이학석사 학위논문

Monotone Function Estimation in Varying Coefficient Model

변수계수모형에서의 단조함수추정

2013년 2월

서울대학교 대학원

통계학과

하 우 석

Monotone Function Estimation
in Varying Coefficient Model

지도교수 박 병 욱

이 논문을 이학석사 학위논문으로 제출함

2012년 10월

서울대학교 대학원

통계학과

하 우 석

하우석의 이학석사 학위논문을 인준함

2012년 12월

위 원 장	김	우	철	(인)
부위원장	박	병	욱	(인)
위 원	김	용	대	(인)

Monotone Function Estimation in Varying Coefficient Model

by

Wooseok Ha

A Dissertation
submitted in fulfillment of the requirement
for the degree of
Master of Science
in
Statistics

The Department of Statistics
College of Natural Sciences
Seoul National University
February, 2013

Abstract

Function estimation under shape constraints is gaining great popularity among statisticians and particularly, monotone function estimation has been studied extensively. This paper introduces function estimation methods in the varying coefficient model when the coefficient functions are monotone. To estimate the coefficient functions, we have to consider the covariate effect, which enter into the weights in the varying coefficient model. We first review the estimation method and algorithm in the traditional regression model and see how this method can be extended to the additive model. Then, we present formulation and algorithm of monotone function estimation in the varying coefficient model by adapting the minimization problem to the traditional case. We also attempt to extend this method to more generalized varying coefficient model where only some part of the coefficient functions enters into the model. Finally, we carry out numerical studies with simulated data.

Keywords: Varying coefficient model; Isotone regression; Monotone function estimation; Backfitting method; Pool Adjacent Violators algorithm

Student Number: 2011-20253

Contents

Chapter 1

Introduction

The varying coefficient model is a generalization of the linear regression model, allowing coefficients to vary depending on different variables. After Hastie and Tibshirani (1993) first introduced the model, it has gained great popularity due to both the flexibility of the model and simplicity of interpretability. Thus, the varying coefficient model is very useful in many real applications and a lot of research has been studied for this model.

In general, coefficient functions in the varying coefficient model is assumed to be smooth, so kernel smoothing is the most popular method to estimate coefficient functions in the varying coefficient model. Fan and Zhang (1999) considered the varying coefficient model

$$Y = \sum_{j=1}^{d} f_j(X) Z_j + \epsilon \tag{1.1}$$

and suggested a two-step estimation method when the coefficient functions have different degrees of smoothness. They also showed that the estimators achieve the optimal rate of convergence. Yang, Park, Xue and Härdle (2006) studied the varying coefficient model where the coefficient functions depend on different variables,

$$Y = \sum_{j=1}^{d} f_j(X_j)Z_j + \epsilon, \tag{1.2}$$

based on the marginal integration method. Lee, Mammen and Park (2012) newly introduced the smooth backfitting method to the model (4.2) and proved the oracle properties of the estimators.

Nonparametric function estimation under shape constraints, such as monotone function estimation or convex/concave function estimation, is very different from local smoothing because we have to take global structure into account. After Brunk (1958) first proposed to minimize the following equation

$$\sum_{i=1}^{n}(y^i - f(x^i))^2 w(x^i) \tag{1.3}$$

over a class of monotone functions, estimation of a monotone function in regression analysis has been studied extensively for a long time. Barlow, Bartholomew, Bremner and Brunk (1972) contains overall explanations about this topic. Brunk (1970) showed that at a fixed point x_0, $n^{1/3}(\hat{f}(x_0) - f(x_0))$ has an asymptotic distribution when $f'(x_0) > 0$. Wright (1981) specified the asymptotic distribution and extended this result to the general case. See also Prakasa (1969), Mukerjee (1988) and Mammen (1991). Mammen and Yu (2007) extended isotone regression model to the nonparametric additive model

$$Y = c + \sum_{j=1}^{d} f_j(X) + \epsilon, \tag{1.4}$$

where additive components f_j, $1 \le j \le d$, are monotone. Based on iterative application of the Pool Adjacent Violators algorithm, they derived a backfitting estimator and showed that the estimator achieves the oracle property, i.e.

asymptotically each component can be estimated as if other components were known. The convergence of the backfitting algorithm was also discussed.

In this paper, we are interested in the varying coefficient model (4.2) in which the coefficient functions f_j's are monotone. Thus, we do not employ localization method such as kernel smoothing. Instead, we use iterative isotonization as in Mammen and Yu (2007) to gain a backfitting coefficient estimators. Unlike the additive model covariates are involved in the varying coefficient model. Thus, covariates enter into weight terms, which usually arises in the varying coefficient model. In section 2, we briefly review and summarize the estimation method of monotone functions. In section 3, we suggest estimation algorithm to estimate monotone coefficient functions in the varying coefficient model. Also, we observe how this method can be combined with the classical backfitting method. In section 4, numerical studies for the proposed algorithms are presented.

Chapter 2

Isotone Regression

In this chapter, we review the estimation method of a regression function in the isotone regression model. We also see how this method can be extended to the nonparametric additive model.

2.1 Optimization under the monotone constraints

Let $I = \{f = (f^1, f^2, \cdots, f^n)^T \in R^n : f^1 \le f^2 \le \cdots \le f^n\}$ be the set of isotonic vectors and $w = (w^1, w^2, \cdots, w^n)^T \in R^n$ be the given weight vector. Given the vector $y = (y^1, y^2, \cdots, y^n)^T \in R^n$, our aim is to minimize in the class of isotonic vectors I the sum

$$\sum_{i=1}^{n} (y^i - f^i)^2 w^i, \tag{2.1}$$

i.e. to find the closest isotonic vector to the given $y \in R^n$.

To see the minimizer of (4.1), we first define $A_j = \sum_{i=1}^{j} w^i$ and $B_j = \sum_{i=1}^{j} w^i y^i$ for $j = 1, 2, \cdots, n$ with $A_0 = B_0 = 0$. Then, the *cumulative sum* *diagram* (CSD) is the plot of $P_j = (A_j, B_j)$ for $j = 0, 1, 2, \cdots, n$. If $y \in R^n$ is a random vector, the CSD can be regarded as a weighted partial sum process. Note that the slope of the chord joining the points $P_i = (A_i, B_i)$ and $P_j = (A_j, B_j)$ is the weighted average

$$\sum_{k=i}^{j} w^k y^k / \sum_{k=i}^{j} w^k. \tag{2.2}$$

If $i = j - 1$, then this is just y^j. The *greatest convex minorant* (GCM) of the CSD is the graph of the supremum of all convex functions whose graphs lie below the CSD. Let $P_j^* = (A_j, B_j^*)$ for $j = 1, 2, \cdots, n$ be the corner points of the GCM of the CSD. Then, we can easily check the following two properties of the GCM.

First, $P_n = P_n^*$. In other words, $B_n = B_n^*$.

Second, if $B_{j-1}^* < B_{j-1}$, then the slope of the GCM at A_{j-1} is equal to the one at A_j for $j = 1, 2, \cdots, n$.

Now, we can derive the solution of the problem (4.1) from the two properties of the GCM. Let y^{*i} be the slope of the GCM between P_{i-1}^* and P_i^*. We will show that $y^* = (y^{*1}, y^{*2}, \cdots, y^{*n})^T \in R^n$ minimizes (4.1), i.e.

$$\sum_{i=1}^{n} (y^i - f^i)^2 w^i \geq \sum_{i=1}^{n} (y^i - y^{*i})^2 w^i + \sum_{i=1}^{n} (y^{*i} - f^i)^2 w^i \tag{2.3}$$

for any $f = (f^1, f^2, \cdots, f^n)^T \in I$. For this, it is suffice to show that

$$\sum_{i=1}^{n} (y^i - y^{*i})(y^{*i} - f^i) w^i \geq 0 \tag{2.4}$$

for any $f = (f^1, f^2, \cdots, f^n)^T \in I$. From the identity $\sum_{i=1}^{n} a^i (b^i - b^{i-1}) = a^n b^n + \sum_{i=1}^{n-1} b^{i-1}(a^{i-1} - a^i)$, we see since $w^i y^i = B^i - B^{i-1}$, for $a^i = y^{*i} - f^i$ and $b^i = B^i - B^{*i}$,

$\sum_{i=1}^{n}(y^i - y^{*i})(y^{*i} - f^i)w^i =$
$\sum_{i=1}^{n-1}[(f^i - f^{i-1}) - (y^{*i} - y^{*i-1})](B^{i-1} - B^{*i-1}) + (y^{*n} - f^n)(B^n - B^{*n}).$

By the properties of the GCM, $(y^{*n}-f^n)(B^n-B^{*n}) = 0$ and $(y^{*i}-y^{*i-1})(B^{i-1}-B^{*i-1})$. Since $f^i \geq f^{i-1}$ and $B^{i-1} > B^{*i-1}$, $i = 1, 2, \cdots, n-1$, we obtain (3.4). The uniqueness of the solution follows easily from (3.3). Thus, the isotonic vector $y^* \in R^n$ minimizes (4.1).

The solution $y^* = (y^{*1}, y^{*2}, \cdots, y^{*n})^T \in R^n$ is expressed by the explicit form

$$y^{*i} = \max_{s \leq i} \min_{t \geq i} \frac{\sum_{j=s}^{t} w^j y^j}{\sum_{j=s}^{t} w^j}, \tag{2.5}$$

which is called max-min formula. Intuitively, (3.5) is to express the way of finding the slope of the GCM. From (3.5), we understand the solution $y^* \in R^n$ as the weighted average of the vector $y \in R^n$ where the average interval is determined by the given $y \in R^n$ and $w \in R^n$. Another way of interpret the solution is that it is the projection of the vector $y \in R^n$ onto the set of isotonic vectors I with respect to the scalar product $\langle f, g \rangle = \sum_{i=1}^{n} f^i g^i w^i$ where f and $g \in R^n$. The projection exists uniquely because the set I is closed and convex.

In order to calculate the solution, we utilize the algorithm, called the *Pool Adjacent Violators algorithm* (PAVA). The underlying idea for this algorithm is that if $y^{i-1} > y^i$ for some i, then $y^{*i-1} = y^{*i}$. Thus, if for some i, then we "pool" these two values y^{i-1} and y^i with associated weights w^{i-1} and w^i.; Replace y^{i-1} and y^i by

$$(w^{i-1}y^{i-1} + w^i y^i)/(w^{i-1} + w^i), \tag{2.6}$$

assigning the associated weight $w^{i-1}+w^i$. With this replaced vector, we iterate the above process until we reach the required order. In short, the algorithm is to pool a pair of adjacent violating values to get a single value. For the description of the algorithm, we define the term *block* for a set of consecutive numbers of $\{1, 2, \cdots, n\}$. Then, finding the solution $y^* = (y^{*1}, y^{*2}, \cdots, y^{*n})^T \in R^n$ is just equal to finding the *solution blocks*, on which y^*'s have the particular value with the monotone order. This value is determined by the weighted average of the given vector $y \in R^n$ within each block. For three consecutive blocks Q_p, Q, Q_n in order, we say the block Q is up-satisfied if the weighted average of $y \in R^n$ on the block Q is less than the one on the block Q_n. Similarly, the block Q is said to be down-satisfied if the weighted average of $y \in R^n$ on the block Q is greater than the one on the block Q_p. Now, we describe the Pool Adjacent Violators algorithm.

Begin with a partition of $\{1, 2, \cdots, n\}$ into blocks which are its individual elements. The algorithm proceeds from the block $\{1\}$ until the *solution blocks* are obtained. Given the blocks $Q_1, \cdots, Q_{m-1}$ in the monotone order,

1. Activate the block $\{r\}$ where $r - 1$ is the greatest number in Q_{m-1}.

2-1. If the active block is up-satisfied, activate the next higher block and go to the step 1.

2-2. If the active block is not up-satisfied, pool the active block with the next higher block and go to the step 3.

3-1. If the active block is down-satisfied, go to the step 2.

3-2. If the active block is not down-satisfied, pool the active block with the next lower block and go to the step 2.

4. Iterate until the active block contains the element n.

2.2 Isotone regression model

Isotone regression model is described as

$$E(Y \mid X) = m(X) \tag{2.7}$$

where the regression function m is a monotone function. Note that if m is non-increasing, then we can recast the regression model by reversing the order of the support of X. Thus, a monotone function hereafter would mean a non-decreasing function. Let $(X^1, Y^1), \cdots, (X^n, Y^n)$ be i.i.d. random samples from the model (3.7). We assume that $(X^i)_{i=1}^n$ are ordered. The least squares estimator is defined by

$$\hat{m} = \arg\min_{f \in I} \sum_{i=1}^{n} (Y^i - f^i)^2 w(X^i) \tag{2.8}$$

where $I = \{f = (f^1, f^2, \cdots, f^n)^T \in R^n : f^1 \leq f^2 \leq \cdots \leq f^n\}$ denotes the set of isotonic vectors and w is a weight function defined on the support of X. Usually, the weight function w is chosen as a constant function.

By section 2.1, the least squares estimator is given by

$$\hat{m}(X^i) = \max_{s \leq i} \min_{t \geq i} \frac{\sum_{j=s}^{t} w(X^j) Y^j}{\sum_{j=s}^{t} w(X^j)}. \tag{2.9}$$

Thus, the least squares estimator is a local weighted average of the response vector. Bandwidth is chosen automatically by the given data and in fact, Wright (1981) showed that the bandwidth is asymptotically $cn^{-1/3}$ where c goes to infinity slowly enough. From this fact, we can guess that the estimator achieves cube root n convergence rate, which is different from convergence rate of usual nonparametric function estimation. We also see that the estimator is interpreted as the slope of the GCM of the weighted partial sum process. The estimator $\hat{m}$ defined on $\{X^i\}$ is extended to the step function:

$$\hat{m}(x) = \hat{m}(X^i), \quad \text{for } X^i \le x < X^{i+1},\ 1 \le i \le n-1,$$
$$\hat{m}(x) = \hat{m}(X^1), \quad \text{for } x < X^1,$$
$$\hat{m}(x) = \hat{m}(X^i), \quad \text{for } x \ge X^n.$$

Wright (1981) proved the asymptotic distribution of $n^{\alpha/(2\alpha+1)}(\hat{m}(x_0) - m(x_0))$ under the assumption $|m(x) - m(x_0)| = A|x - x_0|^\alpha(1 + o(1))$ as $x \to x_0$ for some A and $\alpha > 0$. When $\alpha = 1$, the estimator achieves cube root n convergence rate.

The least squares estimator (3.8) is suffered from boundary problem. This is because if Y^n is larger than any other response value Y^i, $i \le n-1$, then the least squares estimator gives $\hat{m}(X^n) = Y^n$ and thus no average effect occurs by the optimization.

From the viewpoint of smoothness, the estimator is not very satisfactory in that the estimate function gives a step function, not a smooth function. In order to improve smoothness of the estimator, smoothing technique such as kernel smoothing can be considered together with the isotone regression method. Friedman and Tibshirani (1984) suggested the following algorithm (smoothing-isotonization):

1. Smooth the response vector Y on X; $\hat{m}_p(X) = S_X Y$ where S_X is a linear smoother matrix.

2. Apply the Pool Adjacent Violators algorithm to find the closest monotone function $\hat{m}$ to $\hat{m}_p$.

In contrast, Mukerjee (1988) suggested the reverse algorithm (isotonization-smoothing):

1. Apply the Pool Adjacent Violators algorithm to the response vector Y to find the closest monotone function, say $\hat{m}_p$.

2. Smooth $\hat{m}_p(X)$ on X; $\hat{m}(X) = S_X \hat{m}_p(X)$ where S_X is a linear smoother matrix.

The resulting estimator at the reverse algorithm need not be a monotone function because smoothing an isotonic vector does not ensure isotonic property of the result. However, he showed that Nadaraya-Watson linear smoother preserves a monotone property if a kernel function K used in the smoothing step is log-concave. To see this, let $f(x \mid \theta) = K_h(x - \theta)$ for each $x \in \; supp(X)$. Let θ have the prior distribution given by

$$P(\theta = X^i) = 1/n, \; 1 \leq i \leq n$$

Then, the joint density and marginal density functions are respectively

$$f(x, \theta) = f(x \mid \theta) f(\theta) = \sum_{i=1}^{n} I(\theta = X^i) K_h(x - \theta)/n$$

and

$$f(x) = \sum_{i=1}^{n} K_h(x - X^i)/n,$$

Which implies that the posterior density function of θ is $f(\theta \mid x) = \sum_{i=1}^{n} I(\theta = X^i) K_h(x - \theta) / \sum_{i=1}^{n} K_h(x - X^i)$. Note that $\frac{f(\theta \mid x_2)}{f(\theta \mid x_1)}$ is increasing in θ since K is log-concave. Therefore, by Lemma2 (i) in TSH (Lehmann, p.85, 2^{nd} Edition), $\sum_{i=1}^{n} K_h(X^i - x)\hat{m}_p(X^i) / \sum_{i=1}^{n} K_h(X^i - x)$ is increasing in x.

Mammen (1993) compared the above two algorithms and proved that the two estimators are asymptotically first order equivalent.

2.3 Additive isotone regression model

Mammen and Yu (2007) suggested the backfitting method based on the isotone regression method to estimate monotone functions from the additive isotone regression model

$$E(Y \mid X) = c + \sum_{j=1}^{d} m_j(X_j) \tag{2.10}$$

where the additive component functions m_j, $1 \leq j \leq d$, are monotone. Without loos of generality, we assume that m_j's are non-decreasing. For identifiability, we impose the condition

$$\int m_j(x_j)\, dx_j = 0,\ 1 \leq j \leq d. \tag{2.11}$$

Let $(X^1, Y^1), \cdots, (X^n, Y^n)$ be i.i.d. random samples from the model (3.10). Also, let I_j, $1 \leq j \leq d$, be the set of isotonic vectors in R^n corresponding to the ordering of $X_j = (X_j^1, X_j^2, \cdots, X_j^n)^T \in R^n$, i.e.

$$X_j^{i_1} \leq X_j^{i_2} \leq \cdots \leq X_j^{i_n} \implies I_j = \{f \in R^n : f^{i_1} \leq f^{i_2} \leq \cdots \leq f^{i_n}\}.$$

Then, the least squares estimator for the regression model (3.10) is given as minimizer of

$$\sum_{i=1}^{n} (Y^i - c - m_1^i - m_2^i - \cdots - m_d^i)^2 \tag{2.12}$$

where $m_j = (m_j^1, m_j^2, \cdots, m_j^n)^T \in I_j$, $1 \leq j \leq d$. To estimate the additive component functions, we apply the isotone regression method iteratively. Without loss of generality, we drop the constant c.

Given $\hat{m}_1^{[r]}, \cdots, \hat{m}_{j-1}^{[r]}$ and $\hat{m}_{j+1}^{[r-1]}, \cdots, \hat{m}_d^{[r-1]}$, we update $\hat{m}_j^{[r]}$ by the following algorithm:

1. Minimize $\sum_{i=1}^{n} (Y^i - \hat{m}_1^{[r]\,i} - \cdots - \hat{m}_{j-1}^{[r]\,i} - \hat{m}_{j+1}^{[r-1]\,i} - \cdots - \hat{m}_d^{[r-1]\,i} - m_j^i)^2$ subject to $m_j = (m_j^1, m_j^2, \cdots, m_j^n)^T \in I_j$.

2. Apply the Pool Adjacent Violators algorithm to update $\hat{m}_j^{[r]}$.

3. Iterate until $||\hat{m}_j^{[r]} - \hat{m}_j^{[r-1]}||$ converges to zero for all j.

In general, one does not have the unique solution for (3.12). For example, if there exist non constant vectors $g_j = (g_j^1, g_j^2, \cdots, g_j^n)^T \in R^n$, $1 \leq j \leq d$, such that $\hat{m}_j + g_j \in I_j$ for each j and $\sum_{j=1}^d g_j^i = 0$ for $i = 1, 2, \cdots, n$, then $\hat{m}_j + g_j \in I_j$, $1 \leq j \leq d$ also becomes a solution of (3.12). When the least squares estimator exists uniquely, Mammen and Yu (2007) proved the algorithm convergence. They also shows the oracle property of the backfitting estimator.

Chapter 3

Estimation in the Varying Coefficient Model

The main interest of this chapter is to suggest an algorithm to estimate monotone coefficient functions in the varying coefficient model. We start from the basic model and will cover more general models step by step.

3.1 One-dimensional varying coefficient model

In this section, we consider the one-dimensional varying coefficient model

$$E(Y \mid X, Z) = f(X)Z \tag{3.1}$$

where f is a monotone function. Without loss of generality, f is non-decreasing. Let $(X^1, Z^1, Y^1), \cdots, (X^n, Z^n, Y^n)$ be i.i.d. random samples from the model (4.1). We assume that $(X^i)_{i=1}^n$ are ordered. Then, the least squares estimator

$\hat{f}$ is defined by

$$\hat{f} = \arg\min_{f \in I} \sum_{i=1}^{n} (Y^i - f^i Z^i)^2 \tag{3.2}$$

where $I = \{f = (f^1, f^2, \cdots, f^n)^T \in R^n : f^1 \leq f^2 \leq \cdots \leq f^n\}$ denotes the set of isotonic vectors in R^n. Since $(Y^i - f^i Z^i)^2 = (Y^i/Z^i - f^i)^2 (Z^i)^2$, (4.2) is equivalent to

$$\hat{f} = \arg\min_{f \in I} \sum_{i=1}^{n} (Y^i/Z^i - f^i)^2 (Z^i)^2. \tag{3.3}$$

Thus, we can apply the isotone regression theory of section 2.1 and section 2.2 to obtain the least squares estimator. The explicit form of the solution of (3.3) is given by

$$\hat{f}(X^i) = \max_{s \leq i} \min_{t \geq i} \frac{\sum_{j=s}^{t} (Z^j)^2 Y^j / Z^j}{\sum_{j=s}^{t} (Z^j)^2} = \max_{s \leq i} \min_{t \geq i} \frac{\sum_{j=s}^{t} Z^j Y^j}{\sum_{j=s}^{t} (Z^j)^2}. \tag{3.4}$$

We can see that the solution (3.4) is a local weighted average where the weights are supplied by covariate term Z^2, which is also mentioned in Hastie and Tibshirani (1993). For a graphical interpretation, it is straightforward to see the the solution is the slope of the GCM of the CSD $\{(\sum_{i=1}^{j} (Z^i)^2, \sum_{i=1}^{j} (Z^i Y^i)) : j = 0, 1, 2, \cdots, n\}$. In practice, we apply the Pool Adjacent Violators algorithm (PAVA) to compute the least squares estimator.

Actually, the solution (3.4) is interpreted as a projection. K is called *convex cone* if for any a, $b \geq 0$, x, $y \in K$ implies $ax + by \in K$. We can easily check that the set of isotonic vectors I is a convex cone. Recall that the one-dimensional isotonic least squares estimator in the isotone regression model is a projection of the response vector Y onto $I \in R^n$. In the varying coefficient model, the space we are working is closely related to the covariate vector Z.

For given $Z = (Z^1, Z^2, \cdots, Z^n)T \in R^n$, define

$$I_Z = \{h = (h^1, h^2, \cdots, h^n)^T \in R^n : h^i = f^i Z^i, f \in I\}.$$

Then I_Z is also closed convex cone. Thus, the least squares estimator $\hat{f}$ in (4.2) is a projection of the response vector Y onto the space I_Z with respect to the scalar product $\langle f, g\rangle = \sum_{i=1}^n f^i g^i$ where f and $g \in R^n$.

By (3.3), we have another projection interpretation of the least squares estimator $\hat{f}$. If we call $(Y^1/Z^1, Y^2/Z^2, \cdots, Y^n/Z^n)$ the observed coefficient vector, (3.3) tells that the least squares estimator $\hat{f}$ is a projection of the observed coefficient vector onto the space I with respect to the scalar product $\langle f, g\rangle_Z = \sum_{i=1}^n f^i g^i (Z^i)^2$ where f and $g \in R^n$. In other words, we gain the estimator $\hat{f}$ by projecting the observed coefficient vector onto the set of isotonic vectors where the scalar product is given by the covariate weighted scalar product.

3.2 Isotonic varying coefficient model

We consider the varying coefficient model

$$E(Y \mid X, Z) = \sum_{j=1}^{d} f_j(X_j) Z_j \tag{3.5}$$

where the coefficient functions f_j, $1 \le j \le d$, are monotone. Hereafter, we call the model (3.5) the isotonic varying coefficient model.

Let $(X^1, Z^1, Y^1), \cdots, (X^n, Z^n, Y^n)$ be i.i.d. random samples from the model (3.5). We define I_j for $1 \le j \le d$ as in section 2.3. The least squares estimator $\hat{f} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n) \in R^{n\times d}$, $\hat{f}_j \in R^n$, is given as minimizer of

$$\sum_{i=1}^{n}(Y^i - \sum_{j=1}^{d} f_j^i Z_j^i)^2 \tag{3.6}$$

over $f = (f_1, f_2, \cdots, f_n) \in I_1 \times I_2 \times \cdots \times I_n$.

For the least squares estimator $\hat{f} \in R^{n\times d}$, we assume that there does not exist $g_j = (g_j^1, g_j^2, \cdots, g_j^n)^T \in R^n$, $1 \le j \le d$, such that $\hat{f}_j + g_j \in I_j$ for each j and $\sum_{j=1}^{d} g_j^i Z_j^i = 0$ for $i = 1, 2, \cdots, n$. Under this assumption, the following relation

$$\hat{f}_j = \arg\min_{f_j \in I_j} \sum_{i=1}^{n}(Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - f_j^i Z_j^i)^2,\ 1 \le j \le d \tag{3.7}$$

holds. To see this, if $\tilde{f}_j = \arg\min_{f_j \in I_j} \sum_{i=1}^{n}(Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - f_j^i Z_j^i)^2$ for some j, then by the definition of $\hat{f} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n)$,

$$\sum_{i=1}^{n}(Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - \tilde{f}_j^i Z_j^i)^2 = \sum_{i=1}^{n}(Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - \hat{f}_j^i Z_j^i)^2 \tag{3.8}$$

should be hold. By the inequality $(\frac{1}{2}a + \frac{1}{2}b)^2 \le \frac{1}{2}a^2 + \frac{1}{2}b^2$, we have

$$\sum_{i=1}^{n}(Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - \frac{\tilde{f}_j^i + \hat{f}_j^i}{2} Z_j^i)^2 \le \sum_{i=1}^{n}(Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - \tilde{f}_j^i Z_j^i)^2.$$

Thus, the following equality

$$\sum_{i=1}^{n}(Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - \frac{\tilde{f}_j^i + \hat{f}_j^i}{2} Z_j^i)^2 = \sum_{i=1}^{n}(Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - \tilde{f}_j^i Z_j^i)^2$$

holds, which together with (3.8) leads to

$$\sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - \tilde{f}_j^i Z_j^i = \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - \hat{f}_j^i Z_j^i$$

for $i = 1, 2, \cdots, n$. Therefore, the underlying assumption gives $\tilde{f}_j = \hat{f}_j$. This means that the least squares estimator $\hat{f} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n) \in R^{n\times d}$ satisfies the following backfitting equations system,

$$\hat{f}_j = \arg\min_{f_j \in I_j} \sum_{i=1}^{n} (Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i Z_k^i - f_j^i Z_j^i)^2, \; 1 \leq j \leq d. \tag{3.9}$$

Similarly as in the above statements, we can also show that the assumption implies that the least squares estimator $\hat{f} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n) \in R^{n\times d}$ exists uniquely. We will call the assumption the *uniqueness assumption.*

Let $\Pi(\cdot \,|\, I_{Z_j})$ be a projection operator onto the set I_{Z_j} with respect to the scalar product $\langle f, g\rangle = \sum_{i=1}^{n} f^i g^i$ where f and $g \in R^n$.

Proposition 1 *Under the uniqueness assumption, the least squares estimator $\hat{f} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n) \in R^{n\times d}$ given in (3.6) satisfies the following backfitting equations system,*

$$\hat{f}_j = \Pi(Y - \sum_{k\neq j}^{d} \hat{f}_k(X_k) Z_k \,|\, I_{Z_j}), \; 1 \leq j \leq d$$

where $Y - \sum_{k\neq j}^{d} \hat{f}_k(X_k) Z_k \in R^n$ is a vector having its ith component as $Y^i - \sum_{k\neq j}^{d} \hat{f}_k^i(X_k^i) Z_k^i$.

Note that we cannot obtain the explicit form of the solution for (3.9). Numerically to get a solution, we employ the iterative method as in Mammen and Yu (2007). Suppose we want to update the jth coefficient estimate $\hat{f}_j \in R^n$ in the rth iteration and we are given $\hat{f}_k^{[r]}$ for $1 \leq k \leq j-1$ and $\hat{f}_k^{[r-1]}$ for $j+1 \leq k \leq d$. Then, update $\hat{f}_j^{[r]} \in R^n$ by the following algorithm:

1. Minimize $\sum_{i=1}^{n} (Y^i - \hat{f}_1^{[r]\,i} Z_1^i - \cdots - \hat{f}_{j-1}^{[r]\,i} Z_{j-1}^i - \hat{f}_{j+1}^{[r-1]\,i} Z_{j+1}^i - \cdots - \hat{f}_d^{[r-1]\,i} Z_d^i - f_j^i Z_j^i)^2$ subject to $f_j = (f_j^1, f_j^2, \cdots, f_j^n)^T \in I_j$.

2. Apply the Pool Adjacent Violators algorithm discussed in section 3.1 to update $\hat{f}_j^{[r]}$.

3. Iterate until $||\hat{f}_j^{[r]} - \hat{f}_j^{[r-1]}||$ converges to zero for all j.

The algorithm is based on a Dykstra (1983) which presents the algorithm to compute a projection of given vector onto the sum space of closed convex cones. The similar development discussed in Mammen and Yu (2007) shows that the iterative algorithm converges. To see this, the least squares estimator $\hat{f} = (\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_n) \in R^{n \times d}$ given in (3.6) is a projection of the response vector $Y \in R^n$ onto the sum space $I_{Z_j} + I_{Z_2} + \cdots + I_{Z_d}$ where I_{Z_j} for $1 \leq j \leq d$ is defined as I_Z in section 3.1. We observed that $1 \leq j \leq d$ is a closed convex cone given $Z_j \in R^n$. Thus, if $1 \leq j \leq d$ plays the role of H_j in the proof of Theorem 2 in Mammen and Yu (2007), the followed arguments lead to the convergence of the iterative algorithm under the uniqueness assumption.

Proposition 2 *Under the uniqueness assumption, $\hat{f}_j^{[r]}$ converges to $\hat{f}_j$ as r goes to infinity for $1 \leq j \leq d$.*

3.3 Extended varying coefficient model

In some cases, we can require only some of the coefficient functions to be monotone while other coefficient functions are smooth. Thus, the model (3.5) may not be satisfactory. In this section, we consider the varying coefficient model

$$E(Y \mid X, Z) = \sum_{j=1}^{d} f_j(X_j) Z_j + \sum_{k=d+1}^{d'} f_k(X_k) Z_k \tag{3.10}$$

where only a part of the coefficient functions, f_j for $1 \leq j \leq d$, is monotone. The model (3.10) contains both monotone functions and smooth functions as the coefficient functions, so the model (3.10) is more flexible than the model (3.5) discussed in section 3.2. We divide the nonparametric coefficient components into two parts: the dimension index j denotes the monotone coefficient functions and the index k denotes the smooth coefficient functions. In order to estimate the coefficient functions at one step, we constitute the backfitting equations by combining the method in section 3.2 and the classical backfitting method.

Let $(X^1, Z^1, Y^1), \cdots, (X^n, Z^n, Y^n)$ be i.i.d. random samples from the model (3.10). First, note that

$$E(Y - \textstyle\sum_{j=1}^{d} f_j(X_j)Z_j \,|\, X_1, Z_1, \cdots, X_d, Z_d) = \sum_{k=d+1}^{d'} f_k(X_k)Z_k$$
$$E(Y - \textstyle\sum_{k=d+1}^{d'} f_k(X_k)Z_k \,|\, X_{d+1}, Z_{d+1}, \cdots, X_{d'}, Z_{d'}) = \sum_{j=1}^{d} f_j(X_j)Z_j.$$

If $\hat{f}_k$'s for $d+1 \le k \le d'$ are available, we estimate f_j's for $1 \le j \le d$ by

$$(\hat{f}_1, \cdots, \hat{f}_d) = \arg\min_{f_j \in I_j} \sum_{i=1}^{n} (Y^i - \sum_{k=d+1}^{d'} \hat{f}_k(X_k^i)Z_k^i - \sum_{j=1}^{d} f_j^i Z_j^i)^2. \qquad (3.11)$$

If $\hat{f}_j$'s for $1 \le j \le d$ are available, we estimate f_k's for $d+1 \le k \le d'$ by

$$(\hat{f}_{d+1}, \cdots, \hat{f}_{d'}) = \arg\min_{f_k \in R^n} \sum_{i=1}^{n} (Y^i - \sum_{j=1}^{d} \hat{f}_j(X_j^i)Z_j^i - \sum_{k=d+1}^{d'} f_k^i Z_k^i)^2. \qquad (3.12)$$

For (3.11), the part of monotone coefficient functions $(\hat{f}_1, \cdots, \hat{f}_d)$ is the least squares estimator discussed in section 3.2 when the other part $(\hat{f}_{d+1}, \cdots, \hat{f}_{d'})$ is given. Thus, if we denote $\Pi(\cdot \,|\, I_{Z_j})$ as a projection operator onto the set I_{Z_j}, $1 \le j \le d$ with respect to the scalar product $\langle f, g \rangle = \sum_{i=1}^{n} f^i g^i$, we derive the following backfitting equations

$$\hat{f}_j = \Pi(Y - \sum_{j' \ne j}^{d} \hat{f}_{j'}(X_{j'})Z_{j'} - \sum_{k=d+1}^{d'} \hat{f}_k(X_k)Z_k \,|\, I_{Z_j}), \; 1 \le j \le d \qquad (3.13)$$

or equivalently,

$$\hat{f}_j(X_j^i) = \max_{s\le i}\min_{t\ge i}\frac{\sum_{l=s}^t Z_j^l(Y^l - \sum_{j'\neq j}^d \hat{f}_{j'}(X_{j'}^l)Z_{j'}^l - \sum_{k=d+1}^{d'} \hat{f}_k(X_k^l)Z_k^l)}{\sum_{l=s}^t (Z_j^l)^2}$$

for $1 \le j \le d$, where $Y - \sum_{j'\neq j}^d \hat{f}_{j'}(X_{j'})Z_{j'} - \sum_{k=d+1}^{d'} \hat{f}_k(X_k)Z_k \in R^n$ is a vector having its ith component as $Y^i - \sum_{j'\neq j}^d \hat{f}_{j'}(X_{j'}^i)Z_{j'}^i - \sum_{k=d+1}^{d'} \hat{f}_k(X_k^i)Z_k^i$.

For (3.12), we use the classical backfitting method to estimate the smooth coefficient functions $(f_{d+1}, \cdots, f_{d'})$ when the other part $(\hat{f}_1, \cdots, \hat{f}_d)$ is given. Let S_k for $d+1 \le k \le d'$ denotes Nadaraya-Watson linear smoother matrix for kth coefficient function. Then we derive the backfitting equations

$$\hat{f}_k = S_k(Y - \sum_{j=1}^d \hat{f}_j(X_j)Z_j - \sum_{k'\neq k}^{d'} \hat{f}_{k'}(X_{k'})Z_{k'}),\ d+1 \le j \le d' \qquad (3.14)$$

or equivalently,

$$\hat{f}_k(X_k^i) = \max_{s\le i}\min_{t\ge i}\frac{\sum_{l=1}^n Z_k^l(Y^l - \sum_{j=1}^d \hat{f}_j(X_j^l)Z_j^l - \sum_{k'\neq k}^{d'} \hat{f}_{k'}(X_{k'}^l)Z_{k'}^l)K_{h_k}(X_k^i, X_k^l)}{\sum_{l=1}^n (Z_k^l K_{h_k}(X_k^i, X_k^l))^2}$$

for $d+1 \le k \le d'$, where $Y - \sum_{j=1}^d \hat{f}_j(X_j)Z_j - \sum_{k'\neq k}^{d'} \hat{f}_{k'}(X_{k'})Z_{k'} \in R^n$ is a vector having its ith component as $Y^i - \sum_{j=1}^d \hat{f}_j(X_j^i)Z_j^i - \sum_{k'\neq k}^{d'} \hat{f}_{k'}(X_{k'}^i)Z_{k'}^i$.

For computational purpose, we suggest a two-step iterative algorithm. In the first step, let us given $\hat{f}_k^{[r_1-1]}$, $d+1 \le k \le d'$. We update $\hat{f}_j^{[r_1]}$, $1 \le j \le d$, and $\hat{f}_k^{[r_1]}$, $d+1 \le k \le d'$, according to the following algorithm:

1. Update $\hat{f}_j^{[r_1]}$, $1 \le j \le d$, by solving the equations,

$$\hat{f}_j^{[r_1]} = \Pi(Y - \sum_{j'\neq j}^d \hat{f}_{j'}^{[r_1]}(X_{j'})Z_{j'} - \sum_{k=d+1}^{d'} \hat{f}_k^{[r_1-1]}(X_k)Z_k \,|\, I_{Z_j}),\ 1 \le j \le d.$$

2. Update $\hat{f}_k^{[r_1]}$, $d+1 \leq k \leq d'$, by solving the equations,

$$\hat{f}_k^{[r_1]} = S_k(Y - \sum_{j=1}^{d} \hat{f}_j^{[r_1]}(X_j)Z_j - \sum_{k' \neq k}^{d'} \hat{f}_{k'}^{[r_1]}(X_{k'})Z_{k'}),\ d+1 \leq j \leq d'.$$

In the second step, we obtain the solution of 1 and 2 in the first step. Suppose we want to update the jth coefficient estimate $\hat{f}_j^{[r_1]} \in R^n$ for some $j \in \{1, \cdots, d\}$ in the r_2th iteration and we are given $\hat{f}_{j'}^{[r_1,r_2]}$ for $1 \leq j' \leq j-1$ and $\hat{f}_{j'}^{[r_1,r_2-1]}$ for $j+1 \leq j' \leq d$. Then, update $\hat{f}_j^{[r_1,r_2]} \in R^n$ by the following algorithm:

1. Minimize
   $\sum_{i=1}^{n}(Y^i - \hat{f}_1^{[r_1,r_2]}(X_1^i)Z_1^i - \cdots - \hat{f}_{j-1}^{[r_1,r_2]}(X_{j-1}^i)Z_{j-1}^i - \hat{f}_{j+1}^{[r_1,r_2-1]}(X_{j+1}^i)Z_{j+1}^i$
   $- \cdots - \hat{f}_d^{[r_1,r_2-1]}(X_d^i)Z_d^i - \sum_{k=d+1}^{d'} \hat{f}_k^{[r_1-1]}(X_k^i)Z_k^i - f_j^i Z_j^i)^2$
   subject to $f_j = (f_j^1, f_j^2, \cdots, f_j^n)^T \in I_j$.

2. Apply the Pool Adjacent Violators algorithm to update $\hat{f}_j^{[r_1,r_2]}$.

3. Iterate until $||\hat{f}_j^{[r_1,r_2]} - \hat{f}_j^{[r_1,r_2-1]}||$ converges to zero as r_2 goes to infinity for all j.

Suppose we want to update the kth coefficient estimate $\hat{f}_k^{[r_1]} \in R^n$ for some $k \in \{d+1, \cdots, d'\}$ in the r_2th iteration and we are given $\hat{f}_{k'}^{[r_1,r_2]}$ for $d+1 \leq k' \leq k-1$ and $\hat{f}_{k'}^{[r_1,r_2-1]}$ for $d+1 \leq k' \leq d'$. Then, update $\hat{f}_k^{[r_1,r_2]} \in R^n$ by the following algorithm:

1. Minimize
   $\sum_{i=1}^{n}(Y^i - \hat{f}_{d+1}^{[r_1,r_2]}(X_{d+1}^i)Z_{d+1}^i - \cdots - \hat{f}_{k-1}^{[r_1,r_2]}(X_{k-1}^i)Z_{k-1}^i - \hat{f}_{k+1}^{[r_1,r_2-1]}(X_{k+1}^i)Z_{k+1}^i$
   $- \cdots - \hat{f}_{d'}^{[r_1,r_2-1]}(X_{d'}^i)Z_{d'}^i - \sum_{j=1}^{d} \hat{f}_j^{[r_1]}(X_j^i)Z_j^i - f_k^i Z_k^i)^2$
   subject to $f_k = (f_k^1, f_k^2, \cdots, f_k^n)^T \in I_k$.

2. Apply the Pool Adjacent Violators algorithm to update $\hat{f}_k^{[r_1,r_2]}$.

3. Iterate until $||\hat{f}_k^{[r_1,r_2]} - \hat{f}_k^{[r_1,r_2-1]}||$ converges to zero as r_2 goes to infinity for all k.

Chapter 4

Numerical Studies

In this chapter, we check the performance of the algorithms suggested in Chapter 3 with simulation data. We setup and treat two simulation settings; one for the isotonic varying coefficient model of section 3.2 and the other for the extended varying coefficient model of section 3.3.

4.1 Simulated data

In this section, we illustrate simulation data to see how the algorithms suggested in Chapter 3 work. We consider two numerical experiments. For the algorithm suggested in section 3.2, we draw random samples from the following model

$$Y = f_1(X_1)Z_1 + f_2(X_2)Z_2 + \epsilon \tag{4.1}$$

where $(X_1, X_2) \sim U[-1, 1]^2$, $(Z_1, Z_2) \sim N_2(0, 1; \rho)$ and $\epsilon \sim N(0, 0.5^2)$ are independent. We set $f_1(x) = x^3$ and $f_2(x) = \sin(\pi x/2)$. Note that the coefficient functions f_1 and f_2 are non-decreasing functions.

For the algorithm suggested in section 3.3, we draw random samples from the following model

$$Y = f_1(X_1)Z_1 + f_2(X_2)Z_2 + f_3(X_3)Z_3 + f_4(X_4)Z_4 + \epsilon \qquad (4.2)$$

where $(X_1, X_2) \sim U[-1, 1]^2$, $(X_3, X_4) \sim U[0, 1]^2$, $(Z_1, Z_2, Z_3, Z_4) \sim N_4(0, 1; \rho)$ and $\epsilon \sim N(0, 0.5^2)$ are independent. We set $f_1(x) = x^3$, $f_2(x) = \sin(\pi x/2)$, $f_3(x) = \cos(\pi x/2)$ and $f_4(x) = 4(x - 1/2)^2$. The coefficient functions f_1 and f_2 are non-decreasing functions. However, the coefficient functions f_3 and f_4 have no order restriction. Thus, we partition the index set as $\{1, 2\}$ and $\{3, 4\}$. For Nadaraya-Watson smoother, we use Epanechnikov kernel function $K(z) = (3/4)(1 - z^2)I_{[-1,1]}(z)$.

4.2 Simulation Results

In this section, we present some simulation results to understand the finite sample performance. For each simulation setting, we iterate 500 times. Table 1 reports the empirical MISE for $\hat{f}_1$ and $\hat{f}_2$ from the model (4.1). We calculate the empirical MISE on the interior point $[-0.95, 0.95]$ because the least squares estimators for the isotonic varying coefficient model sometimes have serious boundary problem. We observe that MISE goes down as sample size grows. Figure 1 shows the plot of $25\%, 50\%$ and 75% quantile function of $\hat{f}_1$ and $\hat{f}_2$.

Table 2 reports the empirical MISE for $\hat{f}_1$, $\hat{f}_2$, $\hat{f}_3$ and $\hat{f}_4$ from the model

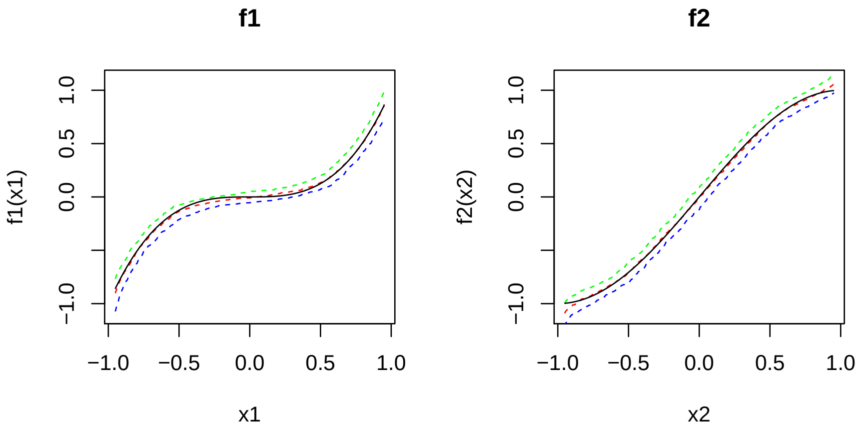

Figure 4.1: The real line shows the true coefficient function and the dashed lines indicate 25%, 50% and 75% quantile functions of the estimates, respectively.

(4.2). we calculate MISE on the interior point $[-0.95, 0.95]$ for $\hat{f}_1$ and $\hat{f}_2$ and on the interior point $[0.05, 0.95]$ for $\hat{f}_3$ and $\hat{f}_4$. We choose bandwidth $h_3 = 0.13$ and $h_4 = 0.15$ for Nadaraya-Watson smoothing. Figure 2 shows the plot of $25\%, 50\%$ and 75% quantile function of $\hat{f}_1$, $\hat{f}_2$, $\hat{f}_3$ and $\hat{f}_4$. From the simulation results, we see the algorithms suggested in section 3.2 and 3.3 works well in the finite sample.

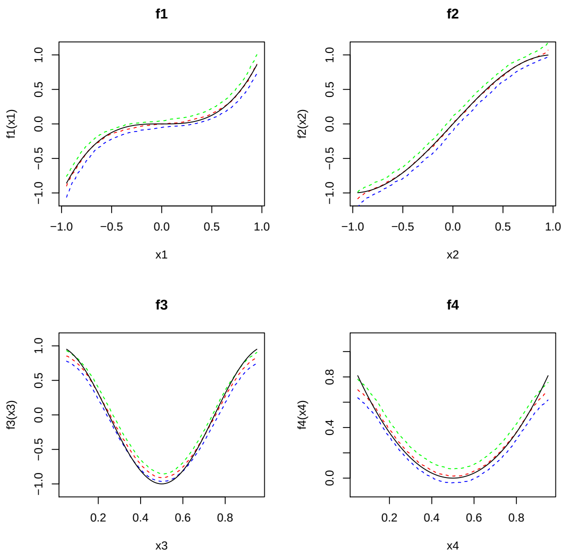

Figure 4.2: The real line shows the true coefficient function and the dashed lines indicate 25%, 50% and 75% quantile functions of the estimates, respectively.

Table 4.1: MISE of $\hat{f}_1$ and $\hat{f}_2$ from the model (4.1) with 500 iterations

	Function	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
$n = 200$	f_1	0.01293142	0.01386029	0.02262113
	f_2	0.01669305	0.01934828	0.02279188
$n = 400$	f_1	0.00769076	0.00790712	0.01118168
	f_2	0.00919209	0.00949084	0.01254080
$n = 800$	f_1	0.00463905	0.00463519	0.00596521
	f_2	0.00549653	0.00566826	0.00716874

Table 4.2: MISE of $\hat{f}_1$, $\hat{f}_2$, $\hat{f}_3$ and $\hat{f}_4$ from the model (4.2) with 500 iterations

	Function	$\rho = 0.1$	$\rho = 0.5$
$n = 200$	f_1	0.01431305	0.02709328
	f_2	0.01825558	0.01938944
	f_3	0.01582735	0.01707882
	f_4	0.00802740	0.00917662
$n = 400$	f_1	0.00808674	0.00842957
	f_2	0.00962990	0.01013734
	f_3	0.00906138	0.00950175
	f_4	0.00417197	0.00462413
$n = 800$	f_1	0.00469586	0.00498268
	f_2	0.00567082	0.00583674
	f_3	0.00521914	0.00553475
	f_4	0.00231001	0.00271022

Chapter 5

Conclusion

In this paper, we have revisited the method of monotone function estimation. First, we studied how to minimize the sum of the squares under the monotone constraints. We associated the solution with the slope of the GCM and obtained the explicit form of a solution. The solution was interpreted as a projection of the given data onto the close convex cone. For computational purpose, we also checked the algorithm called the Pool Adjacent Violators algorithm (PAVA). This algorithm says that to find the solution, we only need to find the *solution blocks*. From this solution, we can estimate the monotone regression function from the isotone regression model. The estimator is given by the slope of the GCM of the weighted partial sum process. We reviewed monotone function estimation in the additive isotone regression model discussed in Mammen and Yu (2007) and derived the backfitting algorithm.

This paper also provides the algorithms to estimate functions in the varying coefficient model when the coefficient functions are monotone. The sum of the error squares in the varying coefficient model is adapted to the minimization problem with random weights associated with the covariate. Thus, we apply the Pool Adjacent Violators algorithm to estimate the monotone coefficient functions. We constituted the backfitting equations for the least squares estimator and provided the iterative algorithm based on the backfitting equations. This algorithm motivated us to consider the varying coefficient model in which only several coefficient functions are under the monotone constraints. We divided the coefficient functions into two parts and suggested an algorithm which is a mixed version of the Pool Adjacent Violators algorithm and the classical backfitting algorithm. The numerical studies demonstrates that these algorithms bring the reasonable results for the finite sample case.

References

[1] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972) *Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley, New York.

[2] Dykstra, R.L. (1983) An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, **78**, 837-842.

[3] Fan, J. and Zhang, W. (1999) Statistical estimation in varying coefficient models. *Annals of Statistics*, **27**, 1491-1518.

[4] Friedman, J. and Tibshirani, R. (1984) The monotone smoothing of scatter plots. *Technometrics*, **26**, 243-350.

[5] Han, S.P. (1988) A successive projection method. *Mathematical Programming*. **40**, 1-14.

[6] Hastie, T. and Tibshirani, R. (1993) Varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, **55**, 757-796.

[7] Lee, Y.K., Mammen, E. and Park, B.U. (2012) Projection-type estimation for varying coefficient regression models. *Bernoulli*, **18**, 177-205.

[8] Leurgans, S. (1982). Asymptotic distributions of slope-of-greatest-convex-minorant estimators. *Annals of Statistics*, **10**, 287-296.

[9] MAMMEN, E. (1991) Estimating a smooth monotone regression function. *Annals of Statistics*, **19**, 724-740.

[10] MAMMEN, E. (1991) Nonparametric regression under qualitative smoothness assumptions. *Annals of Statistics*, **19**, 741-759.

[11] MAMMEN, E., MARRON, J.S., TURLACH, B.A. AND WAND, M.P. (2001) A general projection framework for constrained smoothing. *Statistical Science*, **16**, 232-248.

[12] MAMMEN, E. AND YU, K. (2007) Additive isotone regression. *IMS Lecture Notes-Monograph Series*, **55**, 179-195.

[13] MUKERJEE, H. (1988) Monotone nonparametric regression. *Annals of Statistics*, **16**, 741-750.

[14] OPSOMER, J.D. AND RUPPERT, D. (1997) Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, **25**, 186-211.

[15] PRAKASA RAO, B.L.S. (1969) Estimation of a unimodal density. *The Indian Journal of Statistics, Series A*, **31**, 23-36.

[16] WRIGHT, F.T. (1981) The asymptotic behavior of monotone regression estimates. *Annals of Statistics*, **9**, 443-448.

[17] YANG, L., PARK, B.U., XUE, L. AND HäRDLE, W. (2006) Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association*, **101**, 1212-1227.

국문초록

함수의 형태에 대한 제약하에서 함수를 추정하는 방법은 통계학자들 사이에서 큰 인기를 끌어오고 있으며 특별히 단조함수추정은 오랫동안 광범위하게 연구가 이루어져 왔다. 이 논문에서는 변수계수모형에서 계수함수들이 단조함수일 경우 함수를 추정하는 방법을 제안한다. 일반적으로 계수함수들을 추정하기 위해서는 설명변수의 효과를 고려해야 하는데 변수계수모형에서 설명변수들은 가중치로 들어가게 된다. 이를 위해 먼저 전통적인 회귀함수 모형에서 단조함수를 추정하는 방법과 알고리즘을 자세히 살펴본 후 가법모형으로의 확장을 다룬다. 이를 통해 변수계수모형에서 최적화 문제를 전통적인 방법으로 바꾸어 계수함수가 단조함수일 경우 추정하는 형태와 알고리즘을 제시한다. 또한 이 방법을 변수계수모형에서 계수함수의 일부분만 단조함수일 경우 추정하는 방법으로 확장을 시도한다. 마지막으로, 가상 자료로부터 수치적 연구를 이행한다.

주요어 : 변수계수모형; 단조회귀; 단조함수추정; 역적합 방법; 인접한 순서 위반자료 합침 알고리즘

학 번 : 2011-20253