An Application of Various Nonparametric Techniques by Nonparametric Regression Splines

Memmedaga Memmedli and Akhlitdin Nizamitdinov

***Abstract*—**In this paper we made a comparison study between regression spline, penalized spline, and their Bayesian versions: adaptive Bayesian regression spline and Bayesian penalized spline with a different number of observations. For this purpose we made a simulation study with four different functions with six positions. For regression and penalized splines the important problems are the knot selection and selection of smoothing parameter. For both techniques we used equidistant knot selection as a basis method in regression techniques. The purpose of using different number of sampled observations is to analyze the behavior of utilized techniques. All results are compared with each other by mean value of the MSE (mean squared error). The penalized spline showed one of the best results between spline techniques and their Bayesian versions.

***Keywords*—**Adaptive Bayesian regression spline, Bayesian P-spline, P-spline, Regression spline, Simulation.

I. Introduction

THE nonparametric regression techniques become powerful tools for analyzing, predicting and forecasting for statisticians in recent years.

There are a lot of methods and techniques in nonparametric regression. There exist a various techniques in nonparametric statistics, like regression splines [8], smoothing splines [4], B-splines [16], [17] and P-splines [2], generalized additive models. All these techniques aimed in one problem: to make a precision prediction.

In this study we used techniques that are in our studied research field, such as regression spline, penalized spline and there Bayesian analogues: adaptive Bayesian regression spline and Bayesian penalized spline. For regression and penalized splines we used the 3rd order basis functions with the second order penalties. The selection of proper method to use in specific task is an important problem.

In this study we made the selection of the regression and Bayesian techniques. For this purpose, we made a simulation study with four different functions with six positions. There become 24 different functions totally when changing positions of the functions. For all these functions we sampled n=50, n=100, n=200 and n=400 number of observations with 200 replications. Changing the position of function gives us the different scattered dataset. For comparison of these techniques we used mean value of MSE for each replication. And we showed results graphically with box plot of MSE.

The main idea of regression splines, penalized splines, Bayesian penalized splines and adaptive regression spline has described in problem formulation section. The simulation study and there result are gathered in problem solution section. Finally, conclusion section concludes the paper.

Prof. Dr. M. Memmedli is with the Statistics Department, Faculty of Science, Anadolu University, Eskisehir, TURKEY. (corresponding author to provide phone: +902223350580-4687; e-mail: mmammadov@anadolu.edu.tr).

A. Nizamitdinov is with the Department of Statistics, Faculty of Science, Anadolu University, Eskisehir, TURKEY. (e-mail: ahlidin@gmail.com).

II. Problem Formulation

Firstly, we can summarize nonparametric regression model.

Nonparametric regression model including a predictor (independent) variable x and a response variable y is defined as

$$y_i = f\left(x_i\right) + \varepsilon_i, \quad a < x_1 < \ldots < x_n < b,$$

where $f \in C^2\left[a,b\right]$ is an unknown smooth function, $\left(y_i\right)_{i=1}^n$ are observation values of the response variable y, $\left(x_i\right)_{i=1}^n$ are observation values of the predictor variable x and $\left(\varepsilon_i\right)_{i=1}^n$ are normal distributed random errors with zero mean and common variance $\sigma^2\left(\varepsilon_i \sim N\left(0,\sigma^2\right)\right)$ [9].

In this section we give short information about the utilized techniques. First, it described regression spline function as a basis function for regression techniques. Regression spline chooses a basis amounts to choosing some basis functions, which will be treated as completely known: if $b_j(x)$ is a j^{th} such basis function, then f is assumed to have a representation

$$f(x) = \sum_{j=1}^{q} \beta_j b_j(x) \tag{1}$$

for some values of the unknown parameters, β_j.

The basic aim of the regression spline is to estimate unknown function $f \in C^2[a,b]$ in model (2).

$$y_i = f(x_i) + \varepsilon_i \quad a < x_1 < ... < x_n < b \tag{2}$$

where $f \in C^2[a,b]$ is an unknown smooth function, $y_i, i = 1,...,n$ are observation values of the response variable y, $x_i, i = 1,...,n$ are observation values of the predictor variable x and $\varepsilon_i, i = 1,...,n$ normal distributed random errors with zero mean and common variance σ^2.

Here, smoothing spline is one of the nonparametric regression methods dealing with how to obtain the unknown function, that can be any arbitrary fuction but comes from specific class of function, by imposing a roughness penalty to objective function [11].

Smoothing spline method is one of the most popular methods used for the prediction of the nonparametric regression models. The aim of this method is to estimate the nonparametric function that minimizes penalized least squares criterion. A roughness penalty term multiplied by a positive λ smoothing parameter is added to the residual sum of squares in smoothing spline regression. For this reason, the estimation of the unknown function depends on smoothing parameter λ whose values is generally obtained from data. Therefore, the determination of an optimum smoothing parameter in the interval $(0,\infty)$ has been arisen as an important problem. As related with subject in theory, many studies based on the different selection methods have been discussed for choosing an appropriate smoothing parameter [12].

Smoothing spline [4] estimate of the function arises as a solution to the following minimization problem: Find $f \in C^2[a,b]$ that minimizes the penalized residual sum of squares

$$S(f) = \sum_{i=1}^{n} \{y_i - f(x_i)\}^2 + \lambda \int_a^b \{f''(x)\}^2 dx \tag{3}$$

for value $\lambda > 0$ [10]. The smoothing spline estimate $\hat{f}_\lambda$ for data $y = (y_1,...,y_n)'$ are given by equation $f_\lambda = S_\lambda y$. In equation $\hat{f}_\lambda$ is a natural cubic interpolation spline with knots at $x_1,...,x_n$, vector $\hat{\boldsymbol{f}}_\lambda$ consist of values of $\hat{f}_\lambda$ for a fixed smoothing parameter $\lambda > 0$, S_λ is a well-known positive-definite (symmetrical) smoother matrix which depends on λ and the knot points $x_1,...,x_n$.

Here, if we want to explain the last paragraph, the solution based on smoothing spline for minimum problem in the equation (3) is known as a "natural cubic spline" with knots at $x_1,...,x_n$. From this point of view, a special structured spline interpolation which depends on a chosen value λ becomes a suitable approach of function f in model 1.

Let $f = (f(x_1),...,f(x_n))$ be the vector of values of function f at the knot points $x_1,...,x_n$. The smoothing spline estimate $\hat{f}_\lambda$ of this vector or the fitted values for data $y = (y_1,...,y_n)'$ are given by

$$\hat{f}_\lambda = \begin{bmatrix} \hat{f}_\lambda(x_1) \\ \hat{f}_\lambda(x_2) \\ . \\ . \\ . \\ \hat{f}_\lambda(x_n) \end{bmatrix}_{(n\times 1)} = (S_\lambda)_{(n\times n)} \begin{bmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{bmatrix}_{(n\times 1)}$$

where $\hat{f}_\lambda$ is a natural cubic spline with knots at $x_1,...,x_n$ for a fixed smoothing parameter $\lambda>0$, and S_λ is a well-known positive-definite (symmetrical) smoother matrix which depends on λ and the knot points $x_1,...,x_n$ but not on y.

The next method that we used in study is P-splines [2]. Eilers and Marx make some significant changes in smoothing spline technique. They made following two assumptions: First, they assume that *E(y)=Ba* where $B = (B_1(x), B_2(x),..., B_k(x))$ is an $n \times k$ matrix of B-splines and a is the vector of regression coefficients; secondly, they suppose that the coefficients of adjacent B-splines satisfy certain smoothness conditions that can be expressed in terms of finite differences of the a_is. Thus, from a least-squares perspective, the coefficients are chosen to minimize

$$S = \sum_{i=1}^{m} \left\{ y_i - \sum_{j=1}^{n} a_j B_j(x_i) \right\}^2 + \lambda \sum_{j=k+1}^{n} (\Delta^k a_j)^2 . \tag{4}$$

For least squares smoothing we have to minimize S in (4). The selection of smoothing parameter is one of the main problems in regression technique with penalty. In this study we used a well-known generalized cross-validation method for

choosing of smoothing parameter.Cross-Validation method suppose the following. Let $(S_\lambda)_{ii}$ be the *i*th diagonal element of S_λ. For smoothing splines the usual Cross Validation score function is

$$CV(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{y_i-\hat{f}_\lambda(x_i)}{1-(S_\lambda)_{ii}}\right\}^2. \quad (5)$$

Here λ is chosen to minimize $CV(\lambda)$. The extension of cross-validation is Generalized cross-validation method. The basic idea of Generalized cross validation is to replace the denominators $1-(S_\lambda)_{ii}$ of Cross Validation by their average $1-n^{-1}tr(S_\lambda)$, giving Generalized Cross Validation score function

$$GCV(\lambda)=\frac{1}{n}\frac{\sum_{i=1}^{n}\left\{y_i-\hat{f}_\lambda(x_i)\right\}^2}{\left\{1-n^{-1}tr(S_\lambda)\right\}^2}. \quad (6)$$

In this equation λ is chosen to minimize $GCV(\lambda)$.

The Bayesian versions of regression and penalized spline have a wide implementation in last decades. Among the Bayesian spline techniques we used Bayesian penalized spline and adaptive Bayesian regression spline.

Bayesian P-splines [5] approach by Andreas Brezger and Stefan Lang for additive models and extensions by replacing difference penalties with their stochastic analogues, i.e. Gaussian (intrinsic) random walk priors which serve as smoothness priors for the unknown regression coefficients. Compared to smoothing splines, in a P-splines approach a more parsimonious parameterization is possible, which is of particular advantage in a Bayesian framework where inference is based on MCMC techniques [3].Compared to smoothing splines, in a P-splines approach a more parsimonious parametrization is possible, which is of particular advantage in a Bayesian framework where inference is based on MCMC techniques.

In a Bayesian approach unknown parameters β_j, $j=1,...,p$ and γ are considered as random variables and have to be supplemented with appropriate prior distributions. For the fixed effects parameters γ it assumed independent diffuse priors, i.e. $\gamma_i \propto const$, $j=1....q$.

Priors for the regression parameters of nonlinear functions are defined by replacing the difference penalties by their stochastic analogues. First differences correspond to a first order random walk and second differences to a second order random. Thus, it obtained

$$\beta_{jp}=\beta_{j,p-1}+u_{jp}$$
$$\beta_{jp}=2\beta_{j,p-1}-\beta_{j,p-2}+u_{jp}. \quad (5)$$

The priors can be equivalently written in the form of global smoothness priors.

$$\beta_j\left|\tau_j^2\right. \propto \exp(-\frac{1}{2\tau_j^2}\beta_j'K_j\beta_j) \quad (6)$$

with appropriate penalty matrix $\mathbf{K}_j$.

For full Bayesian inference, the unknown variance parameters τ_j^2 are also considered as random and estimated simultaneously with the unknown β_j. Therefore, hyperpriors are assigned to the variances τ_j^2 (and the overall variance parameter σ^2) in a further stage of the hierarchy by highly dispersed (but proper) inverse Gamma priors $p(\tau_j^2)\sim IG(a_j,b_j)$.

Adaptive Bayesian regression spline was introduced by [1]. He supposed a fully Bayesian approach to regression splines with automatic knot selection. As a basis function representation of the regression spline he used B-spline basis. The reversible jump Markov chain Monte Carlo method allows for estimation both of the number of knots and the knot placement, together with the unknown basis coefficients determining the shape of the spline [1].

Without CV and GCV, we can show some other smoothing parameter selection criterias. These are improved Akaike information criterion, Mallows' C_p criterion and Risk estimation using classical pilots.

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 [15]. This improved criterion is defined as

$$AIC_c=\log\frac{\sum\left\{y_i-\hat{f}_\lambda(x_i)\right\}^2}{n}+1+\frac{2\left\{tr(S_\lambda)+1\right\}}{n-tr(S_\lambda)-2}$$

$$=\log\frac{\left\|(S_\lambda-I)y\right\|^2}{n}+1+\frac{2\left\{tr(S_\lambda)+1\right\}}{n-tr(S_\lambda)-2} \quad (7)$$

This criterion is easy to apply for choosing of smoothing parameter, as can be seen from the equation (7).

Mallows' C_p criterion is known as unbiased risk estimate (UBR) in smoothing spline literature. This type of estimate was suggested by Mallows in 1973 in regression case, and applied to smoothing spline by Craven and Wahba in 1979.

When σ^2 is known, an unbiased estimate of the residual sum of squares is given by C_p criterion:

$$C_p(\lambda) = \frac{1}{n}\left\{\left\|(S_\lambda - I)y\right\|^2 + 2\sigma^2 tr(S_\lambda) + \sigma^2\right\}$$

$$= \frac{1}{n}\left\{\left\|y - \hat{f}_\lambda\right\|^2 + 2\sigma^2 tr(S_\lambda) + \sigma^2\right\} \quad (8)$$

Unless σ^2 is known, in practice an estimate for σ^2 is estimated by

$$\hat{\sigma}^2 = \hat{\sigma}^2_{\hat{\lambda}} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{f}_{\hat{\lambda}}(x_i)\right)^2}{tr\left(I - S_{\hat{\lambda}}\right)} = \frac{\left\|\left(S_{\hat{\lambda}} - I\right)y\right\|^2}{tr\left(I - S_{\hat{\lambda}}\right)} \quad (9)$$

where $\hat{\lambda}$ is pre-chosen with any of the CV, GCV or AIC_c criteria [13], [14].

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

$$R\left(f, \hat{f}_\lambda\right) = \frac{1}{n}E\left\|f - \hat{f}_\lambda\right\|^2$$

$$= \frac{1}{n}\left\{\left\|(S_\lambda - I)f\right\|^2 + \sigma^2 tr\left(S_\lambda S_\lambda^T\right)\right\} \quad (10)$$

It is straightforward to show that $R\left(f, \hat{f}_\lambda\right) = E\left\{C_p(\lambda)\right\}$. Because the risk $R\left(f, \hat{f}_\lambda\right)$ is an unknown quantity, so-called risk is now estimated by computable quantity $R\left(\hat{f}_{\lambda_p}, \hat{f}_\lambda\right)$.

The obtained expression for $R\left(\hat{f}_{\lambda_p}, \hat{f}_\lambda\right)$ is

$$R\left(\hat{f}_{\lambda_p}, \hat{f}_\lambda\right) = \frac{1}{n}E\left\|\hat{f}_{\lambda_p} - \hat{f}_\lambda\right\|^2$$

$$= \frac{1}{n}\left\{\left\|(S_\lambda - I)\hat{f}_{\lambda_p}\right\|^2 + \hat{\sigma}^2_{\lambda_p} tr\left(S_\lambda S_\lambda^T\right)\right\} \quad (11)$$

Where $\hat{\sigma}^2_{\lambda_p}$ and $\hat{f}_{\lambda_p}$ 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 [12].

III. Problem Solution

The simulation study that presents in this study was conducted to evaluate the performances of the above nonparametric techniques. The functions for the simulation were devised in a family-wise manner, where for each family a different factor was used.

Table 1. Specification of simulation setup

Factor	Generic form
Noise level	$y_{ij} = f(x_i) + \sigma_j \varepsilon_i$
Design density	$y_{ij} = f(X_{ji}) + \sigma\varepsilon_i$
Spatial variation	$y_{ij} = f_j(x_i) + \sigma\varepsilon_i$
Variance function	$y_{ij} = f(x_i) + \sqrt{v_j(x_i)}\varepsilon_i$
Particular choices	
$\sigma_j = 0.02 + 0.04(j-1)^2$	
$\sigma = 0.1$, $X_{ji} = F_j^{-1}(X_i)$	
$\sigma = 0.2$, $f_j(x) = \sqrt{x(1-x)}\sin\left[\frac{2\pi\left\{1 + 2^{(9-4j)/5}\right\}}{x + 2^{(9-4j)/5}}\right]$	
$v_j(x) = \left[0.15\left\{1 + 0.4(2j-7)(x-0.5)\right\}\right]^2$	
Other explanations	
$j = 1,...,6;\ n = 50;\quad n = 100;\quad n = 200;$ $n = 400;\ x_i = \frac{i - 0.5}{n};\quad \varepsilon_i \sim iidN(0,1)$	
$f(x) = 1.5\phi\left(\frac{x - 0.35}{0.15}\right) - \phi\left(\frac{x - 0.8}{0.04}\right);$	
$\phi(u) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-u^2}{2}\right)$	
$X_i \sim iid\ \ Uniform[0,1];\quad F_j$ is the $Beta\left(\frac{j+4}{5}, \frac{11-j}{5}\right)$ c.d.f.	

The factors are: noise level, design variable, spatial variation, variance function. For each function we take n=50, n=100, n=200 and n=400 values with 200 replications. The results of simulation are compared using mean value of *MSE* of each replication and graphically showed using box-plots.

The empirical analysis are made by mgcv package [8] of R software, BayesX software[5], http://www.stat.uni-

muenchen.de/~lang/bayesx and for the adaptive Bayesian regression spline we used the functions from the http://www.stat.uni-muenchen.de/sfb386. This setup originally due to Professor Steve Marron, was designed to study the effects of varying the noise level, design density, degree of spatial variation and noise variation function. Totally four sets of numerical experiments are to be performed. Within each set of experiments, the factors under consideration are changed 6 times and there are 24 different configurations. [7] used this functions to make a comparison study of regression spline techniques. The aim of paper [6] was to make a right selection of smoothing parameter using these functions. There is information about used functions in Table 1.

Table 2. Performance criteria of used techniques with n=200 observations

	Penalized spline	Regression spline	Bayesian P-spline	Adaptive Bayesian regression spline
NL 1	**0.000345**	0.011402	0.000664	0.011428
NL 2	0.003114	0.0144	**0.002367**	0.02243
NL 3	0.028975	0.041538	**0.02868**	0.032737
NL 4	0.132932	0.148918	**0.123195**	0.127094
NL 5	0.405746	0.427053	**0.400678**	0.435935
NL 6	0.999366	1.024137	**0.938779**	1.023395
DD 1	**0.008734**	0.021348	0.009247	0.021346
DD 2	0.008946	0.021756	**0.0088**	0.017046
DD 3	**0.008982**	0.021512	0.009403	0.025547
DD 4	**0.008862**	0.021178	0.008991	0.031536
DD 5	**0.00887**	0.021496	0.009291	0.015608
DD 6	**0.008933**	0.022689	0.009236	0.023871
SV 1	0.038036	0.038616	**0.032795**	0.043144
SV 2	**0.036763**	0.041095	0.037599	0.039652
SV 3	0.045477	0.069119	**0.044682**	0.059744
SV 4	0.052796	0.075027	**0.051428**	0.064938
SV 5	**0.054327**	0.076132	0.05751	0.085593
SV 6	**0.052752**	0.074574	0.061656	0.057537
NV 1	**0.02257**	0.034986	0.024483	0.032333
NV 2	**0.022302**	0.034779	0.023689	0.035978
NV 3	**0.0207**	0.032758	0.021413	0.037997
NV 4	0.020484	0.032753	**0.017057**	0.03045
NV 5	**0.022515**	0.03487	0.023706	0.032237
NV 6	0.026817	0.039535	**0.02544**	0.039505

From these functions we have sampled a lot of data. As the results of the simulation we gathered a lot of information. The results of mean values of mean squared consist of 4 tables. The comparison of each technique of sampled observations are shown using box-plots. For one number of observations, for example n=200, we get 24 box-plots. In our study we used n=50, n=100, n=200 and n=400. Because of this, there become a lot of figures of box-plots. As an example here we have showed only results for n=200.

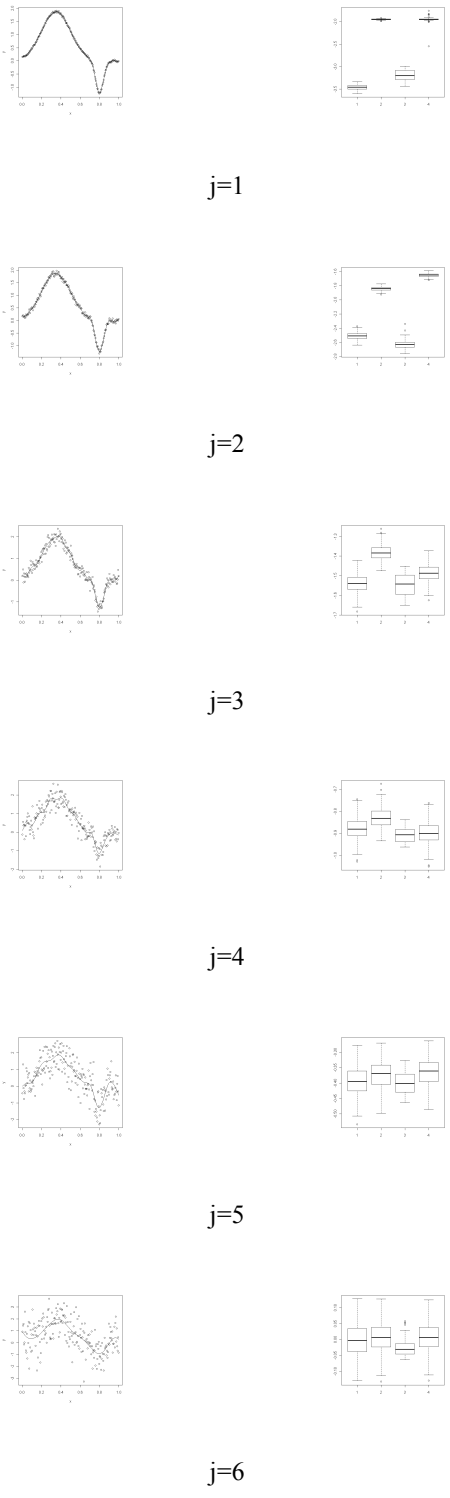

Figure 1. Results of estimation of the Noise level factor function

There are results of performance criteria of n=200

observations sampled from the functions. The abbreviations in first column of the Table means the classification of functions: NL – noise level, DD – design density, SV – spatial variation, NV – noise variation. The numbers 1 from 6 means the position change of functions.

The mean squared error results in Table 2 shows us that the penalized spline and Bayesian penalized spline shows better results that the regression and adaptive Bayesian regression splines.

The reason of better result of penalized techniques is that they used a penalty term based on divided differences of their coefficients. But as you can see from the table, results are become closer when the position of functions changes. In Figure 1, you can realize this difference. Figure 1, shows us the different position of noise function. The observations sampled from position 1 to 6 are different scattered. When $j=1$ observations are close to each other and shows more smoothed function. When j becomes bigger, observations become more scattered. For example, when j=6 observations still to far from each other.

Table 3. Performance criteria of used techniques with n=50 observations

	Penalized spline	Regression spline	Bayesian P-spline	Adaptive Bayesian regression spline
NL 1	0.000166	0.012815	0.000171	0.014353
NL 2	0.00194	0.015287	0.001942	0.016968
NL 3	0.020739	0.038916	0.020762	0.03913
NL 4	0.111144	0.135153	0.020762	0.03913
NL 5	0.325331	0.380682	0.325428	0.422558
NL 6	0.827612	0.934913	0.827968	0.938111
DD 1	0.006489	0.019144	0.006684	0.021441
DD 2	0.006449	0.018522	0.006449	0.019324
DD 3	0.006526	0.018603	0.006527	0.020891
DD 4	0.006693	0.018878	0.006701	0.018942
DD 5	0.006497	0.019705	0.006561	0.020142
DD 6	0.006256	0.020283	0.006277	0.022514
SV 1	0.032824	0.035179	0.033038	0.0353
SV 2	0.028291	0.034114	0.028476	0.034232
SV 3	0.047157	0.061821	0.047465	0.062034
SV 4	0.049601	0.071452	0.049926	0.071699
SV 5	0.050501	0.074163	0.050831	0.07442
SV 6	0.053525	0.075331	0.053875	0.075591
NV 1	0.017834	0.035957	0.017951	0.036081
NV 2	0.015467	0.03275	0.015569	0.032863
NV 3	0.015	0.031669	0.015098	0.031778
NV 4	0.01521	0.031135	0.015309	0.031242
NV 5	0.016077	0.033144	0.016182	0.033258
NV 6	0.019882	0.036777	0.020012	0.036904

From Table 2, we could conclude that when observations are close to each other, penalized spline and Bayesian penalized spline shows good results. But with scattered observations it's preferable to use the proposed techniques, but with revision of the observations.

Results of simulation study with another number observations, i.e. n=50, n=100 and n=400, results differs, but not so much. We could conclude from other results that Bayesian penalized spline shows good results with a large number of observations. The mean values of mean squared error are gathered in Tables 3, 4, 5.

When the numbers of observations are equal to 50, the models obtained by penalized spline technique are close to sampled values. We can conclude this by observing results from the Table 3.

The next table shows us results of performance criteria of used techniques with n=100 observations. When the numbers of observations increase, Bayesian penalized spline method shows better result comparing to previous table with 50 observations.

Table 4. Performance criteria of used techniques with n=100 observations

	Penalized spline	Regression spline	Bayesian P-spline	Adaptive Bayesian regression spline
NL 1	**0.00028**	0.011893	0.000289	0.01204
NL 2	0.002717	0.014706	**0.002716**	0.015691
NL 3	0.026833	0.041373	**0.026726**	0.044145
NL 4	**0.123872**	0.144431	0.124299	0.154108
NL 5	0.385279	0.410246	**0.384608**	0.437732
NL 6	0.926801	0.987671	0.929998	1.053845
DD 1	0.007989	0.020256	0.008017	0.021613
DD 2	0.007981	0.020719	0.008008	0.022108
DD 3	0.007949	0.020937	0.007977	0.022339
DD 4	0.007949	0.021325	0.007977	0.022754
DD 5	0.007897	0.020776	0.007924	0.022168
DD 6	0.007796	0.021915	0.007823	0.023383
SV 1	0.036157	0.037492	0.036282	0.040004
SV 2	0.033841	0.038709	0.033957	0.041302
SV 3	0.044934	0.068767	0.045089	0.073374
SV 4	0.049033	0.07268	0.049203	0.07755
SV 5	0.04804	0.071347	0.048206	0.076128
SV 6	0.047653	0.071625	0.047817	0.076424
NV 1	0.024115	0.038666	0.024198	0.041257
NV 2	0.020269	0.034734	0.020339	0.037061
NV 3	0.018082	0.031393	0.018145	0.033496
NV 4	0.018421	0.032097	0.018485	0.034247
NV 5	0.020609	0.034565	0.020681	0.036881
NV 6	0.023546	0.038501	0.023627	0.041081

By the bold style of some cells in Table 4., it was just shown the comparison of penalized spline and its Bayesian version.

Table 5. Performance criteria of used techniques with n=400 observations

	Penalized spline	Regression spline	Bayesian P-spline	Adaptive Bayesian regression spline
NL 1	0.00037	0.011237	**0.000366**	0.011347
NL 2	0.003339	0.014287	**0.003299**	0.014427
NL 3	**0.030559**	0.042393	0.030692	0.042808
NL 4	0.137062	0.150939	**0.135417**	0.152418
NL 5	0.420344	0.436346	**0.4153**	0.440622
NL 6	1.002157	1.025751	0.990131	1.035803
DD 1	0.009258	0.02137	0.009384	0.021197
DD 2	0.009268	0.021674	0.009394	0.021498
DD 3	0.009382	0.021288	0.009308	0.021308
DD 4	0.009198	0.021373	0.009323	0.021393
DD 5	0.00943	0.021666	0.009355	0.021686
DD 6	0.009437	0.021785	0.009363	0.021805
SV 1	0.039429	0.039725	0.038956	0.039841
SV 2	0.038	0.041681	0.037544	0.041803
SV 3	0.048525	0.07001	0.047943	0.070215
SV 4	0.054663	0.075641	0.054007	0.075863
SV 5	0.054485	0.07578	0.053831	0.076002
SV 6	0.054213	0.075172	0.053562	0.075392
NV 1	0.028689	0.04054	0.028012	0.040633
NV 2	0.023878	0.035536	0.023315	0.035618
NV 3	0.021583	0.033069	0.021074	0.033145
NV 4	0.021658	0.033174	0.021147	0.03325
NV 5	0.023599	0.035006	0.023042	0.035086
NV 6	0.028198	0.040135	0.027533	0.040228

From Table 5, we can realize that when number of observations become bigger, Bayesian penalized spline outperforms other techniques.

IV. Conclusion

In this study we made a simulation study using various nonparametric techniques. The methods that we have used in this study are following: regression spline, penalized spline, and their Bayesian versions: adaptive Bayesian regression spline and Bayesian P-splines. The main goal of our study is to compare nonparametric regression techniques and Bayesian versions of these techniques. For this purpose we made a simulation study with different functions. For each function we sampled n = 50, n = 100, n = 200 and n =400 number of observations. The purpose of using different number of sampled observations is to analyze the behavior of utilized techniques. For simulation study we used dataset which sampled from the functions in the paper of [7]. The results of simulation study are compared with each other using mean value of the MSE (mean squared error) and showed results graphically by using box plot of MSE.

In this paper we made a comparison study between regression spline, penalized spline, and their Bayesian versions: adaptive Bayesian regression spline and Bayesian penalized spline with a different number of observations. For this purpose we made a simulation study with four different functions with six positions. For regression and penalized splines the important problems are the knot selection and selection of smoothing parameter. For both techniques we used equidistant knot selection as a basis method in regression techniques. Of course there are a lot of methods for knot selection, but we will revised it in further researches. The selection of smoothing parameter is made using a well-known generalized cross validation method. For Bayesian techniques important point is selection parameters for prior and posterior distribution. For Bayesian penalized spline was made the different hyperparameter selection.

The best result for Bayesian penalized spline we could find when hyperparameters are a=1, b=0.00005.

The adaptive Bayesian regression spline show the worst result among all techniques used in analysis, because of position of observations. For more scattered observations this technique shows pretty good estimation.

The penalized spline showed one of the best results between spline techniques and their Bayesian versions. But in some functions Bayesian penalized spline outperforms penalized spline. The results of regression spline are close to adaptive Bayesian method.

We can get better performance from Bayesian penalized spline while data size is increasing. For example, when n = 400, Bayesian penalized spline outperforms penalized spline. So, we propose using Bayesian penalized spline for observations with big size.

References

[1] C. Biller, "Adaptive Bayesian Regression splines in semi parametric generalized linear models", *Journal of Computational and Graphical Statistics*, 9 (1), 2000, pp.122-140.G.

[2] P.H.C. Eilers and B.D. Marx, "Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinders)", *Statistical Science*, 11(2), 1996, pp. 89-121.

[3] L. Fahrmeir and S. Lang, "Bayesian Inference for generalized additive mixed models based on Markov Random Field Priors", *Journal of the Royal Statistical Society C (Applied Statistics)*, 50, 2001, pp. 201-220.

[4] P.J. Green and B.M. Silverman, *Nonparametric regression and generalized linear models.* Chapman and Hall, London, 1994.

[5] S. Lang and A. Brezger, "Bayesian P-splines", *Journal of Computational and Graphical Statistics*, 13, 2001, pp. 183-212.

[6] T.C.M. Lee, "Smoothing parameter selection for smoothing splines: A simulation study", *Computational Statistics and Data Analysis*, 2003, pp.139-148.

[7] M.P. Wand, "A comparison of regression spline smoothing procedures", *Computational Statistics*, 15, 2000,pp. 443-462.

[8] S.N. Wood, "Generalized additive models: an introduction with R", *Chapman and Hall*, 2006.

[9] D. Aydın, M. Mammadov, "A Comparative Study of Hybrid, Neural Networks and Nonparametric Regression Models in Time Series Prediction", *WSEAS Transactions on Mathematics*, Issue 10, Volume 8, October 2009, pp. 593-603.

[10] D. Aydın, "Estimation of GDP in Turkey by nonparametric regression models", *Proceedings of the 6.th WSEAS International Conference on Instrumentation, Measurement, Circuits&Systems,* Hangzhou, China, April 15-17, 2007, pp. 221-225.

[11] N. A. Ibrahim, Suliadi, "GEE-Smoothing spline in semiparametric model with correlated nominal Data: Estimation and simulation study", *ASM'10 Proceedings of the 4th International Conference on Applied Mathematics, Simulation, Modeling,* WSEAS Stevens Point, Wisconsin, USA, 2010, pp. 19-26.
[12] D. Aydın, R. E. Omay, "The Empirical Performances of the Selection Criteria for Nonparametric Regression Using Smoothing Spline", *Proceedings of the 5th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics,* Venice, Italy, November 20-22, 2006, pp. 228-233.
[13] T.C.M. Lee, "İmproved Smoothing spline Regression by Combining Estimates of Different Smoothness", *Statistics & Probability Letters,* 2004.
[14] G. Wahba, "Spline Model for Observational Data", Siam, Philadelphia Pa., 1990.
[15] C. M. Hurvich, J. S. Simonoff and C.-L. Tasi, "Smoothing parameter selection in nonparametric regression using an improved Akaike Information criterion", *J. R. Statist. Soc. B,* Vol. 60, 1998, 271-293.
[16] C. De Boor, "A Practical Guide to Splines", Springer, New York, 1978.
[17] P. Dierckx, "Curve and Surface Fitting with Splines", Clarendon Press, Oxford, 1993.

Memmedaga Memmedli was born in Azerbaijan in 1947. He received a diploma at Baku State University in 1971. He received a PhD diploma from Russian Academy of Science in 1977. From 1985-2006, he has worked as a Associate Professor; from 2007, as a Professor Dr. at Department of Statistics at Anadolu University, Eskisehir, Turkey. Now, he is head of Department of Statistics, Faculty of Science, Anadolu University, Eskisehir, Turkey.
He has published many journals and conference papers in his research areas. His research interests are Game Theory, Control Theory, Optimization, Artificial Neural Networks, Fuzzy Modeling and Nonparametric Regression Analysis.