



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Smoothing Parameter Selection Problem in Nonparametric Regression Based on Smoothing Spline: A Simulation Study

¹Dursun Aydin and ²M. Seref Tuzemen

¹Department of Statistics, Mugla University, 48000, Mugla, Turkey

²Department of Industrial Engineering, Anadolu University, 426470, Eskisehir, Turkey

Abstract: This study presented a comparative study of the different smoothing parameter selection methods. The parameter selection problem has been examined in respect to a smoothing spline implementation in predicting nonparametric regression models. For this purpose, a simulation experiment has been carried out by using a program written in MATLAB 6.5. The simulation experiment provides a comparison of the six different selection methods. In this context, 500 replications have been carried out in simulation for sample sets with different sizes. Thus, the empirical performances of the six selection criteria have been investigated and the suitable selection criteria are obtained for an optimum parameter selection.

Key words: Nonparametric regression, smoothing spline, smoothing parameter, selection criteria, cross-validation, generalized cross-validation

INTRODUCTION

Smoothing spline method can be used to estimate the nonparametric regression models. The main goal of this method is to estimate the function that minimizes the penalized residual sum of squares. Basis of the penalized least squares method is to add a penalty term to the residual sum of squares. In the light of this information, the estimation of the unknown function depends on smoothing parameter. Therefore, the determination of an optimum smoothing parameter in the interval $(0, \infty)$ was found to be an underlying complication. In the literature, different selection methods are components of various studies for an appropriate smoothing parameter. Indeed, to a considerable extent (Craven and Wahba, 1979; Hardle, 1991; Hardle *et al.*, 1988; Wahba, 1990; Hurvich *et al.*, 1998; Eubank, 1999; Lee and Solo, 1999; Hastie and Tibshirani, 1990; Schimek, 2000; Cantoni and Ronchetti, 2001; Ruppert *et al.*, 2003; Lee, 2003, 2004; Kou, 2003) supplement on the selection of the smoothing parameter.

The empirical performances of the selection methods used in selection of the smoothing parameter are compared in this paper. Selection methods used in simulation are an improved version of Akaike Information Criterion (AIC_c), Generalized Cross-Validation (GCV), Cross-Validation (CV), Mallows' C_p criterion, risk estimation using classical pilots (RCP) and local risk estimation (LRS). A simulation experiment is achieved to

find out which selection methods are the best in smoothing parameter selection. To throw light on this issue, the samples differing in small and large sizes are secured by means of the above mentioned simulation and moreover, six selection methods are evaluated.

This study is mainly concerned with the selection of smoothing parameter through Monte Carlo simulation. Smoothing parameters play a crucial role in this procedure. Krivobokova and Kauermann (2007) showed that using the REML to estimate smoothing parameter outperforms other methods such as (generalized) CV or Akaike criterion especially when the error correlation structure is misspecified. Krivobokova *et al.* (2008) formulated a hierarchical mixed model to estimate local smoothing parameter to achieve adaptive penalized spline smoothing. Cao *et al.* (2010) discussed different methods of choosing the important smoothing parameter and recommend GCV as the choice for penalized spline smoothing parameter selection for both computational efficiency and accuracy of the functional coefficient regression models. Aydin and Memmedli (2011) recommended GCV and REML as being good smoothing parameter selection criteria for small and medium sized samples.

SMOOTHING SPLINE METHOD

Suppose observed are n pairs of measurements (x_i, y_i) , $i = 1, 2, \dots, n$, satisfying the model:

$$y_i = f(x_i) + \epsilon_i, \quad \alpha < x_1 < \dots < x_n < b \quad (1)$$

where, $f \in C^2 [a, b]$ is an unknown smooth function and $(\epsilon_i)_{i=1}^n$ are normal distributed random errors with zero mean and common variance σ^2 .

The basic aim of the nonparametric regression is to estimate unknown function $f \in C^2 [a, b]$ (the class of all functions f with continuous first and second derivatives) in model 1. Smoothing spline estimate of the f function appears as a solution to the following minimization problem: Find $\hat{f} \in C^2[a, b]$ that minimizes the penalized residual sum of squares:

$$s(f) = \sum_{i=1}^n \{y_i - f(x_i)\}^2 + \lambda \int_a^b \{f''(x)\}^2 dx \quad (2)$$

for pre-specified value $\lambda > 0$. The first term in Eq. 2 denotes the Residual Sum of the Squares (RSS) and it penalizes the lack of fit. The second term which is weighted by λ denotes the roughness penalty and it imposes a penalty on roughness. In other words, it penalizes the curvature of the function f . The λ in Eq. 2 is recognized to be the smoothing parameter. As λ varies from 0 to $+\infty$, the solution varies from interpolation to a linear model. As $\lambda \rightarrow +\infty$, the roughness penalty dominates in (2) and the spline estimate is compelled to be a constant. As $\lambda \rightarrow 0$, the roughness penalty disappears in (2) and the spline estimation interpolates the data. Thus, the smoothing parameter λ plays a key role in controlling the trade-off between the goodness of fit represented by:

$$\sum_{i=1}^n \{y_i - f(x_i)\}^2$$

and smoothness of the estimate measured by:

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

The solution based on smoothing spline for minimum problem in the Eq. 2 is known as a “natural cubic spline” with knots at x_1, \dots, x_n . From this point of view, a special structured spline interpolation which depends on a chosen value λ develops into a suitable approach of function f in model 1. Let $f = (F(x_1), \dots, f(x_n))$ be the vector of values of function f at the knot points x_1, \dots, x_n . The smoothing spline estimate \hat{f}_λ of this vector or the fitted values for data $y = (y_1, \dots, y_n)^T$ are projected by:

$$\hat{f} = \begin{pmatrix} \hat{f}_\lambda(x_1) \\ \hat{f}_\lambda(x_2) \\ \vdots \\ \hat{f}_\lambda(x_n) \end{pmatrix}_{(n \times 1)} = (S_\lambda)_{(n \times n)} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{(n \times 1)} \quad \text{or, in short, } \hat{f}_\lambda = S_\lambda y \quad (3)$$

where, \hat{f}_λ is a natural cubic spline with knots at x_1, \dots, x_n for a fixed $\lambda > 0$ and S_λ is a well-known positive-definite (symmetrical) smoother matrix which depends on λ and the knot points x_1, \dots, x_n but not on y . Function \hat{f}_λ , the estimator of function f is obtained by cubic spline interpolation that rests on condition $\hat{f}(x_i) = (f)_i, i = 1, 2, \dots, n$. To gain better perspective on smoothing spline (Eubank, 1999; Green and Silverman, 1994; Wahba, 1990) state studied opinions.

SMOOTHING PARAMETER SELECTION METHODS

Although, smoothing spline estimator solves the problem of allowing fits with variable slope, a new dilemma emerges. In fact, it generates the determination of the appropriate value for the smoothing parameter λ for a given data set. The same value of λ is unlikely to work equally well with every data set. As such, the estimation methods have been introduced for the selection of smoothing parameter λ in Eq. 2. The positive value λ that minimizes any smoothing parameter selection methods is selected as an appropriate smoothing parameter.

Selection methods used in experiment: Various smoothing parameter selection methods are featured in the literature. Most of these suggested methods were implemented in our simulation study. Moreover, a selection criterion from previous studies in the literature to provide an effective performance was also introduced in this particular study. The selection criteria used in our simulation study are classified as:

Cross-validation: The basic idea of CV is to disregard one of the points $\{x_i, y_i\}_{i=1}^n$ gradually, to select the smoothing parameter λ that minimizes the residual sum of squares and to estimate the squared residual for a smooth function at x_i based on the remaining $(n-1)$ points. The CV score can be translated as:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{f}_\lambda^{(-i)}(x_i)\}^2 \equiv CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{f}_\lambda(x_i)}{1 - (S_\lambda)_{ii}} \right\}^2 \quad (4)$$

where, \hat{f}_λ is the fit (spline smoother) for n pairs of measurements $\{x_i, y_i\}_{i=1}^n$ with smoothing parameter λ and $\hat{f}_\lambda^{(-i)}$ is the fit calculated by leaving out the i th data point and $(S_\lambda)_{ii}$ is the i th diagonal element of smoother matrix S_λ (Wahba, 1990; Green and Silverman, 1994).

Using the approximations $(S_\lambda)_{ii} \approx \{s_\lambda s_\lambda^T\}_{ii}$ and $1/(1 - (S_\lambda)_{ii})^2 \approx 1 + 2(S_\lambda)_{ii}$ signifies that:

$$E\{CV(\lambda)\} \approx PSE(\lambda) + 2/n + \sum_{i=1}^n (S_\lambda)_{ii} b_i^2(\lambda)$$

(Hastie and Tibshirani, 1990).

Generalized cross-validation: GCV is a modified form of the CV which is a conventional method for choosing the smoothing parameter. The GCV score which is constructed by analogy to CV score can be obtained from the ordinary residuals by dividing by the factors $1-(S_\lambda)_{ii}$. The underlying design of GCV is to replace the factors $1-(S_\lambda)_{ii}$ in Eq. 8 with the average score $1-n^{-1} \text{tr}(S_\lambda)$. Thus, by summing the squared corrected residual and factor $1-n^{-1} \text{tr}(S_\lambda)^2$, by the analogy ordinary cross-validation, the GCV score function can be procured as follow (Craven and Wahba, 1979; Wahba, 1990):

$$GCV(\lambda) = \frac{1}{n} \frac{\sum_{i=1}^n \{y_i - \hat{f}_\lambda(x_i)\}^2}{\{1 - n^{-1} \text{tr}(S_\lambda)\}^2} = \frac{n^{-1} \|(I - S_\lambda)y\|^2}{[n^{-1} \text{tr}(I - S_\lambda)]^2} \quad (5)$$

Mallows' C_p criterion: In the literature, C_p criterion is referred to as an unbiased risk estimate (UBR). This type of estimate was suggested by Mallows (1973) in the regression case and applied to smoothing spline by Craven and Wahba (1979). If σ^2 is recognized, an unbiased estimate of the residual sum of squares is provided by C_p criterion:

$$C_p(\lambda) = \frac{1}{n} \{ \|(S_\lambda - I)y\|^2 + 2\sigma^2 \text{tr}(S_\lambda) - \sigma^2 \} = \frac{1}{n} \{ \|y - \hat{f}_\lambda\|^2 + 2\sigma^2 \text{tr}(S_\lambda) - \sigma^2 \} \quad (6)$$

Unless σ^2 is known, in practice an estimation for σ^2 can be given by:

$$\hat{\sigma}_\lambda^2 = \hat{\sigma}_\lambda^2 = \frac{\sum_{i=1}^n (y_i - \hat{f}_\lambda(x_i))^2}{\text{tr}(I - S_\lambda)} = \frac{\|(S_\lambda - I)y\|^2}{\text{tr}(I - S_\lambda)} \quad (7)$$

where, $\hat{\lambda}$ is pre-chosen with any of the CV, GCV or AIC_c criteria ($\hat{\lambda}$ is an estimate of λ) (Lee, 2003, 2004; Wahba, 1990). According to Hastie and Tibshirani (1990) and Ruppert *et al.* (2003), GCV is approximately equal to C_p .

Improved Akaike information criterion: An improved version of a criterion based on the classical Akaike Information Criterion (AIC), AIC_c criterion, is used for choosing the smoothing parameter for nonparametric smoothers (Hurvich *et al.*, 1998). This improved criterion is defined as:

$$AIC_c(\lambda) = \log \frac{\sum_{i=1}^n \{y_i - \hat{f}_\lambda(x_i)\}^2}{n} + 1 + \frac{2\{\text{tr}(S_\lambda) + 1\}}{n - \text{tr}(S_\lambda) - 2} = \log \frac{\|(S_\lambda - I)y\|^2}{n} + 1 + \frac{2\{\text{tr}(S_\lambda) + 1\}}{n - \text{tr}(S_\lambda) - 2} \quad (8)$$

As can be seen from the Eq. 8, this criterion is easy to apply for the selection of smoothing parameter.

Risk estimation using classical pilots: Risk function measures the distance between the actual regression function (f) and its estimation (\hat{f}_λ). Needless to say that, a good estimate must contain minimum risk. A direct computation leads to the bias-variance decomposition for $R(f, \hat{f}_\lambda)$:

$$R(f, \hat{f}_\lambda) = \frac{1}{n} E \|\hat{f}_\lambda - f\|^2 = \frac{1}{n} \{ \|(S_\lambda - I)f\|^2 + \sigma^2 \text{tr}(S_\lambda S_\lambda^T) \} \quad (9)$$

A clear-cut explanation shows that $R(f, \hat{f}_\lambda) = E \{ C_p(\lambda) \}$. Because the risk $R(f, \hat{f}_\lambda)$ is an unknown quantity, so-called risk is now estimated by computable quantity $R(f, \hat{f}_\lambda)$. The obtained expression for $R(f, \hat{f}_\lambda)$ is:

$$R(\hat{f}_{\lambda_p}, \hat{f}_\lambda) = \frac{1}{n} E \|\hat{f}_{\lambda_p} - \hat{f}_\lambda\|^2 = \frac{1}{n} \{ \|(S_\lambda - I)\hat{f}_{\lambda_p}\|^2 + \hat{\sigma}_{\lambda_p}^2 \text{tr}(S_\lambda S_\lambda^T) \} \quad (10)$$

where, $\hat{\sigma}_{\lambda_p}^2$ and \hat{f}_{λ_p} are the appropriate pilot estimates for σ^2 and f , respectively. The pilot λ_p selected by classical methods is used for computation of the pilot estimates.

Local risk estimation: The LRS method proposed by Lee (2004), aims to select the $\hat{f}_\lambda(x_i)$ that minimizes the local risk $R_\lambda(x_i) = E \{ f(x_i) - \hat{f}_\lambda(x_i) \}^2$ for the each knot points x_i . A direct computation leads to the bias-variance decomposition for $R_\lambda(x_i)$:

$$R_\lambda(x_i) = \{ (S_\lambda f)(x_i) - f(x_i) \}^2 + \sigma^2 s_\lambda(x_i) \quad (11)$$

In the above equation, $(S_\lambda f)(x_i)$ is the i th element of vector $S_\lambda f$ and $S_\lambda(x_i)$ is the i th diagonal element of the square matrix $s_\lambda S_\lambda^T$. An estimator for $R_\lambda(x_i)$ is firstly computed and the $\hat{f}_\lambda(x_i)$ is selected in order to minimize it. This process is repeated for all x_i 's and at the end of the process a final mixed estimate for f is derived. The LRS method can be practically performed with the following five steps (Lee, 2003):

- For a set of pre-selected smoothing parameters $\lambda_1 < \dots < \lambda_m$, calculate the corresponding set of smoothing spline estimates: $F = \{ \hat{f}_{\lambda_1}, \dots, \hat{f}_{\lambda_m} \}$
- Select the pilot value λ_p from AIC_c criterion in Eq. 8 by using the elements in F
- For λ_p , calculate the estimates \hat{f}_{λ_p} and $\hat{\sigma}_{\lambda_p}^2$ by using the Eq. 7
- Substitute the pilots \hat{f}_{λ_p} and $\hat{\sigma}_{\lambda_p}^2$ into the expression $\hat{R}_\lambda(x_i) = \{ (S_\lambda \hat{f}_{\lambda_p})(x_i) - \hat{f}_{\lambda_p}(x_i) \}^2 + \hat{\sigma}_{\lambda_p}^2 s_\lambda(x_i)$ and obtain the estimates $\hat{R}_\lambda(x_i)$
- For each x_i , find the $\hat{f}_\lambda(x_i)$ from F which minimizes $\hat{R}_\lambda(x_i)$ and the final estimate accept the appropriate values $\hat{f}_\lambda(x_i)$ for $f(x_i)$

SIMULATION EXPERIMENT

This study was conducted to evaluate the performances of the nine selection methods mentioned above. The experimental setup in this paper is adopted from Professor Steve Marron. By using a program coded in MATLAB, we generated the samples sized $n = 25, 50, 100, 150, 200, 350$ and 400 . The number of replication was 500 for each of the samples. For each simulated data sets, the Mean Squared-Errors (MSE) was used for evaluate the quality of any curve estimate \hat{f} . To find out if the difference between the MSE median values of any two selection methods is significant or not, the paired Wilcoxon tests were assessed. In this way, methods which complement the best smoothing parameter were determined by evaluating so-called selection methods.

Experimental setup: The experimental setup applied at this stage was designed to study the effects of the following three factors which vary an independent and effective approach: noise level, spatial variation, variance function.

The setup specification is listed Table 1. Simulation study was performed according to MATLAB program and the experimental setup was designed in the following framework:

- Totally three sets of numerical experiments were performed. For each set of experiments, only one of the above three experimental factors (e.g., noise level, degree of spatial variation and noise variance function) has been altered while the remaining two have been kept unchanged
- Within each set of experiments, the factor levels was modified four times (i.e., $r = 1, 2, 3, 4$) to detect the effects of any experimental factor in Table 1
- To see the performance of the small and large samples of the selection methods. For each factor level r within each set of experiments, we generated 7 different samples with sample sizes $n = 25, 50, 100, 150, 200, 350$ and 400
- The number of replications was 500 for each of the 84 numerical experiments

- We computed the appropriate smoothing spline estimators \hat{f}_λ in equality (3) by selecting the smoothing parameter λ which minimizes the selection methods
- We used the MSE values to evaluate \hat{f}_λ computed according to each of the selection criterion:

$$MSE = \frac{1}{n} \sum_{i=1}^n \{f(x_i) - \hat{f}_\lambda(x_i)\}^2$$

- $(\hat{f}_\lambda(x_i) = (\hat{f}_\lambda)_i)$. (where $f(x_i)$ is value at knots x_i of the appropriate function f defined in Table 1)
- Paired Wilcoxon test was applied to test whether MSE values was considered as the performance measure of any two methods are significant or not
- The factor levels indicated as r was changed four times (i.e., $r = 1, 2, 3, 4$) in order to detect the effects of any factor from three factors in Table 1
- By considering 3 factors, 4 factor levels and 7 samples, totally, 84 numerical experiments were conducted

Evaluations of the experimental results: For each simulated data set used in the experiments, the MSE values were used in order to evaluate the quality of any curve estimate \hat{f} . Paired Wilcoxon tests were applied to test whether the difference between the median MSE values of any two methods is significant or not. The significance level used was 5%. The selection methods were also ranked as follows: If median MSE value of a method is significantly less than the remaining five, it will be assigned a rank 1. If median MSE value of a method is significantly larger than one but less than the remaining four, it will be assigned a rank 2 and similarly for ranks 3-9. Methods having non-significantly different median values will share the same averaged rank, on the other hand, method or methods having the smallest rank will be superior.

In this simulation study, because totally 84 different configurations are made, it is not possible to display here all these configurations. Therefore, only 12 different configurations are given in Fig. 1-3 for $n = 150$.

Table 1: Specification of the simulation setup

Factors	Generic form	Particular choices
Noise level	$y_{ir} = f(x_i) + \sigma_i \varepsilon_i$	$\sigma_i = 0.02 + 0.04 (r-1)^2, i = 1, \dots, n$
Spatial variation	$y_{ir} = f(x_i) + \sigma \varepsilon_i$	$\sigma_i = 0.02, f_r(x) = \sqrt{x(1-x)} \sin \left[\frac{2\pi \{1 + 2^{(9+4r)/5}\}}{x + 2^{(9+4r)/5}} \right]$
Variance function	$y_{ir} = f(x_i) + \sqrt{v_r(x_i)} \varepsilon_i$	$v_r(x) = [0.15 \{1 + 0.4 (2r-7) (x-0.5)\}]^2$

Where
 $r = 1, \dots, 4; x_i = \frac{i-0.5}{n}, \varepsilon_i \sim \text{iid } N(0,1), f(x) = 1.5\theta \left(\frac{x-0.35}{0.15} \right) - \theta \left(\frac{x-0.8}{0.04} \right); \theta(u) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{u^2}{2} \right)$
 $n = 25, 50, 100, 150, 200, 350$ and 400 (it was taken seven different sample size)

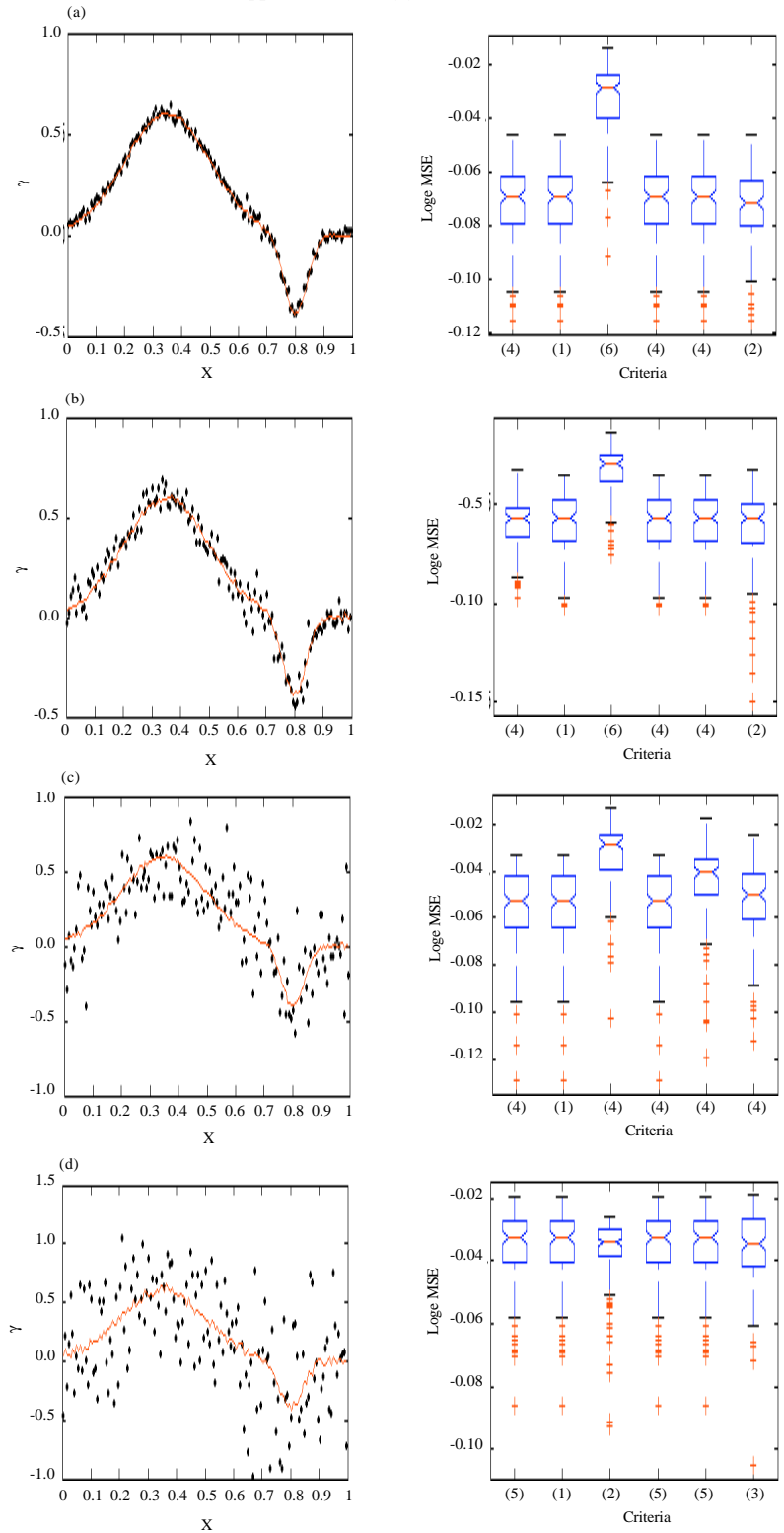


Fig. 1(a-d): Simulation results correspond to the noise level factor for $n = 150$, (a) $1 = 1$, (b) $1 = 2$, (c) $1 = 3$ and (d) $1 = 4$

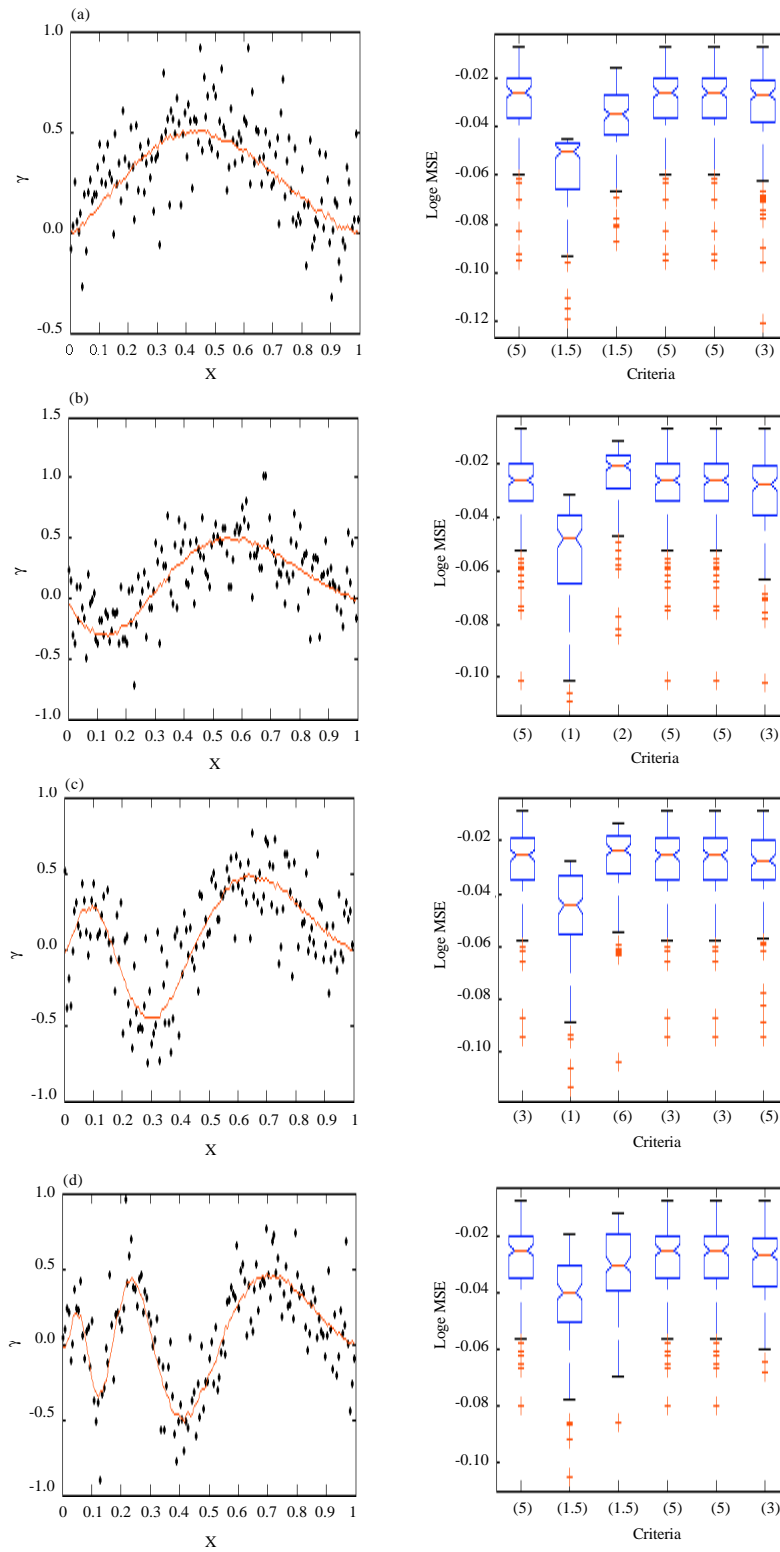


Fig. 2(a-d): Simulation results correspond to the spatial variation factor for $n = 150$, (a) $l = 1$, (b) $l = 2$, (c) $l = 3$ and (d) $l = 4$

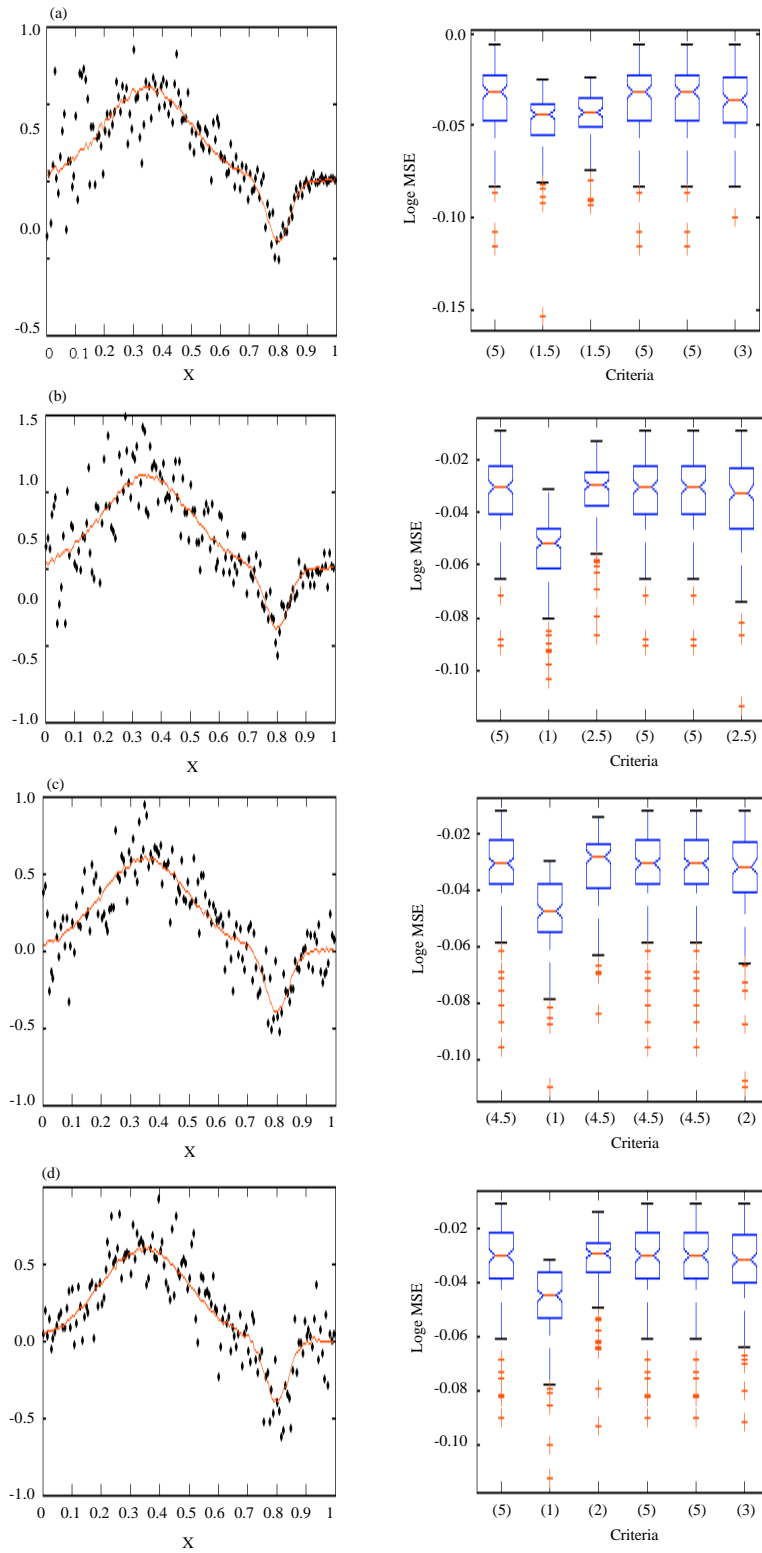


Fig. 3(a-d): Simulation results correspond to the variance factor for $n = 150$, (a) $1 = 1$, (b) $1 = 2$, (c) $1 = 3$ and (d) $1 = 4$

Table 2: Averaged Wilcoxon test ranking values for the six selection methods

Criteria	Noise level	Spatial variation	Variance function	Overall
n = 25				
AIC _c	3.750	4.000	3.500*	3.750
GCV	2.875*	2.375*	3.500*	2.917*
CV	3.625	3.125	3.500*	3.417
C _p	3.750	4.000	3.500*	3.750
RCP	3.750	4.000	3.500*	3.750
LRS	3.250*	3.500	3.500*	3.417
n = 50				
AIC _c	4.250	4.375	4.750	3.750
GCV	2.375*	1.875*	1.500*	2.917*
CV	2.750	3.250	2.250	3.417
C _p	4.250	4.375	4.750	3.750
RCP	4.250	4.375	4.750	3.750
LRS	3.125	2.750	3.000	3.417
n = 100				
AIC _c	4.500	4.750	5.000	4.750
GCV	1.000*	1.375*	1.625*	1.333*
CV	3.250	2.625	2.375	2.750
C _p	4.500	4.750	5.000	4.750
RCP	4.500	4.750	5.000	4.750
LRS	3.250*	2.750	2.750	2.917
n = 150				
AIC _c	4.750	4.500	4.875	4.708
GCV	1.000*	1.250*	1.125*	1.125*
CV	4.500	2.750	2.625	3.292
C _p	4.750	4.500	4.875	4.708
RCP	4.750	4.500	4.875	4.708
LRS	2.750	3.500	2.625	2.958

*Selection method has the best ranking

Table 3: Averaged Wilcoxon test ranking values for the six selection methods in large sample sizes

Criteria	Noise level	Spatial variation	Variance function	Overall
n = 200				
AIC _c	4.000*	3.750*	2.500*	3.417*
GCV	4.000*	3.750*	4.625	4.125
CV	6.250	6.750	6.500	6.500
C _p	4.000*	3.750*	4.625	4.125
RCP	5.000	4.750	6.500	5.417
LRS	4.750	5.625	5.250	5.208
n = 350				
AIC _c	3.875*	3.375*	2.875*	3.375
GCV	3.875*	3.375*	3.750	3.667
CV	7.750	7.250	5.875	6.958
C _p	3.875*	3.375*	3.750	3.667
RCP	5.000	4.625	7.000	5.542
LRS	3.000	4.125	2.875*	3.333*
n = 400				
AIC _c	4.500*	3.375*	3.000*	3.625*
GCV	4.500*	3.375*	3.000*	3.625*
CV	5.750	6.000	5.250	5.667
C _p	4.500*	3.375*	3.000*	3.625*
RCP	5.500	5.000	7.250	5.917
LRS	4.500*	4.750	4.250	4.500

*Selection method has the best ranking

The head row plots of the Fig. 1-3 display the true regression function together with all typical simulated data set. The bottom row plots display the boxplots of the log₁₀ MSE values for, from left to right, AIC_c, GCV, CV, C_p, RCP and LRS. The numbers below the boxplots are the paired Wilcoxon test rankings.

Table 4: Means of the averaged Wilcoxon test ranking values for the six selection methods means for n = 25, 50, 100 and 150

Criteria	Noise level	Spatial variation	Variance function	Overall
AIC _c	4.313	4.500	4.531	4.448
GCV	1.813*	1.719*	1.938*	1.823*
CV	3.531	2.938	2.688	3.052
C _p	4.313	4.500	4.531	4.448
RCP	4.313	4.500	5.531	4.448
LRS	3.344	3.125	2.969	3.146

*Selection method has the best ranking

Table 5: Means of the averaged Wilcoxon test ranking values for the six selection methods means for n = 200, 350 and 400

Criteria	Noise level	Spatial variation	Variance function	Overall
AIC _c	4.125	3.750*	2.792*	3.556*
GCV	4.125	3.750*	3.792	3.889
CV	6.583	6.667	5.875	6.375
C _p	4.125	3.750*	3.792	3.889
RCP	5.167	4.792	6.917	5.625
LRS	4.083*	4.833	4.125	4.347

*Selection method has the best ranking

For 84 different simulation experiments, the averaged ranking values of the selection methods according to Wilcoxon tests are tabulated in Table 2 and 3.

According to the results in Table 2, for small sized samples (for n = 25, 50, 100 and 150), GCV has had the best empirical performance for all factors. In accordance with the overall Wilcoxon test rankings in Table 2, GCV have also displayed a good performance. As shown Table 2, because of $R(\mathbf{r}, \hat{\mathbf{r}}_k) = E\{C_p, (\lambda)\}$, C_p and RCP methods produced the same results under all experimental factors. In this situation, for small sample sizes, for which reason the effects of the replication of simulation, C_p is approximately equal to its E {C_p}. For small samples, it is observed that CV has produced the worst performance.

According to Table 3, for large sized samples (for n = 200, 350 and 400), AIC_c criterion has had the best empirical performance. It is shown that AIC_c and GCV and C_p criteria have shared a better performance for n = 400. According to the overall Wilcoxon test rankings in Table 3, AIC_c, GCV, C_p and LRS criteria can be ranked in terms of the performance. As shown in Table 3, generally, C_p and GCV gave the same results. This can be interpreted to follow the accepted view that GCV is asymptotically equal to GCV (Hastie and Tibshirani, 1990). That is to say that, for large sized samples, because of the effect of the replication of simulation, performance of the GCV and C_p is approximately equal. RCP has also produced the worst performance for large samples.

The scores in Table 4 and 5 are obtained by taking the means of the averaged Wilcoxon test ranking values tie with each of the selection methods in Table 2-3, respectively.

As shown in Table 4, according to the means of the small samples for all factors, GCV criterion has had the

best empirical performance. When it is compared to the other criteria, C_p (RCP), AIC_c criteria have resulted in the worst performance.

According to the means of the large samples for all factors in Table 5, generally it is shown that AIC_c has had the best empirical performance. GCV (C_p) and LRS criteria can be ranked in terms of the performance after than AIC_c . However, CV has produced the worst result.

RECOMMENDATIONS

Finally, by considering the simulation results and evaluations given in the above, the following suggestions have to be taken into account:

- For both large and small samples, GCV are recommended as being the best selection criterion
- For especially large samples, the use AIC_c would seem to be more appropriate. As for small samples, we propose the implementation of GCV criterion
- For large samples, the implementation of AIC_c criterion, in addition to and GCV (C_p) and LRS criteria would be more beneficial. For small samples, GCV criterion should prove fruitful

Naturally, the above recommended suggestions have to be considered with a fair amount of caution as they are only an appraisal based on simulation results.

REFERENCES

Aydin, D. and M. Memmedli, 2011. Optimum smoothing parameter selection for penalized least squares in form of linear mixed effect models. Optimization: J. Math. Program. Oper. Res., 10.1080/02331934.2011.574698.

Cantoni, E. and E. Ronchetti, 2001. Resistant selection of the smoothing parameter for smoothing splines. Stat. Comput., 11: 141-146.

Cao, Y., H. Lin, T.Z. Wu and Y. Yu, 2010. Penalized spline estimation for functional coefficient regression models. Comput. Stat. Data Anal., 54: 891-905.

Craven, P. and G. Wahba, 1979. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. Numer. Math., 31: 377-403.

Eubank, R.L., 1999. Nonparametric Regression and Smoothing Spline. 2nd Edn., Marcel Dekker Inc., New York.

Green, P.J. and B.W. Silverman, 1994. Nonparametric Regression and Generalized Linear Models 1st Edn., Chapman and Hall, London.

Hardle, W., 1991. Applied Nonparametric Regression. 1st Edn., Cambridge University Press, Cambridge.

Hardle, W., P. Hall and J.S. Marron, 1988. How far are automatically chosen regression smoothing parameters from their optimum? J. Am. Stat. Assoc., 83: 86-89.

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

Hurvich, C.M., J.S. Simonoff and C.L. Tsai, 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J. R. Stat. Soc. Ser. B, 60: 271-293.

Kou, S.C., 2003. On the efficiency of selection criteria in spline regression. Probab. Theory Related Fields, 127: 153-176.

Krivobokova, T. and G. Kauermann, 2007. A note on penalized spline smoothing with correlated errors. J. Am. Stat. Assoc., 102: 1328-1337.

Krivobokova, T., C.M. Crainiceanu and G. Kauermann, 2008. Fast adaptive penalized splines. J. Comput. Graphical Stat., 17: 1-20.

Lee, T.C.M. and V. Solo, 1999. Bandwidth selection for local linear regression: A simulation study. Comput. Stat., 14: 515-532.

Lee, T.C.M., 2003. Smoothing parameter selection for smoothing splines: A simulation study. Comput. Stat. Data Anal., 42: 139-148.

Lee, T.C.M., 2004. Improved smoothing spline regression by combining estimates of different smoothness. Stat. Probab. Lett., 67: 133-140.

Mallows, C.L., 1973. Some comments on C_p . Technometrics, 15: 661-676.

Ruppert, D., M.P. Wand and R.J. Carroll, 2003. Semi-Parametric Regression. 1st Edn., Cambridge University Press, Cambridge.

Schimek, M.G., 2000. Smoothing and Regression: Approaches, Computation and Application. John Wiley and Sons Inc., USA., ISBN-13: 9780471179467, pp: 19-39.

Wahba, G., 1990. Spline Models for Observational Data. 1st Edn., SIAM, Philadelphia, Pennsylvania, USA., ISBN: 10: 0898712440, Pages: 180.