Review Article

The Estimation of Smoothing Parameter using Smoothing Techniques on Nonparametric Regression

Autcha Araveepoorn*

Department of Applied Statistics, Faculty of Science,
King Monkut's Institute of Technology Ladkrabang, Thailand
**Corresponding author. E-mail address: kaautcha@kmitl.ac.th*

Received April 22, 2011; Accepted September 12, 2011

Abstract

This article discusses on the smoothing parameter which is controlled by interpolating spline based on the smoothing techniques that consisted of smoothing spline method, kernel regression method, and penalized spline regression method.

The smoothing parameter is controlled the fitting model and the trade of between the bias of the estimator. We also propose the range of smoothing parameter of these methods to fit the smoothing function which data is nonlinear. Therefore, we mention the characteristic of smoothing function when the smoothing parameters have the various values. According to the results, it is concluded that the smoothing parameter of the smoothing spline method is suitable worked between zero to one, the kernel regression is good performance between two to ten, and the penalized spline is useful between one to ten.

Key Words : Smoothing Parameter; Smoothing Technique; Smoothing Spline; Kernel Regression; Penalized Spline Regression

Introduction

The analysis of available explanatory variables has an application in regression function. The parametric and nonparametric method are the choices for estimating regression function between two variables that consisted of predictor variables and a response variable. A parametric regression model requires an assumption that the form of the underlying regression function such as linearity, stationary variance, and independence of explanatory variables. The selection of parametric model depends much on the problem and may be too restrictive in some applications. If an inappropriate parametric model is used, it is possible to produce misleading conclusions. In other situation, a parametric model may not be available to use. To overcome the difficulty caused by the restrictive assumption of the parametric form of the regression function, one may remove the restriction that the regression function belongs to a parametric family. This approach leads to so-called nonparametric regression.

Typically, the nonparametric regression methods are based on a smoothing technique which produces a smoother. A smoother is a tool for summarizing the trend of a response variable as a function of one or more predictor variables.

The single predictor case is called scatterplot smoothing that can be used to enhance the visual appearance of the scatterplot of response versus predictor variable, to help our eyes pick out the trend in the plot (Hastie and Tibshirani,1990). There are many smoothing techniques, e.g., a smoothing spline (Wahba, 1990, Green and Silverman, 1994), a kernel regression (Wand and Jones, 1995, Fan and Gijbels, 1996) , and a penalized spline regression (Ruppert, *et al.,* 2003).

The aim of smoothing techniques is to estimate smoothing estimators or smoothers that controlled by smoothing parameter. There are many methods that can be used to estimate the smoothing parameter. However, different methods of smoothing techniques have different methods of smoothing parameter, if we know the range of smoothing parameter that can be helped user for nonparametric regression analysis.

In this article, we consider the nonparametric regression in Section 2 and apply in smoothing technique methods as a smoothing spline method, a kernel regression, and penalized spline regression in Section 3. In the Section 4, we show the variation of smoothing parameter in example data and conclude in Section 5.

The Nonparametric Regression

The simple nonparametric regression functions written as

$$y_t = f(x_t) + \varepsilon_t \quad ,t = 1,2,...,n \qquad (1)$$

where $x_t, t = 1,2,...,n$ are known the predictor at the time points, $y_t, t = 1,2,...,n$ are the responses at the time points, $f(x_t)$ are the nonparametric regression function that we want to estimate, and $\varepsilon_t, t = 1,2,...,n$ denote the measurement errors.

Smoothing Techniques

The nonparametric regression can be applied the smoothing techniques to fit the nonparametric regression function by using smoothing spline method, kernel regression method, and penalized spline regression method.

The Smoothing Spline Method

Wahba (1990) defined the natural polynomial spline $s(x) = s_n^m(x)$ is a real-valued function on $[a,b]$ with the aid of n so-called knots ; $-\infty \le a < x_1 < ... < x_n < b < \infty$. The class of m -order spline with domain $[a,b]$ will be denoted by $W^m[a,b]$.

The natural measure associated with the function $f \in W^m[a,b]$ that used to measure the roughness of curve which is called the quadratic penalty function given by

$$\int_a^b \left\{ f^{(m)}(x) \right\}^2 dx \qquad (2)$$

Consider the simple nonparametric regression model where the observation y_t at design points $x_t, t = 1,2,...,n$ assumed to satisfy

$$y_t = f(x_t) + \varepsilon_t \quad ,t = 1,2,...,n \qquad (3)$$

where $f(.)$ denotes a smooth function. To estimate $\hat{f}_\lambda(.)$ minimizes $s_n^{(m)}(f)$ over the class of function $f(.)$ following

$$s_n^{(m)}(f) = \sum_{t=1}^{n} \left\{ y_t - f(x_t) \right\}^2 + \lambda \int_a^b \left\{ f^{(m)}(x) \right\}^2 dx \qquad (4)$$

where $\lambda > 0$ denotes a smoothing parameter to be determined by a suitable cross-validation criteria or information criteria. The smoothing parameter controls the trade-off between fidelity to the data and roughness of function, if $\lambda \to \infty$, the $\hat{f}_\lambda(.)$ converges to linear function, if $\lambda \to 0$, the $\hat{f}_\lambda(.)$ converges to interpolating spline.

In this study, we emphasize $m = 2$ so-called the natural cubic spline which is commonly considered in the statistical literature (see Green and Silverman, 1994). We use this class of cubic smoothing spline to fit $\hat{f}_\lambda(.)$ by starting with the simple nonparametric regression model.

The first procedure of smoothing spline is considered a least square problem to fit a function $\hat{f}_\lambda(.)$ that minimizes the residuals sum of squares

$$RSS = \sum_{t=1}^{n} \{ y_t - f(x_t) \}^2 \quad (5)$$

Assuming the range $f(.)$ in (3) is finite interval, $[a,b] = [x_{(1)}, x_{(n)}]$, where $x_{(1)}$ denotes the i th order statistic and the roughness penalty of $f(.)$ is measured by

$$\int_a^b \{ f''(x) \}^2 \, dx \quad (6)$$

This leads to the following penalized least squares regression to find $\hat{f}_\lambda(x_t)$ called the smoothing spline estimator by minimizing

$$\hat{f}_\lambda(x_t) = \arg \min_{\mu} S_\lambda(f) \quad (7)$$

and

$$S_\lambda(f) = \sum_{t=1}^{n} \{ y_t - f(x_t) \}^2 + \lambda \int_a^b \{ f''(x) \}^2 \, dx \quad (8)$$

where $\lambda > 0$ is a smoothing parameter controlling the size of the roughness penalty and used to trade-off the goodness of fit.

In practice, this step can be implemented by using the function "smooth.spline" in the R software.

Kernel Smoothing

Nadaraya (1964) and Watson (1964) suggested the decomposition in estimating smoothing function for nonparametric regression . The smoothing estimator is written as

$$\hat{f}(x) = \frac{\sum_{t=1}^{n} K\left(\frac{x_t - x}{\lambda}\right) y_t}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{\lambda}\right)} \quad (9)$$

$$= \sum_{t=1}^{n} w_t y_t$$

where λ is known as the bandwidth parameter or called the smoothing parameter which is controlled the smoothness of the estimated curve and the kernel weights are given by,

$$w_t = \frac{\sum_{t=1}^{n} \left(\frac{x_t - x}{\lambda}\right)}{\sum_{t=1}^{n} K\left(\frac{x_t - x}{\lambda}\right)}$$

The smoothing parameter can be chosen using a cross-validation criteria. The kernel functions K can be chosen to several kernel functions that are commonly used : Gaussian, uniform, triangle, Epanechnikov, quartic (biweight), and tricube (triweight). Some popular kernel functions are shown in the Table 1.

Table 1 The kernel density functions

Kernel	Kernel Density Function
Gaussian	$K(u)=(2\pi)^{-1/2}e^{-u^2/2}, \quad u\in[-\infty,\infty]$
Uniform	$K(u)=\frac{1}{2}, \quad u\in[-1,1]$
Triangular	$K(u)=1-\lvert u\rvert, \quad u\in[-1,1]$
Epanechnikov	$K(u)=\frac{3}{2}(1-u^2), \quad u\in[-1,1]$
Quartic	$K(u)=\frac{15}{16}(1-u^2)^2, \quad u\in[-1,1]$
Tricube	$K(u)=\frac{35}{32}(1-u^2)^3, \quad u\in[-1,1]$

The selection of the kernel function is not critical for the performance of regression function. However, it will be used the simplified Gaussian kernel that the smoothing estimator can be written in Gaussian density function by

$$\hat{f}_\lambda(x)=\frac{\sum_{t=1}^{n}\frac{1}{2\pi}\exp-\frac{1}{2}\left(\frac{x_t-x}{\lambda}\right)^2 y_t}{\sum_{t=1}^{n}\frac{1}{2\pi}\exp-\frac{1}{2}\left(\frac{x_t-x}{\lambda}\right)^2} \quad (10)$$

In this case, it is used the "ksmooth" function in the R software perform kernel regression estimator.

Penalized Spline Regression

Eubank(1988, 1999) introduced the regression spline that the local neighborhoods are specified by a group of locations:

$$\tau_{0.},\tau_1,\tau_2,\ldots,\tau_K,\tau_{K+1} \quad (11)$$

in the range of interval $[a,b]$, where $a=\tau_0<\tau_1<....<\tau_K<\tau_{K+1}=b$. These locations are known as knots, and $\tau_r, r=1,2,\ldots,K$ are called interior knots.

A regression spline can be constructed using the k-th degree truncated power basis with K knots $\tau_1,\tau_2,\ldots,\tau_K$:

$$1,x,\ldots,x^k,(x-\tau_1)_+^k,\ldots,(x-\tau_K)_+^k \quad (12)$$

where w_+^k denotes k-th power of the positive part of w or written as $w_+^k=\{\max(0,w)\}^k$. The first $(k+1)$ basis functions of the truncated power basis (12) are polynomials of degree up to k, and the others are all the truncated power functions of degree k. Conventionally, the truncated power basis of degree "k= 0,1,2, and 3" is denoted "constant, linear, quadratic" and "cubic" truncated power basis, respectively.

Using the truncated power basis in (12), a regression spline can be expressed as

$$f(x)=\sum_{s=0}^{k}\beta_s x^s+\sum_{r=1}^{K}\beta_{k+r}(x-\tau_r)_+^k \quad (13)$$

where $\beta_0,\beta_1,\ldots,\beta_{k+K}$ are the unknown coefficients to be estimated by a suitable loss minization.

The penalized spline is a method to estimate an unknown smooth function using the truncated power function (Ruppert and Carroll (2000)), and the penalized spline can be expressed as

$$f(x_t) = \sum_{j=0}^{m-1} \alpha_j x_t^j + \sum_{k=1}^{K} \beta_k \left| x_t - \tau_k \right|^{2m-1} \quad (14)$$

where $\alpha = [\alpha_0, ..., \alpha_{m-1}]^T$ is the vector of the coefficients under the truncated power function, $\beta = [\beta_1, ..., \beta_K]^T \sim N(0, \sigma_\beta^2 \Omega^{-1/2} (\Omega^{1/2})^{T})$, and the (l,k)th entry of Ω Ω is $\left| \tau_l - \tau_k \right|^{2m-1}$ and only the coefficient of $\left| x_t - \tau_k \right|^{2m-1}$ are penalized so that a reasonably large order K can be used.

In this case, we focus on m=2, or the so-called low-rank thin-plate spline which tends to have very good numerical properties. The low-rank thin-plate spline representation of $f(.)$ is

$$f(x_t, \theta) = \alpha_0 + \alpha_1 x_t + \sum_{k=1}^{K} \beta_k \left| x_t - \tau_k \right|^3 \quad (15)$$

where $\theta = (\alpha_0, \alpha_1, \beta_1, ..., \beta_K)^T$ is the vector of regression coefficients, and $\tau_1 < \tau_2 < ... < \tau_K$ are fixed knots. The number of knots, K can be selected using a cross validation method or information theoretic methods (e.g., BIC or AIC).

This class of penalized spline smoothers , $\hat{f}_\lambda(.)$, may also be expressed in convenient vector form

$$\hat{f}_\lambda = C(C^T C + \lambda^3 D)^{-1} C^T y \quad (16)$$

where

$$C = \begin{bmatrix} 1 & x_t & \left| x_t - \tau_k \right|^3_{1 \le k \le K} \end{bmatrix}_{1 \le t \le n},$$

$$D = \begin{bmatrix} 0_{2\times 2} & 0_{2\times K} \\ 0_{K\times 2} & (\Omega_K^{1/2})^T \Omega_K^{1/2} \end{bmatrix}.$$

The penalized spline smoothers is estimated by using the “SemiPar” function in the R software.

The Example of Data Analysis

In this section, the data are obtained from “Semiparametric Regression” (Ruppert *et al*, 2003). The data frame has 221 observations from a light detection and ranging (LIDAR) experiment which contains the range distance travelled before the light is reflected back to its source and logratio logarithm of the ratio of received light from two laser sources.

Let y_t denote the logratio logarithm of the range distance travelled with t where $t = 390, ..., 720$ shown the scatterplot of LIDAR data in Figure 1

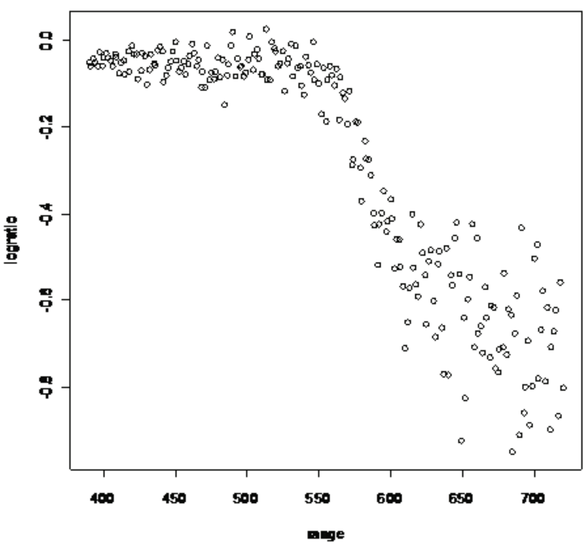

Figure 1 The scatterplot of LIDAR data

From Figure 2-4 show the choice of smoothing parameter that profound influence on the fit using smoothing spline method, kernel regression method, and penalized spline regression method. When smoothing parameter is changed, the fitting lines have an effect on quality of the smoothing. If the smoothing parameter is too small, the fitting line interpolates between observed data. If the smoothing parameter is so large, the fitting line converges to linear function.

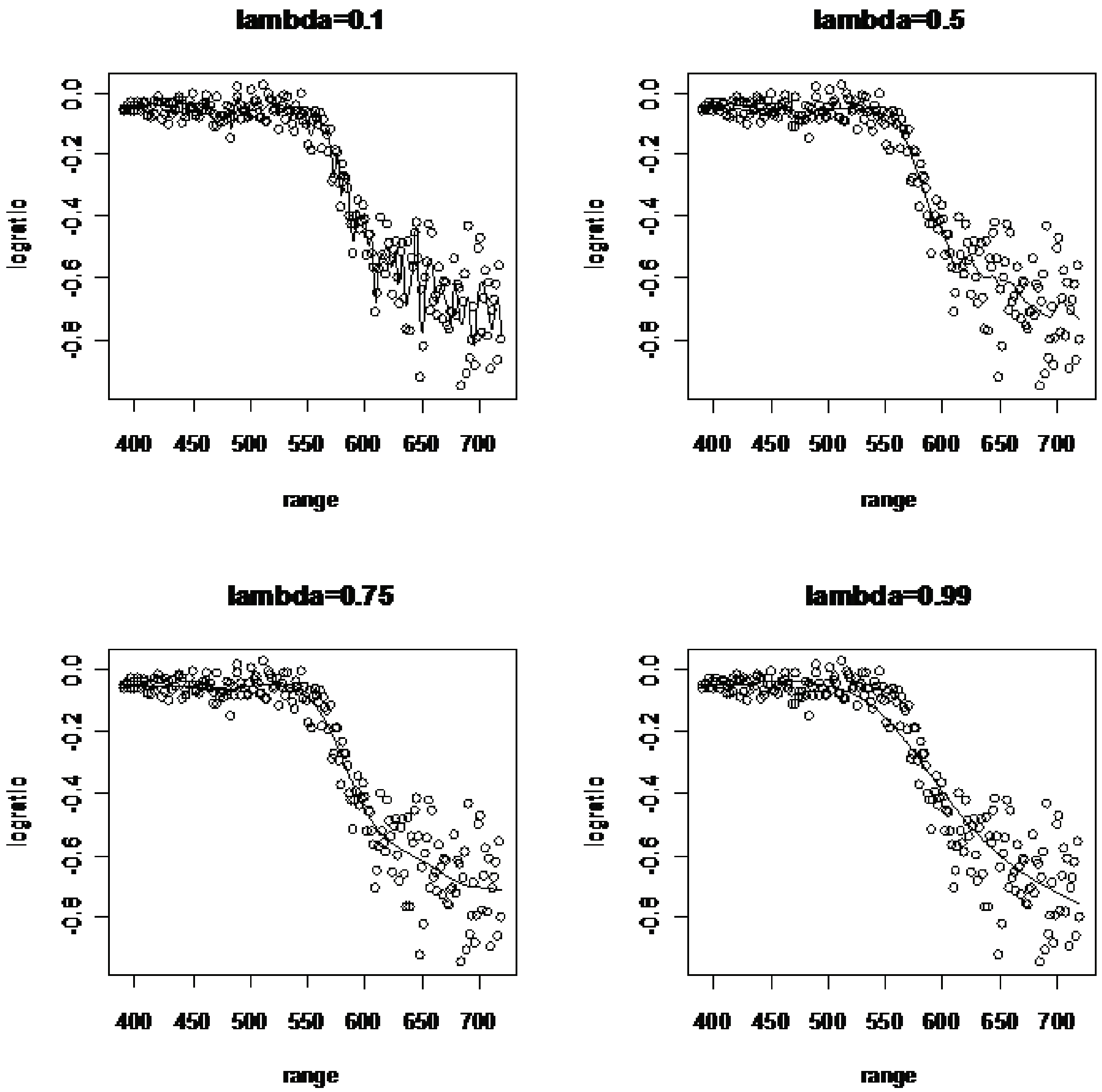

Figure 2 The scatterplot of LIDAR data from smoothing spline method (smoothing parameter or λ = 0.1, 0.5, 0.75, and 0.99)

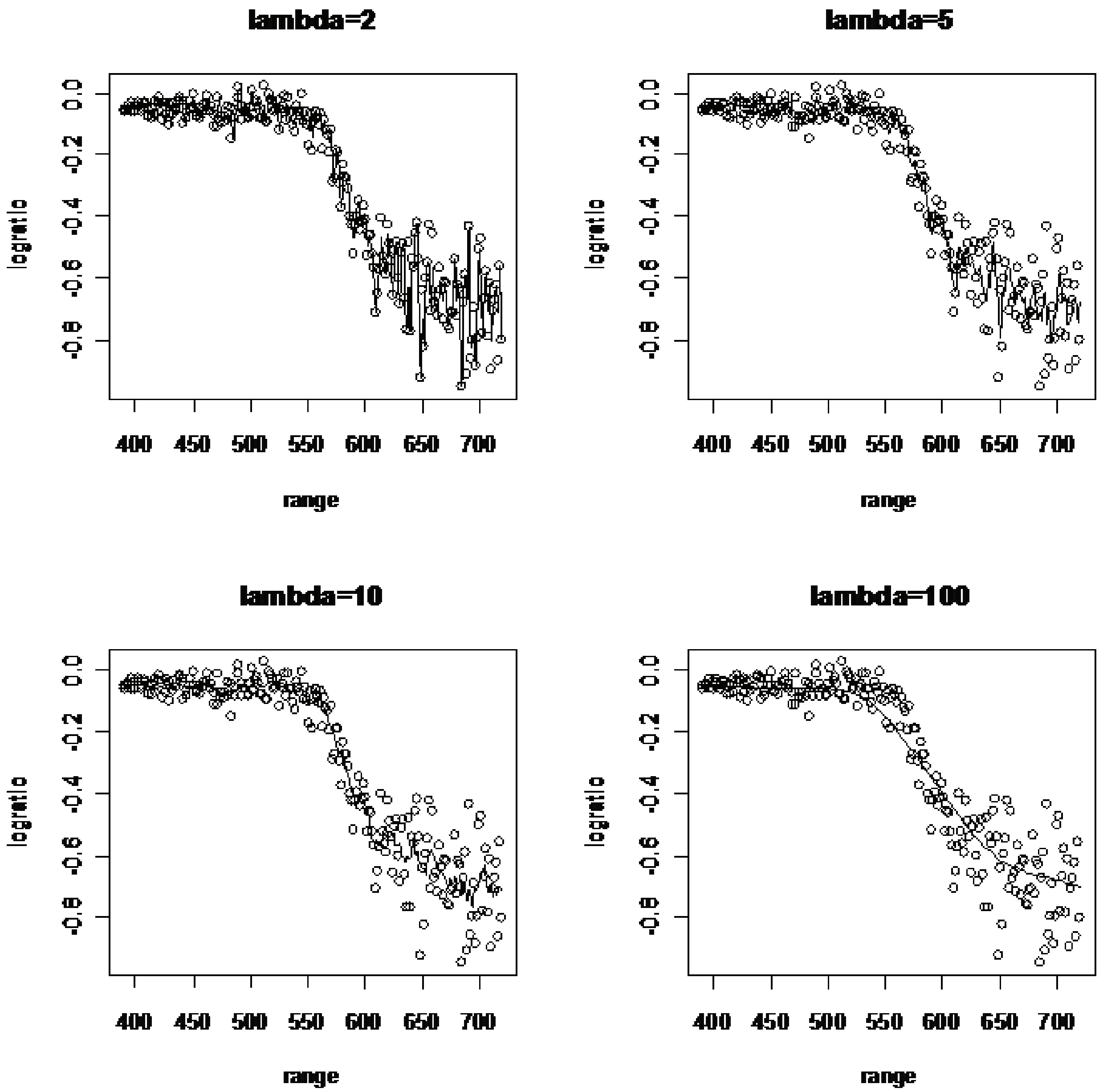

Figure 3 The scatterplot of LIDAR data from kernel regression method (smoothing parameter or λ = 2, 5, 10, and 100)

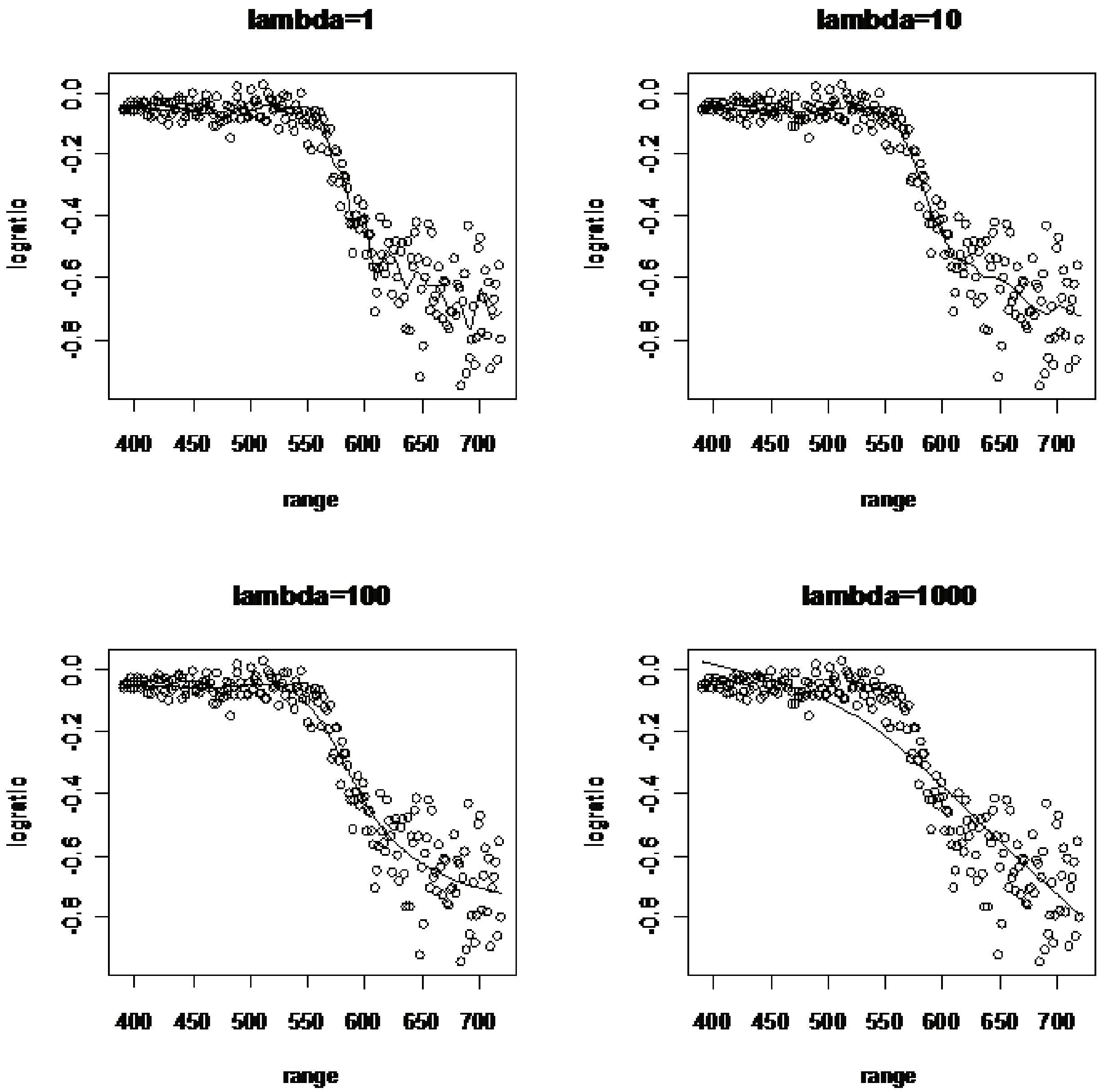

Figure 4 The scatterplot of LIDAR data from penalized spline regression method (smoothing parameter or λ = 1, 10, 100, and 1000)

The performance of the smoothing techniques is related about how close are the estimated values and the observed values. The Mean Square Error (MSE) is the criterion that are used to compare the performances of smoothing spline method, kernel regression, and penalized spline regression. This criterion is defined as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$$

The MSE of these methods are given in Table 2.

Table 2 The MSE values and the smoothing parameter (λ) of the smoothing techniques

Smoothing spline		Kernel regression		Penalized spline	
$\lambda = 0.1$	0.0037	$\lambda = 2$	0.0016	$\lambda = 1$	0.0053
$\lambda = 0.5$	0.0055	$\lambda = 5$	0.0050	$\lambda = 10$	0.0057
$\lambda = 0.75$	0.0059	$\lambda = 10$	0.0054	$\lambda = 100$	0.0062
$\lambda = 0.99$	0.0073	$\lambda = 100$	0.0070	$\lambda = 1000$	0.0109

From the Table 2, it is appeared out that Mean Square Error (MSE) of smoothing spline method, kernel regression method, and penalized spline regression method is slightly different where the smoothing parameters have been changed. The smoothing parameter of smoothing spline performs significantly between zero to one ($0 < \lambda < 1$). The MSE of the kernel regression method is smaller than the MSEs of the other methods especially when smoothing parameter is more than one ($\lambda > 1$) but it should not be more than 10. The panelized spline regression method provides the slightly different in all smoothing parameter except when the smoothing parameter is so large. However, there are no criterion to define the value of smoothing parameter. The range of smoothing parameter can be considered for estimating smoothers.

Conclusion

We have been discussed the nonparametric regression based on the smoothing spline, kernel regression, and penalized spline regression that controlled by smoothing parameter. It is concluded that the smoothing parameter is different depended on the process of these methods. Therefore the estimation of the range of smoothing parameter has indicated a good performance by considering the MSE when the function in R program is worked reasonably when setup the values of smoothing parameter.

References

Eubank, R. L. (1988) *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York.

Eubank, R. L. (1999) *Nonparametric Regression and Spline Smoothing.* Marcel Dekker, New York.

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

Fan, J. and Gijbels, I. (1996) *Local polynomial modeling and its applications.* Chapman and Hall, London.

Hastie, T. J. and Tibshirani, R. (1990) *Generalized Additive Models.* Chapman and Hall, London.

Nadaraya, E. A. (1964) On estimating regression. *Theory of Probability and Its Application*, 9, 141-142.

Ruppert, D. and Carroll, R. J. (2000) Spatial-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics*, 42, 205-224.

Ruppert, D., Wand, M. P., and Carroll. R. J. (2003) *Semiparametric Regression.* Cambridge University Press.

Wahba, G. (1990) *Spline Models for Observational Data.* , SIAM, Philadelphia, PA.

Wand, M. P. and Jones, M. C. (1995) *Kernel Smoothing.* Chapman and Hall, London.

Watson, G.S. (1964) Smooth regression analysis. *Sankhya*, Series A, 26, 359-372.