Journal of Nonparametric Statistics
Vol. 00, No. 00, Month 2019, 1–27

RESEARCH ARTICLE

Estimation of additive frontier functions with shape constraints

Lu Wang[a], Lan Xue[b*] and Lijian Yang[c]

[a]*Beijing International Center for Mathematical Research, Peking University, Beijing, China 100871*; [b]*Department of Statistics, Oregon State University, Corvallis, OR 97330, USA*; [c] *Center for Statistical Science & Department of Industrial Engineering, Tsinghua University, Beijing 100084, China.*

(*v3.7 released September 2009*)

Production frontier is an important concept in modern economics and has been widely used to measure production efficiency. Existing nonparametric frontier models often only allow one or low-dimensional input variables due to "curse-of-dimensionality". In this paper we propose a flexible additive frontier model which quantifies the effects of multiple input variables on the maximum output. In addition, we consider the estimation of the nonparametric frontier functions with shape restrictions. Economic theory often imposes shape constraints on production frontier, such as, monotonicity and concavity. A two-step constrained polynomial spline method is proposed to give smooth estimates that automatically satisfy such shape constraints. The proposed method is not only easy to compute, but also more robust to outliers. In theory, we established uniform consistency of the proposed method. We illustrate the proposed method by both simulation studies and an application to the Norwegian farm data. The numerical studies suggest that the proposed method has superior performance by incorporating shape constraints.

Keywords: Back-fitting; consistency; nonparametric frontier; Norwegian farm data; polynomial Spline

AMS Subject Classification: 62G05; 62G08; 62G20

*Corresponding author. Email: xuel@stat.oregonstate.edu

ISSN: 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/1048525YYxxxxxxxx
http://www.informaworld.com

1. Introduction

Estimation of production frontiers is becoming an increasingly popular research topic in economics and statistics in recent years. Since the seminal work of Debreu (1951) and Koopmans (1951), a large amount of literature has been developed on specification and estimation of production frontiers, and on measurement of the associated technical efficiency of production units.

Consider a non-negative vector $(\mathbf{x}, y) \in \mathcal{R}_+^d \times \mathcal{R}_+$, where $\mathbf{x}$ represents the d inputs used in production and y represents the output. According to economic theory (Koopmans 1951, Shephard 1970), the production set is defined as $\Psi = \left\{(\mathbf{x}, y) \in \mathcal{R}_+^d \times \mathcal{R}_+ | \ \mathbf{x} \text{ can produce } y\right\}$, i.e., the set of physically attainable points $(\mathbf{x}, y)$. The production function or frontier of Ψ is defined as $\rho(\mathbf{x}) = \sup\{y, (\mathbf{x}, y) \in \Psi\}$, which 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 level of the output for a firm working at the level of inputs $\mathbf{x}$, and presents a useful benchmark value or reference frontier which can be used to assess efficiency of 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 $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$.

In the existing literature, two main classes of models have been proposed for the production frontier function: the deterministic frontier model and the stochastic approach. The deterministic frontier models rely on the assumption that all data lie in the production set Ψ, i.e., $P\{(\mathbf{X}_i, Y_i) \in \Psi\} = 1$, for $i = 1, \ldots, n$. Two nonparametric estimators were developed under this framework: data envelopment analysis (DEA) (Charnes et al. 1978, Farrell 1957) and free disposal hull (FDH) (Deprins et al. 1978). 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), Hall and Park (2002), Jeong and Simar (2006) and Knight (2001), which employed methods such as local polynomial or piecewise polynomials to obtain frontier estimates that are smooth. These methods are appealing since they rely on very few assumptions on Ψ 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 envelope all data points by construction. The stochastic approach, initiated by Aigner et al. (1977) and Meeusen and Van den Broeck (1977), allows some observations to be outside of the production set. However, the 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 which is nonparametric in nature and flexible to capture complex structure of the production frontier. It is also 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 polynomial smoothing. However due to "the curse of dimensionality", their method can only accommodate low dimension of input variables. In addition, econometric theories often impose shape constraints on the frontier functions. The general production axiom of free disposability of inputs and outputs (Färe et al. 1985, Shephard 1970) implies that the frontier function $\rho(\cdot)$ is monotone. The convexity of the production set Ψ implies that $\rho(\cdot)$ is also concave, corresponding to diminishing marginal returns. Therefore, it is of interest to provide frontier estimates that automatically satisfy such shape constraints.

In this paper, we extend the nonparametric regression frontier model in Martins-Filho and Yao (2007) for multiple input variables and impose an additive structure on the frontier functions to circumvent "the curse of dimensionality". Furthermore, we consider the estimation of frontier functions under shape constraints, which is not well studied in the existing literature. The nonparametric methods, such as DEA and FDH, give monotone estimates. However their estimates are step-wise and hard to interpret, and as pointed out in Daouia et al. (2016), they often underestimate the true support boundary. More recently, Wu and Sickles (2018) proposed a monotone spline estimator (Ramsay 1988) 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 envelope the data set. However, they only focus on single input case. In addition, similar to DEA and FDH, their method is sensitive to outliers since it is constructed to envelope all data points.

In this paper, we propose an additive frontier model and a two-step constrained polynomial spline method for the frontier estimation. Our proposed method guarantees a smooth estimator and is easy to implement. Most importantly, we incorporate the shape constraints (monotonicity and/or concavity) to capture the shape of frontier more accurately. The simulation studies illustrate that our proposed method with shape constraints is effective in enhancing the estimation accuracy. In addition, the proposed estimator is more robust to the outliers than the estimator proposed in Martins-Filho and Yao (2007).

This paper is organized as follows. Section 2 describes the additive frontier model that we are interested in. In section 3, we propose a two-step polynomial spline method for the estimation of the frontier functions. Asymptotic properties of the proposed estimator are also established. Section 4 introduces several extensions of the additive frontier model to allow interactions among input variables. Section 5 includes simulation studies and the application of the proposed method to a real data set. The assumptions and technical proofs are contained in the appendix.

2. Additive Frontier Model

In this paper, we develop an additive frontier model which 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 observations independently generated from the additive frontier model

$$Y_i = \rho\left(\mathbf{X}_i\right) R_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where Y_i is the output variable and $\mathbf{X}_i = \left(X_{i1}, \ldots, X_{id}\right)^T$ are the input variables of the i-th observation. The unobserved random variable R_i represents the efficiency of $i-$th observation, which takes values in $[0, 1]$. The larger value of R_i indicates more efficient production because the realized output Y_i is closer to the production frontier $\rho(\mathbf{X}_i)$. In a special case with $R_i = 1$, the maximum output is obtained. In addition, we assume the frontier function $\rho\left(\mathbf{X}_i\right)$ is of an additive structure and written as $\rho\left(\mathbf{X}_i\right) =$

$\rho_0+\rho_1(X_{i1})+\ldots+\rho_d(X_{id})$, where ρ_0 is an unknown intercept and $\{\rho_l(\cdot)\}_{l=1}^d$ are unknown univariate nonparametric functions which quantify the effects of input variables on the maximum output. For model identification, we assume that each additive component $\rho_l(\cdot)$ is theoretically centered with $\mathrm{E}[\rho_l(X_l)]=0$, for $l=1,\ldots,d$. Furthermore, we assume that $\mathrm{E}(R_i|\mathbf{X}_i)=\mu_R\in(0,1)$ and $\mathrm{var}(R_i|\mathbf{X}_i)=\sigma_R^2$. The parameter μ_R is viewed as the mean efficiency given the production set and σ_R is the scale parameter for the distribution of R.

In the proposed additive frontier model (2.1), there is no strong parametric restriction on the production set Ψ since the frontier function $\rho(\cdot)$ is not constrained to a specific parametric family. Compared with the nonparametric frontier model in Martins-Filho and Yao (2007), our additive frontier model restricts the contribution of each input variable on the frontier to be additive. The proposed additive frontier model enjoys the advantages of additive models and is unaffected by the curse of dimensionality. However, the additive assumption can be restrictive when the marginal output of a given input depends on other input variables. In section 4, several extensions of the additive frontier model are considered to allow for potential interactions among input variables.

Our main interest lies in the estimation of frontier function $\rho(\cdot)$ of model (2.1). To begin with, we rewrite model (2.1) as

$$Y_i=\rho(\mathbf{X}_i)\mu_R+\rho(\mathbf{X}_i)\varepsilon_i=m(\mathbf{X}_i)+\rho(\mathbf{X}_i)\varepsilon_i,\qquad i=1,\ldots,n, \tag{2.2}$$

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 $\varepsilon_i=R_i-\mu_R$ satisfies $\mathrm{E}(\varepsilon_i|\mathbf{X}_i)=0$ and $var(\varepsilon_i|\mathbf{X}_i)=\sigma_R^2$. In the regression model (2.2), the additive regression function m_j characterizes the shape of ρ_j. These two functions are different only by a scale parameter μ_R. 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 $\{m_l(\cdot)\}_{l=1}^d$ which describe the shape of frontier functions using polynomial splines. Production theory in econometrics often imposes shape constraints on the additive frontier functions $\{\rho_l(\cdot)\}_{l=1}^d$ such as monotonicity and/or concavity. Therefore, in order to capture the shape of frontiers more accurately, we also incorporate shape constraints in the first step. The second step is for estimating the location of the frontier function which is associated with the mean efficiency μ_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. Methodology and Theory

3.1. *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 (2.2) using polynomial splines. Without loss of generality, assume that $\mathbf{X}$ takes value on $[0,1]^d$. Let $u_n=\{0=u_0<u_1<\cdots<u_{N_n}<u_{N_n+1}=1\}$ be a knot sequence with N_n interior knots. The polynomial splines of degree p are polynomial functions with degree p (or less) on the partitioned intervals and $(p-1)$-times differentiable at the interior knots. Using u_n as knots, denote the space of polynomial splines with degree p as $G^p=G^p([0,1],u_n)$ which has di-

mension N_n+p+1. Then 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 $J_n = N_n + p$, see Wang and Xue (2015). 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 $B_{lj} = \tilde{B}_j - n^{-1}\sum_{i=1}^n \tilde{B}_j(X_{il})$. 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 $\hat{m}_0 = \bar{Y} = n^{-1}\sum_{i=1}^n Y_i$. Then we can approximate the theoretically centered nonparametric function $m_l(\cdot)$ by a linear combination of the centered B-spline basis with $m_l(x) \approx \mathbf{B}_l^T(x)\boldsymbol{\beta}_l$ for a set of coefficients $\boldsymbol{\beta}_l = \left(\beta_{l1}, \ldots, \beta_{lJ_n}\right)^T$.

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 $\mathbf{B}_{nl} = (\mathbf{B}_l(X_{1l}), \ldots, \mathbf{B}_l(X_{nl}))^T$, for $l = 1, \ldots, d$. The traditional polynomial spline method (Huang 1998, Stone 1985) estimates the unknown coefficients $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_d^T\right)^T$ by minimizing the sum of squared errors,

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

Then the estimator of unknown function $m_l(\cdot)$ is given as, for $l = 1, \ldots, d$,

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

The traditional polynomial spline estimator (3.4) enjoys the same optimal rate of convergence as an univariate nonparametric function estimator. However, it does not satisfy the shape constraints, such as monotonicity and/or concavity. Therefore, to capture the shape of frontier functions more accurately, we incorporate shape constraints which guarantee our estimates to be monotone and/or concave. In the following, Lemma 3.1 gives a sufficient condition for a polynomial spline to be monotone increasing. 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. Therefore, the sufficient conditions for a polynomial spline to be concave are developed in Lemmas 3.2 and 3.3. For notation simplicity, the sufficient conditions in Lemmas 3.2 and 3.3 are derived using equally spaced knot sequence. Let a polynomial spline $g(x) = \mathbf{B}^T(x)\boldsymbol{\beta}$, where $\mathbf{B}(x)$ is the empirically centered B-spline basis of G^p.

Lemma 3.1: *A sufficient condition for $g(x)$ to be monotone increasing is that its coefficients satisfy $\beta_1 \geq 0$ and $\beta_j \geq \beta_{j-1}$, for $j = 2, ..., N_n + p$.*

Lemma 3.2: *A sufficient condition for a linear spline $g(x)$ to be concave is that its coefficients satisfy the following conditions: if $N_n = 1, \beta_2 - \beta_1 \leq \beta_1$; otherwise, $\beta_2 - \beta_1 \leq \beta_1$ and $\beta_j - \beta_{j-1} \leq \beta_{j-1} - \beta_{j-2}$, for $j = 3, ..., N_n + 1$.*

Lemma 3.3: *A sufficient condition for a polynomial spline $g(x)$ with degree $p \geq 2$ to be concave is that its coefficients satisfy the following conditions:*

$\beta_2 - \beta_1 \leq 2\beta_1$,
$\beta_j - \beta_{j-1} \leq j\beta_{j-1}/(j-1) - \beta_{j-2}$, *for* $j = 3, ..., p-1$
$\beta_p - \beta_{p-1} \leq p\beta_{p-1}/(p-1) - \beta_{p-2}$,
$\beta_j - \beta_{j-1} \leq \beta_{j-1} - \beta_{j-2}$, *for* $j = p+1, ..., N_n + 1$
$\beta_{N_n+2} - \beta_{N_n+1} \leq (p-1)\left(\beta_{N_n+1} - \beta_{N_n}\right)/p$,

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

For the sake of simplicity, we use C_M and C_C to represent the set of spline coefficients that satisfy the monotone increasing conditions and concave conditions respectively. Wang and Xue (2015) developed a one-step backfitted constrained polynomial spline method to estimate monotone additive regression functions. Here, we adopt the same estimation strategy but with more choices of shape constraints. Based on the traditional unconstrained estimator $\{\tilde{m}_l\}_{l=1}^d$, for each $l = 1, ..., d$, we define the l-th pseudo responses $Y_{i,-l}^* = Y_i^* - \sum_{l' \neq l} \tilde{m}_{l'}(X_{il'})$, and $\mathbf{Y}_{-l}^* = \left(Y_{1,-l}^*, ..., Y_{n,-l}^*\right)^T$. Then, to obtain a shape constrained estimator for m_l, we propose to estimate the spline coefficients by minimizing the following constrained least squares

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

where C_l is the set of shape constraints of the spline coefficients for the l-th additive component. For example, taking $C_l = C_M$ gives a monotone estimator, and $C_l = C_M \cap C_C$ guarantees both monotonicity and concavity. In addition, it is flexible to impose different shape constraints for the different additive components. The minimization in (3.5) is a linearly constrained quadratic optimization problem and can be efficiently solved through quadratic programming. Instead of the proposed one-step backfitted procedure, an alternative approach is to impose the shape constraints on all spline coefficients simultaneously in the global least squares (3.3) with one constrained estimation. However, the total number of linear constraints under consideration can be large, therefore lead to unstable numerical performances.

As a result, the shape constrained estimator of $m_l(\cdot)$ is obtained as $\hat{m}_l(x) = \mathbf{B}_l^T(x)\hat{\boldsymbol{\beta}}_l$, for $l = 1, ..., d$, and the regression function $m(\cdot)$ is 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 regression function $m(\cdot)$ and the frontier function $\rho(\cdot)$ are different only by the mean efficiency μ_R which is associated with the location of the frontier function. To estimate μ_R, one observe $\sup \frac{Y}{m(\mathbf{X})} = \sup \frac{\rho(\mathbf{X})R}{\rho(\mathbf{X})\mu_R} = \sup \frac{R}{\mu_R} = \frac{1}{\mu_R}$, since the maximum value of the efficiency R is 1. Therefore, we propose to estimate μ_R by

$$\hat{\mu}_R = \left[\max_i \left(Y_i / \hat{m}\left(\mathbf{X}_i\right)\right)\right]^{-1}. \tag{3.6}$$

In addition, since the estimator in (3.6) is sensitive to the outliers, in our implementation, we propose a robust estimator of μ_R . The details for the robust modification are given in section 3.3. Finally, the regression model in (2.2) implies that $m_l(x_l) = \rho_l(x_l)\mu_R$. Therefore, the additive frontier functions ρ_l can be estimated as $\hat{\rho}_l(x_l) = \hat{m}_l(x_l)/\hat{\mu}_R$ for $l = 1, \ldots, d$, 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 in Martins-Filho and Yao (2007), our proposed method is not only easier to compute, but also allows to capture the shape of frontiers more accurately by the incorporation of the shape constraints. In addition, as illustrated in our simulation studies, the constrained polynomial spline method is more robust to the presence of outliers than

the unconstrained method.

3.2. *Asymptotic Properties*

In this section, we establish the asymptotic properties of the proposed frontier estimator. The assumptions required for asymptotic analysis are presented in the appendix.

Let $\tilde{m}_l$ and $\hat{m}_l$ represent the unconstrained and shape constrained (monotone or concave constrained) estimator of m_l respectively. The corresponding estimators of μ_R and ρ 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 estimator, where we denote the convergence rate $L_n = \sqrt{N_n/n} + N_n^{-p-1}$.

Theorem 3.1 : *Under assumptions (A1)-(A4), (A5) or (A5*), for* $l = 1, ...d$, *as* $n \to \infty$, $\sup_x |\tilde{m}_l(x) - m_l(x)| = O_p(L_n)$.

Theorem 3.2 : *Under assumptions (A1)-(A4), (A5) or (A5*), as* $n \to \infty$, $|\tilde{\mu}_R - \mu_R| = O_p(L_n)$.

Theorem 3.3 : *Under assumptions (A1)-(A4), (A5) or (A5*), as* $n \to \infty$, $\sup_x |\tilde{\rho}(x) - \rho(x)| = O_p(L_n)$.

The results in Theorems 3.1 and 3.2 refer to the estimators $\tilde{m}_l$ and $\tilde{\mu}_R$ that are obtained in the first and second estimation step respectively, but without shape constraints. The asymptotic behavior of our main interest $\tilde{\rho}$ is a combination of the results in the two steps and its characterization is given in Theorem 3.3.

Theorem 3.4 : *Under regularity conditions (A1)-(A5) (for monotone constrained estimator) or (A1)-(A4), (A5*) (for concave constrained estimator), for* $p \leq 3$ *and* $l = 1, ...d$, *as* $n \to \infty$, $\sup_x |\hat{m}_l(x) - m_l(x)| + |\hat{\mu}_R - \mu_R| + \sup_x |\hat{\rho}(x) - \rho(x)| = O_p(L_n)$.

Theorem 3.4 is developed for our proposed shape constrained frontier estimator. When the shape constraints are correctly specified, the results in Theorem 3.4 suggest that the proposed estimator with shape constraints enjoys the same asymptotic properties as the unconstrained one. Theorems 3.1-3.4 establish the uniform rates of convergence for the proposed estimators. On the other hand, local asymptotic results are also desirable for constructing confidence intervals and making inference for polynomial spline estimators. However, it is challenging to establish such results. For additive models, one difficulty is deriving the precise asymptotic expressions for the point-wise bias of polynomial spline estimator since the support of the additive spline basis functions is no longer locally supported. For more detailed discussion, please see section 7 of Huang (2003).

3.3. *Implementation*

In section 3.1, we estimate the mean efficiency by $\hat{\mu}_R = \left[\max_{1 \leq i \leq n} (Y_i/\hat{m}(X_i))\right]^{-1}$. However, this estimator is sensitive to the extreme values or outliers in the data. Therefore, in the implementation, we propose a modified estimator which is more robust to outliers. Let Q_1 and Q_3 be the first and third quantiles of $\{Y_i/\hat{m}(X_i)\}_{i=1}^n$ respectively. The corresponding inter-quantile 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 $\hat{\mu}_R = \left[\max_{i\in S}\left(Y_i/\hat{m}\left(X_i\right)\right)\right]^{-1}$, where $S = \{i : Y_i/\hat{m}\left(X_i\right) \in [Q_1 - 1.5IQR, Q_3 + 1.5IQR]\}$. In our simulation studies, we implement this robust modification and the 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 variable. The optimal N_n is selected using the Bayes Information Criterion (BIC). To be specific, let $\hat{Y}_i\left(N_n\right)$ 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, $\hat{N}_n = argminBIC(N_n) = argmin\left\{\log MSE + \left[(d(N_n + p) + 1)\log n\right]/n\right\}$ with $MSE = \sum_{i=1}^{n}\left\{Y_i - \hat{Y}_i\left(N_n\right)\right\}^2/n$.

4. Extensions

In model (2.1), the frontier function $\rho(\cdot)$ is assumed to be of additive form, which enables easy interpretation and alleviates "curse-of-dimensionality" in estimating multivariate nonparametric functions. However, this additive structure may not be practical in situations when the marginal output of an input depends on other input variables. In this section, we consider several extensions of the additive frontier model to allow for modeling interactions among input variables.

Extension 1: The most direct and simplest extension is to include products of input variables as pseudo inputs in the additive model. Let $X_1, \ldots, X_d$ be d input variables, and $Z_1 = X_1X_2, \ldots, Z_{d^*} = X_{d-1}X_d$ be the pseudo input variables that contain all second order product terms among d input variables with $d^* = d(d-1)/2$. Then one can consider an additive frontier model of the form

$$\rho(\mathbf{X}) = \rho_0 + \rho_1(X_1) + \cdots + \rho_d(X_d) + \rho_{d+1}(Z_1) + \cdots + \rho_{d+d^*}(Z_{d^*}), \tag{4.7}$$

where the pseudo input variables $Z_1, \ldots, Z_{d^*}$ are created to incorporate possible interactions among input variables. Higher order product terms can be included. But it can dramatically increase the number of additive terms and can also introduce collinearity problem.

Model (4.7) is an additive frontier model with a total of $d + d^*$ input variables. If the shape constraints such as monotoncity and/or concavity are imposed on each individual function $\rho_1, \ldots, \rho_{d+d^*}$, then our estimation procedure proposed in section 3 can be directly applied to estimate the unknown functions in above model.

Extension 2: Cobb–Douglas production function has been a popularly used parametric production function, which is of form $\rho(\mathbf{X}) = \beta_0 X_1^{\beta_1}\cdots X_p^{\beta_p}$, where $\beta_0, \beta_1, \ldots, \beta_p$ are unknown parameters. As an extension of the Cobb˘Douglas production function, we consider $\log(\rho(\mathbf{X})) = \rho_0 + \rho_1(X_1) + \cdots + \rho_p(X_p)$. Or equivalently, $Y = \rho(\mathbf{X})R = \exp\left(\rho_0 + \rho_1(X_1) + \cdots + \rho_p(X_p)\right)R$. It is an additive model in log-scale. Similar to Cobb-Douglas function, it is multiplicative in input variables. But the contribution of each input variable to the production function is nonparametric instead. In addition, one observes

that

$$\mu_{\log(Y)}(\mathbf{x}) = E(\log(Y)|\mathbf{X}=\mathbf{x}) = \rho_0 + \rho_1(x_1) + \cdots + \rho_p(x_p) + \mu_{\log(R)}$$
$$= \rho_0^* + \rho_1(x_1) + \cdots + \rho_p(x_p), \tag{4.8}$$

where $\rho_0^* = \rho_0 + \mu_{\log(R)}$. Therefore, the additive functions $\rho_1(x_1), \cdots, \rho_p(x_p)$ can be directly estimated using the proposed method, but with $\log(Y)$ as the response variable instead. In addition, one observes that

$$S_R = \sup \frac{Y}{\exp(\mu_{\log(Y)}(\mathbf{X}))} = \sup \frac{\exp\left(\rho_0 + \rho_1(X_1) + \cdots + \rho_p(X_p)\right) R}{\exp(\rho_0 + \rho_1(x_1) + \cdots + \rho_p(x_p)) \exp(\mu_{log(R)})}$$
$$= \frac{1}{\exp(\mu_{log(R)})}.$$

Therefore the frontier function can be represented as $\rho(\mathbf{x}) = \exp(\mu_{\log(Y)}(\mathbf{x}))S_R$.

The proposed two-step estimation method can be slightly modified to estimate this multiplicative model. In the first step, the constrained polynomial spline method can be used to estimate the additive components in (4.8). One advantage of this multiplicative model is that the monotonicity and concavity of each additive components also guarantees such shape constraints on the total output for each input variable. In the second step, the location of the frontier function can be estimated by $\hat{S}_R = \max_{i=1,\cdots,n} \frac{Y_i}{\exp(\hat{\mu}_{\log_{(Y_i)}}(\mathbf{X_i}))}$. Finally, one combines the results from previous two steps and estimate the frontier function as $\hat{\rho}(\mathbf{x}) = \exp(\hat{\mu}_{\log(Y)}(\mathbf{x}))\hat{S}_R$.

Extension 3: Functional ANOVA model in Huang (1998) and Gu (2013) is an useful extension of the additive model, which allows for interactions among input variables, and at the same time retains the flexibility of nonparametric modeling. Similar to the traditional ANOVA model, it uses functions of a single input to model main effects, and functions of multiple inputs for lower-order interactions. In particular, we consider a second-order functional ANOVA model for the frontier

$$\rho(\mathbf{X}) = \rho_0 + \sum_{l=1}^{d} \rho_l(X_l) + \sum_{1\le l<s\le d} \rho_{ls}(X_l, X_s), \tag{4.9}$$

in which $\{\rho_l(x_l)\}_{1\le l\le d}$ are the main effect terms, and $\{\rho_{ls}(x_l, x_s)\}_{1\le l<s\le d}$ are bivariate functions and flexible in modeling second order interactions among input variables. We assume the main effect terms follow shape constraints such as monotoncity and/or concavity as in the additive frontier model. For the interaction terms, we assume $\rho_{ls,1}(x_s) = \rho_{ls}(x_{l0}, x_s)$ (and $\rho_{ls,2}(x_l) = \rho_{ls}(x_l, x_{s0})$) follows monotoncity and/or concavity constraints as a function of x_s (and x_l) for any fixed x_{l0} (and x_{s0}).

To uniquely identify the functional components in (4.9), one often assumes the identification conditions (Gu 2013) that $A_l\rho_l(X_l) = 0$, and $A_l\rho_{ls}(X_l, X_s) = A_s\rho_{ls}(X_l, X_s) = 0$ for all $1 \le l < s \le d$. Here A_l is an averaging operator that averages out the variable X_l. For simplicity, we consider $A_l\rho_{ls} = \int \rho_{ls}(x_s, x_l)dx_l$. For any second-order functional ANOVA model that satisfies shape constraints, one can center the components in (4.9)

to meet the identification conditions as

$$\rho(\mathbf{X}) = \rho_0^* + \sum_{l=1}^{d} \rho_l^*(X_l) + \sum_{1\le l<s\le d} \rho_{ls}^*(X_l, X_s), \tag{4.10}$$

with $\rho_{ls}^*(X_l, X_s) = (I - A_l)(I - A_s)\rho_{ls}(X_l, X_s)$, $\rho_l^*(X_l) = (I - A_l)\rho_l(X_l) + (I - A_l)\left\{\sum_{l<s} A_s\rho_{ls}(X_l, X_s) + \sum_{s<l} A_s\rho_{ls}(X_s, X_l)\right\}$, and $\rho_0^* = \rho_0 + \sum_{l=1}^{d} A_l\rho_l(X_l) + \sum_{1\le l<s\le d} A_lA_s\rho_{ls}(X_l, X_s)$. Here I is the identity operator.

One notes that after centering, the functions ρ_l^* and ρ_{ls}^* do not necessarily follow the shape constraints, although the original functions ρ_l and ρ_{ls} do. The centering is only for uniquely defining the individual components in the frontier function. The frontier function $\rho(\mathbf{X})$ is unchanged after the centering. In the following, we first extend the proposed constrained polynomial spline method to obtain estimators that satisfy the shape constraints; then center these estimators to uniquely estimate the centered unknown components in (4.10).

The two-step estimation proposed for the additive frontier model can be extended to estimate model (4.9). In particular, the regression function

$$m(\mathbf{X}_i) = E(Y_i|\mathbf{X}_i) = m_0 + \sum_{l=1}^{d} m_l(X_l) + \sum_{1\le l<s\le d} m_{ls}(X_l, X_s), \tag{4.11}$$

with $m_l = \rho_l\mu_R$ for $l = 0, 1, \ldots, d$ and $m_{ls} = \rho_{ls}\mu_R$ for $1 \le l < s \le d$.

To approximate bivariate functions in model (4.11), we first introduce tensor product B-spline basis. Let $\tilde{\mathbf{B}}_l = \left(\tilde{B}_{l1}, \ldots, \tilde{B}_{lJ_n}\right)$ be the uncentered B-spline basis for input x_l defined in section 3.1. Then the tensor product basis constructed from $\tilde{\mathbf{B}}_l$ and $\tilde{\mathbf{B}}_s$ is defined as $\mathbf{G}_{ls} = \left\{\mathbf{G}_{jj'}(x_l, x_s), 1 \le j, j' \le J_n\right\}$ with $G_{jj'}(x_l, x_s) = \tilde{B}_{lj}(x_l)\tilde{B}_{sj'}(x_s)$. Therefore without any shape constraint, the traditional polynomial spline estimator of the regression function can be obtained by solving

$$\left(\tilde{m}_0, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\gamma}}\right) = \underset{m_0,\beta,\gamma}{\arg\min} \sum_{i=1}^{n} \left\{ Y_i - m_0 - \sum_{l=1}^{d} \beta_l^T\tilde{\mathbf{B}}_l(X_{il}) - \sum_{1\le l<s\le d} \gamma_{ls}^T\mathbf{G}_{ls}(X_{il}, X_{is}) \right\}^2 .$$

Then the unknown components in (4.11) can be estimated by $\tilde{m}_l(x_l) = \tilde{\boldsymbol{\beta}}_l^T\mathbf{B}_l(x_l)$, $\tilde{m}_{ls}(x_l, x_s) = \tilde{\boldsymbol{\gamma}}_{ls}^T\mathbf{G}_{ls}(x_l, x_s)$, which also serve as the initial estimator in the constrained estimation. In addition, let $\tilde{m}(\mathbf{x}) = \tilde{m}_0 + \sum_{l=1}^{d} \tilde{m}_l(x_l) + \sum_{1\le l<s\le d} \tilde{m}_{ls}(x_l, x_s)$ be the corresponding estimator of the regression function for any fixed $\mathbf{x} = (x_1, \ldots, x_d)^T$.

In the following, we extend the one-step backfitted procedure to obtain spline estimators with shape constraints. In particular, for each $1 \le l \le d$, to estimate the main effect m_l, we define the l-th pseudo responses $Y_{i,-l} = Y_i - \tilde{m}(\mathbf{X}_i) + \tilde{m}_l(X_{il})$. Then we consider the constrained least squares

$$\hat{\boldsymbol{\beta}}_l = \underset{\beta_l}{\arg\min} \sum_{i=1}^{n} \left\{ Y_{i,-l} - \beta_l^T\tilde{\mathbf{B}}_l(X_{il}) \right\}^2 \quad \text{subject to } \boldsymbol{\beta}_l \in C_l,$$

where C_l is the set of linear constraints for the l-th main effect term.

For the interaction terms, consider a bivariate spline function $g_{ls}(x_s, x_l) = \sum_{j=1}^{J_n}\sum_{j'=1}^{J_n}\gamma_{ls,jj'}\tilde{B}_{lj}(x_l)\tilde{B}_{sj'}(x_s)$. Then any fixed $x_l = x_{l0}$, one can write $g_{ls,1}(x_s) = g_{ls}(x_s, x_{l0}) = \sum_{j'=1}^{J_n}\gamma^*_{sj'}(x_{l0})\tilde{B}_{sj'}(x_s)$ with $\gamma^*_{sj'}(x_{l0}) = \sum_{j=1}^{J_n}\gamma_{ls,jj'}\tilde{B}_{lj}(x_{l0})$. Therefore, the linear constraints can be imposed on the coefficients $\boldsymbol{\gamma}^*_{ls,1}(x_{l0}) = \left\{\gamma^*_{sj'}, 1 \le j' \le J_n\right\}$ to ensure the corresponding shape constraints on $g_{ls,1}(x_s)$ at a fixed x_{l0}. Similarly, for any fixed $x_s = x_{s0}$, let $g_{ls,2}(x_l) = g_{ls}(x_{s0}, x_l) = \sum_{j=1}^{J_n}\gamma^*_{lj}(x_{s0})\tilde{B}_{sj}(x_s)$ with $\gamma^*_{lj}(x_{s0}) = \sum_{j'=1}^{J_n}\gamma_{ls,jj'}\tilde{B}_{sj'}(x_{s0})$. Then the linear constraints on the coefficients $\gamma^*_{ls,2}(x_{s0}) = \left\{\gamma^*_{lj}, j = 1, \ldots, J_n\right\}$ ensure the corresponding shape constraints on $g_{ls,2}(x_l)$.

Therefore, for each interaction term m_{ls}, let $Y_{i,-ls} = Y_i - \tilde{m}(\mathbf{X}_i) + \tilde{m}_{ls}(X_{il}, X_{is})$ be the pseudo responses, and consider

$$\hat{\boldsymbol{\gamma}}_{ls} = \arg\min_{\gamma_{ls}} \sum_{i=1}^{n}\left\{Y_{i,-ls} - \gamma_{ls}^T\mathbf{G}_{ls}(X_{il}, X_{is})\right\}^2, \tag{4.12}$$
$$\text{subject to } \gamma^*_{ls,1}(X_{il}) \in C_{ls}, \gamma^*_{ls,2}(X_{is}) \in C_{ls} \text{ for } i = 1, \ldots, n.$$

As a result, the unknown components in (4.11) can be estimated with shape constraints by

$$\hat{m}_l(x_l) = \hat{\boldsymbol{\beta}}_l^T\mathbf{B}_l(x_l), \quad \hat{m}_{ls}(x_l, x_s) = \hat{\boldsymbol{\gamma}}_{ls}^T\mathbf{G}_{ls}(x_l, x_s). \tag{4.13}$$

In (4.12), there are a total of $2n$ sets of constraints to ensure the spline estimate $\hat{m}_{ls}(x_l, x_s)$ satisfy the shape constraints as a function of x_l when $x_s = X_{is}$ for all $i = 1, \ldots, n$, and as a function of x_s when $x_l = X_{il}$ for all $i = 1, \ldots, n$.

To uniquely estimate the functional components that satisfy the identification conditions, we center the estimators in (4.13) as follows

$$\hat{m}_0^* = \hat{m}_0 + \sum_{l=1}^{d} A_l\hat{m}_l(x_l) + \sum_{1\le l<s\le d} A_lA_s\hat{m}_{ls}(x_l, x_s),$$
$$\hat{m}_l^*(x_l) = (I - A_l)\hat{m}_l(X_l) + (I - A_l)\left\{\sum_{l<s} A_s\hat{m}_{ls}(X_l, X_s) + \sum_{s<l} A_s\hat{m}_{ls}(X_s, X_l)\right\},$$
$$\hat{m}^*_{ls}(x_l, x_s) = (I - A_l)(I - A_s)\hat{m}_{ls}(x_l, x_s).$$

Here A_l is the integration operator with respect to variable x_l with $A_lm(x_s, x_l) = \int m(x_s, x_l)dx_l$ and I is the identity operator.

After the mean function is estimated, one can similarly estimate $\hat{\mu}_R$ using (3.6) with $\hat{m}(\mathbf{x}) = \hat{m}_0 + \sum_{l=1}^{d}\hat{m}_l(x_l) + \sum_{1\le l\le s\le d}\hat{m}_{ls}(x_l, x_s)$. Consequently, the frontier function can be estimated as $\hat{\rho}(\mathbf{x}) = \hat{m}(\mathbf{x})/\hat{\mu}_R$, and the centered components estimated as $\hat{\rho}^*_l(x_l) = \hat{m}^*_l(\mathbf{x})/\hat{\mu}_R$, and $\hat{\rho}^*_{ls}(x_{ls}) = \hat{m}^*_{ls}(x_l, x_s)/\hat{\mu}_R$ for $1 \le l < s \le d$.

In summary, we have discussed three extensions of the additive frontier model to incorporate interactions among input variables. Similar to the additive model, extensions 1 and 2 only involve estimation of univariate component functions, but the interactions are assumed to be of specific structures. Extension 3 is the most flexible in modeling interactions through nonparametric bivariate functions, but with an increased cost in

computation and often requires larger sample sizes due to "curse-of-dimensionality". In addition, in all three extensions, we impose shape constraints on the individual component functions. For extensions 1 and 3, these constraints are only sufficient, but not necessary for ensuring similar constraints on the total production function for each input variable.

5. Empirical Results

In this section, we conduct simulation studies to evaluate the numerical performance of our proposed method with finite samples. In the univariate case, we systematically compare our method to several existing methods in the literature including the nonparametric method using local linear regression in Martins-Filho and Yao (2007), the semiparametric method under shape constraints described in Wu and Sickles (2018), and the constrained polynomial spline estimation of the boundary curve proposed by Daouia et al. (2016). We also study the performance of our proposed method when there are multiple input variables with data being generated from an additive frontier model, and a frontier model with bivarate interaction terms. In addition to the simulation studies, the application of the proposed method to the Norwegian Farm data is also illustrated.

We use the averaged integrated squared errors (AISE) to evaluate the estimation accuracy of different methods. Let $\hat{\rho}_k$ be an estimator of ρ_k in the k-th replication, and $\{x_j\}_{j=1}^{ngrid}$ be a set of grid points. 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 $AISE(\hat{\rho}) = \frac{1}{r}\sum_{k=1}^{r} ISE(\hat{\rho}_k)$.

5.1. *Simulation Study*

5.1.1. Univariate Case

In this example, we adopt the simulation set-up as given in Martins-Filho and Yao (2007) to compare their local linear method with ours. The data is generated from a frontier model with a single input variable $Y = \rho\left(X\right)R$, where the frontier function $\rho\left(X\right) = 3\left(X - 1.5\right)^3 + 0.25X + 1.125$ and the input variable X is generated from the uniform distribution on $[1, 2]$. To generate the efficiency variable R with support on $(0, 1)$, we let $R = \exp\left(-Z\right)$ where Z is an exponential variable with scale parameter $\beta = 1/3, 1,$ and 3. These three choices of scale parameters correspond to three different shapes for the distribution of the production efficiency R: left-skewed, uniform, and right-skewed and enable us to examine the impact of different underlying distributions of R on the performance of frontier estimator. We consider sample sizes $n = 100$, 250, or 500 and $r = 1000$ replications are generated for each set-up.

For each simulated data set, we estimate the frontier function using seven different methods: the nonparametric method via local linear regression (LLR) in Martins-Filho and Yao (2007); our proposed two-step method using unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), monotone and concave constrained linear spline (MCCLS); the integral transformation of linear spline (ITLS) in Wu and Sickles (2018); and quadratic spline with monotonicity constraint (QS-M), or monotonicity and concavity constraints (QS-MC) proposed in Daouia et al. (2016).

In this example, the true frontier function is monotone increasing, but not concave. We consider both monotone constraint, and monotone and concave constraint to illustrate

the effect of incorporating different shape constraints on estimation accuracy. For the MCLS and MCCLS methods, shape constraints are imposed as in section 3.1. For the ITLS method, to estimate the regression function $m(\cdot)$, we consider model (4) in Wu and Sickles (2018), in which we set $g(a) = a^2$ to incorporate the monotonicity and concavity constraints. The model is then estimated by minimizing the sum of squared residuals. In the LLR method, the direct plug-in method is used to select the bandwidth in LLR method, as described by Ruppert (1995). The QS-M and QS-MC methods are implemented by using R package *npbr*.

The simulation results are summarized in Table B1. Overall, it clearly shows that for the proposed methods (ULS, MCLS, MCCLS), the AISEs decrease as the sample size n increases, supporting our asymptotic convergence results. In addition, the method with correct shape constraints (MCLS) always has better performance than the one without constraint (ULS). However, when incorrect shape constraint (MCCLS) is used, the results are mixed. In particular, MCCLS gives generally smaller AISEs for small sample size ($n = 100$), and can lead to larger AISEs than ULS due to the mis-specification of the shape constraint in particular when $\beta = 1/3$.

Among seven different methods, we observe the performances of QS-M and QS-MC are among the best, while all the other methods are generally comparable, except that LLR method has noticeably larger AISEs for $\beta = 1/3$. Different from other methods, QS-M and QS-MC estimate the frontier function directly by finding the shape constrained spline function that is closest to upper boundary of the data points. In contrast, all the other methods are implemented in a two-step procedure, and their estimation accuracies are affected by the estimation of both the mean regression function and the location parameter from the production efficiency variable.

Tables B1 also shows that the underlying distributions of the production efficiency variable R has an impact on the performance of different methods. The proposed spline methods have least favorable performance when $\beta = 3$ compared to the scenario of $\beta = 1$. This phenomena is mainly due to the difficulty in estimating μ_R under the right-skewed distribution of the efficiency variable R where the majority of observations are close to 0. However, the performance of the LLR method is the worst when $\beta = 1/3$. This inferior performance is due to the fact that under this scenario $\sigma_R^2 = 0.04$ which is half of that in the other two scenarios, leading to a larger variability in estimating the frontier function as indicated by Theorem 2 in Martins-Filho and Yao (2007). The simulation results show that the spline methods have great advantage over the LLR method when the efficiency R has a left-skewed distribution, i.e., when $\beta = 1/3$.

Furthermore, Table B1 compares the computation time to perform 1000 Monte-Carlo simulations in R when $\beta = 3$ for different methods. Our proposed methods have the best efficiency in terms of computation time. It is worth mentioning that although the ITLS method is comparable with our proposed MCCLS method regarding their performance and even performs slightly better when sample size is large enough, it comes with extremely low computation efficiency due to computation complexity involved in the double integrals. As a result, Table B1 shows that ITLS method requires much longer computing time than the other methods.

For our two-step method, a robust procedure is proposed in the estimation of μ_R to alleviate the effect of outliers. In order to evaluate its effectiveness, two artificial outliers $(X, Y) = (1.4, 3)$ and $(1.6, 3)$ are added manually to each simulated dataset. Then, all estimation procedures are applied again. To make the LLR and ITLS methods comparable with our robust method, we also apply similar robust modification in estimating σ_R^2 and μ_R, respectively. Table B2 clearly illustrates the effectiveness of the robust procedures.

When there are outliers, our proposed spline methods perform noticeably better than the LLR, QS-M, and QS-MC under all scenarios. For example, when $\beta = 1$ and $n = 50$, the AISEs of the MCLS method decrease by 97.8% of that of the QS-M method.

In Figures $B1$ and $B2$, we plot the estimated frontier functions from one simulated data using LLR (solid grey line), ULS (dashed brown line), MCLS (dotted purple line), MCCLS (long-dashed red line), ITLS (dot-dashed green line), QS-M (dotted orange lines), and QS-MC (two-dashed blue lines) methods, along with the true frontier (solid black line). Without outliers, Figure $B1$ shows that our proposed methods perform very well as the fitted curves are very close to the true curve and they are noticeably better than the LLR method when $\beta = 1/3$ and 1. When there are outliers, Figure $B2$ shows that the fitted curves from LLR, QS-M, and QS-MC are pulled up due to two extreme outliers. After the robust method is applied, the plots indicate that the spline methods work much better than the LLR method especially for smaller sample size. This suggests that our methods are more robust to the outliers.

5.1.2. Multivariate Case

In this example, we consider the 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 $\rho_0 = 8$, $\rho_1(X_1) = 2X_1 - 1$, $\rho_2(X_2) = 2X_2 + [\sin(2\pi X_2)]/2\pi - 1$, $\rho_3(X_3) = 3X_3^{1/3} - 9/4$, and $\rho_4(X_4) = (\log X_4 + 1)/2$. 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 the univariate example. The additive frontier model covers a combination of functions that are either monotone increasing or monotone increasing and concave. Similarly, consider sample size $n = 100$, 250, or 500 and $r = 1000$ replications are generated for each set-up. For each generated data, we estimate the additive frontier function using the proposed two-step method with or without shape constraints: 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 μ_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 the simulated data set.

The simulation results are summarized in Tables $B3$ (without outliers) and $B4$ (with outliers). For three spline methods, Table $B4$ shows that the estimation errors decrease as the sample size increases. 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. In addition, MCCLS 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 μ_R. Again, when using the contaminated data set, without the robust modification, the ULS method gives very large estimation errors with $\beta = 3$ because of large errors in the estimation of μ_R. When the robust method is applied, the ULS method is improved greatly but still worse than the two constrained methods. It shows that the robust method improve the estimation accuracy of μ_R greatly, especially under the case with right-skewed distribution of the efficiency variable R, i.e., when $\beta = 3$. For example, when applying the robust modification, the MSE of $1/\mu_R$ obtained using ULS decreases from 3896.62 to 0.07.

5.1.3. Second order functional ANOVA frontier model

In this example, we consider a functional ANOVA frontier model with a bivariate interaction term $Y = [\rho_0 + \rho_1(X_1) + \rho_2(X_2) + \rho_{12}(X_1, X_2)]R$, where $\rho_0 = 8$, $\rho_1(X_1) = 0.4\left(X_1^2 + 2X_1 - 0.75\right)$, $\rho_2(X_2) = 0.7\left(e^{X_2} - e + 1\right)$, and $\rho_{12}(X_1, X_2) = 4X_1X_2 - 2X_1 -$

$2X_2 + 1$. The input variables X_1 and X_2 are independently generated from the uniform distribution on $[0, 1]$. The efficiency variable R is generated in the same way as that in previous two examples. Again, different sample size $n = 100$, 250, and 500 are considered, and $r = 1000$ replications are generated for each set-up. To estimate the frontier functions, we consider our proposed two-step method using unconstrained linear spline (ULS), and monotone constrained linear spline (MCLS). Similarly, AISEs are used to evaluate the estimation accuracy.

Table B3 shows that the AISEs for both methods decrease with the increase of sample size. Furthermore, the spline method with correctly specified shape constraints (MCLS) performs much better than the one without constraint (ULS) under all scenarios. In addition, similar to the univariate and multivariate cases, the proposed method has better performance for smaller values of β. We also plot the estimated frontier functions using the ULS and MCLS methods, along with the true functions, as illustrated in Figure B3. Obviously, the MCLS method produces better fitted results than the ULS method, since the fitted curves or surface using the MCLS method are more close to the true functions. This visually confirm our numerical findings.

5.2. *Norwegian Farm Data*

We now apply the proposed two-step method to the Norwegian Farm data. This dataset is originally from Norwegian Farm Accountancy Survey collected by Norwegian Agricultural Economics Research Institute in 2004 to 2008. The data has been analyzed by Kumbhakar et al. (2014) and Wang and Xue (2015). We here consider the data on 145 grain farms of year 2007. 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), and their interaction terms. The output variable (Y) is the logarithms of farm revenue measured in Norwegian kroner. In Kumbhakar et al. (2014), they thoroughly studied 6 different parametric stochastic frontier models for the estimation of farm efficiency. Wang and Xue (2015) used a monotone additive regression model to quantify the relationship between the input variables and the output variable. We here consider the additive frontier model

$$Y = [\rho_0 + \rho_1(X_1) + \rho_2(X_2) + \rho_3(X_3) + \rho_4(X_4) + \rho_5(X_1X_2) + \rho_6(X_1X_3) + \rho_7(X_1X_4) + \rho_8(X_2X_3) + \rho_9(X_2X_4) + \rho_{10}(X_3X_4)] \times R$$

to quantify the maximum farm revenue given these four input variables and their interaction terms. Our semi-parametric model is more flexible than the parametric ones considered in Kumbhakar et al. (2014). In addition, Wang and Xue (2015) focused on the estimation of the regression or conditional mean function without consideration of interaction terms.

To estimate the unknown components in the model, we first consider the unconstrained linear spline (ULS) method. 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. Figure $B4$ plots the estimated 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 method (dashed blue line). The 95% point-wise confidence intervals from bootstrap (dotted blue line) using 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 1000 bootstrap samples as the lower and upper bounds respectively. Figure $B4$ shows that the input variables X_1, X_2 and X_3 have monotone increasing main effects on the farm revenue and the rate of change seems to decrease as the input variable increases. However, the interaction terms in general do not follow the monotone increasing constraint. Therefore, we impose monotone increasing or monotone and concave constraints on the main effect terms (ρ_1, ρ_2, ρ_3, ρ_4), while no shape constraint is imposed on the interaction terms (ρ_5, ρ_6, ρ_7, ρ_8, ρ_9), given that some interactions terms may have negative effect on farm revenue. We use the proposed method to estimate the frontier model with shape contraints on the main effect terms. In particular, in the one-step backfitted procedure, we consider monotone constrained linear spline (MCLS) with $C_l = C_M$ or monotone and concave constrained linear spline (MCCLS) $C_l = C_M \cap C_C$ for $l = 1, \ldots, 4$, and $C_l = \emptyset$ for $l = 5, \ldots, 9$ with $\emptyset$ being an empty set. Figure $B4$ plots the estimation results from both MCLS (dot-dashed green line) and MCCLS (long-dashed red line). It shows that three spline methods give similar fitted results.

Furthermore, to assess the production efficiency of each farm, in Figure $B5$ 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 three estimation method give similar results while using the MCCLS method for fitting the main effects produces slightly smaller efficiency estimates compared to the other two methods. Furthermore, it indicates that the majority of 145 farms have high efficiency with the estimates no less than 0.95. In addition, farms with lower revenue tend to have lower production efficiency.

In addition, to further explain the difference in efficiency among farms, we explore the relationships between efficiency and five other explanatory variables of interest including off-farm income share (net off-farm income 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 $B6$ shows the scatter plots of those variables versa the estimated efficiency and the boxplots of efficiency estimates for different education levels of farmers using MCCLS method in fitting the main effects and ULS method in fitting the interaction effects. For the first four continuous 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 farmers to work is reduced by the extra off-farm income. We also notice that as the off-farm income share increases, the efficiency slightly increases. However, it seems that there is no apparent relationship between the efficiency and farmer experience or the farmer's education level.

Acknowledgements

The authors thank Dr. Gudbrand Lien for sharing the Norwegian Farm data. Xue's research was supported by the Simons Foundation award 272556 and National Science Foundation award 1812258, Yang's by National Natural Science Foundation of China award 11371272 and Research Fund for the Doctoral Program of Higher Education of China award 20133201110002.

References

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

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, 78, 3-30.

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

Debreu, G. (1951), 'The coefficient of resource utilization', *Econometrica*, 19, 273-292.

Deprins, D., Simar, L., and Tulkens, H. (1978), '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.

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.

Gu, C. (2013), *Smoothing spline ANOVA models*, New York: Springer-Verlag.

Hall, P., and Park, B. U. (2002), 'New methods for bias correction at endpoints and boundaries', *Annals of Statistics*, 30, 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.

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 nonconvex 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), 'Analysis of production as an efficient combination of activities', in: Koopmans, T.C (Ed.), *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.

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

Meeusen, W., and Van den Broeck, J. (1977), 'Efficiency estimation from Cobb-Douglas production functions with composed error', *International Economic Review*, 18, 435-444.

Ramsay, J. O. (1988), 'Monotone regression splines in action', *Statistical science*, 3, 425-461.

Ruppert, D., Sheather, S. J., Wand, M. P. (1995) 'An effective bandwidth selector for local least squares regression'. *Journal of the American Statistical Association*, 90, 1257-1270.

Schumaker, L. L. (1981), *Spline functions*, New York: Wiley.

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

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

689-705.

Wang, L., and Xue, L. (2015), 'Constrained polynomial spline estimation of monotone additive models', *Journal of Statistical Planning and Inference*, 167, 27-40.

Wang, J, and Yang, L. (2009), 'Polynomial spline confidence bands for regression curves', *Statistica Sinica*, 19, 325-342.

Wu, X., and Sickles, R. (2018), 'Semiparametric estimations under shape constraints with applications to production functions', *Econometrics and Statistics*, 6, 74-89.

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

Appendix A. Assumptions

The asymptotic analysis in subsection 3.2 requires the following assumptions.

(A1) The input variables $\mathbf{X}_i$ are i.i.d. with support set $[0,1]^d$. Its joint 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 inputs and efficiencies $\{\mathbf{X}_i, R_i\}_{i=1}^n$ are i.i.d. with $\mathrm{E}(R_i|\mathbf{X_i}) = \mu_R$ and $\mathrm{var}(R_i|\mathbf{X}_i) = \sigma_R^2 < +\infty$. The efficiency R has a continuous distribution function $F_R(\cdot)$ such that $F_R(0) = 0, F_R(1) = 1$ and that the left derivative $f_R(1) = F_R'(1-)$ exists and is positive.

(A3) The knot sequence $\{0 = u_0 < u_1 < \cdots < u_{N_n} < u_{N_n+1} = 1\}$ is equally spaced on $[0,1]$ with $u_j = j/(N_n+1)$ for $j = 0, 1, \cdots, N_n+1$.

(A4) The number of interior knots N_n satisfies $N_n^{2p+2}/n \to 0$ and $N_n^{2p+5} log(n)/n \to +\infty$, as $n \to +\infty$.

(A5) For $1 \leq l \leq d$, the additive frontier function ρ_l is monotone increasing and $(p+1)$-times continuously differentiable for some integer $p \geq 1$. We assume that there exists a constant $c_3 > 0$, such that $\rho_l'(x) \geq c_3$, for $x \in [0,1]$.

(A5*) For $1 \leq l \leq d$, the additive frontier function ρ_l is concave and $(p+1)$-times continuously differentiable for some integer $p \geq 2$. Furthermore, we assume that there exists a constant $c_4 < 0$, such that $\rho_l^{(2)}(x) \leq c_4$, for $x \in [0,1]$.

Assumption (A1) is the same as condition 1 in Stone (1985). Assumption (A2) requires that the efficiency random variables are i.i.d. with a common distribution. Assumption (A3) assumes that the interior knots are equally spaced in the interval $[0,1]$, which is required to simplify the linear constraints derived in Lemmas 2 and 3 in the main paper. One can relax this condition to the one considered in Huang (1998), Xue and Yang (2006) and Wang and Xue (2015) which requires the knot sequence to be pseudo equally spaced. The assumption for the number of interior knots and samples size are give in the assumption (A4). This assumption is satisfied by N_n with the optimal order that $N_n = O(n^{1/(2p+3)})$. Additionally, for the monotone constrained estimator, Assumption (A5) assumes that each additive frontier function is monotone increasing and its first order derivative is bounded below from zero. Similarly, for the concave constrained estimator, Assumption (A5*) requires that each additive frontier function is concave.

Appendix B. Lemmas and Proofs

Proof of Lemma 3.1. Recall that $\tilde{\mathbf{B}}(x) = \left(\tilde{B}_1(x), \ldots, \tilde{B}_{N_n+p+1}(x)\right)^T$ is the non-centered B-spline basis of G^p. By de Boor (2001) p.115, $\tilde{B}_{j,k}'(x) =$

$(k-1)\left[-\tilde{B}_{j+1,k-1}(x)/(u_{j+1}-u_{j-k+2})+\tilde{B}_{j,k-1}(x)/(u_j-u_{j-k+1})\right]$, for $j=1,\ldots,N+p$, where $k=p+1$ is the order of the polynomial spline. Therefore, the first order derivative of the polynomial spline is

$$g'(x)=\left[\mathbf{B}^T(x)\boldsymbol{\beta}\right]'=\left[\sum_{j=1}^{N_n+p}\beta_j B_{j,k}(x)\right]'=\left[\sum_{j=1}^{N_n+p}\beta_j\left(\tilde{B}_{j,k}-\frac{1}{n}\sum_{i=1}^{n}\tilde{B}_{j,k}(x_i)\right)\right]'$$

$$=(k-1)\left[\left(\beta_1/(u_1-u_{-p+1})\right)\tilde{B}_{1,k-1}+\sum_{j=2}^{N_n+p}\left(\left(\beta_j-\beta_{j-1}\right)/(u_j-u_{j-p})\right)\tilde{B}_{j,k-1}\right]. \tag{B1}$$

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 $\beta_1\geq 0$ and $\beta_j-\beta_{j-1}\geq 0$ for $j=2,...,N_n+p$, and Lemma 3.1 follows. ■

Proof of Lemma 3.2. According to equation (B1), the first order derivative of the linear spline can be written as $g'(x)=\left[\left(\beta_1/(u_1-u_0)\right)\tilde{B}_{1,1}(x)+\sum_{j=2}^{N_n+1}\left(\left(\beta_j-\beta_{j-1}\right)/(u_j-u_{j-1})\right)\tilde{B}_{j,1}(x)\right]$. For a linear spline, the rate of change is a constant in each interval. To ensure that $g(\cdot)$ is a concave function, the rates of change need to be non-increasing in the whole region. Therefore, we'll have $\left(\beta_2-\beta_1\right)/(u_2-u_1)\leq\beta_1/(u_1-u_0)$ and $\left(\beta_j-\beta_{j-1}\right)/(u_j-u_{j-1})\leq\left(\beta_{j-1}-\beta_{j-2}\right)/(u_{j-1}-u_{j-2})$ for $j=3,...,N_n+p$ and $N_n\geq 2$. When $N_n=1$, one has $\left(\beta_2-\beta_1\right)/(u_2-u_1)\leq\beta_1/(u_1-u_0)$. Since the knots are equally spaced, Lemma 3.2 follows. ■

Proof of Lemma 3.3. By equation (B1), we have $g'(x)=(k-1)\left[\left(\beta_1/(u_1-u_{-p+1})\right)\tilde{B}_{1,k-1}+\sum_{j=2}^{N_n+p}\left(\left(\beta_j-\beta_{j-1}\right)/(u_j-u_{j-k+1})\right)\tilde{B}_{j,k-1}\right]$.

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

$$g^{(2)}(x)=(k-1)\left[\left(\beta_1/(u_1-u_{-p+1})\right)\tilde{B}'_{1,k-1}+\sum_{j=2}^{N_n+p}\left(\left(\beta_j-\beta_{j-1}\right)/(u_j-u_{j-k+1})\right)\tilde{B}'_{j,k-1}\right]$$

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

$$+(k-1)(k-2)\frac{-\beta_1}{(u_1-u_{-p+1})(u_1-u_{-p+2})}\tilde{B}_{2,k-2}$$

$$=I+II. \tag{B2}$$

Furthermore, term I in (B2) can be decomposed as

$$I=(k-1)(k-2)\sum_{j=2}^{p-1}\frac{\beta_j-\beta_{j-1}}{u_j-u_{j-k+1}}\left(\frac{-\tilde{B}_{j+1,k-2}}{u_j-u_{j-k+2}}+\frac{\tilde{B}_{j,k-2}}{u_{j-1}-u_{j-k+1}}\right)$$

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

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

$$=I_1+I_2+I_3.$$

Let $h = u_j - u_{j-1}$, for $j = 1, \ldots, N_n + 1$. Then, for terms I_1, I_2, I_3, one has

$$\begin{aligned}
I_1 &= (k-1)(k-2)\frac{\beta_2-\beta_1}{(u_2-u_{-p+2})(u_1-u_{-p+2})}\tilde{B}_{2,k-2} \\
&\quad + (k-1)(k-2)\sum_{j=3}^{p-1}\left[\frac{\beta_j-\beta_{j-1}}{(u_j-u_{j-k+1})(u_{j-1}-u_{j-k+1})} - \frac{\beta_{j-1}-\beta_{j-2}}{(u_{j-1}-u_{j-k})(u_{j-1}-u_{j-k+1})}\right]\tilde{B}_{j,k-2} \\
&\quad + (k-1)(k-2)\frac{-(\beta_{p-1}-\beta_{p-2})}{(u_{k-2}-u_{-1})(u_{k-2}-u_0)}\tilde{B}_{p,k-2} \\
&= (k-1)(k-2)\frac{\beta_2-\beta_1}{2h^2}\tilde{B}_{2,k-2} + (k-1)(k-2)\sum_{j=3}^{p-1}\left[\frac{\beta_j-\beta_{j-1}}{j(j-1)h^2} - \frac{\beta_{j-1}-\beta_{j-2}}{(j-1)^2h^2}\right]\tilde{B}_{j,k-2} \\
&\quad + (k-1)\frac{-(\beta_{p-1}-\beta_{p-2})}{(k-2)h^2}\tilde{B}_{p,k-2}, \qquad \text{(B3)}
\end{aligned}$$

$$I_2 = \frac{\beta_p-\beta_{p-1}}{h^2}\tilde{B}_{p,k-2} + \sum_{j=p+1}^{N_n+1}\frac{\beta_j-\beta_{j-1}-(\beta_{j-1}-\beta_{j-2})}{h^2}\tilde{B}_{j,k-2} - \frac{\beta_{N_n+1}-\beta_{N_n}}{h^2}\tilde{B}_{N_n+2,k-2}, \qquad \text{(B4)}$$

and

$$\begin{aligned}
I_3 &= (k-1)(k-2)\frac{(\beta_{N_n+2}-\beta_{N_n+1})}{(u_{N_n+2}-u_{N_n-k+3})(u_{N_n+1}-u_{N_n-k+3})}\tilde{B}_{N_n+2,k-2} + (k-1)(k-2)\times \\
&\quad \sum_{j=N_n+3}^{N_n+p}\left[\frac{\beta_j-\beta_{j-1}}{(u_j-u_{j-k+1})(u_{j-1}-u_{j-k+1})} - \frac{\beta_{j-1}-\beta_{j-2}}{(u_{j-1}-u_{j-k})(u_{j-1}-u_{j-k+1})}\right]\tilde{B}_{j,k-2} \\
&= (k-1)\frac{(\beta_{N_n+2}-\beta_{N_n+1})}{(k-2)h^2}\tilde{B}_{N_n+2,k-2} + (k-1)(k-2)\times \\
&\quad \sum_{j=N_n+3}^{N_n+p}\left[\frac{\beta_j-\beta_{j-1}}{(N_n+1-j+p)^2h^2} - \frac{\beta_{j-1}-\beta_{j-2}}{(N_n+2-j+p)(N_n+1-j+p)h^2}\right]\tilde{B}_{j,k-2}. \qquad \text{(B5)}
\end{aligned}$$

Also, II in (B2) can be written as

$$II = (k-1)(k-2)\frac{-\beta_1}{h^2}\tilde{B}_{2,k-2}. \qquad \text{(B6)}$$

By 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 (B4), (B4), (B5) and (B6), we have the following set of inequalities:

$\beta_2-\beta_1 \le 2\beta_1$,
$\beta_j-\beta_{j-1} \le j(\beta_{j-1}-\beta_{j-2})/(j-1)$, for $j = 3, ..., p-1$
$\beta_p-\beta_{p-1} \le p(\beta_{p-1}-\beta_{p-2})/(p-1)$,
$\beta_j-\beta_{j-1} \le \beta_{j-1}-\beta_{j-2}$, for $j = p+1, ..., N_n+1$
$\beta_{N_n+2}-\beta_{N_n+1} \le (p-1)(\beta_{N_n+1}-\beta_{N_n})/p$,
$\beta_j-\beta_{j-1} \le (N_n+1-j+p)(\beta_{j-1}-\beta_{j-2,})/(N_n+2-j+p)$, for $j = N_n+3, ..., N_n+p$,
for degree $p \ge 2$ and Lemma 3.3 follows. ■

In the following, we present the exact forms of constraints for quadratic and cubic splines to be concave. These results are used in the numerical studies.

Lemma B.1: *A sufficient condition for a quadratic spline* $g(x) = \mathbf{B}^T(x)\boldsymbol{\beta}$ *, 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:*

$\beta_2 - \beta_1 \leq 2\beta_1$,
$\beta_3 - \beta_2 \leq (\beta_2 - \beta_1)/2$, *for* $N_n = 1$, *and*
$\beta_2 - \beta_1 \leq 2\beta_1$,
$\beta_j - \beta_{j-1} \leq \beta_{j-1} - \beta_{j-2}$,
$\beta_{N+2} - \beta_{N+1} \leq (\beta_{N+1} - \beta_N)/2$, *for* $N \geq 2$, $j = 3, ..., N+1$.

Proof. It follows directly from Lemma 3.3. ■

Lemma B.2: *A sufficient condition for a cubic spline* $g(x) = \mathbf{B}^T(x)\boldsymbol{\beta}$ *, 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:*

$\beta_2 - \beta_1 \leq 2\beta_1$,
$\beta_3 - \beta_2 \leq \beta_2 - \beta_1$,
$\beta_4 - \beta_3 \leq (\beta_3 - \beta_2)/2$, *for* $N_n = 1$,
$\beta_2 - \beta_1 \leq 2\beta_1$,
$\beta_3 - \beta_2 \leq 3(\beta_2 - \beta_1)/2$,
$\beta_4 - \beta_3 \leq 2(\beta_3 - \beta_2)/3$,
$\beta_5 - \beta_4 \leq (\beta_4 - \beta_3)/2$, *for* $N_n = 2$ *and*
$\beta_2 - \beta_1 \leq 2\beta_1$,
$\beta_3 - \beta_2 \leq 3(\beta_2 - \beta_1)/2$,
$\beta_j - \beta_{j-1} \leq \beta_{j-1} - \beta_{j-2}$,
$\beta_{N+2} - \beta_{N+1} \leq 2(\beta_{N+1} - \beta_N)/3$,
$\beta_{N+3} - \beta_{N+2} \leq (\beta_{N+2} - \beta_{N+1})/2$, *for* $N_n \geq 3$, $j = 4, ..., N+1$.

Proof. It follows directly from Lemma 3.3. ■

Lemma B.3: *For any* m *that is centered with* $E(m(X_l)) = 0$ *and satisfies condition (A5*) , there exists a concave function* $g \in G^{(p)}$ *that is empirically centered with* $\sum_{i=1}^n g(X_{il})/n = 0$*, such that* $\|m - g\|_\infty \leq c\left\|m^{(p+1)}\right\|_\infty N_n^{-p-1}$*, and* $\left\|m^{(2)} - g^{(2)}\right\|_\infty \leq c\left\|m^{(p+1)}\right\|_\infty N_n^{-p+1}$ *for some constant* $c > 0$ *and large enough* n*.*

Proof. According to Theorem 1.51 in Schumaker (1981), for any $m \in C^{p+1}[0,1]$, there exists a function $g \in G^{(p)}$, such that $\|m - g\|_\infty \leq c\left\|m^{(p+1)}\right\|_\infty N_n^{-p-1}$ and $\left\|m^{(2)} - g^{(2)}\right\|_\infty \leq c\left\|m^{(p+1)}\right\|_\infty N_n^{-p+1}$ for some constant $c > 0$ and $p > 1$. 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$g^{(2)} \leq \left\|m^{(2)} - g^{(2)}\right\|_\infty + m^{(2)} \leq c_1/2 < 0$. In addition, if m is centered with $E(m(X_l)) = 0$, then

$$
\begin{aligned}
|\overline{g}_n| = \left|\sum_{i=1}^n g(x_{il})/n\right| &\leq \left|\sum_{i=1}^n [g(x_{il}) - m(x_{il})]/n\right| + \left|\sum_{i=1}^n m(x_{il})/n\right| \\
&\leq c\left\|m^{(p+1)}\right\|_\infty N_n^{-p-1} + O_p\left(n^{-1/2}\right) = O_p\left(N_n^{-p-1}\right).
\end{aligned}
$$

Let $g^* = g - \overline{g}_n$. Then there exists a constant $c > 0$, such that

$$
\begin{aligned}
\|m - g^*\|_\infty \leq \|m - g\|_\infty + |\overline{g}_n| &\leq c\left\|m^{(p+1)}\right\|_\infty N_n^{-p-1}, \\
\left\|m^{(2)} - g^{*(2)}\right\|_\infty = \left\|m^{(2)} - g^{(2)}\right\|_\infty &\leq c\left\|m^{(p+1)}\right\|_\infty N_n^{-p+1}.
\end{aligned}
$$

Therefore Lemma B.3 follows for sample size large enough. ■

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 $Y_{i,-l}^* = Y_i - \sum_{l' \neq l} m_{l'}(X_{il'}) = m_l(X_{il}) + \rho(\mathbf{X}_i)\varepsilon_i$ as the pseudo response, and for $l = 1, ..., d$ let $\mathbf{Y}_{-l}^* = \left(Y_{1,-l}^*, ..., Y_{n,-l}^*\right)^T$. Then $m_l^*(x) = \mathbf{B}_l^T(x)\boldsymbol{\beta}_l^*$ where $\boldsymbol{\beta}_l^* = \left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\mathbf{Y}_{-l}^*$. Note that $\{m_l^*(x)\}_{l=1}^d$ are unavailable for real data analysis and are constructed only for proving the theoretical results.

Lemma B.4: *Under regularity conditions (A1)-(A4),(A5*), for $l = 1, \ldots, d$ and $L_n = \sqrt{N_n/n} + N_n^{-p-1}$, one has* $\sup_{x\in[0,1]}\left|m_l^{*(2)}(x) - m_l^{(2)}(x)\right| = O_p\left(L_n N_n^2\right)$.

Proof. By Lemma B.3, for each $l = 1, \ldots, d$, there exists a concave function g_l, such that $\|m_l - g_l\|_\infty \leq c\left\|m_l^{(p+1)}\right\|_\infty N_n^{-p-1}$, and $\left\|m_l^{(2)} - g_l^{(2)}\right\|_\infty \leq c\left\|m_l^{(p+1)}\right\|_\infty N_n^{-p+1}$. By definition, one has

$$\begin{aligned}
& m_l^{*(2)}(x) \\
&= \mathbf{B}_l^{(2)}(x)^T\boldsymbol{\beta}_l^* = \mathbf{B}_l^{(2)}(x)^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\mathbf{Y}_{-l}^* = \tilde{\mathbf{B}}_l^{(2)}(x)^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T(\mathbf{m}_l + \rho\varepsilon) \\
&= \mathbf{B}_l^{(2)}(x)^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T(\mathbf{g}_l + \mathbf{m}_l - \mathbf{g}_l + \rho\varepsilon) \\
&= \mathbf{B}_l^{(2)}(x)^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\mathbf{g}_l + \mathbf{B}_l^{(2)}(x)^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T(\mathbf{m}_l - \mathbf{g}_l) \\
&\quad + \mathbf{B}_l^{(2)}(x)^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\rho\varepsilon \\
&= I(x) + II(x) + III(x). \qquad \text{(B7)}
\end{aligned}$$

Since we can write $\mathbf{g}_l = \mathbf{B}_{nl}\boldsymbol{\gamma}_l$ for some coefficient $\boldsymbol{\gamma}_l$, $I(x) = \mathbf{B}_l^{(2)}(x)^T\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\mathbf{B}_{nl}\boldsymbol{\gamma}_l = \mathbf{B}_l^{(2)}(x)^T\boldsymbol{\gamma}_l = g_l^{(2)}(x)$. Then by Lemma B.3, we have

$$\sup_{x\in[0,1]}\left|I(x) - m_l^{(2)}(x)\right| = \sup_{x\in[0,1]}\left|g_l^{(2)}(x) - m_l^{(2)}(x)\right| \leq c\left\|m_l^{(p+1)}\right\|_\infty N_n^{-p+1}. \qquad \text{(B8)}$$

Note that $\sup_x\left|\mathbf{B}_l^{(2)}(x)\right| = O\left(N_n^2\right)$ and that the number of nonzero elements in $\mathbf{B}_l^{(2)}(x)$ is bounded

$$\begin{aligned}
\sup_{x\in[0,1]}|II(x)| &\leq CN_n^2\left|\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T(\mathbf{m}_l - \mathbf{g}_l)\right| \\
&= CN_n^2\left|\left(n^{-1}\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1} \times n^{-1}\mathbf{B}_{nl}^T(\mathbf{m}_l - \mathbf{g}_l)\right|.
\end{aligned}$$

According to equation (A.10) in Lemma A.3 of Wang and Yang (2009), $\left\|\left(n^{-1}\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\right\|_\infty = O_{a.s}(N_n)$, hence

$$\begin{aligned}
\sup_{x\in[0,1]}|II(x)| &\leq CN_n^2 \times O_{a.s}(N_n) \times \left|n^{-1}\mathbf{B}_{nl}^T(\mathbf{m}_l - \mathbf{g}_l)\right| = O_{a.s}\left(N_n^3\right) \times \left|n^{-1}\mathbf{B}_{nl}^T(\mathbf{m}_l - \mathbf{g}_l)\right| \\
&\leq O_{a.s}\left(N_n^3\right) \times \left|\langle\mathbf{B}_l, 1\rangle_n\right| \times CN_n^{-p-1} \leq O_{a.s}\left(N_n^{2-p}\right) \times \left|\langle\mathbf{B}_l, 1\rangle_n\right|.
\end{aligned}$$

Applying the inequality $\sup_{\varphi\in G^{(p)}}\left|\langle\varphi,1\rangle_n-\langle\varphi,1\rangle\right|\|\varphi\|^{-1}=O_p\left(n^{-1/2}N_n^{1/2}\log^{1/2}n\right)$, one obtains that

$$\begin{aligned}\left|\langle\mathbf{B}_l,1\rangle_n\right| &\leq |\langle\mathbf{B}_l,1\rangle|+\|\mathbf{B}_l\|\times O_p\left(n^{-1/2}N_n^{1/2}\log^{1/2}n\right)\\ &= O_p\left(N_n^{-1}+N_n^{-1/2}\times n^{-1/2}N_n^{1/2}\log^{1/2}n\right)=O_p\left(N_n^{-1}\right).\end{aligned}$$

Therefore,

$$\sup_{x\in[0,1]}|II(x)|\leq O_{a.s}\left(N_n^{2-p}\right)\times O_p\left(N_n^{-1}\right)=O_p\left(N_n^{-p+1}\right). \tag{B9}$$

Since $\sup_{x\in[0,1]}\left|\mathbf{B}_l^{(2)}(x)\right|=O_p\left(N_n^2\right)$, and the element of $n^{-1}\mathbf{B}_{nl}^T\rho\varepsilon$ is $n^{-1}\sum_{i=1}^n B_{lj}(x_i)\rho(x_i)\varepsilon_i=O_p\left(n^{-1/2}N_n^{-1/2}\right)$, therefore

$$\sup_{x\in[0,1]}|III(x)|=O_p\left(N_n^2\times N_n\times n^{-1/2}N_n^{-1/2}\right)=O_p\left(n^{-1/2}N_n^{5/2}\right). \tag{B10}$$

Finally, Lemma B.4 follows from equations (B7), (B8), (B9), and (B10). ■

Lemma B.5: *For any function m that satisfies condition (A5*) with $p=2,3$ and large enough n, there exists a concave function $g\in G^{(p)}$ whose coefficients satisfy the concave constraints such that $\|m-g\|_\infty\leq c\left\|m^{(p+1)}\right\|_\infty N_n^{-p-1}$, for some constant $c>0$.*

Proof. By Lemma B.3, for any $m\in C^{p+1}[0,1]$ that satisfies condition (A5*) and large sample size, there exists a concave function $g\in G^{(p)}$, such that $\|m-g\|_\infty\leq c_1\left\|m^{(p+1)}\right\|_\infty N_n^{-p-1}$ for some constant $c_1>0$. Since $m\in C^{p+1}[0,1]$, its first order derivative $m'\in C^p[0,1]$. According to Lemma 3 from Wang and Xue (2015), for $p\leq 3$ and large enough n, there exists a $\hat{g}'\in G^{(p-1)}$ whose coefficients satisfy the linear constraints, such that $\|g'-\hat{g}'\|_\infty\leq c_2\left\|m^{(p+1)}\right\|_\infty N_n^{-p}$ for some constant $c_2>0$. With $\hat{g}(x)=\int\hat{g}'(x)\,dx\in G^{(p)}$ and follow the similar arguments in the proof of Lemma 3 in Wang and Xue (2015), we can easily show that $\|g-\hat{g}\|_\infty\leq c_2\left\|m^{(p+1)}\right\|_\infty N_n^{-p-1}$. Furthermore, in the proof of Lemma 3 in Wang and Xue (2015), we showed that the spline coefficients of $\hat{g}^{(2)}$ are non-positive which suggests that the coefficients of $\hat{g}$ satisfy the concave constraints. Therefore, the Lemma B.5 follows. ■

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

Lemma B.6: *Under regularity conditions (A1)-(A4),(A5*), for $l=1,\ldots,d$ and $p=2,3$, there exists a spline function $g_l=\sum_{j=1}^{J_n}\gamma_{lj}\tilde{B}_j$ 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 $p=2,3$, by Lemma B.5, there exists a spline function g_l whose coefficients satisfy the

concave constraints such that $\|m_l - g_l\|_\infty = O_p\left(N_n^{-p-1}\right)$. By definition, we have

$$\begin{aligned}\boldsymbol{\beta}_l^* &= \left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\left(\mathbf{m}_l + \rho\varepsilon\right) = \left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\left(\mathbf{g}_l + \mathbf{m}_l - \mathbf{g}_l + \rho\varepsilon\right)\\ &= \left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\mathbf{g}_l + \left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\left(\mathbf{m}_l - \mathbf{g}_l\right) + \left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\rho\varepsilon\\ &= I + II + III. \end{aligned} \tag{B11}$$

Write $\mathbf{g}_l = \mathbf{B}_{nl}\boldsymbol{\gamma}_l$, then

$$I = \left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\mathbf{B}_{nl}\boldsymbol{\gamma}_l = \boldsymbol{\gamma}_l. \tag{B12}$$

Since $|\left(n^{-1}\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}|_\infty = O_{a.s.}\left(N_n\right)$ and $\|m_l - g_l\|_\infty = O_p\left(N_n^{-p-1}\right)$, similar arguments as these in (B9) gives that

$$|II|_\infty = O_p\left(N_n^{-p-1}\right). \tag{B13}$$

Then, follow the similar argument in the proof of Lemma 4 in Wang and Xue (2015),

$$|III|_\infty = O_p\left(\sqrt{\frac{N_n^3 \log n}{n}}\right). \tag{B14}$$

From equations (B11), (B12), (B13) and (B14) and assumption (A4), we have $\sup|\boldsymbol{\beta}_l^* - \boldsymbol{\gamma}_l|_* = \sup|II + III|_* = O_p\left(\sqrt{\frac{N_n^3 \log n}{n}}\right)$ and then the Lemma B.6 follows. ■

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

Lemma B.7: *Under regularity conditions (A1)-(A5*), for $l = 1, \ldots, d$ and $p \leq 3$, there exists a spline function $g_l = \sum_{j=1}^{J_n}\gamma_{lj}B_j$ whose coefficients satisfy the concave constraints such that*

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

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

Proof. We have

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

By Lemma B.6, for any fixed $l = 1, ...d$, the coefficient $\boldsymbol{\beta}_l^*$ satisfies the concave constraints with probability approaching to 1 as $n \to \infty$. Furthermore, similar arguments as these in (B9) gives that $|I|_\infty = O_p\left(L_n\right)$. Hence the Lemma B.7 follows. ■

Lemma B.8: *Under assumption (A2), as $n \to \infty$, $1 - \max_{1 \leq i \leq n} R_i = O_p\left(n^{-1}\right)$.*

Proof. Denote $M_n = \max_{1\leq i\leq n} R_i$, then $\forall r > 0, P[M_n \leq r] = F_R(r)^n$. Take $r_n = 1 - a_n f_R^{-1}(1)/n$ for some $a_n > 0, n > a_n, a_n \to a > 0$. Then $r_n - 1 = O\left(n^{-1}\right)$ and for large enough n, $F_R(r_n) - F_R(1) = f_R(1)(r_n - 1) + o\left(n^{-1}\right)$. Then as $n \to \infty$, $F_R(r_n) = 1 + f_R(1)\left\{-a_n f_R^{-1}(1)/n\right\} + o\left(n^{-1}\right) = 1 - a_n/n + o\left(n^{-1}\right)$, and $P[M_n \leq r_n] = \left\{1 - a_n/n + o\left(n^{-1}\right)\right\}^n \to e^{-a}$.

Then as $n \to \infty$, $P[n(1 - M_n) \leq n(1 - r_n)] = 1 - P[M_n \leq r_n] \to 1 - e^{-a}$. So $P\left[n(1 - M_n) \leq a_n f_R^{-1}(1)\right] \to 1 - e^{-a}$. Hence $P[n(1 - M_n) \leq x] \to 1 - e^{-x f_R(1)}, \forall x > 0$. Therefore $n(1 - M_n)$ converges in distribution, and $1 - M_n = O_p\left(n^{-1}\right)$. ■

Proof of Theorem 3.1

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

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

Then we have

$$\begin{aligned}
&\sup_x|\tilde{m}_l(x) - m_l(x)| \leq \sup_x\left|\mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T(\mathbf{m}_l + \boldsymbol{\rho}\boldsymbol{\varepsilon}) - m_l(x)\right| \\
&+\sup_x\left|\mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\sum_{l'\neq l}(\mathbf{m}_{l'} - \tilde{\mathbf{m}}_{l'})\right| \\
&= I + II. \qquad \text{(B15)}
\end{aligned}$$

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), $\sup_x\left|\mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\boldsymbol{\rho}\boldsymbol{\varepsilon}\right| = O_p\left(\sqrt{N_n/n}\right)$ and $\sup_x|\bar{m}_l(x) - m_l(x)| = O_p\left(N_n^{-p-1}\right)$. Therefore,

$$\begin{aligned}
I &\leq \sup_x\left|\mathbf{B}_l^T(x)\left(\mathbf{B}_{nl}^T\mathbf{B}_{nl}\right)^{-1}\mathbf{B}_{nl}^T\boldsymbol{\rho}\boldsymbol{\varepsilon}\right| + \sup_x|\bar{m}_l(x) - m_l(x)| \\
&= O_p\left(\sqrt{N_n/n} + N_n^{-p-1}\right). \qquad \text{(B16)}
\end{aligned}$$

For II, applying similar techniques for proving (B9) and (B10), one obtains that

$$II = O_p\left(\sqrt{N_n/n} + N_n^{-p-1}\right). \qquad \text{(B17)}$$

Theorem 1 follows from equations (B15), (B16) and (B17). ■

Proof of Theorem 3.2

By the definition of $\tilde{\mu}_R$ and Lemma B.8, 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 1. 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 $\left|\tilde{\mu}_R^{-1} - \mu_R^{-1}\right| = O_p(L_n)$, where $L_n =$

$\sqrt{N_n/n} + N_n^{-p-1}$. 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 $\left|\max_i \left[m\left(X_i\right) R_i/\tilde{m}\left(X_i\right)\right] - 1\right| = O_p\left(L_n\right)$. By Theorem 1, one has $\sup_x \left|\tilde{m}\left(x\right) - m\left(x\right)\right| = O_p\left(L_n\right)$. Therefore, for any $\varepsilon > 0$, there exists $\delta > 0$, such that for all $n > N_\varepsilon$,

$$P\left(L_n^{-1} \sup_x \left|m\left(x\right)/\tilde{m}\left(x\right) - 1\right| < \delta\right) > 1 - \varepsilon. \tag{B18}$$

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 (B18), $P\left(L_n^{-1}\left|\max_i \left[m\left(X_i\right) R_i/\tilde{m}\left(X_i\right)\right] - 1\right| < \delta\right) > 1 - \varepsilon$. 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 Lemma 4 that $1 - \max_i R_i = O_p\left(1/n\right) = o_p\left(L_n\right)$ and inequality (B18), for any $\varepsilon > 0$, there exists δ_1 and $\delta_2 > 0$, such that for all $n > N_\varepsilon$, $P\left(\inf_x \left[m\left(x\right)/\tilde{m}\left(x\right)\right] > 1 - L_n\delta_1\right) > 1 - \varepsilon$ 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 $n > N_\varepsilon$, $P\left(\max_i \left[m\left(X_i\right) R_i/\tilde{m}\left(X_i\right)\right] > 1 - L_n\delta_3\right) > 1 - \varepsilon$. ■

Proof of Theorem 3.3

By definition, we have

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

By Theorems 1 and 2, $\sup_x \left\{\mu_R^{-1}\left[\tilde{m}\left(x\right) - m\left(x\right)\right]\right\} = O_p\left(L_n\right)$ and $\sup_x \left[\tilde{m}\left(x\right)\left(\tilde{\mu}_R^{-1} - \mu_R^{-1}\right)\right] = O_p\left(L_n\right)$. One then concludes that $\sup_x \left[\tilde{\rho}\left(x\right) - \rho\left(x\right)\right] = O_p\left(L_n\right)$. ■

Proof of Theorem 3.4

When each m_l is monotone increasing, Theorem 2 in Wang and Xue (2015) showed 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 B.7 indicates that the coefficients of the $\tilde{m}_l$ satisfy the concave constraints with probability approaching one as $n \to \infty$. These imply that for $p \leq 3$ the unconstrained estimator $\tilde{m}_l$ and the shape constrained estimator $\hat{m}_l$ are identical with probability approaching one as $n \to \infty$. Therefore, $\hat{m}_l$ enjoys the same asymptotic properties as $\tilde{m}_l$. Then by Theorem 1, we have $\sup_x \left|\hat{m}_l\left(x\right) - m_l\left(x\right)\right| = O_p\left(L_n\right)$. Moreover, by Lemma 4 that $1 - \max_i R_i = O_p\left(1/n\right)$, the results follow directly from Theorems 2 and 3. ■

Table B1. Averaged integrated squared errors (AISE) for the estimation of frontier function using seven different methods: local linear regression (LLR), unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), monotone and concave constrained linear spline (MCCLS), integral transformation of linear spline (ITLS), quadratic spline with monotonicity constraint (QS-M), and quadratic spline with monotonicity and concavity constraints (QS-MC) under three scenarios **without outliers**, and corresponding computation time in R for scenario where $\beta = 3$ based on 1000 Monte-Carlo simulations.

Method	n	$\beta = \frac{1}{3}$	$\beta = 1$	$\beta = 3$	Computation time for scenario $\beta = 3$
LLR	100	0.3432	0.1046	0.0911	1.15 mins
	250	0.1505	0.0437	0.0331	9.55 mins
	500	0.0819	0.0188	0.0168	1.93 hrs
ULS	100	0.0188	0.0552	0.1105	1.46 secs
	250	0.0128	0.0323	0.0481	1.67 secs
	500	0.0090	0.0231	0.0260	33.83 secs
MCLS	100	0.0156	0.0347	0.0786	23.07 secs
	250	0.0117	0.0255	0.0366	23.44 secs
	500	0.0080	0.0199	0.0225	1.65 mins
MCCLS	100	0.0152	0.0308	0.0711	1.24 mins
	250	0.0142	0.0229	0.0317	1.53 mins
	500	0.0141	0.0191	0.0198	4.95 mins
ITLS	100	0.0135	0.0359	0.1269	1116.19 hrs
	250	0.0121	0.0199	0.0324	1309.83 hrs
	500	0.0122	0.0168	0.0182	1729.63 hrs
QS-M	100	0.0014	0.0072	0.0420	7.43 mins
	250	0.0004	0.0019	0.0102	34.06 mins
	500	0.0001	0.0007	0.0042	2.48 hrs
QS-MC	100	0.0045	0.0124	0.0263	5.59 mins
	250	0.0057	0.0043	0.0084	17.35 mins
	500	0.0072	0.0050	0.0046	55.07 mins

Table B2. Averaged integrated squared errors (AISE) for the estimation of frontier functions using seven different methods: local linear regression (LLR), unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), monotone and concave constrained linear spline (MCCLS), integral transformation of linear spline (ITLS), quadratic spline with monotonicity constraint (QS-M), and quadratic spline with monotonicity and concavity constraints (QS-MC), under three scenarios **with outliers** for sample size $n = 50$ and 250.

		$n = 50$	$n = 250$
Method	β	AISE	AISE
LLR	1/3	3.6121	0.5105
	1	1.2315	0.2116
	3	0.6078	0.0962
ULS	1/3	0.0614	0.0174
	1	0.2425	0.0401
	3	0.3060	0.0505
MCLS	1/3	0.0350	0.0166
	1	0.1116	0.0298
	3	0.1274	0.0338
MCCLS	1/3	0.0350	0.0187
	1	0.1120	0.0291
	3	0.1274	0.0325
ITLS	1/3	0.0458	0.0141
	1	0.1296	0.0236
	3	0.2375	0.0333
QS-M	1/3	2.3263	1.2948
	1	5.1084	1.3044
	3	15.6574	1.3389
QS-MC	1/3	5.3029	1.6103
	1	4.9842	1.6017
	3	4.1513	1.5909

Table B3. Averaged integrated squared errors (AISE) for the estimation of $\{\rho_l(\cdot)\}_{l=1}^4$ and $m(\mathbf{X})$, and mean squared errors (MSE) for the estimation of $1/\mu_R$ using three different spline methods: unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), and monotone and concave constrained linear spline (MCCLS), under three scenarios **without outliers**.

	Method	n	$\rho_1(X_1)$	$\rho_2(X_2)$	$\rho_3(X_3)$	$\rho_4(X_4)$	$m(\mathbf{X})$	$1/\mu_R$
$\beta=\frac{1}{3}$	ULS	100	0.3455	0.3507	0.3610	0.3888	0.6745	0.0322
		250	0.1100	0.1121	0.1188	0.1272	0.2159	0.0303
		500	0.0522	0.0515	0.0605	0.0701	0.1220	0.0209
	MCLS	100	0.1157	0.1148	0.1227	0.1208	0.2654	0.0197
		250	0.0625	0.0607	0.0632	0.0648	0.1377	0.0184
		500	0.0354	0.0353	0.0422	0.0446	0.0964	0.0135
	MCCLS	100	0.0833	0.0850	0.0929	0.0951	0.2191	0.0170
		250	0.0439	0.0444	0.0493	0.0547	0.1245	0.0148
		500	0.0228	0.0263	0.0354	0.0399	0.0930	0.0108
$\beta=1$	ULS	100	1.8027	1.8407	1.8196	1.8220	1.4581	0.1530
		250	0.5798	0.6050	0.6049	0.5930	0.3933	0.2209
		500	0.2398	0.2419	0.2523	0.2707	0.1773	0.1619
	MCLS	100	0.4466	0.4314	0.4074	0.3944	0.4230	0.1252
		250	0.2127	0.2108	0.2140	0.1962	0.1839	0.1070
		500	0.1152	0.1199	0.1235	0.1268	0.1082	0.0820
	MCCLS	100	0.3465	0.3361	0.3236	0.3197	0.3373	0.1259
		250	0.1713	0.1617	0.1765	0.1650	0.1593	0.0987
		500	0.0866	0.0934	0.0981	0.1036	0.0939	0.0696
$\beta=3$	ULS	100	3.8643	3.6581	3.6481	3.7902	1.3299	0.4146
		250	1.2790	1.2714	1.2420	1.2874	0.3517	0.0620
		500	0.5809	0.5251	0.5071	0.5742	0.1570	0.0250
	MCLS	100	0.8374	0.6910	0.7359	0.7785	0.3868	0.1686
		250	0.3736	0.3366	0.3823	0.3633	0.1445	0.0469
		500	0.2464	0.2192	0.2131	0.2189	0.0819	0.0234
	MCCLS	100	0.6503	0.5326	0.5540	0.5868	0.3069	0.1753
		250	0.3009	0.2768	0.3040	0.2871	0.1169	0.0465
		500	0.1998	0.1782	0.1669	0.1839	0.0679	0.0234

Table B4. Averaged integrated squared errors (AISE) for the estimation of $\{\rho_l(\cdot)\}_{l=1}^4$ and $m(\mathbf{X})$, and mean squared errors (MSE) for the estimation of $1/\mu_R$ using three different spline methods: unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), and monotone and concave constrained linear spline (MCCLS), under scenarios **with outliers** and sample size $n = 250$.

Robust (With/ Without)	Method	β	$\rho_1(X_1)$	$\rho_2(X_2)$	$\rho_3(X_3)$	$\rho_4(X_4)$	$m(\mathbf{X})$	$1/\mu_R$
Without	ULS	1/3	0.2636	0.2444	0.2950	0.2859	0.2664	0.4318
		1	1.6994	1.8022	1.8713	1.8441	0.4639	3.5821
		3	746.4053	422.9166	1272.1126	922.8275	0.4574	3896.6163
	MCLS	1/3	0.2168	0.1855	0.2284	0.2032	0.1689	0.4780
		1	0.5150	0.4835	0.5287	0.4606	0.2014	1.2736
		3	1.6137	1.4409	1.6388	1.5131	0.1636	7.3188
	MCCLS	1/3	0.1982	0.1680	0.2129	0.1897	0.1709	0.4743
		1	0.4262	0.3852	0.4440	0.3968	0.2032	1.0842
		3	1.8907	1.3553	2.3452	1.9947	0.1600	9.8067
With	ULS	1/3	0.1278	0.1279	0.1476	0.1530	0.2664	0.0728
		1	0.6613	0.6640	0.7003	0.6881	0.4639	0.4431
		3	1.2378	1.2569	1.2498	1.2803	0.4574	0.0713
	MCLS	1/3	0.0872	0.0814	0.0915	0.0896	0.1689	0.0755
		1	0.3063	0.2936	0.3173	0.2862	0.2014	0.4722
		3	0.3627	0.3346	0.3656	0.3510	0.1636	0.0641
	MCCLS	1/3	0.0846	0.0782	0.0917	0.0904	0.1709	0.0950
		1	0.2793	0.2595	0.2970	0.2693	0.2032	0.4921
		3	0.3415	0.3105	0.3386	0.3197	0.1600	0.0609

Table B5. Averaged integrated squared errors (AISE) for the estimation of functions ρ_1, ρ_2 and ρ_{12} using two different spline methods: unconstrained linear spline (ULS) and monotone constrained linear spline (MCLS) under three scenarios.

	Method	n	$\rho_1(X_1)$	$\rho_2(X_2)$	$\rho_{12}(X_1, X_2)$
$\beta = \frac{1}{3}$	ULS	100	0.1227	0.1238	0.2494
		250	0.0467	0.0450	0.0881
		500	0.0223	0.0227	0.0468
	MCLS	100	0.0745	0.0750	0.0437
		250	0.0337	0.0343	0.0261
		500	0.0205	0.0216	0.0192
$\beta = 1$	ULS	100	0.6748	0.6762	1.3895
		250	0.2632	0.2501	0.4899
		500	0.1228	0.1248	0.2576
	MCLS	100	0.3877	0.3689	0.1140
		250	0.1531	0.1514	0.0584
		500	0.0898	0.0882	0.0430
$\beta = 3$	ULS	100	1.5858	1.5691	3.2641
		250	0.6939	0.6777	1.3145
		500	0.3443	0.3481	0.7017
	MCLS	100	0.8986	0.8586	0.2366
		250	0.3944	0.3844	0.1140
		500	0.2402	0.2351	0.0802

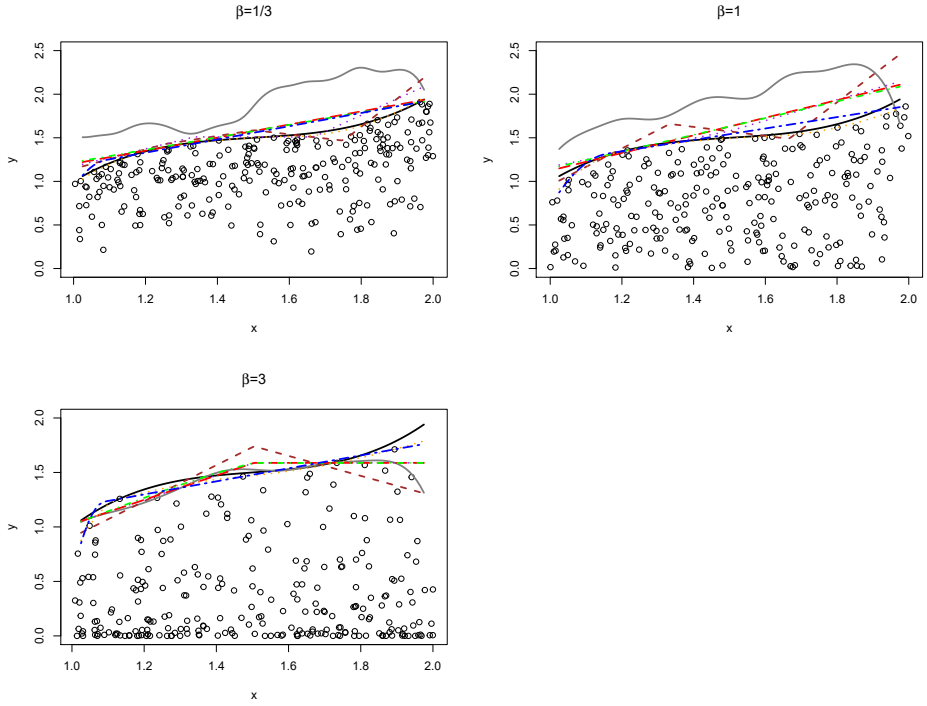

Figure B1. Simulation results of the frontier estimation for sample size $n = 250$ **without outliers** when the parameter $\beta = 1/3$, 1, and 3. In each plot, solid black line represents the true curve, and solid grey, dashed brown, dotted purple, long-dashed red, dot-dashed green, dotted orange, and two-dashed blue lines represent the fitted curves obtained using local linear regression (LLR), unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), monotone and concave constrained linear spline (MCCLS), integral transformation of linear spline (ITLS), quadratic spline with monotonicity constraint (QS-M), and quadratic spline with monotonicity and concavity constraints (QS-MC), respectively.

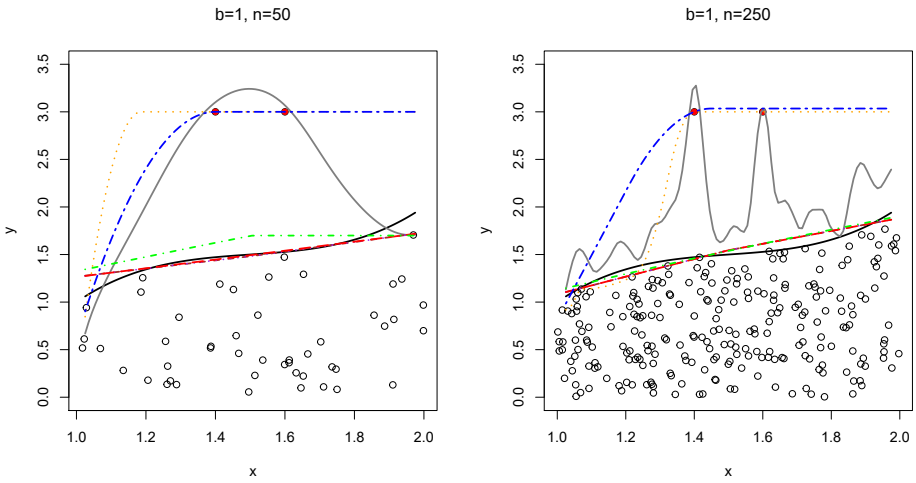

Figure B2. Simulation results of the frontier estimation for sample size $n = 50$ and 250 **with outliers** under the scenario where $\beta = 1$. In each plot, solid black line represents the true curve, and solid grey, dashed brown, dotted purple, long-dashed red, dot-dashed green, dotted orange, and two-dashed blue lines represent the fitted curves obtained using local linear regression (LLR), unconstrained linear spline (ULS), monotone constrained linear spline (MCLS), monotone and concave constrained linear spline (MCCLS), integral transformation of linear spline (ITLS), quadratic spline with monotonicity constraint (QS-M), and quadratic spline with monotonicity and concavity constraints (QS-MC), respectively. Solid red circle indicates the locations where twoartificial outliers are added.

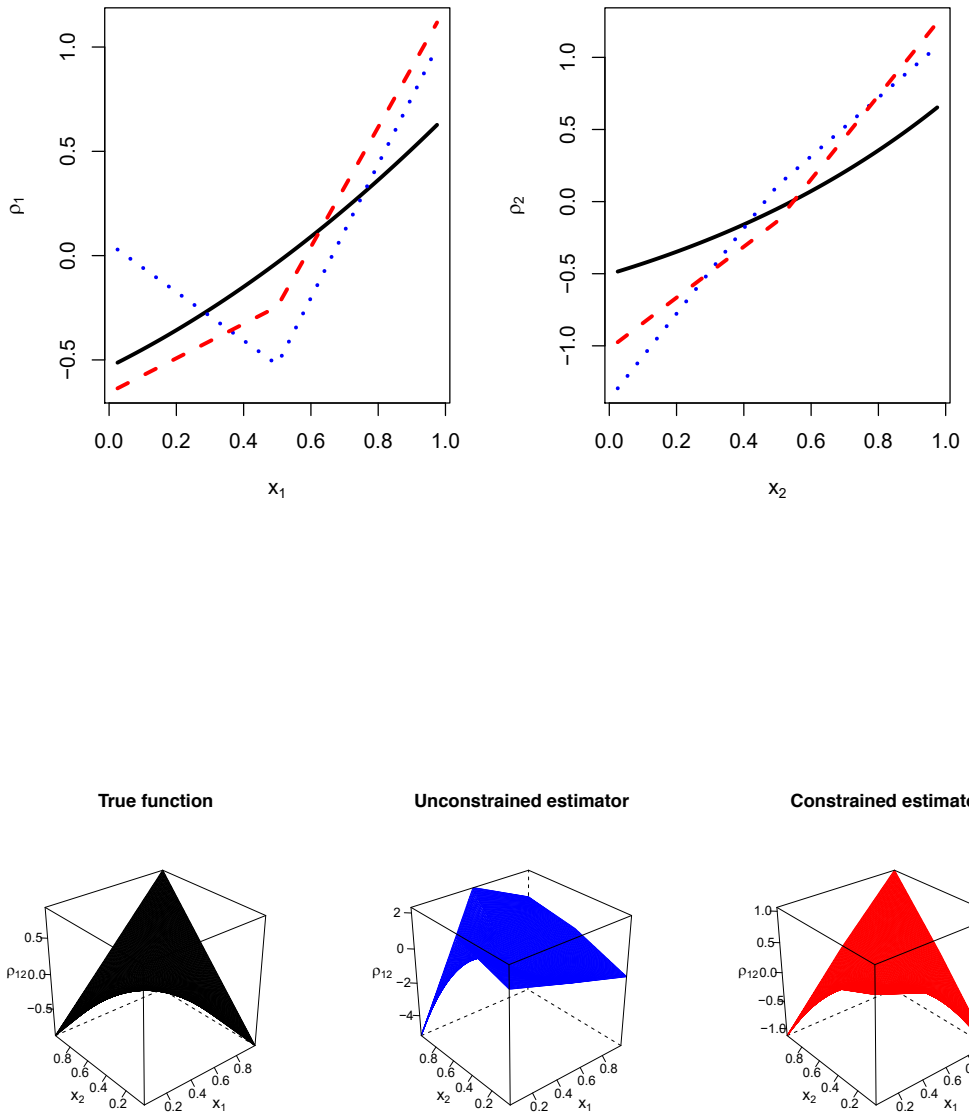

Figure B3. Simulation results of the frontier estimation for sample size $n = 250$ under the scenario where $\beta = 1$. In the first row, solid black, dotted blue, and dashed red lines represent the true function, fitted curves obtained using unconstrained linear spline (ULS), and monotone constrained linear spline (MCLS), respectively. In the second row, black, blue and red surfaces represent the true function, fitted results obtainted using ULS and MCLS, respectivley.

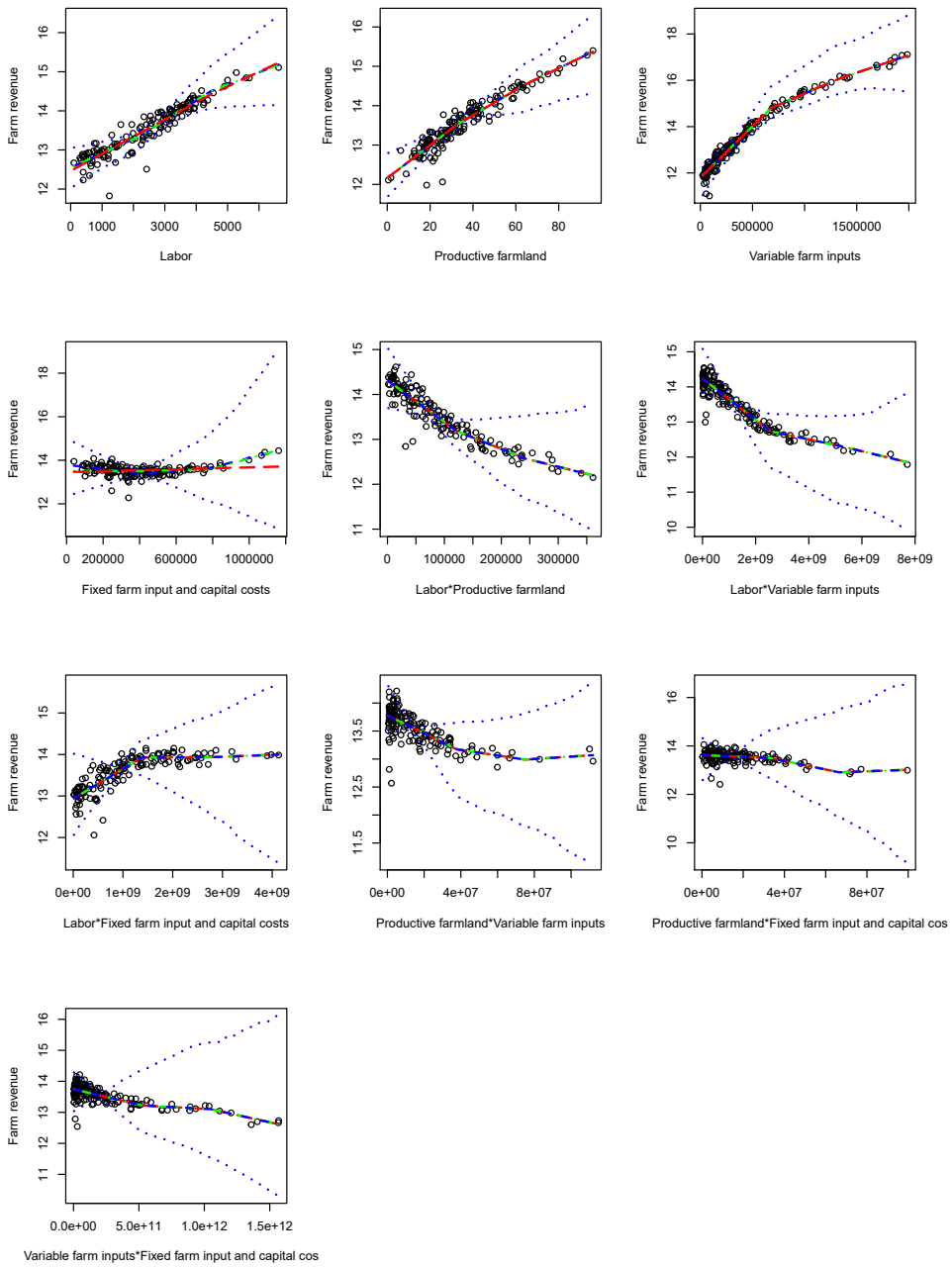

Figure B4. Norwegian Farm data: Pseudo responses for each input variable are denoted by black circle. The dashed blue line denotes estimated mean function using unconstrained linear spline (ULS) method. The dot-dashed green and long-dashed red lines denote estimated mean function using monotone constrained linear spline (MCLS) and monotone and concave constrained linear spline (MCCLS) in fitting main effects and using ULS in fitting interaction effects, respectively. The dotted blue lines represent the 95% point-wise confidence intervals from 1000 bootstrap samples using the ULS method.

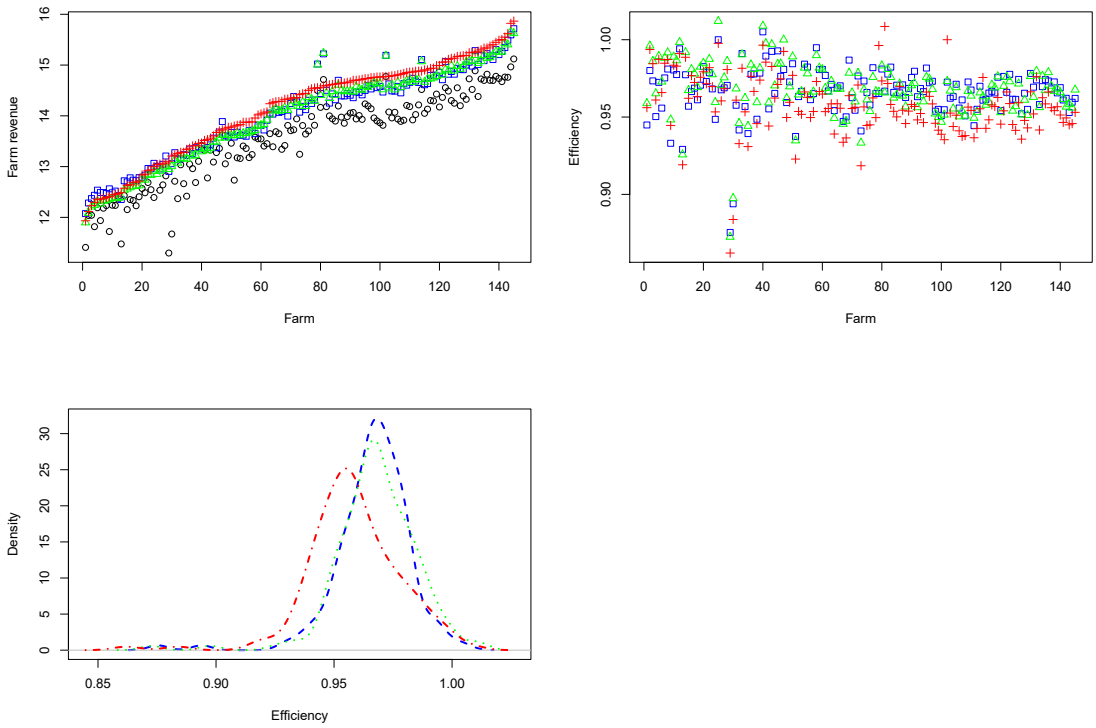

Figure B5. Norwegian Farm data: Estimated maximum farm revenue (left top), efficiency estimates (right top), and the kernel density distribution of the efficiency estimates (left bottom). In the top figures, the blue rectangle, green triangle and red plus represent estimated maximum revenue or efficiency of all 145 farms using unconstrained linear spline (ULS), using monotone constrained linear spline (MCLS) and monotone and concave constrained linear spline (MCCLS) in fitting main effects and ULS in fitting interaction effects, respectively. The true farm revenue is denoted by 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.

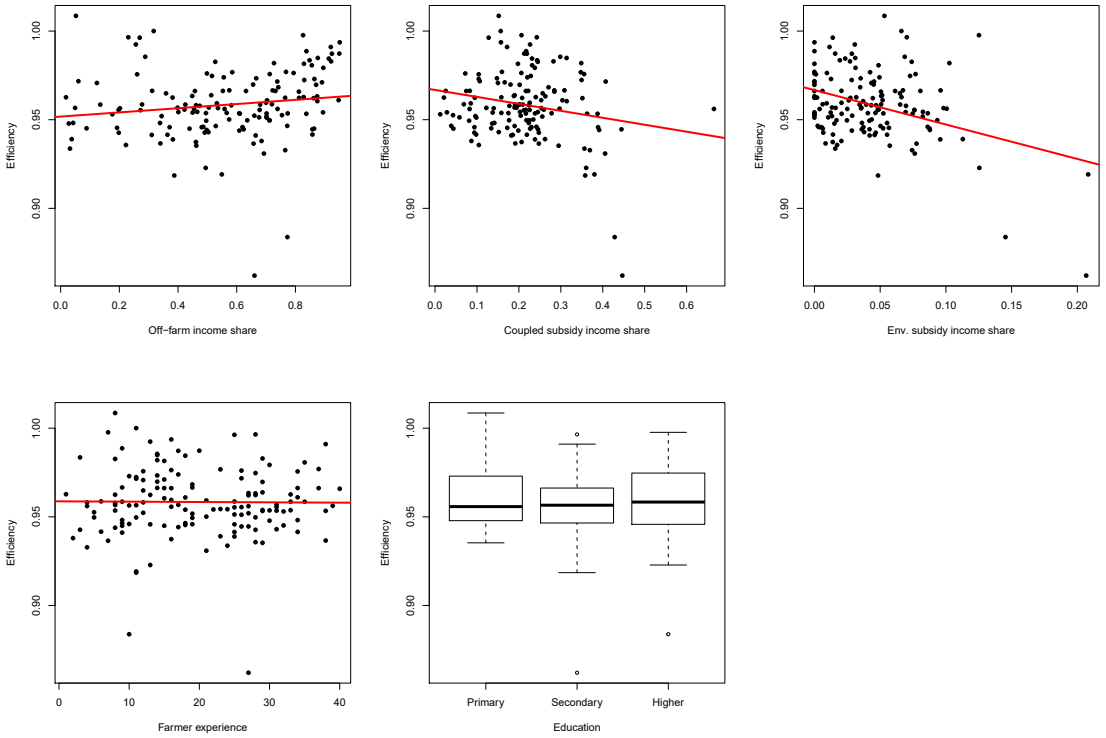

Figure B6. Norwegian Farm data using MCCLS method in fitting main effects and ULS method in fitting interaction effects: Off-farm income share, coupled subsidy income share, environmental subsidy income share, and farmer experience versa the efficiency estimates for the 145 farms, in which solid red lines represent linear least squares fit. Boxplots of the estimated efficiency in the subgroups with different education levels.