MONOTONE *B*-SPLINE ESTIMATORS OF THE CONDITIONAL MEAN FUNCTION

YOUSEF JAWAD ATATRAH

Department of Mathematics and Statistics, Ministry of Education, United Arab Emirates

yousef.atatrah@moe.gov.ae

Abstract

In this research, regression and smoothing spline approximations are used and compared for estimating the conditional mean function. Major attention is directed towards shape constrained estimation. In many applications monotonicity is an integrated part of the regression functions $g(\cdot)$ being fitted. Monotonicity is obtained here free of charge in the sense that the constrained fits inherit the asymptotic properties of the unconstrained estimates. The main tool is the use of quadratic *B*-splines. Some simulation experiments have been undertaken to evaluate finite-sample performance of the presented monotone 'regression' and 'smoothing' spline estimators $\hat{g}_r^*$ and $\hat{g}_s^*$. The monotone estimator $\hat{g}_{rearr}^*$ obtained by applying the modern rearrangement technique is used as a benchmark in various constrained (linear monotone, monotone concave and only monotone) scenarios, for different sample sizes. The resulting Mean Squared Error estimates indicate that $\hat{g}_{rearr}^*$ is the winner only when the true regression function is linear monotone. The smoothing spline $\hat{g}_s^*$ is superior in the other scenarios when it comes to estimate the regression mean. Practical guidelines to effect the necessary computations and comparisons of the different estimators are provided by making use of the R software.

Keywords: Monotonicity, *B*-splines Approximations, Regression Spline Estimator, Smoothing Spline Estimator, Rearrangement Estimator.

الملخص

في هذا البحث, تم استخدام ومقارنة تقريبات regression and smoothing spline وذلك لتقريب اقتران المتوسط الشرطي. الاهتمام الرئيسي يتمثل نحو تقدير الشكل المقيد(المشروط). في العديد من التطبيقات، تعتبر Monotonicity جزءًا لا يتجزأ من دالة الانحدار (٠)g . يتم الحصول على Monotonicity هنا مجانًا بمعنى أن القيم المقيدة ترث الخصائص المقاربة للتقديرات غير المقيدة. الأداة الرئيسية هي استخدام دالة B-spline التربيعية. تم إجراء بعض تجارب المحاكاة لتقييم أداء عينة منتهية للتقريبات monotone 'regression' and 'smoothing' spline estimators $\hat{g}_r^*$ and $\hat{g}_s^*$. تم الحصول على تقريب $\hat{g}_{rearr}^*$ بواسطة تطبيق تقنية إعادة الترتيب الحديثة وتم استخدامها كمعيار في سيناريوهات مختلفة مشروطة (خطي monotone ، مقعر monotone و monotone فقط)، لأحجام عينات مختلفة. تقريب Mean Squared Error الناتج يشير الى أن $\hat{g}_{rearr}^*$ هو الأفضل فقط عندما تكون دالة الانحدار الأصلية $g(x)$ خطي monotone. $\hat{g}_s^*$ يتصرف بشكل أفضل في السيناريوهات الأخرى عندما يتعلق الأمر بتقريب دالة متوسط الانحدار. تم توفير إرشادات عملية للتأثير على الحسابات والمقارنات اللازمة للتقريبات المختلفة من خلال استخدام برنامج الإحصاء R.

الكلمات المفتاحية: Monotonicity , تقريبات spline $B-$, تقريب Regression Spline , تقريب Smoothing Spline , تقريب إعادة الترتيب.

1. Introduction

Nonparametric regression analysis is an increasingly popular tool for the purpose of data smoothing including kernel estimators (Gasser and Muller (1979)), smoothing spline estimators (Eubank (1988)), regression spline estimators (Friedman and Silverman (1989)) and B-spline estimators (He and Shi (1994)). We refer to the books of Wahba (1990) for an overview on the topic of spline models in traditional regression analysis, and to the books of de Boor (2001) and Schumaker (2007) for a modern treatment of splines.

A basic problem in many areas of statistics is the estimation of an unknown target function $g(x)$. In this paper we focus on the problem of estimating the conditional mean function using regression and smoothing B-spline approximations of different orders in the unconstrained case and under the monotonicity constraint.

The conditional mean function describes how the mean of a response variable Y changes with a vector of covariates X and minimizes a sum of least-square errors. In many applications, monotonicity is an integrated part of the function being fitted. For example, growth curves (e.g., weight or height of growing objects over time) are known to be increasing. Typical examples in economics include the evolution of outputs (Y) versus the stock of capital (X) at the country level, expenditures (Y) versus incomes (X) at the household level (Lee et al. (2009)). Other practical applications appear in medical sciences where the probability of contracting a certain disease (say cancer) depends monotonically on certain factors (say smoking frequency, drinking frequency and weight) (Dette and Scheder (2006)). Such examples are abundant in economics, environment, medical sciences and other areas (see, e.g., Ramsay (1988)).

The motivation of using splines in this work lies in their unmatched flexibility and adaptivity as well as their great approximation power. Splines are constructed as piecewise polynomials with specified continuity constraints. These continuity characteristics and the number of parameters defining a spline function depend on a knot mesh at which the polynomial pieces are connected. The main challenge when optimizing splines is determining the number and the locations of the knots. This requires a good initial guess of the knot locations (Ruppert (2002)). Once the sequence of knots is given, the splines can easily be computed for any desired order. There are three general approaches to spline fitting: regression splines, smoothing splines and penalized splines. The fundamental difference between the regression and smoothing splines is that smoothing splines explicitly penalize roughness and use the data points themselves as potential knots whereas regression splines place knots at equidistant or equiquantile points. A special class of splines, called B-splines, is a generalization of the Bezier curve (Racine). B-spline estimates are defined as the scalar product of their normalized basis functions [having order $(p + 1)$ and number of inter-knots segments k_n] and the coefficients of these basis functions which are obtained by solving a programming problem (de Boor (2001)).

The monotonization method we propose is inspired from He and She (1998). It is based on the use of quadratic B-splines on a selected set of knots.

Monotone regression and smoothing splines can be obtained by adding simple linear constrains to the program already in use for calculating the unconstrained estimator. Cubic and higher-order splines are more appealing for smoothness, but monotonicity can no longer be characterized as linear constraints at the knots. For our purpose of estimating the regression mean curve, we will compare the presented monotone quadratic B-spline fits with a monotonic estimate obtained by applying the promising rearrangement technique initiated by Dette, Neumeyer and Pilz (2006) and popularized by Chernozhukov, Fernandez-Val and Galichon (2009). This benchmark estimator is obtained via a rearrangement transformation of the original unrestricted estimate, say $\hat{g}(x)$ of the target function $g(x)$ to a monotonic estimate $\hat{g}^*_{rearr}$. The increasing rearrangement operator simply transforms a function g to the quantile function g^* of the random variable $g(U)$ when $U \sim \mathcal{U}(0,1)$. The rearranged estimator $\hat{g}^*_{rearr}$ has the advantage over the original estimator to be monotone whenever the latter is not monotonic, but also to have a smaller estimation error in the L^p norm.

A huge amount of research has been carried out in the past few decades on nonparametric estimation of the conditional mean function based on the idea of regression and smoothing splines. More recent references on the topic of smoothing splines include Wahba (1990), Hardle (1990), Hastie and Tibshirani (1990), Green and Silverman (1994) and Eubank (1999). The choice of the smoothing parameter in connection with the averaged mean squared error was initiated in a series of early papers, including Wahba and Wold (1975) and Craven and Wahba (1978).

A number of authors have come up with different solutions to the problem of estimating regression curves using regression spline techniques including Stone (1985), Stone (1994), and Huang (2003). More recent attempts of using monotonized spline smoothers can be found in the context of mean regression problems including Lu et al. (2007), Meyer (2008), Wang and Yang (2009) and Pya and Wood (2015). Early works combining smoothness with multiple shape restrictions in the regression setting include, for instance, Wright and Wegman (1980) and Ramsay (1988).

Restricted spline smoothing has also been studied by Mammen (1991) and Mammen and Thomas-Agnan (1999), Pal and Woodroofe (2007) with minimizing a penalized least squares criterion incorporating monotonicity constraints. Finally, Chernozhukov, Fernandez-Val and Galichon (2009) proposed point and interval estimators of monotone functions using the rearrangement method, first initiated by Dette, Neumeyer and Pilz (2006).

To evaluate finite-sample performance of the presented monotone 'regression' and 'smoothing' spline estimators $\hat{g}_r^*$ and $\hat{g}_s^*$, we have undertaken some simulation experiments. The experiments employ three constrained scenarios: linear monotonicity, monotone concavity and single monotonicity. We compare the accuracy of these spline smoothers relative to the 'rearranged' estimator $\hat{g}_{rearr}^*$ by computing Monte Carlo estimates of their bias and mean-squared error for different sample sizes. The choice of smoothing parameters for regularizing both estimated quadratic spline functions is a major issue in practice, but the monotonicity constraint makes this selection easier than the unconstrained smoothing problem: it reduces sharp changes in the slope and curvature of the estimated regression functions. Considering a set of knots equally spaced in percentile ranks, an adequate number k of inter-knot segments in the 'regression' spline can be determined by analogy to the popular Akaike information criterion (AIC). In what concerns the 'smoothing' spline estimator, we implement a Schwarz information criterion (SIC) to select the optimal smoothing parameter λ.

2. Estimating the Conditional Mean

As the main purpose of this study is to use B-spline approximations for estimating conditional mean function, we shall first give the general definition of a polynomial spline.

Definition

Denote a partition of an interval $[a, b]$ by $a = t_0 < t_1 < \ldots t_{k_n} = b$. For an integer $p \geq 0$, a polynomial spline of order $(p + 1)$ with simple knots $t_1, \ldots, t_{k_n - 1}$ is any function $s(\cdot)$ from $[a, b]$ to $\mathbb{R}$ such that

- $s(\cdot)$ is continuously differentiable until order $(p + 1)$ [if $p \geq 1$],

- The restriction of $s(\cdot)$ to inter-knot intervals $(a, t_1], \ldots, (t_i, t_{i+1}], \ldots, (t_{k_n-1}, b]$, coincides with a polynomial of degree less than or equal to p.

In this research we focus on estimating the regression mean function using both regression and smoothing B-spline approximations under the unconstrained and the monotonicity constraint based on the least-squares principle. We restrict ourselves to the interval $[a, b] \equiv [0, 1]$.

The Problem

Suppose that n pairs of observations $\{(x_i, y_i), i = 1, 2, \ldots, n\}$, with $a = x_0 < x_1 < \ldots < x_n < x_{n+1} = 1$, are available to estimate the mean function

$$g(x) = \arg \min_{\theta \in \mathbb{R}} \mathbb{E}[(Y - \theta)^2 | X = x] \tag{1}$$

We consider the following regression model:

$$y_i = g(x_i) + u_i, i = 1, 2, \ldots, n, \tag{2}$$

where the regression errors u_i represent a random noise with mean 0 given the covariate x_i.

We assume that the function g has a uniformly continuous and bounded second order derivatives. It is then well known that the functions g and g' can be uniformly approximated by quadratic B-splines and their derivatives. Here we restrict ourselves to $x \in [0,1]$.

Let $0 = t_0 < t_1 < \ldots t_{k_n} = 1$ be a partition of $[0, 1]$, and let $N = k_n + p$, where k_n represents the number of inter-knot segments and $(p + 1)$ defines the order of the spline approximation. We will denote by $S_{P,T}$ the space of polynomial splines of order $(p + 1)$ with knot mesh $(t_i)_{i=0}^{k_n}$.

Let $\pi(x) = (\pi_1(x), \ldots, \pi_N(x))^T$, with $\pi_j(x)$ being the normalized B-spline basis functions. The motivation here is to estimate the conditional mean using regression and smoothing B-spline approximations.

2.1. Regression Spline Estimators

2.1.1. Unconstrained spline smoothers

According to the regression model (2), a regression B-spline estimate $\hat{g}_r \in S_{P,T}$ of the conditional mean $g(x)$ can be defined as

$$\hat{g}_r = \pi(x)^T \hat{\alpha}_r, \tag{3}$$

where $\hat{\alpha}_r$ can be formulated as follows

$$\hat{\alpha}_r = \arg\min_{\theta \in \mathbb{R}^N} \sum_{i=1}^{n} (y_i - \pi(x_i)^T \alpha)^2. \tag{4}$$

The least-squares minimization problem (4) represents a special subclasses of convex optimization problems without constraints. Write $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^N$, and define the basis matrix $\pi \in \mathbb{R}^{n \times N}$ by

$$\pi_{i,j} = \pi_j(x_i), i = 1,,\ldots,n, \; j = 1,\ldots,N$$

(i.e. jth column of π gives the evaluations of π_j over the points $x_1, \ldots, x_n$). Then, according to Boyd and Vandenberghe (2004), the analytical solution of this minimization problem is given by

$$\hat{\alpha} = (\pi^T \pi)^{-1} \pi^T y \tag{5}$$

There are good algorithms and software implementations for solving least-squares problem to high accuracy, with very high reliability. Least-squares problems are used widely in statistical applications such as interpolation, extrapolation and smoothing of data, and also they used in statistical interpretations, for example in the R package "crs" in order to build the B-spline estimator of the regression mean function.

The selection of knots

For the selection of the initial knots $T = (t_i)_{i=0}^{k_n}$, we can perform "crs" function through "crs" package using the options "uniform" or "quantile", but for choosing the optimal number of knots

we use AIC criteria based on least-squares deviation which is defined as follows

$$AIC(T) = \log\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{g}_r(x_i))^2\right] + 2(k_n + p)/n \,, \qquad (6)$$

where k_n and p are defined previously.

The unconstrained regression B-spline estimate of the conditional mean can be implemented easily using **crs** function through **crs** package (Racine, J., Nie, Z. and Ripley, B. D. (2015)) in R software. The coefficients $\hat{\alpha}_r$ of B-spline basis functions estimate can also be estimated though crs function. The number of knots can be calculated through crs function using the option 'segments', where the number of segments is equal to the number of knots -1. The knots can be generated using (default knots $= 0$ quantiles 0) specifying where knots are to be placed. Quantiles specifies knots placed at equally spaced quantiles and knots $= 0$ quantiles 0 specifies knots placed at equally spaced intervals, we can also use the option degree for specifying the polynomial degree of the B-spline basis for each dimension of the continuous viable x (default degree $= 3$, i.e. cubic spline). Finally, we perform AIC criterion for selecting the optimal number of knots by computing the regression B-spline estimators for $N = 0,\ldots,$nknots, and then select N that corresponds to the smallest AIC.

The quadratic regression B-spline estimator is a good tool for estimating the unknown regression curves, but of course higher-order splines are often more appealing for smoothness. This section carry out comparison between quadratic and cubic regression B-spline estimators of the mean function.

2.1.2. Constrained quadratic spline smoothers

With the monotonicity constraint in mind, the isotonization methods that have been proposed and studied in the literature by most authors tend to be either much more computationally expensive or less flexible for modeling or harder to analyze mathematically. We propose in this section a simple but effective monotone smoothing method based on constrained least-squares deviation principle. More specifically, we focus on quadratic constrained B-spline estimation.

A quadratic monotone regression B-spline estimate $\hat{g}_r^* \in S_{P,T}$ of the conditional mean $g(x)$ can be defined as

$$\hat{g}_r^* = \pi(x)^T \hat{\alpha}_r^*, \tag{7}$$

where $\hat{\alpha}_r^*$ can be formulated as follows

$$\hat{\alpha}_r^* = \arg \min_{\theta \in \mathbb{R}^N} \sum_{i=1}^{n} (y_i - \pi(x_i)^T \alpha)^2. \tag{8}$$

 Subject to

$$(\pi'(t_j))^T \alpha \geq 0, j = 0,1,\ldots, k_n, \tag{9}$$

For the computation of the coefficients $\hat{\alpha}_r^*$ and the estimator itself, first we create the basis functions of order 2 of the B-spline using 6 knots via the quantile method. Afterwards, we used the constrained least-squares algorithm with $(k_n + 1)$ linear constraints for solving the quadratic minimization problem. The resulting solution gives us a vector including the coefficients of the basis functions of the B-spline, and then we make scalar product between the coefficients and the basis functions in order to get the quadratic monotone B-spline curve. The implementation of this estimator was already done by Jeffrey S. Racine.

2.2. Smoothing Spline Estimators

In this section, we use the smoothing spline technique for estimating the conditional mean which is based on the least-squares principle.

2.2.1. Unconstrained spline smoothers

Using the regression model (2), the smoothing B-spline estimate $\hat{g}_s(x)$ of the conditional mean $g(x)$ can be defined as

$$\hat{g}_s(x) = \pi(x)^T \hat{\alpha}_s, \tag{10}$$

where

$$\hat{\alpha}_s = \arg \min_{\theta \in \mathbb{R}^N} \sum_{i=1}^{n} (y_i - \pi(x_i)^T \alpha)^2 + \lambda \int_0^1 \left[\pi^/(t)^T \alpha \right]^2 dt \quad (11)$$

Subject to the monotonicity constraints

$$(\pi^/(t_j))^T \alpha \geq 0 \, , j = 0,1,\ldots, k_n \, , \quad (12)$$

The first term of (11) captures the fit to the data and the second one penalizes the curvature. The penalty constant $\lambda > 0$ plays the role of the smoothing parameter which controls the tradeoff between the two terms. A λ that is too close to zero will yield an estimate that interpolates the data, and a λ that is too big will produce an estimate practically equivalent to the linear regression estimate of the data.

For the computation of the coefficients $\hat{\alpha}_s$, we used the code that was implemented by Mary C. Meyer which is based on least-squares algorithm. In this code, the "coneproj" package has been employed in R software with the function "penspl" which takes the arguments, (1 for monotone increasing and 2 for monotone decreasing), x, y, k (the number of knots chosen by "uniform" method), q (degree of the penalty) and ($\lambda > 0$).

Smoothing parameter selection

In all smoothing techniques, a critical problem is the selection of the smoothing parameter λ. When using smoothing splines one does not need to choose the location of knots, since the knots are chosen to be typically the design points or the number of knots is too large, and the smoothness of the estimate is controlled only via the smoothness parameter.

In the literature, there are several methods for choosing λ including Schwarz-type information criterion (SIC) used in Koenker et al. (1994), and He, Ng and Portnoy (1998), Cross validation (CV) and generalized cross-validation (GCV) (Wahba (1985)). Here we will restrict our attention to (SIC) criterion that is defined as follows

$$SIC(\lambda) = \log \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{g}_\lambda(x_i))^2 \right] + \frac{1}{2} p_\lambda \log(n)/n \, , \quad (13)$$

where $\hat{g}_\lambda$ is the smoothing spline estimator $\hat{g}_s$ that corresponds to the specified smoothing parameter λ and p_λ is the number of interpolated data points that serves as a dimensionality measure of the fitted model.

It is important to note that the first term of SIC becomes infinitely small if $\hat{g}_\lambda$ interpolates every single data point. As a result, the λ that minimizes SIC could be too small for unconstrained fits.

2.2.2. Constrained quadratic spline smoothers

When taking into account the monotonicity constraint, then the quadratic smoothing B-spline estimate of the conditional mean can be defined as

$$\hat{g}_s^*(x) = \pi(x)^T \hat{\alpha}_s^*, \tag{14}$$

where

$$\hat{\alpha}_s^* = \arg\min_{\theta \in \mathbb{R}^N} \sum_{i=1}^{n} (y_i - \pi(x_i)^T \alpha)^2 + \lambda \int_0^1 \left[\pi'(t)^T \alpha\right]^2 dt. \tag{15}$$

Subject to the monotonicity constraints

$$(\pi'(t_j))^T \alpha \geq 0 \,, j = 0,1,\ldots, k_n \,, \tag{16}$$

For the calculation of $\hat{g}_s^*(x)$ estimator, we use the function **penspl** with the same arguments as in the unconstrained case and with the constrained fit (cfit).

3. Results and Discussion

The following examples will be studied with two real datasets in order to understand the monotonicity and non-monotonicity of B-spline smoothing estimators of the conditional mean function.

Example 1: This data example serves to illustrate the case where the regression B-spline estimate doesn't enjoy the monotonicity property even if $g(\cdot)$ does. We use the dataset "iris" in the R

package MASS, this dataset contains 150 pairs of observations of the measurements in centimeters of the variables, Sepal Length (x) and Petal Length (y). We use the regression B-spline of orders 2 and 3 to estimate the conditional mean function as appears in figure 1.
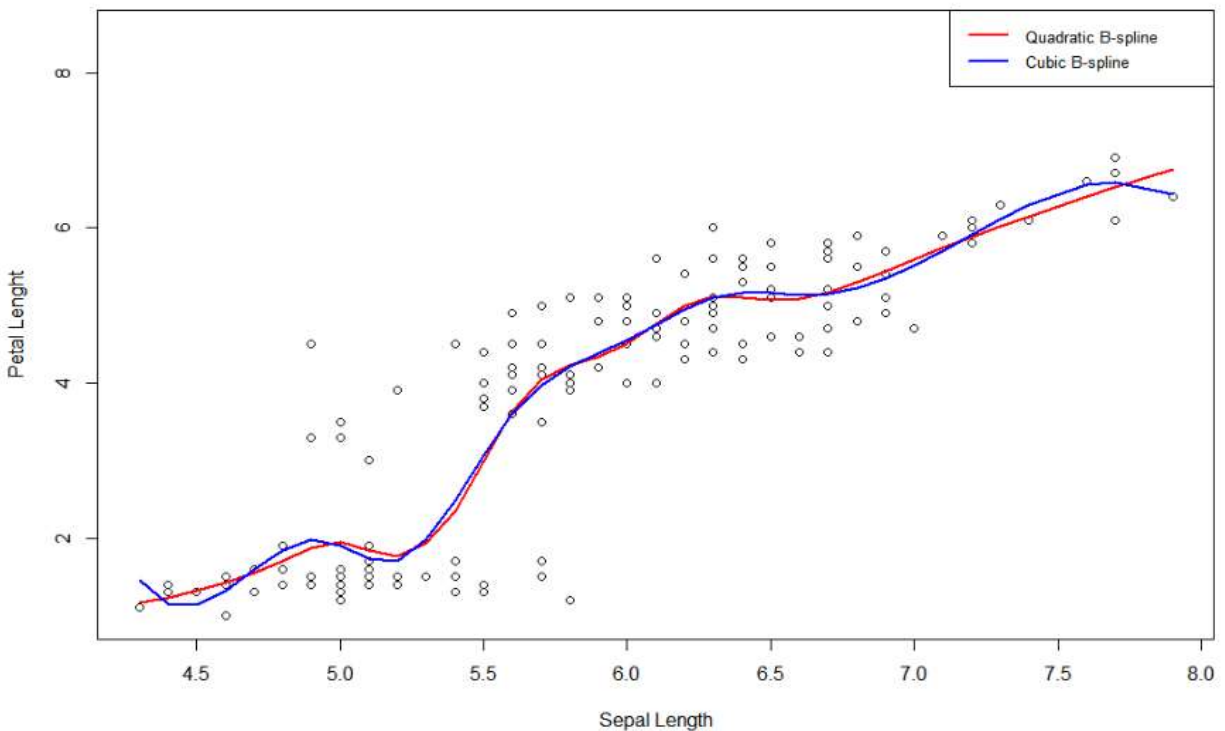

Figure 1: Scatter plot of the iris data, along with unconstrained quadratic and cubic regression B-spline of the conditional mean using 9 initial knots.

For the curves obtained in figure 1, we generate 9 initial knots via the quantile method. We choose the optimal number of knots that corresponds to the smallest value of AIC (see the R code).The final knots selected for the computations of the quadratic B-spline estimate are still 9 points located at (4.3, 4.9, 5.1, 5.5, 5.8, 6.1, 6.4, 6.8, 7.9), while the final knots selected for the cubic B-spline are located at (4.3, 4.93, 5.2, 5.6, 6, 6.3, 6.7, 7.9).

Finally, we can notice from figure 1 that the fitted curves are not over all monotone, whearas the true regression is believed to be monotone. In the next section, we try to solve this problem by introducing the monotonicity constrained regression B-spline estimate of the regression function.

Example 2: Figure 2 provides the monotone quadratic regression B-spline fit of the regression function for the "iris" dataset.
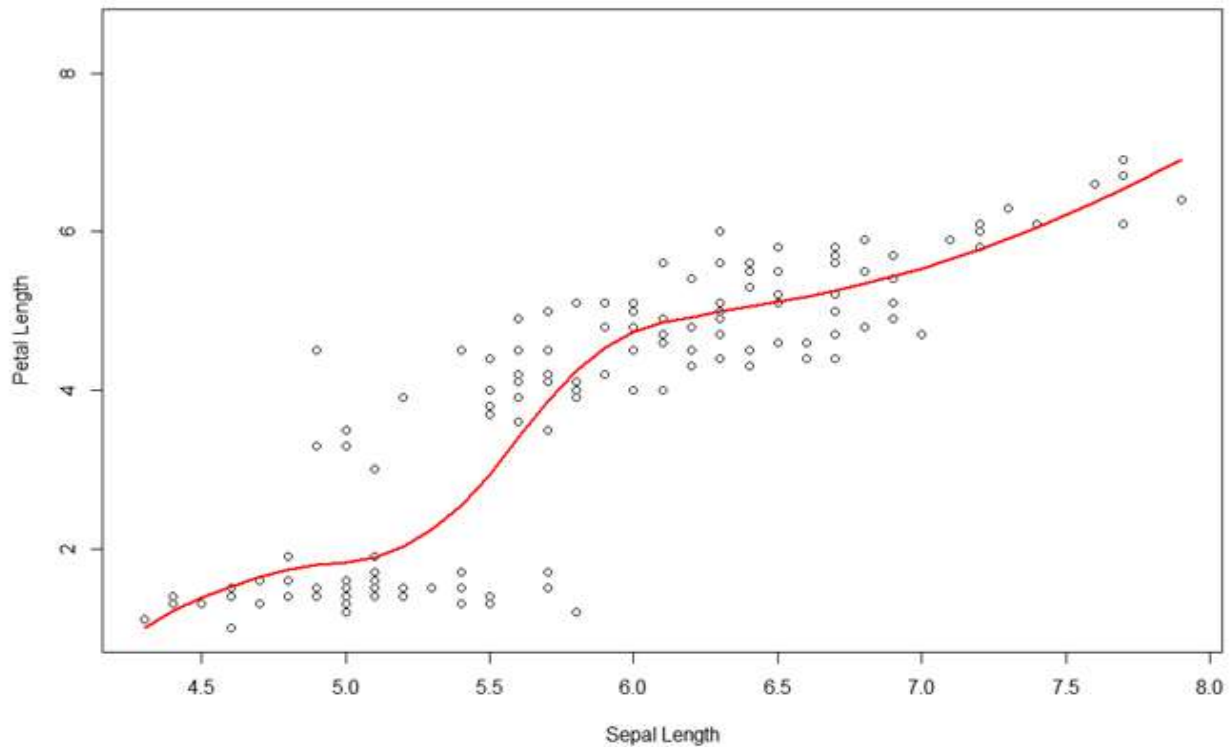

Figure 2: Scatter plot of iris, along with monotone quadratic regression B-spline of the conditional   mean using 6 initial knots.

For the computation of our monotone estimator we generate 6 initial knots via the quantile method. The optimal number of knots was selected using AIC criteria are still 6 points located at (4.3, 5, 5.6, 6.1 ,6.52, 7.9). We can see from figure 2 that this estimator fits the data well.

Example 3: In this example, we use the "onion" dataset where the trend of the data is enjoying the monotonicity property. This dataset has been employed by Mary C. Meyer (2008) and exits in the R package "SemiPar". This dataset contains 84 sets of observations from an experiment involving the production of white Spanish onions in two South Australian locations with two variables, dens (x): a real density of plants (plants per square meter) and yield (y): onion yield (grams per plant).

The obtained regression and smoothing spline fits of orders 2 of the conditional mean are graphed in figure 3.
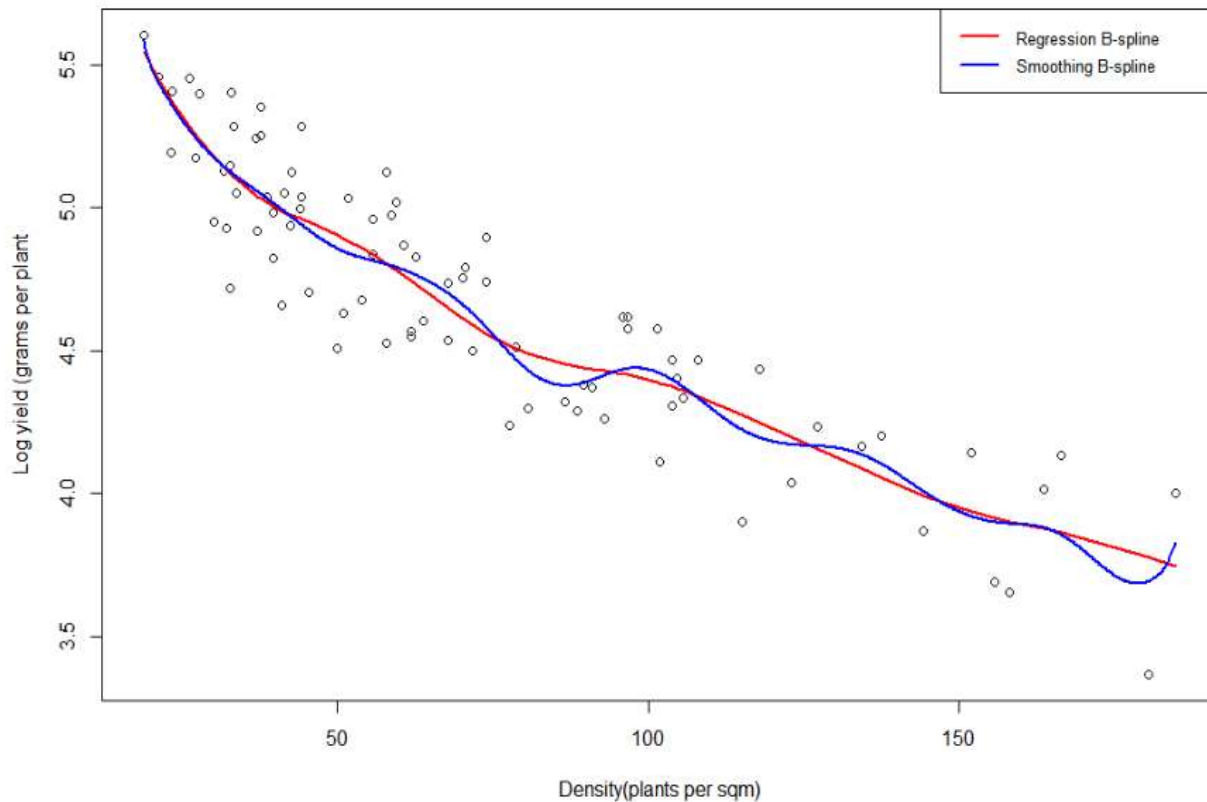

Figure 3: Comparison of fits to the onion data. The unconstrained quadratic regression and smoothing spline estimates of the conditional mean.

For the regression B-spline estimate we use 9 initial knots chosen by making use of the uniform method. The final knots are located at (18.78, 42.5, 66.2, 89.9, 113.6, 137.3, 161.04, 184.75), while the optimal λ in the smoothing spline was selected via SIC.

Example 4: Figure 4 provides the monotone quadratic regression and smoothing B-spline fits of the mean function for the "onion" dataset.
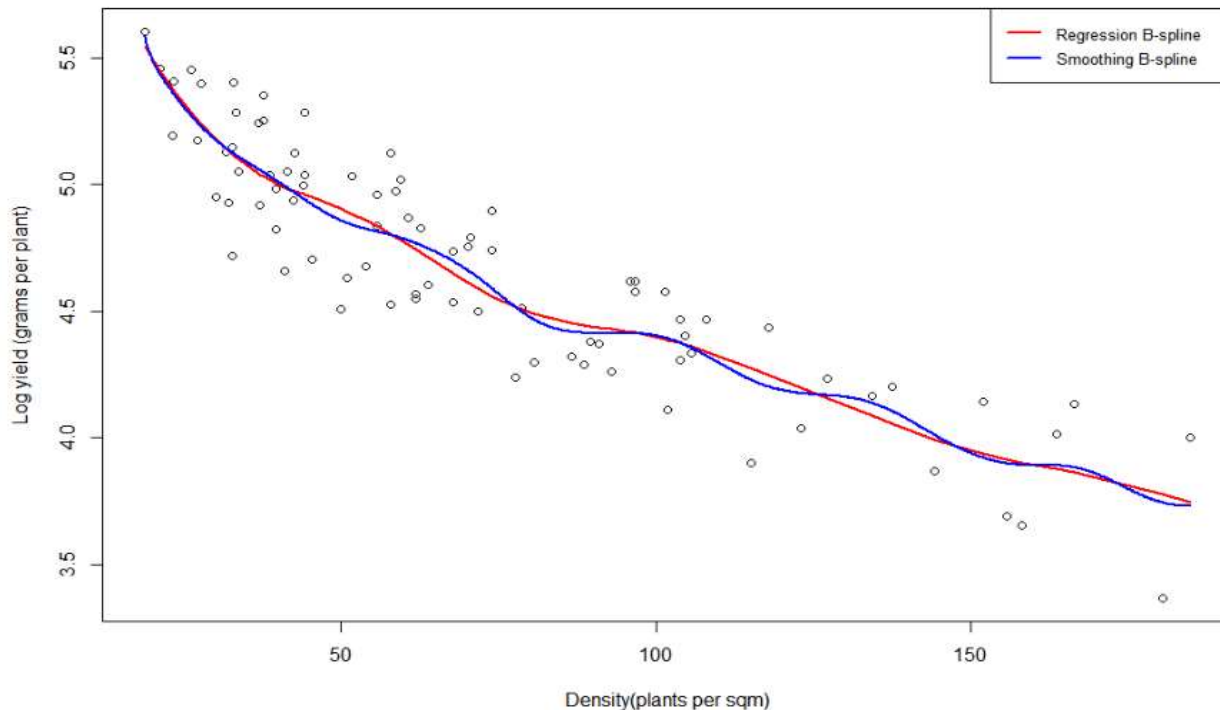
Figure 4: Comparison of fits to the onion data. The monotone quadratic regression and smoothing spline estimates of the conditional mean.

For the regression B-spline estimate we used 9 initial knots chosen by making use of the uniform method. The final knots are still 9 points located at (18.78, 39.52, 60.27, 81.01, 101.76, 122.51, 143.25, 164, 184.75), while the optimal λ in the smoothing spline was selected via SIC.

4. Some Monte-Carlo Evidence

In this study, some simulation experiments were used to compare the performance of the monotone estimators: regression B-spline estimator $\hat{g}_r^*(x)$ and the smoothing B-spline estimator $\hat{g}_s^*(x)$ of the conditional mean in terms of the Mean Squared Error (MSE) and the $Bias$. Both estimators are quadratic and constrained to be monotone increasing. These two estimators will be compared with the monotone estimator $\hat{g}_{rearr}^*(x)$. The experiments all employ the regression model

$$y_i = g(x_i) + u_i, i = 1, 2, \ldots, n, \tag{17}$$

where $x_i's$ are uniformly distributed $U(0,1)$, $u_i's$ represent random errors generated from $N(0,0.1)$, and the mean function is linear monotone $g(x) = x$, or monotone concave $g(x) = \sqrt{x}$, or only monotone $g(x) = exp(-5 + 10x)/(1 + exp(-5 + 10x))$. The experiments use different sample sizes $n = 10, 20, 30, 50, 100, 200$.

In order to obtain the estimated MSE and $Bias$ of all estimators, one can proceed as follows:

a) Generate 1000 samples of size n from the model (17).

b) For each simulated sample, calculate the MSE and $Bias$ for all estimators as follows

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{g}(x_i) - g(x_i))^2, \quad Bias = \frac{1}{n}\sum_{i=1}^{n}(\hat{g}(x_i) - g(x_i)) \qquad (18)$$

where $\hat{g}(x_i)\epsilon\{\hat{g}_r^*(x), \hat{g}_s^*(x), \hat{g}_{rearr}^*(x)\}$

c) Compare the averaged MSE and $Bias$ over the 1000 replications, for each sample size n.

Table A: Model $g(x) = x$. Averaged MSE and $Bias$ for the $\hat{g}_r^*(x)$, $\hat{g}_s^*(x)$, $\hat{g}_{rearr}^*(x)$ estimators of the conditional mean.

	MSE			*Bias*		
N	$\hat{g}_r^*(x)$	$\hat{g}_s^*(x)$	$\hat{g}_{rearr}^*(x)$	$\hat{g}_r^*(x)$	$\hat{g}_s^*(x)$	$\hat{g}_{rearr}^*(x)$
10	0.0071062	0.0033908	0.0019476	0.0003996	0.0010785	-0.0002623
20	0.0064112	0.0016253	0.0009999	-0.0012291	0.0003488	-0.0000716
30	0.0059819	0.0011350	0.0006869	-0.0024229	0.0004659	-0.0008236
50	0.0056175	0.0009196	0.0004199	-0.0016280	-0.0000701	0.0003146
100	0.0053281	0.0007542	0.0001977	-0.0011094	0.0000646	0.0000582
200	0.0051531	0.0006201	0.0000984	-0.0008347	-0.0001930	-0.0001368

Table B: Model $g(x) = \sqrt{x}$. Averaged MSE and $Bias$ for the $\hat{g}_r^*(x)$, $\hat{g}_s^*(x)$, $\hat{g}_{rearr}^*(x)$ estimators of the conditional mean

	MSE			Bias		
n	$\hat{g}_r^*(x)$	$\hat{g}_s^*(x)$	$\hat{g}_{rearr}^*(x)$	$\hat{g}_r^*(x)$	$\hat{g}_s^*(x)$	$\hat{g}_{rearr}^*(x)$
10	0.0069589	0.0028722	0.0037107	0.0002577	0.0010681	0.0006360
20	0.0060667	0.0014676	0.0029485	0.0002866	-0.0003633	-0.0003750
30	0.0056609	0.0011707	0.0026401	-0.0013210	0.0000956	0.0001269
50	0.0053708	0.0009417	0.0024789	0.0000805	-0.0004291	-0.0004736
100	0.0052085	0.0007365	0.0023909	0.0001741	0.0000987	-0.0002787
200	0.0051078	0.0005773	0.0022791	0.0000662	0.0000942	0.0001721

Table C: Model $g(x) = exp(-5 + 10x)/(1 + exp(-5 + 10x))$. Averaged MSE and $Bias$ for the $\hat{g}_r^*(x)$, $\hat{g}_s^*(x)$, $\hat{g}_{rearr}^*(x)$ estimators of the conditional mean.

	MSE			Bias		
n	$\hat{g}_r^*(x)$	$\hat{g}_s^*(x)$	$\hat{g}_{rearr}^*(x)$	$\hat{g}_r^*(x)$	$\hat{g}_s^*(x)$	$\hat{g}_{rearr}^*(x)$
10	0.0140371	0.0114539	0.0080644	-0.0096964	0.0001058	0.0001285
20	0.0080728	0.0076476	0.0080416	-0.0018572	-0.0007189	0.0002731
30	0.0039777	0.0025782	0.0079643	-0.0001857	0.0004105	-0.0007188
50	0.0061333	0.0009160	0.0078653	0.0011354	0.0001406	-0.0006060
100	0.0057008	0.0006304	0.0078629	-0.0000828	0.0001089	0.0001308
200	0.0053320	0.0005269	0.0078231	0.0007071	-0.0002570	-0.0001597

5. Conclusions

In this paper, we studied and compared the performance of the monotone estimators: regression B-spline estimator $\hat{g}_r^*(x)$, the smoothing B-spline estimator $\hat{g}_s^*(x)$ and the rearrangement estimator $\hat{g}_{rearr}^*(x)$ of the conditional mean in terms of the Mean Squared Error (MSE) and the $Bias$. The results are displayed in Tables A, B and C. We can conclude from these tables that the rearrangement estimator $\hat{g}_{rearr}^*(x)$ performs better in terms of MSE followed by the smoothing spline estimator $\hat{g}_s^*(x)$ when the regression function is linear monotone, whereas the $\hat{g}_s^*(x)$ estimator performs better in terms of MSE compared with other estimators among other scenarios. In addition, the bias of all monotone estimators is too small (negligible), and hence most of the error comes from the variance. As a future work, I will extend my work for the problem of estimating regression curves using regression and penalized cubic splines under single and multiple shape constraints.

6. References

[1] Chernozhukov, V., Fernandez-Val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement, Biometrika, 96(3), 559-575.

[2] Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions, Numerische Mathematik, 31(4), 377-403.

[3] Daouia, A., Noh, H. and Park, B.U. (2016). Data envelope fitting with constrained polynomial splines, J. Royal Statist. Soc. - Series B, 78(1), 3-30.

[4] De Boor, C. (2001). A practical guide to splines, Springer.

[5] Dette, H. and Scheder, R. (2006). Strictly monotone and smooth nonparametric regression for two or more variables, Canad. J. Statist, 34, 535-561.

[6] Dette, H., N. Neumeyer, and K. F. Pilz (2006). A simple nonparametric estimator of a strictly monotone regression function, Bernoulli, 12(3), 469-490.

[7] Eubank, R. L. (1988). Spline Smoothing and Nonparametric regression, New York, and Basil, New York, Marcel Dekker.

[8] Eubank, R. L. (1999). Nonparametric Regression and Spline Smoothing, Marcel Dekkar, New York.

[9] Friedman, J. H. and Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling (with discussion). Technometrics, 31, 3-21.

[10] Gasser, Th. and Muller, H. G. (1979). Kernel estimation of regression functions.In Smoothing Techniques for Curve Estimation, Springer-Verlag, New York, 23-68.

[11] Green, P. J. and Silverman, B. W. (1994). Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Chapman Hall, London.

[12] Hardle, W. (1990). Applied Nonparametric Regression, Cambridge: Cambridge University Press.

[13] Hastie, T. and Tibshirani, R. (1990). Generalized additive model, Chapman and Hall.

[14] He, X. and Shi, P. (1994). B-spline estimates of conditional quantile function, Nonparametric Statistics, 3, 299-308.

[15] He, X. and Shi, P. (1998). Monotone B-spline smoothing. Journal of the American Statistical Association, 93(442), 643-650.

[16] Huang, J. Z. (2003). Local Asymptotics For Polynomial Spline Regression, The Annals of Statistics, 31(5), 1600-1635.

[17] Kelly, C. and Rice, J. (1990). Monotone smoothing with application to dose response curves and the assessment of synergism, Biometrics, 46, 1071-1085.

[18] Lee, S., Linton, O., and Whang, Y.J. (2009). Testing for stochastic Monotonicity, Econometrica, 77(2), 585-602.

[19] Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines, Biometrika, 94, 705-718.

[20] Mammen, E. (1991). Estimating a smooth monotone regression function, The Annals of Statistics, 19, 724-740.

[21] Mammen, E. and Thomas-Agnan, C. (1999). Smoothing splines and shape restrictions, Scand. J. Statist., 26, 239-252.

[22] Meyer, M. C. (2008). Inference using shape-restricted regression splines, The Annals of Applied Statistics, 2(3), 1013-1033.

[23] Ng, P. (1996). An algorithm for quantile smoothing splines, Computational Statistics and Data Analysis, 22, 99-118.

[24] Pal, J. K. and Woodroofe, M. (2007). Large sample properties of shape restricted regression estimators with smoothness adjustments, Statist. Sinica, 17, 1601-1616.

[25] Pya, N. and Wood, S.N. (2015). Shape constrained additive models, Statistical Computing, 25(3), 543-559.

[26] Racine, J. S. A Primer On Regression Splines.

[27] Racine, J., Nie, Z. and Ripley, B. D. (2015). crs package: Categorical Regression Splines, URL: http://github.com/JefferyRacine/R-Package-crs/.

[28] Ramsay, J. O. (1988). Monotone regression splines in action, Statist. Sci., 425-461.

[29] Ruppert, D. (2002). Selecting the number of knots for penalized splines, J. Comp. Graph. Statist., 11, 735-757.

[30] Schumaker, L.L. (2007). Spline Functions: Basic Theory, 3rd edition, Cambridge University Press.

[31] Stone, C. J. (1985). Additive regression and other nonparametric models, Ann. Statist., 13, 689-705.

[32] Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion), Ann. Statist., 22, 118-184.

[33] Wahba, G. and Wold, S. (1975). A completely automatic French curve: fitting spline functions by cross-validation. Communications in Statistics, 4, 1-17.

[34] Wahba, G. (1985). A Comparison of GCV and GML for Choosing the Smoothness Parameter in the Generalized Spline Smoothing Problem. The Annals of Statistics, 13, 1378-1402.

[35] Wahba, G. (1990). Spline models for observational data, SIAM, Philadelphia PA.

[36] Wang, J. and Yang, L. (2009). Polynomial Spline Confidence Bands For Regression Curves, Statistica Sinica, 19, 325-342.

[37] Wright, I. W. and Wegman, E. J. (1980). Isotonic, convex and related splines, Ann. Statist., 8, 1023-1035.