilon, \tag{2.10}$$

where Y is the scalar response variable, and $X(t)$ is the covariate function. For the purpose of simplification, we assume that the predictor X and the coefficient function β are defined in the same domain $\mathcal{I}$, and ϵ is a Gaussian random error.

To answer our ecological question, we are interested in estimating the second derivative (curvature) of the link function g. The estimate of the coefficient function β is denoted as $\hat{\beta}$. Define a Hilbert space $\mathcal{B}$ as the set of the coefficient functions β, where $\hat{\beta}(t), \beta(t) \in \mathcal{B}$.

To estimate β, g and g'', we use a local quadratic approximation. We assume that link function g is sufficiently smooth, such that the curvature, g'' is continuous. By Taylor expansion, at a fixed point x, the link function g can be approximated by

$$g(x) \approx g(x_0) + g'(x_0)(x - x_0) + g''(x_0)\frac{(x - x_0)^2}{2}. \tag{2.11}$$

Fix u, where u is in the domain of the link function g, the curvature, denoted as $\hat{g}''$, is estimated by minimizing the weighted sum of squares

$$\left(\hat{a}, \hat{b}, \hat{c}\right) = \tag{2.12}$$
$$\inf_{a,b,c} \sum_{i=1}^{n} \left\{ \left[Y_i - a - b\left(\int X_i(t)\beta(t)\,\mathrm{d}t - u\right) - c\frac{\left(\int X_i(t)\beta(t)\,\mathrm{d}t - u\right)^2}{2} \right]^2 K\left(\frac{\int X_i(t)\beta(t)\,\mathrm{d}t - u}{h_n}\right) \right\},$$

where K is a kernel function and h_n is the bandwidth. The estimators are then $(\hat{g}(u), \hat{g}'(u), \hat{g}''(u)) = \left(\hat{a}, \hat{b}, \hat{c}\right)$.

The coefficient function $\beta(t)$ is unknown in the penalized weighted sum of squares (2.12). We estimate it by minimizing the MSE

$$\hat{\beta} = \inf_{\beta \in \mathcal{B}} \sum_{i=1}^{n} \left[Y_i - \hat{g}\left(\int X_i \beta \right) \right]^2, \tag{2.13}$$

where

$$\hat{g}\left(\int X_i \beta \right) = \frac{\sum\limits_{j:j\neq i} Y_j K\left(\frac{\int X_i \beta - \int X_j \beta}{h_n} \right)}{\sum\limits_{j:j\neq i} K\left(\frac{\int X_i \beta - \int X_j \beta}{h_n} \right)}. \tag{2.14}$$

Since the kernel function K is only defined in $[-1, 1]$, we constrain the domain of the estimate of g or g'' to be in $[-1, 1]$ by normalizing the coefficients of g or g'' under an orthonormal basis to be 1 after optimizing procedure.

Denote a column vector $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^\top$. Fix $j \in \{1, \cdots, n\}$, and $\beta \in \mathcal{B}$, the estimated $\hat{g}\left(\int X_j \beta \right)$, $\hat{g}'\left(\int X_j \beta \right)$ and $\hat{g}''\left(\int X_j \beta \right)$ can be calculated as

$$\left(\hat{g}\left(\int X_j \beta \right), \hat{g}'\left(\int X_j \beta \right), \hat{g}''\left(\int X_j \beta \right) \right)^\top = \left(\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j} \mathbb{X}_{\beta,j} \right)^{-1} \left(\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j} \right) \boldsymbol{Y}, \tag{2.15}$$

where

$$\hat{g}'\left(\int X_j \beta \right) = \left(\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j} \mathbb{X}_{\beta,j} \right)_{2\cdot}^{-1} \left(\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j} \right) \boldsymbol{Y} \doteq S_1(\beta; j)\, \boldsymbol{Y}, \tag{2.16}$$

$$\hat{g}''\left(\int X_j \beta \right) = \left(\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j} \mathbb{X}_{\beta,j} \right)_{3\cdot}^{-1} \left(\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j} \right) \boldsymbol{Y} \doteq S_2(\beta; j)\, \boldsymbol{Y}, \tag{2.17}$$

where $\mathbb{A}_{k\cdot}$ denotes the k^{th} row of a matrix $\mathbb{A}$, and the $(n \times 3)$-dimensional matrix $\mathbb{X}_{\beta,j}$ is

$$\mathbb{X}_{\beta,j} = \left(\mathbf{1}, \int \boldsymbol{X}\beta - \left(\int X_j \beta \right) \mathbf{1}, \frac{\left(\int \boldsymbol{X}\beta - \left(\int X_j \beta \right) \mathbf{1} \right)^2}{2} \right), \tag{2.18}$$

with $\int \boldsymbol{X}\beta \doteq \left(\int X_1 \beta, \cdots, \int X_n \beta \right)^\top$, $\mathbf{1}$ is a n-dimensional column vector of ones, and the $(n \times n)$-dimensional matrix $\mathbb{K}_{\beta,j}$ is

$$\mathbb{K}_{\beta,j} = \operatorname{diag}\left(K\left(\frac{\int X_1 \beta - \int X_j \beta}{h_n} \right), \cdots, K\left(\frac{\int X_n \beta - \int X_j \beta}{h_n} \right) \right). \tag{2.19}$$

The estimation of the coefficient function β and the link function g is therefore a nested procedure, summarized in (2.13), (2.14) and (2.15). Following Ma (2016), the identifiability of the model is ensured by adding a constraint on the coefficient function, such that $\int \beta^2(t)\,\mathrm{d}t = 1$.

2.3 Assumptions

In deriving a convergence rate for $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\hat{g}''\left(\int X_i\hat{\beta}\right) - g''\left(\int X_i\beta\right)\right]^2$, we make the following assumptions in the functional single index model.

1. The observations $(X_i(t), Y_i)$, where $i = 1, \cdots, n$, are independent and identically distributed. Each covariate function $X_i(t)$ is a square-integrable function defined in the interval $\mathcal{I}$. The random error ϵ is independent from X, and has zero mean and variance σ^2.

2. The dependent variable Y has the mth-order absolute moment, where $m \geq 2$. This is an assumption from Ichimura (1993). The finite moment m is used in establishing the main convergence theorem.

3. The link function g and the curvature g'' are bounded and satisfy Lipschitz condition such that

$$\left|g^{(k)}(u) - g^{(k)}(v)\right| \leq D_2 |u - v|, \tag{2.20}$$

   for all u and v, where $D_2 > 0$ and $k = 0, 2$. The Lipschitz condition ensures that if β can be estimated root-n consistently, the distance between $g''\left(\int X\beta\right)$ and $g''\left(\int X\hat{\beta}\right)$ can be controlled.

4. The kernel function K is nonnegative and symmetric with support $[-1, 1]$, and $\int_{-1}^{1} K(s)\,\mathrm{d}s = 1$. Assume that K is three times continuously differentiable, with

$\left|K^{(3)}(s)\right| \leq D_3$, for any $s \in [-1, 1]$ and $D_3 > 0$.

5. For some orthonormal basis $\{\phi_k(t) : k = 1, 2, \cdots\}$, for each $i = 1, \cdots, n$, there exists a sequence of random variables $\left\{c_{ij}\right\}_{j=1}^{\infty}$, such that

$$X_i(t) = \sum_{j=1}^{\infty} c_{ij}\phi_j(t), \tag{2.21}$$

and

$$\beta(t) = \sum_{j=1}^{\infty} b_j\phi_j(t). \tag{2.22}$$

Assume that $\mathrm{E}\left(c_{ij}\right) = 0$.

In particular, we have

$$\hat{\beta}(t) = \sum_{j=1}^{\infty} \hat{b}_j\phi_j(t), \qquad \beta(t) = \sum_{j=1}^{\infty} b_j^0\phi_j(t). \tag{2.23}$$

For any $\beta \in \mathcal{B}$, we can write

$$\int X_i\beta = \sum_{j=1}^{\infty} c_{ij}b_j. \tag{2.24}$$

We observe that an orthonormal basis approximation of the covariate function and coefficient function transforms an integration to an infinite sum. In addition, define a sequence p_n such that $p_n \to \infty$ as $n \to \infty$, we require

$$\sum_{j=p_n+1}^{\infty} c_{ij}b_j = O\left(p_n^{-\lambda}\right), \tag{2.25}$$

where $\lambda > 1$, and $p_n = o\left(\frac{1}{h_n}\right)$. Condition (2.25) ensures that the integration $\int X\beta$ can be approximated by a finite sum of coefficients under an orthonormal basis.

6. Since the kernel function K is continuous with support $[-1, 1]$, we can conclude that K is bounded. We set a threshold C_0 for the response Y such that:

$$Y = \begin{cases} Y, & \text{if } |Y| < C_0, \\ C_0, & \text{otherwise.} \end{cases} \tag{2.26}$$

Therefore, Y is bounded. A Nadaraya-Watson estimator of g is:

$$\hat{g}\left(\int X_i\beta\right) = \frac{\sum\limits_{j:j\neq i} Y_j K\left(\frac{\int X_i\beta - \int X_j\beta}{h_n}\right)}{\sum\limits_{j:j\neq i} K\left(\frac{\int X_i\beta - \int X_j\beta}{h_n}\right)}, \tag{2.27}$$

For the true β, we can conclude that $\sum\limits_{j:j\neq i} K\left(\frac{\int X_i\beta - \int X_j\beta}{h_n}\right) \neq 0$ because $\frac{\int X_i\beta - \int X_j\beta}{h_n} = O\left(h_n^{\lambda-1}\right)$. Since the kernel function K satisfies the Lipschitz condition, the estimated $\hat{g}$ also satisfies the Lipschitz condition on β.

7. Assume that $\sup_{\beta\in\mathcal{B};x} f(x|\beta) < \infty$, where $f(x|\beta)$ is the probability density of $\int X\beta$.

2.4 Convergence Rates

By the definition of $\mathbb{X}_\beta^\top$ and $\mathbb{K}_\beta$, we can calculate

$$\begin{aligned}\mathbb{X}_\beta^\top \mathbb{K}_\beta &= \begin{pmatrix} \mathbf{1}^\top \\ \left(\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}\right)^\top \\ \left[\frac{\left(\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}\right)^2}{2}\right]^\top \end{pmatrix} \begin{pmatrix} K\left(\frac{\int X_1\beta - \int X_j\beta}{h_n}\right) & \mathbf{0}^\top & 0 \\ \mathbf{0} & \ddots & \mathbf{0} \\ 0 & \mathbf{0}^\top & K\left(\frac{\int X_n\beta - \int X_j\beta}{h_n}\right) \end{pmatrix} \\ &= \begin{pmatrix} \left[K\left(\frac{\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}}{h_n}\right)\right]^\top \\ \left[\left(\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}\right) K\left(\frac{\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}}{h_n}\right)\right]^\top \\ \left[\frac{\left(\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}\right)^2}{2} K\left(\frac{\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}}{h_n}\right)\right]^\top \end{pmatrix}. \end{aligned} \tag{2.28}$$

Denote

$$T_j^p(\beta) = \sum_{i=1}^n \left(\int X_i\beta - \int X_j\beta\right)^p K\left(\frac{\int X_i\beta - \int X_j\beta}{h_n}\right), \tag{2.29}$$

and

$$T_j^0(\beta) = \sum_{i=1}^n K\left(\frac{\int X_i\beta - \int X_j\beta}{h_n}\right). \tag{2.30}$$

Denote $u_j \doteq \int X_j\beta$, we have

$$\begin{aligned} T_j^0(\beta) =& \sum_{i=1}^{n} K\left(\frac{\int X_i\beta - \int X_j\beta}{h_n}\right) \\ =& n\int K\left(\frac{z-u_j}{h_n}\right) f(z|\beta)\,\mathrm{d}z + O(1) \\ =& n\int K(m)\, f\left(u_j + h_n m|\beta\right) h_n \mathrm{d}m + O(1) \\ =& nh_n f\left(u_j|\beta\right)\int K(m)\,\mathrm{d}m + O\left(h_n^2\right) + O(1) \\ =& nh_n f\left(\int X_j\beta|\beta\right) + O(1), \end{aligned} \tag{2.31}$$

and

$$\begin{aligned} T_j^p(\beta) =& \sum_{i=1}^{n}\left(\int X_i\beta - \int X_j\beta\right)^p K\left(\frac{\int X_i\beta - \int X_j\beta}{h_n}\right) \\ =& n\int \left(z-u_j\right)^p K\left(\frac{z-u_j}{h_n}\right) f(z|\beta)\,\mathrm{d}z + O(1) \\ =& n\int (h_n m)^p K(m)\, f\left(u_j + h_n m|\beta\right) h_n \mathrm{d}m + O(1) \\ =& nh_n^{p+1} f\left(u_j|\beta\right)\int m^p K(m)\,\mathrm{d}m + O\left(h_n^{p+2}\right) + O(1) \\ =& nh_n^{p+1} f\left(\int X_j\beta|\beta\right)\mu_p(K) + O(1), \end{aligned} \tag{2.32}$$

where $\mu_p(K) \doteq \int_{-1}^{1} m^p K(m)\,\mathrm{d}m$. Since the kernel function K is symmetric, $T_j^p(\beta) = 0$ if p is an odd number.

We have

$$\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j}\mathbb{X}_{\beta,j} \approx \begin{pmatrix} T_j^0(\beta) & T_j^1(\beta) & \frac{T_j^2(\beta)}{2} \\ T_j^1(\beta) & T_j^2(\beta) & \frac{T_j^3(\beta)}{2} \\ \frac{T_j^2(\beta)}{2} & \frac{T_j^3(\beta)}{2} & \frac{T_j^4(\beta)}{4} \end{pmatrix} = \begin{pmatrix} T_j^0(\beta) & 0 & \frac{T_j^2(\beta)}{2} \\ 0 & T_j^2(\beta) & 0 \\ \frac{T_j^2(\beta)}{2} & 0 & \frac{T_j^4(\beta)}{4} \end{pmatrix}. \tag{2.33}$$

The determinant of the matrix $\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j}\mathbb{X}_{\beta,j}$ is

$$\left|\mathbb{X}_{\beta,j}^\top \mathbb{K}_{\beta,j}\mathbb{X}_{\beta,j}\right| = \frac{T_j^0(\beta)\,T_j^2(\beta)\,T_j^4(\beta) - \left(T_j^2(\beta)\right)^3}{4}. \tag{2.34}$$

By definitions of $S_1(\beta; j)$ and $S_2(\beta; j)$ in (2.16) and (2.17), we can get

$$S_1(\beta; j) = \frac{\left[T_j^0(\beta) T_j^4(\beta) - \left(T_j^2(\beta)\right)^2\right]\left[\left(\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}\right) K\left(\frac{\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}}{h_n}\right)\right]^\top}{T_j^0(\beta) T_j^2(\beta) T_j^4(\beta) - \left(T_j^2(\beta)\right)^3}, \quad (2.35)$$

$$S_2(\beta; j) = \frac{2T_j^0(\beta)\left[\left(\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}\right)^2 K\left(\frac{\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}}{h_n}\right)\right]^\top - 2T_j^2(\beta)\left(K\left(\frac{\int \boldsymbol{X}\beta - \left(\int X_j\beta\right)\mathbf{1}}{h_n}\right)\right)^\top}{T_j^0(\beta) T_j^4(\beta) - \left(T_j^2(\beta)\right)^2}. \quad (2.36)$$

2.4.1 Convergence Rate of β

For any $\beta \in \mathcal{B}$, define

$$J_n(\beta) \doteq \frac{1}{n}\sum_{i=1}^n \left[Y_i - \hat{g}\left(\int X_i\beta\right)\right]^2 = \frac{1}{n}\sum_{i=1}^n \left[Y_i - \hat{g}\left(\sum_{j=1}^\infty c_{ij} b_j\right)\right]^2, \quad (2.37)$$

and

$$J_{n,p_n}(\beta) \doteq \frac{1}{n}\sum_{i=1}^n \left[Y_i - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij} b_j\right)\right]^2. \quad (2.38)$$

Denote $\boldsymbol{c}_i = (c_{i1}, \cdots)^\top$ and $\boldsymbol{b} = (b_1, \cdots)^\top$, where $i = 1, \cdots, n$. Define the subspaces $C, B \subset R^\infty$ such that $\boldsymbol{c}_i \in C$ and $\boldsymbol{b} \in B$.

Lemma 1. For a sequence of positive numbers M_n, suppose that $nh_n^2\epsilon_{0n}^2 M_n^{-2} \to \infty$, then

$$\mathrm{P}\left\{\sup_{C\times B}\left|\sum_{i=1}^n\left[\hat{g}_{ni}\left(\int x\beta\right) - \mathrm{E}\left(\hat{g}_{ni}\left(\int x\beta\right)\right)\right]\right| \geq \epsilon_{0n}\right\} \to 0, \quad (2.39)$$

where

$$\hat{g}_{ni}\left(\int x\beta\right) = \frac{1}{h_n} Y_i I\left(Y_i \in [-M_n, M_n]\right) K\left(\frac{\int x\beta - \int X_i\beta}{h_n}\right), \quad (2.40)$$

as $n \to \infty$.

Proof. We have

$$\begin{aligned}
&\mathrm{P}\left\{\sup_{C\times B}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}_{ni}\left(\int x\beta\right)-\mathrm{E}\left(\hat{g}_{ni}\left(\int x\beta\right)\right)\right]\right|\geq\epsilon_{0n}\right\}\\
&\quad=\mathrm{P}\left\{\sup_{C\times B}\left|\sum_{i=1}^{n}\left[Y_iI\left(Y_i\in[-M_n,M_n]\right)K\left(\frac{\int x\beta-\int X_i\beta}{h_n}\right)\right.\right.\right.\\
&\qquad\left.\left.\left.-\ \mathrm{E}\left(Y_iI\left(Y_i\in[-M_n,M_n]\right)K\left(\frac{\int x\beta-\int X_i\beta}{h_n}\right)\right)\right]\right|\geq nh_n\epsilon_{0n}\right\}\\
&\quad\doteq\mathrm{P}\left(\sup_{C\times B}\left|\sum_{i=1}^{n}A_i\right|\geq nh_n\epsilon_{0n}\right),
\end{aligned}$$

where

$$A_i=Y_iI\left(Y_i\in[-M_n,M_n]\right)K\left(\frac{\int x\beta-\int X_i\beta}{h_n}\right)-\mathrm{E}\left(Y_iI\left(Y_i\in[-M_n,M_n]\right)K\left(\frac{\int x\beta-\int X_i\beta}{h_n}\right)\right).$$

By Assumption 4, the kernel function K is bounded. We will apply Bernstein's inequality (see Appendix A.1) to the above equation with

$$\eta_n=nh_n\epsilon_{0n},$$

$$|A_i|\leq 2\left|Y_i\right|K_1\leq 2M_nC_1\doteq c_n,$$

and

$$\begin{aligned}
\mathrm{var}\left(A_i\right)&\leq M_n^2C_2,\\
V_n\doteq nM_n^2C_2&\geq\sum_{i=1}^{n}\mathrm{var}\left(A_i\right),
\end{aligned}$$

where K_1, C_1 and C_2 are constants. By Bernstein's inequality, we can get

$$\begin{aligned}
&\mathrm{P}\left\{\sup_{C\times B}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}_{ni}\left(\int x\beta\right)-\mathrm{E}\left(\hat{g}_{ni}\left(\int x\beta\right)\right)\right]\right|\geq\epsilon_{0n}\right\}\\
&\quad\leq\exp\left[-\frac{\eta_n^2}{2\left(V_n+\frac{1}{3}c_n\eta_n\right)}\right]\\
&\quad=\exp\left[-\frac{\left(nh_n\epsilon_{0n}\right)^2}{2\left(nM_n^2C_2+\frac{2}{3}M_nC_1nh_n\epsilon_{0n}\right)}\right]\\
&\quad=\exp\left[-\frac{nh_n\epsilon_{0n}^2}{\frac{2M_n^2C_2}{h_n}+\frac{4}{3}M_nC_1\epsilon_{0n}}\right].
\end{aligned}$$

The assumption of Lemma *A*.5 in Ichimura (1993) is that the sequence $\{M_n\}_{n=1}^{\infty}$ should satisfy $\epsilon_{0n} h_n M_n^{m-1} \to \infty$. Since $h_n \to 0$ and $\epsilon_{0n} \to 0$, we need $M_n \to \infty$. Therefore, in the denominator, we have $\frac{4}{3} M_n K_1 \epsilon_{0n} = o\left(\frac{2M_n^2 K_2}{h_n}\right)$. If $n h_n^2 \epsilon_{0n}^2 M_n^{-2} \to \infty$, then

$$-\frac{n h_n \epsilon_{0n}^2}{\frac{2M_n^2 K_2}{h_n} + \frac{4}{3} M_n K_1 \epsilon_{0n}} \to -\infty,$$

and

$$\mathrm{P}\left\{\sup_{C\times B}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}_{ni}\left(\int x\beta\right) - \mathrm{E}\left(\hat{g}_{ni}\left(\int x\beta\right)\right)\right]\right| \geq \epsilon_{0n}\right\} \to 0.$$

□

Following similar arguments, we can derive Lemmas 2 and 3 below.

Lemma 2. For a sequence of positive numbers M_n, suppose that $n h_n^4 \epsilon_{1n}^2 M_n^{-2} \to \infty$, then

$$\mathrm{P}\left\{\sup_{C\times B}\left|\sum_{i=1}^{n}\left[\hat{g}'_{ni}\left(\int x\beta\right) - \mathrm{E}\left(\hat{g}'_{ni}\left(\int x\beta\right)\right)\right]\right| \geq \epsilon_{1n}\right\} \to 0, \tag{2.41}$$

where

$$\begin{aligned}\hat{g}'_{ni}\left(\int x\beta\right) &= \left[\frac{1}{h_n} Y_i I\left(Y_i \in [-M_n, M_n]\right) K\left(\frac{\int x\beta - \int X_i\beta}{h_n}\right)\right]' \\ &\doteq \frac{1}{h_n^2} Y_i I\left(Y_i \in [-M_n, M_n]\right) H_1\left(x, X_i, \beta\right) K'\left(\frac{\int x\beta - \int X_i\beta}{h_n}\right),\end{aligned} \tag{2.42}$$

as $n \to \infty$.

Lemma 3. For a sequence of positive numbers M_n, suppose that $n h_n^6 \epsilon_{2n}^2 M_n^{-2} \to \infty$, then

$$\mathrm{P}\left\{\sup_{C\times B}\left|\sum_{i=1}^{n}\left[\hat{g}''_{ni}\left(\int x\beta\right) - \mathrm{E}\left(\hat{g}''_{ni}\left(\int x\beta\right)\right)\right]\right| \geq \epsilon_{2n}\right\} \to 0, \tag{2.43}$$

where

$$\begin{aligned}\hat{g}''_{ni}\left(\int x\beta\right) &= \left[\frac{1}{h_n} Y_i I\left(Y_i \in [-M_n, M_n]\right) K\left(\frac{\int x\beta - \int X_i\beta}{h_n}\right)\right]'' \\ &\doteq \frac{1}{h_n^3} Y_i I\left(Y_i \in [-M_n, M_n]\right) H_2\left(x, X_i, \beta\right) K''\left(\frac{\int x\beta - \int X_i\beta}{h_n}\right),\end{aligned} \tag{2.44}$$

as $n \to \infty$.

Now, we show the root-n consistency of the estimator $\hat{\beta}$.

Theorem 4. The estimator $\hat{\beta}$ is consistent if Assumptions 1 – 6 hold, and the bandwidth sequence satisfies $nh_n^8 \to 0$ and $nh_n^6 \to \infty$.

Proof. Theorem 5.1 in Ichimura (1993) states the consistency of $\hat{\beta}$, while β is a coefficient vector. Theorem 5.1 is based on Lemma 5.1, and Lemma 5.1 is based on Lemmas $A.2 - A.10$. The proof of Lemmas $A.2 - A.7$ will be the same whenever β is a vector or a function. Lemmas 1 – 3 above are the functional version of Lemmas $A.8 - A.10$. We need to figure out the constraint on the smoothing parameter h_n. The constraints are

- Lemma $A.2 - A.4$

$$nh_n^8 \to 0.$$

- Lemma $A.5 - A.7$

$$\epsilon_{0n} h_n M_n^{m-1} \to \infty, \quad \epsilon_{1n} h_n^2 M_n^{m-1} \to \infty, \quad \epsilon_{2n} h_n^3 M_n^{m-1} \to \infty.$$

- Lemma $A.8 - A.10$

$$n\epsilon_{0n}^2 h_n^2 M_n^{-2} \to \infty, \quad n\epsilon_{1n}^2 h_n^4 M_n^{-2} \to \infty, \quad n\epsilon_{2n}^2 h_n^6 M_n^{-2} \to \infty.$$

We require $nh_n^2 \epsilon_{0n}^2 M_n^{-2} \to \infty$. Since $\epsilon_{0n} \to 0$ and $M_n^{-2} \to 0$, we need to have $nh_n^2 \to \infty$. Following the same argument, we need $nh_n^4 \to \infty$ and $nh_n^6 \to \infty$. □

To prove the convergence rate of the functional single index model, we need to find a convergence rate for $\hat{\beta}$, $\hat{\boldsymbol{b}}$ or $\int X\hat{\beta}$.

Theorem 5. Suppose that $nh_n^6 \to \infty$, $nh_n^8 \to 0$ and $\frac{nh_n^{3+\frac{3}{m-1}}}{-\log h_n} \to \infty$, then we have

$$\sqrt{n}\left(\int X_i\hat{\beta} - \int X_i\beta\right) = O(1), \tag{2.45}$$

for $i = 1, \cdots, n$.

Proof. For each p_n such that $p_n \to \infty$ as $n \to \infty$, define

$$\hat{\boldsymbol{b}}_{p_n} = \left(\hat{b}_1, \cdots, \hat{b}_{p_n}\right), \text{ and, } \boldsymbol{b}^0_{p_n} = \left(b^0_1, \cdots, b^0_{p_n}\right).$$

By assumption 5, the estimated $\hat{g}$ satisfies the Lipschitz condition on the true coefficients b_j^0. Since $\sum_{j=p_n+1}^{\infty} c_{ij}b_j^0 = O\left(p_n^{-\lambda}\right)$, we have

$$\left|\hat{g}\left(\sum_{j=1}^{\infty} c_{ij}b_j^0\right) - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right| \le K_2\left|\sum_{j=p_n+1}^{\infty} c_{ij}b_j^0\right| = O\left(p_n^{-\lambda}\right), \tag{2.46}$$

as $n \to \infty$, where K_2 is a constant. Therefore, $\hat{g}\left(\sum_{j=1}^{\infty} c_{ij}b_j^0\right) - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right) \to 0$. By the definition of $J_{n,p_n}(\beta)$, we can get $J_{n,p_n}(\beta) \to J_n(\beta)$ as $n \to \infty$, for any $\beta \in \mathcal{B}$.

Hence, by Lemma 5.4 in Ichimura (1993), we have

$$\sqrt{n}\left(\hat{\boldsymbol{b}}_{p_n} - \boldsymbol{b}^0_{p_n}\right) \to D \doteq \mathrm{N}\left(\mathbf{0}_{p_n}, \mathbb{V}^{-1}\Sigma\mathbb{V}^{-1}\right),$$

where $\mathbf{0}_{p_n}$ is a p_n-dimensional mean vector, and $\mathbb{V}^{-1}\Sigma\mathbb{V}^{-1}$ is a $(p_n \times p_n)$-dimensional covariance matrix. Suppose that $\mathcal{X}$ is the σ-algebra generated by $(X_1, \cdots, X_n)$, the (k, m)th-term of the matrix $\mathbb{V}$ is

$$\begin{aligned}
\mathbb{V}_{km} =& \mathrm{E}\left[\frac{\partial^2 J_{n,p_n}(\beta)}{\partial b_k^0 \partial b_m^0}\,|\mathcal{X}\right] \\
=& \mathrm{E}\left\{\frac{2}{n}\sum_{i=1}^{n}\left[\hat{g}'\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]^2 c_{ik}c_{im} - \frac{2}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]\hat{g}''\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)c_{ik}c_{im}\,|\mathcal{X}\right\} \\
=& \mathrm{E}\left\{\frac{2}{n}\sum_{i=1}^{n}\left[\hat{g}'\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]^2 c_{ik}c_{im}\,|\mathcal{X}\right\} - \mathrm{E}\left\{\frac{2}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]\hat{g}''\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)c_{ik}c_{im}\,|\mathcal{X}\right\}.
\end{aligned}$$

For any $i = 1, \cdots, n$, we have

$$\mathrm{E}\left(Y_i \,|\mathcal{X}\right) = g\left(\int X_i\beta\right) = \mathrm{E}\left[\hat{g}\left(\int X_i\beta\right)|\mathcal{X}\right] + O\left(h_n^2\right), \tag{2.47}$$

where the second equality is the bias property of the kernel density estimate. We can calculate

$$\begin{aligned}
&\mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]|\mathcal{X}\right\} \\
&\quad = \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}\left(\int X_i\beta\right)\right]|\mathcal{X}\right\} + \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{g}\left(\int X_i\beta\right) - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]|\mathcal{X}\right\},
\end{aligned}$$

where the first term converges to 0 by the equation (2.47), since $h_n \to 0$ as $n \to \infty$, and the second term converges to 0 by (2.46). Therefore, we have

$$\mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]|\mathcal{X}\right\} \to 0,$$

as $n \to \infty$. By the Slutsky's Theorem, we can get

$$\mathrm{E}\left\{\frac{2}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]\hat{g}''\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)c_{ik}c_{im}\,|\mathcal{X}\right\} \to 0.$$

Therefore,

$$\mathbb{V}_{km} = \mathrm{E}\left\{\frac{2}{n}\sum_{i=1}^{n}\left[\hat{g}'\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]^2 c_{ik}c_{im}\,|\mathcal{X}\right\}.$$

For any $\beta \in \mathcal{B}$ and any $j \in \{1, \cdots, n\}$, we have

$$\begin{aligned}
\left|\hat{g}'\left(\int X_j\beta\right)\right| &= |S_1(\beta; j)\, \boldsymbol{Y}| \\
&= \left|\frac{\left[T_j^0(\beta) T_j^4(\beta) - \left(T_j^2(\beta)\right)^2\right] \sum_{i=1}^{n} \left(\int X_i\beta - \int X_j\beta\right) K\left(\frac{\int X_i\beta - \int X_j\beta}{h_n}\right) Y_i}{T_j^0(\beta) T_j^2(\beta) T_j^4(\beta) - \left(T_j^2(\beta)\right)^3}\right| \\
&\leq \left|\frac{\left[T_j^0(\beta) T_j^4(\beta) - \left(T_j^2(\beta)\right)^2\right] T_j^1(\beta) \left|\max_{i\in\{1,\cdots,n\}} Y_i\right|}{T_j^0\left(\hat{\beta}\right) T_j^2(\beta) T_j^4(\beta) - \left(T_j^2(\beta)\right)^3}\right| \\
&\approx \left|\frac{\left[nh_n f(u|\beta)\, nh_n^5 f(u|\beta)\, \mu_4(K) - \left(nh_n^3 f(u|\beta)\, \mu_2(K)\right)^2\right] nh_n^2 f(u|\beta)\, \mu_1(K) \left|\max_{i\in\{1,\cdots,n\}} Y_i\right|}{nh_n f(u|\beta)\, nh_n^3 f(u|\beta)\, \mu_2(K)\, nh_n^5 f(u|\beta)\, \mu_4(K) - \left(nh_n^3 f(u|\beta)\, \mu_2(K)\right)^3}\right| \\
&\sim \frac{1}{h_n},
\end{aligned}$$

where $u = \int X_j\beta$.

For any $i = 1, \cdots, n$, denote $c_{\cdot\cdot} = c_{ik}$, for any $k \in \{1, \cdots, n\}$, and $c_{\cdot k} = c_{ik}$ for a fix k. For any $k, m \in \{1, \cdots, n\}$, we have

$$\mathbb{V}_{km} \sim \frac{\mathrm{E}(c_{ik}c_{im}|\mathcal{X})}{h_n^2} \sim \frac{\mathrm{cov}(c_{ik}, c_{im})}{h_n^2} \sim \frac{\mathrm{cov}(c_{\cdot k}, c_{\cdot m})}{h_n^2},$$

since X_i are independent and identically distributed with $\mathrm{E}\left(c_{ij}\right) = 0$, for any $i = 1, \cdots, n$ and $j = 1, \cdots$.

The (k, m)th-term of the matrix Σ is

$$\begin{aligned}
\Sigma_{km} =& \mathrm{E}\left[\sigma\frac{\partial J_{n,p_n}(\beta)}{\partial b_k^0} \cdot \sigma\frac{\partial J_{n,p_n}(\beta)}{\partial b_m^0}\right] \\
=& \mathrm{E}\left[\frac{2\sigma}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]\hat{g}'\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)c_{ik} \cdot \frac{2\sigma}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)\right]\hat{g}'\left(\sum_{j=1}^{p_n} c_{ij}b_j^0\right)c_{im}\right] \\
\sim& \frac{\mathrm{cov}(c_{\cdot k}, c_{\cdot m})}{h_n^2}.
\end{aligned}$$

The diagonal term of the covariance matrix $\mathbb{V}^{-1}\Sigma\mathbb{V}^{-1}$ is

$$\left(\mathbb{V}^{-1}\Sigma\mathbb{V}^{-1}\right)_{kk} \sim \frac{p_n^2 h_n^2}{\mathrm{var}(c_{\cdot\cdot})}.$$

Define the truncated version of $X_i(t)$ and $\beta(t)$ as

$$X_{i,p_n}(t) = \sum_{j=1}^{p_n} c_{ij}\phi_j(t),$$
$$\beta_{p_n}(t) = \sum_{j=1}^{p_n} b_j\phi_j(t),$$

where $i = 1, \cdots, n$.

We have for any $i = 1, \cdots, n$, define $\boldsymbol{c}_{i,p_n} = \left(c_{i1}, \cdots, c_{ip_n}\right)$,

$$\sqrt{n}\left(\int X_{i,p_n}\hat{\beta}_{p_n} - \int X_{i,p_n}\beta_{p_n}\right) = \sqrt{n}\left(\boldsymbol{c}_{i,p_n}\hat{\boldsymbol{b}}_{p_n} - \boldsymbol{c}_{i,p_n}\boldsymbol{b}^0_{p_n}\right) = \boldsymbol{c}_{i,p_n}\sqrt{n}\left(\hat{\boldsymbol{b}}_{p_n} - \boldsymbol{b}^0_{p_n}\right).$$

Therefore, $\sqrt{n}\left(\int X_{i,p_n}\hat{\beta}_{p_n} - \int X_{i,p_n}\beta_{p_n}\right)$ converges to a normal distribution with the covariance matrix $\mathrm{var}(c_{..}) \cdot \frac{p_n^2 h_n^2}{\mathrm{var}(c_{..})} = p_n^2 h_n^2$. By Assumption 5, since $p_n = o\left(\frac{1}{h_n}\right)$, we can get $\mathrm{var}(c_{..}) \cdot \frac{p_n^2 h_n^2}{\mathrm{var}(c_{..})} = o(1)$. When $n \to \infty$, $p_n \to \infty$, we have

$$\sqrt{n}\left(\int X_i\hat{\beta} - \int X_i\beta\right) = O(1),$$

for $i = 1, \cdots, n$. □

2.4.2 Main Theorem

Theorem 6. If $nh_n^6 \to \infty$, $nh_n^8 \to 0$ and $\frac{nh_n^{3+\frac{3}{m-1}}}{-\log h_n} \to \infty$, we have

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\hat{g}''\left(\int X_i\hat{\beta}\right) - g''\left(\int X_i\beta\right)\right]^2 = O\left(h_n^4 + \frac{1}{nh_n^4}\right). \tag{2.48}$$

Proof. Since X_i are independent and identically distributed, for any $j \in \{1, \cdots, n\}$, we only need to find the convergence rate of $\mathrm{E}\left[\hat{g}''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\beta\right)\right]^2$. We can decom-

pose it into three terms:

$$\begin{aligned}
&\mathrm{E}\left[\hat{g}''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\beta\right)\right]^2 \\
&\quad \le \mathrm{E}\left[\hat{g}''\left(\int X_j\hat{\beta}\right) - \bar{g}''\left(\int X_j\hat{\beta}\right)\right]^2 + \mathrm{E}\left[\bar{g}''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\hat{\beta}\right)\right]^2 \\
&\qquad + \mathrm{E}\left[g''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\beta\right)\right]^2.
\end{aligned}$$

where

$$\bar{g}''\left(\int X_j\hat{\beta}\right) \doteq S_2\left(\hat{\beta}; j\right)\boldsymbol{g},$$

and

$$\boldsymbol{g} \doteq \left(g\left(\int X_1\beta\right), \cdots, g\left(\int X_n\beta\right)\right)^\top.$$

By Lemma 7, 8 and 9 in the Appendix A.2, we have

$$\begin{aligned}
\mathrm{E}\left[\hat{g}''\left(\int X_j\hat{\beta}\right) - \bar{g}''\left(\int X_j\hat{\beta}\right)\right]^2 &= O\left(\frac{1}{nh_n^4}\right). \\
\mathrm{E}\left[\bar{g}''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\hat{\beta}\right)\right]^2 &= O\left(h_n^4 + \frac{1}{nh_n^4}\right). \\
\mathrm{E}\left[g''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\beta\right)\right]^2 &= O\left(\frac{1}{n}\right).
\end{aligned}$$

Combining these three terms, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\hat{g}''\left(\int X_i\hat{\beta}\right) - g''\left(\int X_i\beta\right)\right]^2 = O\left(h_n^4 + \frac{1}{nh_n^4}\right).$$

□

2.5 Practical Implementation

2.5.1 Initialization

In (2.13), the coefficient function β is estimated by minimizing the mean square error of g, which is a nonlinear optimization problem. The coefficient function is approximated by a K-dimensional Fourier basis:

$$\beta(t) = \boldsymbol{\psi}^{\top}(t)\,\boldsymbol{c}, \tag{2.49}$$

where $\boldsymbol{c}$ is a K-dimensional column vector, and $t \in [0, 1]$. In order to ensure identifiability, we constrain the coefficient function β to have $\int \beta = 0$ and $\|\beta\| = 1$. Since the Fourier basis is an orthonormal basis, a constraint on β is equivalent to a constraint on the coefficient vector, such that $\|\boldsymbol{c}\| = 1$. The first constraint can be enforced by dropping any constant terms from the Fourier basis. To compensate for rescaling $\boldsymbol{c}$, we also rescale the bandwidth h such that $h = h\,\|\boldsymbol{c}\|$ in the optimization step.

We use `R` function `optim` to minimize the MSE, and an initial value of the coefficient vector $\boldsymbol{c}$ is needed, denoted as $\boldsymbol{c}_{\text{init}}$. We use three different methods to select the initial vector:

1. Assume each item in $\boldsymbol{c}_{\text{init}}$ is equal, such that $\boldsymbol{c}_{\text{init}} = \left(\frac{1}{\sqrt{K}}, \cdots, \frac{1}{\sqrt{K}}\right)$. We don't have any previous knowledge about the true functional form, so for simplicity, we assume that all coefficients in the initial vector are equal to each other.

2. Assume $g(s) = s$, the functional single index model reduced to a functional linear model. The coefficient vector $\boldsymbol{c}_{\text{init}}$ is estimated by minimizing MSE of a linear regression model, and normalize it to be $\|\boldsymbol{c}_{\text{init}}\| = 1$. To obtain an initial vector for β, we need to specify a structure for the link function g. A linear structure of g is

obviously the simplest, and could be calculated easily.

3. Generate 1000 different standard normal distribution random initial vectors (with the same length as $\boldsymbol{c}$), and select the first 10 initial vectors ordered by mean squared error of g, see (2.13), increasingly, when the bandwidth is fixed as the mean of bandwidth sequence. In the cross-validation step, we select a bandwidth h and an initial vector among the 10 vectors, that minimizes the cross-validation values jointly.

2.5.2 Cross-validation

We need to select the bandwidth h_n in the kernel density estimation of g and g''. We examine two cross-validation methods:

1. 10-fold cross-validation. We partition the dataset into 10 subsamples. Each time, we use 9 subsamples as the training set to fit the model and the remaining subsample as the validation set to calculate prediction error. We select the bandwidth h_n that minimizes the prediction error. We will observe that the 10-fold cross-validation method produces similar results with the GCV method introduced below.

2. Fix $j \in \{1, \cdots, n\}$, define

$$\hat{g}\left(\int X_j \beta\right) = \left(\mathbb{X}_{\beta,j}^{\top} \mathbb{K}_{\beta,j} \mathbb{X}_{\beta,j}\right)_{1\cdot}^{-1} \left(\mathbb{X}_{\beta,j}^{\top} \mathbb{K}_{\beta,j}\right) \boldsymbol{Y} \doteq S_0(\beta; j)\, \boldsymbol{Y}, \tag{2.50}$$

where $S_0(\beta; j)$ is a n-dimensional row vector. Denote a $(n \times n)$-dimensional smoother matrix $\mathbb{S}_{h_n} = \left(S_0\left(\hat{\beta}; 1\right), \cdots, S_0\left(\hat{\beta}; n\right)\right)^{\top}$. We can get $\hat{\boldsymbol{Y}} = \mathbb{S}_{h_n} \boldsymbol{Y}$, where

$\hat{\boldsymbol{Y}} \doteq \left(\hat{g}\left(\int X_1\hat{\beta}\right), \cdots, \hat{g}\left(\int X_n\hat{\beta}\right)\right)^\top$. The generalized cross-validation criterion is

$$\text{GCV}\left(h_n\right) \doteq \frac{\frac{1}{n}\left\|\left(\mathbb{I} - \mathbb{S}_{h_n}\right)\boldsymbol{Y}\right\|_2^2}{\left[\frac{1}{n}\text{tr}\left(\mathbb{I} - \mathbb{S}_{h_n}\right)\right]^2}, \tag{2.51}$$

where $\mathbb{I}$ is a n-dimensional identity matrix. We choose the bandwidth h_n that minimizes the GCV (h_n). Note that this does not account for estimating $\hat{\beta}$.

2.5.3 Simulation Study

In order to obtain simulated functional data, we defined a 25-dimensional Fourier basis $\boldsymbol{\psi}(t)$, where $t \in [0, 1]$. The covariate function $X(t)$ is defined in $[0, 1]$, such that

$$X(t) = \sum_{i=1}^{25} \eta_i \psi_i(t), \tag{2.52}$$

where $\eta_i \sim \text{N}\left(0, \left(\frac{i-1}{24}\right)^2\right)$. The coefficient function is

$$\beta(t) = \boldsymbol{\psi}^\top(t)\,\boldsymbol{c}, \tag{2.53}$$

where $\boldsymbol{c} \doteq (0, 1, 1, 0.5, 0, \cdots, 0)^\top$ chosen to yield a $\beta(t)$ that is reasonably. We use three different link functions:

1. $g(s) = e^{-s}$.
2. $g(s) = -s^2$.
3. $g(s) = s$.

In order to measure the performance of our estimators, we define the MSE of the estimated β and $g^{(k)}$ to be

$$\text{RSE} \doteq \left[\int \left(\hat{\beta}(t) - \beta(t)\right)^2 \text{d}t\right]^{\frac{1}{2}}, \tag{2.54}$$

and

$$\text{RASE}(k) \doteq \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\hat{Y}_i^{(k)} - g^{(k)} \left(\int X_i(t) \beta(t) \, \mathrm{d}t \right) \right]^2 \right\}^{\frac{1}{2}}, \tag{2.55}$$

where $\hat{Y}_i^{(k)} = \hat{g}^{(k)} \left(\int X_i(t) \hat{\beta}(t) \, \mathrm{d}t \right)$.

Theorem 6 shows that for consistency, the bandwidth h_n for local quadratic approximation must scale between $O\left(n^{-\frac{1}{6}}\right)$ and $O\left(n^{-\frac{1}{8}}\right)$. If we assume that the optimal bandwidth for the curvature g'' is $n^{-\frac{1}{7}}$, by Chen et al. (2011), we can expect that the optimal bandwidth for the link function g is $n^{-\frac{1}{5}}$. We use either the GCV or the 10-fold cross-validation to select a bandwidth h, which is considered as the optimal bandwidth for g. To obtain g'', we re-estimate g, keeping $\hat{\beta}$ fixed using $h^{\frac{5}{7}}$ as a re-scaled bandwidth. As we have discussed before, we constrain the coefficient $\|\beta\| = 1$; this was achieved by rescaling the bandwidth in our objective function. After solving the nonlinear optimization problem, we rescale both $\boldsymbol{c}$ and the bandwidth. The following table shows the simulation results of the GCV method. The number of data points we use are $n = 100$ and $n = 1000$. In Table 2.2, we also show the RASE2 results without rescaling the bandwidth h to be $h\|\boldsymbol{c}\|$ and $h^{\frac{5}{7}}$, if we start from random initial vector. We can conclude that re-scaling does matter to the final results and it reduces the error of RASE2. The 10-fold cross-validation results in the Appendix A.3 also confirm that by re-scaling the bandwidth, we improve our estimate for g''. We observe that RSE and RASE, as expected, achieve best performance when we initialize our optimizer at the true values. However, the more natural Linear initialization strategy does not outperform initializing at Equal coefficients. For both RASE and RASE2, a more intensive search over initializations pays off; in the case of RASE2 this even outperforms starting from true values. We suspect that this is associated with differing optimal smoothness criteria. Table 2.2 compares our results to when we do not use the $h^{\frac{5}{7}}$ re-scaling where applying this has a significant effect; by starting from a position far from the optimum, a large band-

Table 2.1: Simulation results by GCV using rescaled bandwidth $h\|\boldsymbol{c}\|$ and constraint $\hat{\beta}$.

		g1			g2			g3		
Initial	**n**	**RSE**	**RASE**	**RASE2**	**RSE**	**RASE**	**RASE2**	**RSE**	**RASE**	**RASE2**
True	100	0.2887	0.0959	9.4156	0.2785	0.0945	7.2648	0.3442	0.0918	5.0951
	1000	0.0829	0.0336	0.9003	0.0731	0.0334	0.5429	0.1095	0.0276	0.8620
Linear	100	1.1401	0.3401	1.3869	1.3569	0.4166	1.2014	0.7049	0.1393	0.2507
	1000	1.9010	0.2340	0.5205	1.0719	0.3421	1.3075	0.5287	0.0891	0.0729
Equal	100	0.8389	0.2820	1.0933	0.8342	0.2566	1.4349	0.8329	0.1857	0.2589
	1000	0.8502	0.3159	1.0133	0.8437	0.2838	1.4043	0.8508	0.1921	0.0905
Random	100	1.1911	0.1461	0.5597	1.1617	0.1979	0.6186	1.2603	0.0942	0.2043
	1000	1.2704	0.1320	0.4354	1.3021	0.1890	0.4413	1.2300	0.0796	0.0637

width may have the effect of smoothing the objective function, indirectly improving our re-estimate of g".

Overall, the random initial produces a much better result compared to other initial strategies. Starting from 1000 initial vectors, we increase the probability of selecting a "good" starting point and decrease the chance of converging to a local minimum for the nonlinear optimization problem, although this comes at a significant computational cost. In addition, we do observe an improvement of RASE and RASE2 when $n = 1000$ compared to $n = 100$.

Table 2.2: Simulation results of random starting value with rescaled and original bandwidths selected by GCV

	g1		g2		g3	
n	**original**	**rescaled**	**original**	**rescaled**	**original**	**rescaled**
100	1.2793	0.5597	2.1497	0.6186	0.2360	0.2043
1000	1.1297	0.4354	2.2081	0.4413	0.0703	0.0637

In order to provide a visual sense of the performance of our estimate, in Figures 2.1 and 2.2, we plot the estimates of the link function $g(s) = e^{-s}$ and $g(s) = s$, respectively. We observe that the estimate and the true curve of the link function g almost overlap with each other, but the second derivative has significantly larger error relative to the truth.

2.6 Ecological Data

2.6.1 Model Formulation

Examining our ecological questions, the purpose of estimating the second derivative of the link function in a functional single index model is to figure out whether the link function g is convex or concave. Then, we can answer the question: in which environment, constant or varying, the plant will grow better.

We apply our methods to data collected between 1926 and 1957 at the U.S. Sheep Experiment Station. Each individual plant in the species *Artemisia tripartita* (ARTR)

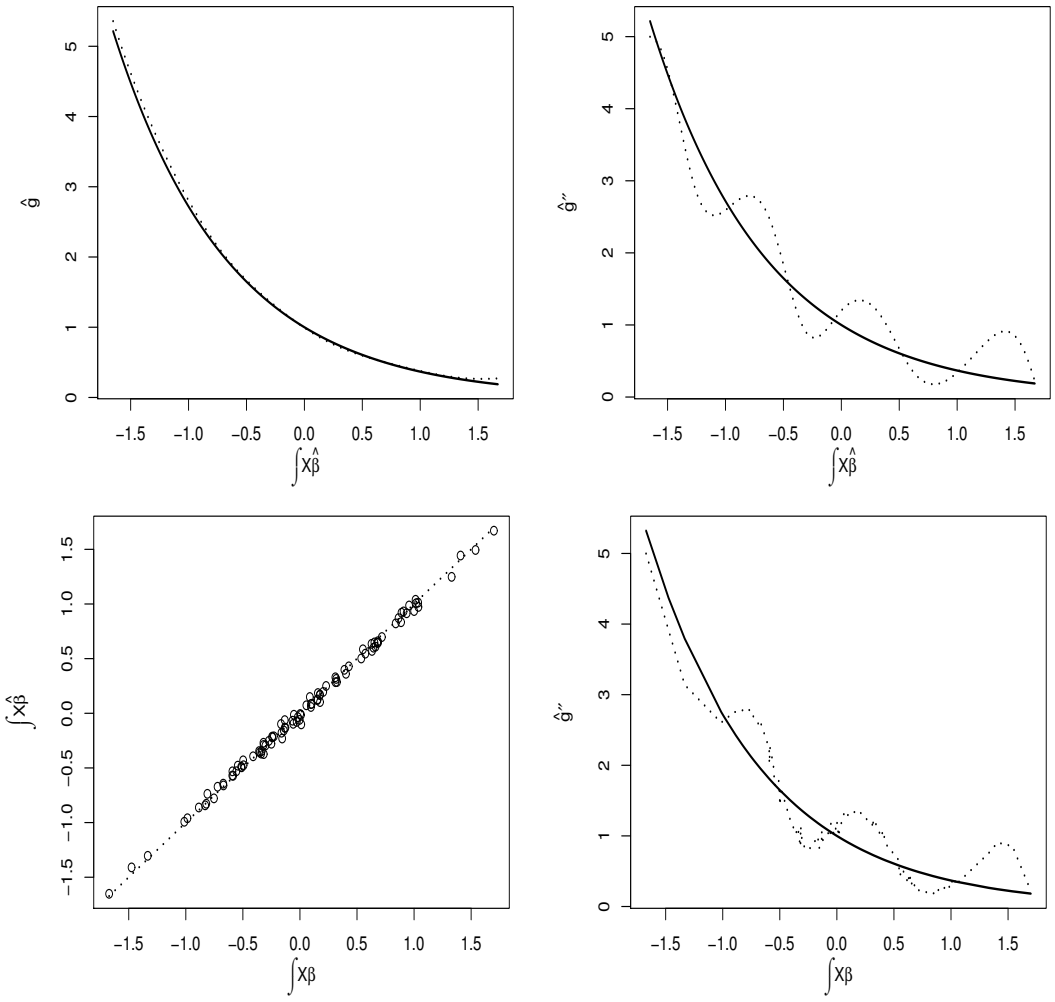

Figure 2.1: Example results using $g(s) = e^{-s}$. The top-left and right panel are the plots of g and g'' over 1000 equally-spaced grid points, while the lower and upper bound are the minimum and maximum of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$. The bottom-right panel gives $g''\left(\int X(t)\hat{\beta}(t)\,\mathrm{d}t\right)$ plotted against the true $\int X(t)\beta(t)\,\mathrm{d}t$, providing a visual representation of the error we control in Theorem 6. In all plots, the true curve is given by the black line, while the dotted line is the estimated curve. The bottom-left panel is the plot of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$ versus $\int X(t)\beta(t)\,\mathrm{d}t$, while the dotted line is $y = x$.

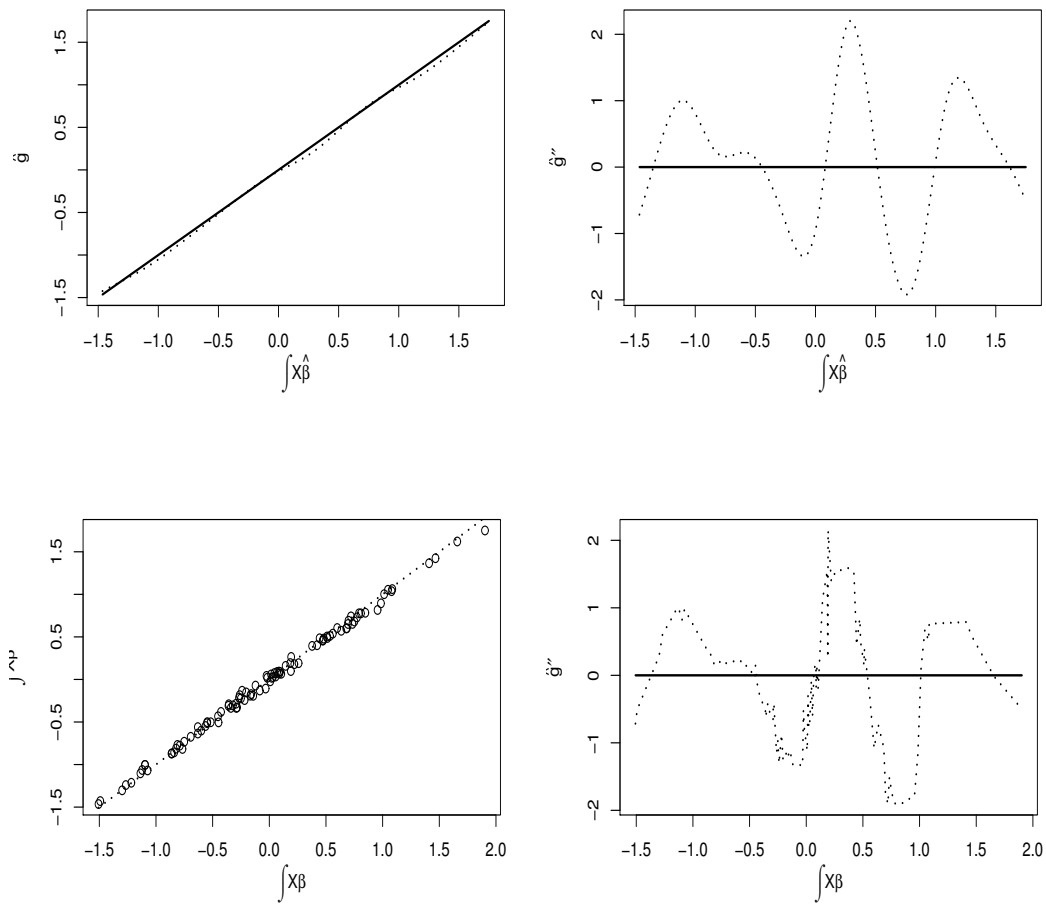

Figure 2.2: Example results using $g(s) = s$. The top-left and right panel are the plots of g and g'' over 1000 equally-spaced grid points, while the lower and upper bound are the minimum and maximum of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$. The bottom-right panel gives $g''\left(\int X(t)\hat{\beta}(t)\,\mathrm{d}t\right)$ plotted against the true $\int X(t)\beta(t)\,\mathrm{d}t$, providing a visual representation of the error we control in Theorem 6. In all plots, the true curve is given by the black line, while the dotted line is the estimated curve. The bottom-left panel is the plot of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$ versus $\int X(t)\beta(t)\,\mathrm{d}t$, while the dotted line is $y = x$.

was mapped in each of 26 $1m^2$ ungrazed quadrats allowing plant growth (the increase in ground area covered by the plant canopy) to be measured for 22 year-to-year transitions. In total, this yielded 1003 total data points with the following variables:

1. `logarea.t1`, `logarea.t0` : the plant's logarithm of area at time t_0 and t_1, where t_0 is the observation start time and t_1 is the end time. A relatively large difference indicates a high relative growth rate of the plant at that time.
2. `W`: a measure of plant competition. Taken to be a scalar covariate.
3. `p.00` – `p.36`: monthly total precipitation over the previous 36 months observed in Dubois, ID, USA, denoted as $p(s)$.
4. `t.00` – `t.36`: averaged temperature (within a month) over the previous 36 months observed in Dubois, ID, USA, denoted as $t(s)$.

See Teller et al. (2016) for further details. The precipitation and temperature histories are modeled as two covariate functions. Assume that the response variable is $logarea.t1 - logarea.t0$, a Functional Single Index model is:

$$logarea.t1 - logarea.t0 = g\left(\alpha \cdot W + \int p\beta_1 + \int t\beta_2\right), \tag{2.56}$$

where the coefficient α, the functions g, β_1 and β_2 need to be estimated.

2.6.2 Results

We used 1000 randomly chosen initial starting vectors, as well as the equal and linear values described above. Among these, one of the random starting points was chosen as having the smallest GCV after optimization. The plot of the estimated g, g'', and the coefficient functions β_1 and β_2 is in Figure 2.3. Since the estimated g'' is always

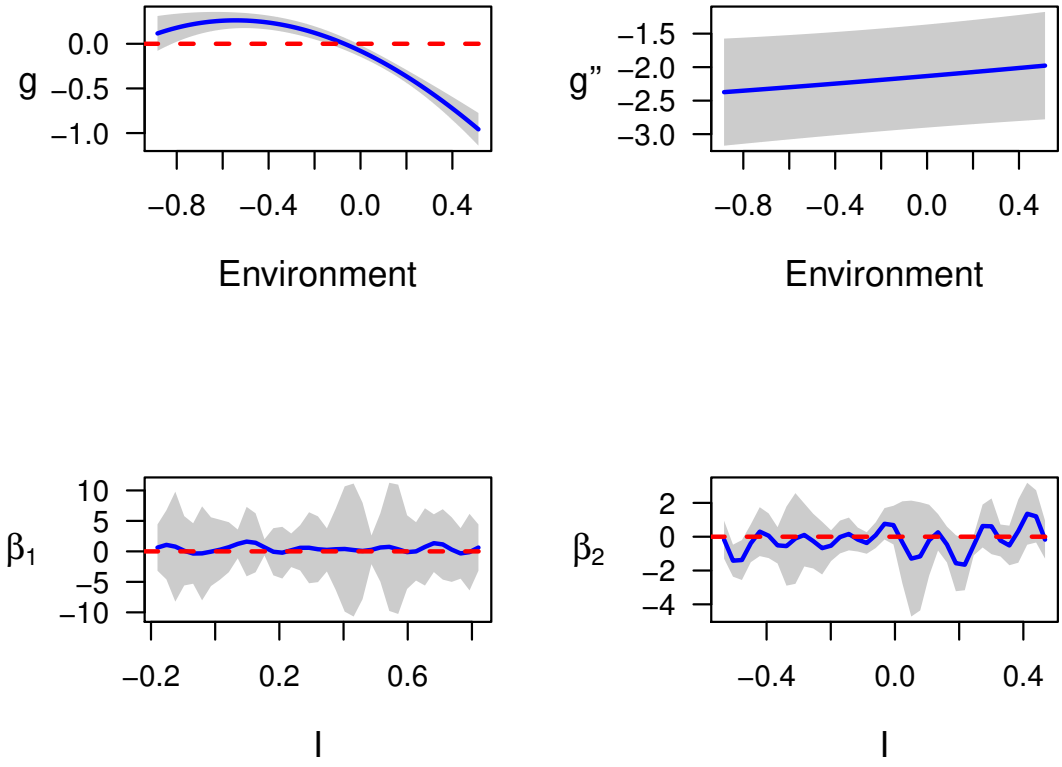

Figure 2.3: The top-left panel is the estimate of g, while the top-right panel is the estimate of g''. The bottom panel are the plots of the estimated β_1 and β_2.

negative, the link function g is concave. We could conclude that the species will grow better with a constant environment.

2.7 Conclusion

To answer the ecological question, we need to figure out the convexity or concavity of the link function g, or equivalently, find whether the second derivative is positive

or negative. In this paper, we used the local quadratic method to approximate the link function g, and estimated the curvature of g and the coefficient function β by a nested optimization procedure. Under some assumptions, we showed that the coefficient function β could be estimated root-n consistently. In addition, the rate of convergence of the curvature g'' is $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\hat{g}''\left(\int X_i\hat{\beta}\right) - g''\left(\int X_i\beta\right)\right]^2 = O\left(h_n^4 + \frac{1}{nh_n^4}\right)$.

In the simulation study, we used three different link functions, convex, concave and neither convex nor concave. While we derive convergence rates for the curvature of g, our simulation results demonstrate the numerical challenges that accompany Functional Single Index models. We can estimate g fairly well, but our estimates of g'' are sensitive to the choice of initial condition, requiring considerable care in optimization. We expect that these numerical challenges are specific to the estimators employed.

In our case study, the conclusions from Figure 2.3 suggest that strong concavity in the response of *Artemesia tripartita* indicating that temperature and rainfall variability has a negative effect that was not captured in the models used in Adler et al. (2010). We conclude that average relative growth will be greater in a constant environment; although we note that this conclusion may not apply to absolute growth.

CHAPTER 3

JENSEN EFFECT IN FUNCTIONAL SINGLE INDEX MODEL

3.1 Introduction

3.1.1 Ecology Questions

In natural ecosystems, environmental conditions are highly variable over time and space (e.g., Vasseur and McCann, 2007) and many classical questions in ecology and evolution are therefore concerned with the potential consequences of this variation. Two important topics have been how species' traits and life histories evolve so that species can persist in environments that may be favorable at some times and unfavorable at others (see Cohen (1966) and Koons et al. (2008)). Nonlinear responses to environmental variability can contribute to maintaining the biodiversity of competing species, allowing them to coexist stably (see for example Hutchinson (1961); Chesson and Warner (1981); Ellner (1987); Chesson (1994, 2000b,a)). Nonlinear responses to environmental conditions are also important for forecasting responses to climate change, as environmental variability can either increase population growth rate (Drake (2005); Koons et al. (2009)) or decrease it (Lewontin and Cohen (1969)), depending on the shape of the norm of reaction between environmental variables and the components of population growth rate (survival, growth, and reproduction).

The goal of this paper is to develop methods for determining whether environmental variability is beneficial or harmful for some component of population growth rate, which will depend on whether the response curve is concave-up or concave-down over the range of environmental variability. Common statistical models for species growth and

survival make parametric assumptions involving specific forms of nonlinearity, which can pre-determine the effect of environmental variability in the model. For example, in a logistic regression model for survival where the logit(survival) is a linear function of covariates, small variance in one covariate about its mean will increase average survivorship if survival at the mean covariate is below 0.5, and decrease average survivorship if survival at the mean covariate is above 0.5. The conclusion is then driven more by statistical convention than by the data. Recent statistical research in semi-parametric or nonparametric methods inspire us to understand the effect of environmental variation via nonparametric models that do not impose these assumptions. Specifically, we consider spline-based methods (see Wood (2000), Ramsay (2006), Ramsay et al. (2009) and Ruppert et al. (2003)) to predict nonlinear responses under environmental fluctuation.

The nonparametric model considered here is

$$\boldsymbol{G} = g(\boldsymbol{E}) + \epsilon, \tag{3.1}$$

where $\boldsymbol{G}$ and $\boldsymbol{E}$ are the growth (or future size) and environment of a plant. The function g is the link function to be estimated, and ϵ is random error. We assume that the environmental $\boldsymbol{E}$ is described by a climate history, such as temperature or precipitation observed at a fine time-scale, such as daily resolution. Following Teller et al. (2016), these are thought of as functional covariates, leading to representation of $\boldsymbol{E}$ as a functional linear term

$$\boldsymbol{E} = \int X(t)\beta(t)\,\mathrm{d}t, \tag{3.2}$$

where $\beta(t)$ is the coefficient function to be estimated, and $X(t)$ is the observed climate history.

To assess the impact of variability in $\boldsymbol{E}$ on the growth rate $\boldsymbol{G}$, we need to compare $g[\mathrm{E}(\boldsymbol{E})]$ (growth under constant conditions) and $\mathrm{E}[g(\boldsymbol{E})]$. By the Jensen's inequality,

we have that $g\left[\mathrm{E}\left(\boldsymbol{E}\right)\right] \leq \mathrm{E}\left[g\left(\boldsymbol{E}\right)\right]$ if g is a convex function, and a varying environment will accelerate the growth. Otherwise, if the link function g is concave, a constant environment at the average of environmental factor $\boldsymbol{E}$ would be more beneficial.

3.1.2 Model Formulation

Combining (3.1) and (4.4), a functional model for observed species growth is

$$Y = g\left(\int X\left(t\right)\beta\left(t\right)\mathrm{d}t\right) + \epsilon. \tag{3.3}$$

This is the Functional Single Index model introduced in Chen et al. (2011) and Ma (2016). The functional single index model is an extension of the single index model to a functional covariate via $\int X\left(t\right)\beta\left(t\right)\mathrm{d}t$. This model allows a nonlinear relationship between response Y and covariate function $X\left(t\right)$. In addition, it improves stability in estimating link function g through imposing smoothness on $\beta\left(t\right)$.

To assess the impact of environmental variability, we need to test the convexity or concavity of the link function g. A first approach would be to estimate the curvature of link function g over a domain. We illustrate our estimation procedure in Section 2, where we show that estimating second derivatives requires unrealistic sample sizes.

However, we observed that estimate of the link function g obtained by our procedure is quite accurate. So instead of evaluating the curvature of g, we examine the "Jensen Effect" directly, and estimate the quantity $g\left[\mathrm{E}\left(\boldsymbol{E}\right)\right] - \mathrm{E}\left[g\left(\boldsymbol{E}\right)\right]$, or the sign of Jensen's inequality, which only involves an estimate of g. Inspired by the SiZer method Chaudhuri and Marron (1999), we derive a test based on the maximum estimated value over a range of smoothing parameters. We show that our estimate and hypothesis test work well on both simulated and real data set.

3.1.3 Related Literature

There is a large literature on the single index model examining both applied methodology and theoretical properties. The link function g and the coefficient vector $\boldsymbol{\beta}$ have been estimated by three different methods. (1) The most widely used is the Projection Pursuit Regression (PPR) approach introduced in Hardle et al. (1993). This method is a nested estimation procedure, with the link function g estimated by local polynomial approximation and the coefficient function by minimizing the MSE. Theoretical properties were studied in Hardle et al. (1993) and Ichimura (1993). (2) The Average Derivative approach was introduced in Hristache et al. (2001). (3) Li (1991) introduced the Sliced Inverse Regression method, which considered the estimation of the coefficient vector as a dimension-reduction problem.

In contrast, there are few studies of the functional single index model. A counterpart to the Projection Pursuit Regression was introduced in Chen et al. (2011), where the coefficient function β was approximated by a spline basis and the coefficient vector is estimated. In addition, a convergence rate was found for this method. Ma (2016) used two spline bases to approximate the coefficient function and the link function, respectively, and derived some asymptotic properties of the resulting estimate.

The SiZer (SIgnificant ZERo crossings of derivative) method that we adopt to assess significance, introduced in Chaudhuri and Marron (1999), was designed to bypass smoothing parameter selection by comparing the estimates of a curve over a range of smoothing parameters. Our test statistic is inspired by the SiZer method. Instead of trying to select an optimal smoothing parameter for estimation and inference, we examine estimates over a range of smoothing parameters and do inference based on maximizing a test statistic over that range.

3.2 Estimating Curvature

This section provides a short demonstration of the difficulties of estimating curvature nonparametrically in a functional single index model, which would be necessary for directly using Jensen's inequality to examine the ecological consequences of environmental variability. We first define an estimate for g and discuss some of the numerical challenges it involves, and then provide some results that attempt to estimate g'' directly when selecting smoothing parameters by GCV.

3.2.1 Estimation Procedure

We consider estimating β, g and g'' via a smoothed basis expansion. Assume that n independent and identically distributed data pairs $(X_1(t), Y_1), \cdots, (X_n(t), Y_n)$ are observed where $X_j(t)$ is a real-valued function on $[0, 1]$. The Functional Single Index model is

$$Y = g\left(\int X(t)\beta(t)\,\mathrm{d}t\right) + \epsilon, \tag{3.4}$$

where the coefficient function $\beta(t)$ is, like $X(t)$, defined on the interval $[0, 1]$. ϵ is assumed to be a Gaussian random error.

This integral may need to be evaluated numerically, depending on the representations used for X_i and β, and we assume that this is done up to ignorable error throughout the calculations below. To ensure identifiability of the model, we require that $\int \beta^2(t)\,\mathrm{d}t = 1$.

We use a K_1-dimensional B-spline basis for the link function g. For any s in the range of possible $\int X\beta$ values, the link function g can be written as

$$g(s) = \boldsymbol{\phi}^\top(s)\,\boldsymbol{d}, \tag{3.5}$$

where $\boldsymbol{d}$ is a K_1-dimensional column coefficient vector.

We use a K_2-dimensional Fourier basis for the coefficient function β, such that

$$\beta(t) = \boldsymbol{\psi}^\top(t)\,\boldsymbol{c}, \tag{3.6}$$

where $\boldsymbol{c}$ is a K_2-dimensional column coefficient vector, and $t \in [0, 1]$. The constraint $\|\beta\| = 1$ is reduced to requiring $\|\boldsymbol{c}\| = 1$ in this case.

The coefficient vectors $\boldsymbol{c}$ and $\boldsymbol{d}$ are estimated by minimizing a penalized sum of squares

$$\begin{aligned} \text{2PLS} &\doteq \sum_{i=1}^{n} (Y_i - g_i)^2 + \lambda_g \int \left(g^{(2)}(s)\right)^2 \mathrm{d}s + \lambda_\beta \int \left(\beta^{(2)}(t)\right)^2 \mathrm{d}t \\ &= \sum_{i=1}^{n} \left\{ Y_i - \boldsymbol{\phi}^\top \left[\left(\int X_i \boldsymbol{\psi}^\top \right) \boldsymbol{c} \right] \boldsymbol{d} \right\}^2 + \lambda_g \boldsymbol{d}^\top \mathbb{P}_g \boldsymbol{d} + \lambda_\beta \boldsymbol{c}^\top \mathbb{P}_\beta \boldsymbol{c}, \end{aligned} \tag{3.7}$$

where $g_i \doteq g\left(\int X_i \beta\right)$ and the penalty matrices $\left[\mathbb{P}_\beta\right]_{ij} = \int \psi_i^{(2)}(t)\,\psi_j^{(2)}(t)\,\mathrm{d}t$ and $\left[\mathbb{P}_g\right]_{ij} = \int \phi_i^{(2)}(t)\,\phi_j^{(2)}(t)\,\mathrm{d}t$ are available analytically for most common choices of basis expansion.

Equation (3.7) specifies a nonlinear optimization problem, which we solve numerically using built-in optimizers in R (see below). Denoting the estimated coefficients as $\hat{\boldsymbol{c}}$ and $\hat{\boldsymbol{d}}$, the estimates are

$$\hat{\beta}(t) = \boldsymbol{\psi}^\top(t)\,\hat{\boldsymbol{c}}, \tag{3.8}$$

and

$$\hat{g}_i \doteq \hat{g}\left(\int X_i \hat{\beta}\right) = \boldsymbol{\phi}^\top \left[\left(\int X_i(t)\,\boldsymbol{\psi}^\top(t)\,\mathrm{d}t \right) \hat{\boldsymbol{c}} \right] \hat{\boldsymbol{d}}, \tag{3.9}$$

where $i = 1, \cdots, n$.

3.2.2 Bases, Optimization and Cross-Validation

Our objective criterion (3.7) requires nonlinear numerical optimization. In our experiments below we have used the R function `optim` with some additional modifications.

The simulations reported in Section 3.4 used the BFGS gradient-based optimizer. However, at large values of λ we find that very tight convergence criteria are needed to reduce the numerical error below that of the estimated noise. This was mitigated with two strategies:

1. We initialize our optimization with $\boldsymbol{d}$ chosen so that $\hat{g}$ is exactly linear and $\boldsymbol{c}$ is obtained from functional linear regression.
2. We re-initialize BFGS once it converges, and run it a second time. BFGS uses a sequentially-calculated approximate Hessian, and can stop early due to poor estimation of this Hessian. Re-initialization resets the approximate Hessian to the identity, so that optimization is restarted with a steepest descent step.

To maintain identifiability of our model, we normalize our estimate of $\beta(t)$ within each evaluation of the objective function and multiply by $\text{sign}(\beta(0))$.

In order to represent $\hat{g}$ with a basis expansion, we need to control the range of its arguments. Throughout our estimates below, we have used the identifiability requirement that $\|\beta\| = 1$ to use a range of $[-S, S]$ where S is the largest score for the maximum eigenvalue from a principal components decomposition of the X_i. If $\left|\int X_i(t)\beta(t)\mathrm{d}t\right| > S$ we replace the argument with the corresponding end-point of the range and add a penalty of $\left|\int X_i(t)\beta(t)\,\mathrm{d}t\right| - S$ to the objective (3.7). In practice, while we find that this exceedance can occur during optimization, it never appears in the final result.

To illustrate the impact of initial conditions on our estimates, as well as the challenge of estimating curvature in functional single index models, Section 3.2.3 below presents examples of estimating g'' with initial conditions either given by the true basis coefficients $\boldsymbol{c}$ and $\boldsymbol{d}$ or by starting from $\boldsymbol{c} = \left(\frac{1}{\sqrt{K_2}}, \cdots, \frac{1}{\sqrt{K_2}}\right)$, using Nelder-Mead optimization.

In each case, different initialization procedures yield quite different performance statistics, but all perform poorly in estimating curvature.

While our assessment of statistical significance avoids selecting λ, it will be useful to have a value for visualization and for an estimate of residual variance. To choose λ, we define a smoother matrix $\mathbb{S}_\lambda$ associated with λ, and the GCV value for selecting λ is calculated from

$$\mathrm{GCV}(\lambda) \doteq \frac{\frac{1}{n}\left\|(\mathbb{I}-\mathbb{S}_\lambda)\boldsymbol{Y}\right\|^2}{\left[\frac{1}{n}\mathrm{tr}(\mathbb{I}-\mathbb{S}_\lambda)\right]^2}, \tag{3.10}$$

where $\mathbb{I}$ is the $(n \times n)$-dimensional identity matrix. We derive $\mathbb{S}_\lambda$ from a Taylor expansion in (3.29) below.

3.2.3 A Simulated Demonstration

We present here a brief simulation study to observe the accuracy of curvature estimates. The covariate function $X(t)$ was generated based on a Fourier basis

$$X_i(t) = \mu(t) + \sum_{k=1}^{4} \xi_{ik}\eta_k(t), \qquad i = 1, \cdots, n, \tag{3.11}$$

where $\mu(t) = t$, $\eta_1(t) = \frac{1}{\sqrt{2}}\sin(2\pi t)$, $\eta_2(t) = \frac{1}{\sqrt{2}}\cos(2\pi t)$, $\eta_3(t) = \frac{1}{\sqrt{2}}\sin(4\pi t)$, $\eta_4(t) = \frac{1}{\sqrt{2}}\cos(4\pi t)$, and ξ_{ik} are i.i.d $\mathrm{N}(0, \gamma_k)$ with $\gamma_1 = 1$, $\gamma_2 = \frac{1}{2}$, $\gamma_3 = \frac{1}{4}$, $\gamma_4 = \frac{1}{8}$. The coefficient function is

$$\beta(t) = 2\left[\frac{1}{\sqrt{12}}\eta_1(t) + \frac{1}{\sqrt{12}}\eta_2(t) + \frac{1}{\sqrt{6}}\eta_3(t) + \frac{1}{\sqrt{6}}\eta_4(t)\right]. \tag{3.12}$$

We observe that the coefficients for β satisfy $\|\boldsymbol{c}\| = 1$, under an orthonormal basis. The random errors ϵ_i are simulated as i.i.d. Gaussian noise with mean 0 and $\mathrm{var}(\epsilon) = 0.1\mathrm{var}\left[g\left(\int X\beta\right)\right]$.

We selected the sample size as $n = 100$ and examined three link functions:

1. $g(s) = e^{-s}$.

2. $g(s) = -s^2$.

3. $g(s) = s$.

To measure the performance of our estimators we define the MSE of the estimated β and $g^{(k)}$ to be

$$\text{RSE} = \left[\int \left(\hat{\beta}(t) - \beta(t)\right)^2 \mathrm{d}t\right]^{\frac{1}{2}}, \tag{3.13}$$

and

$$\text{RASE}(k) = \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{Y}_i^{(k)} - g^{(k)}\left(\int X_i(t)\beta(t)\,\mathrm{d}t\right)\right]^2\right\}^{\frac{1}{2}}, \tag{3.14}$$

where $\hat{Y}_i^{(k)} = \hat{g}^{(k)}\left(\int X_i(t)\hat{\beta}(t)\,\mathrm{d}t\right)$ for $k = 0, 1, \ldots$.

Of particular concern in the results (Table 3.1) is the substantial discrepancy between estimates from different initial conditions. Ye and Hooker (2018) similarly observed that second derivative estimates were highly sensitive to the effort placed into optimization.

	g1			**g2**			**g3**		
Initial	**RSE**	**RASE**	**RASE2**	**RSE**	**RASE**	**RASE2**	**RSE**	**RASE**	**RASE2**
True	1.1213	0.0921	5.2517	0.5385	0.0490	2.9079	0.7608	0.0706	5.4393
Equal	0.6417	0.0800	3.0516	0.6980	0.0730	4.1328	0.7024	0.0764	1.2170

Table 3.1: Simulation results with $\left(\lambda_g, \lambda_\beta\right)$ selected by GCV. Values in the Table are averages over 100 simulations.

The plots in Figure 3.1 provide an example of our results. The estimate of the link function nearly overlaps the true curve, indicating that our estimate of the link function is quite accurate. However, for the second derivative, the estimate deviates from the true curve, in fact becoming negative towards the right-hand limit. This reduced accuracy

is also evident in the results in Table 3.1. These plots indicate that our estimate of the curvature is not good enough to use as a basic for decisions on the convexity of g. In addition, the performance of the estimators varies a lot from different initial values. In Figure 3.2 we see that different initial conditions can lead to either over- or under-fitting g". Further examples are provided in Appendix B.1. We therefore directly assess the Jensen Effect via other methods, introduced in the next section.

3.3 Jensen Effect

The ecological interest in g'' is in the comparison of $g\left[\mathrm{E}\left(\boldsymbol{E}\right)\right]$ and $\mathrm{E}\left[g\left(\boldsymbol{E}\right)\right]$. Because reliable estimate of g'' requires unrealistic sample sizes, we instead compare these quantities directly to estimate what we call the "Jensen Effect".

We define a difference statistic by

$$\delta = \frac{1}{n}\sum_{i=1}^{n} g\left(\int X_i\beta\right) - g\left(\int \bar{X}\beta\right), \tag{3.15}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. If the link function g is convex, then $\delta > 0$ which indicates better growth with a varying environment; otherwise, the difference $\delta < 0$ and a constant environment is better for growth. However, this estimate still depends on the smoothing parameters λ_g and λ_β.

To account for the choice of smoothing parameters, we extend the SiZer method, first introduced for local linear regression in Chaudhuri and Marron (1999), and then extended to smoothing splines setting in Marron and Zhang (2005). SiZer is a graphical tool to make inference about features on an estimated curve, and detect features of that curve in an innovative way. Traditional nonparametric methods have difficulties in selecting a appropriate bandwidth in a local linear approximation, or a smoothing

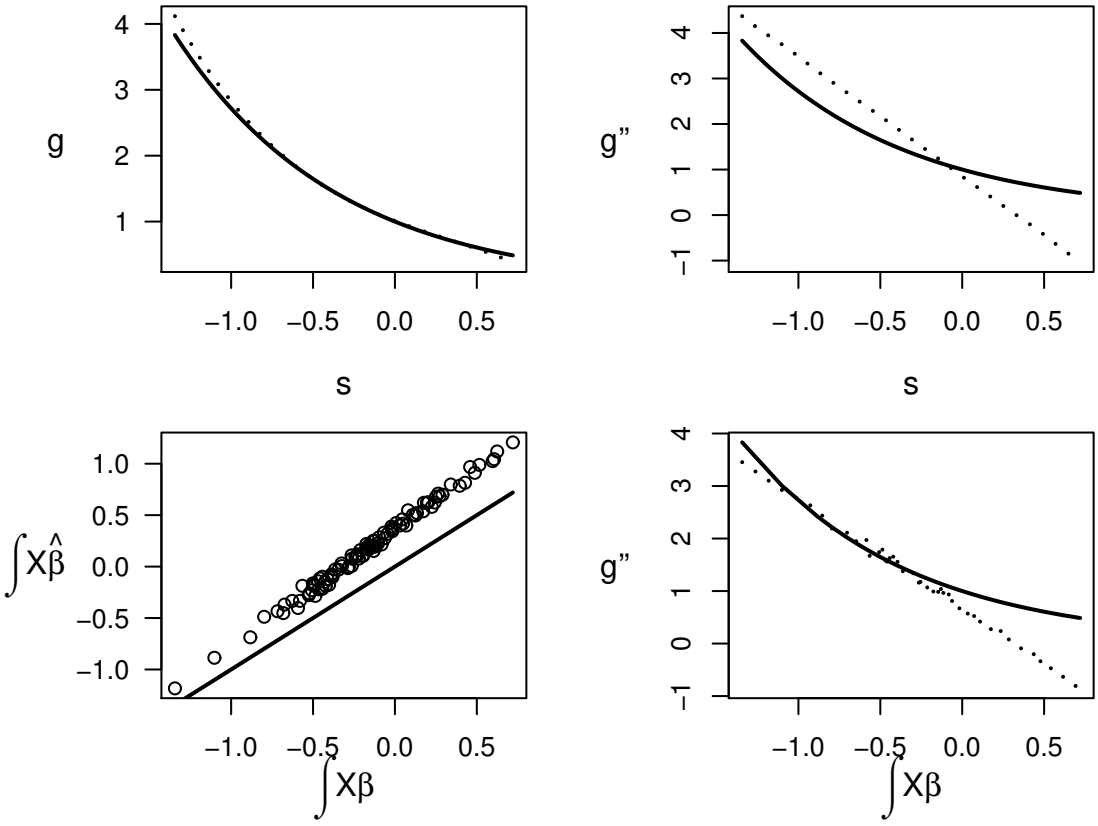

Figure 3.1: Example estimate for the link function $g(s) = e^{-s}$. Top-left and right panels plot g and g'' over 1000 equally-spaced grid points between the minimum and maximum of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$. Dotted curves are estimated values, and the solid curves are the truth. The bottom-left panel plot is $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$ versus $\int X(t)\beta(t)\,\mathrm{d}t$; the solid line is the 1:1 line. The bottom-right panel presents g'' (solid) and $\hat{g}''$ (dotted) evaluated at $\int X(t)\beta(t)\,\mathrm{d}t$ and $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$ respectively but plotted against the true argument.

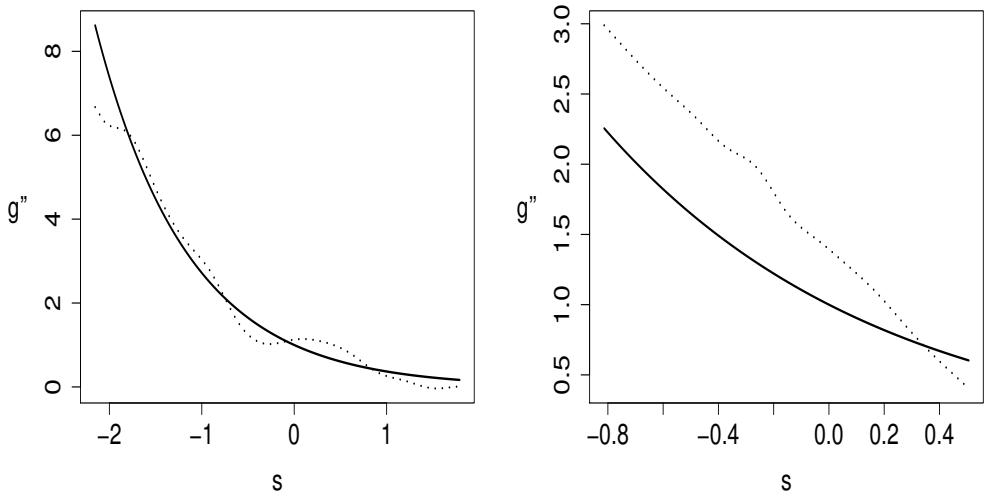

Figure 3.2: Estimates for g'' from different initial conditions. Left, using the known true g as the initial condition. Right, starting from equal values of the coefficients. Note that this example was chosen for illustrative purposes and does not use the same data as Figure 3.1.

parameter in a smoothing spline. Classical methods such that 10-fold cross-validation and generalized cross-validation provide point estimates, but it is difficult to incorporate their uncertainty into inference about the resulting curves. The SiZer method skips the process of selecting a tuning parameter. Instead, it examines the estimate of a curve over a range of tuning parameters, and observes features of the curve for each combination of observation point and tuning parameter. Inspired by the SiZer method, we look the difference δ over a range of λ values for g and β, and generate hypothesis tests using the maximum or minimum value of δ as a function of λ.

3.3.1 Hypothesis Test 1: Nonparametric Smoothing

We begin by briefly developing our SiZer-inspired test for a standard smoothing spline (treating the environment $\boldsymbol{E}$ as known) before developing the test for a functional single index model. Here defining Φ to be matrix of evaluations, $\Phi_{ij} = \phi_j(E_i)$ and $\mathbb{P}$ to be the

second derivative penalty matrix, the standard smoothing spline estimate is

$$\hat{g}_\lambda(e) = \phi(e)^\top \left(\Phi^\top \Phi + \lambda \mathbb{P}\right)^{-1} \Phi^\top \mathbf{Y}. \tag{3.16}$$

Define the $(n+1)$-dimensional column vector $\boldsymbol{a} \doteq \left(\frac{1}{n}, \cdots, \frac{1}{n}, -1\right)^\top$ and the augmented set of evaluation points $\mathbf{e} = \left(E_1, \ldots, E_n, \bar{E}\right)^\top$, where $E_i = \int X_i \hat{\beta}$ (at each observed environment value) and $\bar{E} = \frac{1}{n} \sum_{i=1}^{n} \int X_i \hat{\beta}$ (averaged across all environment values), with corresponding evaluation matrix Φ^+. We can write

$$\delta_\lambda = \frac{1}{n} \sum \hat{g}_\lambda(E_i) - \hat{g}_\lambda(\bar{E}) = \boldsymbol{a}^\top \Phi^+ \left(\Phi^\top \Phi + \lambda \mathbb{P}\right)^{-1} \Phi^\top \mathbf{Y} = \boldsymbol{u}_\lambda \boldsymbol{Y}, \tag{3.17}$$

which we can standardize to obtain the t-statistic

$$t_\lambda = \frac{\boldsymbol{u}_\lambda \boldsymbol{Y}}{\sigma \sqrt{\boldsymbol{u}_\lambda \boldsymbol{u}_\lambda^\top}}. \tag{3.18}$$

Since the response variable Y is a Gaussian process, the test statistics t_λ is also a Gaussian process with the covariance function

$$\Sigma(\lambda_1, \lambda_2) = \frac{\boldsymbol{u}_{\lambda_1} \boldsymbol{u}_{\lambda_2}}{\left\|\boldsymbol{u}_{\lambda_1}\right\| \left\|\boldsymbol{u}_{\lambda_2}\right\|} \tag{3.19}$$

which involves no unknown parameters. We can thus use $\max_\lambda |t_\lambda|$ as a test statistic, obtaining critical values by simulating from the Gaussian process $\mathbb{GP}(\mathbf{0}, \Sigma)$. Under the null hypothesis $\delta = 0$, t_λ is a Gaussian process with mean $\mathbf{0}$. In practice, we need to estimate $\hat{\sigma}$, the estimate of the standard deviation of the random error ϵ, which we do based on the value λ selected by GCV.

An important consideration here is that we expect δ_λ to inherit smoothing bias, but this should result in under-estimation of the Jensen Effect because it will shrink the estimated second derivative. By examining t_λ over the whole range of λ we can assess this effect at various levels of smoothing while maintaining a conservative test. We do still need to choose λ by GCV in our estimate $\hat{\sigma}^2$, because we use the same $\hat{\sigma}^2$

when calculating the covariance matrix Σ. We expect $\hat{\sigma}^2$ to be relatively insensitive to the specific λ chosen; maintaining a constant $\hat{\sigma}^2$ in the t-statistic removes the need to account for changes in $\hat{\sigma}^2$ across λ.

3.3.2 Hypothesis Test 2: Functional Single Index Model

The functional single index model complicates the process described above by including two smoothing parameters and nonlinear effects of $\hat{\beta}$, necessitating a Taylor expansion to approximate the recipe above. For each pair of smoothing parameters $\left(\lambda_g, \lambda_\beta\right)$, we obtain an estimate of β, g, and δ, denoted as $\hat{\delta}\left(\lambda_g, \lambda_\beta\right)$.

Defining

$$\boldsymbol{i} = \left(\int X_1\hat{\beta}, \cdots, \int X_n\hat{\beta}, \int \bar{X}\hat{\beta}\right)^\top, \tag{3.20}$$

$$\boldsymbol{i}_{-1} = \left(\int X_1\hat{\beta}, \cdots, \int X_n\hat{\beta}\right)^\top, \tag{3.21}$$

$$\boldsymbol{v} = \left(\hat{g}\left(\int X_1\hat{\beta}\right), \cdots, \hat{g}\left(\int X_n\hat{\beta}\right), \hat{g}\left(\int \bar{X}\hat{\beta}\right)\right)^\top, \tag{3.22}$$

the estimated difference function given $\left(\lambda_g, \lambda_\beta\right)$ is

$$\begin{aligned}\hat{\delta}\left(\lambda_g, \lambda_\beta\right) &= \boldsymbol{a}^\top \boldsymbol{v} = \boldsymbol{a}^\top \boldsymbol{\phi}\left(\boldsymbol{i}\right)\hat{\boldsymbol{d}} \\ &= \boldsymbol{a}^\top \boldsymbol{\phi}\left(\boldsymbol{i}\right)\left(\boldsymbol{\phi}\left(\boldsymbol{i}_{-1}\right)^\top \boldsymbol{\phi}\left(\boldsymbol{i}_{-1}\right) + \lambda_g \mathbb{P}_g\right)^{-1} \boldsymbol{\phi}\left(\boldsymbol{i}_{-1}\right)^\top \boldsymbol{Y}.\end{aligned} \tag{3.23}$$

To construct a t-statistic to test the significance of δ, an estimate of the variance of the difference function δ is needed. The estimated difference function $\hat{\delta}\left(\lambda_g, \lambda_\beta\right)$ is defined on an estimate of $\hat{\boldsymbol{c}}$ and $\hat{\boldsymbol{d}}$, which are the coefficients of β and g respectively. Therefore, we need to calculate the covariance of the estimated $\hat{\boldsymbol{c}}$ and $\hat{\boldsymbol{d}}$.

Recall (3.7), the penalized least squares criterion to be minimized, and define the matrices of linear basis effects $\Psi_{ij} = \int X_i\left(t\right)\psi_j\left(t\right)\mathrm{d}t$ and evaluations of the link function

bases and derivatives $\Phi_{ij}^{(k)} = \phi_j^{(k)}(\Psi_{i\cdot}\boldsymbol{c})$ with $\boldsymbol{c}$ taken at its expected estimate. We derive gradients of PLS as

$$\begin{pmatrix} \nabla_{\boldsymbol{d}} \\ \nabla_{\boldsymbol{c}} \end{pmatrix} = \begin{pmatrix} \Phi^\top \{\boldsymbol{Y} - \Phi\boldsymbol{d}\} + \lambda_g \mathbb{P}_g \boldsymbol{d} \\ \Psi^\top \mathrm{diag}\left\{\Phi^{(1)}\boldsymbol{d}\right\} \{\boldsymbol{Y} - \Phi^\top \boldsymbol{d}\} + \lambda_\beta \mathbb{P}_\beta \boldsymbol{c} \end{pmatrix} = \begin{pmatrix} \mathbb{Z}_g \\ \mathbb{Z}_\beta \end{pmatrix} (\boldsymbol{Y} - \Phi\boldsymbol{d}) + \begin{pmatrix} \lambda_g \mathbb{P}_g \boldsymbol{d} \\ \lambda_\beta \mathbb{P}_\beta \boldsymbol{c} \end{pmatrix} \tag{3.24}$$

and expected Hessian

$$\mathbb{H} = \begin{pmatrix} \mathbb{Z}_g^\top \mathbb{Z}_g + \lambda_g \mathbb{P}_g & \mathbb{Z}_g^\top \mathbb{Z}_\beta \\ \mathbb{Z}_\beta^\top \mathbb{Z}_g & \mathbb{Z}_\beta^\top \mathbb{Z}_\beta + \lambda_\beta \mathbb{P}_\beta \end{pmatrix}. \tag{3.25}$$

We can now obtain the sandwich covariance

$$\mathrm{cov}\begin{pmatrix} \boldsymbol{d} \\ \boldsymbol{c} \end{pmatrix} = \hat{\sigma}^2 \mathbb{H}^{-1} \begin{pmatrix} \mathbb{Z}_g \\ \mathbb{Z}_\beta \end{pmatrix}^\top \begin{pmatrix} \mathbb{Z}_g \\ \mathbb{Z}_\beta \end{pmatrix} \mathbb{H}^{-1} \tag{3.26}$$

where we estimate σ^2 from

$$\hat{\sigma}^2 = \frac{1}{\mathrm{df}_{\mathrm{res}}} \sum_{i=1}^{n} \left[Y_i - \hat{g}\left(\int X_i \hat{\beta} \right) \right]^2. \tag{3.27}$$

As as in Ruppert et al. (2003), the residual degree of freedom is defined as

$$\mathrm{df}_{\mathrm{res}} = n - 2\mathrm{tr}(\mathbb{S}) + \mathrm{tr}(\mathbb{S}\mathbb{S}^\top) - K_1, \tag{3.28}$$

where

$$\mathbb{S} \doteq \mathbb{S}\left(\lambda_g, \lambda_\beta\right) = \begin{pmatrix} \mathbb{Z}_g \\ \mathbb{Z}_\beta \end{pmatrix} \mathbb{H}\left(\mathbb{Z}_g^\top,\ \mathbb{Z}_\beta^\top\right) \tag{3.29}$$

is an approximate smoother matrix in which we use the values of $(\lambda_g, \lambda_\beta)$ selected by GCV.

We now define a t-statistic for δ as a function of λ,

$$t \doteq t\left(\lambda_g, \lambda_\beta\right) \doteq \frac{\hat{\delta}\left(\lambda_g, \lambda_\beta\right)}{\mathrm{sd}\left[\hat{\delta}\left(\lambda_g, \lambda_\beta\right)\right]}. \tag{3.30}$$

where $\mathrm{sd}\left[\hat{\delta}\left(\lambda_g, \lambda_\beta\right)\right]$ is given by

$$\mathrm{sd}\left[\hat{\delta}\left(\lambda_g, \lambda_\beta\right)\right] = \left\{\left[\boldsymbol{a}^\top \boldsymbol{\phi}(\boldsymbol{i})\right] \mathrm{cov}\left(\hat{\boldsymbol{d}}\right) \left[\boldsymbol{a}^\top \boldsymbol{\phi}(\boldsymbol{i})\right]^\top\right\}^{\frac{1}{2}}. \tag{3.31}$$

Defining the n-dimensional row vector $\boldsymbol{u}_\lambda$ as

$$\boldsymbol{u}_\lambda = \boldsymbol{a}^\top \boldsymbol{\phi}(\boldsymbol{i})\left(\boldsymbol{\phi}(\boldsymbol{i}_{-1})^\top \boldsymbol{\phi}(\boldsymbol{i}_{-1}) + \lambda \mathbb{P}\right)^{-1} \boldsymbol{\phi}(\boldsymbol{i}_{-1})^\top, \tag{3.32}$$

the estimated covariance matrix of $\hat{\delta}_\lambda$ is $\hat{\sigma}^2 \boldsymbol{u}_\lambda \boldsymbol{u}_\lambda^\top$. δ_λ is therefore approximately a Gaussian process indexed by λ and we can get the estimated variance of t_λ from $\frac{\boldsymbol{u}_\lambda \boldsymbol{u}_\lambda^\top}{\|\boldsymbol{u}_\lambda\|^2}$.

We want to test if $\delta \equiv 0$. Denote the number of the smoothing parameters λ_g as m, we test H_0: $\left(\hat{\delta}_{\lambda_1}, \cdots, \hat{\delta}_{\lambda_m}\right)^\top = \mathbf{0}_m$. Under H_0, $\left(t_{\lambda_1}, \cdots, t_{\lambda_m}\right)^\top \sim \mathrm{N}\left(\mathbf{0}_m, \mathbb{A}_{mm}\right)$, where $\mathbf{0}_m$ is a m-dimensional column vector, and the covariance matrix $\mathbb{A}$ is $(m \times m)$-dimensional with the (i, j) term equals to $\frac{\boldsymbol{u}_{\lambda_i} \boldsymbol{u}_{\lambda_j}^\top}{\|\boldsymbol{u}_{\lambda_i}\| \|\boldsymbol{u}_{\lambda_i}\|}$. The test statistic that we examine is $T = \max\left\{t_{\lambda_1}, \cdots, t_{\lambda_m}\right\}$.

In order to obtain a critical value for this statistic, we repeatedly simulate t_λ from $\mathrm{N}\left(\mathbf{0}_m, \mathbb{A}_{mm}\right)$ and obtain a distribution for $\max_\lambda |t_\lambda|$.

3.4 Simulation Study

In this section, we use simulated data to explore the power of our test for both single index and functional single index models.

3.4.1 Single Index Model

We test for a Jensen Effect by calculating the difference function δ over a range of smoothing parameters. If the link function g is convex, the δ function will be positive

for most of λ values, although it may have high variance at low λ and high bias at high λ. For each simulation, we conduct the hypothesis test introduced in previous section.

Our simulation study starts with the single index model with $p = 5$ covariates generated uniformly on $[-0.5, 0.5]$, and the coefficient $\boldsymbol{\beta} = \frac{1}{\sqrt{p}}\mathbf{1}_p$ so that $\|\boldsymbol{\beta}\| = 1$.

To illustrate the Jensen Effect, we choose three different link functions, (1) $g(s) = e^s$, (2) $g(s) = -s^2$, (3) $g(s) = s$. We represented g by a 25-dimensional quintic B-spline basis. For each link function, we simulated 1000 data sets of size 100, with error standard deviation 0.1. We obtained critical values for our test by simulating 5000 normal samples from the null distribution. Figure 3.3 presents a sample of δ_λ and t_λ functions functions versus $\log(\lambda)$ for $g(s) = e^s$; plots for the other link functions are in Appendix C.2. The rejection rates for these functions are: 99.2%, 99.3% and 5.7% respectively.

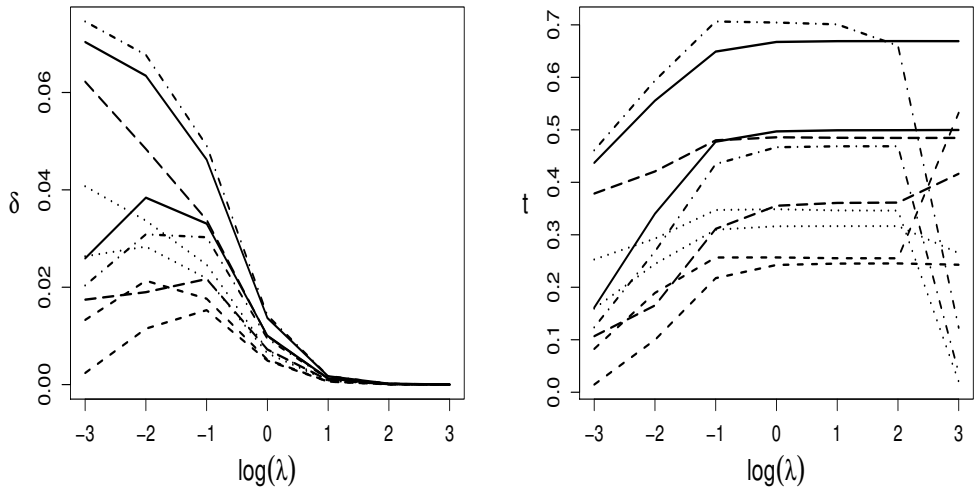

Figure 3.3: Left: a sample of δ_λ as a function of λ in a single index model with link function $g(s) = e^s$. Right: the corresponding t_λ functions.

3.4.2 Functional Single Index Model

To define a distribution for the functional covariates, we use a 25-dimensional Fourier basis $\boldsymbol{\psi}(t)$, where $t \in [0, 1]$. The covariate functions $X(t)$ are generated as

$$X(t) = \sum_{i=1}^{25} \xi_i \psi_i(t), \tag{3.33}$$

where $\xi_i \sim \mathrm{N}\left(0, e^{-(i-1)/12}\right)$. The coefficient function is

$$\beta(t) = \boldsymbol{c}^\top \boldsymbol{\psi}(t), \tag{3.34}$$

where $\boldsymbol{c} = (0, 1, 1, 0.5, 0, \cdots, 0)^\top$.

Again we used the three link functions $g(s) = e^2$, $g(s) = -s^2$, $g(s) = s$. We represented g by a 25-dimensional quintic B-spline basis. For each link function, we generated 1000 simulated data sets of size 100 with error standard deviation 0.1, and for each such data set we generated 5000 normal samples from the null distribution to obtain critical values.

A plot of the δ_λ and t_λ functions for $g(s) = e^s$ is presented in Figure 3.4. We have placed equivalent plots for $g(s) = -s^2$ and $g(s) = s$ in Appendix B.3. The rejection rates for the three link functions were 100%, 100% and 7.3%, showing very good power with a reasonable sample size and close to nominal rate when the null hypothesis is true (no curvature).

3.4.3 Power Analysis

To investigate the power of our test in more detail we consider a series of increasingly nonlinear link functions

$$g(s) = s + \eta e^{-s}, \tag{3.35}$$

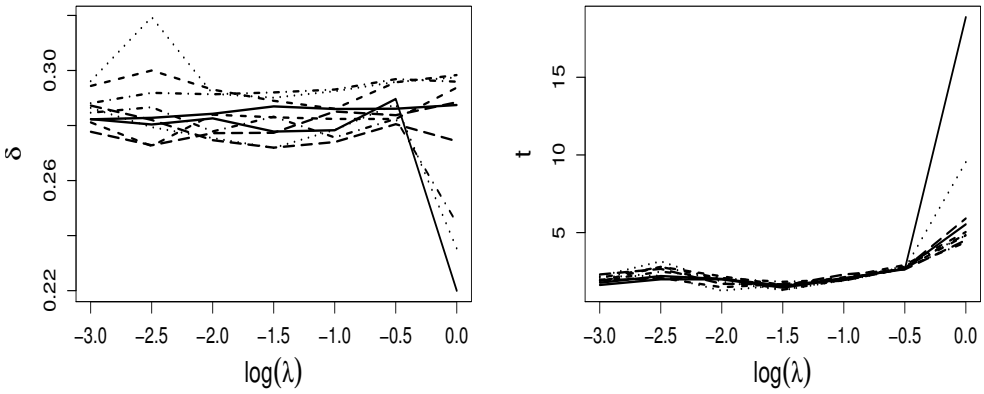

Figure 3.4: Left: a sample of δ_λ as a function of λ in a functional single index model with link function $g(s) = e^s$. Right: the corresponding t_λ functions.

with $0 \leq \eta \leq 1.2$ for the single index model and $0 \leq \eta \leq 0.8$ for the functional single index model. As η increases, g becomes strongly convex. For each η, we generate 1000 simulated data sets and again used 5000 normal samples under the null distribution to obtain critical values.

Figure 3.5 presents the rejection rate plotted against η. We observe a sharp increase as η increases, as expected. As the link function g becomes more and more convex, the rejection rate will converge to 1.

3.5 Application to real ecological data

To demonstrate application of our tests, we analyze the North Temperate Lakes LTER: Zooplankton - Trout Lake Area data set (https://portal.edirepository.org/nis/mapbrowse?scope=knb-lter-ntl&identifier=37&revision=29). An earlier version of these data were analyzed by Drake (2005) to examine the temperature-dependence of copepod populations. In our

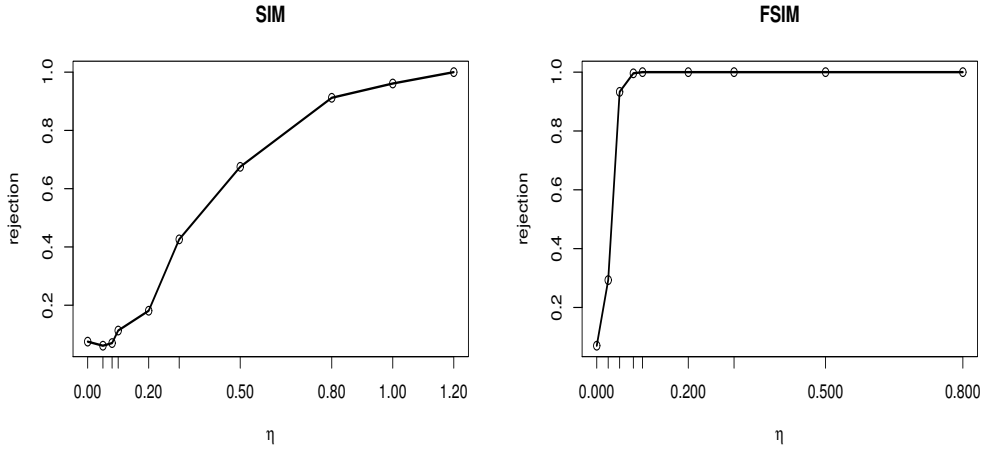

Figure 3.5: The power function of the Jensen Effect test plotted against η in the link function $g(s) = s + \eta e^{-s}$. A single index model using the simulation settings from Section 3.4.1 produced the left plot, while the right-hand plot is obtained using a functional single index model with settings described in Section 3.4.2.

data set, the density of the populations of nine species of copepods and rotifers, along with water temperature, were recorded from 1981 to 2015 in 8 different lakes. Our choice of species was determined based on the number and length of observations available and differs from those studied in Drake (2005). Note that the growth response in these data is not the growth (in size) of an individual, but the growth (in numbers) of a population, but our functional single index model is still appropriate for this setting.

The values recorded in the original data set are:

1. `d`: species' density at a specific time and lake.
2. `t`: record of temperature corresponding to the same time as `d`. The temperature was recorded on irregular time points among different years and lakes, so we preprocessed the temperature data by fitting a smoothing spline.

At the time of each observed response, we used lake temperature values over the 60

Species	Sample Size	p-value
Diacyclops thomasi	1482	0.170
Filinia terminalis	461	0.512
Gastropus stylifer	1083	0.140
Kellicottia longispina	1541	0.014*
Keratella cochlearis	1654	0.726
Keratella earlinae	1092	0.425
Keratella quadrata	689	0.059
Polyarthra remata	1553	0.003*
Polyarthra vulgaris	1771	0.338

Table 3.2: Sample size and p-value for each species in LTER dataset. $*$ indicates that the p-value is significant for the hypothesis test at level $\alpha = 0.05$.

preceding days as the climate history covariate $X(t)$. For each lake, we fit a penalized spline functional single index model for the growth in population density as a function of temperature history in each lake. We used 37 Fourier basis functions to represent $\beta(t)$ in each case and a 21-dimensional cubic B-spline basis to represent g, and searched over values of $\log(\lambda)$ from 10^{-6} to 10^{6}.

The p-values for 9 species are given in Table 3.2; plots of the estimates of g, g'' and δ functions are given in Appendix B.4. *Kellicottia longispina* and *Polyarthra remata* have significantly nonzero Jensen Effects ($p < 0.05$), indicating evidence of nonlinear responses by those species. The δ function of these two species are different from 0 over the pre-specified range of smoothing parameters. In the plots of the δ function of these two species, we observe that their δ functions are above zero over the range of λ, showing that their link functions are convex. Thus, their average growth rate is increased by the presence of environmental variation.

3.6 Conclusion

To determine whether a species' growth rate is increased or decreased by environmental variability, we examine the Jensen Effect. Our first attempt in this direction was to estimate the curvature of the link function in a Functional Single Index model. In our penalized spline based method, we found that unrealistic sample size was needed to obtain accurate estimates. So instead we investigated the net effect of the Jensen's inequality, by comparing the expected response (averaged across the environmental variation) to the response at the expected environment. Inspired by the SiZer method, our test is based on maximizing across a wide range encompassing all plausible smoothing parameter values, thus avoiding the need for smoothing parameter selection. We have shown that our method and test work well on both simulated and real data.

There are multiple potential extensions of this methodology. We have used observed data as representative of the covariates of interest. However, the test can be conducted for any assumed distribution of covariates and it may be of interest to describe regions of single index values in which the estimated link function exhibits the Jensen's inequality. We have also applied this model to growth data, using least-squares. For survival data all standard link functions impose particular curvature structures. These can be modified by adding a nonparametric function within the link, but obtaining equivalent tests – since the relevant effect is on survival itself, not its logit-transform – requires further development.

CHAPTER 4

JENSEN EFFECT IN EXPONENTIAL SINGLE INDEX MODEL

4.1 Introduction

In natural ecosystems, environmental conditions are highly variable over time and space (e.g., Vasseur and McCann, 2007) and many classical questions in ecology and evolution are therefore concerned with the potential consequences of this variation to different organisms. We want to know whether environmental variability is beneficial or harmful for some components of species growth rate, survival probability and number of offspring. The nonlinear response to environment factor is modeled in a (functional) single index model for growth rates, or its generalized version for binary or count response variables, and the impact of environmental variability will depend on whether the nonlinear response curve is concave-up or concave-down.

Nonlinear responses to environmental variability can maintain the coexistence of competing species, (see for example Hutchinson (1961); Chesson and Warner (1981); Ellner (1987); Chesson (1994, 2000b,a)), if their growth rates are maximized under varying environment. Environmental variability can either increase population growth rate (Drake (2005); Koons et al. (2009)) or decrease it (Lewontin and Cohen (1969)), depending on the shape of reaction norm of nonlinear response. The nonlinear response can also predict the effect of projected increases or decreases in environmental variability due to climate change.

We first looked at the species growth model, with only one link function involved. Since we would like to find out the convexity of the response curve, one idea was to estimate the curvature of the nonlinear curve in a functional single index model directly.

In Ye and Hooker (2018), we estimated the impact of environmental variability through a nested nonlinear optimization algorithm, with local quadratic approximation. Even though we showed convergence rates and consistency of the estimators, simulation results required unrealistic sample sizes and were sensitive to the choices of bandwidths and initial values. Instead, Ye et al. (2019) developed methods to evaluate the effect of environmental variability on species success by evaluating the Jensen Effect, or the sign of the Jensen's inequality directly. Inspired by the SiZer method introduced in Chaudhuri and Marron (1999) and Marron and Zhang (2005), we estimated the quantity of the Jensen Effect over a range of smoothing parameters, which skipped the cross-validation procedure, and proposed hypothesis test to see the impact of environmental variability.

The nonparametric growth model is

$$\boldsymbol{G} = g(\boldsymbol{E}) + \epsilon, \tag{4.1}$$

where $\boldsymbol{G}$ and $\boldsymbol{E}$ are the growth (or future size) and environment of a species. The function g is the link function to be estimated, and ϵ is a Gaussian random error.

To analyze the situation of variance heteroscedasticity in ecology data, we will take a nonlinear transformation of response variables, such as logarithm. For example, in Ye and Hooker (2018), the plant data was collected between 1926 and 1957 at the U.S. Sheep Experiment Station. We took a logarithm of plant's area at time t_0 and t_1, where t_0 is the observation start time and t_1 is the end time. A relatively large difference of logarithm responses between t_0 and t_1 indicates a high growth rate of the plant within that time period. We want to investigate the Jensen Effect based on a log-transformation response variable directly. In that paper, we obtain a concave relationship between environmental variability and logarithm of growth rate, which indicates that a constant environment will benefit the logarithm of the growth rates. However, we should notice that the log-transformation will naturally impose a convex relationship. We want to fig-

ure out the Jensen Effect based on the original response variable. Denote the logarithm of growth rates as $\boldsymbol{G}^* = \log(\boldsymbol{G})$, which we will observe $\boldsymbol{G}^*$ in real ecology data.

By taking the exponential on both sides, this leads to an exponential single index model as

$$\boldsymbol{G} = \exp\left[g(\boldsymbol{E}) + \epsilon\right], \tag{4.2}$$

where the composite link function is $g^* \doteq \exp(g)$. By investigating this model, we can compare the growth rates under constant environment $\exp(g)[\mathrm{E}(\boldsymbol{E})]$ with respect to varying environment $\mathrm{E}[\exp(g)(\boldsymbol{E})]$.

Write the transformed model as

$$\boldsymbol{G}^* = g(\boldsymbol{E}) + \epsilon, \tag{4.3}$$

which is the single index model we have discussed before.

To complete the model, the environmental $\boldsymbol{E}$ is described by climate histories (such as temperature and precipitation) observed up to daily resolution. Following Teller et al. (2016), these are thought of as functional covariates leading to a representation of $\boldsymbol{E}$ as a functional linear term. Here for simplicity, we will model environmental variability $\boldsymbol{E}$ as a single index, such that

$$\boldsymbol{E} \doteq X\beta, \tag{4.4}$$

where β are the coefficients to be estimated. In this chapter, we will define an exponential single index model, with an additional exponential link, and investigate the impact of environmental variability by looking at the logarithm of species growth. We will evaluate the Jensen Effect by estimating the quantity $\delta \doteq \mathrm{E}[g^*(\boldsymbol{E})] - g^*[\mathrm{E}(\boldsymbol{E})]$. One difficulty on estimating the Jensen Effect in the exponential model is that the curvature of the additional exponential link implies a natural null hypothesis on convexity, as the

exponential function is convex. We are looking for evidence that the curvature might be incorrect. Applying the SiZer method, we will estimate the values of δ over a range of smoothing parameters λ, which only depend on estimates of the link function g. To test if the curve g^* is concave, or $\delta < 0$, we will develop a one-sided hypothesis test, with critical values simulating from a Gaussian process.

We would also like to investigate the impact of environmental variability on the survival probability or offspring count. In Chapter 5, we will extend the single index model to its generalized version. We will define the generalized single index model, and estimate the Jensen Effect for count or binary response variables. Section 5.2 will introduce the Poisson model for count response variables, while Section 5.3 will discuss the logistic model for binary response.

4.2 Exponential Model

4.2.1 Model Formulation and Estimation

In an exponential single index model, g^* is the link function to be compared, and we can estimate it by an estimate of the link function g in a single index model. To assess the impact of environment variability in $\boldsymbol{E}$ on the growth rate $\boldsymbol{G}$, we need to compare $g^*\left[\mathrm{E}\left(\boldsymbol{E}\right)\right]$ (growth under constant environment) and $\mathrm{E}\left[g^*\left(\boldsymbol{E}\right)\right]$ (growth under varying environment). By the Jensen's inequality, we have that $g^*\left[\mathrm{E}\left(\boldsymbol{E}\right)\right] \leq \mathrm{E}\left[g^*\left(\boldsymbol{E}\right)\right]$, if g^* is a convex function. That confirms that, a varying environment will accelerate the growth of a species. Otherwise, if the link function g^* is concave, then we prefer a constant environment at the average of environmental factor $\boldsymbol{E}$.

To investigate whether the composite link function $g^* \doteq \exp(g)$ is convex or concave, we can estimate the second derivative of the link function g in a single index model, to obtain an estimate of $(g^*)''$. As presented in Ye and Hooker (2018), an estimate of g'' by a nested nonlinear optimization algorithm obtains nice theoretical properties, but empirically, is not good enough to determine the convexity of the curve. To improve estimation accuracy of the nested nonlinear optimization problem, it requires large sample sizes. Initial values also influence the estimation accuracy. In addition, cross-validation procedure sometimes fails to select the right bandwidth or smoothing parameter, or increases computational cost of simulation study.

However, it is easier for us to obtain a good estimate of g, or say g^*, compared to an estimate of g''. So we estimate the Jensen Effect quantity $\delta \doteq \mathrm{E}\left[g^*(\boldsymbol{E})\right] - g^*\left[\mathrm{E}(\boldsymbol{E})\right]$ directly, or the sign of the Jensen's inequality, which only involves an estimate of g. If this quantity δ is negative over a range of smoothing parameter λ values (usually $\left[-e^{10}, e^{6}\right]$), the link function g^* is concave. We also develop a hypothesis test to see if the quantity is statistically significant less than 0, which indicates that growth rates will decrease in a varying environment.

Combining (4.2) and (4.4), the exponential single index model is formulated as

$$Y = \exp\left[g(X\beta) + \epsilon\right]. \tag{4.5}$$

Denote $Y^* = \log(Y)$ and we will observe Y^*. Assume that n data points $\left(X_1, Y_1^*\right), \ldots, \left(X_n, Y_n^*\right)$ are observed, and the number of covariates is p. They are independent and identically distributed as (X, Y^*). The logarithm response can be written in a single index model as

$$Y^* = g(X\beta) + \epsilon. \tag{4.6}$$

To ensure the identifiability of the model, we require that $\|\beta\| = 1$. We will estimate β and g through the machinery of smoothing splines.

We use a K_1-dimensional B-spline basis for the link function g. For any s in the domain of $\mathcal{I} \doteq \left\{X_i\hat{\beta}, i = 1, \ldots, n\right\}$, the link function g can be written as

$$g(s) = \phi^\top(s)\,\boldsymbol{d}, \tag{4.7}$$

where $\phi(\cdot)$ and $\boldsymbol{d}$ are K_1-dimensional column vectors.

The coefficients β and $\boldsymbol{d}$ are estimated by minimizing a penalized sum of squares

$$\begin{aligned} \text{PLS} &\doteq \sum_{i=1}^{n}\left[Y_i^* - g(X_i\beta)\right]^2 + \lambda\int_{\mathcal{I}}\left(g^{(2)}(s)\right)^2 \mathrm{d}s \\ &= \sum_{i=1}^{n}\left[Y_i^* - \phi^\top(X_i\beta)\,\boldsymbol{d}\right]^2 + \lambda\boldsymbol{d}^\top\mathbb{P}_g\boldsymbol{d}, \end{aligned} \tag{4.8}$$

where the lower and upper bound of $\hat{g}$ remains unknown until we get an estimate of β. In (4.8), the $(K_1 \times K_1)$-dimensional penalty matrix $\mathbb{P}_g$ is

$$\mathbb{P}_g \doteq \sum_{i=1}^{n}\phi^{(2)}(s_i)\left(\phi^{(2)}\right)^\top(s_i), \tag{4.9}$$

where

$$s_i \doteq X_i\hat{\beta}, \quad i = 1, \ldots, n. \tag{4.10}$$

The equation (4.8) is a nonlinear optimization problem, which we solve numerically using a BFGS algorithm as implemented in the `R` function `optim`. Denote the estimated coefficients as $\hat{\beta}$ and $\hat{\boldsymbol{d}}$, the estimate of the link function g is

$$\hat{g}_i \doteq \hat{g}\left(X_i\hat{\beta}\right) = \phi^\top\left(X_i\hat{\beta}\right)\hat{\boldsymbol{d}}, \tag{4.11}$$

where $i = 1, \ldots, n$.

4.2.2 Jensen Effect

To investigate the Jensen Effect, we need to get an estimate of the quantity $\mathrm{E}\left[g^*\left(\boldsymbol{E}\right)\right] - g^*\left[\mathrm{E}\left(\boldsymbol{E}\right)\right]$, or the sign of the quantity. We can define δ in the exponential model as

$$\begin{aligned}\delta \doteq & \frac{1}{n}\sum_{i=1}^{n} g^*\left(X_i\beta\right) - g^*\left(\bar{X}\beta\right) \\ = & \frac{1}{n}\sum_{i=1}^{n} \exp\left[g\left(X_i\beta\right)\right] - \exp\left[g\left(\bar{X}\beta\right)\right],\end{aligned} \tag{4.12}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Here defining Φ to be matrix of evaluations of basis functions for g, such that $E_i = X_i\hat{\beta}$ and $\Phi_{ij} \doteq \left[\phi_j\left(X_i\hat{\beta}\right)\right]_{ij} = \left[\phi_j\left(E_i\right)\right]_{ij}$, where $i = 1, \cdots, n$ and $j = 1, \cdots, K_1$. At fixed e and λ, the standard smoothing spline estimate is

$$\hat{g}_\lambda\left(e\right) = \phi\left(e\right)^\top\left(\Phi^\top\Phi + \lambda\mathbb{P}_g\right)^{-1}\Phi^\top\mathbf{Y}^*, \tag{4.13}$$

and the estimate of the coefficients $\boldsymbol{d}$ is

$$\hat{\boldsymbol{d}}_\lambda = \left(\Phi^\top\Phi + \lambda\mathbb{P}_g\right)^{-1}\Phi^\top\mathbf{Y}^*. \tag{4.14}$$

We have

$$\hat{g}^*_\lambda\left(e\right) = \exp\left[\hat{g}_\lambda\left(e\right)\right] = \exp\left[\phi\left(e\right)^\top\left(\Phi^\top\Phi + \lambda\mathbb{P}_g\right)^{-1}\Phi^\top\mathbf{Y}^*\right] = \exp\left[\phi\left(e\right)^\top\hat{\boldsymbol{d}}_\lambda\right]. \tag{4.15}$$

Define a $(n+1)$-dimensional column vector $\boldsymbol{a} \doteq \left(\frac{1}{n}, \cdots, \frac{1}{n}, -1\right)^\top$ and the augmented set of evaluation points $\mathbf{e} \doteq \left(E_1, \ldots, E_n, \bar{E}\right)^\top$ with corresponding evaluation matrix Φ^+, where $E_i = X_i\hat{\beta}$ (at each environment value) and $\bar{E} = \frac{1}{n}\sum_{i=1}^{n} X_i\hat{\beta}$ (averaged across all

environment values). We can write

$$\begin{aligned}\hat{\delta}_\lambda \doteq & \frac{1}{n}\sum_{i=1}^{n}\exp\left[\hat{g}_\lambda\left(E_i\right)\right]-\exp\left[\hat{g}_\lambda\left(\bar{E}\right)\right] \\ = & \boldsymbol{a}^\top \exp\left[\Phi_\lambda^+\left(\Phi_\lambda^\top\Phi_\lambda+\lambda\mathbb{P}_g\right)^{-1}\Phi_\lambda^\top\mathbf{Y}^*\right] \\ = & \boldsymbol{a}^\top \exp\left(\Phi_\lambda^+\hat{\boldsymbol{d}}_\lambda\right).\end{aligned} \tag{4.16}$$

Choose two smoothing parameters λ_1 and λ_2, and define the covariance matrix of δ as Σ_δ. By the Delta method (see Appendix C.1) we can approximate Σ_δ as follows:

$$\begin{aligned}\Sigma_\delta\left(\lambda_1,\lambda_2\right) \doteq & \operatorname{cov}\left(\hat{\delta}_{\lambda_1},\hat{\delta}_{\lambda_2}\right) \\ = & \left[\boldsymbol{a}\cdot\exp\left(\Phi_{\lambda_1}^+\hat{\boldsymbol{d}}_{\lambda_1}\right)\right]^\top\Phi_{\lambda_1}^+\operatorname{cov}\left(\hat{\boldsymbol{d}}_{\lambda_1},\hat{\boldsymbol{d}}_{\lambda_2}\right)\left(\Phi_{\lambda_2}^+\right)^\top\left[\boldsymbol{a}\cdot\exp\left(\Phi_{\lambda_2}^+\hat{\boldsymbol{d}}_{\lambda_2}\right)\right],\end{aligned} \tag{4.17}$$

where $\operatorname{cov}\left(\hat{\boldsymbol{d}}_{\lambda_1},\hat{\boldsymbol{d}}_{\lambda_2}\right)$ is calculated in Ye et al. (2019). Recall in (4.8), the penalized least squares criterion to be minimized, and define the matrix of evaluations of the link function bases and derivatives $\Phi_{ij}^{(k)} \doteq \phi_j^{(k)}\left(X_i\hat{\beta}\right)$ and covariates matrix $\mathbb{X}$. We derive gradient of (4.8) as

$$\begin{pmatrix}\nabla_{\boldsymbol{d}}\\ \nabla_\beta\end{pmatrix}=\begin{pmatrix}\Phi^\top\left(\boldsymbol{Y}^*-\Phi^\top\boldsymbol{d}\right)+\lambda\mathbb{P}_g\boldsymbol{d}\\ \mathbb{X}^\top\operatorname{diag}\left[\left(\Phi^{(1)}\right)^\top\boldsymbol{d}\right]\left(\boldsymbol{Y}^*-\Phi^\top\boldsymbol{d}\right)\end{pmatrix}\doteq\begin{pmatrix}\mathbb{Z}_g\\ \mathbb{Z}_\beta\end{pmatrix}\left(\boldsymbol{Y}^*-\Phi^\top\boldsymbol{d}\right)+\begin{pmatrix}\lambda\mathbb{P}_g\boldsymbol{d}\\ \mathbf{0}_p\end{pmatrix}, \tag{4.18}$$

and expected Hessian

$$\mathbb{H}\doteq\begin{pmatrix}\mathbb{Z}_g^\top\mathbb{Z}_g+\lambda\mathbb{P}_g & \mathbb{Z}_g^\top\mathbb{Z}_\beta\\ \mathbb{Z}_\beta^\top\mathbb{Z}_g & \mathbb{Z}_\beta^\top\mathbb{Z}_\beta\end{pmatrix}. \tag{4.19}$$

We can now obtain the sandwich covariance

$$\operatorname{cov}\begin{pmatrix}\boldsymbol{d}\\ \beta\end{pmatrix}=\hat{\sigma}^2\mathbb{H}^{-1}\begin{pmatrix}\mathbb{Z}_g\\ \mathbb{Z}_\beta\end{pmatrix}^\top\begin{pmatrix}\mathbb{Z}_g\\ \mathbb{Z}_\beta\end{pmatrix}\mathbb{H}^{-1}. \tag{4.20}$$

The variance of the random error σ^2 is estimated as

$$\hat{\sigma}^2 = \frac{1}{\mathrm{df}_{\mathrm{res}}} \sum_{i=1}^{n} \left[Y_i^* - \hat{g}\left(X_i \hat{\beta} \right) \right]^2, \tag{4.21}$$

and in Ruppert et al. (2003), the residual degree of freedom is defined as

$$\mathrm{df}_{\mathrm{res}} = n - 2\mathrm{tr}\left(\mathbb{S} \right) + \mathrm{tr}\left(\mathbb{S}\mathbb{S}^\top \right) - p, \tag{4.22}$$

where $\mathbb{S} \doteq \mathbb{S}\left(\lambda \right)$ is the approximate smoother matrix defined in the model.

The estimate of the covariance of the t-statistics is

$$\mathrm{cov}\left(t_{\lambda_1}, t_{\lambda_2} \right) = \frac{\mathrm{cov}\left(\hat{\delta}_{\lambda_1}, \hat{\delta}_{\lambda_2} \right)}{\sqrt{\mathrm{var}\left(\hat{\delta}_{\lambda_1} \right)} \sqrt{\mathrm{var}\left(\hat{\delta}_{\lambda_2} \right)}}, \tag{4.23}$$

$$\Sigma_t = \left[\mathrm{diag}\left(\Sigma_\delta \right) \right]^{-\frac{1}{2}} \Sigma_\delta \left[\mathrm{diag}\left(\Sigma_\delta \right) \right]^{-\frac{1}{2}}. \tag{4.24}$$

The t-statistics is generated by $\boldsymbol{t} \doteq \Sigma_\delta^{-\frac{1}{2}} \hat{\boldsymbol{\delta}}$, where $\hat{\boldsymbol{\delta}} \doteq \left(\hat{\delta}_{\lambda_1}, \cdots, \hat{\delta}_{\lambda_m} \right)$. The null distribution is simulated as a multivariate normal distribution $\boldsymbol{t}_{\mathrm{null}} \sim \mathrm{N}\left(\mathbf{0}_m, \Sigma_t \right)$. The test statistics is defined as $\min_\lambda \boldsymbol{t}$.

4.2.3 Simulation Results

Single Index Model

We test for the Jensen Effect by estimating the quantity δ over a range of smoothing parameters λ (especially at smaller values of λ). In the exponential model, if the composite link function $g^* = \exp\left(g \right)$ is convex, the δ function will be positive for most of λ values, although it may have high variance at low λ and high bias at high λ. For each simulation, we conduct the hypothesis test introduced in previous section.

Our simulation study starts with $p = 5$ covariates generated uniformly on $[0, 0.5]$. Since we need to do the log-transformation, our covariates could only define on the positive real line. The coefficients $\beta = \frac{1}{\sqrt{p}}\mathbf{1}_p$ so that $\|\beta\| = 1$.

To illustrate the Jensen Effect, we choose 4 different link functions, (1) $g^*(s) = \exp(s)$, (2) $g^*(s) = \sqrt{s}$, (3) $g^*(s) = \sin(s)$, (4) $g^*(s) = s$, where the domain is $s \in [0, 0.5]$. In the exponential model, we first need to get an estimate of the link function g in the single index model. So we represent g by a 25-dimensional quintic B-spline basis. For each link function, we simulate 1000 data sets of size 1000, with error standard deviation 0.01. We obtain critical values for our test by simulating 5000 normal samples from the null distribution. Figure 4.1 presents a sample of δ_λ and t_λ functions versus $\log(\lambda)$ for $g^*(s) = \exp(s)$; plots for the other link functions are in Appendix C.2. The rejection rates for these functions are: 0%, 100%, 98.3%, 0.1% respectively. We can see that the hypothesis test will reject $\delta = 0$ when the link function $g^* = \exp(g)$ is concave. We will confirm a convex relationship for example (1) when the rejection rate is nearly 0, and most of the δ values are positive.

Power Analysis

To investigate the statistical power of our hypothesis test, we consider a series of increasing values of σ, which is the standard deviation of random error ϵ. The value of σ is chosen to be $\{0.01, 0.05, 0.08, 0.1, 0.2\}$.

Figure 4.2 presents the rejection rates plotted against different values of σ. We observe a sharp decrease as σ increases, as expected. As the response variables contain more random noise, the rejection rate will decrease.

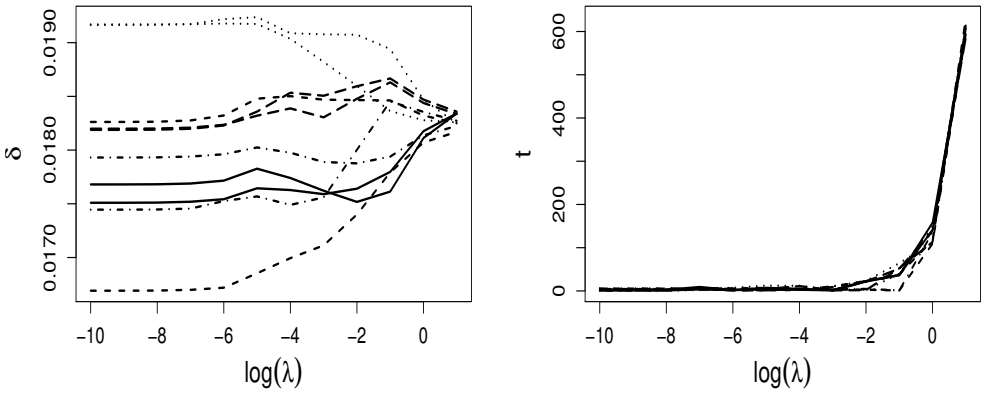

Figure 4.1: This figure illustrates an example of convex function in the exponential single index model, where the link function is $g^*(s) = \exp(s)$. Left: a sample of the Jensen Effect values δ_λ as a function of smoothing parameters λ. Right: the corresponding t-statistics t_λ functions.

Real Data Analysis

Back to our ecology questions, we look at the plant *Artemisia tripartita* (ARTR) observed in the U.S. Sheep Experiment Station (USSES) between 1926 and 1957. Each individual plant was mapped in each of 26 $1m^2$ ungrazed quadrats allowing plant growth (the increase in ground area covered by the plant canopy) to be measured for 22 year-to-year transitions. The response variable are the plant area difference between time t_0 and t_1. From the Q-Q plot in Figure 4.3, there exists some variance heteroscedasticity in the response variable. To fix it, we will take the logarithm transformation and we can see that the nonlinear transformation will fix the problem. Since it uses the logarithm of response variable, we can use the exponential model formulation to find out the Jensen Effect of the real data set. We can answer the question: in which environment, constant or varying, the organisms will grow better.

In total, the data set has 1003 data points with the following variables:

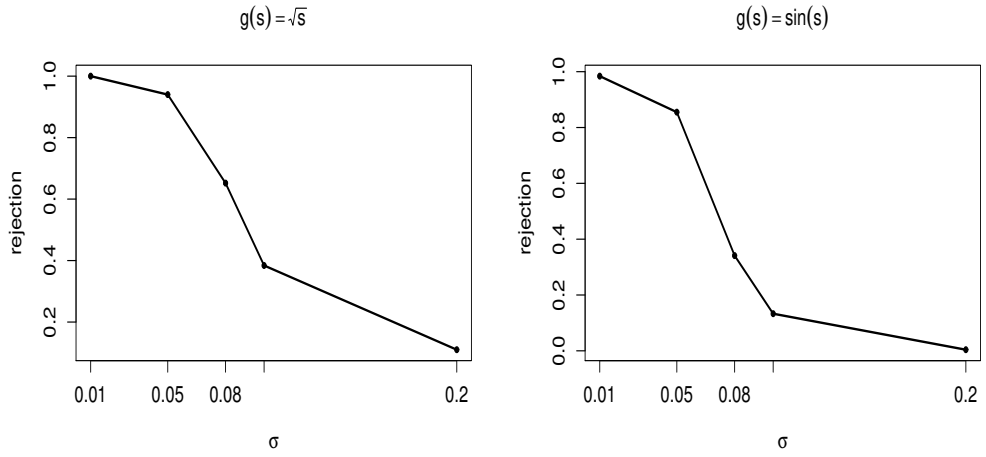

Figure 4.2: This figure illustrates statistical power of the hypothesis test for the Jensen Effect in the exponential single index model. The rejection rates of the Jensen Effect test are plotted against σ values as the standard deviation of the random error. Left: concave function $g^*(s) = \sqrt{s}$. Right: concave function $g^*(s) = \sin(s)$.

1. `logarea.t1` – `logarea.t0`: the difference of plant's logarithm of area between time t_0 and t_1, where t_0 is the observation start time and t_1 is the end time. A relatively large difference indicates a high growth rate of the plant within that time period.

2. `W`: a measure of plant competition. Taken to be a scalar covariate.

3. `p.00` – `p.36`: monthly total precipitation over the previous 36 months observed in Dubois, ID, USA, denoted as $p(s)$.

4. `t.00` – `t.36`: averaged temperature (within a month) over the previous 36 months observed in Dubois, ID, USA, denoted as $t(s)$.

See Teller et al. (2016) for further details. The precipitation and temperature histories are modeled as two covariate functions. We also add an intercept to the single index

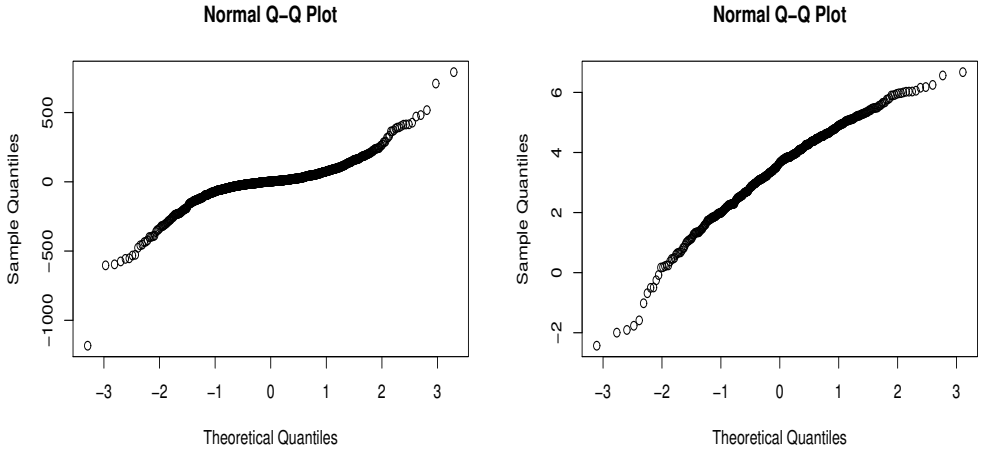

Figure 4.3: This figure illustrates the QQ-plots of the original and log-transformation response variable. The left plot is the original plot, while the right plot is the log-transformation plot. We can observe that the log-transformation greatly reduces the variance heteroscedasticity.

values corresponding to the plant competition variable. The model is

$$\log\left(A_{t_1}\right) - \log\left(A_{t_0}\right) = g\left(\alpha W + \int p\beta_1 + \int t\beta_2\right), \tag{4.25}$$

$$\frac{A_{t_1}}{A_{t_0}} = \exp\left[g\left(\alpha W + \int p\beta_1 + \int t\beta_2\right)\right], \tag{4.26}$$

where $\log\left(A_{t_1}\right) - \log\left(A_{t_0}\right)$ is the difference of plant's logarithm of area between time t_0 and t_1. (4.26) is an exponential functional single index model, and the intercept α, the functions g, β_1 and β_2 need to be estimated. To investigate the Jensen Effect, we need to get an estimate of the composite link function $g^* \doteq \exp\left(g\right)$.

We use the same Fourier basis for coefficient functions β_1 and β_2, where the number of bases are 15. We generate initial values to solve the nonlinear optimization problem by assuming that the composite link function is linear, such that $g^*\left(s\right) = s$. By the simulation results, we will reject the null hypothesis $\delta = 0$, and confirm that environmental variability has some effect on plant growth. Since the hypothesis test considered here is a one-sided test, it also concludes that the link function g^* is concave. It indicates that

the plant observed in this station will grow better in a constant environment. In Chapter 2, we define the Jensen Effect based on the log-transformation of the response variables. However, we would like to know the effect of climate histories on the original response variables. Here, we have answered this question by estimating the Jensen Effect in an exponential single index model.

We observe that most of the Jensen Effect δ values are negative in Figure 4.4, when they are estimated by different λ values. That also confirms that the composite link function g^* is concave over a certain range, and environmental variability is harmful for plant growth in this region. The rest of plots are estimates of functions g^*, β_1 and β_2, where the smoothing parameter λ is selected by generalized cross-validation.

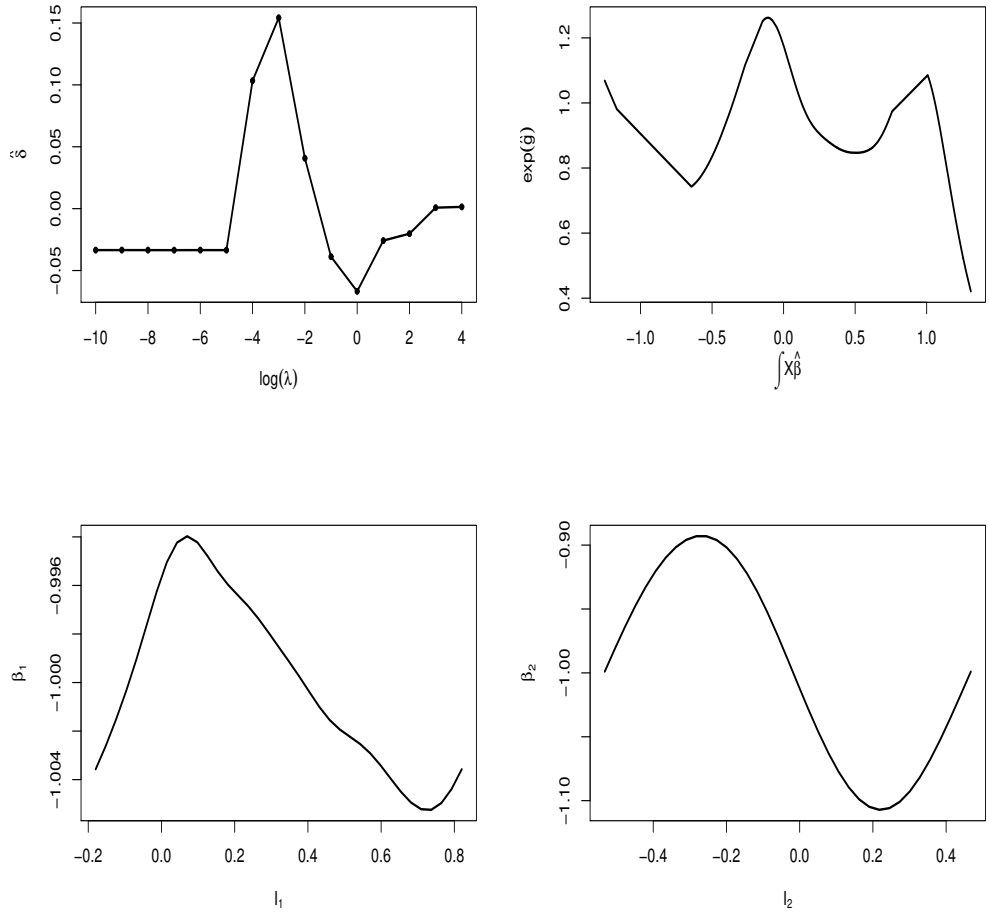

Figure 4.4: This figure illustrates the Jensen Effect for the USSES data set on logarithm of growth rates, using the exponential single index model formulation. Top-left: plot of the Jensen Effect values δ against the logarithm of smoothing parameters $\log(\lambda)$. Top-right: plot of the estimate of the link function g^* corresponding to the λ selected by GCV. Bottom: plots of estimates of the coefficient functions β_1 and β_2 selected by GCV.

CHAPTER 5

JENSEN EFFECT IN GENERALIZED SINGLE INDEX MODEL

5.1 Introduction

In Chapter 4, we define the logarithm of growth rates in an exponential single index model, and estimate the Jensen Effect using the transformed response variables directly. Sometimes, we will observe a binary response variable, such as individual survival, or a Poisson count, such as the number of seedlings or offspring. It is natural to extend our single index model to generalized linear model. We can define a Generalized Single Index model as

$$\eta := g(X\beta), \tag{5.1}$$

$$\mu = h(\eta) = h[g(X\beta)], \tag{5.2}$$

$$Y \sim \text{Distribution}(\mu). \tag{5.3}$$

The quantity η is defined by a single index model. h is the GLM link function, and the composite link function is defined as $h \star g$. In Chapter 4, the link function h is exponential. The response variable is generated according to the GLM distribution.

In Section 5.2, the GLM link function h is exponential, and the response variable Y is generated from Poisson distribution to model the number of offspring in ecology questions. Since the Poisson model shares the same link function h with the exponential model, we can borrow some ideas from the exponential model to find out the Jensen Effect, or the sign of the Jensen's inequality. Then we can answer the ecological question: whether environmental variability will increase or decrease the number of offspring of a species.

In Section 5.3, we will investigate the logistic single index model, where the GLM link h is a logistic function, and we will observe a binary response, such as individual survival. We can determine the impact of environmental variability on survival probability.

5.2 Poisson Model

5.2.1 Model Formulation and Estimation

The count response variable Y is generated by a Poisson(μ) distribution, where η is generated by a single index model, such as

$$\eta := g(X\beta), \tag{5.4}$$

$$\mu = \exp(\eta) = \exp[g(X\beta)]. \tag{5.5}$$

Recall that μ is the mean and variance of a Poisson distribution and the GLM link function h is exponential.

Fix the number of observations n, define $X_1, \cdots, X_n$ as the covariates, and $Y_1, \cdots, Y_n$ as the response generated by a Poisson distribution. The Jensen Effect δ is defined as

$$\delta = \frac{1}{n}\sum_{i=1}^{n} \exp[g(X_i\beta)] - \exp\left[g\left(\bar{X}\beta\right)\right], \tag{5.6}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Since we are dealing with generalized linear model, the nonlinear optimization objective function will include the log-likelihood of the model. The log-likelihood of a

Poisson single index model is

$$\begin{aligned}\ell &= \sum_{i=1}^{n} \left[Y_i \eta_i - \exp\left(\eta_i\right) \right] \\ &= \sum_{i=1}^{n} \left\{ Y_i g\left(X_i \beta\right) - \exp\left[g\left(X_i \beta\right) \right] \right\}.\end{aligned} \tag{5.7}$$

Use a B-spline basis to represent the link function g, such that $g(s) = \phi^{\top}(s)\boldsymbol{d}$. Assume $\mathbb{P}_g$ as the second derivative penalty matrix for g, the penalized log-likelihood is

$$\begin{aligned}\text{PLL} &= -\ell + \lambda \boldsymbol{d}^{\top} \mathbb{P}_g \boldsymbol{d} \\ &= -\sum_{i=1}^{n} \left\{ Y_i \phi^{\top}\left(X_i \beta\right) \boldsymbol{d} - \exp\left[\phi^{\top}\left(X_i \beta\right) \boldsymbol{d} \right] \right\} + \lambda \boldsymbol{d}^{\top} \mathbb{P}_g \boldsymbol{d}.\end{aligned} \tag{5.8}$$

For a fixed smoothing parameter λ, we can estimate the coefficients β and $\boldsymbol{d}$ by minimizing the penalized log-likelihood (5.8). We can obtain an estimate of the link function g, μ and η, when we get an estimate of β and $\boldsymbol{d}$. Denote the estimates as $\hat{\beta}_\lambda$ and $\hat{\boldsymbol{d}}_\lambda$, we have $\hat{\boldsymbol{g}}_\lambda = \hat{\boldsymbol{\eta}}_\lambda = \phi^{\top}\left(X\hat{\beta}_\lambda\right)\hat{\boldsymbol{d}}_\lambda = \Phi_\lambda \hat{\boldsymbol{d}}_\lambda$ and $\hat{\boldsymbol{\mu}}_\lambda = \exp\left(\hat{\boldsymbol{g}}_\lambda\right)$, where $\hat{\boldsymbol{g}}_\lambda$ and $\hat{\boldsymbol{\mu}}_\lambda$ are n-dimensional column vectors evaluated at each observation and estimated β.

Define a $(n+1)$-dimensional column vector $\boldsymbol{a} = \left(\frac{1}{n}, \cdots, \frac{1}{n}, -1\right)^{\top}$ and the augmented set of evaluation points $\mathbf{e}_\lambda = \left(E_1, \ldots, E_n, \bar{E}\right)^{\top}$, where $E_i = X_i \hat{\beta}_\lambda$, with corresponding evaluation matrix Φ_λ^{+}. We can write

$$\hat{\delta}_\lambda = \frac{1}{n}\sum_{i=1}^{n} \exp\left[\hat{g}_\lambda\left(E_i\right)\right] - \exp\left[\hat{g}_\lambda\left(\bar{E}\right)\right] = \boldsymbol{a}^{\top} \exp\left(\Phi_\lambda^{+} \hat{\boldsymbol{d}}_\lambda\right). \tag{5.9}$$

Define the adjusted response vector as

$$\boldsymbol{z}_\lambda = \hat{\boldsymbol{\eta}}_\lambda + \frac{\boldsymbol{Y} - \hat{\boldsymbol{\mu}}_\lambda}{\hat{\boldsymbol{\mu}}_\lambda}, \tag{5.10}$$

and a weight matrix $\mathbb{W}_\lambda = \text{diag}\left(\hat{\boldsymbol{\mu}}_\lambda\right)$, then an estimate of the coefficients $\boldsymbol{d}$ is

$$\hat{\boldsymbol{d}}_\lambda = \left(\Phi_\lambda^{\top} \mathbb{W}_\lambda \Phi_\lambda + \lambda \mathbb{P}_g\right)^{-1} \Phi_\lambda^{\top} \mathbb{W}_\lambda \boldsymbol{z}_\lambda. \tag{5.11}$$

By the Delta method introduced in Appendix C.1, fix smoothing parameters λ_1 and λ_2, the covariance between δ_{λ_1} and δ_{λ_2} is

$$\begin{aligned}\Sigma_\delta(\lambda_1,\lambda_2) =& \operatorname{cov}\left(\hat{\delta}_{\lambda_1},\hat{\delta}_{\lambda_2}\right) \\ =& \left[\boldsymbol{a}\cdot\exp\left(\Phi^+_{\lambda_1}\hat{\boldsymbol{d}}_{\lambda_1}\right)\right]^\top \Phi^+_{\lambda_1}\operatorname{cov}\left(\hat{\boldsymbol{d}}_{\lambda_1},\hat{\boldsymbol{d}}_{\lambda_2}\right)\left(\Phi^+_{\lambda_2}\right)^\top\left[\boldsymbol{a}\cdot\exp\left(\Phi^+_{\lambda_2}\hat{\boldsymbol{d}}_{\lambda_2}\right)\right].\end{aligned} \tag{5.12}$$

By Friedman et al. (2001), the first and second derivative of the log-likelihood ℓ with respect to $\boldsymbol{d}$ in a Poisson model are

$$\frac{\partial\ell}{\partial\boldsymbol{d}} = \Phi^\top(\boldsymbol{Y}-\boldsymbol{\mu}), \tag{5.13}$$

$$\frac{\partial^2\ell}{\partial\boldsymbol{d}^2} = -\Phi^\top\mathbb{W}\Phi. \tag{5.14}$$

Then we have

$$\frac{\partial(\mathrm{PLL})}{\partial\boldsymbol{d}} = \Phi^\top(\boldsymbol{\mu}-\boldsymbol{Y})+\lambda\boldsymbol{d}^\top\mathbb{P}_g, \tag{5.15}$$

$$\frac{\partial^2(\mathrm{PLL})}{\partial\boldsymbol{d}^2} = \Phi^\top\mathbb{W}\Phi+\lambda\mathbb{P}_g. \tag{5.16}$$

By the Taylor expansion, at fixed points $\boldsymbol{d}$ and $\boldsymbol{d}_0$, we can approximate

$$\mathrm{PLL}(\boldsymbol{d}) \approx \mathrm{PLL}(\boldsymbol{d}_0)+\frac{\partial(\mathrm{PLL})}{\partial\boldsymbol{d}_0}(\boldsymbol{d}-\boldsymbol{d}_0)+\frac{\partial^2(\mathrm{PLL})}{\partial\boldsymbol{d}_0^2}\frac{(\boldsymbol{d}-\boldsymbol{d}_0)^2}{2}. \tag{5.17}$$

In equation (5.17), we assume that $\frac{\partial^2(\mathrm{PLL})}{\partial\boldsymbol{d}_0^2} \to \frac{\partial^2(\mathrm{PLL})}{\partial\boldsymbol{d}^2}$ and we have the covariance between $\hat{\boldsymbol{d}}_{\lambda_1}$ and $\hat{\boldsymbol{d}}_{\lambda_2}$ is

$$\begin{aligned}\operatorname{cov}\left(\hat{\boldsymbol{d}}_{\lambda_1},\hat{\boldsymbol{d}}_{\lambda_2}\right) \approx& \left(\frac{\partial^2(\mathrm{PLL})}{\partial\hat{\boldsymbol{d}}^2_{\lambda_1}}\right)^{-1}\operatorname{cov}\left(\frac{\partial(\mathrm{PLL})}{\partial\hat{\boldsymbol{d}}_{\lambda_1}},\frac{\partial(\mathrm{PLL})}{\partial\hat{\boldsymbol{d}}_{\lambda_2}}\right)\left(\frac{\partial^2(\mathrm{PLL})}{\partial\hat{\boldsymbol{d}}^2_{\lambda_2}}\right)^{-1} \\ \approx& \left(\Phi^\top_{\lambda_1}\mathbb{W}_{\lambda_1}\Phi_{\lambda_1}+\lambda_1\mathbb{P}_g\right)^{-1}\Phi^\top_{\lambda_1}\operatorname{cov}(Y)\Phi_{\lambda_2}\left(\Phi^\top_{\lambda_2}\mathbb{W}_{\lambda_2}\Phi_{\lambda_2}+\lambda_2\mathbb{P}_g\right)^{-1}.\end{aligned} \tag{5.18}$$

We know that $\operatorname{cov}(Y)=\boldsymbol{\mu}$, and $\mathbb{W}_\lambda = \operatorname{diag}(\hat{\boldsymbol{\mu}}_\lambda)$. In our simulations, we use the λ selected by generalized cross-validation. By Wood (2006), the GCV value in a penalized

generalized linear model is

$$\mathrm{GCV}(\lambda) = \frac{n \left\| \sqrt{\mathbb{W}_\lambda} \left(\boldsymbol{z}_\lambda - \Phi_\lambda \hat{\boldsymbol{d}}_\lambda \right) \right\|^2}{[n - \mathrm{tr}(\mathbb{S}_\lambda)]^2}, \tag{5.19}$$

where the smoother matrix is

$$\mathbb{S}_\lambda = \Phi_\lambda^\top \left(\Phi_\lambda^\top \mathbb{W}_\lambda \Phi_\lambda + \lambda \mathbb{P}_g \right)^{-1} \Phi_\lambda^\top \mathbb{W}_\lambda, \tag{5.20}$$

and λ_{gcv} is selected by minimizing GCV(λ).

Therefore, the covariance matrix is formulated as

$$\mathrm{cov}\left(\hat{\boldsymbol{d}}_{\lambda_1}, \hat{\boldsymbol{d}}_{\lambda_2} \right) \approx \left(\Phi_{\lambda_1}^\top \mathbb{W}_{\lambda_1} \Phi_{\lambda_1} + \lambda_1 \mathbb{P}_g \right)^{-1} \Phi_{\lambda_1}^\top \mathbb{W}_{\lambda_{\mathrm{gcv}}} \Phi_{\lambda_2} \left(\Phi_{\lambda_2}^\top \mathbb{W}_{\lambda_2} \Phi_{\lambda_2} + \lambda_2 \mathbb{P}_g \right)^{-1}. \tag{5.21}$$

Putting (5.12) and (5.21) together, we have an estimate of the covariance matrix of $\hat{\delta}$. The estimate of the covariance of the t-statistics is

$$\mathrm{cov}(t_{\lambda_1}, t_{\lambda_2}) = \frac{\mathrm{cov}\left(\hat{\delta}_{\lambda_1}, \hat{\delta}_{\lambda_2} \right)}{\sqrt{\mathrm{var}\left(\hat{\delta}_{\lambda_1} \right)} \sqrt{\mathrm{var}\left(\hat{\delta}_{\lambda_2} \right)}}, \tag{5.22}$$

$$\Sigma_t = \left[\mathrm{diag}(\Sigma_\delta) \right]^{-\frac{1}{2}} \Sigma_\delta \left[\mathrm{diag}(\Sigma_\delta) \right]^{-\frac{1}{2}}.$$

The t-statistics is generated by $\boldsymbol{t} = \Sigma_\delta^{-\frac{1}{2}} \hat{\boldsymbol{\delta}}$. The null distribution is simulated as a multivariate normal distribution $\boldsymbol{t}_{\mathrm{null}} \sim \mathrm{N}(\boldsymbol{0}, \Sigma_t)$. The test statistics is defined as $\min_\lambda \boldsymbol{t}$.

5.2.2 Simulation Results

Like the exponential model, we test for the Jensen Effect by calculating the quantity δ over a range of smoothing parameters λ. In the Poisson single index model, if the composite link function $g^* = \exp(g)$ is convex, the δ function will be positive for most of λ values. For each simulation, we simulate the null distribution from a Gaussian process, and conduct a one-sided hypothesis test to see if $\delta < 0$.

Our simulation study starts with $p = 5$ covariates generated uniformly on $[0, 0.5]$. The coefficients $\beta = \frac{1}{\sqrt{p}}\mathbf{1}_p$ so that $\|\beta\| = 1$.

To illustrate the Jensen Effect, we choose 3 different link functions, (1) $g^*(s) = \exp\left(\frac{s}{8}\right)$, (2) $g^*(s) = \frac{30}{1+\exp\left(-\frac{s}{8}\right)}$, (3) $g^*(s) = s$. In the Poisson single index model, we first need to get an estimate of the link function g, so we represented g by a 25-dimensional quintic B-spline basis. For each link function, we simulated 200 data sets of size 1000, with error standard deviation 0.01. We obtained critical values for our test by simulating 5000 normal samples from the null distribution. Figure 5.1 presents a sample of δ_λ and t_λ functions versus $\log(\lambda)$ for $g^*(s) = \exp\left(\frac{s}{8}\right)$; plots for the other link functions are in Appendix D.1. The rejection rates for these functions are: 0%, 82.5%, 1.5% respectively. The convex function has positive Jensen Effect value δ so that we observe a zero rejection rate as expected.

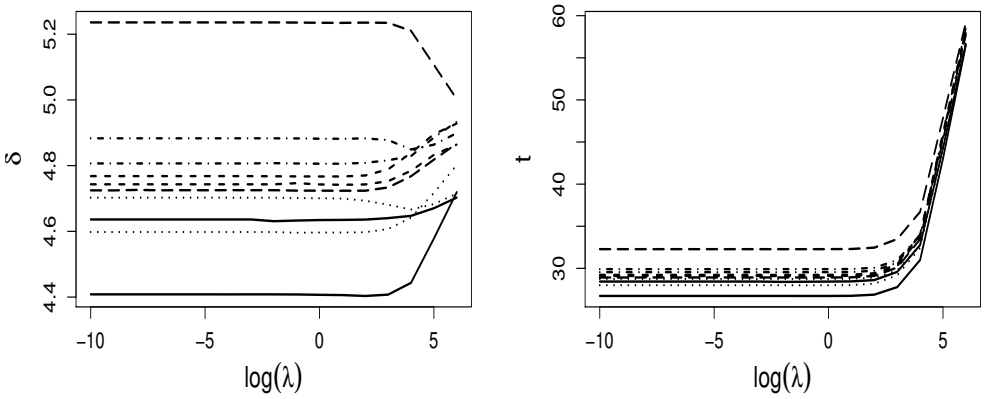

Figure 5.1: This figure illustrates an example of convex function in the Poisson single index model, where the link function is $g^*(s) = \exp\left(\frac{s}{8}\right)$. Left: a sample of the Jensen Effect values δ_λ as a function of smoothing parameters λ. Right: the corresponding t-statistics t_λ functions.

5.2.3 Power Analysis

To investigate the power of the hypothesis test in a Poisson single index model, we use a concave link function $h \star g$ such that

$$(h \star g)(s) = \exp(g)(s) = \frac{30}{1 + \exp\left(-\frac{s}{a}\right)}. \tag{5.23}$$

The values of a are $\{2, 4, 6, 8, 10, 12, 14, 16\}$. From the plots of g^* in Figure 5.2, we can observe that g^* becomes more concave for the first half values of a, and becomes less concave for the second half values of a. The rejection rates and the true Jensen Effect δ values are in Figure 5.3. We can see that as the link function becomes more concave, the rejection rate will converge to 1; while when it becomes less concave, the rejection rate will converge to 0.

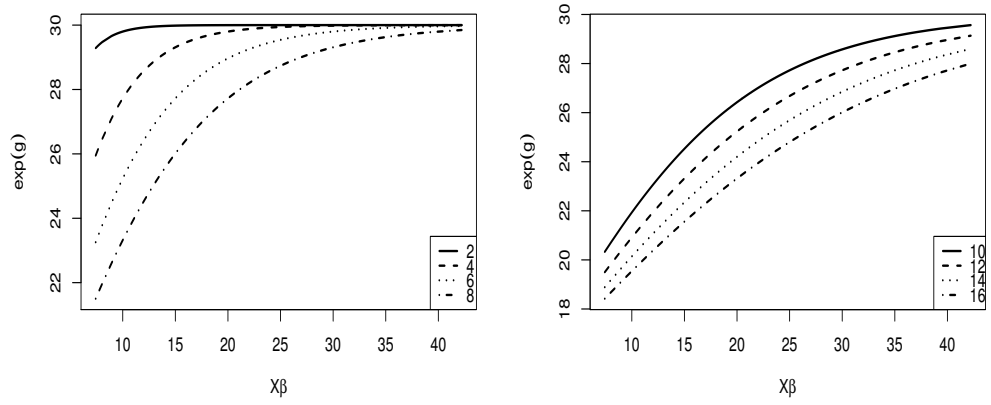

Figure 5.2: This figure illustrates the concavity of link functions g^* in a Poisson single index model, with $(h \star g)(s) = \exp(g)(s) = \frac{30}{1+\exp\left(-\frac{s}{a}\right)}$. Left: the values of a are $\{2, 4, 6, 8\}$. The link function becomes more concave as the value of a increases. Right: the values of a are $\{10, 12, 14, 16\}$. The link function becomes less concave as the value of a increases.

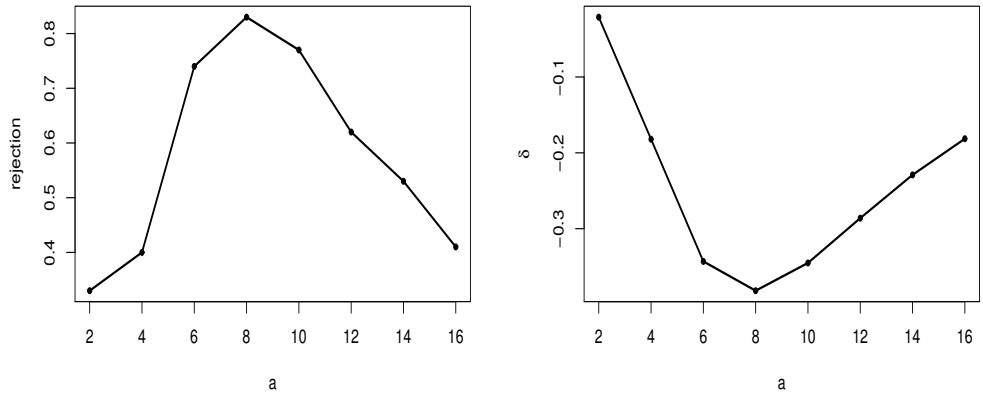

Figure 5.3: This figure illustrates statistical power of the hypothesis test for the Jensen Effect in the Poisson single index model. The Poisson link function is $\exp(g)(s) = \frac{30}{1+\exp\left(-\frac{s}{a}\right)}$, where the values of a determine the concavity of the link functions. Left: the rejection rates plotted against the values of a. Right: the true Jensen Effect δ values plotted against the values of a.

5.2.4 Real Data Analysis

We analyzed ecological data from the U.S. Sheep Experiment Station (USSES), which is located 9.6 km north of Dubois, Idaho (44.2°N, 112.1°W), 1500m above sea level. Ecologists at the USSES established 26 1 m^2 quadrats between 1926 and 1932. Among them, 18 quadrats were distributed among 4 ungrazed exclosures, and 8 were distributed in 2 pastures grazed at medium intensity spring through fall. All quadrats were located on similar topography and soils.

During the period of data collection, mean annual precipitation was 270 mm and mean temperatures ranged from −8°C (January) to 21°C (July). The vegetation is dominated by the shrub, *Artemisia tripartita* (ARTR), and the C3 perennial bunchgrasses *Pseudoroegneria spicata*, *Hesperostipa comata*, and *Poa secunda* (PSSP, HECO, and POSE, respectively). These 4 species, the focus of our analyses here, comprised over

70% of basal cover (grasses) and 60% of canopy cover (shrubs and forbs).

For each plant, we use the number of seedlings as the response variable, which is `NRquad` in the data set. The 5 covariates are: `totParea` (total area), `ppt1`, `ppt2` (precipitation), `TmeanSpr1`, `TmeanSpr2` (temperature). Since the response variable is the seedlings count, we will establish a Poisson single index model for the data. Define η as

$$\eta = \alpha W + g\left(X\beta\right), \tag{5.24}$$

where W denotes the total area, and X is the covariate vector for temperature and precipitation.

The covariance between W and $g\left(X\beta\right)$ has been investigated as a separate mechanism that can serve to stabilize species coexistence, generally described as the *storage effect* (Chesson, 1994; Ellner et al., 2016) in addition to the *relative nonlinearity* provided by differing shapes of g. We note the potential of these methods to also investigate the storage effect but do not pursue this here. In this paper, we will estimate the Jensen Effect only for the link function g and perform the hypothesis test to see the impact of environmental variability on the number of seedlings for each plant.

Hypothesis tests show that plants *Hesperostipa comata* (HECO) and *Pseudoroegneria spicata* (PSSP) reject the null hypothesis, which indicates that the nonlinear curve is concave in a certain environmental range, and most of the Jensen Effect values δ are negative for these plants. The number of HECO and PSSP seedlings will decrease in a varying environment. In Figure 5.4 and D.5, we can observe that nearly all δ values are negative, which also confirms that a constant environment will benefit the plants reproduction. The right plots in these figures illustrate an example of $\exp\left(g\right)$ estimate, corresponding to the minimum of δ. We can see that there are several peaks corresponding to some single index values, or say environmental factor. Previous ecology theories

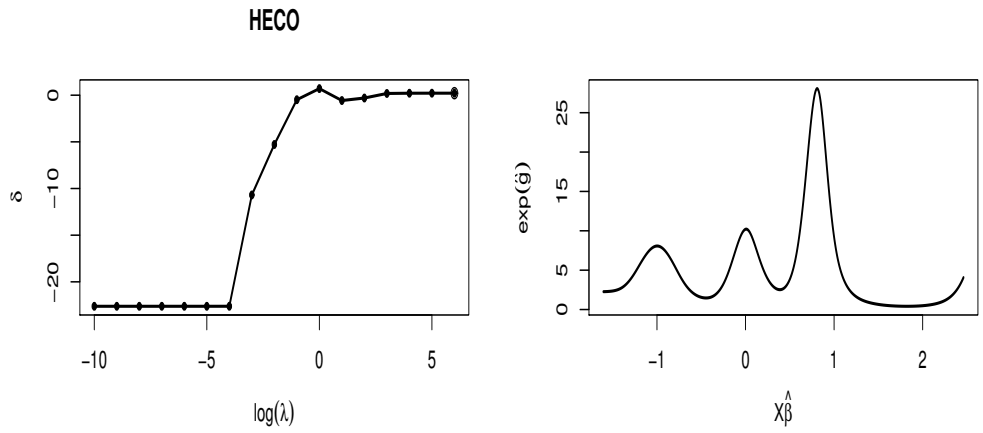

Figure 5.4: This figure illustrates the Jensen Effect for the plant *Hesperostipa comata* (HECO) in USSES data set, where we formulate the data in a Poisson single index model. Left: the Jensen Effect values δ plotted against the values of λ. Right: the curve of the exponential link function $\exp(\hat{g})$, where λ is corresponding to the minimum of δ.

also showed that in some environmental ranges, the number of seedlings of a plant will be abundant.

For all other plants, the nonlinear curves are not concave by the hypothesis test, and most of their δ estimates appear to be positive. It concludes that environmental variability will increase the number of seedlings. Figures D.3 and D.4 in Appendix D.2 show the estimated link functions $\exp(g)$ with λ selected by generalized cross-validation, and we can observe that all these curves are convex.

5.3 Logistic Model

In the previous Section 5.2, we investigate the relationship between the number of offspring of a species and environmental variability. We define the Poisson single index model, and estimate the Jensen Effect to determine that whether varying environment

will increase or decrease the number of offspring.

Sometimes, we will observe a binary response variable, such as individual survival. The binary response can be described by a logistic single index model. We want to predict the probability of survival with respect to environmental factor. We know that logistic model can capture the success probability of a binary response (survival or not), and we will use the logistic link to model the survival data. The standard logistic model enforces concavity above $\mathrm{P}(Y = 1) = 0.5$.

5.3.1 Model Formulation and Estimation

Denote 1 as individual survival and 0 as death. The survival status (binary response) of a species $Y \in \{1, 0\}$ is generated by a Bernoulli(π) distribution, where π is the probability of survival, such as

$$\eta = g(X\beta), \tag{5.25}$$

$$\pi = \text{logistic}(\eta) = \frac{1}{1 + \exp[-g(X\beta)]}, \tag{5.26}$$

where the logistic link is defined as $\text{logistic}(x) = \frac{1}{1+\exp(-x)}$. The binary response Y and the single index model is related by the logistic link.

Fix the number of observations n, define $X_1, \cdots, X_n$ as the covariates, and $Y_1, \cdots, Y_n$ as response generated by a Bernoulli distribution. The Jensen Effect quantity δ is defined as

$$\delta = \frac{1}{n}\sum_{i=1}^{n} \text{logistic}[g(X_i\beta)] - \text{logistic}\left[g\left(\bar{X}\beta\right)\right], \tag{5.27}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

The log-likelihood of a logistic model is

$$\begin{aligned}\ell &= \sum_{i=1}^{n} \{Y_i \eta_i - \log\left[1 + \exp\left(\eta_i\right)\right]\} \\ &= \sum_{i=1}^{n} \{Y_i g\left(X_i \beta\right) - \log\left[1 + \exp\left(g\left(X_i \beta\right)\right)\right]\}.\end{aligned} \tag{5.28}$$

Use a B-spline basis function to represent the link function g, such that $g(s) = \phi^{\top}(s)\,\boldsymbol{d}$. Assume $\mathbb{P}_g$ as the second derivative penalty matrix for g, the penalized log-likelihood is

$$\begin{aligned}\text{PLL} &= -\ell + \lambda \boldsymbol{d}^{\top} \mathbb{P}_g \boldsymbol{d} \\ &= -\sum_{i=1}^{n} \{Y_i \phi^{\top}\left(X_i \beta\right) \boldsymbol{d} - \log\left[1 + \exp\left(\phi^{\top}\left(X_i \beta\right) \boldsymbol{d}\right)\right]\} + \lambda \boldsymbol{d}^{\top} \mathbb{P}_g \boldsymbol{d}.\end{aligned} \tag{5.29}$$

For a fixed smoothing parameter λ, we can estimate the coefficients β and $\boldsymbol{d}$ by minimizing (5.29). We can obtain an estimate of the link function g, π and η, when we get an estimate of β and $\boldsymbol{d}$. Denote the estimates as $\hat{\beta}_\lambda$ and $\hat{\boldsymbol{d}}_\lambda$, we have $\hat{\boldsymbol{g}}_\lambda = \hat{\boldsymbol{\eta}}_\lambda = \phi^{\top}\left(X \hat{\beta}_\lambda\right) \hat{\boldsymbol{d}}_\lambda = \Phi_\lambda \hat{\boldsymbol{d}}_\lambda$ and $\hat{\boldsymbol{\pi}}_\lambda = \text{logistic}\left(\hat{\boldsymbol{g}}_\lambda\right)$, where $\hat{\boldsymbol{g}}_\lambda$ and $\hat{\boldsymbol{\pi}}_\lambda$ are n-dimensional column vectors evaluated at each observation.

At fixed λ, an estimate of the Jensen Effect is

$$\hat{\delta}_\lambda = \frac{1}{n} \sum_{i=1}^{n} \text{logistic}\left[\hat{g}_\lambda\left(E_i\right)\right] - \text{logistic}\left[\hat{g}_\lambda\left(\bar{E}\right)\right] = \boldsymbol{a}^{\top} \text{logistic}\left(\Phi_\lambda^{\dagger} \hat{\boldsymbol{d}}_\lambda\right). \tag{5.30}$$

Define the adjusted response vector as

$$\boldsymbol{z}_\lambda = \hat{\boldsymbol{\eta}}_\lambda + \frac{\boldsymbol{Y} - \hat{\boldsymbol{\pi}}_\lambda}{\hat{\boldsymbol{\pi}}_\lambda\left(\mathbf{1}_n - \hat{\boldsymbol{\pi}}_\lambda\right)}, \tag{5.31}$$

and a weight matrix $\mathbb{W}_\lambda = \text{diag}\left(\hat{\boldsymbol{\pi}}_\lambda\left(\mathbf{1}_n - \hat{\boldsymbol{\pi}}_\lambda\right)\right)$, then an estimate of the coefficients $\boldsymbol{d}$ is

$$\hat{\boldsymbol{d}}_\lambda = \left(\Phi_\lambda^{\top} \mathbb{W}_\lambda \Phi_\lambda + \lambda \mathbb{P}_g\right)^{-1} \Phi_\lambda^{\top} \mathbb{W}_\lambda \boldsymbol{z}_\lambda. \tag{5.32}$$

The covariance matrices Σ_δ and $\Sigma_{\boldsymbol{d}}$ are estimated in (5.12) and (5.18). Sometimes, we observe a non-symmetric covariance matrix $\Sigma_{\boldsymbol{d}}$ for simulation studies in a logistic single index model. If that happens, we can symmetrize it by $\Sigma_{\boldsymbol{d}} \approx \frac{1}{2}\left(\Sigma_{\boldsymbol{d}} + \Sigma_{\boldsymbol{d}}^\top\right)$.

Denote the singular value decomposition (SVD) of the covariance matrix Σ_δ as

$$\Sigma_\delta = \mathbb{U}_\delta \mathbb{V}_\delta \mathbb{U}_\delta^\top. \tag{5.33}$$

If some of eigenvalues of Σ_δ in (5.33) are not positive, we will only consider the positive eigenvalues, and the corresponding eigenvectors. We will remove the non-positive eigenvalues and eigenvectors, and denote the new set of eigenvalues and eigenvectors as $\mathbb{V}_\delta^+$ and $\mathbb{U}_\delta^+$. Then we can reformulate the covariance matrix of δ by

$$\Sigma_\delta = \mathbb{U}_\delta^+ \mathbb{V}_\delta^+ \left(\mathbb{U}_\delta^+\right)^\top. \tag{5.34}$$

The t-statistics is generated by $\boldsymbol{t} = \Sigma_\delta^{-\frac{1}{2}} \hat{\boldsymbol{\delta}}$. The null distribution is simulated as a multivariate normal distribution $\boldsymbol{t}_{\text{null}} \sim \mathrm{N}(\mathbf{0}, \Sigma_t)$. The test statistics is defined as $\max_\lambda \boldsymbol{t}$.

5.3.2 Power Analysis

As in the Poisson model, we test for the Jensen Effect by calculating the quantity δ over a range of smoothing parameters λ. In the logistic single index model, if the composite link function $h \star g = \text{logistic}(g)$ is convex, the δ function will be positive for most of λ values. For each simulation, we simulate the null hypothesis from a Gaussian process, and conduct the hypothesis test introduced in previous section.

The $p = 5$ covariates are generated uniformly on $[0, 0.5]$. The coefficient $\beta = \frac{1}{\sqrt{p}}\mathbf{1}_p$ so that $\|\beta\| = 1$. To estimate g and β, we represent g by a 25-dimensional quintic B-spline basis.

To illustrate the statistical power of the hypothesis test, we choose the convex link functions as

$$(h \star g)(s) = \text{logistic}(g)(s) = \exp(-as), \tag{5.35}$$

where the coefficient a is selected from $\{.1, .2, .3, .4, .5, 1, 2, 3, 4, 5\}$. For each link function, we simulated 100 data sets of size 1000, with error standard deviation 0.01. We obtain critical values for our test by simulating 5000 normal samples from the null distribution. Figure 5.5 presents the rejection rates of the hypothesis test.

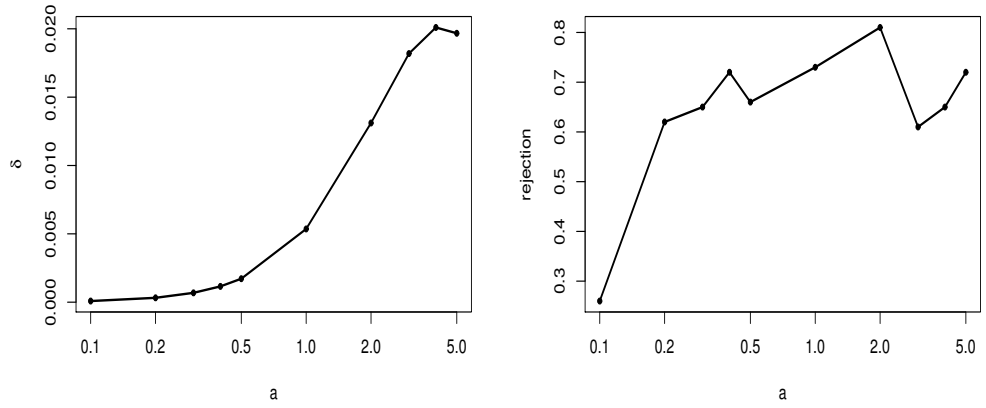

Figure 5.5: This figure illustrates statistical power of the hypothesis test for the Jensen Effect in the logistic single index model. The logistic link function is $\text{logistic}(g)(s) = \exp(-as)$, where the values a are $\{.1, .2, .3, .4, .5, 1, 2, 3, 4, 5\}$. The values a determine the convexity of the link functions. Left: the true Jensen Effect value δ plotted against each value of a. Right: the rejection rate corresponding to each value of a. The values of a are plotted on a log scale.

CHAPTER 6

CONCLUSION

Environment are varying over time and space, and ecology data confirms that environmental variability has impacts on some components of species growth rates. Our research is inspired by an interesting ecology question: whether a constant or varying environment will increase the growth rates of a species. We modeled the nonlinear response of environmental factor on growth rates in a functional single index model. To answer the ecology question, we would like to know the convexity of the nonlinear curve, or the sign of the Jensen's inequality.

It is known to us that the sign of the second derivative of a function determines the convexity of the curve. In Chapter 2, we used a nested nonlinear optimization algorithm with local quadratic approximation to estimate the curvature. We showed the convergence rates and consistency of our estimators. However, empirical implementation of the algorithm requires unrealistic sample sizes for initial value and bandwidth selections. Our estimates for the second derivative may mislead us to answer the ecology questions to some extent.

Instead, we estimated the Jensen Effect quantity directly in Chapter 3, which only involved an estimate of the curve g. One big obstacle of our estimation procedure is to select a good smoothing parameter. Inspired by the SiZer method, we estimated the Jensen Effect over a range of smoothing parameters, which we can observe the feature of the curve more accurately. In addition, we proposed a hypothesis test to investigate the sign of the Jensen's inequality. Our estimates and test procedure worked both on simulated and real ecology data set.

Ecology data may need some nonlinear transformations of the response variables to

fix variance heteroscedasticity. We considered the logarithm of the response variables in Chapter 4, and defined the exponential single index model. We estimated the Jensen Effect based on the log-transformed response variables directly. We also developed a hypothesis test to see if the curve is concave or not. Our method obtained high statistical power for the simulated data set.

Apart from growth rates of a species, we also observed the number of offspring of a species, or individual survival. We extended our model to generalized single index model in Chapter 5, where the response variables are generated from an exponential family. We estimated the Jensen Effect by minimizing a penalized log-likelihood, and tested if the quantity is statistically significant different from 0.

APPENDIX A

CHAPTER 2 OF APPENDIX

A.1 Bernstein's Inequality

Bernstein's Inequality. Let $Y_{1n}, \cdots, Y_{nn}$ be independent random variables with means 0 and bounded ranges, that is $|Y_{in}| \leq c_n$. Write σ_{in}^2 for the variance of Y_{in}. Suppose $V_n \geq \sigma_{1n}^2 + \cdots + \sigma_{nn}^2$. Then for each $\eta_n > 0$,

$$\mathrm{P}\left(|Y_{1n} + \cdots + Y_{nn}| > \eta_n\right) \leq \exp\left[-\frac{\eta_n^2}{2\left(V_n + \frac{1}{3}c_n\eta_n\right)}\right]. \tag{A.1}$$

A.2 Lemmas for Theorem 6

Lemma 7. If $nh_n^6 \to \infty$, $nh_n^8 \to 0$ and $\frac{nh_n^{3+\frac{3}{m-1}}}{-\log h_n} \to \infty$, then we have

$$\mathrm{E}\left[\hat{g}''\left(\int X_j\hat{\beta}\right) - \bar{g}''\left(\int X_j\hat{\beta}\right)\right]^2 = O\left(\frac{1}{nh_n^4}\right). \tag{A.2}$$

Proof. Denote $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^\top$. The term could be expressed as

$$\begin{aligned}
&\mathrm{E}\left[\hat{g}''\left(\int X_j\hat{\beta}\right) - \bar{g}''\left(\int X_j\hat{\beta}\right)\right]^2 \\
&\quad = \mathrm{E}\left[S_2\left(\hat{\beta}; j\right)\boldsymbol{\epsilon}\right]^2 \\
&\quad = \mathrm{E}\left[\frac{2T_j^0\left(\hat{\beta}\right)\sum_{i=1}^n\left(\int X_i\hat{\beta} - \int X_j\hat{\beta}\right)^2 K\left(\frac{\int X_i\hat{\beta} - \int X_j\hat{\beta}}{h_n}\right)\epsilon_i - 2T_j^2\left(\hat{\beta}\right)\sum_{i=1}^n K\left(\frac{\int X_i\hat{\beta} - \int X_j\hat{\beta}}{h_n}\right)\epsilon_i}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right) - \left(T_j^2\left(\hat{\beta}\right)\right)^2}\right]^2 \\
&\quad \leq \mathrm{E}\left[\frac{2T_j^0\left(\hat{\beta}\right)\sum_{i=1}^n\left(\int X_i\hat{\beta} - \int X_j\hat{\beta}\right)^2 K\left(\frac{\int X_i\hat{\beta} - \int X_j\hat{\beta}}{h_n}\right)\epsilon_i}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right) - \left(T_j^2\left(\hat{\beta}\right)\right)^2}\right]^2 + \mathrm{E}\left[\frac{2T_j^2\left(\hat{\beta}\right)\sum_{i=1}^n K\left(\frac{\int X_i\hat{\beta} - \int X_j\hat{\beta}}{h_n}\right)\epsilon_i}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right) - \left(T_j^2\left(\hat{\beta}\right)\right)^2}\right]^2.
\end{aligned}$$

We bound the next two terms as:

$$
\begin{aligned}
&\mathrm{E}\left[\frac{T_j^0\left(\hat{\beta}\right)\sum_{i=1}^{n}\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2 K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)\epsilon_i}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}\right]^2 \\
&\quad=\mathrm{E}\left[\frac{1}{h_n^5}\cdot\frac{\frac{1}{nh_nf\left(u|\hat{\beta}\right)}T_j^0\left(\hat{\beta}\right)\cdot\frac{1}{n}\sum_{i=1}^{n}\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2 K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)\epsilon_i}{\frac{1}{nh_nf\left(u|\hat{\beta}\right)}T_j^0\left(\hat{\beta}\right)\cdot\frac{1}{nh_n^5f\left(u|\hat{\beta}\right)}T_j^4\left(\hat{\beta}\right)-\left(\frac{1}{nh_n^3f\left(u|\hat{\beta}\right)}T_j^2\left(\hat{\beta}\right)\right)^2}\right]^2 \\
&\quad=\left(\frac{1}{h_n^5}\cdot\frac{1}{\mu_4(K)-(\mu_2(K))^2}\cdot\frac{1}{f\left(u\left|\hat{\beta}\right.\right)}\right)^2\cdot\frac{1}{n}\left[\int(z-u)^2K\left(\frac{z-u}{h_n}\right)f\left(z\left|\hat{\beta}\right.\right)\mathrm{d}z\right]^2\cdot\mathrm{E}\left(\epsilon^2\right) \\
&\quad\doteq M_1\cdot\frac{1}{nh_n^{10}f^2\left(u\left|\hat{\beta}\right.\right)}\left[\int(h_nm)^2K(m)f\left(u+h_nm\left|\hat{\beta}\right.\right)h_n\mathrm{d}m\right]^2\cdot\mathrm{E}\left(\epsilon^2\right) \\
&\quad\sim M_1\cdot\frac{1}{nh_n^{10}f^2\left(u\left|\hat{\beta}\right.\right)}\left[h_n^3f\left(u\left|\hat{\beta}\right.\right)\int m^2K(m)\,\mathrm{d}m\right]^2\cdot\mathrm{E}\left(\epsilon^2\right) \\
&\quad\sim\frac{1}{nh_n^4},
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{E}\left[\frac{T_j^2\left(\hat{\beta}\right)\sum_{i=1}^{n}K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)\epsilon_i}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}\right]^2 \\
&\quad=\mathrm{E}\left[\frac{1}{h_n^3}\cdot\frac{\frac{1}{nh_n^3(u)}T_j^2\left(\hat{\beta}\right)\cdot\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)\epsilon_i}{\frac{1}{nh_nf\left(u|\hat{\beta}\right)}T_j^0\left(\hat{\beta}\right)\cdot\frac{1}{nh_n^5f\left(u|\hat{\beta}\right)}T_j^4\left(\hat{\beta}\right)-\left(\frac{1}{nh_n^3f(u)}T_j^2\left(\hat{\beta}\right)\right)^2}\right]^2 \\
&\quad=\left(\frac{1}{h_n^3}\cdot\frac{\mu_2(K)}{\mu_4(K)-(\mu_2(K))^2}\cdot\frac{1}{f\left(u\left|\hat{\beta}\right.\right)}\right)^2\cdot\frac{1}{n}\left[\int K\left(\frac{z-u}{h_n}\right)f\left(z\left|\hat{\beta}\right.\right)\mathrm{d}z\right]^2\cdot\mathrm{E}\left(\epsilon^2\right) \\
&\quad\doteq M_2\cdot\frac{1}{nh_n^6f^2\left(u\left|\hat{\beta}\right.\right)}\left[\int K(m)f\left(u+h_nm\left|\hat{\beta}\right.\right)h_n\mathrm{d}m\right]^2\cdot\mathrm{E}\left(\epsilon^2\right) \\
&\quad\sim M_2\cdot\frac{1}{nh_n^6f^2\left(u\left|\hat{\beta}\right.\right)}\left[h_nf\left(u\left|\hat{\beta}\right.\right)\int K(m)\,\mathrm{d}m\right]^2\cdot\mathrm{E}\left(\epsilon^2\right) \\
&\quad\sim\frac{1}{nh_n^4}.
\end{aligned}
$$

where M_1, M_2 are constants and $u=\int X_j\hat{\beta}$.

Therefore, we have

$$\mathrm{E}\left[\hat{g}''\left(\int X_j\hat{\beta}\right)-\bar{g}''\left(\int X_j\hat{\beta}\right)\right]^2=O\left(\frac{1}{nh_n^4}\right).$$

□

Lemma 8. If $nh_n^6\to\infty$, $nh_n^8\to 0$ and $\frac{nh_n^{3+\frac{3}{m-1}}}{-\log h_n}\to\infty$, then we have

$$\mathrm{E}\left[\bar{g}''\left(\int X_j\hat{\beta}\right)-g''\left(\int X_j\hat{\beta}\right)\right]^2=O\left(h_n^4+\frac{1}{nh_n^4}\right). \tag{A.3}$$

Proof. Calculate

$$\begin{aligned}
&\bar{g}''\left(\int X_j\hat{\beta}\right)-g''\left(\int X_j\hat{\beta}\right)\\
&=S_2\left(\hat{\beta};j\right)\boldsymbol{g}-g''\left(\int X_j\hat{\beta}\right)\\
&=\frac{2T_j^0\left(\hat{\beta}\right)\sum_{i=1}^n\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)g\left(\int X_i\beta\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}\\
&\quad-\frac{-2T_j^2\left(\hat{\beta}\right)\sum_{i=1}^n K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)g\left(\int X_i\beta\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}-g''\left(\int X_j\hat{\beta}\right).
\end{aligned}$$

Write

$$g\left(\int X_i\beta\right)=g\left(\int X_i\hat{\beta}\right)+\left(g\left(\int X_i\beta\right)-g\left(\int X_i\hat{\beta}\right)\right),$$

for $i=1,\cdots,n$, we decompose it into two terms A and B, where A is

$$\begin{aligned}
&\frac{2T_j^0\left(\hat{\beta}\right)\sum_{i=1}^n\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)g\left(\int X_i\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}\\
&\quad-\frac{2T_j^2\left(\hat{\beta}\right)\sum_{i=1}^n K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)g\left(\int X_i\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}-g''\left(\int X_j\hat{\beta}\right),
\end{aligned}$$

and B is

$$\frac{\sum_{i=1}^{n}\left[2T_j^0\left(\hat{\beta}\right)\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2 K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)-2T_j^2\left(\hat{\beta}\right)K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)\right]\left(g\left(\int X_i\beta\right)-g\left(\int X_i\hat{\beta}\right)\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}.$$

Denote $K_{ij}\doteq K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)$, the term A can be calculated as

$$\begin{aligned}
&\frac{2T_j^0\left(\hat{\beta}\right)\sum_{i=1}^{n}\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2 K_{ij}g\left(\int X_i\hat{\beta}\right)-2T_j^2\left(\hat{\beta}\right)\sum_{i=1}^{n}K_{ij}g\left(\int X_i\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}-g''\left(\int X_j\hat{\beta}\right)\\
&=\frac{2T_j^0\left(\hat{\beta}\right)\sum_{i=1}^{n}\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2 K_{ij}g\left(\int X_i\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}-\frac{2T_j^2\left(\hat{\beta}\right)\sum_{i=1}^{n}K_{ij}g\left(\int X_i\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}-g''\left(\int X_j\hat{\beta}\right),
\end{aligned}$$

where

$$\begin{aligned}
&\frac{2T_j^0\left(\hat{\beta}\right)\sum_{i=1}^{n}\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2 K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)g\left(\int X_i\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}\\
&=\frac{2T_j^0\left(\hat{\beta}\right)\sum_{i=1}^{n}\left(\int X_i\hat{\beta}-u\right)^2 K_{ij}\left[g(u)+g'(u)\left(\int X_i\hat{\beta}-u\right)+\frac{1}{2}g''(u)\left(\int X_i\hat{\beta}-u\right)^2+O\left(\int X_i\hat{\beta}-u\right)^4\right]}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}\\
&=\frac{2T_j^0\left(\hat{\beta}\right)\left(T_j^2\left(\hat{\beta}\right)g(u)+T_j^3\left(\hat{\beta}\right)g'(u)+\frac{1}{2}T_j^4\left(\hat{\beta}\right)g''(u)\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}+O\left(h_n^2\right)\\
&=\frac{2T_j^0\left(\hat{\beta}\right)T_j^2\left(\hat{\beta}\right)g\left(\int X_j\hat{\beta}\right)+T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)g''\left(\int X_j\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}+O\left(h_n^2\right),
\end{aligned}$$

and

$$\frac{2T_j^2\left(\hat{\beta}\right)\sum_{i=1}^{n}K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)g\left(\int X_i\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}$$

$$=\frac{2T_j^2\left(\hat{\beta}\right)\sum_{i=1}^{n}K\left(\frac{\int X_i\hat{\beta}-u}{h_n}\right)\left[g\left(u\right)+g'\left(u\right)\left(\int X_i\hat{\beta}-u\right)+\frac{1}{2}g''\left(u\right)\left(\int X_i\hat{\beta}-u\right)^2+O\left(\int X_i\hat{\beta}-u\right)^4\right]}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}$$

$$=\frac{2T_j^2\left(\hat{\beta}\right)\left(T_j^0\left(\hat{\beta}\right)g\left(u\right)+T_j^1\left(\hat{\beta}\right)g'\left(u\right)+\frac{1}{2}T_j^2\left(\hat{\beta}\right)g''\left(u\right)\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}+O\left(h_n^2\right)$$

$$=\frac{2T_j^2\left(\hat{\beta}\right)T_j^0\left(\hat{\beta}\right)g\left(\int X_j\hat{\beta}\right)+\left(T_j^2\left(\hat{\beta}\right)\right)^2g''\left(\int X_j\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}+O\left(h_n^2\right).$$

Therefore, the term A is

$$\frac{2T_j^0\left(\hat{\beta}\right)T_j^2\left(\hat{\beta}\right)g\left(\int X_j\hat{\beta}\right)+T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)g''\left(\int X_j\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}$$

$$-\frac{2T_j^2\left(\hat{\beta}\right)T_j^0\left(\hat{\beta}\right)g\left(\int X_j\hat{\beta}\right)+\left(T_j^2\left(\hat{\beta}\right)\right)^2g''\left(\int X_j\hat{\beta}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}-g''\left(\int X_j\hat{\beta}\right)+O\left(h_n^2\right)$$

$$=g''\left(\int X_j\beta\right)-g''\left(\int X_j\beta\right)+O\left(h_n^2\right)$$

$$=O\left(h_n^2\right).$$

The term B is bounded by

$$|B|\leq\left|\frac{\sum_{i=1}^{n}\left[2T_j^0\left(\hat{\beta}\right)\left(\int X_i\hat{\beta}-\int X_j\hat{\beta}\right)^2K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)-2T_j^2\left(\hat{\beta}\right)K\left(\frac{\int X_i\hat{\beta}-\int X_j\hat{\beta}}{h_n}\right)\right]\left(E_1\frac{1}{\sqrt{n}}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}\right|$$

$$=\frac{\left(2T_j^0\left(\hat{\beta}\right)T_j^2\left(\hat{\beta}\right)-2T_j^0\left(\hat{\beta}\right)T_j^2\left(\hat{\beta}\right)\right)\left(E_1\frac{1}{\sqrt{n}}\right)}{T_j^0\left(\hat{\beta}\right)T_j^4\left(\hat{\beta}\right)-\left(T_j^2\left(\hat{\beta}\right)\right)^2}$$

$$\sim\frac{1}{\sqrt{n}h_n^2},$$

where E_1 is a constant.

Combining the terms A and B, we have

$$\mathrm{E}\left[\bar{g}''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\hat{\beta}\right)\right]^2 = O\left(h_n^4 + \frac{1}{nh_n^4}\right).$$

□

Lemma 9. If $nh_n^6 \to \infty$, $nh_n^8 \to 0$ and $\frac{nh_n^{3+\frac{3}{m-1}}}{-\log h_n} \to \infty$, then we have

$$\mathrm{E}\left[g''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\beta\right)\right]^2 = O\left(\frac{1}{n}\right). \tag{A.4}$$

Proof. By Theorem 5, we have $\left(\int X_i\hat{\beta} - \int X_i\beta\right) = O\left(\frac{1}{\sqrt{n}}\right)$. Since g'' satisfies the Lipschitz condition, we can get

$$\mathrm{E}\left[g''\left(\int X_j\hat{\beta}\right) - g''\left(\int X_j\beta\right)\right]^2 \leq K_3\left(\int X_i\hat{\beta} - \int X_i\beta\right)^2 = O\left(\frac{1}{n}\right),$$

where K_3 is a constant. □

A.3 10-fold CV Results

The simulation results using the 10-fold cross-validation is:

Table A.1: Simulation results by 10-fold CV using rescaled bandwidth $h\|\boldsymbol{c}\|$ and constraint $\hat{\beta}$.

		g1			g2			g3		
Initial	**n**	**RSE**	**RASE**	**RASE2**	**RSE**	**RASE**	**RASE2**	**RSE**	**RASE**	**RASE2**
True	100	0.1921	0.0868	1.2106	0.2052	0.0904	4.5719	0.1982	0.0690	0.7954
	1000	0.0680	0.0290	0.4551	0.0620	0.0271	0.4016	0.0774	0.0243	0.2762
Linear	100	1.1383	0.3294	1.4303	1.3370	0.4245	1.3388	0.6995	0.1331	0.2692
	1000	1.9015	0.2232	0.5060	1.0419	0.3185	1.3652	0.5297	0.0852	0.0786
Equal	100	0.8486	0.2789	1.0731	0.8385	0.2509	1.4728	0.8349	0.1836	0.3349
	1000	0.8442	0.2977	1.0181	0.8500	0.2755	1.5396	0.8447	0.1827	0.1045
Random	100	1.2998	0.1339	0.5191	1.2086	0.1979	0.7160	1.4296	0.0875	0.2362
	1000	1.5607	0.1288	0.4209	1.3540	0.1687	0.4418	1.7622	0.0798	0.0744

For comparison purpose, we estimate the curvature with or without rescaling the bandwidth, where the results are in Table A.2

Table A.2: Simulation results of random starting value with rescaled and original bandwidths selected by 10-fold cross-validation

	g1		g2		g3	
n	**original**	**rescaled**	**original**	**rescaled**	**original**	**rescaled**
100	1.2710	0.5191	2.2005	0.7160	0.2768	0.2362
1000	1.0145	0.4209	2.2812	0.4418	0.0752	0.0744

Table A.3: Simulation results of true starting value with rescaled and original bandwidths selected by 10-fold cross-validation

	g1		g2		g3	
n	**original**	**rescaled**	**original**	**rescaled**	**original**	**rescaled**
100	5.9302	1.9266	33.811	7.9015	4.5630	0.9791
1000	2.4004	1.0660	2.2494	1.3416	1.2444	0.3558

A.4 True Starting Value

In Table A.3 and A.4, we compare the RASE2 results of true starting value with rescaled and original bandwidths.

Table A.4: Simulation results of true starting value with rescaled and original bandwidths selected by GCV

	g1		g2		g3	
n	**original**	**rescaled**	**original**	**rescaled**	**original**	**rescaled**
100	51.397	7.3786	38.007	6.5712	50.816	5.9508
1000	9.0220	1.5416	8.9590	1.8170	7.8245	1.1493

A.5 Cross-validation Values

The cross-validation values for both GCV and 10-fold cross-validation are in Table A.5.

Table A.5: Cross-validation Values

		g1		g2		g3	
Initial	**n**	**GCV**	**10-fold**	**GCV**	**10-fold**	**GCV**	**10-fold**
True	100	0.0398	3.8686	0.0399	3.3823	0.0377	1.4970
	1000	0.0401	0.9468	0.0402	0.7882	0.0400	0.4781
Linear	100	0.1559	5.2259	0.2184	4.7373	0.0498	1.6148
	1000	0.0793	2.0295	0.1272	2.2896	0.0439	0.7265
Equal	100	0.1401	4.7640	0.1046	3.9670	0.0795	1.8530
	1000	0.1358	2.6925	0.1318	2.1995	0.0751	1.1081
Random	100	0.0611	4.2920	0.0957	3.9025	0.0459	1.6237
	1000	0.0591	2.4463	0.0850	2.2702	0.0468	1.0243

APPENDIX B

CHAPTER 3 OF APPENDIX

B.1 Plots of Estimated Curvature

This section provides plots provide examples of the accuracy of estimating a second derivative in a functional single index model. In each we plot the true and estimated link function and its second derivative. We also plot our estimated functional single index values versus their truth and the distortion that produces for the estimated second derivative.

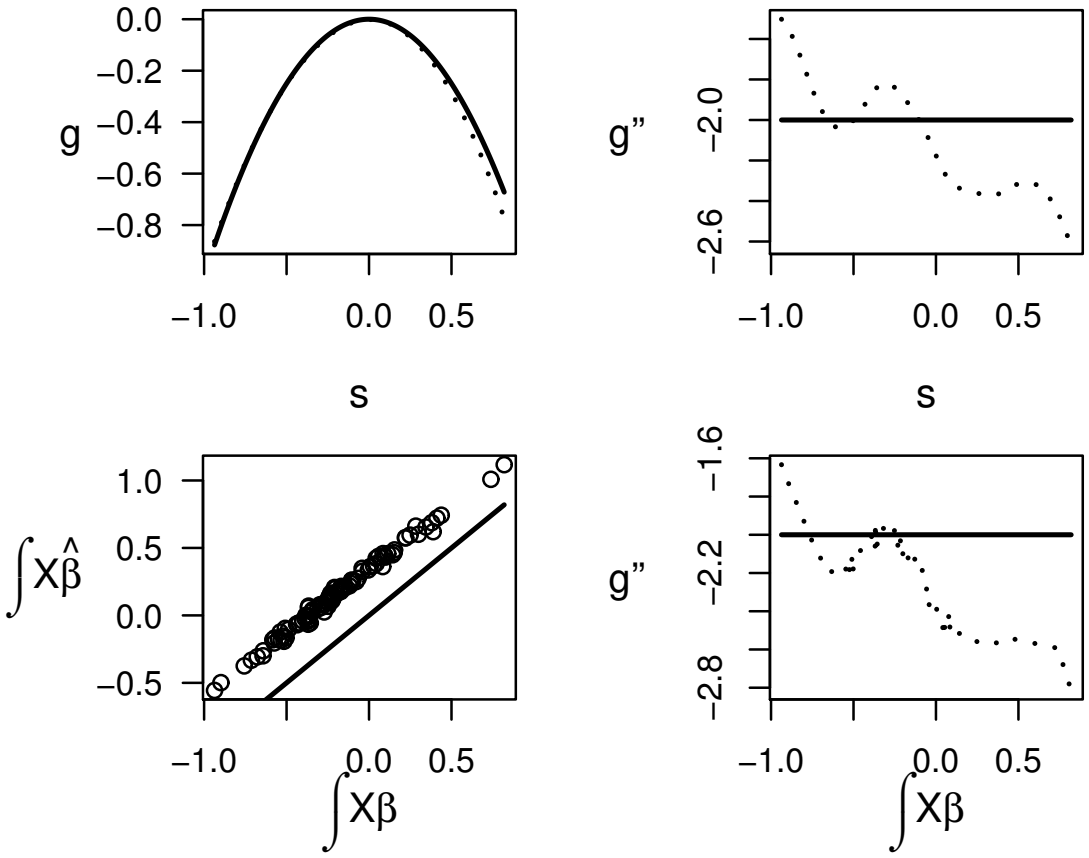

Figure B.1: The link function is $g(s) = -s^2$. The top-left and right panel are the plots of g and g" over 1000 equally-spaced grid points, while the lower and upper bound are the minimum and maximum of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$. The bottom-right panel is the plot of g" over the true $\int X(t)\beta(t)\,\mathrm{d}t$. The black line is the true curve, while the red line is the estimated curve. The bottom-left panel is the plot of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$ versus $\int X(t)\beta(t)\,\mathrm{d}t$, while the red line is $y = x$.

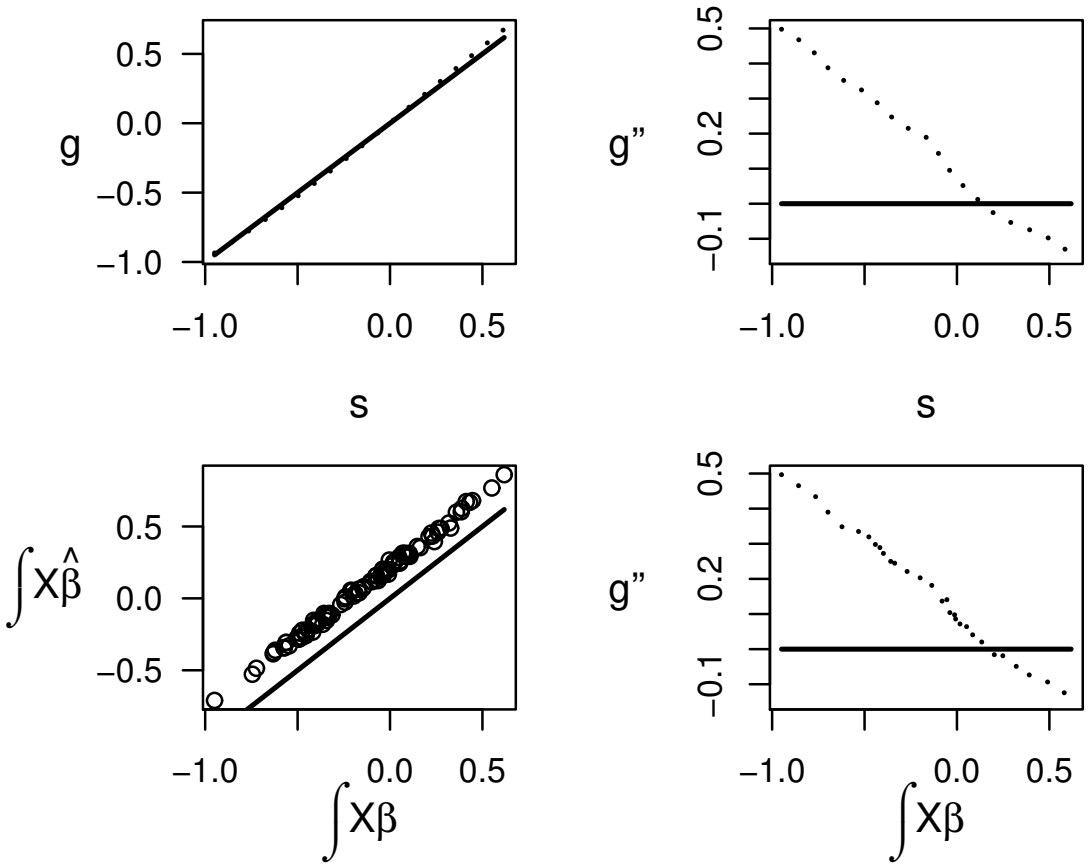

Figure B.2: The link function is $g(s) = s$. The top-left and right panel are the plots of g and g" over 1000 equally-spaced grid points, while the lower and upper bound are the minimum and maximum of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$. The bottom-right panel is the plot of g" over the true $\int X(t)\beta(t)\,\mathrm{d}t$. The black line is the true curve, while the red line is the estimated curve. The bottom-left panel is the plot of $\int X(t)\hat{\beta}(t)\,\mathrm{d}t$ versus $\int X(t)\beta(t)\,\mathrm{d}t$, while the red line is $y = x$.

B.2 Diagnostic Plots for the Jensen Effect: Single Index Model

These plots give example δ functions using a single index model and the corresponding t functions for links $g(s) = -s^2$ and $g(s) = s$ respectively.

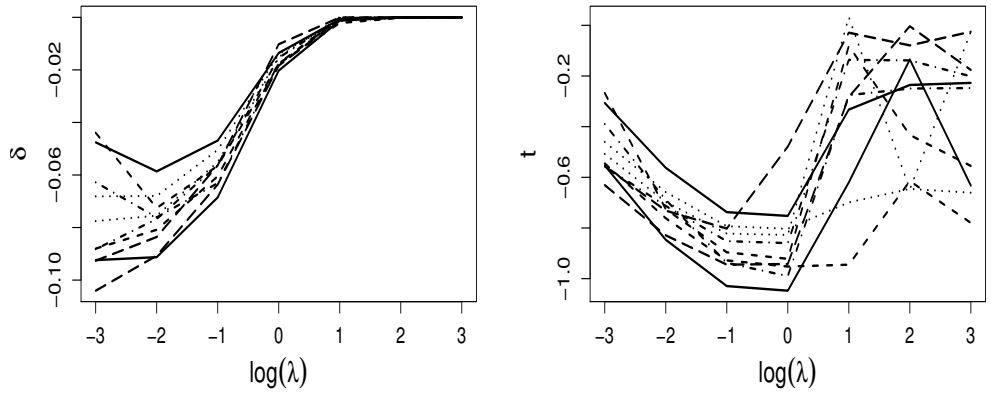

Figure B.3: Left: a sample of δ_λ as a function of λ in a single index model with link function $g(s) = -s^2$. Right: the corresponding t_λ functions.

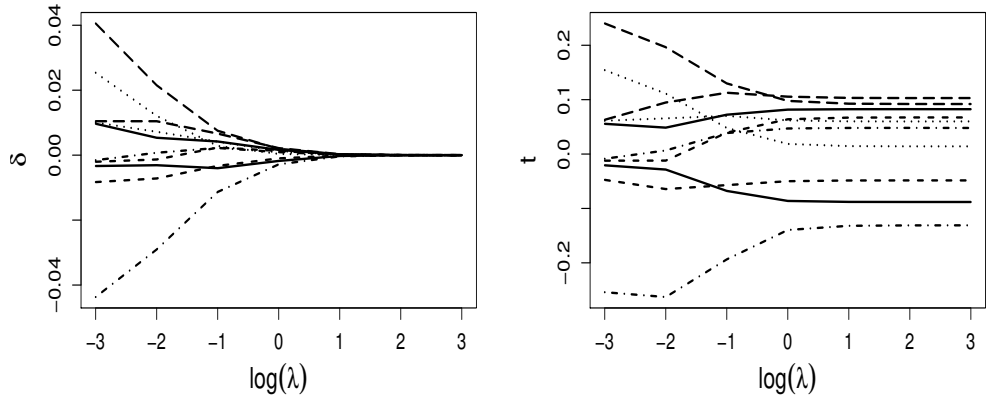

Figure B.4: Left: a sample of δ_λ as a function of λ in a single index model with link function $g(s) = s$. Right: the corresponding t_λ functions.

B.3 Diagnostic Plots for the Functional Single Index Model

These plots give example δ functions using a single index model and the corresponding t functions for links $g(s) = -s^2$ and $g(s) = s$ respectively.

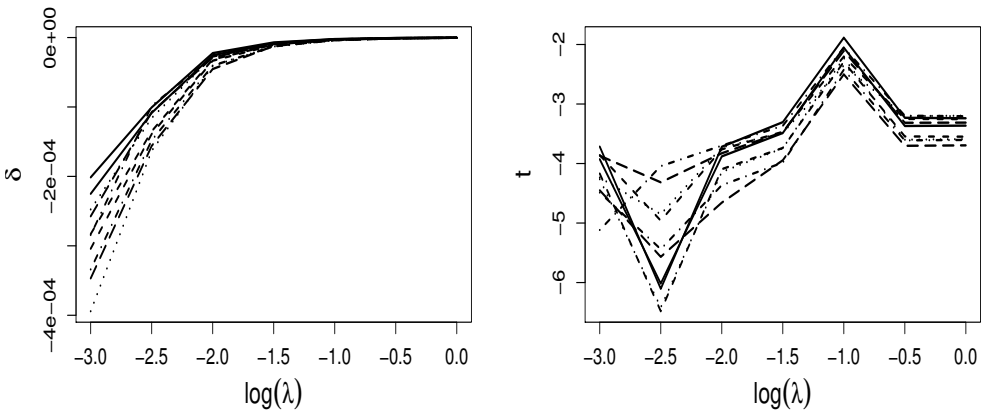

Figure B.5: Left: a sample of δ_λ as a function of λ in a functional single index model with link function $g(s) = -s^2$. Right: the corresponding t_λ functions.

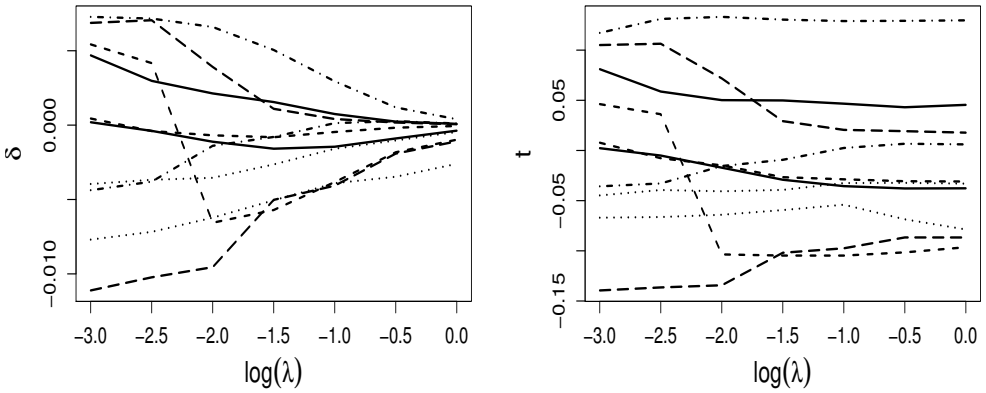

Figure B.6: Left: a sample of δ_λ as a function of λ in a functional single index model with link function $g(s) = s$. Right: the corresponding t_λ functions.

B.4 Plots for the copepod data

Plots 12 through 20 provide descriptive plots for each species from the copepod study. We provide a plot of δ versus $\log(\lambda)$ with the value selected by GCV in red. We also provide plots of $\hat{\beta}$, $\hat{g}$ and $\hat{g}''$ at the value of λ selected by GCV along with approximate point-wise confidence intervals.

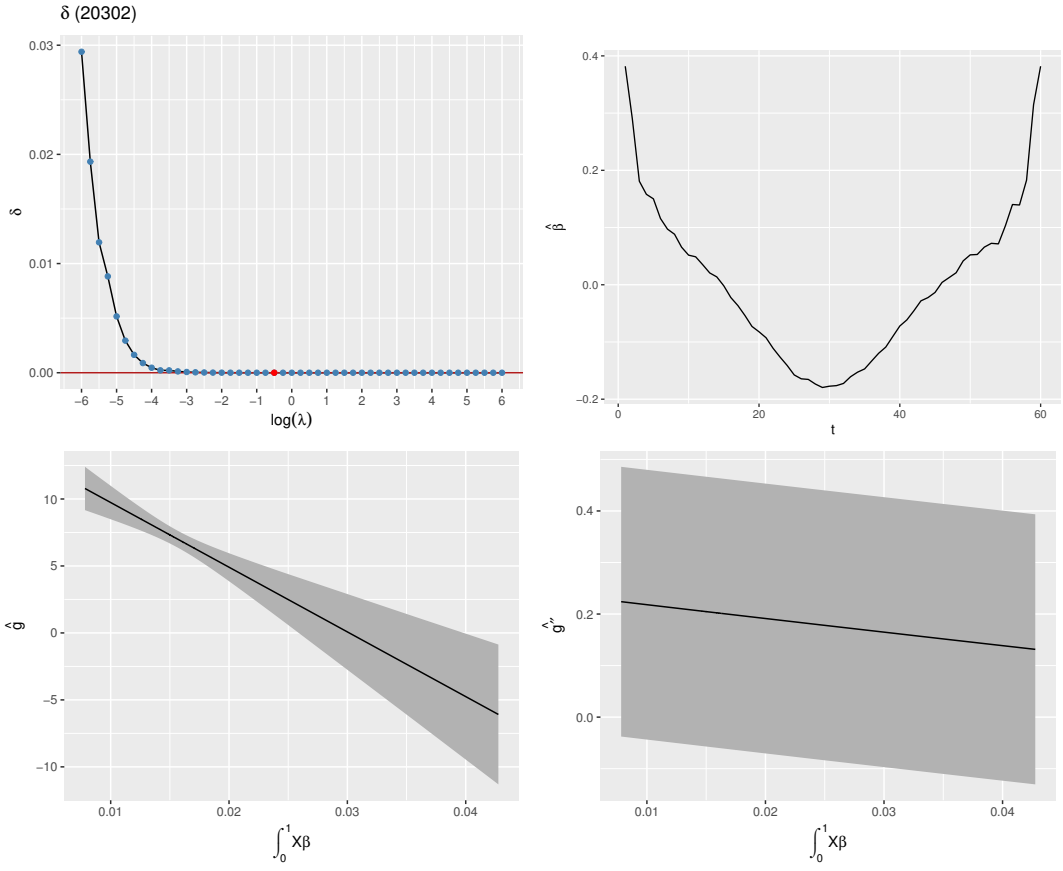

Figure B.7: Plots for *Diacyclops Thomasi*.

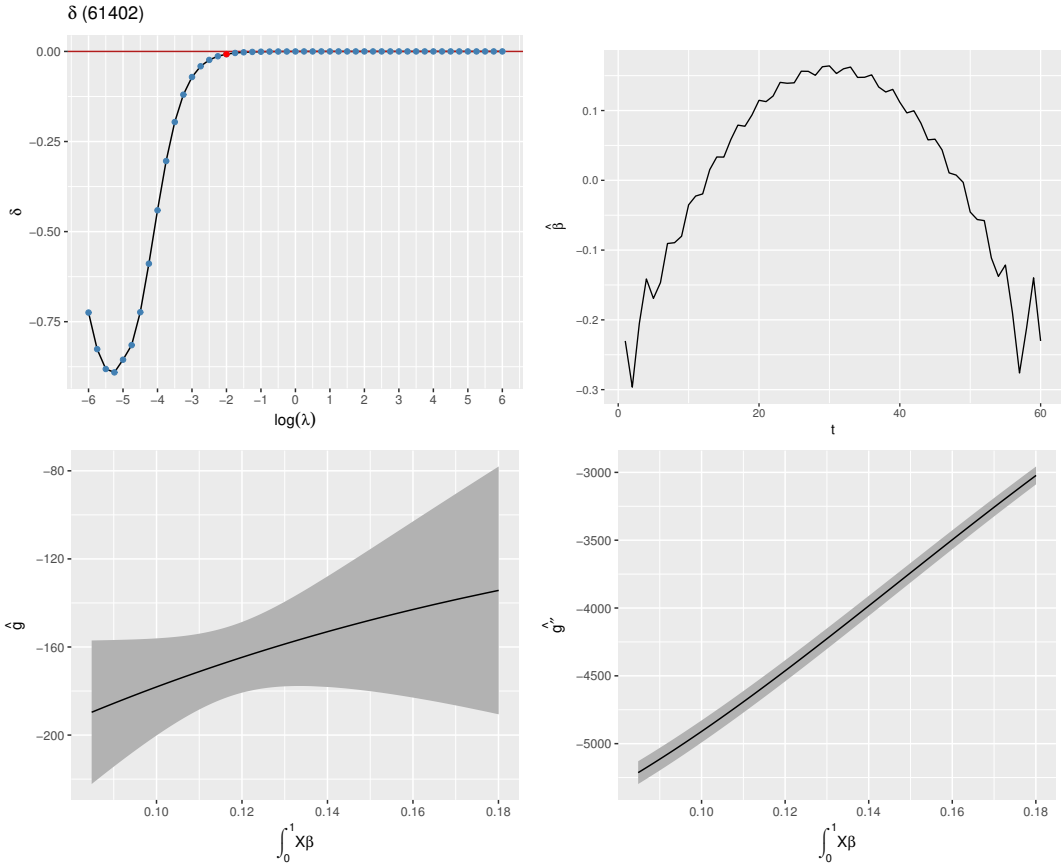

Figure B.8: Plots for *Filinia Terminalis*.

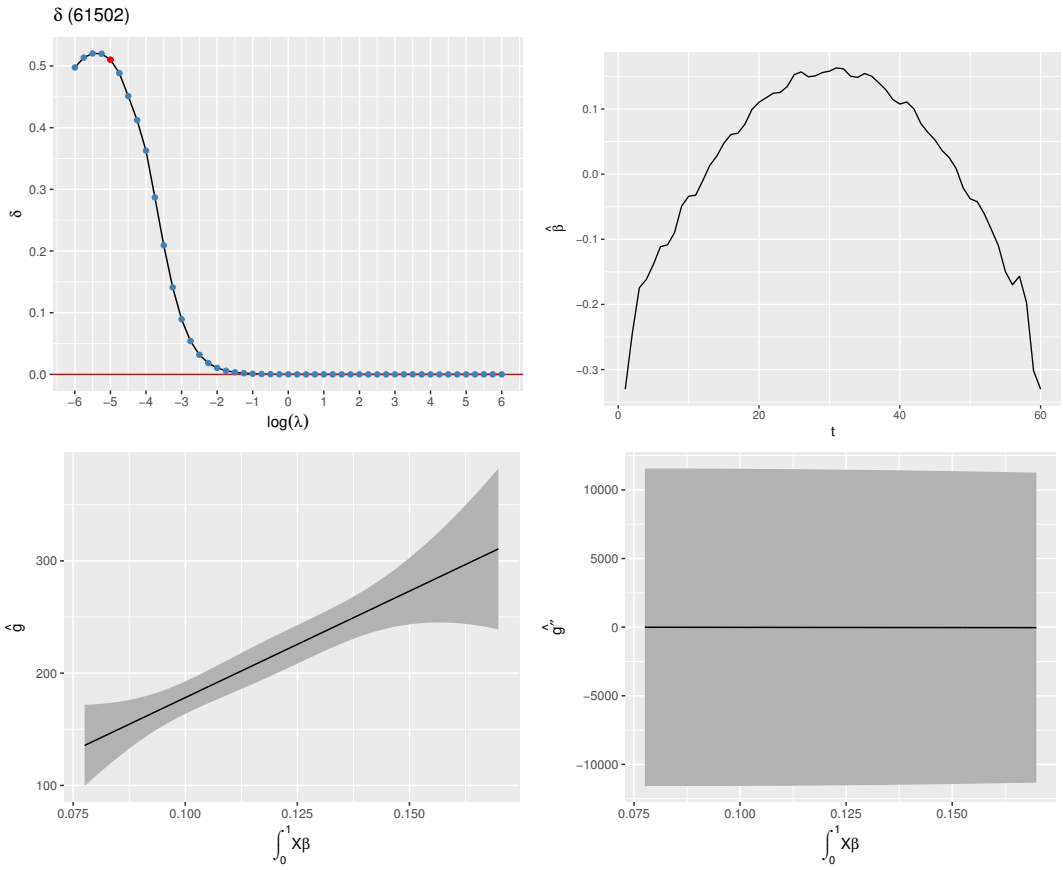

Figure B.9: Plots for *Gastropus Stylifer*.

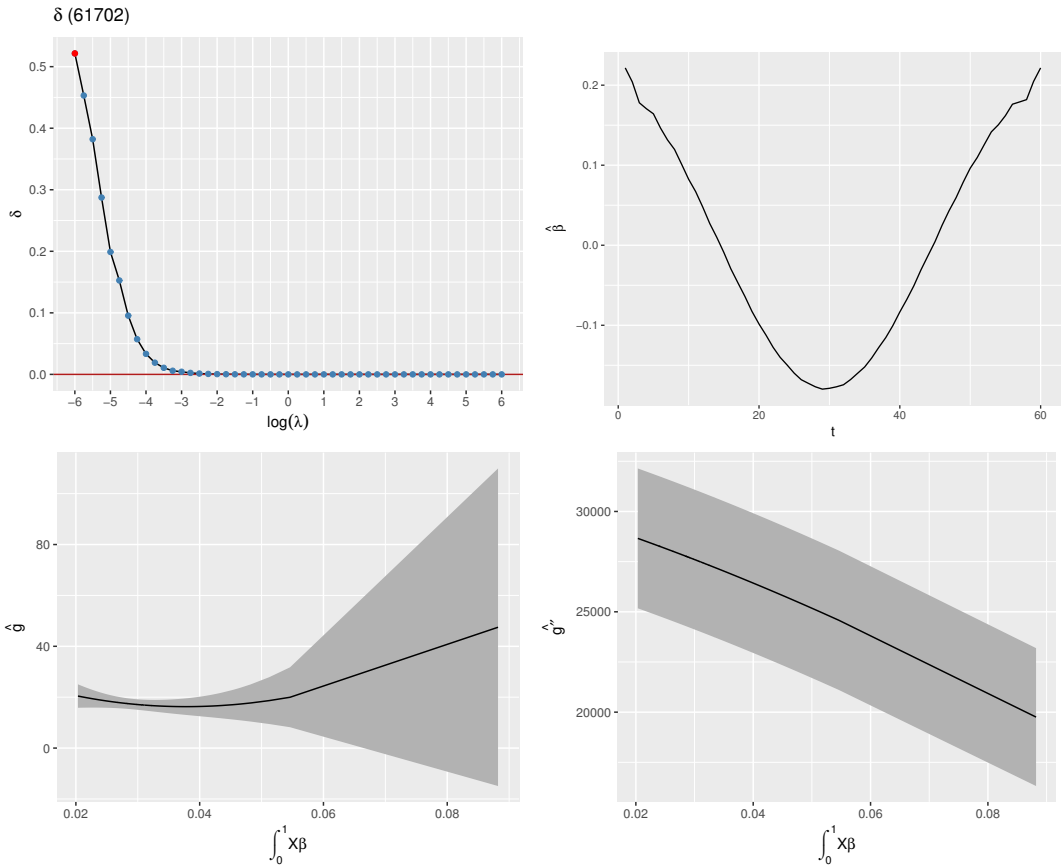

Figure B.10: Plots for *Kellicottia Longispina*.

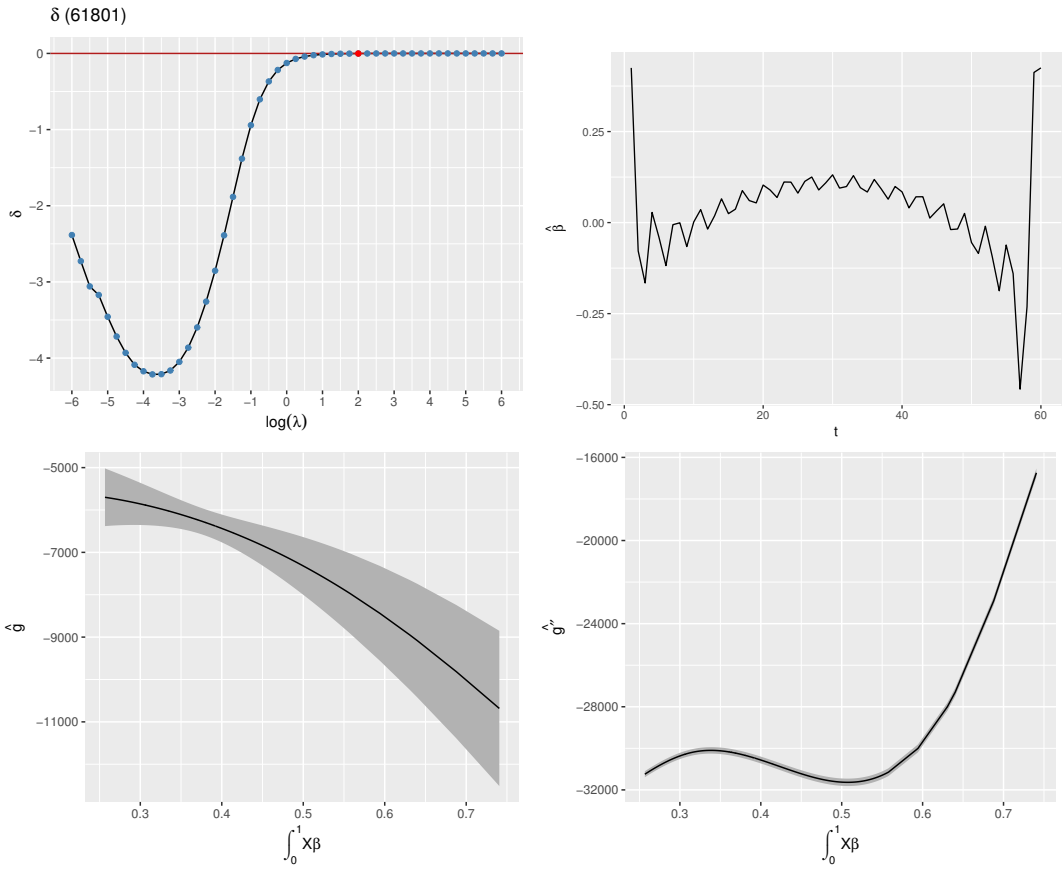

Figure B.11: Plot for *Keratella Cochlearis*.

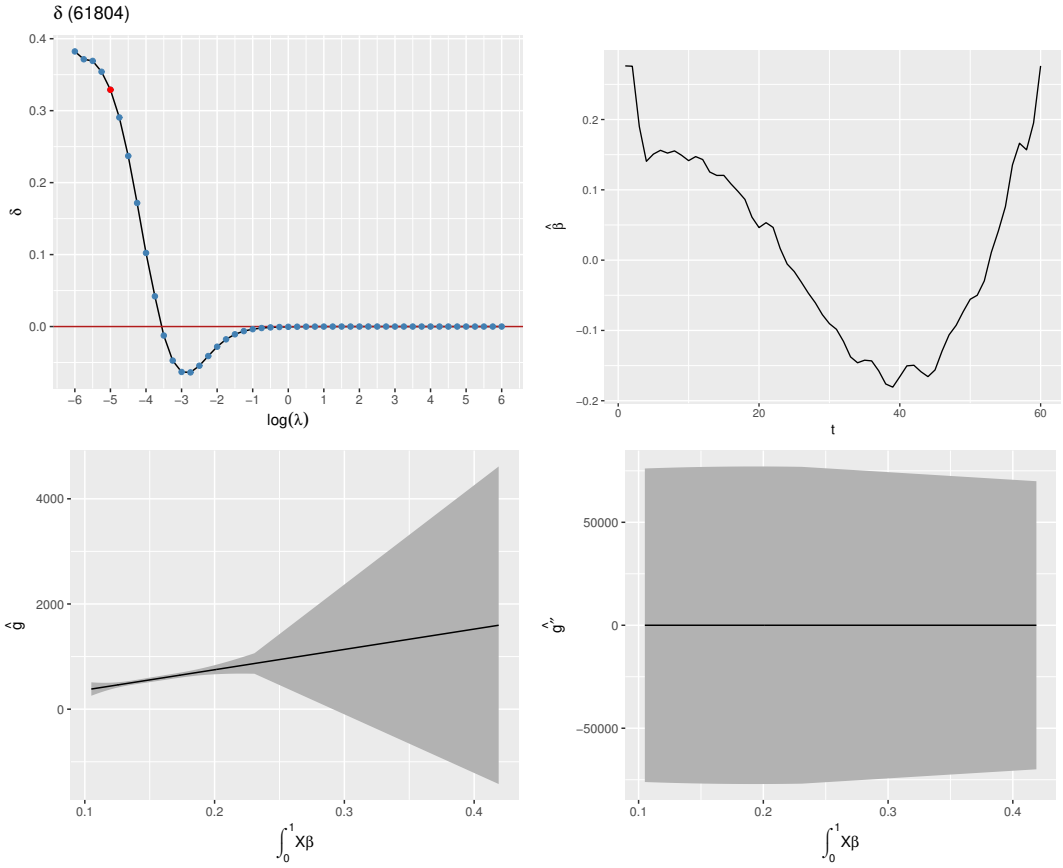

Figure B.12: Plot for *Keratella Earlinae*.

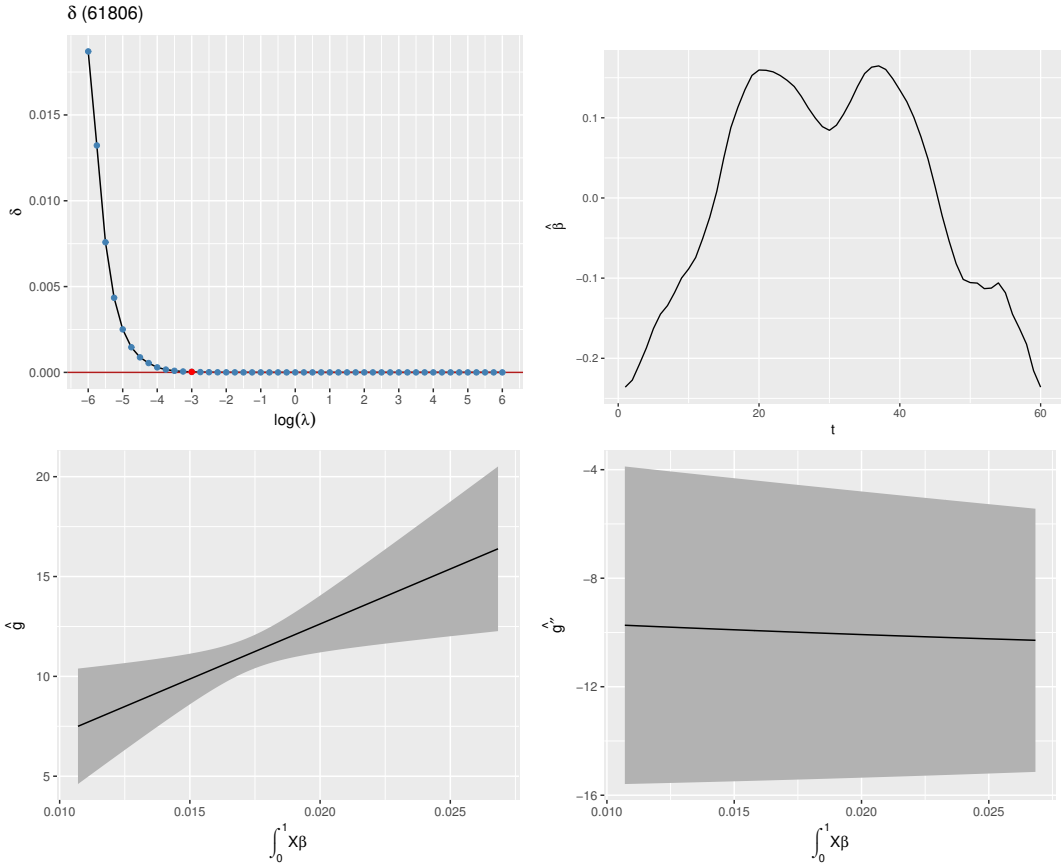

Figure B.13: Plot for *Keratella Quadrata*.

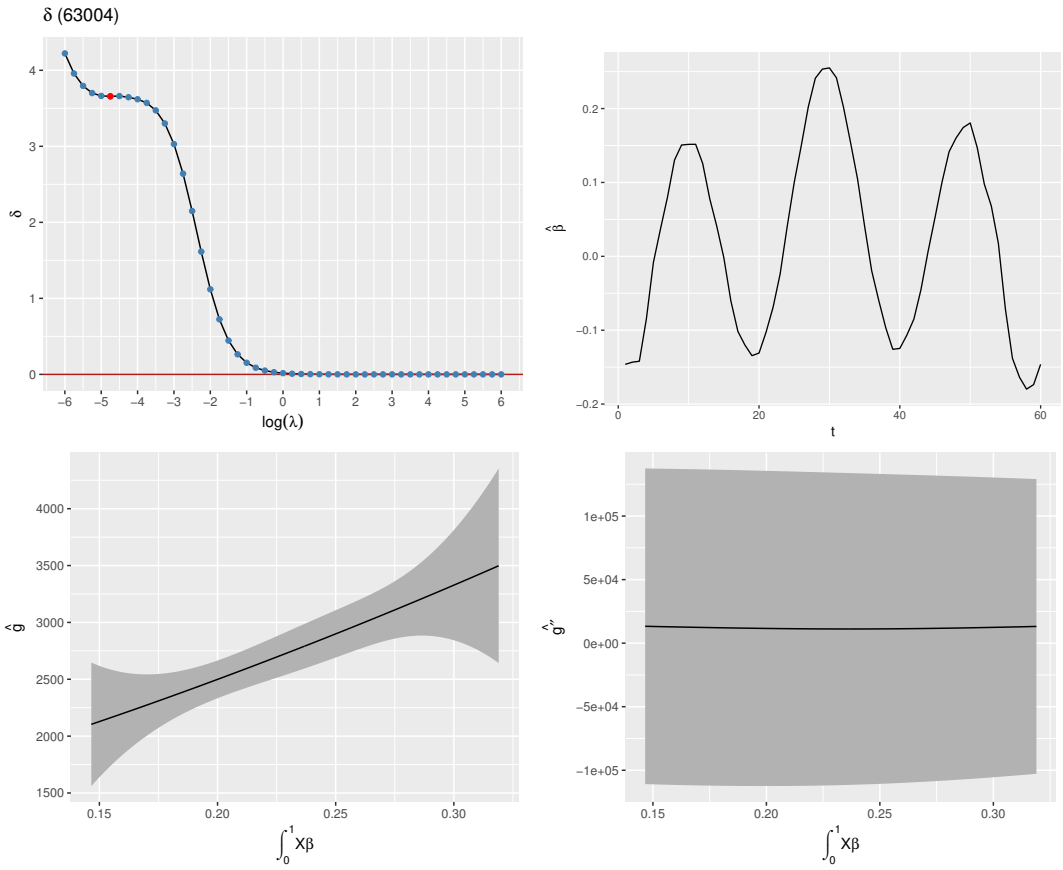

Figure B.14: Plot for *Polyarthra Remata*.

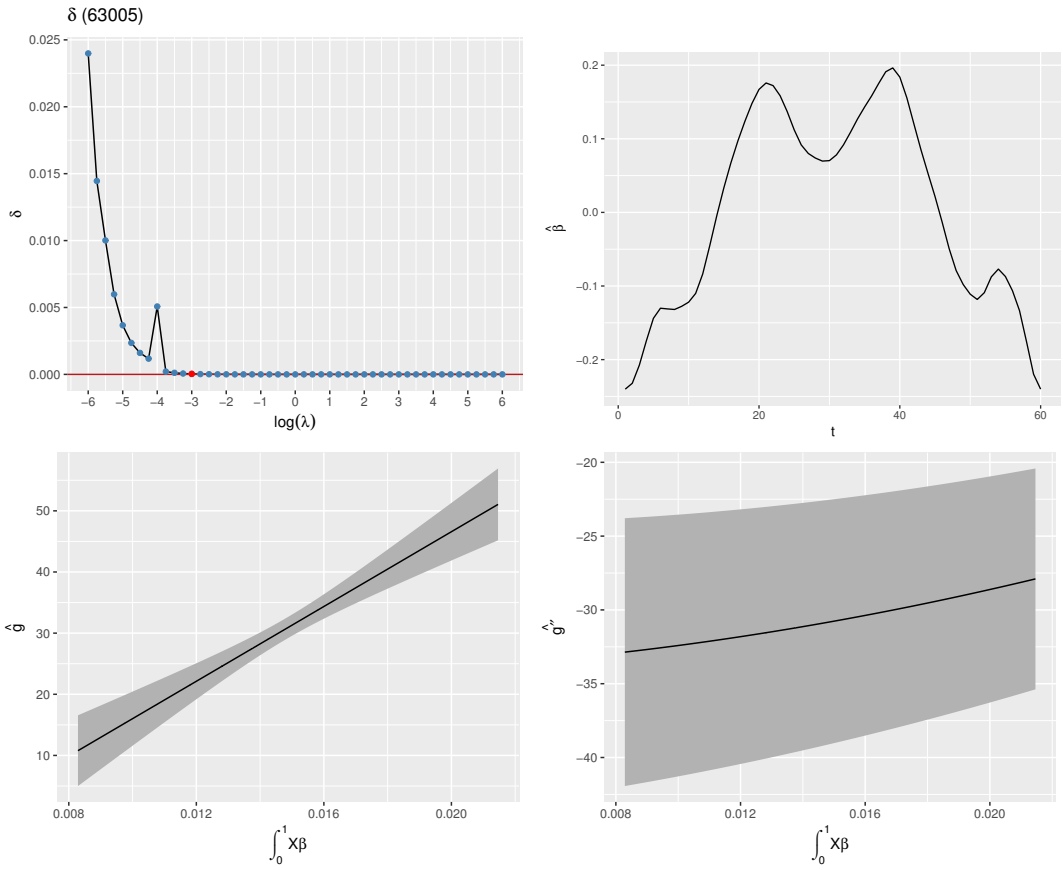

Figure B.15: Plot for *Polyarthra Vulgaris*.

B.5 GCV Calculations

Here we provide an approximate smoothing matrix from which to calculate GCV for the functional single index model. We derive the approximate smoothing matrix using the Taylor expansion. Writing

$$\boldsymbol{c} = \hat{\boldsymbol{c}} + \boldsymbol{c}^*, \qquad \boldsymbol{d} = \hat{\boldsymbol{d}} + \boldsymbol{d}^*, \tag{B.1}$$

where $\boldsymbol{c}$ and $\boldsymbol{d}$ are the true coefficient vectors, we have

$$\begin{aligned} Y_i =& \boldsymbol{\phi}^\top \left[\int X_i(t)\, \boldsymbol{\psi}^\top(t)\, (\hat{\boldsymbol{c}} + \boldsymbol{c}^*)\, \mathrm{d}t \right] \left(\hat{\boldsymbol{d}} + \boldsymbol{d}^* \right) + \epsilon_i \\ \approx& \boldsymbol{\phi}^\top \left[\int X_i(t)\, \boldsymbol{\psi}^\top(t)\, \mathrm{d}t \cdot \hat{\boldsymbol{c}} \right] \hat{\boldsymbol{d}} + \boldsymbol{\phi}^\top \left[\int X_i(t)\, \boldsymbol{\psi}^\top(t)\, \mathrm{d}t \cdot \hat{\boldsymbol{c}} \right] \boldsymbol{d}^* \\ &+ \left\{ (\boldsymbol{\phi}')^\top \left[\int X_i(t)\, \boldsymbol{\psi}^\top(t)\, \mathrm{d}t \cdot \hat{\boldsymbol{c}} \right] \hat{\boldsymbol{d}} \cdot \int X_i(t)\, \boldsymbol{\psi}^\top(t)\, \mathrm{d}t \right\} \boldsymbol{c}^* + \epsilon_i. \end{aligned} \tag{B.2}$$

Assume that the $(n \times p)$-dimensional matrix $\mathbb{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)^\top$, where $\boldsymbol{X}_i = \left(X_i(t_1), \cdots, X_i\left(t_p\right) \right)^\top$. The $(p \times K_2)$-dimensional Fourier basis matrix is $\Psi = \left(\boldsymbol{\psi}^\top(t_1), \cdots, \boldsymbol{\psi}^\top\left(t_p\right) \right)^\top$. We can calculate

$$\begin{pmatrix} \int X_1(t)\, \boldsymbol{\psi}^\top(t)\, \mathrm{d}t \\ \vdots \\ \int X_n(t)\, \boldsymbol{\psi}^\top(t)\, \mathrm{d}t \end{pmatrix} \approx \frac{1}{p} \mathbb{X}\Psi. \tag{B.3}$$

Denote

$$\begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix} \doteq \begin{pmatrix} \int X_1(t)\, \boldsymbol{\psi}^\top(t)\, \mathrm{d}t \cdot \boldsymbol{c} \\ \vdots \\ \int X_n(t)\, \boldsymbol{\psi}^\top(t)\, \mathrm{d}t \cdot \boldsymbol{c} \end{pmatrix}, \tag{B.4}$$

then the $(n \times K_1)$-dimensional B-spline basis matrices are $\Phi = (\boldsymbol{\phi}(s_1), \cdots, \boldsymbol{\phi}(s_n))^\top$, and $\Phi^{(k)} = \left(\boldsymbol{\phi}^{(k)}(s_1), \cdots, \boldsymbol{\phi}^{(k)}(s_n) \right)^\top$, where $\cdot^{(k)}$ denotes the k^{th} order of derivative.

Denote the response vector as $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^\top$. Define

$$\boldsymbol{Y}^* = \boldsymbol{Y} - \Phi \boldsymbol{d}, \tag{B.5}$$

and

$$\mathbb{Z} = [\mathbb{Z}_1, \mathbb{Z}_2] = \left[\Phi, \mathrm{diag}\left(\Phi^{(1)} \boldsymbol{d} \right) \cdot \left(\frac{1}{p} \mathbb{X}\Psi \right) \right]. \tag{B.6}$$

The linear model with the response vector $\boldsymbol{Y}^*$, predictors $\boldsymbol{c}^*$ and $\boldsymbol{d}^*$ is

$$\boldsymbol{Y}^* = \mathbb{Z}_1 \boldsymbol{d}^* + \mathbb{Z}_2 \boldsymbol{c}^* + \boldsymbol{\epsilon}. \tag{B.7}$$

The $(K_1 \times K_1)$-dimensional penalized matrix of the function g is

$$\mathbb{P}_g = \left(\Phi^{(2)}\right)^\top \left(\Phi^{(2)}\right), \tag{B.8}$$

and the $(K_2 \times K_2)$-dimensional penalized matrix of the coefficient function β is

$$\mathbb{P}_\beta = \left[\mathrm{diag}\left(\Phi^{(3)}\boldsymbol{d}\right) \cdot \left(\frac{1}{p}\mathbb{X}\Psi\right)\right]^\top \left[\mathrm{diag}\left(\Phi^{(3)}\boldsymbol{d}\right) \cdot \left(\frac{1}{p}\mathbb{X}\Psi\right)\right]. \tag{B.9}$$

The penalized sum of squares are

$$\mathrm{PLS} \doteq \left\| \boldsymbol{Y}^* - \mathbb{Z} \begin{pmatrix} \boldsymbol{d}^* \\ \boldsymbol{c}^* \end{pmatrix} \right\|^2 + \lambda_g \left(\boldsymbol{d}^*\right)^\top \mathbb{P}_g \boldsymbol{d}^* + \lambda_\beta \left(\boldsymbol{c}^*\right)^\top \mathbb{P}_\beta \boldsymbol{c}^*. \tag{B.10}$$

A smoothing matrix can be written as

$$\mathbb{S}\left(\lambda_g, \lambda_\beta\right) = \mathbb{Z}\left[\mathbb{Z}^\top\mathbb{Z} + \begin{pmatrix} \lambda_g\mathbb{P}_g & \\ & \lambda_\beta\mathbb{P}_\beta \end{pmatrix}\right]^{-1} \mathbb{Z}^\top. \tag{B.11}$$

We substitute this smoothing matrix into the GCV formula, and find a pair of $\left(\lambda_g, \lambda_\beta\right)$ minimizing the GCV.

APPENDIX C

CHAPTER 4 OF APPENDIX

C.1 Delta Method

In Casella and Berger (2002), the Delta Method is

Theorem 1. Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \rightarrow \mathrm{N}\left(0, \sigma^2\right)$ in distribution. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \rightarrow \mathrm{N}\left(0, \sigma^2 [g'(\theta)]^2\right) \quad \text{in distribution.} \tag{C.1}$$

C.2 Diagnostic Plots for the Jensen Effect: Exponential Model

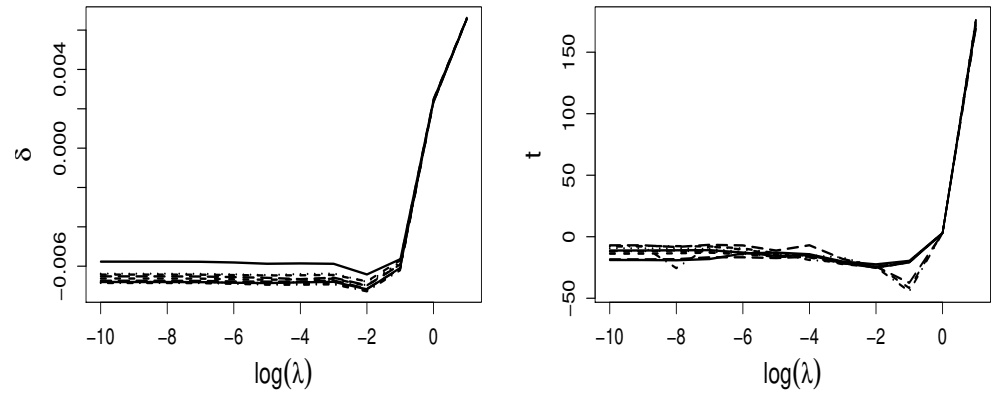

Figure C.1: This figure illustrates an example of concave function in the exponential single index model, where the link function is $g^*(s) = \sqrt{s}$. Left: a sample of the Jensen Effect values δ_λ as a function of smoothing parameters λ. Right: the corresponding t-statistics t_λ functions.

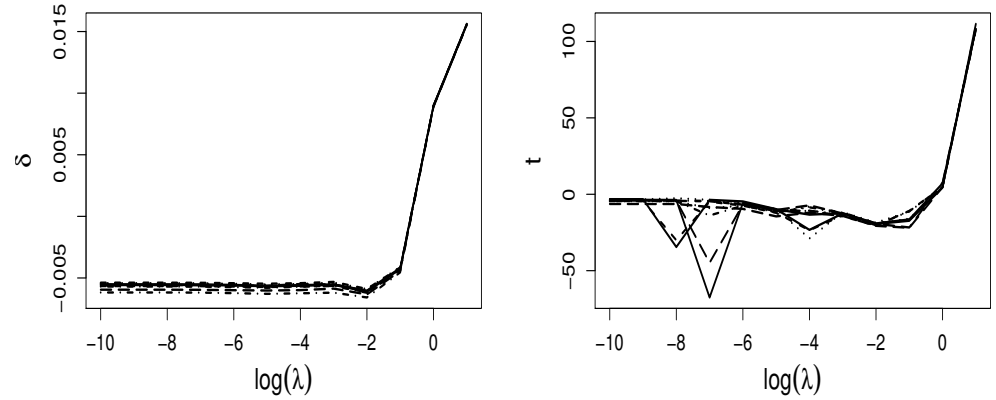

Figure C.2: This figure illustrates another example of concave function in the exponential single index model, where the link function is $g^*(s) = \sin(s)$. Left: a sample of the Jensen Effect values δ_λ as a function of smoothing parameters λ. Right: the corresponding t-statistics t_λ functions.

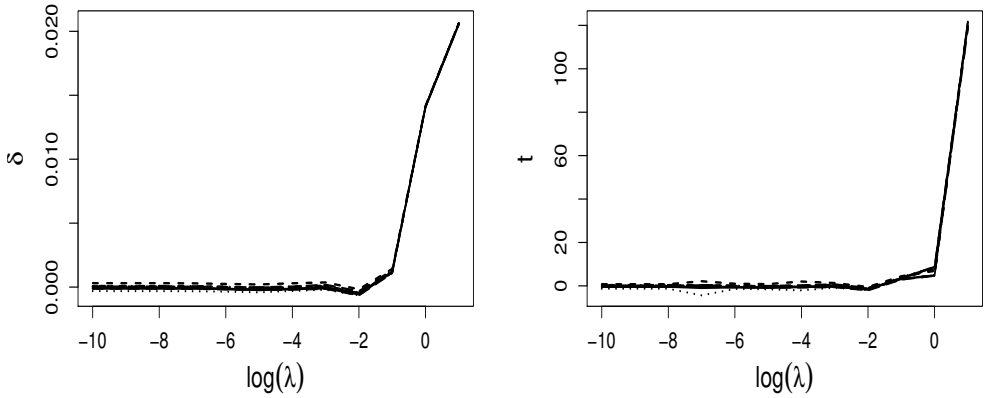

Figure C.3: This figure illustrates a linear function in the exponential single index model, where the link function is $g^*(s) = s$. Left: a sample of the Jensen Effect values δ_λ as a function of smoothing parameters λ. Right: the corresponding t-statistics t_λ functions.

APPENDIX D

CHAPTER 5 OF APPENDIX

D.1 Diagnostic Plots for the Jensen Effect: Poisson Model

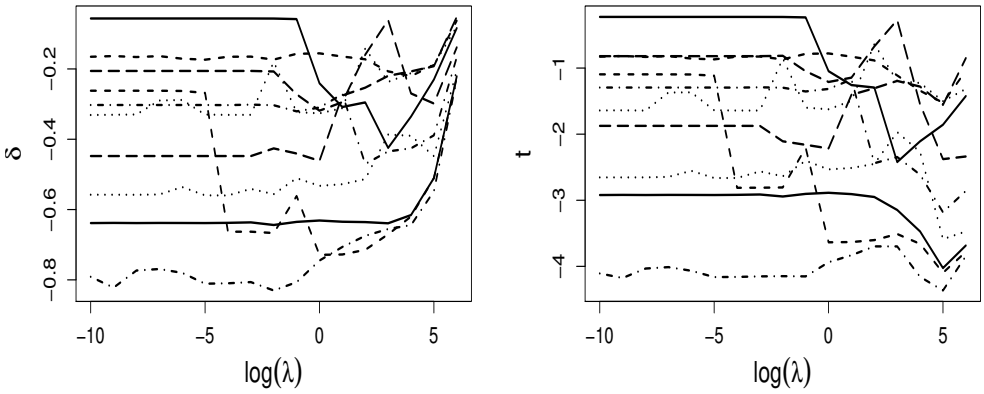

Figure D.1: This figure illustrates an example of concave function in the Poisson single index model, where the link function is $g^*(s) = \frac{30}{1+\exp\left(-\frac{s}{8}\right)}$. Left: a sample of the Jensen Effect values δ_λ as a function of smoothing parameters λ. Right: the corresponding t-statistics t_λ functions.

D.2 Plots for USSES Plants data in Poisson Single Index Model

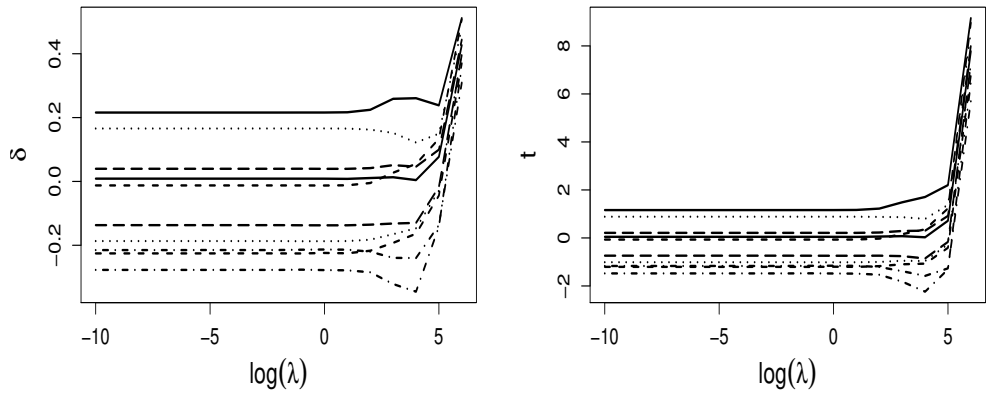

Figure D.2: This figure illustrates an example of linear function in the Poisson single index model, where the link function is $g^*(s) = s$. Left: a sample of the Jensen Effect values δ_λ as a function of smoothing parameters λ. Right: the corresponding t-statistics t_λ functions.

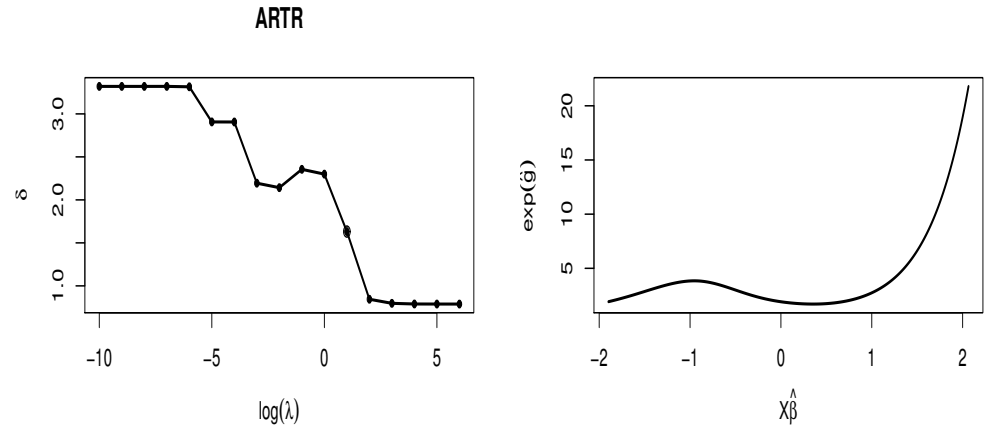

Figure D.3: This figure illustrates the Jensen Effect for the plant *Artemisia tripartita* (ARTR) in USSES data set, where we formulate the data in a Poisson single index model. Left: the Jensen Effect values δ plotted against the values of λ. Right: the curve of the exponential link function $\exp(\hat{g})$, where λ is selected by GCV.

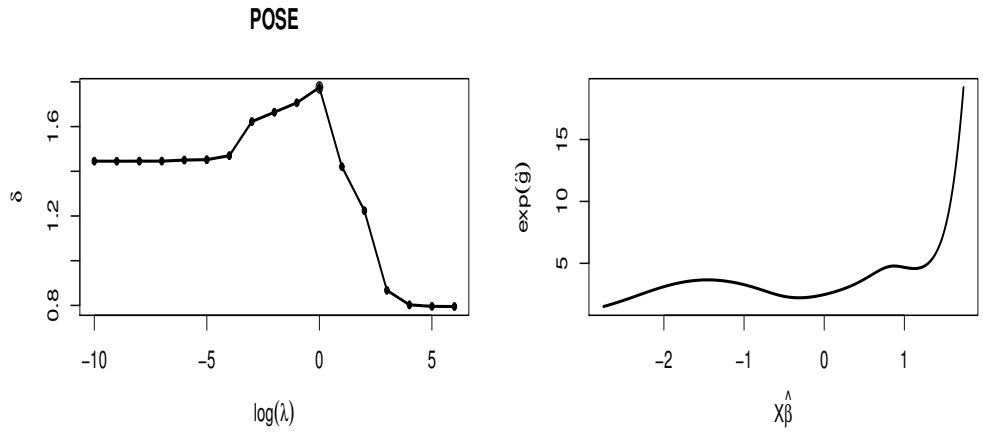

Figure D.4: This figure illustrates the Jensen Effect for the plant *Poa secunda* (POSE) in USSES data set, where we formulate the data in a Poisson single index model. Left: the Jensen Effect values δ plotted against the values of λ. Right: the curve of the exponential link function $\exp(\hat{g})$, where λ is selected by GCV.

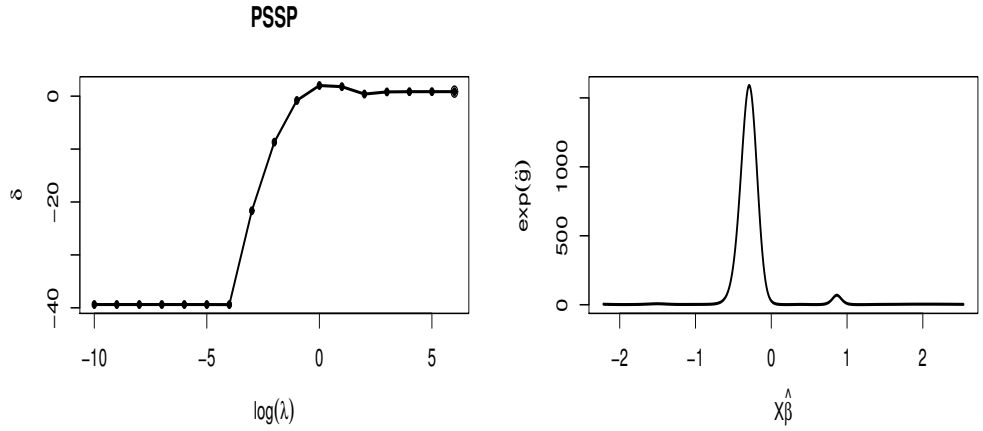

Figure D.5: This figure illustrates the Jensen Effect for the plant *Pseudoroegneria spicata* (PSSP) in USSES data set, where we formulate the data in a Poisson single index model. Left: the Jensen Effect values δ plotted against the values of λ. Right: the curve of the exponential link function $\exp(\hat{g})$, where λ is corresponding to the minimum of δ.

BIBLIOGRAPHY

Adler, P. B., S. P. Ellner, and J. M. Levine (2010). Coexistence of perennial plants: an embarrassment of niches. *Ecology letters 13*(8), 1019–1029.

Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.

Chaudhuri, P. and J. S. Marron (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association 94*(447), 807–823.

Chen, D., P. Hall, H.-G. Müller, et al. (2011). Single and multiple index functional regression models with nonparametric link. *The Annals of Statistics 39*(3), 1720–1747.

Chesson, P. (1994). Multispecies competition in variable environments. *Theoretical population biology 45*(3), 227–276.

Chesson, P. (2000a). General theory of competitive coexistence in spatially-varying environments. *Theoretical population biology 58*(3), 211–237.

Chesson, P. (2000b). Mechanisms of maintenance of species diversity. *Annual review of Ecology and Systematics 31*(1), 343–366.

Chesson, P. L. and R. R. Warner (1981). Environmental variability promotes coexistence in lottery competitive systems. *The American Naturalist 117*(6), 923–943.

Clark, J. S., D. M. Bell, M. H. Hersh, and L. Nichols (2011). Climate change vulnerability of forest biodiversity: climate and competition tracking of demographic rates. *Global Change Biology 17*(5), 1834–1849.

Cohen, D. (1966). Optimizing reproduction in a randomly varying environment. *Journal of theoretical biology 12*(1), 119–129.

Dahlgren, J. P. and J. Ehrlén (2011). Incorporating environmental change over succession in an integral projection model of population dynamics of a forest herb. *Oikos 120*(8), 1183–1190.

Drake, J. M. (2005). Population effects of increased climate variation. *Proceedings of the Royal Society of London B: Biological Sciences 272*(1574), 1823–1827.

Ellner, S. (1987). Alternate plant life history strategies and coexistence in randomly varying environments. In *Theory and models in vegetation science*, pp. 199–208. Springer.

Ellner, S. P., R. E. Snyder, and P. B. Adler (2016). How to quantify the temporal storage effect using simulations instead of math. *Ecology letters 19*(11), 1333–1342.

Escabias, M., A. M. Aguilera, and M. J. Valderrama (2007). Functional PLS logit regression model. *Computational Statistics & Data Analysis 51*(10), 4891–4902.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics New York.

Hardle, W., P. Hall, H. Ichimura, et al. (1993). Optimal smoothing in Single-Index Models. *The annals of Statistics 21*(1), 157–178.

Hristache, M., A. Juditsky, and V. Spokoiny (2001). Direct estimation of the index coefficient in a Single-Index Models. *Annals of Statistics 29*(3), 595–623.

Hutchinson, G. E. (1961). The paradox of the plankton. *The American Naturalist 95*(882), 137–145.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of Single-Index Models. *Journal of Econometrics 58*(1-2), 71–120.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(3), 411–432.

Koons, D. N., C. J. E. Metcalf, and S. Tuljapurkar (2008). Evolution of delayed reproduction in uncertain environments: a life-history perspective. *The American Naturalist 172*(6), 797–805.

Koons, D. N., S. Pavard, A. Baudisch, J. E. Metcalf, et al. (2009). Is life-history buffering or lability adaptive in stochastic environments? *Oikos 118*(7), 972–980.

Lewontin, R. C. and D. Cohen (1969). On population growth in a randomly varying environment. *Proceedings of the National Academy of Sciences 62*(4), 1056–1060.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association 86*(414), 316–327.

Ma, S. (2016). Estimation and inference in Functional Single-Index Models. *Annals of the Institute of Statistical Mathematics 68*(1), 181–208.

Marron, J. and J. T. Zhang (2005). SiZer for smoothing splines. *Computational Statistics 20*(3), 481–502.

Müller, H.-G. and U. Stadtmüller (2005). Generalized functional linear models. *Annals of Statistics*, 774–805.

Ramsay, J. O. (2006). *Functional data analysis*. Wiley Online Library.

Ramsay, J. O., G. Hooker, and S. Graves (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Cambridge university press.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica: Journal of the Econometric Society*, 1461–1481.

Teller, B. J., P. B. Adler, C. B. Edwards, G. Hooker, and S. P. Ellner (2016). Linking demography with drivers: climate and competition. *Methods in Ecology and Evolution 7*(2), 171–183.

Tuljapurkar, S. (2013). *Population dynamics in variable environments*, Volume 85. Springer Science & Business Media.

Vasseur, D. A. and K. S. McCann (2007). *The Impact of Environmental Variability on Ecological Systems*. Springer Netherlands.

Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(2), 413–428.

Wood, S. N. (2006). Generalized Additive Models: An introduction with R.

Ye, Z. and G. Hooker (2018). Local quadratic estimation of the curvature in a Functional Single Index Model. *arXiv preprint arXiv:1803.09321*.

Ye, Z., G. Hooker, and S. Ellner (2019). The Jensen Effect and Functional Single Index Models: Estimating the ecological implications of nonlinear reaction norms. *arXiv preprint arXiv:1901.01864*.