AN ABSTRACT OF THE DISSERTATION OF

Lu Wang for the degree of Doctor of Philosophy in Statistics presented on May 24, 2016.

Title: Nonparametric Estimation of Additive Models with Shape Constraints

Abstract approved: _____

Lan Xue

Monotone additive models are useful in estimating productivity curves or analyzing disease risk where the predictors are known to have monotonic effects on the response. Existing literature mainly focuses on univariate monotone smoothing. Available methods for the estimation of monotone additive models are either difficult to interpret or have no asymptotic guarantees. In the first part of this dissertation, we propose a one-step backfitted constrained polynomial spline method for the estimation of monotone additive models. In our proposed method, we obtain monotone estimators by imposing a set of linear constraints on the spline coefficients for each additive component. In the second part of the dissertation, we extend the constrained polynomial spline method to estimate the production frontier that is used to quantify the maximum production output in econometrics. The estimation of frontier functions is more challenging since it is the boundary of the support rather than the mean output function to be estimated. Here, we develop a two-step shape constrained polynomial spline method for the frontier estima-

tion. The first step is to capture the shape of frontier while the second step is to estimate the location of frontier. Both proposed methods in this dissertation give smooth estimators with the desired shape constraints (monotonicity or/and concavity). They are easily implementable and computationally efficient by taking advantage of linear programming. Most importantly, our methods are applicable for multi-dimensions where some existing methods fail to work. For the assessment of properties of the proposed estimators, asymptotic theory is also developed. In addition, the simulation studies and application of our methods to analyze Norwegian Farm data in both parts suggest that our proposed methods have better numerical performance than the existing methods, especially when the data has outliers.

# Nonparametric Estimation of Additive Models with Shape Constraints

by

Lu Wang

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 24, 2016
Commencement June 2016

Doctor of Philosophy dissertation of Lu Wang presented on May 24, 2016.

APPROVED:

---

Major Professor, representing Statistics

---

Chair of the Department of Statistics

---

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

---

Lu Wang, Author

# ACKNOWLEDGEMENTS

I would like to take this opportunity to extend my gratitude and appreciation to everyone who made this Ph.D. dissertation possible.

First, I would like to express my sincere gratitude to my advisor, Dr. Lan Xue, for her unreserved guidance and help throughout my doctoral studies. With endless patience and encouragement, she has taught me how to become a good statistician and researcher step-by-step. She has always been there to give invaluable advice and unconditional support when I encounter difficulties. I am thankful for all her contributions that have made my graduate journey fruitful and enjoyable. As a passionate scientist and excellent educator, she has also energized me to continue in academic research in the future.

I would also like to thank my committee members, Dr. Virginia Lesser, Dr. Lisa Madsen, and Dr. Yuan Jiang, for their time and insightful comments through my research process. I appreciate Dr. Virginia Lesser for providing me with support and advice during my job searching process. I want to thank Dr. Lisa Madsen and Dr. Yuan Jiang for showing me the beautiful world of statistics. I am fortunate to have been in their classrooms because they have tremendously influenced my academic career.

I also thank all professors in the Department of Statistics for their devoted instruction. They are role models who have shaped my mind and even my future.

I am truly grateful to Dr. Gudbrand Lien from Lillehammer University College for sharing the Norwegian Farm data with me.

I want to thank all my friends for their endless encouragement and trust. Their timely help will always be remembered.

Last but not least, my deepest gratitude goes to my parents for their unflagging support throughout my life and studies. Thank you Mom and Dad for giving me strength to chase my dreams and inspiring me on my way. Without you, I would not have become who I am. I know that your unconditional love will be with me all my life.

# TABLE OF CONTENTS

# LIST OF FIGURES

LIST OF FIGURES (Continued)

# LIST OF TABLES

## LIST OF TABLES (Continued)

# Nonparametric Estimation of Additive Models with Shape Constraints

## 1 Introduction

Non- and semi-parametric models are useful in modeling nonlinear dynamics in data sets where linear models fail to take it into account. Compared with linear models, non- and semi-parametric models are more flexible since they relax the strong assumption on linear relationships between predictors and the response variable. Due to their flexibility, non- and semi-parametric models have been widely used in many scientific areas, such as genetics (Boni et al. 2007), chemistry (Hnizdo et al. 2007), medical imaging (Sled et al. 1998), environmental (Van Bergeijk et al. 1992) and social sciences (Leys and Schumann 2010).

Nonparametric methods are also used in production frontier analysis. In economics, it is often of interest to investigate the relationship between input and output variables. A production frontier quantifies the maximum output that can be obtained with limited inputs. It provides a useful reference in production efficiency analysis. Many nonparametric frontier methods have been proposed in recent years including Barr et al. (1994), Cazals et al. (2002), Aragon et al. (2005) and Martins-Filho and Yao (2007).

A variety of non- and semi-parametric models have been developed in the existing literature. Among them, additive models have gained increasing popularity in recent years due to their flexibility and interpretability. Without a linear restriction, additive models describe the relationship between each predictor and response via an unknown

nonparametric function. The contribution of each predictor is additive that enables easy interpretation of the fitted results.

In recent decades, a vast amount of literature has been published on the estimation of additive models. Various estimation methods have been proposed, including kernel backfitting (Hastie and Tibshirani 1990, Opsomer and Ruppert 1997, Horowitz and Mammen 2004), polynomial spline smoothing (Stone 1985, Huang 1998, Ma and Racine 2013), marginal integration (Linton and Nielsen 1995, Linton and Härdle 1996), kernel backfitting and projection (Mammen, et al. 1999), spline backfitted kernel smoothing (Wang and Yang 2007, Liu and Yang 2010, Ma and Yang 2011, Liu, et al. 2013), and Bayesian backfitting (Hastie and Tibshirani 2000). Among these estimation methods, the backfitting algorithm performs an iterative procedure to update the estimate until convergence. The polynomial spline method is easy to implement and requires less computation.

In this dissertation, we are particularly interested in the estimation of additive models when each additive component is monotone. The monotone additive model is useful in many areas where the predictors are known or required to have monotone effects on the response. For example, Morton-Jones et al. (2000) used the monotone additive model to assess disease risk in epidemiology; de Boer et al. (2002) applied the monotone additive model to explore the relationship between the toxicity and the degree of contamination of aquatic sediments.

The monotone additive model is not well represented in the literature compared with the additive model. A widely used approach to estimate the monotone additive model is to iteratively apply the Pool Adjacent Violator Algorithm (PAVA) in a backfitting

procedure (Mammen and Yu 2007). The resulting estimator proposed in Mammen and Yu (2007) enjoys the oracle property, i.e., each additive component can be estimated with the same asymptotic accuracy as if the other components were known. However, this method only gives a step-wise estimate that is discontinuous so it is difficult to interpret the estimation result. Leitenstorfer and Tutz (2007) and Tutz and Leitenstorfer (2007) proposed a continuous estimate using boosting based on B-spline basis functions or monotonic basis functions but without asymptotic guarantees.

In the first part of this dissertation, we propose a one-step backfitted constrained polynomial spline estimator for monotone additive models. The proposed method approximates nonparametric components using polynomial splines and obtains smooth monotone estimates by applying the constrained polynomial spline method. The proposed constrained polynomial spline method imposes a set of simple linear constraints on spline coefficients in a backfitting procedure. This method takes advantage of linear programming and is efficient to compute. In univariate smoothing with only one predictor and with quadratic spline, our proposed estimator is reduced to the one proposed by He and Shi (1998). However, their method is only developed for the univariate case and quadratic spline. Our proposed estimation method can be used for multi-dimensions and any order of polynomial splines.

Our simulation studies show that the proposed method has better numerical performance than some existing methods. It is particularly useful and has advantages when the data has outliers. In addition, we establish the asymptotic properties of our proposed smooth monotone estimator and prove that it has the optimal rate of convergence in terms of $L_2$ norm for large sample sizes.

As an extension and application of our proposed constrained polynomial spline method in econometrics, we are also interested in developing a method for the estimation of production frontier functions. Consider a nonnegative vector $(\mathbf{x}, y) \in \mathscr{R}_+^d \times \mathscr{R}_+$, where $\mathbf{x}$ represents the $d$ inputs used in production and $y$ represents the output of a production unit. According to economic theory (Koopmans 1951, Shephard 1970), the production set is defined as $\Psi = \left\{(\mathbf{x}, y) \in \mathscr{R}_+^d \times \mathscr{R}_+ |\ \mathbf{x} \text{ can produce } y\right\}$, i.e., the set of physically attainable points $(\mathbf{x}, y)$. The production frontier function of $\Psi$ is defined as $\rho(\mathbf{x}) = \sup\{y, (\mathbf{x}, y) \in \Psi\}$ that is the upper boundary of the production set. The output efficiency measure is defined as $R = y/\rho(\mathbf{x}) \in [0, 1]$. The production function specifies the maximal achievable output for a firm working at the level of inputs $\mathbf{x}$ and presents a useful benchmark value or reference frontier that can be used to assess efficiency for firms operating at the same level of inputs. The main focus of frontier analysis is on the specification and estimation of the production frontier function $\rho(\cdot)$ given a random sample of the production units $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$.

Since the seminal work of Koopmans (1951) and Debreu (1951), an expanding area of research has focused on the estimation of production frontiers and on the measurement of the corresponding efficiency of production units. Two main approaches have been developed for the estimation of production frontier functions: the deterministic frontier model and the stochastic approach. Deterministic frontier models rely on the assumption that all data lie in the production set $\Psi$, i.e., $P\{(\mathbf{X}_i, Y_i) \in \Psi\} = 1$, for $i = 1, \ldots, n$. Two well-known nonparametric estimation methods have been developed under this framework: data envelopment analysis (DEA) (Farrell 1957, Charnes et al. 1978) and full disposal hull (FDH) (Deprins et al. 1984). Both methods employ linear

programming to find the smallest free disposal set or the smallest free disposal convex set covering all data points. Their improved versions can be found in Hall et al. (1998), Knight (2001), Hall and Park (2002) and Jeong and Simar (2006). These methods are attractive because they rely on very few assumptions on the production set $\Psi$ and the joint distribution of $\{\mathbf{X}, Y\}$. However, both methods can be severely influenced by the presence of outliers or extreme values since they envelop all data points in their construction. The stochastic approach, initiated by Aigner et al. (1977) and Meeusen and van den Broek (1977), allows some observations to be outside of the production set. However, stochastic frontier models often assume strong parametric restrictions on the shape of the frontier function $\rho(\cdot)$ and a misspecified model can lead to invalid estimation and inference results.

Martins-Filho and Yao (2007) proposed an appealing deterministic frontier regression model that is nonparametric in nature and flexible to capture complex structures of the production frontier. Their method is more robust to extreme values and outliers than the DEA and FDH approaches. Martins-Filho and Yao (2007) estimated the frontier using a three-step procedure based on local linear smoothing. However, due to the curse of dimensionality, their method can only accommodate low dimensions of input variables. In addition, econometric theories often impose shape constraints on the frontier function. The general production axiom of free disposability of inputs and outputs (Färe et al. 1985 and Shephard 1970) implies that the frontier function $\rho(\cdot)$ is monotone. The convexity of the production set $\Psi$ implies that $\rho(\cdot)$ is also concave, corresponding to diminish marginal returns. Therefore, it is of interest to provide frontier estimates that automatically satisfy such shape constraints.

The nonparametric methods, such as DEA and FDH, give monotone estimates. However, their estimates are step-wise and hard to interpret. As pointed out in Daouia et al. (2016), they often underestimate the true support boundary. More recently, Wu and Sickles (2013) proposed a monotone spline estimator (Ramsay 1998) for a semi-parametric frontier model. It achieves monotonicity and concavity by using integral transformations of non-constrained second order derivative functions. But the class of such integral transformed functions is relatively small compared to the class of all monotone and concave functions. Better estimation results can be obtained by using more general spline functions. Daouia et al. (2016) extended the idea in Hall et al. (1998) and proposed a novel method to estimate the boundary of the production set using constrained spline methods. They used linear programming or second-order cone programming to find the closest constrained spline function that envelops the data set. However, their method focuses on the single input case and only work well for one or low dimensional input variables.

In the second part of this dissertation, we extend the nonparametric regression frontier model in Martins-Filho and Yao (2007) for multiple input variables and impose an additive structure on frontier functions to partially alleviate the curse of dimensionality. Furthermore, we consider to estimate the frontier functions under shape constraints that is not well represented in the existing literature. Here, we propose a two-step constrained polynomial spline method for the frontier estimation. Our proposed method guarantees a smooth estimator and is easy to implement. It also can be applied to multi-dimensions where many existing methods do not work. Most importantly, we incorporate the shape constraints (monotonicity or/and concavity) in the estimation procedure to capture the

shape of frontiers more accurately. The simulation studies illustrate that our proposed method with shape constraints is effective in enhancing estimation accuracy. In addition, the proposed estimator is more robust to outliers than the estimator proposed by Martins-Filho and Yao (2007).

This dissertation is organized as follows. In Chapter 2, we propose a one-step constrained polynomial estimation method for the monotone additive models. In Chapter 3, we extend the constrained nonparametric method to estimate the production frontier functions and develop a two-step shape constrained estimator. In both Chapters 2 and 3, we examine our proposed methods by performing simulation studies and analyzing a real data set, the Norwegian Farm data. A summary of our findings and discussion are provided in Chapter 4.

# 2 Constrained Polynomial Spline Estimation of Monotone Additive Models

## 2.1 Additive Models

The additive model that was made popular by Hastie and Tibshirani (1990) has been widely used in multivariate nonparametric modeling and received considerable attention in recent decades. As a nonparametric regression method, the additive model allows the effect of each predictor variable to be modeled non-parametrically while requiring an additive structure.

Suppose there are $n$ independent and identically distributed (i.i.d.) observations generated from the additive model

$$Y_i = \alpha_0 + \alpha_1(X_{i1}) + \ldots + \alpha_d(X_{id}) + \varepsilon_i \quad i = 1, \ldots, n, \tag{2.1}$$

where $Y_i$ is the response variable and $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^T$ are the predictor variables for the $i$ -th observation. Furthermore, in model (2.1), $\alpha_0$ is the unknown intercept and $\{\alpha_l(\cdot)\}_{l=1}^d$ are unknown univariate smooth nonparametric functions. Without loss of generality, we assume that the predictors are distributed on the compact support $[0,1]^d$. In addition, for model identification the nonparametric function $\alpha_l(\cdot)$ is assumed to be theoretically centered with $E[\alpha_l(X_l)] = 0$, for $l = 1, \ldots, d$. As a result, one has $E(Y) = \alpha_0$. Therefore, the intercept term $\alpha_0$ can be consistently estimated as $\hat{\alpha}_0 = \bar{Y}$

$= \frac{1}{n}\sum_{i=1}^{n} Y_i$ at the parametric convergence rate of $\sqrt{n}$ that is faster than the convergence rate for nonparametric function estimation. For the sake of simplicity, in the following we assume $\alpha_0 = 0$ in (2.1) and focus on the estimation of nonparametric functions $\{\alpha_l(\cdot)\}_{l=1}^{d}$ in the simplified model

$$Y_i = \alpha_1(X_{i1}) + \ldots + \alpha_d(X_{id}) + \varepsilon_i \quad i = 1, \ldots, n, \tag{2.2}$$

The additive model is a flexible generalization of the linear regression model. On one hand, it relaxes the restriction on the linear relationship between the predictors and response. On the other hand, the additive model also retains the important feature of the linear regression model: the model is additive in the predictor effects. The additive functional form enables easy interpretation of the fitted results since each predictor has a marginal effect on the response. Most importantly, the additive model avoids curse of dimensionality since each additive component can be estimated using univariate smoothing.

In this chapter, we are interested in the estimation of model (2.2) in the situation where the nonparametric functions $\{\alpha_l(\cdot)\}_{l=1}^{d}$ are monotone increasing or decreasing. Our question of interest is how to estimate the nonparametric functions $\{\alpha_l(\cdot)\}_{l=1}^{d}$ under the monotone constraints. In the following, we propose a one-step back-fitted constrained polynomial spline procedure to estimate such monotone additive models. It first uses the traditional polynomial spline method to obtain an initial estimator of each additive component. Then a back-fitted constrained polynomial spline method is used to produce the final estimator for each component that is guaranteed to be monotone.

## 2.2 Methodology and Theory

In our proposed method, we estimate the nonparametric functions $\{\alpha_l(\cdot)\}_{l=1}^d$ in model (2.2) using polynomial splines. Let $u_n = \{0 = u_0 < u_1 < \cdots < u_{N_n} < u_{N_n+1} = 1\}$ be a partition of the interval $[0,1]$, with $N_n$ interior knots. With this knot sequence, the interval $[0,1]$ is partitioned into $N_n + 1$ intervals and we write them as $I_k = [u_k, u_{k+1})$ for $k = 0, ..., N_n - 1$ and $I_{N_n} = [u_{N_n}, u_{N_n+1}]$. Using $u_n$ as knots, the polynomial splines of order $p+1$ are polynomial functions with degree $p$ (or less) on the intervals $I_k$ for $k = 0, ..., N_n$, and $(p-1)$-times differentiable at the interior knots. In this dissertation, we denote the space of $p$-times continuously differentiable real-valued functions in $[0,1]$ as $C^p[0,1]$ and define the space of polynomial splines of order $p+1$ (or degree $p$) based on the knots $u_n$ as $G^p = G^p([0,1], u_n)$.

### 2.2.1 An Initial Estimator

Let $\tilde{\mathbf{B}}(x) = \left(\tilde{B}_1(x), \ldots, \tilde{B}_{J_n+1}(x)\right)^T$ be the set of B-spline basis of $G^p$ with dimension $J_n = N_n + p$. Because of the fact that $\sum_{j=1}^{J_n+1} \tilde{B}_j(x) = 1$ for any $x \in [0,1]$. Without loss of generality, we focus on the first $J_n$ basis and create an empirically centered B-spline basis for each variable $X_l$ by taking $B_{lj} = \tilde{B}_j - \frac{1}{n}\sum_{i=1}^n \tilde{B}_j(x_{il})$ for model identification. We denote the empirically centered B-spline basis as $\mathbf{B}_l(x) = (B_{l1}(x), \ldots, B_{lJ_n}(x))^T$, for $l = 1, \cdots, d$. Lemma 2 of Xue and Yang (2006) ensures that the theoretically centered function $\alpha_l(\cdot)$ in model (2.2) can be approximated well by a linear combination of the empirically centered B-spline basis. Therefore, one can write $\alpha_l(x) \approx \mathbf{B}_l^T(x)\beta_l$ with a set of coefficients $\beta_l = \left(\beta_{l1}, \ldots, \beta_{lJ_n}\right)^T$, for $l = 1, \cdots, d$. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ and

$\mathbf{B}_n = (\mathbf{B}_{n1}, \ldots, \mathbf{B}_{nd})$, where $\mathbf{B}_{nl} = (\mathbf{B}_l(x_{1l}), \ldots, \mathbf{B}_l(x_{nl}))^T$. The traditional polynomial spline method (Stone 1985 and Huang 1998) estimates the unknown coefficients $\beta = \left(\beta_1^T, \ldots, \beta_d^T\right)^T$ by taking

$$\tilde{\beta} = \underset{\beta \in R^{dJ_n}}{\arg\min} (\mathbf{Y} - \mathbf{B}_n\beta)^T (\mathbf{Y} - \mathbf{B}_n\beta) = \left(\mathbf{B}_n^T\mathbf{B}_n\right)^{-1}\mathbf{B}_n^T\mathbf{Y}. \tag{2.3}$$

Then the unknown function $\alpha_l(\cdot)$ is estimated as

$$\tilde{\alpha}_l(x) = \mathbf{B}_l^T(x)\tilde{\beta}_l, \tag{2.4}$$

for $l = 1, \ldots, d$. Stone (1985) and Huang (1998) established the optimal $L_2$ rate of convergence for the polynomial spline estimator (2.4). They showed that the estimate of each nonparametric component in the additive model enjoys the same optimal convergence rate as a univariate nonparametric function estimator. The local asymptotic theory of the polynomial spline estimator was developed in Huang (2003). However, this traditional polynomial spline method is unable to give shape constrained estimates, such as monotone increasing or decreasing function estimates. Therefore we only use the traditional polynomial spline estimator (2.4) as an initial estimator and propose a constrained polynomial spline method to get a monotone estimator. In the following, we focus on estimating monotone increasing functions. The estimation of monotone decreasing functions follows similarly.

### 2.2.2 Proposed Method

Using the traditional polynomial spline estimator $\{\tilde{\alpha}_l(\cdot)\}_{l=1}^d$ as an initial estimator, we define the pseudo response $Y_{i,-l} = Y_i - \sum_{l' \neq l} \tilde{\alpha}_{l'}(x_{il'})$ and $\mathbf{Y}_{-l} = \left(Y_{1,-l}, ..., Y_{n,-l}\right)^T$, for $l = 1, ..., d$. Then $\mathbf{Y}_{-l}$ can be viewed as an approximation of $\alpha_l(\cdot)$ at $n$ data points. In addition, by Example 4.25 in Schumaker (2007), a sufficient condition for a polynomial spline $g(x) = \tilde{\mathbf{B}}^T(x)\beta$ to be monotone increasing is that its coefficients satisfy $\beta_j \geq \beta_{j-1}$, for $j = 2, ..., J_n + 1$. When using the empirically centered B-spline basis, Lemma 4 gives a modified sufficient condition that is $\beta_1 \geq 0$ and $\beta_j \geq \beta_{j-1}$, for $j = 2, ..., J_n$. Denote the set of spline coefficients that satisfy the monotone constraints as

$$C_l = \left\{ \left(\beta_{l1}, \ldots, \beta_{lJ_n}\right) \in R^{J_n} \mid \beta_{l1} \geq 0 \text{ and } \beta_{lj} \geq \beta_{l(j-1)}, \text{ for } j = 2, \ldots, J_n \right\}.$$

Then, we propose to estimate the coefficients $\beta_l$ by minimizing the following constrained least squares to ensure monotonicity,

$$\hat{\beta}_l \;=\; \arg\min_{\beta_l} \left(\mathbf{Y}_{-l} - \mathbf{B}_{nl}\beta_l\right)^T \left(\mathbf{Y}_{-l} - \mathbf{B}_{nl}\beta_l\right), \quad \text{subject to } \beta_l \in C_l. \tag{2.5}$$

As a result, with the constrained least squares coefficient estimator $\hat{\beta}_l$, the monotone polynomial spline estimation of $\alpha_l(\cdot)$ is obtained by

$$\hat{\alpha}_l(x) = \mathbf{B}_l^T(x)\hat{\beta}_l. \tag{2.6}$$

For $p = 2$ and $d = 1$, the constrained estimator (2.6) is reduced to the estimator

proposed by He and Shi (1998) and its asymptotic theory has been developed. The estimator developed by He and Shi (1998) is obtained by constraining the first order derivative of a quadratic spline function to be positive. Their approach, however, is only feasible for the univariate case and quadratic spline with degree $p = 2$ when its first order derivative is linear in unknown parameters. Our approach is generally applicable for any order of polynomial splines and any dimension of predictor variables. Here, we extend the work for any integer $p$ and $d$ and are also interested in exploring the asymptotic properties of the generalized constrained estimator (2.6).

Furthermore, we could also obtain the estimate of the coefficients by directly minimizing the global sum of squares as in equation (2.3), but subject to all $d$ constraints $\beta_l \in C_l$ for each $l = 1, \ldots, d$ simultaneously. However, this approach is not numerically stable due to the difficulty in conducting the constrained optimization with a large number of parameters. Therefore, we choose to estimate each additive component separately in a backfitting procedure. In addition, one can also use the constrained estimator (2.6) as an another initial value and repeat the backfitting procedure in (2.5) until convergence. However, our experience indicates that its numerical performance is very similar to the one-step backfitted constrained polynomial spline estimation. Therefore, we will focus on the one-step backfitted estimator (2.6).

### 2.2.3 Asymptotic Properties

To establish the theoretical results of our proposed method, we write the traditional polynomial spline estimator $\tilde{\alpha}_l(x)$ as an unconstrained one-step backfitted spline esti-

mator of each additive component. Let $\mathbf{Y}_{-l}$ be the pseudo response for the predictor variable $X_l$. The coefficients estimator $\tilde{\beta}_l$ can be viewed as the solution that minimizes the sum of squares $(\mathbf{Y}_{-l} - \mathbf{B}_{nl}\beta_l)^T (\mathbf{Y}_{-l} - \mathbf{B}_{nl}\beta_l)$ without any constraints and $\tilde{\alpha}_l(x) = \mathbf{B}_l^T(x)\tilde{\beta}_l$ is the corresponding one-step backfitted spline estimator of $\alpha_l(\cdot)$, for each $l = 1, \ldots, d$. For finite samples, $\tilde{\alpha}_l(\cdot)$ is not always monotone when $\alpha_l(\cdot)$ is actually monotone. But in the following, we show that $\tilde{\alpha}_l(\cdot)$ is monotone increasing with probability approaching to one as the sample size increases.

For our asymptotic analysis, we need the following assumptions:

(A1) The errors $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. distributed with $E(\varepsilon_i|X) = 0$ and $Var(\varepsilon_i|X) = \sigma^2 < +\infty$. There exists positive constant $c_1$, such that $E\left(|\varepsilon_i|^{2+\eta}|X\right) \leq c_1$ *a.s.* for some $\eta > 0$.

(A2) The predictor variables $\mathbf{X}_i$ are i.i.d. distributed on a compact support. Without loss of generality, we assume that the support is $[0,1]^d$. Its density function, denoted by $f(\mathbf{x})$, is continuous and $0 < c_2 \leq f(\mathbf{x}) \leq c_3 < \infty$, for $\mathbf{x} \in [0,1]^d$ and positive constants $c_2$ and $c_3$.

(A3) For each $l = 1, \ldots, d$, the additive function $\alpha_l$ is strictly monotone increasing and $(p+1)$-times continuously differentiable for some integer $p \geq 1$. Furthermore, we assume that there exists a constant $c_4 > 0$, such that $\alpha_l'(x) \geq c_4$, for $x \in [0,1]$.

(A4) For the set of knots $\{0 = u_0 < u_1 < \cdots < u_{N_n} < u_{N_n+1} = 1\}$, there exists a constant $c_5 > 0$, such that $\frac{\max(u_{j+1}-u_j, j=0,\ldots,N_n)}{\min(u_{j+1}-u_j, j=0,\ldots,N_n)} \leq c_5$.

(A5) The number of interior knots $N_n$ satisfies

$$N_n \to +\infty \text{ and } \frac{N_n^4}{n} \to 0, \text{ as } n \to +\infty.$$

Condition (A1) requires that the error terms are i.i.d. distributed with a common distribution that is not necessarily a Normal distribution. Assumption (A2) is the same as condition 1 of Stone (1985). Assumption (A3) requires that the regression function is strictly monotone increasing and its first order derivative is lower bounded. For the monotone decreasing functions, the corresponding theorems can be developed in a similar way. Assumption (A4) requires that the interior knots are pseudo equally spaced in the interval $[0,1]$. The same condition is also considered in Huang (1998). Assumption (A5) is the conditions for the number of interior knots and samples size under those the following theorems are developed.

**Theorem 1** *Under regularity conditions (A1)-(A5), for $l = 1, ...d$, one has,*

$$\sup_{x\in[0,1]} \left|\tilde{\alpha}_l'(x) - \alpha_l'(x)\right| = O_p\left(\sqrt{N_n^{-2p+1} + N_n^4/n}\right).$$

Write $\tilde{\alpha}_l(x) = \tilde{\mathbf{B}}_l^T(x)\tilde{\beta}_l$ with coefficients $\tilde{\beta}_l = \left(\tilde{\beta}_{l1}, \ldots, \tilde{\beta}_{l(J_n+1)}\right)^T$.

**Theorem 2** *Under regularity conditions (A1)-(A5), one has, for $l = 1, \ldots, d$ and $p \leq 3$, there exists a spline function $g_l$ with $g_l = \mathbf{B}_l^T \gamma_l$ and the coefficients $\gamma_l = \left(\gamma_{l1}, \ldots, \gamma_{l(J_n+1)}\right)^T$ satisfy the monotone constraints $\gamma_{lj} - \gamma_{l(j-1)} \geq 0$, $j = 2, \ldots, J_n + 1$, such that*

$$\sup_{j=1,\ldots,J_n+1} \left|\tilde{\beta}_{lj} - \gamma_{lj}\right| = O_p\left(\sqrt{\frac{N_n^3 \log n}{n}}\right).$$

*Therefore, the coefficients of $\tilde{\alpha}_l(x)$ satisfy the monotone constraints $\tilde{\beta}_{lj} - \tilde{\beta}_{l(j-1)} \geq 0$, for $j = 2, \ldots, J_n + 1$ with probability approaching to 1 as $n \to \infty$.*

**Theorem 3** *Under regularity conditions (A1)-(A5), for $p \leq 3$ and any $l = 1, \ldots d$, one has,*

$$\|\hat{\alpha}_l - \alpha_l\| = O_p\left(\sqrt{N_n^{-2p-2} + N_n/n}\right).$$

Theorem 1 implies that the one-step backfitted unconstrained estimator $\tilde{\alpha}_l(\cdot)$ is actually monotone increasing for large sample size. Furthermore, we prove that for degree $p \leq 3$, the coefficients of $\tilde{\alpha}_l(\cdot)$ satisfy the monotone constraints with probability approaching to 1 as the sample size goes to infinity. This suggests that the unconstrained estimator $\tilde{\alpha}_l(\cdot)$ and constrained estimator $\hat{\alpha}_l(\cdot)$ are identical when the sample size is large enough. Therefore, naturally the constrained estimator $\hat{\alpha}_l(\cdot)$ enjoys the same asymptotic properties as the unconstrained estimator and the result is stated in Theorem 3.

## 2.3 Empirical Results

In this section, we conduct simulation studies to evaluate the numerical performance of our proposed method with finite samples. We also illustrate the application of the proposed method by analyzing a real data set, the Norwegian Farm data.

### 2.3.1 Simulation Study

We consider a monotone additive model with five additive components, $Y = \alpha_1(X_1) + \alpha_2(X_2) + \alpha_3(X_3) + \alpha_4(X_4) + \alpha_5(X_5) + \varepsilon$, where the five monotone increasing functions $\{\alpha_l(\cdot)\}_{l=1}^5$ are given in Table 2.1. The predictors $\{X_l\}_{l=1}^5$ are independently generated

from the uniform distribution on $[0,1]$ and the error $\varepsilon$ follows a normal distribution with mean 0 and variance $\sigma^2$. We set $\sigma^2 = \frac{1}{4}\sum_{l=1}^{5}\int_0^1 \alpha_l^2(x)\,dx$ so that the signal to noise ratio is 4.

For this additive model, $r = 100$ samples of size $n = 50$, 100, 200 and 500 are generated, respectively. For each generated data, we estimate the monotone functions using the proposed constrained quadratic spline method (CQS) with $p = 2$ as defined in (2.6). We also consider the unconstrained quadratic spline (UQS) estimator described in (2.4) and an estimator obtained by applying Pool Adjacent Violator Algorithm (PAVA) in a one-step backfitting procedure. For $l = 1,\ldots,5$, let $x_{\min,l}$ and $x_{\max,l}$ denote the smallest and largest observations of the variable $x_l$, respectively. Knots are equally placed in the interval $\left[x_{\min,l}, x_{\max,l}\right]$ and the number of the interior knots $N_n$ is selected as the integer part of $n^{\frac{1}{2p+3}}$. We evaluate the estimation accuracy of three different estimators by their averaged integrated squared errors (AISE). Let $\hat{\alpha}_{k,l}$ be an estimator of $\alpha_l$ in the $k$-th replication and $\left\{x_j\right\}_{j=1}^{ngrid}$ be a set of grid points in the interval $[0,1]$ where the functions are evaluated. Then we define the integrated squared error as $ISE(\hat{\alpha}_{k,l}) = \frac{1}{ngrid}\sum_{j=1}^{ngrid}\left[\hat{\alpha}_{k,l}\left(x_j\right) - \alpha_l\left(x_j\right)\right]^2$ and averaged integrated squared error as $AISE(\hat{\alpha}_l) = \frac{1}{r}\sum_{k=1}^{r} ISE(\hat{\alpha}_{k,l})$. Furthermore, in order to check the robustness of the methods, ten outliers are added manually to each generated data set. For the 100 replications, the range of simulated response values varies. But for majority of the replications, the range is between $[-3,3]$. To make it simple, ten observations $y_i$ are randomly selected and their response values $y_i$ are replaced with $y_{i,new} = y_i - 5$ or $y_{i,new} = y_i + 5$. Then, all three estimators (CQS, UQS, PAVA) are calculated again using the contaminated data sets and their AISEs are reevaluated.

The simulation results are summarized in Table 2.2. Overall, it clearly shows that the AISEs decrease when the sample size $n$ increases. This verifies our asymptotic theory. Moreover, when there are no outliers, the AISEs of the constrained estimator are slightly smaller than that of the unconstrained estimator and both of them are much smaller than that of the PAVA estimator. The difference among these three estimators diminishes as the sample size increases. When there are outliers, our proposed constrained method performs noticeably much better than the other two methods, especially for small and moderate sample sizes. For example, when the sample size $n = 50$, the AISEs decrease by about 50% if we use the constrained estimation method instead of the unconstrained method. It clearly shows that our proposed constrained method is more robust to the outliers and is particularly useful when the sample size is small or moderate. Furthermore, with the increase of sample size, the unconstrained estimator tends to be similar to the constrained estimator and both are much better than the PAVA estimator, particularly for large sample sizes.

In addition to the numerical tables, we also plot the typically estimated functions in Figure 2.1 for the data without outliers and in Figure 2.2 for the data with outliers, for sample size $n = 50$. The typically estimated function is the one whose ISE is the median in the 100 replications. In each plot, we plot the three typically estimated curves using CQS (dot-dashed red line), UQS (dashed blue line) and PAVA (long-dashed green line) along with the true functions (solid black line). We also plot the 95% point-wise confidence intervals for each additive function (dotted brown line) using constrained polynomial spline method. At each grid point $\{x_j\}_{j=1}^{ngrid}$, the 95% point-wise confidence interval uses the 2.5% and 97.5% sample quantiles of the 100 estimates from CQS

as the lower and upper bounds, respectively. From Figure 2.1, we can see that our proposed estimation method works very well since the fitted functions are very close to the true functions. The PAVA method has the worst performance and its estimation is not smooth. When there are outliers, Figure 2.2 shows that there are more fluctuation for the fitted curves of the unconstrained and PAVA method since the outliers tend to pull the curve down or up. However, the constrained estimator performs better than the other two methods and there is no apparent influence of the outliers. It graphically confirms the numerical findings observed in Table 2.2.

### 2.3.2 Norwegian Farm Data

In addition to the simulation study, we are also interested in applying our proposed method to a real data set. Here, we consider the Norwegian Farm data that was used in Kumbhakar, Lien and Hardaker (2014). This data set is originally obtained from the Norwegian Farm Accountancy Survey and includes farm production and economic data collected annually. In our analysis, we use 2007 data that contains observations on 151 grain farms in Norway. The data consists of one response variable and four predictors. The response $Y$ is the farm revenue measured in Norwegian kroner (2008 NOK). The predictors are total number of hours worked (labor) on the farm ($X_1$), productive farmland in hectares ($X_2$), variable farm inputs ($X_3$) and fixed farm input and capital costs ($X_4$). A monotone additive model with four additive components $Y = \alpha_1(X_1) + \alpha_2(X_2) + \alpha_3(X_3) + \alpha_4(X_4) + \varepsilon$ is considered to model the monotone increasing effects of predictor variables on farm revenue.

In our analysis, we randomly partition the data into two parts: the training and the test sets. The 120 observations in the training set are used to obtain the unconstrained and constrained estimators and the remaining 31 observations in the test set are for the prediction. The PAVA method is not considered in this analysis because of its inferior performance in our simulation study. For both unconstrained and constrained methods, we consider linear ($p = 1$) and cubic ($p = 3$) splines with two different knot sequences. The knots are equally spaced in the range of each variable $x_l$ or equally spaced in the percentile ranks by taking the $\frac{j}{N_n+1}$-th quantile of the observations of each variable $x_l$ for $j = 1,\ldots,N_n$. The number of the interior knots $N_n$ is chosen as the integer part of $n^{\frac{1}{2p+3}}$ or $n^{\frac{1}{2p+3}}+1$ . The performance of estimators is assessed by the averaged mean squared estimation error (AMSEE) and averaged mean squared prediction error (AMSPE). Let $\{\mathbf{x}_i = (x_{i1},\ldots,x_{i5}), y_i\}_{i=1}^{120}$ and $\{\mathbf{x}_i = (x_{i1},\ldots,x_{i5}), y_i\}_{i=121}^{151}$ be the data points used for estimation and prediction, respectively. Denote $\hat{\alpha}_l$ as an estimator of $\alpha_l$ and $\hat{\alpha} = \sum_{l=1}^{d} \hat{\alpha}_l$ as an estimator of the regression function using a given partition of the data set. We define the mean squared estimation error (MSEE) and mean squared prediction error (MSPE) as $MSEE(\hat{\alpha}) = \frac{1}{120}\sum_{i=1}^{120}\left[y_i - \sum_{l=1}^{d}\hat{\alpha}_l(x_{il})\right]^2$ and $MSPE(\hat{\alpha}) = \frac{1}{31}\sum_{i=121}^{151}\left[y_i - \sum_{l=1}^{d}\hat{\alpha}_l(x_{il})\right]^2$, respectively. Then we repeat this process 100 times and obtain the averaged mean squared estimation error (AMSEE) and the averaged mean squared prediction error (AMSPE) by $AMSEE = \frac{1}{100}\sum_{r=1}^{100} MSEE(\hat{\alpha}^r)$ and $AMSPE = \frac{1}{100}\sum_{r=1}^{100} MSPE(\hat{\alpha}^r)$, respectively, where $\hat{\alpha}^r$ is an estimator of the regression function in the $r$-th replication.

The AMSEEs and AMSPEs for both methods are summarised in Table 2.3. We observe that the AMSEEs of the unconstrained method are smaller than that of the con-

strained method. This is not surprising since when the constraints are added, the sum of squares in (2.5) is minimized over a smaller parameter space and this will lead to larger estimation errors. However, when we compare the AMSPEs, it clearly shows that the constrained estimator gives smaller AMSPEs. Furthermore, the estimated additive components are plotted in Figure 2.3 (equally spaced knots) and Figure 2.4 (equal quantile knots). In both plots, the estimates using constrained linear spline (dot red line), unconstrained linear spline (dashed blue line), constrained cubic spline (solid red line), and unconstrained cubic spline (dot-dashed blue line) are plotted. The cubic splines method ($p = 3$) provide more smooth function estimates compared with linear splines ($p = 1$). Furthermore, for the productive farmland variable ($X_2$) in Figure 2.3, we can see a clear difference in the fitted curves between the unconstrained and constrained methods due to one particular data point at the right bottom of the plot. This point dramatically pulls down the fitted curve of the unconstrained method and the resulting estimate is no longer monotone. However, our proposed method is less sensitive to this data point. Additionally, when using the equal quantile knots, Figure 2.3 illustrates that unconstrained and constrained methods produce very similar fitted results.

## 2.4 Proof of Lemmas and Theorems

### 2.4.1 Preliminary Lemmas and Proof

**Lemma 4** *A sufficient condition for a polynomial spline $g(x) = \mathbf{B}^T(x)\beta$ , where $\mathbf{B}(x)$ is the empirically centered B-spline basis of $G^p$, to be monotone increasing is that its*

*coefficients satisfy the following conditions:* $\beta_1 \geq 0$ *and* $\beta_j \geq \beta_{j-1}$, *for* $j = 2, ..., N_n + p$.

**Proof**. Recall that $\tilde{\mathbf{B}}(x) = \left(\tilde{B}_{-p}(x), \ldots, \tilde{B}_{N_n}(x)\right)^T$ is the non-centered B-spline basis of $G^p$. By de Boor (2001, p.115), one has

$$\tilde{B}'_{j,k}(x) = (k-1)\left[-\tilde{B}_{j+1,k-1}(x) / \left(u_{j+k} - u_{j+1}\right) + \tilde{B}_{j,k-1}(x) / \left(u_{j+k-1} - u_j\right)\right],$$

where $k = p+1$ is the order of the polynomial spline. Therefore, the first order derivative of the polynomial spline is

$$\begin{aligned}
& g'(x) \\
= \; & \left[\mathbf{B}^T(x)\beta\right]' = \left[\sum_{j=-p+1}^{N_n} \beta_j B_{j,k}(x)\right]' \\
= \; & \left[\sum_{j=-p+1}^{N_n} \beta_j \left(\tilde{B}_{j,k} - \frac{1}{n}\sum_{i=1}^{n} \tilde{B}_{j,k}(x_i)\right)\right]' \\
= \; & \left[\sum_{j=-p+1}^{N_n} \beta_j \tilde{B}_{j,k}\right]' = \sum_{j=-p+1}^{N_n} \beta_j \tilde{B}'_{j,k} \\
= \; & (k-1)\left[\frac{\beta_{-p+1}}{u_1 - u_{-p+1}} \tilde{B}_{-p+1,k-1} + \sum_{j=-p+2}^{N_n} \frac{\beta_j - \beta_{j-1}}{u_{j+k-1} - u_j} \tilde{B}_{j,k-1}\right]. \qquad (2.7)
\end{aligned}$$

Because the B-spline basis is positive, a sufficient condition to guarantee the monotonicity of the polynomial spline is that the coefficients of the basis are non-negative. Therefore, one has $\beta_{-p+1} \geq 0$ and $\beta_j - \beta_{j-1} \geq 0$, for $j = -p+2, ..., N_n$, and Lemma 4 follows. ■

**Lemma 5** *For any function* $\alpha(\cdot)$ *that satisfies condition (A3), there exists a monotone increasing function* $g \in G^{(p)}$, *such that* $\|\alpha - g\|_\infty \leq c \left\|\alpha^{(p+1)}\right\|_\infty / N_n^{p+1}$ *and* $\|\alpha' - g'\|_\infty \leq$

$c\left\|\alpha^{(p+1)}\right\|_{\infty}/N_n^p$ *for large sample size and some constant* $c>0$.

**Proof**. According to Theorem 1.51 in Schumaker (2015), for any $\alpha \in C^{p+1}[0,1]$, there exists a function $g \in G^{(p)}$, such that $\|\alpha - g\|_{\infty} \leq c\left\|\alpha^{(p+1)}\right\|_{\infty}/N_n^{p+1}$ and $\|\alpha' - g'\|_{\infty}$ $\leq c\left\|\alpha^{(p+1)}\right\|_{\infty}/N_n^p$ for some constant $c>0$. Next, we need to show that such $g$ is monotone increasing. By condition (A3), $\alpha'(x) \geq c_1 > 0$ for some constant $c_1$ and any $x \in [0,1]$. When the sample size $n$ is large enough, one has

$$g'(x) \geq \alpha'(x) - \left\|\alpha' - g'\right\|_{\infty} \geq c_1/2 > 0,$$

and Lemma 5 follows. ■

For each $l = 1,\ldots,d$, let $\alpha_l^*$ be the one-step backfitted estimate of $\alpha_l$ with all the other additive components known. In this case, it is reduced to a univariate polynomial spline smoothing for the $l^{th}$ variable. Let $Y_{i,-l}^* = Y_i - \sum_{l' \neq l}\alpha_{l'}(x_{il'}) = \alpha_l(x_{il}) + \varepsilon_i$ be the pseudo response with the other additive component known and $\mathbf{Y}_{-l}^* = \left(Y_{1,-l}^*, ..., Y_{n,-l}^*\right)^T$, for $l = 1, ..., d$. Then $\alpha_l^*(x) = \mathbf{B}_l^T(x)\beta_l^*$ with $\beta_l^* = \left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T\mathbf{Y}_{-l}^*$. One notes that $\left\{\alpha_l^*\right\}_{l=1}^d$ are not available in our analysis and are constructed to prove our theory only.

**Lemma 6** *Under regularity conditions (A1)-(A5), one has, for* $l = 1,\ldots,d$,

$$\sup_{x\in[0,1]}\left|\alpha_l^{*\prime}(x) - \alpha_l'(x)\right| = O_p\left(N_n^{\frac{1}{2}-p} + N_n^{\frac{3}{2}}/\sqrt{n}\right).$$

**Proof**. By condition (A3) and Lemma 5, for each $l = 1,\ldots,d$, there exists a monotone increasing function $g_l$, such that $\|\alpha_l - g_l\|_{\infty} \leq c\left\|\alpha_l^{(p+1)}\right\|_{\infty}/N_n^{p+1}$ and $\left\|\alpha_l' - g_l'\right\|_{\infty} \leq c\,\|$ $\alpha_l^{(p+1)}\,\|_{\infty}/N_n^p$. Then by definition, we have

$$
\begin{aligned}
& \alpha_l^{*\prime}(x) \\
= & \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \beta_l^* = \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \mathbf{Y}_{-l}^* \\
= & \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\alpha_l + \varepsilon) \\
= & \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{g}_l + \alpha_l - \mathbf{g}_l + \varepsilon) \\
= & \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \mathbf{g}_l + \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\alpha_l - \mathbf{g}_l) \\
& + \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \varepsilon \\
= & I(x) + II(x) + III(x). \qquad (2.8)
\end{aligned}
$$

One can write $\mathbf{g}_l = \tilde{\mathbf{B}}_{nl}\gamma_l$ for some coefficient $\gamma_l$. Therefore,

$$
I(x) = \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl} \beta_l^* = \left[\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \gamma_l = g_l^{\prime}(x),
$$

then by Lemma 5, one has

$$
\sup_{x\in[0,1]} \left|I(x) - \alpha_l^{\prime}(x)\right| = \sup_{x\in[0,1]} \left|g_l^{\prime}(x) - \alpha_l^{\prime}(x)\right| \le c \left\|\alpha_l^{(p+1)}\right\|_{\infty} / N_n^p. \qquad (2.9)
$$

For $II(x)$, Cauchy-Schwarz inequality gives that

$$
\begin{aligned}
\sup_x II(x) \quad \le \quad & \sqrt{\sup_x \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l^{\prime}(x)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l^{\prime}(x)\right]} \times \\
& \sqrt{\left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T (\alpha_l - \mathbf{g}_l)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T (\alpha_l - \mathbf{g}_l)\right]}.
\end{aligned}
$$

Let $\lambda_{\min}$ and $\lambda_{\max}$ be the smallest and largest eigenvalues of $\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}$, respectively.

Then by Theorem 5.4.2 in Devore and Lorentz (1993), we have $\lambda_{\min} \asymp \lambda_{\max} \asymp 1/N_n$ and $\sup_x \left[\tilde{\mathbf{B}}_l'(x)\right]^T \tilde{\mathbf{B}}_l'(x) = O_P\left(N_n^2\right)$, where $\asymp$ means both sides have the same order. Therefore

$$\left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l'(x)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l'(x)\right] \leq \frac{1}{n\lambda_{\min}} \left[\tilde{\mathbf{B}}_l'(x)\right]^T \tilde{\mathbf{B}}_l'(x) = O_P\left(N_n^3/n\right). \tag{2.10}$$

Similarly, by Lemma 5, $\sup|\alpha_l - g_l| = O_p\left(1/N_n^{p+1}\right)$, so we have

$$\begin{aligned} & \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T(\alpha_l - \mathbf{g}_l)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T(\alpha_l - \mathbf{g}_l)\right] \\ \leq & \frac{1}{n\lambda_{\min}}(\alpha_l - \mathbf{g}_l)^T \tilde{\mathbf{B}}_{nl}\tilde{\mathbf{B}}_{nl}^T(\alpha_l - \mathbf{g}_l) \\ \leq & \frac{\lambda_{\max}}{\lambda_{\min}}(\alpha_l - \mathbf{g}_l)^T(\alpha_l - \mathbf{g}_l) = O_p\left(n/N_n^{2p+2}\right). \end{aligned}$$

Therefore,

$$\sup_x III(x) = O_p\left(N_n^{\frac{1}{2}-p}\right). \tag{2.11}$$

Next, for $III(x)$, one notes $\sup_x \left[\frac{1}{\sqrt{n}}B_l'(x)\right]^T = O_p\left(N_n/\sqrt{n}\right)$ and the element of $\frac{1}{\sqrt{n}}B_{nl}^T\varepsilon$ is $\frac{1}{\sqrt{n}}\sum_{i=1}^n B_{lj}(x_i)\varepsilon_i = O_p\left(1/\sqrt{N_n}\right)$, we have

$$\sup_x III(x) = O_p\left(N_n^{\frac{3}{2}}/\sqrt{n}\right). \tag{2.12}$$

Finally, Lemma 6 follows from equations (2.8), (2.9), (2.11), (2.12) and condition (A5). ■

**Lemma 7** *For any function $\alpha \in C^{p+1}[0,1]$ that satisfies condition (A3) with $p \leq 3$,*

*there exists a monotone increasing function* $g \in G^{(p)}$ *and* $\hat{g} \in G^{(p)}$ *whose coefficients satisfy the linear constraints, such that* $\|\alpha - g\|_\infty \leq c_1 \left\|\alpha^{(p+1)}\right\|_\infty / N_n^{p+1}$ *and* $\|g - \hat{g}\|_\infty \leq c_2 \left\|\alpha^{(p+1)}\right\|_\infty / N_n^{p+1}$*, for large sample size and some constants* $c_1, c_2 > 0$*.*

**Proof**. According to Lemma 5, for any function $\alpha \in C^{p+1}[0,1]$ that satisfies condition (A3), there exists a monotone increasing function $g \in G^{(p)}$, such that $\|\alpha - g\|_\infty \leq c_1 \left\|\alpha^{(p+1)}\right\|_\infty / N_n^{p+1}$, for some constant $c_1 > 0$. For $p = 1$ and 2, it is easy to see that the coefficients of a monotone spline function satisfy the linear constraints. So by taking $\hat{g} = g$, Lemma 7 directly follows. For $p = 3$, since $\alpha \in C^4[0,1]$, its first order derivative $\alpha' \in C^3[0,1]$. The quadratic spline Hermite interpolation $g^* = \sum_{j=-2}^{N_n} \gamma_j^* \tilde{B}_{j,2}$ of $\alpha'$ satisfies $\alpha'(u_i) = g^*(u_i)$ and $\alpha^{(2)}(u_i) = g^{*\prime}(u_i)$, for $i = 0, \ldots, N_n + 1$, where

$$\begin{aligned} \gamma_{2k}^* &= \alpha'(u_{2k+2}) - (u_{2k+2} - u_{2k+1})\,\alpha^{(2)}(u_{2k+2})/2 \\ \gamma_{2k+1}^* &= \alpha'(u_{2k+2}) + (u_{2k+3} - u_{2k+2})\,\alpha^{(2)}(u_{2k+2})/2, \end{aligned} \tag{2.13}$$

for $k = -1, \ldots, (N_n - 1)/2$. By Theorem 6 of Dubeau and Savoie (1996), one has $\|\alpha' - g^*\|_\infty \leq c \left\|\alpha^{(4)}\right\|_\infty / N_n^3$, for some constant $c > 0$. Let $\hat{g}(x) = \int g^*(x)\,dx \in G^{(3)}$. One has

$$\begin{aligned} \|\alpha - \hat{g}\|_\infty &= \sup_x |\alpha(x) - \hat{g}(x)| = \sup_j \sup_{u_j \leq x \leq u_{j+1}} |\alpha(x) - \hat{g}(x)| \\ &= \sup_j \sup_{u_j \leq x \leq u_{j+1}} \left| \int_{u_j}^{x} \left[\alpha'(u) - \hat{g}'(u)\right] du \right| \end{aligned}$$

$$
\begin{aligned}
&\leq \sup_j \sup_{u_j \leq x \leq u_{j+1}} \int_{u_j}^{x} \left| \left[ \alpha'(u) - \hat{g}'(u) \right] \right| du \\
&= \sup_j \int_{u_j}^{u_{j+1}} \left| \left[ \alpha'(u) - g^*(u) \right] \right| du \\
&\leq c \left\| \alpha^{(4)} \right\|_\infty / N_n^4.
\end{aligned}
$$

Therefore, we have $\|g - \hat{g}\|_\infty \leq \|g - \alpha\|_\infty + \|\alpha - \hat{g}\|_\infty \leq c_2 \left\| \alpha^{(4)} \right\|_\infty / N_n^4$, for some constant $c_2 > 0$.

For $\hat{g}$ constructed above, we write $\hat{g} = \sum_{j=-3}^{N} \gamma_j \tilde{B}_{j,3}$ for a set of coefficients $\left\{ \gamma_j \right\}_{j=-3}^{N}$. Now we show that the spline coefficients satisfy the linear constraints that $\gamma_j \geq \gamma_{j-1}$, for $j = -2, \ldots, N_n$ when the sample size is large enough. By de Boor (2001, p.116), we have

$$
\begin{aligned}
\hat{g}' &= \left( \sum_{j=-3}^{N_n} \gamma_j \tilde{B}_{j,3} \right)' = \sum_{j=-3}^{N_n} \gamma_j \tilde{B}'_{j,3} \\
&= 3 \sum_{j=-2}^{N} \left( \gamma_j - \gamma_{j-1} \right) \tilde{B}_{j,2} / \left( u_{j+3} - u_j \right) \\
&= \sum_{j=-2}^{N_n} \gamma_j^* \tilde{B}_{j,2} = g^*.
\end{aligned}
$$

So the constraints $\gamma_j \geq \gamma_{j-1}$ are equivalent to $\gamma_j^* \geq 0$, for $j = -2, \ldots, N_n$. By the equation (2.13), $\min_j \gamma_j^* > \min_x \alpha'(x)/2 > c_3/2 > 0$ when the sample size $n$ is large enough. Therefore, $\hat{g}$ satisfies the linear constraints and the proof of Lemma 7 is complete. ■

Recall that $\alpha_l^* = \tilde{\mathbf{B}}_l^T \beta_l^*$ is the one-step backfitted estimate of $\alpha_l$ with all other additive components known.

**Lemma 8** *Under regularity conditions (A1)-(A5), one has, for $l = 1,\ldots,d$ and $p \leq 3$, there exists a spline function $g_l$ with $g_l = \mathbf{B}_l^T \gamma_l$ and the coefficients $\gamma_l = \left(\gamma_{l1},\ldots,\gamma_{l(J_n+1)}\right)^T$ that satisfy the monotone constraints $\gamma_{lj} - \gamma_{l(j-1)} \geq 0$, for $j = 2,\ldots,J_n+1$, such that*

$$\sup_j \left|\beta_{lj}^* - \gamma_{lj}\right| = O_p\left(\sqrt{\frac{N_n^3 \log n}{n}}\right).$$

*Therefore, the coefficients of $\alpha_l^*$ satisfy the monotone constraints $\beta_{lj}^* - \beta_{l(j-1)}^* \geq 0$, for $j = 2,\ldots,J_n+$ with probability approaching to 1 as $n \to \infty$.*

**Proof**. Let $\varepsilon_0 = \min\left\{\alpha_l'(x), x \in [0,1]\right\} \geq c > 0$, then $\alpha_l'(x) \geq \varepsilon_0$ for all $x \in [0,1]$. Lemma 6 suggests that there exists an integer $n(\varepsilon_0)$, such that $\sup_{x\in[0,1]}\left|\alpha_l^{*\prime} - \alpha_l'\right| < \varepsilon_0/2$, when $n > n(\varepsilon_0)$. Therefore when $n > n(\varepsilon_0)$, $\alpha_l^{*\prime} \geq \alpha_l' - \varepsilon_0/2 \geq \varepsilon_0/2 > 0$ for all $x \in [0,1]$. This implies that $\alpha_l^*$ is actually monotone increasing when the sample size is large enough. For $p = 1$ and 2, the coefficients of monotone spline satisfy the linear constraints, so Lemma 8 follows. For $p = 3$, we show that the coefficients of $\alpha_l^{*\prime}$ are positive with probability approaching to 1 as the sample size goes to infinity. By Lemma 7, there exists $g_l = \sum_{j=1}^{J_n+1} \gamma_{lj}\tilde{B}$ such that $\|\alpha_l - g_l\|_\infty = O_p\left(N_n^{-p-1}\right)$. We next check the coefficients of $\alpha_l^*$,

$$\begin{aligned}
\beta_l^* &= \left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T(\alpha_l + \varepsilon) \\
&= \left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T(\mathbf{g}_l + \alpha_l - \mathbf{g}_l + \varepsilon) \\
&= \left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T\mathbf{g}_l + \left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T(\alpha_l - \mathbf{g}_l) + \left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T\varepsilon \\
&= I + II + III. \qquad (2.14)
\end{aligned}$$

Since $\mathbf{g}_l = \tilde{\mathbf{B}}_{nl}\gamma_l$, we have

$$I = \left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\gamma_l = \gamma_l. \tag{2.15}$$

Let $|\cdot|_*$ denotes element-wise absolute value of a matrix. With the fact that $\sup_x|\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}|_* = O_P\left(N_n^{-1}\right)$ and $\left\|\alpha_l - g_l^*\right\|_\infty = O_p\left(N_n^{-p-1}\right)$,

$$\sup|II|_* = \sup\left|\left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T\left(\alpha_l - \mathbf{g}_l^*\right)\right|_* = O_p\left(n^{-1}N_n^{-p}\right). \tag{2.16}$$

Then, we write part $III$ as $III = \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\varepsilon$. The $j^{th}$ element of $\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\varepsilon$ is $\xi_j = \frac{1}{n}\sum_{i=1}^n\tilde{B}_{lj}(x_{il})\varepsilon_i$. Let $\xi_{ij} = \tilde{B}_{lj}(x_{il})\varepsilon_i$ and define

$$\tilde{\xi}_{ij} = \tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \le n^\delta\right) - E\left[\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \le n^\delta\right)\right]$$

for some $\delta > 0$.

$$\begin{aligned} E\tilde{\xi}_{ij}^2 &\ge \frac{1}{2}E\left[\tilde{B}_{lj}^2(x_{il})\varepsilon_i^2 I\left(|\varepsilon_i| \le n^\delta\right)\right] - \left[E\left(\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \le n^\delta\right)\right)\right]^2 \\ &= \frac{1}{2}E\left[\tilde{B}_{lj}^2(x_{il})\varepsilon_i^2\right] - \frac{1}{2}E\left[\tilde{B}_{lj}^2(x_{il})\varepsilon_i^2 I\left(|\varepsilon_i| \ge n^\delta\right)\right] \\ &\quad - \left[E\left(\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \ge n^\delta\right)\right)\right]^2, \end{aligned}$$

in which, under condition (A1),

$$E\left[\tilde{B}_{lj}^2(x_{il})\varepsilon_i^2 I\left(|\varepsilon_i| \ge n^\delta\right)\right] = E\left[\tilde{B}_{lj}^2(x_{il})\right]E\left[\varepsilon_i^2 I\left(|\varepsilon_i| \ge n^\delta\right)|X\right]$$

$$\begin{aligned} &\leq E\left[\tilde{B}_{lj}^2(x_{il})\right] E\left(\frac{|\varepsilon_i|^{2+\eta}}{n^{\eta\delta}}|X\right) \\ &\leq \frac{c_1 E\left[\tilde{B}_{lj}^2(x_{il})\right]}{n^{\eta\delta}} \\ &\leq \frac{c_1}{n^{\eta\delta}N_n}. \end{aligned}$$

Moreover,

$$E\left[\tilde{B}_{lj}^2(x_{il})\varepsilon_i^2\right] \geq c_2 E\left[\tilde{B}_{lj}^2(x_{il})\right] \geq \frac{c_3}{N_n},$$

$$\left[E\left(\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \geq n^{\delta}\right)\right)\right]^2 \leq E\left[\tilde{B}_{lj}^2(x_{il})\varepsilon_i^2 I\left(|\varepsilon_i| \geq n^{\delta}\right)\right] \leq \frac{c_1}{n^{\eta\delta}N_n}.$$

One has $E\tilde{\xi}_{ij}^2 \geq c_3/2N_n - 3c_1/2n^{\eta\delta}N_n \geq c_4/N_n$. Then by Minkowski's inequality, for any integer $k \geq 3$,

$$\begin{aligned} E|\tilde{\xi}_{ij}|^k &\leq 2^{k-1}\left\{E\left[\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \leq n^{\delta}\right)\right]^k + \left[E\left(\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \leq n^{\delta}\right)\right)\right]^k\right\} \\ &\leq 2^{k-1}\left[\frac{c_5 n^{\delta k}}{N_n} + \left(\frac{c_5}{N_n}\right)^k\right] \\ &\leq \frac{c_6 n^{\delta k}}{N_n} = \frac{c_6 n^{3\delta(k-2)}k!}{c_4 n^{2\delta(k-3)}k!}\cdot\frac{c_4}{N_n} \\ &\leq \frac{c_6 n^{3\delta(k-2)}}{c_4}k!E\tilde{\xi}_{ij}^2 \\ &= c_r^{k-2}k!E\tilde{\xi}_{ij}^2. \end{aligned}$$

So, Cramer's conditions are satisfied with the Cramer constant $c_r = cn^{3\delta}$ and we can apply Bernstein's inequality to $\sum_{i=1}^n \tilde{\xi}_{ij}$. For any $\varepsilon > 0$ and $n$ is large enough, one has

$$
\begin{aligned}
& P\left(\frac{1}{n}\left|\sum_{i=1}^{n}\tilde{\xi}_{ij}\right| \geq \varepsilon\sqrt{\frac{N_n \log n}{n}}\right) \\
\leq\ & 2\exp\left(-\frac{\varepsilon^2 n N_n \log n}{4\sum_{i=1}^{n} E\tilde{\xi}_{ij}^2 + 2cn^{3\delta}\varepsilon\sqrt{nN_n \log n}}\right) \\
\leq\ & 2\exp\left(-\frac{\varepsilon^2 n N_n \log n}{4nc_4/N_n + 2cn^{3\delta}\varepsilon\sqrt{(nN_n \log n)}}\right) \\
\leq\ & 2\exp\left(-3\log n\right) = 2n^{-3},
\end{aligned}
$$

for $0 < \delta < (p+1)/(3(2p+3))$. Furthermore, by Markov's inequality, for any $\varepsilon > 0$,

$$
\begin{aligned}
& P\left(\frac{1}{n}\left|\sum_{i=1}^{n}\left(\xi_{ij} - \tilde{\xi}_{ij}\right)\right| > \varepsilon\sqrt{\frac{N_n \log n}{n}}\right) \\
\leq\ & \frac{1}{\varepsilon\sqrt{nN_n \log n}} E\left|\sum_{i=1}^{n}\left(\xi_{ij} - \tilde{\xi}_{ij}\right)\right| \\
\leq\ & \frac{1}{\varepsilon\sqrt{nN_n \log n}} \sum_{i=1}^{n} E\left|\xi_{ij} - \tilde{\xi}_{ij}\right| \\
=\ & \frac{1}{\varepsilon\sqrt{nN_n \log n}} \sum_{i=1}^{n} E\left|\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \geq n^{\delta}\right) - E\left[\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \geq n^{\delta}\right)\right]\right| \\
\leq\ & \frac{2}{\varepsilon\sqrt{nN_n \log n}} \sum_{i=1}^{n} E\left|\tilde{B}_{lj}(x_{il})\varepsilon_i I\left(|\varepsilon_i| \geq n^{\delta}\right)\right| \\
\leq\ & \frac{c_7}{\varepsilon\sqrt{(N_n \log n)/n}\, n^{(\eta+1)\delta}} \leq cn^{-3},
\end{aligned}
$$

for $\eta$ that is large enough. Since

$$
\sup\left|\frac{1}{n}\tilde{\mathbf{B}}_{nl}^{T}\varepsilon\right|_{*} = \sup_j \frac{1}{n}\left|\sum_{i=1}^{n}\xi_{ij}\right| \leq \sup_j \frac{1}{n}\left|\sum_{i=1}^{n}\left(\xi_{ij} - \tilde{\xi}_{ij}\right)\right| + \sup_j \frac{1}{n}\left|\sum_{i=1}^{n}\tilde{\xi}_{ij}\right|,
$$

we have

$$\begin{aligned}
& \sum_{i=1}^{n} P\left(\sup\left|\frac{1}{n}\tilde{\mathbf{B}}_{nl}^{T}\varepsilon\right|_{*} \geq 2\varepsilon\sqrt{\frac{N_n\log n}{n}}\right) \\
\leq\ & \sum_{i=1}^{n} P\left(\sup_j \frac{1}{n}\left|\sum_{i=1}^{n}\left(\xi_{ij}-\tilde{\xi}_{ij}\right)\right| \geq \varepsilon\sqrt{\frac{N_n\log n}{n}}\right) \\
& + \sum_{i=1}^{n} P\left(\sup_j \frac{1}{n}\left|\sum_{i=1}^{n}\tilde{\xi}_{ij}\right| \geq \varepsilon\sqrt{\frac{N_n\log n}{n}}\right) \\
\leq\ & \sum_{i=1}^{n}\sum_{j=1}^{J_n+1} P\left(\frac{1}{n}\left|\sum_{i=1}^{n}\left(\xi_{ij}-\tilde{\xi}_{ij}\right)\right| \geq \varepsilon\sqrt{\frac{N_n\log n}{n}}\right) \\
& + \sum_{i=1}^{n}\sum_{j=1}^{J_n+1} P\left(\frac{1}{n}\left|\sum_{i=1}^{n}\tilde{\xi}_{ij}\right| \geq \varepsilon\sqrt{\frac{N_n\log n}{n}}\right) \\
\leq\ & \sum_{i=1}^{n} cN_n n^{-3} < +\infty.
\end{aligned}$$

Therefore by Borel-Cantelli Lemma,

$$\sup|III|_{*} = \sup\left|\left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^{T}\tilde{\mathbf{B}}_{nl}\right)^{-1}\frac{1}{n}\tilde{\mathbf{B}}_{nl}^{T}\tilde{\varepsilon}\right|_{*} = O_p\left(\sqrt{\frac{N_n^3\log n}{n}}\right). \tag{2.17}$$

From equations (2.14), (2.15), (2.16) and (2.17), we have $\sup|\beta_l^* - \gamma_l|_* = \sup|II + III|_* = O_p\left(\sqrt{\frac{N_n^3\log n}{n}}\right)$ and then Lemma 8 follows. ■

### 2.4.2 Proof of Theorems

## Proof of Theorem 1

Define $\tilde{\alpha}_{i,-l} = \sum_{l' \neq l} \tilde{\alpha}_{l'}(x_{il'})$ and $\tilde{\alpha}_{-l} = \left(\tilde{\alpha}_{1,-l}, ..., \tilde{\alpha}_{n,-l}\right)^T$. The first order derivative of the unconstrained estimator is

$$\begin{aligned}
\tilde{\alpha}_l'(x) &= \left[\tilde{\mathbf{B}}_l'(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{y} - \tilde{\alpha}_{-l}) \\
&= \left[\tilde{\mathbf{B}}_l'(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \left(\sum_{l=1}^{d} \alpha_l + \varepsilon - \tilde{\alpha}_{-l}\right) \\
&= \left[\tilde{\mathbf{B}}_l'(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\alpha_l + \varepsilon) + \left[\tilde{\mathbf{B}}_l'(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\alpha_{-l} - \tilde{\alpha}_{-l}) \\
&= \alpha_l^{*\prime}(x) + I(x).
\end{aligned}$$

According to Lemma 6, for any fixed $l = 1, ...d$,

$$\sup_{x \in [0,1]} \left|\alpha_l^{*\prime}(x) - \alpha_l'(x)\right| = O_p\left(N_n^{\frac{1}{2}-p} + N_n^{\frac{3}{2}}/\sqrt{n}\right). \tag{2.18}$$

Then, for part $I(x)$, we have

$$\sup_x I(x) \leq \sqrt{\sup_x \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l'(x)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l'(x)\right]} \times \sqrt{\left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T (\alpha_{-l} - \tilde{\alpha}_{-l})\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T (\alpha_{-l} - \tilde{\alpha}_{-l})\right]},$$

in which, the first term of right hand side is $O_P(\sqrt{N_n^3/n})$ by (2.10) in the proof of Lemma 6, and

$$
\begin{aligned}
& \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^{T}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)\right]^{T}\left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^{T}\tilde{\mathbf{B}}_{nl}\right)^{-1}\left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^{T}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)\right] \\
\leq & \frac{1}{n\lambda_{\min}}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)^{T}\tilde{\mathbf{B}}_{nl}\tilde{\mathbf{B}}_{nl}^{T}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right) \\
\leq & \frac{\lambda_{\max}}{\lambda_{\min}}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)^{T}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right) \\
= & O_{p}(nN_{n}^{-2p-2}+N_{n}),
\end{aligned}
$$

in which $\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)^{T}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)=O_{p}\left(nN_{n}^{-2p-2}+N_{n}\right)$ by Theorem 1 of Huang (1998). Therefore, we have

$$
\sup_{x} I(x)=O_{p}\left(\sqrt{N_{n}^{1-2p}+\frac{N_{n}^{4}}{n}}\right). \tag{2.19}
$$

Finally, Theorem 1 follows from equations (2.18), (2.19) and condition (A5). ■

## Proof of Theorem 2

By definition, we have

$$
\begin{aligned}
\tilde{\beta}_{l} &= \left(\tilde{\mathbf{B}}_{nl}^{T}\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^{T}\left(\mathbf{y}-\tilde{\alpha}_{-l}\right) \\
&= \left(\tilde{\mathbf{B}}_{nl}^{T}\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^{T}\left(\sum_{l=1}^{d}\alpha_{l}+\varepsilon-\tilde{\alpha}_{-l}\right) \\
&= \left(\tilde{\mathbf{B}}_{nl}^{T}\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^{T}\left(\alpha_{l}+\varepsilon\right)+\left(\tilde{\mathbf{B}}_{nl}^{T}\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^{T}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right) \\
&= \beta_{l}^{*}+I.
\end{aligned}
$$

According to Lemma 8, for any fixed $l=1,\ldots d$, the coefficient $\beta_{l}^{*}$ satisfies the monotone constraints with probability approaching to one. Furthermore,

$$\sup\left|\left(\tilde{\mathbf{B}}_{nl}^{T}\tilde{\mathbf{B}}_{nl}\right)^{-1}\right|_{*}=O_{P}\left(N_{n}/n\right),$$

$$\begin{aligned}\sup\left|\tilde{\mathbf{B}}_{nl}^{T}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)\right|_{*} &\leq \sqrt{\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)^{T}\tilde{\mathbf{B}}_{nl}\tilde{\mathbf{B}}_{nl}^{T}\left(\alpha_{-l}-\tilde{\alpha}_{-l}\right)}\\ &= O_{p}\left(\sqrt{\left(nN_{n}^{-2p-2}+N_{n}\right)n/N_{n}}\right).\end{aligned}$$

So, $\sup|I|_{*}=O_{p}\left(\sqrt{N_{n}^{-2p-1}+N_{n}^{2}/n}\right)$ and then Theorem 2 directly follows. ■

## Proof of Theorem 3

Let $\varepsilon_{0}=\min\left\{\alpha_{l}'\left(x\right),x\in[0,1]\right\}$. Under condition (A3), each $\alpha_{l}$ is monotone increasing, then we have $\alpha_{l}'\left(x\right)\geq\varepsilon_{0}\geq c_{4}>0$, for all $x\in[0,1]$. Theorem 1 states that there exists an integer $n\left(\varepsilon_{0}\right)$, such that when $n>n\left(\varepsilon_{0}\right)$, $\sup_{x\in[0,1]}\left|\tilde{\alpha}_{l}'\left(x\right)-\alpha_{l}'\left(x\right)\right|<\varepsilon_{0}/2$. Therefore when $n>n\left(\varepsilon_{0}\right)$, $\tilde{\alpha}_{l}'\left(x\right)\geq\alpha_{l}'\left(x\right)-\varepsilon_{0}/2\geq\varepsilon_{0}/2>0$ for all $x\in[0,1]$. This implies that $\tilde{\alpha}_{l}\left(x\right)$ is actually monotone increasing when the sample size is large enough. For $p\leq 3$, Theorem 2 indicates that the coefficients of $\tilde{\alpha}_{l}$ satisfy the linear constraints for large sample size. It implies that the constrained estimator $\hat{\alpha}_{l}$ and unconstrained estimator $\tilde{\alpha}_{l}$ are identical when sample size is large enough, for $p\leq 3$. Therefore, $\hat{\alpha}_{l}$ enjoys the same asymptotic properties with $\tilde{\alpha}_{l}$. According to Theorem 1 in Huang (1998), one has $\|\tilde{\alpha}_{l}-\alpha_{l}\|=O_{p}\left(\sqrt{N_{n}^{-2p-2}+N_{n}/n}\right)$, for $l=1,\cdots,d$. The constrained estimator $\hat{\alpha}_{l}$ also enjoys this optimal rate of convergence and then Theorem 3 follows. ■

## 2.5 Tables and Figures

Table 2.1: Monotone increasing functions used in the simulation study.

| |
|---|
| $\alpha_1(x) = 2x - 1$ |
| $\alpha_2(x) = \frac{3}{2}x^2 - \frac{1}{2}$ |
| $\alpha_3(x) = e^x + 1 - e$ |
| $\alpha_4(x) = \frac{1}{2}(\log x + 1)$ |
| $\alpha_5(x) = 2x + \frac{\sin(2\pi x)}{2\pi} - 1$ |

Table 2.2: Simulation results: averaged integrated squared errors (AISE) using unconstrained quadratic spline (UQS), constrained quadratic spline (CQS), and pool adjacent violator algorithm (PAVA).

| Outliers (N/Y) | Method | n | $\alpha_1(x)$ | $\alpha_2(x)$ | $\alpha_3(x)$ | $\alpha_4(x)$ | $\alpha_5(x)$ |
|---|---|---|---|---|---|---|---|
| N | UQS | 50 | 0.0384 | 0.0311 | 0.0301 | 0.0424 | 0.0345 |
| | | 100 | 0.0152 | 0.0133 | 0.0117 | 0.0250 | 0.0143 |
| | | 200 | 0.0101 | 0.0087 | 0.0095 | 0.0194 | 0.0101 |
| | | 500 | 0.0037 | 0.0043 | 0.0036 | 0.0148 | 0.0044 |
| | CQS | 50 | 0.0306 | 0.0236 | 0.0244 | 0.0315 | 0.0280 |
| | | 100 | 0.0140 | 0.0111 | 0.0102 | 0.0207 | 0.0128 |
| | | 200 | 0.0090 | 0.0072 | 0.0085 | 0.0173 | 0.0092 |
| | | 500 | 0.0037 | 0.0039 | 0.0033 | 0.0141 | 0.0043 |
| | PAVA | 50 | 0.0694 | 0.0658 | 0.0634 | 0.0791 | 0.0781 |
| | | 100 | 0.0353 | 0.0324 | 0.0332 | 0.0356 | 0.0344 |
| | | 200 | 0.0211 | 0.0194 | 0.0211 | 0.0217 | 0.0220 |
| | | 500 | 0.0098 | 0.0101 | 0.0102 | 0.0100 | 0.0103 |
| Y | UQS | 50 | 0.5237 | 0.4343 | 0.4308 | 0.4679 | 0.3951 |
| | | 100 | 0.0988 | 0.1045 | 0.0876 | 0.1037 | 0.0971 |
| | | 200 | 0.0330 | 0.0322 | 0.0371 | 0.0463 | 0.0362 |
| | | 500 | 0.0067 | 0.0081 | 0.0075 | 0.0174 | 0.0085 |
| | CQS | 50 | 0.1971 | 0.2132 | 0.1847 | 0.2234 | 0.1368 |
| | | 100 | 0.0673 | 0.0602 | 0.0555 | 0.0624 | 0.0577 |
| | | 200 | 0.0244 | 0.0203 | 0.0251 | 0.0336 | 0.0265 |
| | | 500 | 0.0063 | 0.0062 | 0.0064 | 0.0156 | 0.0078 |
| | PAVA | 50 | 0.5296 | 0.5604 | 0.4484 | 0.5336 | 0.4682 |
| | | 100 | 0.1764 | 0.1592 | 0.1455 | 0.1602 | 0.1579 |
| | | 200 | 0.0521 | 0.0522 | 0.0566 | 0.0566 | 0.0563 |
| | | 500 | 0.0177 | 0.0168 | 0.0172 | 0.0173 | 0.0189 |

Table 2.3: Norwegian Farm data: averaged mean squared estimation errors (AMSEE) and averaged mean squared prediction errors (AMSPE) from constrained (C) and unconstrained (U) polynomial spline methods. Both linear ($p = 1$) and cubic ($p = 3$) splines are considered with equally spaced (ES) or equal quantile (EQ) knots.

| Knots | Method | $p$ | AMSEE | AMSPE |
|---|---|---|---|---|
| ES | U | 1 | 0.0313 | 0.1501 |
| | C | 1 | 0.0319 | 0.1479 |
| | U | 3 | 0.0220 | 0.4648 |
| | C | 3 | 0.0723 | 0.3330 |
| EQ | U | 1 | 0.0897 | 0.1925 |
| | C | 1 | 0.0908 | 0.1867 |
| | U | 3 | 0.0234 | 0.1550 |
| | C | 3 | 0.0299 | 0.1487 |

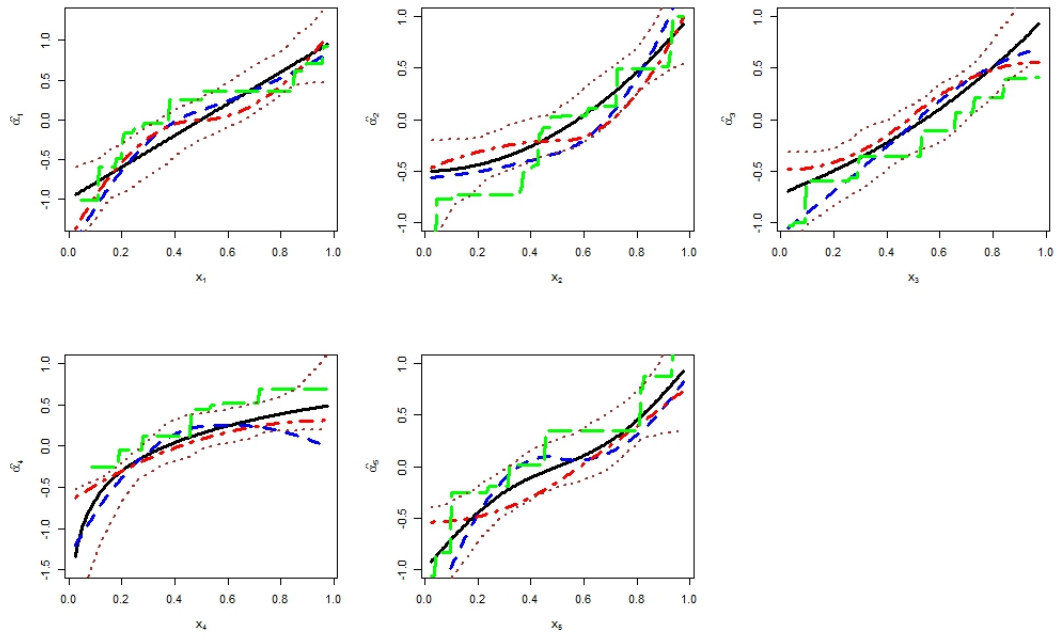

Figure 2.1: Simulation results of the monotone increasing functions **without outliers** and sample size $n = 50$. In each plot, the solid black line represents the true curve, while the dashed blue, dot-dashed red and long-dashed green lines represent the typically fitted curves obtained using unconstrained quadratic spline (UQS), constrained quadratic spline (CQS) and pool adjacent violator algorithm (PAVA), respectively. The dotted lines represent the 95% empirical point-wise confidence intervals using the CQS method.

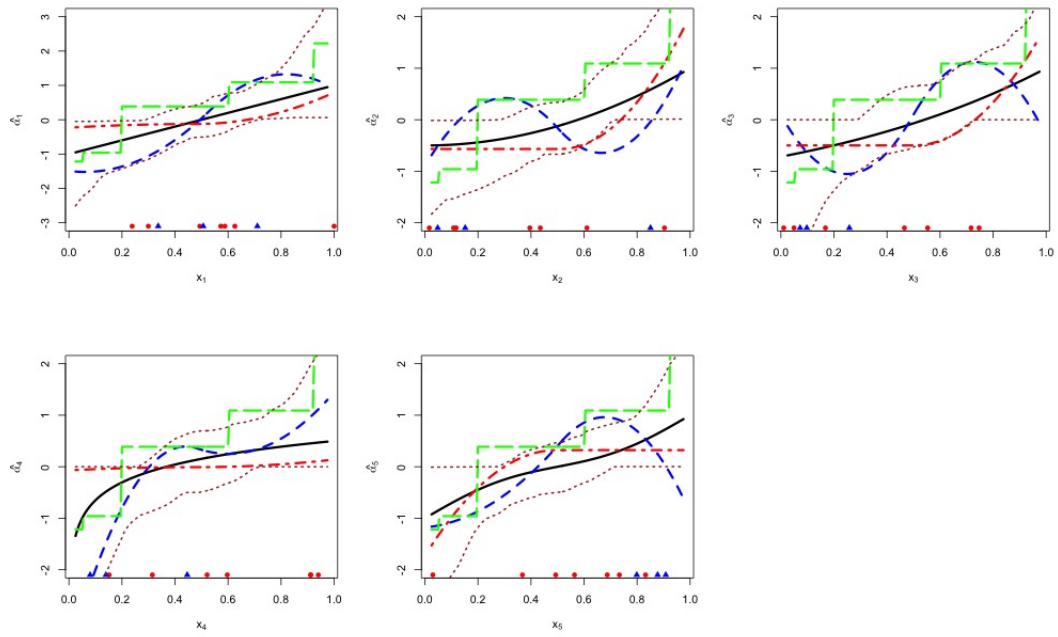

Figure 2.2: Simulation results of the monotone increasing functions **with outliers** and sample size $n = 50$. The solid black line represents the true curve, while the dashed blue, dot-dashed red and long-dashed green lines represent the typically fitted curves obtained using unconstrained quadratic spline (UQS), constrained quadratic spline (CQS) and pool adjacent violator algorithm (PAVA), respectively. The dotted lines represent the 95% empirical point-wise confidence intervals using the CQS method. The solid triangles and circles locate the positions of ten outliers: the triangle (or circle) indicates the location where the response is manually decreased (or increased) by five.

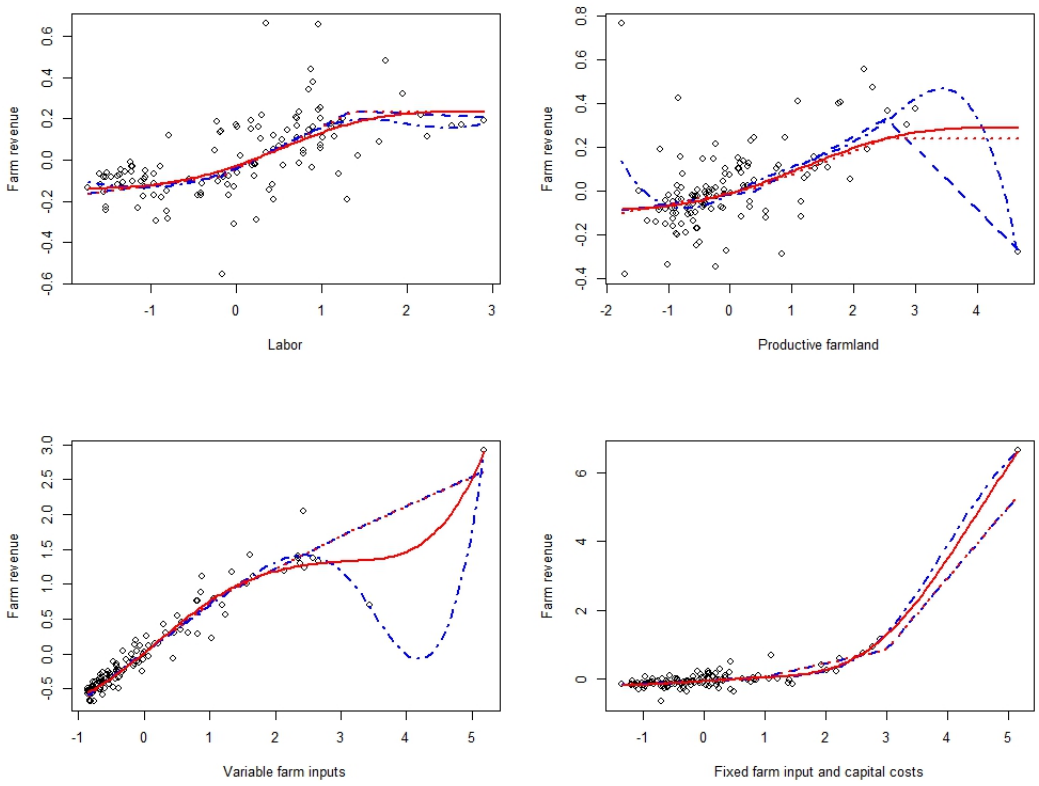


Figure 2.3: Norwegian Farm data: fitted results for each component using **equally spaced knots**. The black circle represents the pseudo response for each predictor. The dashed and dot-dashed blue lines represent fitted curves using unconstrained method for $p = 1$ and $p = 3$, respectively. The dotted and solid red lines represent fitted curves using constrained method for $p = 1$ and $p = 3$, respectively.

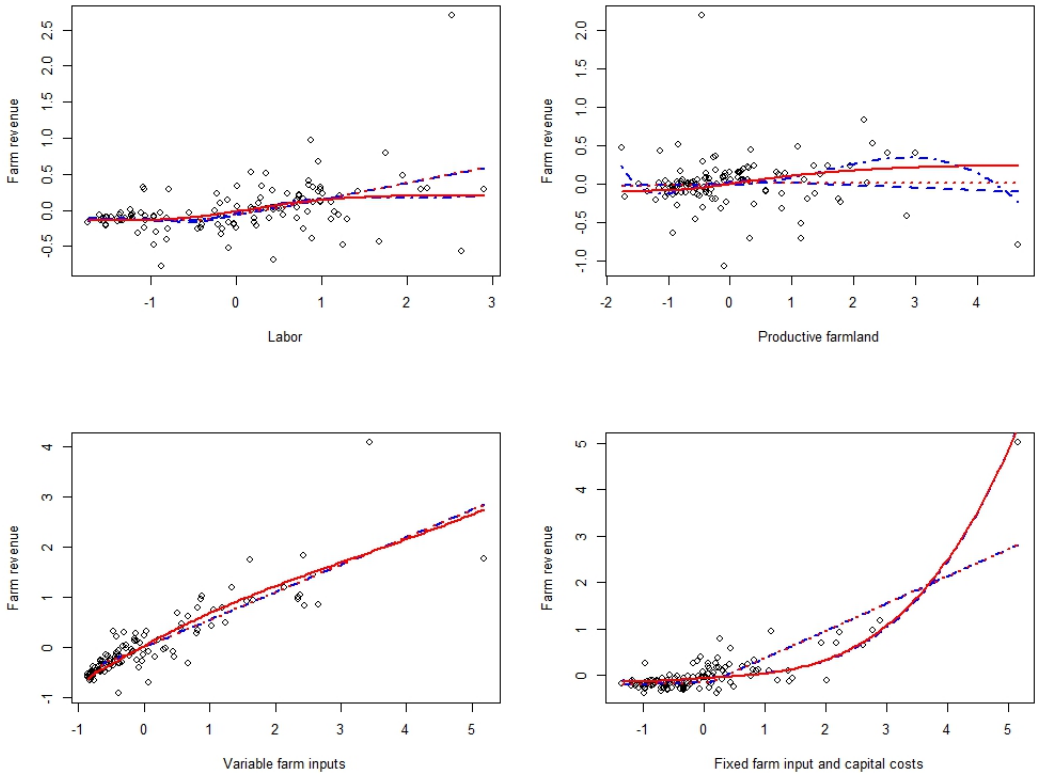


Figure 2.4: Norwegian Farm data: fitted results for each component using **equal quantile knots**. The black circle represents the pseudo response for each predictor. The dashed and dot-dashed blue lines represent fitted curves using unconstrained method for $p = 1$ and $p = 3$, respectively. The dotted and solid red lines represent fitted curves using constrained method for $p = 1$ and $p = 3$, respectively.

# 3 Estimation of Additive Frontier Functions with Shape Constraints

## 3.1 Additive Frontier Models

In this chapter, we extend the constrained polynomial spline method described in Chapter 2 to estimate the production frontier functions. Consider a nonnegative vector $(\mathbf{x}, y) \in \mathscr{R}_+^d \times \mathscr{R}_+$, where $\mathbf{x}$ represents the $d$ inputs and $y$ represents the output of a production unit. The production set is defined as the set of physically attainable points $(\mathbf{x}, y)$, i.e., $\Psi = \big\{(\mathbf{x}, y) \in \mathscr{R}_+^d \times \mathscr{R}_+ |\ \mathbf{x} \text{ can produce } y\big\}$. The production frontier function of $\Psi$ is defined as $\rho(\mathbf{x}) = \sup\{y, (\mathbf{x}, y) \in \Psi\}$ that is the upper boundary of the production set. In production analysis, the main interest lies on the specification and estimation of the production frontier function $\rho(\cdot)$ given a random sample of the production units $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$.

First of all, we develop an additive frontier model that is inspired by the deterministic nonparametric regression frontier model proposed in Martins-Filho and Yao (2007). The construction of the additive frontier model is as follows. Suppose there are $n$ independent distributed observations generated from the additive frontier model

$$Y_i = \rho(\mathbf{X_i}) R_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where $Y_i$ is the output variable and $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^T$ are the input variables of the $i$-th observation. We assume that the frontier function $\rho(\mathbf{X}_i)$ is of an additive structure and

written as $\rho(\mathbf{X}_i) = \rho_0 + \rho_1(X_{i1}) + \ldots + \rho_d(X_{id})$, where $\rho_0$ is an unknown intercept and $\{\rho_l(\cdot)\}_{l=1}^d$ are unknown univariate nonparametric functions that quantify the effect of input variables on the maximum output. For model identification, we assume that each additive component $\rho_l(\cdot)$ is theoretically centered with $E[\rho_l(X_l)] = 0$, for $l = 1, \ldots, d$. The efficiency $R$ is an unobserved random variable that takes values in $[0,1]$. The larger value of $R$ indicates more efficient production because the realized output $Y$ is closer to the production frontier $\rho(\mathbf{X})$. In a special case with $R = 1$, the maximum output is obtained. Furthermore, we assume that $E(R|\mathbf{X}) = \mu_R \in (0,1)$ and $Var(R|\mathbf{X}) = \sigma_R^2$. The parameter $\mu_R$ is viewed as the mean efficiency given the production set and $\sigma_R$ is the scale parameter for the distribution of $R$.

The proposed additive frontier model (3.1) has a number of desirable properties. First, there is no restriction on the production set $\Psi$ since the frontier function $\rho(\cdot)$ is not constrained to a specific parametric family. Furthermore, compared with the nonparametric frontier model described in Martins-Filho and Yao (2007), our proposed additive frontier model allows that the contribution of each input variable is additive. Therefore, the proposed frontier model enjoys the advantages of additive models and is especially less affected by the curse of dimensionality.

In this chapter, our main interest lies in the estimation of frontier function $\rho(\cdot)$ of model (3.1). To begin with, we rewrite model (3.1) as

$$Y_i = \rho(\mathbf{X}_i)\mu_R + \rho(\mathbf{X}_i)\varepsilon_i = m(\mathbf{X}_i) + \rho(\mathbf{X}_i)\varepsilon_i, \quad i = 1, \ldots, n, \tag{3.2}$$

where $m(\mathbf{X}_i) = \rho(\mathbf{X}_i)\mu_R = m_0 + m_1(X_{i1}) + \ldots + m_d(X_{id})$ and $m_j = \rho_j \mu_R$, for $j =$

$0,\ldots,d$. The error term $\varepsilon_i = R_i - \mu_R$ is with $E(\varepsilon_i|\mathbf{X}_i) = 0$ and $Var(\varepsilon_i|\mathbf{X}_i) = \sigma_R^2$. In the regression model (3.2), the additive regression function $m_j$ characterizes the shape of $\rho_j$. These two functions are different only by a scale parameter $\mu_R$ that represents the mean efficiency. In the following section, we propose a two-step polynomial spline method to estimate the additive frontier functions $\{\rho_l(\cdot)\}_{l=1}^d$. In the first step, we estimate the additive regression functions $\{m_l(\cdot)\}_{l=1}^d$ using the polynomial spline method. The mean functions $\{m_l(\cdot)\}_{l=1}^d$ describe the shape of frontier functions. In addition, production theory in econometrics often imposes shape constraints on the additive frontier functions $\{\rho_l(\cdot)\}_{l=1}^d$ such as monotonicity or concavity. Therefore, in order to capture the shape of frontiers more accurately, we also consider incorporating shape constraints in the first estimation step. The second step is for estimating the location of the frontier function that is associated with the mean efficiency $\mu_R$. Finally, with the fact that $m(\cdot) = \rho(\cdot)\mu_R$, the estimator of frontier functions can be obtained by combining estimators in the previous two steps.

## 3.2 Methodology and Theory

### 3.2.1 Proposed Method

We propose to estimate the frontier functions in two easily implementable steps. In the first step, we estimate the mean functions $\{m_l(\cdot)\}_{l=1}^d$ in model (3.2) using the polynomial spline method. Let $u_n = \{0 = u_0 < u_1 < \cdots < u_{N_n} < u_{N_n+1} = 1\}$ be a knot sequence on the interval $[0,1]$, with $N_n$ interior knots. With the knot sequence $u_n$, the

interval $[0,1]$ is partitioned into $N_n+1$ smaller intervals. The polynomial splines of order $p+1$ are polynomial functions with degree $p$ (or less) on the partitioned intervals and $(p-1)$-times differentiable at the interior knots. Follow the notation in Chapter 2, we denote the space of polynomial splines with degree $p$ based on the knots $u_n$ as $G^p = G^p([0,1],u_n)$ that is with dimension $N_n+p+1$. Then we denote B-spline basis of $G^p$ as $\tilde{\mathbf{B}}(x) = \left(\tilde{B}_1(x),\ldots,\tilde{B}_{J_n+1}(x)\right)^T$, where $J_n = N_n+p$. Due to the fact that $\sum_{j=1}^{J_n+1}\tilde{B}_j(x)=1$, without loss of generality, we focus on the first $J_n$ basis and create empirically centered B-spline basis by taking $B_j = \tilde{B}_j - \frac{1}{n}\sum_{i=1}^n \tilde{B}_j(x_i)$. Let $\mathbf{B}_l(x) = (B_{l1}(x),\ldots,B_{lJ_n}(x))^T$ be the centered basis for the input variable $X_{l,}$ for $l=1,\ldots,d$. Under the assumption that the additive frontier functions $\{\rho_l(\cdot)\}_{l=1}^d$ are theoretically centered, the intercept term $m_0$ can be consistently estimated as $\hat{m}_0 = \bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. Then we approximate the theoretically centered nonparametric function $m_l(\cdot)$ by a linear combination of the centered B-spline basis, i.e., $m_l(x) \approx \mathbf{B}_l^T(x)\beta_l$, where $\beta_l = \left(\beta_{l1},\ldots,\beta_{lJ_n}\right)^T$ is a set of coefficients. Let $\mathbf{Y}^* = (Y_1-\hat{m}_0,\ldots,Y_n-\hat{m}_0)^T$ and $\mathbf{B}_n = (\mathbf{B}_{n1},\ldots,\mathbf{B}_{nd})$, where $\mathbf{B}_{nl} = (\mathbf{B}_l(x_{1l}),\ldots,\mathbf{B}_l(x_{nl}))^T$. The polynomial spline method (Stone 1985 and Huang 1998) estimates the unknown coefficients $\beta = \left(\beta_1^T,\ldots,\beta_d^T\right)^T$ by minimizing the sum of squares, i.e.,

$$\tilde{\beta} = \underset{\beta\in R^{dJ_n}}{\arg\min}\,(\mathbf{Y}^*-\mathbf{B}_n\beta)^T(\mathbf{Y}^*-\mathbf{B}_n\beta) = \left(\mathbf{B}_n^T\mathbf{B}_n\right)^{-1}\mathbf{B}_n^T\mathbf{Y}^*. \tag{3.3}$$

Then the unknown function $m_l(\cdot)$ is given as

$$\tilde{m}_l(x) = \mathbf{B}_l^T(x)\tilde{\beta}_l, \tag{3.4}$$

for $l = 1, \ldots, d$.

The traditional polynomial spline estimator (3.4) enjoys the same optimal rate of convergence as an univariate nonparametric function estimator, however, it can not give shape constrained estimates, such as monotone increasing or concave function estimates. To capture the shape of frontier functions more accurately, we consider to incorporate shape constraints that guarantee our estimates to be monotone or concave. In the following, Lemma 9 gives a sufficient condition for a polynomial spline to be monotone increasing that is considered in Chapter 2. In addition to the monotone constraints, we are also interested in imposing concave constraints to control the rate of increasing since the rate of technology productivity change is commonly decreasing in practice. As a result, the sufficient conditions for a polynomial spline to be concave are developed in Lemmas 10 and 11.

**Lemma 9** *A sufficient condition for a polynomial spline* $g(x) = \mathbf{B}^T(x)\beta$ *, where* $\mathbf{B}(x)$ *is the empirically centered B-spline basis of* $G^p$*, to be monotone increasing is that its coefficients satisfy the following conditions:* $\beta_1 \geq 0$ *and* $\beta_j \geq \beta_{j-1}$*, for* $j = 2, ..., N_n + p$.

**Lemma 10** *A sufficient condition for a linear spline* $g(x) = \mathbf{B}^T(x)\beta$ *, where* $\mathbf{B}(x)$ *is the empirically centered B-spline basis of* $G^1$*, to be concave is that its coefficients satisfy the following conditions:* $\beta_1 - \beta_0 \leq \beta_0$ *and* $\beta_j - \beta_{j-1} \leq \beta_{j-1} - \beta_{j-2}$*, for* $N \geq 2$*,* $j = 2, ..., N_n$ *or* $\beta_1 - \beta_0 \leq \beta_0$*, for* $N = 1$.

**Lemma 11** *A sufficient condition for a polynomial spline* $g(x) = \mathbf{B}^T(x)\beta$ *with degree* $p \geq 2$*, where* $\mathbf{B}(x)$ *is the empirically centered B-spline basis of* $G^p$*, to be concave is that its coefficients satisfy the following conditions:*

$$\beta_{-p+2}-\beta_{-p+1} \leq 2\beta_{-p+1},$$
$$\beta_j-\beta_{j-1} \leq (j+p)\,\beta_{j-1}/(j+p-1)-\beta_{j-2}, \textit{for } j=-p+3,...,-1$$
$$\beta_0-\beta_{-1} \leq p\beta_{-1}/(p-1)-\beta_{-2},$$
$$\beta_j-\beta_{j-1} \leq \beta_{j-1}-\beta_{j-2}, \textit{for } j=1,...,N_n-p+1$$
$$\beta_{N_n-p+2}-\beta_{N_n-p+1} \leq (p-1)\left(\beta_{N_n-p+1}-\beta_{N_n-p}\right)/p,$$
$$\beta_j-\beta_{j-1} \leq (N_n+1-j)\left(\beta_{j-1}-\beta_{j-2}\right)/(N_n+2-j), \textit{for } j=N_n-p+3,...,N_n$$

For the sake of simplicity, we use $C_M$ and $C_C$ to represent the set of spline coefficients that satisfies the monotone increasing conditions and concave conditions, respectively. In Chapter 2, we have developed a one-step backfitted constrained polynomial spline method to estimate monotone additive regression functions. In this chapter, we adopt the same estimation method but with more options of shape constraints. With the traditional unconstrained estimator $\{\tilde{m}_l\}_{l=1}^d$, we define the pseudo response $Y^*_{i,-l} = Y^*_i - \sum_{l'\neq l}\tilde{m}_{l'}(x_{il'})$ and $\mathbf{Y}^*_{-l} = \left(Y^*_{1,-l},...,Y^*_{n,-l}\right)^T$, $l=1,\ldots,d$. Then to produce a shape constrained estimator, we propose to estimate the coefficients by minimizing the following constrained least squares

$$\hat{\beta}_l = \underset{\beta_l}{\arg\min}\left(\mathbf{Y}^*_{-l}-\mathbf{B}_{nl}\beta_l\right)^T\left(\mathbf{Y}^*_{-l}-\mathbf{B}_{nl}\beta_l\right), \quad \text{subject to } \beta_l \in C_l, \tag{3.5}$$

where $C_l$ is the set of shape constraints of the spline coefficients for the $l$-th additive component. Taking $C_l = C_M$ gives a monotone estimator. In addition, to ensure monotonicity and concavity simultaneously, we propose to impose the shape constraints $C_l = C_M \cap C_C$ to guarantee a monotone and concave estimator. Furthermore, we are also able to impose different sets of shape constraint for different additive components. As a result, the

shape constrained polynomial spline estimator of $m_l(\cdot)$ is obtained by

$$\hat{m}_l(x) = \mathbf{B}_l^T(x)\hat{\beta}_l, \tag{3.6}$$

for $l = 1,\ldots,d$. Then the mean function $m(\cdot)$ is naturally estimated as $\hat{m}(\mathbf{x}) = \hat{m}_0 + \sum_{l=1}^{d} \hat{m}_l(x_l)$.

In the second step, we observe that the mean function $m(\cdot)$ and the frontier function $\rho(\cdot)$ are different only by the parameter $\mu_R$ that is associated with the location of the frontier function. So we consider to estimate the parameter $\mu_R$. Notice that $Y/m(\mathbf{X}) = \rho(\mathbf{X})R/\rho(\mathbf{X})\mu_R = R/\mu_R$ and its supremum is $1/\mu_R$ since the maximum value of the efficiency variable $R$ is 1. Therefore, we propose to estimate the mean efficiency by taking

$$\hat{\mu}_R = \left[\max_i \left(Y_i/\hat{m}(\mathbf{X}_i)\right)\right]^{-1}. \tag{3.7}$$

In addition, since this estimator is sensitive to the outliers, in our implementation, we propose a robust estimator of $\mu_R$. The details for the robust modification are given in section 3.2.3.

Finally, one observes that the regression model (3.2) implies that $m_l(x_l) = \rho_l(x_l)\mu_R$. Therefore, the additive frontier functions $\rho_l$ can be estimated as $\hat{\rho}_l(x_l) = \hat{m}_l(x_l)/\hat{\mu}_R$, for $l = 0,\ldots,d$, and $\hat{\rho}(\mathbf{x}) = \hat{\rho}_0 + \sum_{l=1}^{d} \hat{\rho}_l(x_l)$.

The proposed two-step polynomial spline estimation method gives a smooth estimator of the frontier function and is easy to implement. Compared with the local linear regression method given in Martins-Filho and Yan (2007), our proposed method is computationally easier and faster. Furthermore, the incorporation of monotonicity or

concavity constraints allows us to capture the shape of frontiers more accurately. In addition, as illustrated in Chapter 2, the constrained polynomial spline method is more robust to the outliers than the unconstrained method. Most importantly, our proposed method is applicable for the multi-dimensions where there are multiple input variables and less affected by the curse of dimensionality.

### 3.2.2 Asymptotic Properties

In this section, we establish the asymptotic properties of our proposed frontier estimator. For the asymptotic analysis, the following assumptions are required:

(A1) The input variables $\mathbf{X}_i$ are i.i.d. distributed on a compact support. Without loss of generality, we assume that the support is $[0,1]^d$. Its density function, denoted by $f(\mathbf{x})$, is continuous and $0 < c_1 \leq f(\mathbf{x}) \leq c_2 < \infty$, for $\mathbf{x} \in [0,1]^d$ and positive constants $c_1$ and $c_2$.

(A2) The efficiency variables $\{R_i\}_{i=1}^n$ are i.i.d. distributed with $E(R_i|\mathbf{X_i}) = \mu_R$ and $Var(R_i|\mathbf{X}_i) = \sigma_R^2 < +\infty$. There exists positive constant $c_3$, such that $E\left(|R_i - \mu_R|^{2+\eta}|\mathbf{X}_i\right) \leq c_3$ a.s. for some $\eta > 0$.

(A3) For the set of knots $\{0 = u_0 < u_1 < \cdots < u_{N_n} < u_{N_n+1} = 1\}$, there exists a constant $c_4 > 0$, such that $\max(u_{j+1} - u_j, j = 0, \ldots, N_n) / \min(u_{j+1} - u_j, j = 0, \ldots, N_n) \leq c_4$.

(A4) The number of interior knots $N_n$ satisfies

$$N_n \to +\infty \text{ and } \frac{N_n^4}{n} \to 0, \text{ as } n \to +\infty.$$

(A5) For each $l = 1, \ldots, d$, the additive frontier function $\rho_l$ is monotone increasing and $(p+1)$-times continuously differentiable for some integer $p \geq 1$. Furthermore, we assume that there exists a constant $c_5 > 0$, such that $\rho_l'(x) \geq c_5$, for $x \in [0,1]$.

(A5*) For each $l = 1, \ldots, d$, the additive frontier function $\rho_l$ is concave and $(p+2)$-times continuously differentiable for some integer $p \geq 1$. Furthermore, we assume that there exists a constant $c_6 < 0$, such that $\rho_l^{(2)}(x) \leq c_6$, for $x \in [0,1]$.

Assumption (A1) is the same as condition 1 in Stone (1985). Condition (A2) requires that the unobserved efficiency random variables are i.i.d. distributed with a common distribution. Condition (A3) assumes that the interior knots are pseudo equally spaced in the interval $[0,1]$. This condition is also considered in Chapter 2 and Huang (1998). The conditions for the number of interior knots and samples size are give in the assumption (A4). Additionally, for the monotone constrained estimator, condition (A5) assumes that each additive frontier function is monotone increasing and its first order derivative is lower bounded. Similarly, for the concave constrained estimator, condition (A5*) requires that each additive frontier function is concave and its second order derivative is upper bounded.

Let $\tilde{m}_l$ and $\hat{m}_l$ represent the unconstrained and shape constrained (monotone or concave constrained) estimators of $m_l$, respectively. The corresponding estimators of $\mu_R$ and $\rho$ are $\tilde{\mu}_R$ (without constraint) or $\hat{\mu}_R$ (with constraint) and $\tilde{\rho}$ (without constraint) or $\hat{\rho}$ (with constraint), respectively. The following theorems are developed for both unconstrained and shape constrained estimators.

**Theorem 12** *Under regularity conditions (A1)-(A4), one has, for* $l = 1, \ldots, d$,

$$\sup_x |\tilde{m}_l(x) - m_l(x)| = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right).$$

**Theorem 13** *Let* $L_n$ *be a sequence such that* $L_n > 0$ *and* $L_n$ *converges to* 0 *as the sample size goes to infinity. Suppose* $\sup_x |\tilde{m}(x) - m(x)| = O_p(L_n)$ *and* $1 - \max_i R_i = O_p(L_n)$, *then one has*

$$|\tilde{\mu}_R - \mu_R| = O_p(L_n).$$

**Theorem 14** *Under the assumptions given in Theorem 13, one has,*

$$\sup_x |\tilde{\rho}(x) - \rho(x)| = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right).$$

The results in Theorems 12 and 13 refer to the estimators $\tilde{m}_l$ and $\tilde{\mu}_R$ that are obtained in the first and second estimation steps, respectively. The asymptotic behavior of our main interest $\tilde{\rho}$ is a combination of the results in these two steps and its characterization is given in Theorem 14. Note that in Theorems 13 and 14, an additional restriction on the distribution of the efficiency variable $1 - \max_i R_i = O_p(L_n)$ is required.

**Theorem 15** *Under regularity conditions (A1)-(A5) (for monotone constrained estimator) or (A1)-(A4), (A5*) (for concave constrained estimator), for* $p \leq 3$ *and* $l = 1, \ldots d$, *one has*

$$\sup_x |\hat{m}_l(x) - m_l(x)| = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right).$$

*Furthermore, assume that* $1-\max_i R_i = O_p\left(\sqrt{N_n^2/n+N_n^{-2p-1}}\right)$, *then one has*

$$|\hat{\mu}_R-\mu_R| = O_p\left(\sqrt{N_n^2/n+N_n^{-2p-1}}\right)$$

*and*

$$\sup_x |\hat{\rho}(x)-\rho(x)| = O_p\left(\sqrt{N_n^2/n+N_n^{-2p-1}}\right).$$

Theorem 15 is developed for our proposed two-step shape constrained polynomial spline estimator. We prove that when the shape constraints are correctly specified, the constrained estimator of the mean function $\hat{m}_l$ has the same asymptotic properties with the unconstrained estimator $\tilde{m}_l$, for $l = 1,\ldots,d$. With this property and the asymptotic results described in Theorems 13 and 14, we naturally obtain the asymptotic behavior of the parameter estimator $\hat{\mu}_R$ and our final frontier estimator $\hat{\rho}$.

### 3.2.3 Implementation

In section 3.2.1, we estimate the mean efficiency $\mu_R$ in the second step by taking $\hat{\mu}_R = \left[\max_{1\le i\le n}\left(Y_i/\hat{m}(\mathbf{X}_i)\right)\right]^{-1}$. We observe that this estimator is sensitive to the extreme large values or outliers in the data sets. Therefore, in the implementation, we propose a modified estimator that is more robust to outliers. Let $Q_1$ and $Q_3$ be the first and third quantiles of $\max_{1\le i\le n}\left(Y_i/\hat{m}(\mathbf{X}_i)\right)$, respectively. The corresponding interquantile range (IQR) is defined as $IQR = Q_3 - Q_1$. The data points that are beyond 1.5 times interquantile range (IQR) are regarded as outliers. Then the adjusted estimator is $\hat{\mu}_R = \left[\max_{i\in S}\left(Y_i/\hat{m}(\mathbf{X}_i)\right)\right]^{-1}$, where $S = [Q_1 - 1.5IQR, Q_3 + 1.5IQR]$. In our simulation

studies, we implement this robust modification and the numerical results show that this method is effective especially when the data has outliers.

In addition, in our proposed polynomial spline estimation method, the appropriate selection of the knot sequences is crucial. To reduce computational complexity, we use the same knot sequences for both initial and constrained polynomial spline estimation procedures. The knot sequences are equally spaced in the range of each input variable. The same number of interior knots $N_n$ is used for all input variables. For the simulation studies in section 3.3.1, we adopt the optimal $N_n$ that is selected using the Bayes Information Criterion (BIC). To be specific, let $\hat{Y}_i(N_n)$ denote the estimator of the $i$-th observation using $N_n$ as the number of interior knots. Then the selected $\hat{N}_n$ is the one that minimizes the BIC value, i.e., $\hat{N}_n = \underset{N_n \in (1,2,3,4,5)}{\arg\min}\ BIC(N_n) = \underset{N_n \in (1,2,3,4,5)}{\arg\min}\ \{\log MSE + [(N_n + p + 1)\log n]/n\}$, where $MSE = \sum_{i=1}^{n} \left[Y_i - \hat{Y}_i(N_n)\right]^2 / n$.

## 3.3 Empirical Results

In this section, we conduct simulation studies to evaluate the numerical performance of our proposed method with finite samples. Both univariate and multivariate cases are considered. For comparison purpose, in the univariate case, we also include a nonparametric estimation method via local linear regression described in Martins-Filho and Yao (2007). In addition to the simulation studies, the application of our proposed method to the Norwegian Farm data is also illustrated.

### 3.3.1 Simulation Study

#### 3.3.1.1 Univariate Case

In this example, we adopt a simulation set-up as given in Martins-Filho and Yao (2007) to compare our proposed method with their nonparametric method based on local linear regression. The data is generated from a frontier model with a single input variable $Y = \rho(X)R$, where $\rho(X)$ is the frontier function and $R$ is the efficiency variable. The input variable $X$ is generated from the uniform distribution on $[1,2]$. Let $R = \exp(-Z)$, where $Z$ is an exponential variable with the scale parameter $\beta$, therefore $R$ has support on $(0,1)$. We choose $\rho(x) = 3(x-1.5)^3 + 0.25x + 1.125$ and consider three different parameters for the exponential distribution with $\beta = 1/3, 1$, and 3. These choices of the parameter $\beta$ lead to three different shapes for the distribution of the production efficiency variable $R$: left-skewed, uniform, and right-skewed for $\beta = 1/3, 1$, and 3, respectively. We aim to examine the impact of different underlying distributions of $R$ on the performance of frontier estimator. Three sample sizes $n = 100$, 250, and 500 are considered and $r = 100$ replications are generated for each design.

For each simulated data set, we estimate the frontier function using both unconstrained (ULS) and constrained linear spline estimation methods. Two types of shape constraints are considered: monotonicity only (MCLS), and monotonicity and concavity (MCCLS). In this example, the true frontier function is monotone increasing. The above spline methods are considered to illustrate the effect of incorporating shape constraints on improving estimation accuracy. For the purpose of comparison, we also consider the nonparametric estimator using local linear regression as proposed in Martins-Filho and

Yao (2007).

We use the averaged integrated squared errors (AISE) to evaluate the estimation accuracy of these four different estimation methods. Let $\hat{\rho}_k$ be an estimator of $\rho_k$ in the $k$-th replication and $\{x_j\}_{j=1}^{ngrid}$ be a set of grid points where the functions are evaluated. Then the integrated squared error (ISE) of $\hat{\rho}_k$ is defined as $ISE(\hat{\rho}_k) = \frac{1}{ngrid}\sum_{j=1}^{ngrid}\left[\hat{\rho}_k\left(x_j\right) - \rho\left(x_j\right)\right]^2$ and the averaged integrated squared error is calculated as $AISE(\hat{\rho}) = \frac{1}{r}\sum_{k=1}^{r} ISE(\hat{\rho}_k)$. To better understand the performance of our proposed two-step polynomial spline estimation methods, in addition to the AISEs of $\hat{\rho}(x)$, we also report the AISEs of $\hat{m}(x)$ from the first estimation step and the mean squared errors (MSE) of $\hat{\mu}_R$ from the second estimation step.

The simulation results are summarized in Tables 3.1 and 3.2. Overall, it clearly shows that for all four methods, the AISEs decrease as the sample size $n$ increases, supporting our asymptotic convergence results. When comparing the three spline methods, we observe that the monotone constrained spline methods with correctly specified constraints have better performance than the one without constraint under all three designs. Furthermore, Table 3.2 shows that the incorporation of shape constraints enhance the estimation accuracy in both first and second estimation steps. In addition, the MCLS method with correctly specified constrains has the smallest AISEs under the scenario where $\beta = 1/3$, while the MCCLS method has the best performance when $\beta = 1$ and 3 although the additional constraint on concavity is misspecified. For the local linear regression method (LLR), the AISEs of the LLR method is the largest when $\beta = 1/3$. For the other two cases when $\beta = 1$ and 3, the LLR method is comparable with the spline methods and even has slightly better performance than the other two when $\beta = 3$.

These results also suggest the impact of different underlying distributions of the production efficiency variable $R$ on the performance of different estimation methods. Since the LLR and our proposed spline methods estimate different components of the model, the distributions of $R$ have different influences on their final estimation of frontiers. The spline methods have least favorable performance when $\beta = 3$ compared with the case when $\beta = 1/3$ or 1. This phenomena is mainly due to the difficulty in estimating $\mu_R$ under the right-skewed distribution of the efficiency variable $R$ where the majority of observations are close to 0. However, the LLR method has the worst performance when $\beta = 1/3$. This inferior performance is due to the fact that under this design $\sigma_R^2 = 0.04$ that is half of that in the other two cases, leading to a larger variability in estimating the frontier function as indicated by Theorem 2 in Martins-Filho and Yao (2007). In this simulated example, Table 3.1 shows the three spline methods have great advantage over the LLR method when the efficiency variable $R$ has a left-skewed distribution ($\beta = 1/3$).

In addition, for our spline methods, a robust procedure is proposed in the estimation of the location parameter $\mu_R$ to alleviate the effect of outliers. In order to evaluate its effectiveness, two artificial outliers located at $(X,Y) = (1.4,3)$ and $(1.6,3)$ are added manually to each simulated data set. Then, all four estimation procedures are performed again and compared with the ones without robust modification. Notice that, for the LLR method we only report the AISEs of frontier function since this method estimates $\sigma_R^2$ instead of $\mu_R$. To make the LLR method comparable with our robust method, we apply similar robust modification in estimating $\sigma_R^2$. Table 3.3 clearly illustrates the effectiveness of the robust procedure when there are outliers. In particular, the AISEs of the frontier estimator reduce dramatically for all four methods after the robust modification

is applied. For the spline methods, it suggests that the robust method is effective in estimating $\mu_R$ thus improve the performance of the frontier estimation. Also, we can easily see that when there are outliers, our proposed spline methods are noticeably much better than the LLR method under all three designs. For example, in the case when $n = 50$ and $\beta = 1/3$, the AISEs of the frontier estimation using MCLS method decrease by 99.07% as compared to the LLR method.

In Figures 3.1 and 3.2, we plot the estimated frontier functions from one simulated data set using four different methods: LLR (solid blue line), ULS (dashed green line), MCLS (dotted purple line), and MCCLS (dot-dashed red line), along with the true frontier (solid black line). Without outliers, Figure 3.1 shows that our proposed methods perform very well as the fitted curves are very close to the true curve and they are noticeably much better than the LLR method in the case where $\beta = 1/3$. When there are outliers and without robust method, by Figure 3.2 we can see that all fitted curves are pulled up due to those two extreme outliers. When the robust method is applied, the plots indicate that all four methods work much better and the three spline methods are even better than the LLR method especially for the smaller sample size. This suggests that our proposed method is more robust to the outliers. It graphically verify our numerical findings.

#### 3.3.1.2 Multivariate Case

In addition to the univariate case, here we consider an additive frontier model with multiple input variables, $Y = [\rho_0 + \rho_1(X_1) + \rho_2(X_2) + \rho_3(X_3) + \rho_4(X_4)]R$, where $\rho_0 =$

8, $\rho_1(x) = 2x - 1$, $\rho_2(x) = 2x + [\sin(2\pi x)]/2\pi - 1$, $\rho_3(x) = 3x^{1/3} - 9/4$, and $\rho_4(x) = (\log x + 1)/2$. The input variables $\{X_l\}_{l=1}^4$ are independently generated from the uniform distribution on $[0,1]$. The efficiency variable $R$ is generated in the same way as that in the univariate example. The selected additive frontier functions cover a combination of the functions that are either monotone increasing or monotone increasing and concave. Similarly, three sample sizes $n = 100$, 250, and 500 are considered and $r = 100$ replications are generated for each set-up. For each generated data, we estimate the additive frontier functions using three linear splines methods: ULS, MCLS and MCCLS as described in the univariate case. Again, AISE and MSE are used to evaluate the estimation accuracy of the nonparametric functions and the parameter $\mu_R$, respectively. Two artificial outliers located at $(X_1, X_2, X_3, X_4, Y) = (0.4, 0.4, 0.4, 0.4, 12)$ and $(0.6, 0.6, 0.6, 0.6, 12)$ are added manually to each simulated data set.

The simulation results are summarized in Tables 3.4 (without outliers) and 3.5 (with outliers). For all three spline methods, Table 3.4 shows that the estimation errors decrease as the sample size increases. Also, the two spline methods with shape constraints (MCLS and MCCLS) perform much better than the unconstrained method (ULS) in both first and second step estimation. These findings are consistent with that in the univariate case. In addition, the MCCLS method gives the least AISEs among these three spline methods. Furthermore, similar to the univariate case, the spline methods have the worst performance when $\beta = 3$ due to the difficulty in estimating $\mu_R$. Again, when using the contaminated data set, without the robust modification, Table 3.5 shows that the ULS method gives very large estimation errors when $\beta = 3$ because of large errors in the estimation of $\mu_R$. When the robust method is applied, the ULS method is im-

proved greatly but still worse than the other two constrained methods. It illustrates that the robust method improve the estimation accuracy of $\mu_R$ greatly, especially under the case with right-skewed distribution of the efficiency variable $R$ ($\beta = 3$). For example, when applying the robust modification, the MSE of $1/\mu_R$ obtained using ULS decreases from 800.34 to 0.076.

### 3.3.2 Norwegian Farm Data

In this real data example, we apply our proposed method to the Norwegian Farm data. This data set is originally from Norwegian Farm Accountancy Survey collected by Norwegian Agricultural Economics Research Institute. Recall that a monotone additive model has been used to analyze this data set in Chapter 2. Here we revisit the same data set but the additive frontier model is applied. We consider the data of year 2007 that includes observations on 151 grain farms in Norway. Similarly, we focus on four input variables: total number of hours worked (labor) on the farm ($X_1$), productive farmland in hectares ($X_2$), variable farm inputs ($X_3$) and fixed farm input and capital costs ($X_4$). The output variable ($Y$) is the logarithm of farm revenue measured in Norwegian kroner. Kumbhakar, Lien and Hardaker (2014) thoroughly studied 6 different parametric stochastic frontier models for the estimation of farm efficiency. In Chapter 2, we use a monotone additive regression model to quantify the relationship between the input variables and the output variable. In this chapter, we consider the additive frontier model $Y = [\rho_0 + \rho_1(X_1) + \rho_2(X_2) + \rho_3(X_3) + \rho_4(X_4)]R$ to quantify the maximum farm revenue given these four input variables. Our semi-parametric model is more flexible that

the parametric ones considered in Kumbhakar, Lien and Hardaker (2014). In addition, in Chapter 2 we focus on the estimation of the regression or conditional mean function, not the frontier function.

To estimate the unknown components in the model, three linear spline methods: unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), and monotone and concave constrained linear spline (MCCLS) are considered. The number of interior knots $N_n$ is taken as the integer part of $n^{1/(2p+3)}$. The knot sequence is selected to be equally spaced in the range of each input variable $X_l$.

Figure 3.3 plots the estimates of the conditional mean function obtained in the first estimation step. We plot the pseudo responses (black circle) along with the estimated mean functions using the ULS (dashed blue line), MCLS (dot-dashed green line), and MCCLS (long-dashed red line) methods. The 95% point-wise confidence intervals from boostrapping for each input variable (dotted blue line) of the ULS method are also plotted. At each grid point, the 95% point-wise confidence intervals use the 2.5% and 97.5% sample quantiles of the ULS estimates obtained from 100 boostrapped samples as the lower and upper bounds, respectively. It clearly shows that the three spline methods give very similar estimation results. Additionally, all four input variables have monotonic effects on the farm revenue and the increasing rates decrease as the input variables increase.

Furthermore, to assess the production efficiency of each farm, in Figure 3.4 we plot the estimated maximum farm revenue (left top), efficiency estimates (right top), and the kernel density distribution of the efficiency estimates (left bottom). We can easily see that the estimation results are very similar for the three spline methods: the ULS

(blue rectangle or line), MCLS (green triangle or line), and MCCLS (red plus or line) methods. Furthermore, it indicates that the majority of 151 farms have high efficiency with the estimate above 0.95. In addition, farms with lower revenue tend to have lower production efficiency.

In addition, to explain the difference in efficiency among these farms, we also explore the relationships between the efficiency and 5 other explanatory variables of interest that include off-farm income share (net income off the farm as a proportion of the total net income), coupled subsidy income share (coupled subsidies as a proportion of the total farm net income), environmental subsidy income share (farm environmental payments as a proportion of the total farm net income), farmer experience (number of years as a farmer) and the farmers' education level. Figure 3.5 shows the scatter plots of those variables versa the estimated efficiency and the box plots of the estimated efficiency for the three different groups of farmers with different education levels, using the MCCLS method. For the first four explanatory variables, we also fit the least squares regression lines (solid red lines). Clearly, it indicates that larger coupled subsidy income share or environmental subsidy income share are associated with lower efficiency. These negative influences may be due to the reason that the motivation of the farmers to work is reduced by the extra off-farm income. However, it seems that there is no apparent relationship between the efficiency and off-farm income share, farmer experience or the farmer's education level.

## 3.4 Proof of Lemmas and Theorems

### 3.4.1 Preliminary Lemmas and Proof

**Proof of Lemma 9**.

See the proof of Lemma 4 in section 2.4.1. ■

**Proof of Lemma 10**.

According to equation (2.7), the first order derivative of the linear spline can be written as

$$g'(x) = \left[ \left(\beta_0 / \left(u_1 - u_0\right)\right) \tilde{B}_{0,1}(x) + \sum_{j=1}^{N_n} \left( \left(\beta_j - \beta_{j-1}\right) / \left(u_{j+1} - u_j\right) \right) \tilde{B}_{j,1}(x) \right].$$

For a linear spline, the rate of change is a constant in each interval. To ensure that $g(\cdot)$ is a concave function, the rate of change needs to be non-increasing in the whole region. Therefore, we have $\left(\beta_1 - \beta_0\right) / \left(u_2 - u_1\right) \leq \beta_0 / \left(u_1 - u_0\right)$ and $\left(\beta_j - \beta_{j-1}\right) / \left(u_{j+1} - u_j\right)$ $\leq \left(\beta_{j-1} - \beta_{j-2}\right) / \left(u_j - u_{j-1}\right)$, for $j = 2, ..., N$ and $N \geq 2$. When $N = 1$, one has $\left(\beta_1 - \beta_0\right) / \left(u_2 - u_1\right) \leq \beta_0 / \left(u_1 - u_0\right)$. Since the knots are equally space, Lemma 9 follows. ■

**Proof of Lemma 11**.

By equation (2.7), we have

$$g'(x) = (k-1) \left[ \frac{\beta_{-p+1}}{u_1 - u_{-p+1}} \tilde{B}_{-p+1,k-1} + \sum_{j=-p+2}^{N_n} \frac{\beta_j - \beta_{j-1}}{u_{j+k-1} - u_j} \tilde{B}_{j,k-1} \right].$$

When degree $p \geq 2$, we can take the second order derivative and obtain

$$
\begin{aligned}
& g^{(2)}(x) \\
= \; & (k-1)\left[\frac{\beta_{-p+1}}{u_1-u_{-p+1}}\tilde{B}'_{-p+1,k-1}+\sum_{j=-p+2}^{N_n}\frac{\beta_j-\beta_{j-1}}{u_{j+k-1}-u_j}\tilde{B}'_{j,k-1}\right] \\
= \; & (k-1)(k-2)\sum_{j=-p+2}^{N_n}\frac{\beta_j-\beta_{j-1}}{u_{j+k-1}-u_j}\left(\frac{-\tilde{B}_{j+1,k-2}}{u_{j+k-1}-u_{j+1}}+\frac{\tilde{B}_{j,k-2}}{u_{j+k-2}-u_j}\right) \\
& +(k-1)(k-2)\frac{-\beta_{-p+1}}{\left(u_1-u_{-p+1}\right)\left(u_1-u_{-p+2}\right)}\tilde{B}_{-p+2,k-2} \\
= \; & I+II.
\end{aligned}
$$

Furthermore, part $I$ can be decomposed as

$$
\begin{aligned}
I \; = \; & (k-1)(k-2)\sum_{j=-p+2}^{-1}\frac{\beta_j-\beta_{j-1}}{u_{j+k-1}-u_j}\left(\frac{-\tilde{B}_{j+1,k-2}}{u_{j+k-1}-u_{j+1}}+\frac{\tilde{B}_{j,k-2}}{u_{j+k-2}-u_j}\right) \\
& +(k-1)(k-2)\sum_{j=0}^{N_n-k+2}\frac{\beta_j-\beta_{j-1}}{u_{j+k-1}-u_j}\left(\frac{-\tilde{B}_{j+1,k-2}}{u_{j+k-1}-u_{j+1}}+\frac{\tilde{B}_{j,k-2}}{u_{j+k-2}-u_j}\right) \\
& +(k-1)(k-2)\sum_{j=N_n-k+3}^{N_n}\frac{\beta_j-\beta_{j-1}}{u_{j+k-1}-u_j}\left(\frac{-\tilde{B}_{j+1,k-2}}{u_{j+k-1}-u_{j+1}}+\frac{\tilde{B}_{j,k-2}}{u_{j+k-2}-u_j}\right) \\
= \; & I_1+I_2+I_3.
\end{aligned}
$$

Then, we check each partition and get

$$
\begin{aligned}
I_1 &= (k-1)(k-2)\frac{\beta_{-p+2}-\beta_{-p+1}}{\left(u_2-u_{-p+2}\right)\left(u_1-u_{-p+2}\right)}\tilde{B}_{-p+2,k-2} \\
&\quad +(k-1)(k-2) \\
&\quad \times\sum_{j=-p+3}^{-1}\left[\frac{\beta_j-\beta_{j-1}}{\left(u_{j+k-1}-u_j\right)\left(u_{j+k-2}-u_j\right)}-\frac{\beta_{j-1}-\beta_{j-2}}{\left(u_{j+k-2}-u_{j-1}\right)\left(u_{j+k-2}-u_j\right)}\right]\tilde{B}_{j,k-2} \\
&\quad +(k-1)(k-2)\frac{-\left(\beta_{-1}-\beta_{-2}\right)}{\left(u_{k-2}-u_{-1}\right)\left(u_{k-2}-u_0\right)}\tilde{B}_{0,k-2} \\
&= (k-1)(k-2)\frac{\beta_{-p+2}-\beta_{-p+1}}{2h^2}\tilde{B}_{-p+2,k-2} \\
&\quad +(k-1)(k-2)\sum_{j=-p+3}^{-1}\left[\frac{\beta_j-\beta_{j-1}}{(j+k-1)(j+k-2)h^2}-\frac{\beta_{j-1}-\beta_{j-2}}{(j+k-2)^2h^2}\right]\tilde{B}_{j,k-2} \\
&\quad +(k-1)\frac{-\left(\beta_{-1}-\beta_{-2}\right)}{(k-2)h^2}\tilde{B}_{0,k-2}, \qquad (3.8)
\end{aligned}
$$

$$
\begin{aligned}
I_2 &= \frac{\beta_0-\beta_{-1}}{h^2}\tilde{B}_{0,k-2}+\sum_{j=1}^{N_n-k+2}\frac{\beta_j-\beta_{j-1}-\left(\beta_{j-1}-\beta_{j-2}\right)}{h^2}\tilde{B}_{j,k-2} \\
&\quad +\frac{-\left(\beta_{N_n-k+2}-\beta_{N_n-k+1}\right)}{h^2}\tilde{B}_{N_n-k+3,k-2}, \qquad (3.9)
\end{aligned}
$$

and

$$
I_3 = (k-1)(k-2)\frac{\left(\beta_{N_n-k+3}-\beta_{N_n-k+2}\right)}{\left(u_{N_n+2}-u_{N_n-k+3}\right)\left(u_{N_n+1}-u_{N_n-k+3}\right)}\tilde{B}_{N_n-k+3,k-2}
$$

$$
\begin{aligned}
& + (k-1)(k-2) \\
& \times \sum_{j=N_n-k+4}^{N_n} \left[ \frac{\beta_j - \beta_{j-1}}{(u_{j+k-1} - u_j)(u_{j+k-2} - u_j)} - \frac{\beta_{j-1} - \beta_{j-2}}{(u_{j+k-2} - u_{j-1})(u_{j+k-2} - u_j)} \right] \tilde{B}_{j,k-2} \\
= \quad & (k-1) \frac{(\beta_{N_n-k+3} - \beta_{N_n-k+2})}{(k-2)h^2} \tilde{B}_{N_n-k+3,k-2} \\
& + (k-1)(k-2) \\
& \times \sum_{j=N_n-k+4}^{N_n} \left[ \frac{\beta_j - \beta_{j-1}}{(N_n+1-j)^2 h^2} - \frac{\beta_{j-1} - \beta_{j-2}}{(N_n+2-j)(N_n+1-j)h^2} \right] \tilde{B}_{j,k-2}. \qquad (3.10)
\end{aligned}
$$

Also, $II$ can be written as

$$
II = (k-1)(k-2) \frac{-\beta_{-p+1}}{h^2} \tilde{B}_{-p+2,k-2}. \qquad (3.11)
$$

By conducting the above decomposition, the second order derivative of the polynomial spline of order $k$ can be written as a linear combination of B-spline basis in the polynomial spline space with order $k-2$. Since the B-spline basis is positive, a sufficient condition to guarantee the concavity of the polynomial spline is that the coefficients of the distinct basis are all non-positive. Now, combine with (3.8), (3.9), (3.10) and (3.11), we get the following set of inequalities:

$$
\begin{aligned}
& \beta_{-p+2} - \beta_{-p+1} \leq 2\beta_{-p+1}, \\
& \beta_j - \beta_{j-1} \leq (j+p) \left( \beta_{j-1} - \beta_{j-2} \right) / (j+p-1), \text{ for } j = -p+3, ..., -1 \\
& \beta_0 - \beta_{-1} \leq p \left( \beta_{-1} - \beta_{-2} \right) / (p-1), \\
& \beta_j - \beta_{j-1} \leq \beta_{j-1} - \beta_{j-2}, \text{ for } j = 1, ..., N_n - p + 1 \\
& \beta_{N_n-p+2} - \beta_{N_n-p+1} \leq (p-1) \left( \beta_{N_n-p+1} - \beta_{N_n-p} \right) / p,
\end{aligned}
$$

$\beta_j - \beta_{j-1} \leq (N_n + 1 - j)\left(\beta_{j-1} - \beta_{j-2,}\right) / (N_n + 2 - j)$, for $j = N_n - p + 3, ..., N_n$ with degree $p \geq 2$ and Lemma 11 follows. ■

**Lemma 16** *A sufficient condition for a quadratic spline $g(x) = \mathbf{B}^T(x)\beta$ , where $\mathbf{B}(x)$ is the empirically centered B-spline basis of $G^2$, to be concave is that its coefficients satisfy the following conditions:*

$\beta_0 - \beta_{-1} \leq 2\beta_{-1}$,

$\beta_1 - \beta_0 \leq (\beta_0 - \beta_{-1})/2$, *for* $N = 1$, *and*

$\beta_0 - \beta_{-1} \leq 2\beta_{-1}$,

$\beta_j - \beta_{j-1} \leq \beta_{j-1} - \beta_{j-2}$,

$\beta_N - \beta_{N-1} \leq (\beta_{N-1} - \beta_{N-2})/2$, *for* $N \geq 2$, $j = 1, ..., N-1$.

**Proof**. Following from Lemma 11, we get the results. ■

**Lemma 17** *A sufficient condition for a cubic spline $g(x) = \mathbf{B}^T(x)\beta$ , where $\mathbf{B}(x)$ is the empirically centered B-spline basis of $G^3$, to be concave is that its coefficients satisfy the following conditions:*

$\beta_{-1} - \beta_{-2} \leq 2\beta_{-2}$,

$\beta_0 - \beta_{-1} \leq \beta_{-1} - \beta_{-2}$,

$\beta_1 - \beta_0 \leq (\beta_0 - \beta_{-1})/2$, *for* $N = 1$,

$\beta_{-1} - \beta_{-2} \leq 2\beta_{-2}$,

$\beta_0 - \beta_{-1} \leq 3(\beta_{-1} - \beta_{-2})/2$,

$\beta_1 - \beta_0 \leq 2(\beta_0 - \beta_{-1})/3$,

$\beta_2 - \beta_1 \leq (\beta_1 - \beta_0)/2$, *for* $N = 2$ *and*

$\beta_{-1} - \beta_{-2} \leq 2\beta_{-2}$,

$\beta_0-\beta_{-1} \leq 3\left(\beta_{-1}-\beta_{-2}\right)/2,$

$\beta_j-\beta_{j-1} \leq \beta_{j-1}-\beta_{j-2},$

$\beta_{N-1}-\beta_{N-2} \leq 2\left(\beta_{N-2}-\beta_{N-3}\right)/3,$

$\beta_N-\beta_{N-1} \leq \left(\beta_{N-1}-\beta_{N-2}\right)/2$, *for* $N \geq 3$, $j = 1,...,N-2.$

**Proof**. Following from Lemma 11, we have the results. ■

**Lemma 18** *For any m that satisfies condition (A5*) and large sample size, there exists a concave function* $g \in G^{(p+1)}$, *such that* $\|m-g\|_\infty \leq c\left\|m^{(p+2)}\right\|_\infty / N_n^{p+2}$, *and* $\left\|m^{(2)}-g^{(2)}\right\|_\infty \leq c\left\|m^{(p+2)}\right\|_\infty / N_n^p$ *for some constant* $c > 0$.

**Proof**. According to Theorem 1.51 in Schumaker (2015), for any $m \in C^{p+2}[0,1]$, there exists a function $g \in G^{(p+1)}$, such that $\|m-g\|_\infty \leq c\left\|m^{(p+2)}\right\|_\infty / N_n^{p+2}$ and $\left\|m^{(2)}-g^{(2)}\right\|_\infty \leq c\left\|m^{(p+2)}\right\|_\infty / N_n^p$ for some constant $c > 0$. By condition (A5*), $m^{(2)}(x)$ $\leq c_1 < 0$ for some constant $c_1$ and any $x \in [0,1]$. Then for large sample size, one has

$$g^{(2)} \leq \left\|m^{(2)}-g^{(2)}\right\|_\infty + m^{(2)} \leq c_1/2 < 0.$$

Therefore, the spline function $g$ is concave and Lemma 18 follows. ■

For each $l = 1,\ldots,d$, let $m_l^*$ be the one-step backfitted estimate of $m_l$ when all the other additive components are known. In this case, it reduces to a univariate polynomial spline smoothing. Define $Y_{i,-l}^* = Y_i - \sum_{l' \neq l} m_{l'}(x_{il'}) = m_l(x_{il}) + \rho(x_i)\varepsilon_i$ as the pseudo response, and for $l = 1,\ldots,d$ let $\mathbf{Y}_{-l}^* = \left(Y_{1,-l}^*,...,Y_{n,-l}^*\right)^T$. Then $m_l^*(x) = \tilde{\mathbf{B}}_l^T(x)\beta_l^*$ where $\beta_l^* = \left(\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1}\tilde{\mathbf{B}}_{nl}^T\mathbf{Y}_{-l}^*$. Note that $\left\{m_l^*\right\}_{l=1}^d$ are only constructed to prove our theorem.

**Lemma 19** *Under regularity conditions (A1)-(A4),(A5*), one has, for* $l = 1, \ldots, d$,

$$\sup_{x\in[0,1]} \left| m_l^{*(2)}(x) - m_l^{(2)}(x) \right| = O_p\left( N_n^{\frac{1}{2}-p} + N_n^{\frac{5}{2}}/\sqrt{n} \right).$$

**Proof**. By Lemma 18, for each $l = 1, \ldots, d$, there exists a concave function $g_l$, such that $\|m_l - g_l\|_\infty \leq c \left\| m_l^{(p+2)} \right\|_\infty / N_n^{p+2}$ and $\left\| m_l^{(2)} - g_l^{(2)} \right\|_\infty \leq c \left\| m_l^{(p+2)} \right\|_\infty / N_n^p$. By definition, one has

$$\begin{aligned}
& m_l^{*(2)}(x) \\
= & \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \beta_l^* = \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \mathbf{Y}_{-l}^* \\
= & \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{m}_l + \rho\varepsilon) \\
= & \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{g}_l + \mathbf{m}_l - \mathbf{g}_l + \rho\varepsilon) \\
= & \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \mathbf{g}_l + \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{m}_l - \mathbf{g}_l) \\
& + \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \rho\varepsilon \\
= & I(x) + II(x) + III(x). \qquad (3.12)
\end{aligned}$$

Since we can write $\mathbf{g}_l = \tilde{\mathbf{B}}_{nl}\gamma_l$ for some coefficient $\gamma_l$,

$$I(x) = \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\gamma_l = \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \gamma_l = g_l^{(2)}(x),$$

then by Lemma 18, we have

$$\sup_{x\in[0,1]} \left| I(x) - m_l^{(2)}(x) \right| = \sup_{x\in[0,1]} \left| g_l^{(2)}(x) - m_l^{(2)}(x) \right| \leq c \left\| m_l^{(p+2)} \right\|_\infty / N_n^p. \qquad (3.13)$$

Next, apply Cauchy-Schwarz inequality to $II(x)$ and we get

$$\begin{aligned}\sup_x II(x) \quad \leq \quad & \sqrt{\sup_x \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l^{(2)}(x)\right]} \times \\ & \sqrt{\left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T(\mathbf{m}_l-\mathbf{g}_l)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T(\mathbf{m}_l-\mathbf{g}_l)\right]}.\end{aligned}$$

Let $\lambda_{\min}$ and $\lambda_{\max}$ be the smallest and largest eigenvalues of $\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}/n$, respectively. According to Theorem 5.4.2 in Devore and Lorentz (1993), $\lambda_{\min} \asymp \lambda_{\max} = O_p(1/N_n)$, where $\asymp$ means both sides have the same order. With the fact that $\sup_x \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \tilde{\mathbf{B}}_l^{(2)}(x) = O_P\left(N_n^4\right)$, one has

$$\left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_l^{(2)}(x)\right] \leq \frac{1}{n\lambda_{\min}}\left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T \tilde{\mathbf{B}}_l^{(2)}(x) = O_P\left(N_n^5/n\right).$$

Also,

$$\begin{aligned} & \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T(\mathbf{m}_l-\mathbf{g}_l)\right]^T \left(\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\tilde{\mathbf{B}}_{nl}\right)^{-1} \left[\frac{1}{\sqrt{n}}\tilde{\mathbf{B}}_{nl}^T(\mathbf{m}_l-\mathbf{g}_l)\right] \\ \leq \quad & \frac{1}{n\lambda_{\min}}(\mathbf{m}_l-\mathbf{g}_l)^T\tilde{\mathbf{B}}_{nl}\tilde{\mathbf{B}}_{nl}^T(\mathbf{m}_l-\mathbf{g}_l) \\ \leq \quad & \frac{\lambda_{\max}}{\lambda_{\min}}(\mathbf{m}_l-\mathbf{g}_l)^T(\mathbf{m}_l-\mathbf{g}_l) = O_p\left(n/N_n^{2p+4}\right).\end{aligned}$$

So,

$$\sup_x II(x) = O_p\left(N_n^{\frac{1}{2}-p}\right). \tag{3.14}$$

Next, we notice that $\sup_x \left[\tilde{\mathbf{B}}_l^{(2)}(x)\right]^T = O_p\left(N_n^2\right)$, and the element of $\frac{1}{n}\tilde{\mathbf{B}}_{nl}^T\rho\varepsilon$ is

$\frac{1}{n}\sum_{i=1}^{n} B_{lj}(x_i)\rho(x_i)\varepsilon_i = O_p\left(1/\sqrt{nN_n}\right)$, therefore

$$\sup_x III(x) = O_p\left(N_n^{\frac{5}{2}}/\sqrt{n}\right). \tag{3.15}$$

Finally, Lemma 19 follows from equations (3.12), (3.13), (3.14), (3.15). ■

**Lemma 20** *For any function $m \in C^{p+2}[0,1]$ that satisfies condition (A5*) with $p \leq 3$ and large sample size, there exists a concave function $g \in G^{(p+1)}$ and $\hat{g} \in G^{(p+1)}$ whose coefficients satisfy the concave constraints, such that $\|m-g\|_\infty \leq c_1 \left\|m^{(p+2)}\right\|_\infty / N_n^{p+2}$ and $\|g-\hat{g}\|_\infty \leq c_2 \left\|m^{(p+2)}\right\|_\infty / N_n^{p+2}$, for some constant $c_1, c_2 > 0$.*

**Proof**. By Lemma 18, for any $m \in C^{p+2}[0,1]$ that satisfies condition (A5*) and large sample size, there exists a concave function $g \in G^{(p+1)}$, such that $\|m-g\|_\infty \leq c_1 \left\|m^{(p+2)}\right\|_\infty / N_n^{p+2}$ for some constant $c_1 > 0$. Since $m \in C^{p+2}[0,1]$, its first order derivative $m' \in C^{p+1}[0,1]$. According to Lemma 7 in Chapter 2, for $p \leq 3$ and large sample size, there exists a $\hat{g}' \in G^{(p)}$ whose coefficients satisfy the linear constraints, such that $\|g'-\hat{g}'\|_\infty \leq c_2 \left\|m^{(p+2)}\right\|_\infty / N_n^{p+1}$, for some constant $c_2 > 0$. With $\hat{g}(x) = \int \hat{g}'(x)\,dx \in G^{(p+1)}$ and follow the similar arguments in the proof of Lemma 7, we can easily show that $\|g-\hat{g}\|_\infty \leq c_2 \left\|m^{(p+2)}\right\|_\infty / N_n^{p+2}$. Moreover, in the proof of Lemma 7, we showed that the spline coefficients of $\hat{g}^{(2)}$ are non-positive that suggests that the coefficients of $\hat{g}$ satisfy the concave constraints. Therefore, Lemma 20 follows. ■

Recall that $m_l^* = \sum_{j=1}^{J_n+1} \beta_{lj}^* \tilde{B}_j$ is the one-step backfitted estimate of $m_l$ with all other additive components known, for $l = 1, \ldots, d$.

**Lemma 21** *Under regularity conditions (A1)-(A4),(A5*), one has, for* $l = 1, \ldots, d$ *and* $p \leq 3$*, there exists a spline function* $g_l = \sum_{j=1}^{J_n+1} \gamma_{lj} \tilde{B}_j$ *whose coefficients satisfy the concave constraints such that*

$$\sup_j \left| \gamma_{lj} - \beta^*_{lj} \right| = O_p \left( \sqrt{\frac{N_n^3 \log n}{n}} \right).$$

*Therefore, the coefficients of* $m_l^*$ *satisfies the concave constraints with probability approaching to* 1 *as* $n \to \infty$*.*

**Proof**. For $p \leq 3$, by Lemma 20, there exists a spline function $g_l$ whose coefficients satisfy the concave constraints such that $\|m_l - g_l\|_\infty = O_p\left(N_n^{-p-2}\right)$. By definition, we have

$$\begin{aligned}
\beta_l^* &= \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{m}_l + \rho\varepsilon) \\
&= \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{g}_l + \mathbf{m}_l - \mathbf{g}_l + \rho\varepsilon) \\
&= \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \mathbf{g}_l + \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{m}_l - \mathbf{g}_l) + \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \rho\varepsilon \\
&= I + II + III. \qquad (3.16)
\end{aligned}$$

Write $\mathbf{g}_l = \tilde{\mathbf{B}}_{nl} \gamma_l$, then

$$I = \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl} \gamma_l = \gamma_l. \qquad (3.17)$$

Let $|\cdot|_*$ denotes the element-wise absolute value of a matrix. Since $\sup_x |\frac{1}{n} \tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}|_* = O_P\left(N_n^{-1}\right)$ and $\|m_l - g_l\|_\infty = O_p\left(N_n^{-p-2}\right)$, we have

$$\sup |II|_* = O_p\left(n^{-1} N_n^{-p-1}\right). \qquad (3.18)$$

Then, follow the similar argument in the proof of Lemma 8, one has

$$\sup |III|_* = O_p\left(\sqrt{\frac{N_n^3 \log n}{n}}\right). \tag{3.19}$$

From equations (3.16), (3.17), (3.18) and (3.19), we have $\sup|\beta_l^* - \gamma_l|_* = \sup|II + III|_* = O_p\left(\sqrt{\frac{N_n^3 \log n}{n}}\right)$ and then Lemma 21 follows. ■

Recall that $\tilde{m}_l = \sum_{j=1}^{J_n+1} \tilde{\beta}_{lj}\tilde{B}_j$ is the one-step backfitted unconstrained estimate of $m_l$ for $l = 1,\ldots,d$.

**Lemma 22** *Under regularity conditions (A1)-(A5*), one has, for $l = 1,\ldots,d$ and $p \leq 3$, there exists a spline function $g_l = \sum_{j=1}^{J_n+1} \gamma_{lj}\tilde{B}_j$ whose coefficients satisfy the concave constraints such that*

$$\sup_j \left|\gamma_{lj} - \tilde{\beta}_{lj}\right| = O_p\left(\sqrt{\frac{N_n^3 \log n}{n}}\right).$$

*Therefore, the coefficients of $\tilde{m}_l$ satisfies the concave constraints with probability approaching to 1 as $n \to \infty$.*

**Proof**. We have

$$\begin{aligned}
\tilde{\beta}_l &= \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{y} - \tilde{\mathbf{m}}_{-l}) \\
&= \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T \left(\sum_{l=1}^{d} \mathbf{m}_l + \rho\varepsilon - \tilde{\mathbf{m}}_{-l}\right) \\
&= \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{m}_l + \rho\varepsilon) + \left(\tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl}\right)^{-1} \tilde{\mathbf{B}}_{nl}^T (\mathbf{m}_{-l} - \tilde{\mathbf{m}}_{-l}) \\
&= \beta_l^* + I.
\end{aligned}$$

By Lemma 21, for any fixed $l = 1, ...d$, the coefficient $\beta_l^*$ satisfies the concave constraints with probability approaching to 1 as the sample size goes to infinity. Moreover,

$$\sup \left| \left( \tilde{\mathbf{B}}_{nl}^T \tilde{\mathbf{B}}_{nl} \right)^{-1} \right|_* = O_P \left( N_n / n \right),$$

$$\begin{aligned} \sup \left| \tilde{\mathbf{B}}_{nl}^T \left( \mathbf{m}_{-l} - \tilde{\mathbf{m}}_{-l} \right) \right|_* &\leq \sqrt{\left( \mathbf{m}_{-l} - \tilde{\mathbf{m}}_{-l} \right)^T \tilde{\mathbf{B}}_{nl} \tilde{\mathbf{B}}_{nl}^T \left( \mathbf{m}_{-l} - \tilde{\mathbf{m}}_{-l} \right)} \\ &= O_p \left( \sqrt{\left( n N_n^{-2p-2} + N_n \right) n / N_n} \right). \end{aligned}$$

Therefore, $\sup |I|_* = O_p \left( \sqrt{N_n^{-2p-1} + N_n^2 / n} \right)$ and then Lemma 22 follows. ■

### 3.4.2 Proof of Theorems

## Proof of Theorem 12

By the definition of $\tilde{m}_l(x)$, one has

$$\begin{aligned} \tilde{m}_l(x) &= \mathbf{B}_l^T(x) \tilde{\beta}_l = \mathbf{B}_l^T(x) \left( \mathbf{B}_{nl}^T \mathbf{B}_{nl} \right)^{-1} \mathbf{B}_{nl}^T \mathbf{Y}_{-l} \\ &= \mathbf{B}_l^T(x) \left( \mathbf{B}_{nl}^T \mathbf{B}_{nl} \right)^{-1} \mathbf{B}_{nl}^T \left( \mathbf{Y} - \sum_{l' \neq l} \tilde{\mathbf{m}}_{l'} \right) \\ &= \mathbf{B}_l^T(x) \left( \mathbf{B}_{nl}^T \mathbf{B}_{nl} \right)^{-1} \mathbf{B}_{nl}^T \left( \mathbf{m}_l + \rho \varepsilon \right) \\ &\quad + \mathbf{B}_l^T(x) \left( \mathbf{B}_{nl}^T \mathbf{B}_{nl} \right)^{-1} \mathbf{B}_{nl}^T \sum_{l' \neq l} \left( \mathbf{m}_{l'} - \tilde{\mathbf{m}}_{l'} \right). \end{aligned}$$

Then we have

$$
\begin{aligned}
& \sup_x \left|\tilde{m}_l(x) - m_l(x)\right| \\
\leq\ & \sup_x \left|\mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\left(\mathbf{m}_l + \rho\varepsilon\right) - m_l(x)\right| \\
& + \sup_x \left|\mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\sum_{l'\neq l}\left(\mathbf{m}_{l'} - \tilde{\mathbf{m}}_{l'}\right)\right| \\
=\ & I(x) + II(x). \qquad (3.20)
\end{aligned}
$$

Let $\bar{m}_l(x) = \mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\mathbf{m}_l$, then by Corollary 3.1 and Theorem 5.1 of Huang (2003), $\sup_x \left|\mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\left(\mathbf{m}_l + \rho\varepsilon\right) - \bar{m}_l(x)\right| = O_p\left(\sqrt{N_n/n}\right)$ and $\sup_x \left|\bar{m}_l(x) - m_l(x)\right| = O_p\left(N_n^{-p-1}\right)$. Therefore,

$$
\begin{aligned}
I(x) &\leq \sup_x \left|\mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\left(\mathbf{m}_l + \rho\varepsilon\right) - \bar{m}_l(x)\right| + \sup_x \left|\bar{m}_l(x) - m_l(x)\right| \\
&= O_p\left(\sqrt{N_n/n} + N_n^{-p-1}\right). \qquad (3.21)
\end{aligned}
$$

For $II(x)$, Cauchy-Schwarz inequality gives that

$$
\begin{aligned}
II(x) &\leq \sup_x \sqrt{\left[\mathbf{B}_l(x)/\sqrt{n}\right]^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}/n\right)^{-1}\left[\mathbf{B}_l(x)/\sqrt{n}\right]} \times \\
&\quad \sqrt{\left[\mathbf{B}_{nl}^T\sum_{l'\neq l}\left(\mathbf{m}_{l'} - \tilde{\mathbf{m}}_{l'}\right)/\sqrt{n}\right]^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}/n\right)^{-1}\left[\mathbf{B}_{nl}^T\sum_{l'\neq l}\left(\mathbf{m}_{l'} - \tilde{\mathbf{m}}_{l'}\right)/\sqrt{n}\right]} \\
&= II_1(x) \times II_2 \qquad (3.22)
\end{aligned}
$$

Let $\lambda_{\min}$ and $\lambda_{\max}$ be the smallest and largest eigenvalues of $\mathbf{B}_{nl}^T\mathbf{B}_{nl}/n$, respectively. Then by Theorem 5.4.2 in Devore and Lorentz (1993), $\lambda_{\min} \asymp \lambda_{\max} = O_p(1/N_n)$ where $\asymp$ means both sides have the same order. Along with the fact that $\sup_x \mathbf{B}_l^T(x)\mathbf{B}_l(x) =$

$O_P(1)$, one has

$$II_1(x) \leq \sup_x \sqrt{\mathbf{B}_l^T(x)\mathbf{B}_l(x)/n\lambda_{\min}} = O_p\left(\sqrt{N_n/n}\right). \tag{3.23}$$

Similarly,

$$\begin{aligned} II_2 &\leq \sqrt{\left[\sum_{l'\neq l}(\mathbf{m}_{l'}-\tilde{\mathbf{m}}_{l'})\right]^T \mathbf{B}_{nl}\mathbf{B}_{nl}^T \left[\sum_{l'\neq l}(\mathbf{m}_{l'}-\tilde{\mathbf{m}}_{l'})\right]/n\lambda_{\min}} \\ &\leq \sqrt{\lambda_{\max}\left[\sum_{l'\neq l}(\mathbf{m}_{l'}-\tilde{\mathbf{m}}_{l'})\right]^T \left[\sum_{l'\neq l}(\mathbf{m}_{l'}-\tilde{\mathbf{m}}_{l'})\right]/\lambda_{\min}} \\ &= O_p\left(\sqrt{N_n + nN_n^{-2p-2}}\right), \end{aligned} \tag{3.24}$$

in which $\left[\sum_{l'\neq l}(\mathbf{m}_{l'}-\tilde{\mathbf{m}}_{l'})\right]^T \left[\sum_{l'\neq l}(\mathbf{m}_{l'}-\tilde{\mathbf{m}}_{l'})\right] = O_p\left(N_n + nN_n^{-2p-2}\right)$ is given by Theorem 1 of Huang (1998). Therefore, according to (3.22), ( 3.23) and (3.24), we conclude that

$$II(x) = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right). \tag{3.25}$$

Finally, Theorem 1 follows from equations (3.20), (3.21) and ( 3.25). ■

## Proof of Theorem 13

By the definition of $\tilde{\mu}_R$, we have $\tilde{\mu}_R = \left\{\max_i [Y_i/\tilde{m}(X_i)]\right\}^{-1} \leq \sup_x |\tilde{m}(x)| \left(\max_i Y_i\right)^{-1}$ $= O_p(1)$, in which $\sup_x |\tilde{m}(x)| = O_p(1)$ is given by Theorem 12. Since $|\tilde{\mu}_R - \mu_R| = \tilde{\mu}_R\mu_R\left|\tilde{\mu}_R^{-1} - \mu_R^{-1}\right|$, we only need to show that $\left|\tilde{\mu}_R^{-1} - \mu_R^{-1}\right| = O_p(L_n)$. We note that

$$\begin{aligned}\left|\tilde{\mu}_R^{-1}-\mu_R^{-1}\right| &= \mu_R^{-1}\left|\mu_R \max_i\left[Y_i/\tilde{m}\left(X_i\right)\right]-1\right| \\ &= \mu_R^{-1}\left|\mu_R \max_i\left[\rho\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1\right| \\ &= \mu_R^{-1}\left|\max_i\left[m\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1\right|,\end{aligned}$$

therefore it is sufficient to show that $\left|\max_i\left[m\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1\right| = O_p\left(L_n\right)$. Given the assumption that $\sup_x\left|\tilde{m}\left(x\right)-m\left(x\right)\right| = O_p\left(L_n\right)$, for any $\varepsilon>0$, there exists $\delta>0$, such that for all $n>N_\varepsilon$,

$$P\left(L_n^{-1}\sup_x\left|m\left(x\right)/\tilde{m}\left(x\right)-1\right|<\delta\right)>1-\varepsilon. \tag{3.26}$$

When $\max_i\left[m\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1\geq 0$, $L_n^{-1}\left|\max_i\left[m\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1\right|$ $\leq L_n^{-1}\sup_x\left|m\left(x\right)/\tilde{m}\left(x\right)-1\right|$. By inequality (3.26),

$$P\left(L_n^{-1}\left|\max_i\left[m\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1\right|<\delta\right)>1-\varepsilon.$$

When $\max_i\left[m\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1<0$,

$$\left|\max_i\left[m\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1\right|\leq 1-\max_i R_i\inf_x\left[m\left(x\right)/\tilde{m}\left(x\right)\right]$$

and

$$P\left(L_n^{-1}\left|\max_i\left[m\left(X_i\right)R_i/\tilde{m}\left(X_i\right)\right]-1\right|<\delta\right)\geq P\left(\max_i R_i\inf_x\left[m\left(x\right)/\tilde{m}\left(x\right)\right]>1-L_n\delta\right).$$

By the assumption that $1-\max_i R_i = O_p(L_n)$ and inequality (3.26), for any $\varepsilon > 0$, there exists $\delta_1$ and $\delta_2 > 0$, such that for all $n > N_\varepsilon$, $P\left(\inf_x [m(x)/\tilde{m}(x)] > 1 - L_n\delta_1\right) > 1-\varepsilon$ and $P\left(\max_i R_i > 1 - L_n\delta_2\right) > 1-\varepsilon$. Therefore, for any $\varepsilon > 0$, there exists $\delta_3 > 0$, such that for all $n > N_\varepsilon$, $P\left(\max_i [m(X_i) R_i/\tilde{m}(X_i)] > 1 - L_n\delta_3\right) > 1-\varepsilon$. ■

## Proof of Theorem 14

By definition, we have

$$\begin{aligned}
\tilde{\rho}(x) - \rho(x) &= \tilde{m}(x)/\tilde{\mu}_R - m(x)/\mu_R \\
&= \tilde{m}(x)/\tilde{\mu}_R - \tilde{m}(x)/\mu_R + \tilde{m}(x)/\mu_R - m(x)/\mu_R \\
&= \tilde{m}(x)\left(\tilde{\mu}_R^{-1} - \mu_R^{-1}\right) + \mu_R^{-1}[\tilde{m}(x) - m(x)].
\end{aligned}$$

By Theorems 12 and 13 ,

$$\sup_x \left\{\mu_R^{-1}[\tilde{m}(x) - m(x)]\right\} = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right)$$

and

$$\sup_x \left[\tilde{m}(x)\left(\tilde{\mu}_R^{-1} - \mu_R^{-1}\right)\right] = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right).$$

Therefore, we conclude that $\sup_x [\tilde{\rho}(x) - \rho(x)] = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right)$. ■

## Proof of Theorem 15

When each $m_l$ is monotone increasing, Theorem 2 in Chapter 2 states that for $p \leq 3$, the coefficients of the unconstrained estimator $\tilde{m}_l$ satisfy the monotone constraints for large sample size. Similarly, when each $m_l$ is concave, for $p \leq 3$ Lemma 21 indicates that the coefficients of the $\tilde{m}_l$ satisfy the concave constraints with large sample size. These imply that for $p \leq 3$ the unconstrained estimator $\tilde{m}_l$ and the shape constrained estimator $\hat{m}_l$ are identical when the sample size is large enough. Therefore, $\hat{m}_l$ enjoys the same asymptotic properties with $\tilde{m}_l$, then by Theorem 12, we have $\sup_x |\hat{m}_l(x) - m_l(x)| = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right)$. Moreover, under the assumption that $1 - \max_i R_i = O_p\left(\sqrt{N_n^2/n + N_n^{-2p-1}}\right)$, the results follow directly from Theorem 13 and 14. ■

## 3.5 Tables and Figures

Table 3.1: Univariate case: averaged integrated squared errors (AISE) of frontier functions under three different experimental designs **without outliers** using four estimation methods: local linear regression (LLR), unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), and monotone and concave constrained linear spline (MCCLS).

| Method | n | $\beta = \frac{1}{3}$ | $\beta = 1$ | $\beta = 3$ |
|---|---|---|---|---|
| LLR | 100 | 0.3259 | 0.1193 | 0.0856 |
| | 250 | 0.1843 | 0.0412 | 0.0313 |
| | 500 | 0.0771 | 0.0227 | 0.0152 |
| ULS | 100 | 0.0207 | 0.0485 | 0.1176 |
| | 250 | 0.0129 | 0.0325 | 0.0438 |
| | 500 | 0.0082 | 0.0270 | 0.0246 |
| MCLS | 100 | 0.0174 | 0.0322 | 0.0876 |
| | 250 | 0.0114 | 0.0285 | 0.0358 |
| | 500 | 0.0075 | 0.0242 | 0.0218 |
| MCCLS | 100 | 0.0174 | 0.0282 | 0.0796 |
| | 250 | 0.0140 | 0.0263 | 0.0313 |
| | 500 | 0.0139 | 0.0215 | 0.0184 |

Table 3.2: Univariate case: averaged integrated squared errors (AISE) of the mean function $m(\cdot)$ and mean squared errors (MSE) of the parameter $1/\mu_R$ under three different experimental designs **without outliers** using three spline estimation methods: unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), and monotone and concave constrained linear spline (MCCLS).

| | | $\beta = \frac{1}{3}$ | | $\beta = 1$ | | $\beta = 3$ | |
|---|---|---|---|---|---|---|---|
| Method | n | $m(x)$ | $1/\mu_R$ | $m(x)$ | $1/\mu_R$ | $m(x)$ | $1/\mu_R$ |
| ULS | 100 | 0.0051 | 0.0117 | 0.0074 | 0.0653 | 0.0062 | 0.2196 |
| | 250 | 0.0026 | 0.0063 | 0.0036 | 0.0456 | 0.0024 | 0.0557 |
| | 500 | 0.0015 | 0.0042 | 0.0023 | 0.0321 | 0.0014 | 0.0151 |
| MCLS | 100 | 0.0042 | 0.0101 | 0.0061 | 0.0405 | 0.0040 | 0.2046 |
| | 250 | 0.0023 | 0.0057 | 0.0034 | 0.0404 | 0.0019 | 0.0610 |
| | 500 | 0.0015 | 0.0039 | 0.0021 | 0.0301 | 0.0013 | 0.0147 |
| MCCLS | 100 | 0.0038 | 0.0105 | 0.0052 | 0.0425 | 0.0033 | 0.2171 |
| | 250 | 0.0025 | 0.0074 | 0.0029 | 0.0397 | 0.0016 | 0.0560 |
| | 500 | 0.0021 | 0.0078 | 0.0019 | 0.0285 | 0.0011 | 0.0163 |

Table 3.3: Univariate case: averaged integrated squared errors (AISE) of the frontier function and mean squared errors (MSE) of the parameter $1/\mu_R$ under three different experimental designs **with outliers** and sample size $n = 50$ and 250.

| | n | | 50 | | 250 | |
|---|---|---|---|---|---|---|
| Robust (N/Y) | Method | $\beta$ | Frontier | $1/\mu_R$ | Frontier | $1/\mu_R$ |
| N | LLR | 1/3 | 68.3250 | NA | 3.8018 | NA |
| | | 1 | 13.6266 | NA | 1.3963 | NA |
| | | 3 | 3.9625 | NA | 1.2244 | NA |
| | ULS | 1/3 | 2.1778 | 1.2629 | 2.5235 | 1.8358 |
| | | 1 | 2.3620 | 2.5041 | 2.6167 | 4.1562 |
| | | 3 | 3.9186 | 9.1302 | 2.5637 | 14.3460 |
| | MCLS | 1/3 | 2.3131 | 1.3654 | 2.4847 | 1.8079 |
| | | 1 | 2.3626 | 2.6171 | 2.6368 | 4.2044 |
| | | 3 | 3.0650 | 7.9891 | 2.5305 | 14.4676 |
| | MCCLS | 1/3 | 2.3104 | 1.3637 | 2.6178 | 1.9081 |
| | | 1 | 2.3446 | 2.5998 | 2.5661 | 4.0956 |
| | | 3 | 3.0595 | 7.9839 | 2.4068 | 13.7622 |
| Y | LLR | 1/3 | 3.1842 | NA | 0.5935 | NA |
| | | 1 | 1.1308 | NA | 0.2340 | NA |
| | | 3 | 0.5706 | NA | 0.0983 | NA |
| | ULS | 1/3 | 0.0526 | 0.0096 | 0.0161 | 0.0058 |
| | | 1 | 0.1930 | 0.0962 | 0.0391 | 0.0370 |
| | | 3 | 0.2622 | 1.1732 | 0.0467 | 0.1023 |
| | MCLS | 1/3 | 0.0296 | 0.0044 | 0.0153 | 0.0056 |
| | | 1 | 0.0747 | 0.0639 | 0.0327 | 0.0295 |
| | | 3 | 0.1180 | 1.3090 | 0.0336 | 0.1074 |
| | MCCLS | 1/3 | 0.0296 | 0.0044 | 0.0178 | 0.0069 |
| | | 1 | 0.0748 | 0.0636 | 0.0320 | 0.0296 |
| | | 3 | 0.1183 | 1.3069 | 0.0332 | 0.1060 |

Table 3.4: Multivariate case: averaged integrated squared errors (AISE) of frontier functions $\{\rho_l(\cdot)\}_{l=1}^4$ and the mean function $m(\cdot)$, and mean squared errors (MSE) of the parameter $1/\mu_R$ under three different experimental designs **without outliers** using three spline estimation methods: unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), and monotone and concave constrained linear spline (MCCLS).

| Setting | Method | n | $\rho_1(x)$ | $\rho_2(x)$ | $\rho_3(x)$ | $\rho_4(x)$ | $m(x)$ | $1/\mu_R$ |
|---|---|---|---|---|---|---|---|---|
| $\beta = 1/3$ | ULS | 100 | 0.3709 | 0.3913 | 0.4200 | 0.4211 | 0.7733 | 0.0335 |
| | | 250 | 0.0972 | 0.1000 | 0.1136 | 0.1110 | 0.1964 | 0.0273 |
| | | 500 | 0.0455 | 0.0535 | 0.0665 | 0.0685 | 0.1183 | 0.0192 |
| | MCLS | 100 | 0.1063 | 0.1116 | 0.1222 | 0.1072 | 0.2411 | 0.0184 |
| | | 250 | 0.0593 | 0.0623 | 0.0579 | 0.0669 | 0.1389 | 0.0158 |
| | | 500 | 0.0321 | 0.0365 | 0.0461 | 0.0446 | 0.0952 | 0.0127 |
| | MCCLS | 100 | 0.0729 | 0.0819 | 0.0878 | 0.0729 | 0.1924 | 0.0144 |
| | | 250 | 0.0381 | 0.0471 | 0.0437 | 0.0559 | 0.1254 | 0.0127 |
| | | 500 | 0.0226 | 0.0273 | 0.0381 | 0.0415 | 0.0938 | 0.0112 |
| $\beta = 1$ | ULS | 100 | 1.7966 | 1.9123 | 1.8255 | 1.9192 | 1.4766 | 0.1613 |
| | | 250 | 0.6512 | 0.7733 | 0.7137 | 0.5614 | 0.4102 | 0.2310 |
| | | 500 | 0.2122 | 0.2206 | 0.2495 | 0.2664 | 0.1757 | 0.1388 |
| | MCLS | 100 | 0.4333 | 0.4107 | 0.3987 | 0.4365 | 0.4129 | 0.1528 |
| | | 250 | 0.2151 | 0.2661 | 0.2681 | 0.1865 | 0.2149 | 0.1215 |
| | | 500 | 0.1151 | 0.1030 | 0.1301 | 0.1196 | 0.1039 | 0.0739 |
| | MCCLS | 100 | 0.3164 | 0.3023 | 0.3085 | 0.3441 | 0.3319 | 0.1346 |
| | | 250 | 0.1771 | 0.2127 | 0.2436 | 0.1682 | 0.1921 | 0.1254 |
| | | 500 | 0.0896 | 0.0857 | 0.1095 | 0.0985 | 0.0991 | 0.0611 |
| $\beta = 3$ | ULS | 100 | 3.9584 | 3.7585 | 4.1963 | 3.2029 | 1.3848 | 0.3845 |
| | | 250 | 1.3849 | 1.2420 | 1.2552 | 1.3002 | 0.3562 | 0.0525 |
| | | 500 | 0.4669 | 0.4942 | 0.4227 | 0.4770 | 0.1394 | 0.0253 |
| | MCLS | 100 | 0.7481 | 0.5459 | 0.9112 | 0.6893 | 0.3529 | 0.1692 |
| | | 250 | 0.4383 | 0.3251 | 0.3839 | 0.3972 | 0.1506 | 0.0421 |
| | | 500 | 0.2391 | 0.2305 | 0.2258 | 0.1882 | 0.0829 | 0.0280 |
| | MCCLS | 100 | 0.5816 | 0.4565 | 0.6605 | 0.5216 | 0.2830 | 0.1932 |
| | | 250 | 0.3184 | 0.2727 | 0.2938 | 0.3279 | 0.1165 | 0.0409 |
| | | 500 | 0.1933 | 0.1733 | 0.1861 | 0.1413 | 0.0659 | 0.0305 |

Table 3.5: Multivariate case: averaged integrated squared errors (AISE) of frontier functions $\{\rho_l(\cdot)\}_{l=1}^4$ and mean squared errors (MSE) of the parameter $1/\mu_R$ under three different experimental designs **with outliers** and sample size $n = 250$.

| Robust (N/Y) | Method | $\beta$ | $\rho_1(x)$ | $\rho_2(x)$ | $\rho_3(x)$ | $\rho_4(x)$ | $1/\mu_R$ |
|---|---|---|---|---|---|---|---|
| N | ULS | 1/3 | 0.2443 | 0.2497 | 0.2801 | 0.2773 | 0.4312 |
| | | 1 | 3.5493 | 1.9495 | 3.5332 | 3.2385 | 5.9337 |
| | | 3 | 228.5200 | 193.8965 | 64.1956 | 140.7903 | 800.3386 |
| | MCLS | 1/3 | 0.2137 | 0.2122 | 0.2152 | 0.2195 | 0.4809 |
| | | 1 | 0.5556 | 0.6298 | 0.7567 | 0.5675 | 1.4698 |
| | | 3 | 1.7420 | 1.1424 | 1.2625 | 1.5039 | 6.0256 |
| | MCCLS | 1/3 | 0.1932 | 0.1969 | 0.1977 | 0.2054 | 0.4782 |
| | | 1 | 0.4385 | 0.4831 | 0.6144 | 0.4373 | 1.1451 |
| | | 3 | 1.4518 | 1.0076 | 1.0545 | 1.2604 | 4.9243 |
| Y | ULS | 1/3 | 0.1064 | 0.1187 | 0.1334 | 0.1387 | 0.0613 |
| | | 1 | 0.7613 | 0.8911 | 0.8691 | 0.7025 | 0.4747 |
| | | 3 | 1.4437 | 1.2105 | 1.4161 | 1.3748 | 0.0759 |
| | MCLS | 1/3 | 0.0762 | 0.0871 | 0.0790 | 0.0919 | 0.0654 |
| | | 1 | 0.2925 | 0.3704 | 0.3811 | 0.2692 | 0.4612 |
| | | 3 | 0.3869 | 0.3133 | 0.3522 | 0.4041 | 0.0616 |
| | MCCLS | 1/3 | 0.0729 | 0.0839 | 0.0798 | 0.0941 | 0.0873 |
| | | 1 | 0.2673 | 0.3254 | 0.3656 | 0.2541 | 0.4577 |
| | | 3 | 0.3612 | 0.2964 | 0.3256 | 0.3819 | 0.0630 |

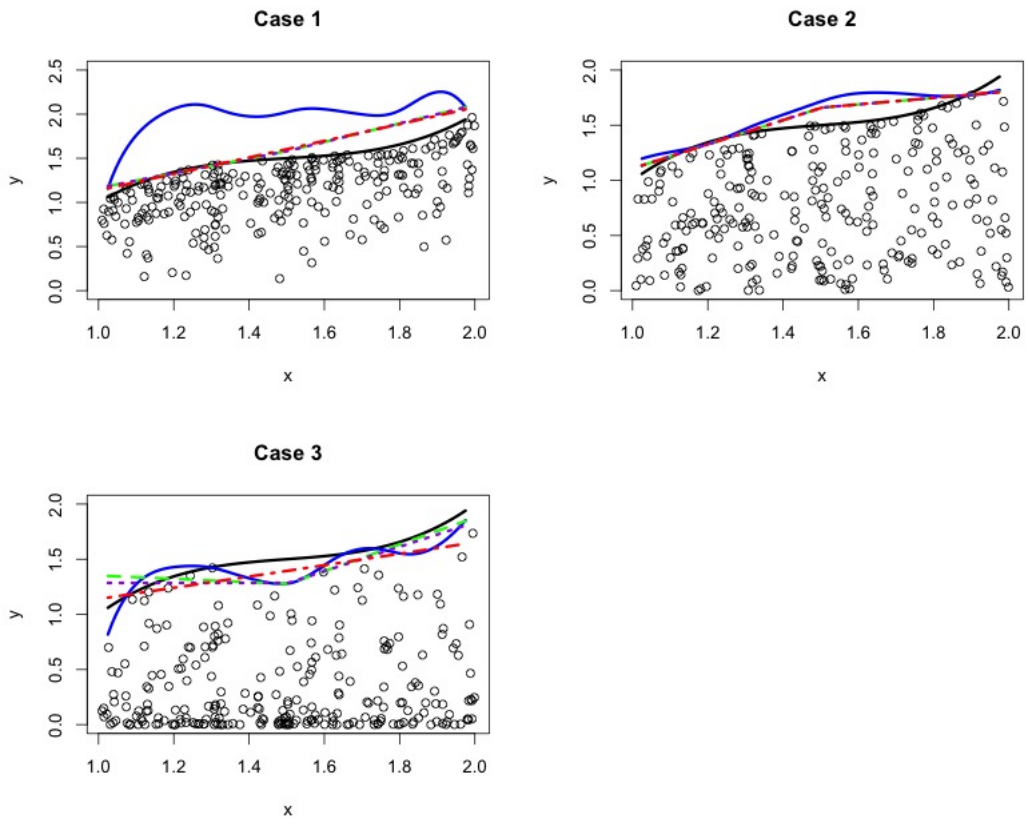


Figure 3.1: Simulation results of the frontier functions **without outliers** and sample size $n = 250$. Cases 1, 2, and 3 correspond to the three different experimental designs with the scale parameter $\beta = 1/3$, 1, and 3, respectively. The solid black line represents the true curve, while the solid blue, dashed green, dotted purple and dot-dashed red lines represent the fitted curves of one simulated data set obtained using local linear regression (LLR), unconstrained linear spline (ULS), monotone constrained linear spline (MCLS) and monotone and concave constrained linear spline (MCCLS), respectively.

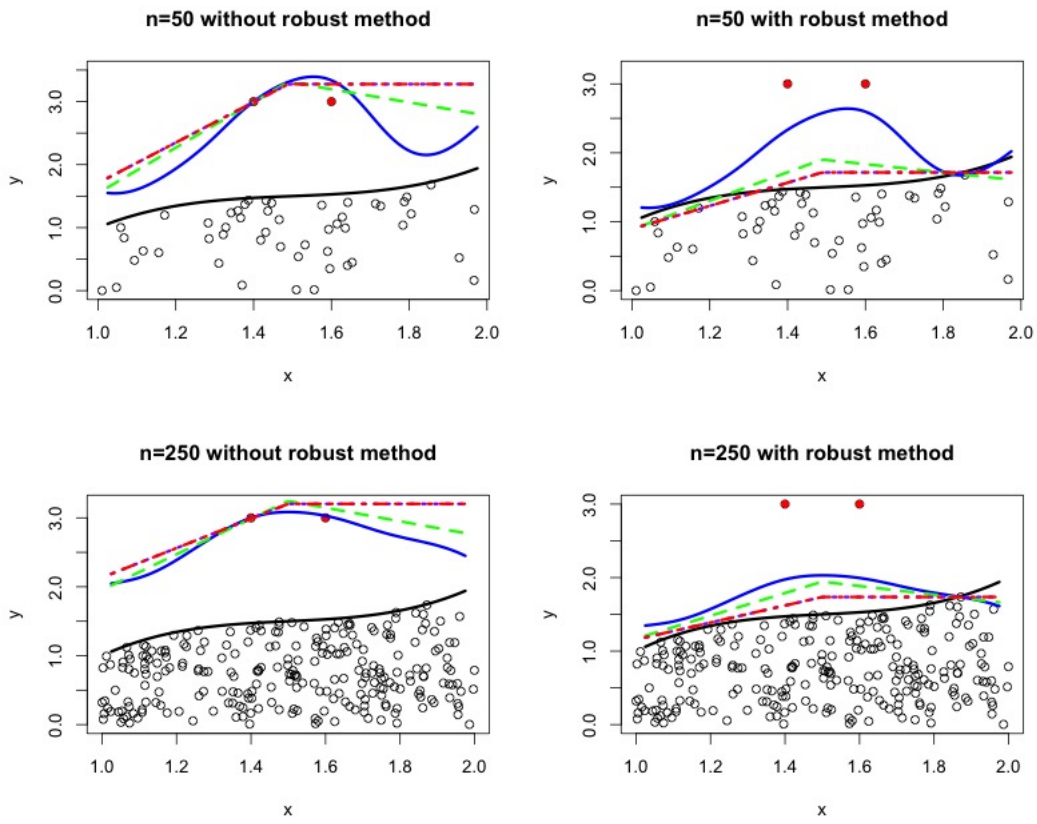


Figure 3.2: Simulation results of the frontier functions under the experimental design where $\beta = 1$ **with outliers** and for sample size $n = 50$ and 250. The solid black line represents the true curve, while the solid blue, dashed green, dotted purple and dot-dashed red lines represent fitted curves of one simulated data set obtained using local linear regression (LLR), unconstrained linear spline (ULS), monotone constrained linear spline (MCLS) and monotone and concave constrained linear spline (MCCLS), respectively. The solid red circles represent two artificial outliers.

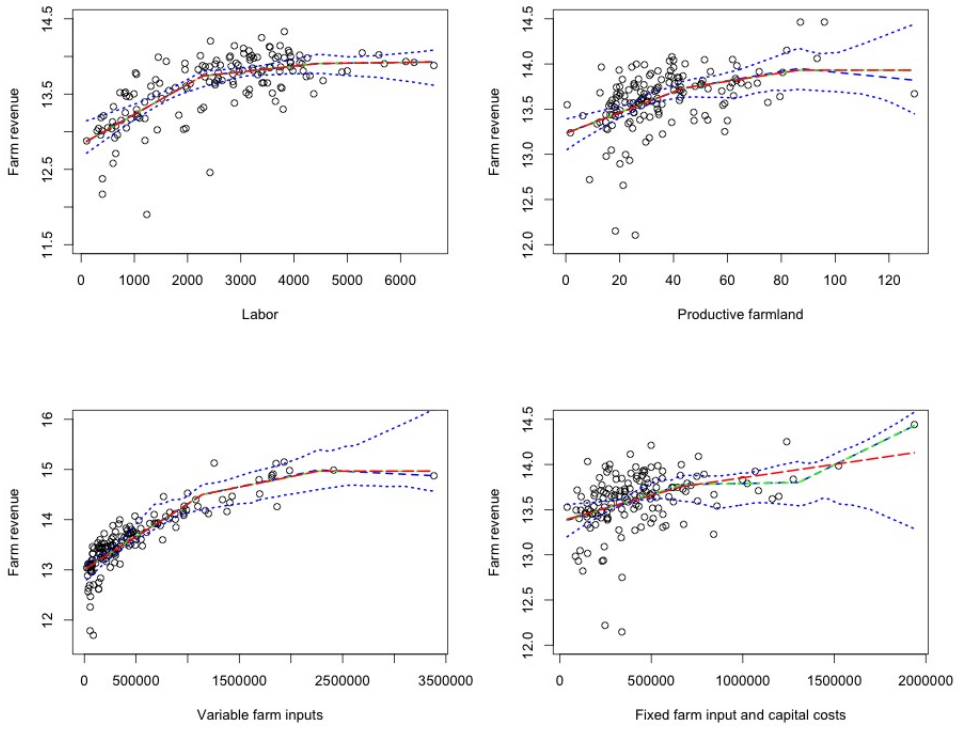


Figure 3.3: Norwegian Farm data: fitted results for each input variable obtained in the first estimation step. The black circle represents the pseudo response. In each plot, the dashed blue, dot-dashed green, and long-dashed red lines represent estimated mean function using unconstrained linear spline (ULS), monotone constrained linear spline (MCLS) and monotone and concave constrained linear spline (MCCLS), respectively. The dotted blue lines represent the 95% point-wise confidence intervals from 100 boostrapped samples using the ULS method.

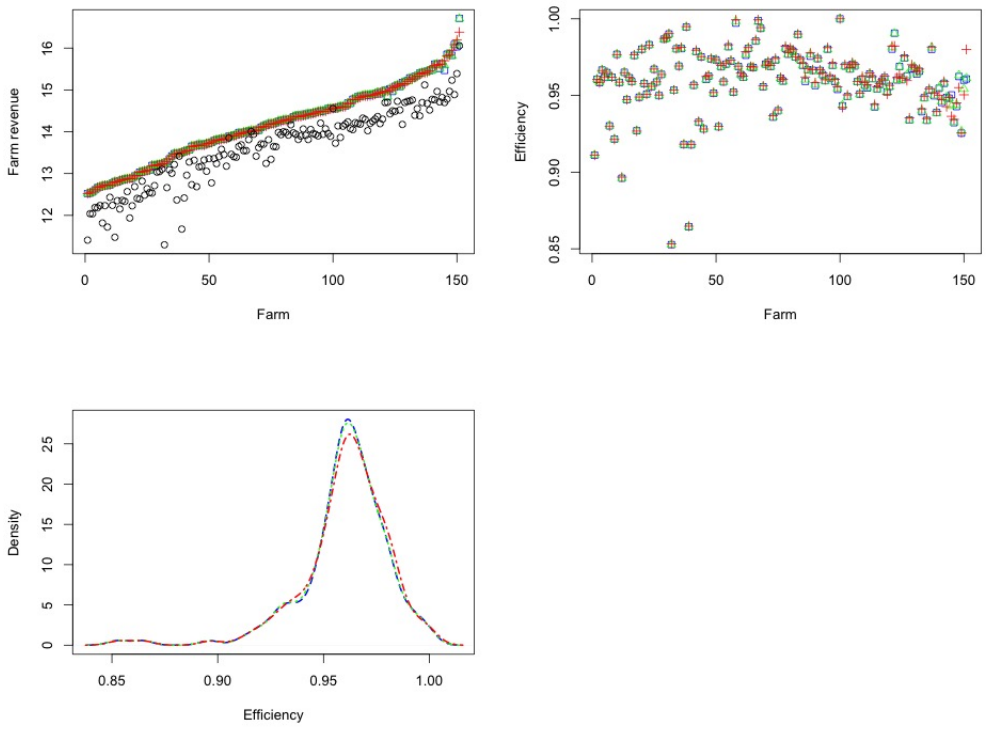


Figure 3.4: Norwegian Farm data: estimated maximum farm revenue (left top), efficiency estimates (right top), and the kernel density distribution of the efficiency estimates (left bottom). In the top figures, the blue rectangle, green triangle and red plus represent estimated maximum revenue or efficiency of all 151 farms using unconstrained linear spline (ULS), monotone constrained linear spline (MCLS) and monotone and concave constrained linear spline (MCCLS), respectively. The true farm revenue is denoted by the black circle. In the left bottom figure, the dashed blue, dot-dashed green, and long-dashed red lines represent kernel density distribution of the efficiency estimates using ULS, MCLS, and MCCLS, respectively.

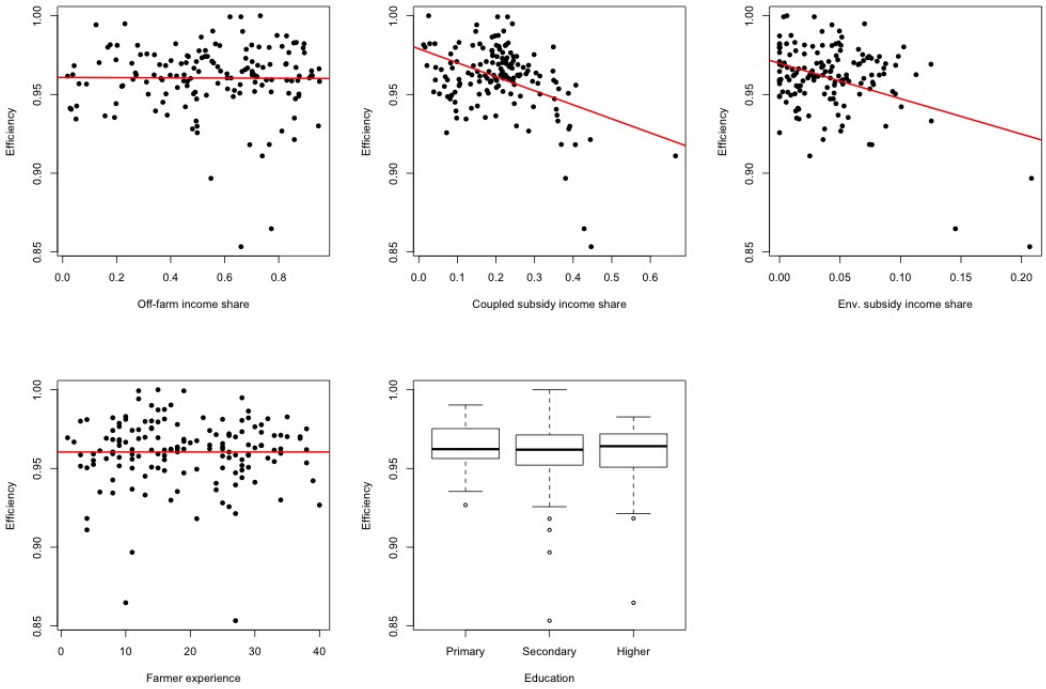


Figure 3.5: Norwegian Farm data: the exploration of relationships between the efficiency and 5 other explanatory variables of interest. The scatter plots of off-farm income share, coupled subsidy income share, environmental subsidy income share, and farmer experience versa the efficiency estimates obtained using the **MCCLS** method for the 151 farms, in which the solid red lines represent the linear least squares regression lines. Box plots of the estimated efficiency in the subgroups with different education levels of the farmers.

# 4 Discussion

In this dissertation, we first propose a one-step backfitted constrained polynomial spline estimation method for the monotone additive models. The proposed method approximates nonparametric functions via polynomial splines and obtains monotone estimates by applying a set of simple linear constraints on the spline coefficients. We then extend this constrained method to estimate the production frontier functions and propose a two-step polynomial spline estimation method with monotone or/and concave constraints.

Both proposed estimation methods give smooth estimators with desirable asymptotic properties. They are also easy to implement and fast to compute. Furthermore, our methods are designed for data sets with multi-dimensions of predictor variables while some existing methods only work the univariate case with only one predictor. In addition, the empirical results in Chapters 2 and 3 illustrate that compared with the existing methods (Mammen and Yu 2007 and Martins-Filho and Yao 2007), ours have better numerical performance and are very competitive in terms of computational time and estimation accuracy. In particular, our proposed methods outperform other methods when there are outliers or extreme values.

We believe that we have clearly demonstrated that our proposed one-step backfitted constrained polynomial spline method makes an excellent candidate to solve the problem of estimating additive models with known shape constraints. Our work con-

tributes to the literature by providing a desirable solution to this less studied problem. The proposed method is prospectively promising since it is useful in many scientific areas, such as estimating growth curves or predicting disease risk. A particular application presented in this dissertation is the estimation of production frontier functions. Our proposed frontier estimator can be viewed as a useful alternative to the DEA, FDH and other estimators (Hall et al. 1998, Knight 2001 and Cazals et al. 2002) that have been widely used in practice. In addition, our proposed method takes the shape of frontiers into consideration and this provides a great improvement compared with the existing approaches.

There are few areas that our work can be extended. First, the linear constraints developed in Lemmas 9, 10 and 11 are sufficient but not necessary conditions for polynomial splines to be monotone or concave. This implies that our constrained optimization is performed over a subspace of the true monotone or concave spline functions. Better estimation results can be obtained if the constrained optimization is performed over the larger space of all monotone or concave spline functions. Papp and Alizadeh (2012) gave the sufficient and necessary conditions for spline functions to be monotone or concave. Their approach is based on a characterization of nonnegative polynomials that leads to the requirement for solving optimization problems with semi-definite and second order conic constraints. For future work, it is desirable to adopt such approaches in conducting constrained optimization and compare it with our current results. But the computation of their method will be more challenging than our approach since our method takes advantage of linear programming.

Furthermore, we note that the asymptotic properties of our proposed shape con-

strained estimators are developed only up to cubic splines. It is desirable to develop a more general theoretical result that is applicable to higher orders of polynomial splines. The challenge here is to construct a spline function that satisfies our linear constraints and still provides good approximation to smooth functions with shape constraints.

In addition, the Norwegian Farm data contains an unbalanced set of farm-level panel data. In our analysis, we only focus on a one-year data rather than the whole panel. It will be interesting in future research to generalize our methods to analyze panel data.

Lastly, there is some other potential work to consider regarding the estimation of frontier functions. Since our method is constructed under the deterministic frontier framework, the extension or generalization of the additive frontier model and proposed estimator to accommodate stochastic frontier analysis is desirable. Furthermore, another direction for future research is to make an extension to the case with multiple output variables since this is not well explored in the existing literature.

# Bibliography

Aigner, D., Lovell, C. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6, 21 - 37.

Aragon, Y., Daouia, A., and Thomas-Agnan, C. (2005). Nonparametric frontier estimation: a conditional quantile-based approach. *Econometric Theory*, 21, 358 - 389.

Barr, R. S., Seiford, L. M., and Siems, T. F. (1994). Forecasting bank failure: a non-parametric frontier estimation approach. *Recherches Économiques de Louvain/Louvain Economic Review*, 417 - 429.

Boni, M. F., Posada, D., and Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, 176, 1035 - 1047.

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

de Boer, W. J., Besten, P. J., and Ter Braak, C. F. (2002). Statistical analysis of sediment toxicity by additive monotone regression splines. *Ecotoxicology*, 11, 435 - 450.

Cazals, C., Florens, J. P., and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106, 1 - 25.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*. 2, 429 - 444.

Daouia, A., Noh, H., and Park, B. U. (2016). Data envelope fitting with constrained polynomial splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 3 - 30.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica: Journal of the Econometric Society*, 19, 273 - 292.

Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labor-efficiency in post offices. In: Marchand, M., Pestiau, P., Tulkens, H. (Eds.). *The Performance of Public Enterprises: Concepts and Measurements*. North-Holland, Amsterdam.

Devore, R. A., and Lorentz, G. G. (1993). *Constructive Approximation*. Berlin Heidelberg: Springer-Varlag.

Dubeau, F., and Savoie, J. (1996). Optimal error bounds for quadratic spline interpolation. *Journal of Mathematical Analysis and Applications*, 198, 49 - 63.

Färe, R., Grosskopf, S. and Lovell, C. A. K. (1985). *The Measurement of Efficiency of Production*. Boston: Kluwer-Nijhoff.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Series A , 120, 253 - 290.

Hall, P., and Park, B. U. (2002). New methods for bias correction at endpoints and boundaries. *Annals of Statistics*, 1460 - 1479.

Hall, P., Park, B. U., and Stern, S. E. (1998). On polynomial estimators of frontiers and boundaries. *Journal of Multivariate Analysis*, 66, 71-98.

Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized additive models*. London: Chapman and Hall.

Hastie, T., and Tibshirani, R. (2000). Bayesian backfitting. *Statistical Science*, 15, 196 - 223.

He, X., and Shi, P. (1998). Monotone B-spline smoothing. *Journal of the American Statistical Association*, 93, 643 - 650.

Hnizdo, V., Darian, E., Fedorowicz, A., Demchuk, E., Li, S., and Singh, H. (2007). Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *Journal of Computational Chemistry*, 28, 655 - 668.

Horowitz, J. L., and Mammen, E. (2004). Nonparametric estimation of an additive model with a link function. *The Annals of Statistics*, 32, 2412 - 2443.

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

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics*, 31, 1600 - 1635.

Jeong, S. O., and Simar, L. (2006). Linearly interpolated FDH efficiency score for non-convex frontiers. *Journal of Multivariate Analysis*, 97, 2141 - 2161.

Knight, K. (2001). Limiting distributions of linear programming estimators. *Extremes*, 4, 87 - 103.

Koopmans, T.C. (1951). *Activity Analysis of Production and Allocation*. Wiley: New York.

Kumbhakar, S. C., Lien, G., and Hardaker, J. B. (2014). Technical efficiency in competing panel data models: a study of Norwegian grain farming. *Journal of Productivity Analysis*, 41, 111 - 129.

Leitenstorfer, F., and Tutz, G. (2007). Generalized monotonic regression based on B-splines with an application to air pollution data. *Biostatistics*, 8, 654 - 673.

Leys, C., and Schumann, S. (2010). A nonparametric method to analyze interactions: The adjusted rank transform test. *Journal of Experimental Social Psychology*, 46, 684 - 688.

Linton, O. B., and Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika*, 83, 529 - 540.

Linton, O., and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82, 93 - 100.

Liu, R., and Yang, L.(2010). Spline-backfitted kernel smoothing of additive coefficient model. *Econometric Theory*, 26, 29 - 59.

Liu, R., Yang, L., and Härdle, W. (2013). Oracally efficient two-step estimation of generalized additive model. *Journal of the American Statistical Association*, 108, 619 - 631.

Ma, S., and Racine, J. (2013). Additive Regression Splines with Irrelevant Categorical and Continuous Regressors. *Statistica Sinica*, 23, 515-541.

Ma, S., and Yang, L. (2011). Spline-backfitted Kernel Smoothing of Partially Linear Additive Model. *Journal of Statistical Planning and Inference*, 141, 204-219.

Mammen, E., Linton, O., and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics*, 27, 1443 - 1490.

Mammen, E., and Yu, K. (2007). Additive isotone regression. *Lecture Notes-Monograph Series*, 55, 179 -195.

Martins-Filho, C., and Yao. F (2007). Nonparametric frontier estimation via local linear regression. *Journal of Econometrics*, 141(1), 283 - 319.

Meeusen, W., and Van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 435 - 444.

Morton-Jones, T., Diggle, P., Parker, L., Dickinson, H. O. and Blinks, K. (2000). Additive isotonic regression models in epidemiology. *Statistics in Medicine*, 19, 849 - 859.

Opsomer, J. D., and Ruppert D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, 25, 186 - 211.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, 425 - 441.

Schumaker, L. L. (2007). *Spline Functions: Basic Theory*. Cambridge: Cambridge University Press.

Schumaker, L. L. (2015). *Spline Functions: Computational Methods*. Vol. 142. SIAM, 2015.

Shephard, R.W. (1970). *Theory of Cost and Production Functions*. Princeton: Princeton University Press.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *Medical Imaging, IEEE Transactions on*, 17, 87 - 97.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics*, 13, 689 - 705.

Tutz, G., and Leitenstorfer, F. (2007). Generalized smooth monotonic regression in additive modeling. *Journal of Computational and Graphical Statistics*, 16, 165 - 188.

Van Bergeijk, K. E., Noordijk, V. H., Lembrechts, J., and Frissel, M. J. (1992). Influence of pH, soil type and soil organic matter content on soil-to-plant transfer of radiocesium and-strontium as analyzed by a nonparametric method. *Journal of Environmental Radioactivity*, 15, 265 - 276.

Wang, L., and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *The Annals of Statistics*, 35, 2474 - 2503.

Wu, X., and Sickles, R. (2013), Semiparametric estimations under shape constraints with applications to production functions. Working paper, Department of Agricultural Economics, Texas A&M University, College Station, Texas.

Xue, L., and Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, 16, 1423 - 1446.