Journal of Scientific Research & Reports
4(2): 153-161, 2015; Article no.JSRR.2015.017
ISSN: 2320–0227

SCIENCEDOMAIN *international*
www.sciencedomain.org

SCIENCEDOMAIN

The Use of Spline, Bayesian Spline and Penalized Bayesian Spline Regression for Modeling

M. Sami Erdoğan[1] and Özlem Ege Oruç[1*]

[1]Department of Statistics, Faculty of Sciences, Dokuz Eylül University, Turkey.

Authors' contributions

This work was carried out in collaboration between both authors. Author MSE designed the study, and wrote the first draft of the manuscript. Author ÖEO checked the results of the study performed the Spline and Bayesian Spline regression analysis. Authors MSE and ÖEO managed the draft corrections. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JSRR/2015/11712
Editor(s):
(1) Narcisa C. Apreutesei, Technical University of Iasi, Romania.
(2) Luigi Rodino, Faculty of Mathematical Analysis, Dipartimento di Matematica, Università di Torino, Italy.
Reviewers:
(1) Anonymous, Kookmin University, South Korea.
(2) Anonymous, Central South University, China.
(3) Maria Hussain, Department of Mathematics Lahore College for Women University, Lahore, Pakistan.
(4) Mohan Delampady, Statistics and Mathematics Unit, Indian Statistical Institute, Bangalore, India.
(5) Anonymous, Slovak University of Technology, Bratislava, Slovak Republic.
(6) Anonymous, Atatürk University, Turkey.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=697&id=22&aid=6384

Short Research Article

Received 30th May 2014
Accepted 13th September 2014
Published 7th October 2014

ABSTRACT

Aims: The aim of this study is modeled the ratios of export to imports data in Turkey by using nonparametric regression methods.
Study Design: This was Spline, Bayesian Spline and Penalized Spline Regression modeling study.
Place and Duration of Study: Turkish Statistical Institute. The ratios of export to import data consist of sixty-seven month periods (May 2007 to November 2012).
Methodology: In this study, distribution graph of ratios of export to import between 2007 to 2012 years in Turkey is modeled using spline and Bayesian spline regression methods. The results of these regression models are compared. Then Penalized spline regression is examined with Bayesian approach and models are established for the different values of the smoothing parameter which obtained using prior distributions. We proposed a new smoothing parameter using the information content of normal distribution. Under the assumption of the coefficients of basis functions are normally distributed, the new smoothing parameter (λ^*) is defined as the ratio of the

**Corresponding author: E-mail: ozlem.ege@deu.edu.tr;*

information content of normal distribution.
Results: When we compared the spline and Bayesian spline regression models, both models were shown similar characteristics. The coefficients of *β* and *b* parameter vectors were very similar and the coefficients of determination of two models were obtained same. But, the standard errors of parameter estimations of Bayesian spline regression were smaller than spline regression models. For this reason, we conclude that Bayesian spline regression model parameter estimation is more reliable then spline regression model. We also compared penalized Bayesian spline models using different penalty terms. The different models on the same data set have been set up using different value of λ. From the results, observe that the absolute value of the coefficients of basis functions decrease as the penalty term 1/λ increase. Also, the coefficient of determination of the model gradually diminishes. In addition, we proposed a new smoothing parameter using the information content of normal distribution. According to results, small changes in λ* have made drastic changes in smoothing of the model. So, we conclude that λ* is more sensitive than traditional smoothing parameter (λ).
Conclusion: We investigated the three most common nonparametric regression models, which are called spline, Bayesian spline and penalized Bayesian spline, discussing advantages and disadvantages of them using real data. We conclude that Bayesian spline regression model parameter estimation is more reliable than other models. In addition, we conclude that λ* is more sensitive than traditional smoothing parameter (λ).

Keywords: Spline regression; bayesian spline regression; penalized bayesian spline regression; smoothing parameter.

1. INTRODUCTION

There are basically two different philosophical approaches in science statistics. The classical (Frequentist) approach and Bayesian approach. In the past, while the Bayesian approach has always been powerful, it has not always been practical. This is due in large part to the relatively high computational overhead of performing the integrations and summations which lie at the heart of the Bayesian approach. In recent years Bayesian methods have become widespread in many areas such as data analysis and regression analysis. The availability of fast computers allows the required computations to be performed in reasonable time, and thereby makes the benefits of a Bayesian treatment accessible to an ever broadening range of applications.

The need for parsimonious statistical models is well-known and parametric models are often a convenient method for achieving parsimony. However, nonparametric models exist because there are many examples where parametric models do not provide a suitable fit to the data. The main advantage of nonparametric over parametric models is their flexibility. In the nonparametric framework the shape of the functional relationship between covariates and the dependent variables is determined by the data, whereas in the parametric framework the shape is determined by the model. Semiparametric modelling allows a researcher to have the best of both worlds: the parametric and the nonparametric. Those features of the data that are suitable for parametric modelling are modeled that way and nonparametric components are used only where needed.

The Bayesian inference for nonparametric models enjoys the flexibility of nonparametric models and the exact inference provided by the Bayesian inferential machinery. It is this combination that makes Bayesian nonparametric modeling so attractive [1,2]. In this work, we address the problem of nonparametric and semi parametric regression methods in the framework of the Bayesian approach. The flexibility of these methods has a great advantage in the modeling. Nonparametric and semi parametric regression methods have been discussed extensively in the statistics literature. Some of the studies for these regression models can be listed as [3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19]. In this study, distribution graph of ratios of export to import between May 2007 and November 2012 is modeled using spline and Bayesian spline regression methods. The model is constructed using R Language programming. The results of these regression models are compared. Then Penalized spline regression is examined with Bayesian approach and models are established for the different values of the smoothing parameter which obtained using prior distributions. Finally, we proposed a new smoothing parameter using the information content of normal distribution. Under the assumption of the coefficients of basis functions are normally distributed, the new smoothing parameter (λ*) is defined as the ratio of the information content of normal distribution.

2. MATERIALS AND METHODS

2.1. Bayesian Regression

In the classical regression, the distribution of X is assumed to provide no information about *X* the conditional distribution of y given X [20]. But Bayesian regression, the distribution of the independent variables is included the likelihood function. For this reason regarding distribution of the independent variable is eliminated with proportional expression. As a result, Bayesian regression does not deal with the distribution of the independent variable. The mathematical presentation of this situation is given below.

Let ψ denote the distribution of *X*. If prior distribution is independent, $\beta_1, \beta_2, \ldots, \beta_k$, σ^2 and ψ, we can write this equation

$$p(\psi,\beta,\sigma^2) = p(\psi)p(\beta,\sigma^2) \quad (1)$$

Then, posterior distributions can be divided two factors,

$$p(\psi,\beta,\sigma^2|X,y) = p(\psi|X)p(\beta,\sigma^2|X,y) \quad (2)$$

Since the distribution of the independent variables is included the likelihood function, we can write proportional equation is given below

$$p(\psi,\beta,\sigma^2|X,y) = p(\psi|X)p(\beta,\sigma^2|X,y) \quad (3)$$

Therefore, Bayesian regression is not interested with the distribution of vector *X*. In linear regression, the model that relates observations and parameters is written

$$y|\beta,\sigma^2,X \sim N(X\beta,\sigma^2 I) \quad (4)$$

and

$$P(y|\beta,\sigma^2,X) = (\tfrac{1}{2\pi\sigma^2})^{n/2}\exp\left\{-\tfrac{1}{2\sigma^2}(y-X\beta)^{'}(y-X\beta)\right\} \quad (5)$$

Bayesian regression analysis begins with a prior distribution. Since a non informative prior distribution assigns the same probability to each possible value of the parameters, it is most commonly used in linear regression. A non informative prior distribution that is commonly used for linear regression is

$$p(\beta,\sigma) \propto \tfrac{1}{\sigma} \quad (6)$$

We are obtained the posterior distribution of β given σ^2 using the likelihood function and prior distributions. The posterior distribution is given in equation (7).

$$p(\beta,\sigma|y,X) \propto \tfrac{1}{\sigma^{n+1}}\exp\left\{-\tfrac{1}{2\sigma^2}\left[vs^2 + (\beta-\hat{\beta})^{'}X'X(\beta-\hat{\beta})\right]\right\} \quad (7)$$

where $vS^2 = (y-X\beta)' \ (y-X\beta)$ and $v = n-k.$

The marginal posterior probability distribution of β is derived by integrating the posterior distribution of β given σ^2 is as follows.

$$p(\beta|y) = \int_0^\infty p(\beta,\sigma|y)d\sigma = \int_0^\infty p(\beta|\sigma,y)p(\sigma|y)d\sigma \quad (8)$$

$$p(\beta|y) \propto [1 + \tfrac{1}{v}(\beta-\hat{\beta})'\tfrac{X'X}{s^2}(\beta-\hat{\beta})]^{-\frac{(k+v)}{2}} \quad (9)$$

A similar process can be follow for σ^2. The marginal posterior distribution of σ^2 (i.e. the integral over all possible values of β of the joint distribution of β and σ^2) is

$$p(\sigma|y) = \int_{-\infty}^{\infty} p(\beta,\sigma|y)d\beta \quad (10)$$

$$p(\sigma|y) \propto \tfrac{1}{\sigma^{v+1}}\exp\left(-\tfrac{vs^2}{2\sigma^2}\right) \quad (11)$$

2.2 Spline Regression

Spline Regression model is a semiparametric regression method that reveals relationships between variables which can be called. This regression method is generally preferred when classical methods like linear, quadratic and cubic are not good results. Nowadays, it is used often from time dependent data sets to economy fields. The following are the general form of the spline regression model.

$$m(x_i) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p + \textstyle\sum_{k=1}^{K} b_k (x-t_k)_+^p \quad (12)$$

Where t_K are fixed and known knots, $K$are the number of knots, $\beta_0, \beta_1, \ldots, \beta_p, b_1, b_2, \ldots, b_K$ are unknown regression coefficients in the model. Also, p indicate the degree of spline regression model and $(x-t_k)_+^p$ statement is included as basis function in the model. An important characteristic of function $(x-t_k)_+^p$ is that equal to *0* value as minimum and it is positive definite. If the value of independent variable is smaller than knot value; the value of function will be equal to *0*. Otherwise, if the value of independent variable is greater than knot value, the value of function will be equal to the degree of *pth* of the value of independent variable minus knot value [21].

Spline regression method disintegrates the data set with respect to knots and a function is obtained for each segment. It continues to analysis by way of these segments. Then, the degree of the spline is determined such as linear, quadratic; and analysis is performed.

2.3 Penalized Spline Regression

Penalized spline regression models are a popular statistical tool for curve fitting problems due to their flexibility and computational efficiency. It is a nonparametric regression technique that relies on principles of statistical theory to minimize the possibility of overfitting [2]. The basic idea behind penalized regression methods is to quantify the notion of roughness of a curve through a suitable penalty functional and then to pose the estimation problem in a way that makes explicit the necessary compromise between bias and variability in curve fitting. Spline regression needs to choose the number of knots and their positions but estimation is sensitive to this choice. Penalized spline regression is used a penalization parameter (λ), which is related to the fluctuations of the regression function, to reduce the impact of this choice. Consider the regression model;

$$y_i = m(x_i) + \varepsilon_i \tag{13}$$

Where *m(.)* is a smooth function it is defined as,

$$m(x;\theta) = \beta_0 + \beta_1 X + \sum_{j=1}^{K} b_j (x - K_{ij})_+ \tag{14}$$

Where $\theta = (\beta_0, \beta_1, b_1 \dots, b_K)$. The aim of the regression analysis to estimate the regression function f, where $E(Y|X) = m(x)$. Here, we directly solve for the function f that minimizes the following objective function, a penalized version of the least squares objective:

$$\sum_{i=1}^{n} \{y_i - m(x_i)\}^2 + \frac{1}{\lambda}\theta^T D \tag{15}$$

The first term captures the fit to the data, while the second penalizes curvature. Here, λ is the smoothing parameter, the selection of theλ smooth parameter is of great importance in penalized spline regression. The case $\lambda = 0$ corresponds to the unconstrained case. Increasing the value ofλdown weights the influence of the knots and gives a less rough fit. If we takeλ to be very large, then the effect of the knots diminishes and the least-squares line is approached. There exist some methods for choosingλ and the knot locations from the data.

In (15), *D* is a known positive semi-definite penalty matrix. It is defined as follows;

$$D = \begin{bmatrix} 0_{(p+1)\times(p+1)} & 0_{(p+1)\times K} \\ 0_{K\times(p+1)} & \Sigma^{-1} \end{bmatrix} \tag{16}$$

3. RESULTS AND DISCUSSION

Export and import constitute a country's foreign trade. Assessment can be made about the country's economic situation by examining foreign trade statistics. One of these statistics is the ratio of export to import. In this study, monthly changes in ratios of export to import of Turkey between the years 2007-2012 were examined with spline, Bayesian spline and penalized Bayesian spline models. Examining the distribution chart of data, ratios of export to import show a wavy distribution between the years 2007-2012. In models where the distribution is wavy, it is more advantageous to use nonparametric and semiparametric regression rather than a single model.

In this study, monthly changes in ratios of export to import of Turkey between the years 2007-2012 were examined with spline, Bayesian spline and penalized Bayesian spline models. This chapter presents an application to compare the performance of spline, Bayesian spline and penalized Bayesian spline models on this data set.

Our aim compares the performance of spline regression and Bayesian Spline regression model in terms of their value of coefficient of determination. So is to demonstrate how to use the penalty term in Bayesian approach and make a new proposal on penalty term.

The models are illustrated with an application to ratios of export to import data set given in Turkish Statistical Institute (TUIK). These data consist of sixty-seven month periods (May 2007 to November 2012).In this model the dependent variable (Y) and independent variable (X) are denoted by the ratios of export to import and months, respectively.

To knots to determine, Adaptive Spline Regression method was used. With this method, knots that minimize the lack of fit and the number of its are determined. Thus, in this data set we reached interior knots given by (17, 49, 53, 57) and the degree of the spline is one. Using the R code and then uses least squares to construct the regression model for ratios of export to import data set. The results of spline regression are given in Table 1.

From the Table 1, intercept, the coefficient of independent variable and coefficient of the basis functions in the model were obtained. Coefficient of determination (R-Squared) for this model is obtained 0.6691. All of these coefficients are statistically significant. According the value of F-statistic, the model is valid. The graph of this model is given in Fig. 1.

Then, we applied Bayesian spline regression analysis for same data set. Prior distributions are determined for each parameter which is considered as random variables in the model. Parameters and it's a prior distributions are summarized in (17).

$$\begin{gathered} y_i = \beta_0 + \beta_1 x + \sum_{k=1}^{4} b_k (x - t_k)_+ + \varepsilon_i \\ \beta_0, \beta_1 \sim N(0, 10^6) \\ b_k \sim N(0, 10^6) \\ \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \end{gathered} \quad (17)$$

The different prior distribution can be selected for variance in the literature. In this analysis, the distribution of precision parameter $\tau_\varepsilon = (\sigma_\varepsilon)^{-2}$ was taken a gamma distribution. This is because variance will have the inverse gamma distribution conjunction with select the gamma distribution for precision parameter and conjugate structure have created while getting the posterior distribution. Using the WinBUGS code then construct the Bayesian spline regression model for ratios of export to import data set. There are three different stages of WinBUGS program. These are, writing code for interest model, loading data and creating the initial values for the parameters respectively. The burn-in period, which was used to eliminate the effect of the initial values, was consisted 2000 iterations in this example. The reason is that the number of iterations used in literature reviews based on. The results of Bayesian spline regression are given in Table 2.

The values related to the posterior distribution such as posterior mean, posterior median, MC error, *2.5%* and *97.5%* quantiles were obtained. MC error is used to decide the parameters convergence or not. If this value is smaller than *0.05* we can decide parameter convergence. MC values of all parameters in Bayesian Spline Regression model is smaller than *0.05*. So we were decided that parameters of the models convergence. The *R-Squared* of the model was obtained *0.6697*. The graph of Bayesian spline regression model is given in Fig. 2.

Table 1. The results of spline regression

Coefficients	Estimate	Std. error	t value	Pr(>|t|)
(Intercept)	70.5783	2.0850	33.850	<2e-16***
X	-1.0270	0.1634	-6.285	3.86e-12***
b_1	1.7940	0.2119	8.466	6.91e-09***
b_2	-5.7076	0.8489	-6.723	6.91e-09***
b_3	7.2988	1.6698	4.371	4.89e-05***
b_4	-3.1246	1.2499	-2.500	0.0151*

Residual standard error: 4.475 on 61 degrees of freedom; Multiple R-squared: 0.6691, Adjusted R-squared: 0.6419; F-statistic: 24.66 on 5 and 61 DF, p-value: 1.698e-13

Table 2. The results of bayesian spline regression

Node	Mean	Sd	MC error	2.5%	Median	97.5%	Start	Sample
Beta[0]	70.57	1.831	0.0129	66.93	70.57	74.16	2001	20000
Beta[1]	-1.027	0.144	0.001075	-1.307	-1.028	-0.742	2001	20000
b1	1.795	0.187	0.001448	1.427	1.796	2.16	2001	20000
b2	-5.703	0.750	0.005345	-7.175	-5.704	-4.2	2001	20000
b3	7.285	1.471	0.01035	4.343	7.279	10.16	2001	20000
b4	-3.113	1.097	0.007845	-5.294	-3.111	-0.943	2001	20000
Sigmaeps	3.924	0.315	0.002527	3.364	3.903	4.602	2001	20000

When we compared the two regression models, both models were shown similar characteristics. The coefficients of β and b parameter vectors were very similar and the coefficients of determination of two models were obtained same. Where there are lots of data such as this data set, Bayesian and classical approach eventually results in the same conclusion. But in this comparison, the standard errors of parameter estimations of Bayesian spline regression were smaller than spline regression models. For this reason, we conclude that Bayesian spline regression model parameter estimation is more reliable then spline regression model.

Akaike information criterion (AIC) [22] is the one of the most popular information criteria in the literature. These information criterias are often used model selection and variable selection in Bayesian analysis. To investigate this further we computed the value of the AIC for the spline regression model and for the Bayesian spline regression model results are presented in Table 3. The results show that the spline regression model provides a better fit to the data in terms of lower AIC.

The penalty term, $\lambda=(\sigma_b^2) \,/\, (\sigma_\varepsilon^2)$ which was restrict fluctuations of $\hat{m}$ was added to Bayesian spline model. The coefficient of determination and regression coefficients of this model were obtained for different penalty term $1/\lambda$. The results are given in Table 4.

From the Table 4, observe that the absolute value of the coefficients of basis functions decrease as the penalty term $1/\lambda$ increase. Also, the coefficient of determination of the model gradually diminishes. Another point is that if $1/\lambda$ is large, then the effect of the knots diminishes and the model approaches to the least-squares line. The selection of the smooth parameter $\lambda=\tau_\varepsilon/\tau_\beta=(\sigma_b^2)/(\sigma_\varepsilon^2)$ is of great importance in penalized Bayesian spline regression. The small value of λ corresponds over smoothing. The large value of λ corresponds under smoothing.

In this study, we proposed a new smoothing parameter using the information content of normal distribution. Under the assumption of the coefficients of basis functions are normally distributed, the new smoothing parameter is defined as the ratio of the information content of normal distribution,

$$(\lambda^*=\log_2[\sigma_b(2\pi e)^{(1/2)}]/\log_2[\sigma_\varepsilon\,(2\pi e)^{(1/2)}]).$$

We calculated the coefficient of determination and regression coefficients for penalized Bayesian regression model to investigate the performance of the new smoothing parameter. The results are given in Table 5.

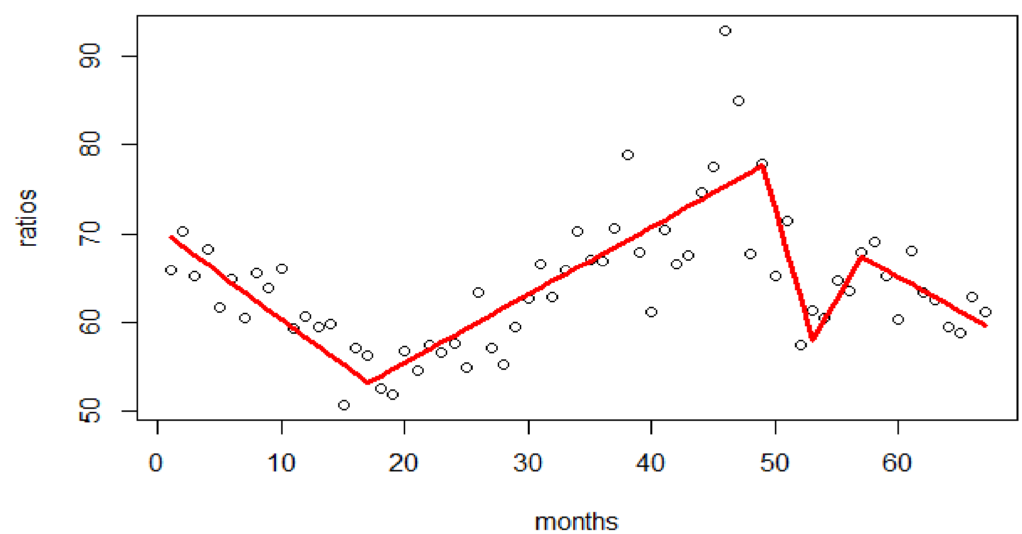

Fig. 1. The spline regression model of data using the manually-selected knots (17, 49, 53, 57)

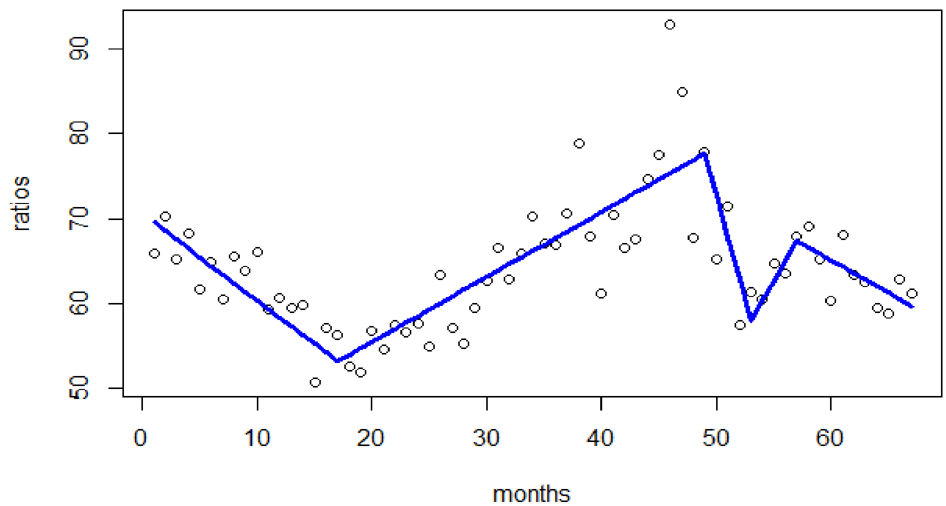

Fig. 2. The bayesian spline regression model of data using the manually-selected knots (17, 49, 53, 57)

Table 3. The values of AIC for spline and bayesian spline regression

Model	AIC
Spline regression	398.637
Bayesian spline regression	406.726

Table 4. Penalized bayesian spline regression model for different λ

Parameter	1/λ=0.85	1/λ=2.25	1/λ=17.1	1/λ=267.6
b_1	1.736	1.689	1.47	0.764
b_2	-4.583	-4.239	-2.456	-0.755
b_3	5.518	4.311	1.352	-0.146
b_4	-2.003	-1.302	-0.003	-0.006
σ_ε^2	20.912	21.669	58.339	357.588
σ_b^2	24.581	9.610	3.411	1.336
R^2	0.612	0.575	0.458	0.211

Table 5. Penalized Bayesian spline regression model for different λ^*

Parameter	$1/\lambda^*=1$	$1/\lambda^*=1.164$	$1/\lambda^*=1.695$	$1/\lambda^*=2.79$
b_1	1.726	1.692	1.472	0.7641
b_2	-4.737	-4.212	-2.465	-0.7558
b_3	5.293	4.257	1.364	-0.1454
b_4	-1.876	-1.273	-0.006	-0.0057
σ_ε^2	21.169	21.734	58.125	356.454
σ_b^2	21.603	9.437	3.433	1.336
R^2	0.604	0.578	0.460	0.212

According to Table 5, small changes in λ^* have made drastic changes in smoothing of the model. So, we conclude that λ^* is more sensitive than λ. If the amount of information contained of the distribution of basis functions increases, the value of $1/\lambda^*$ decreases. It corresponds under smoothing. If the information contained of the distribution of error term decreases, the value of $1/\lambda^*$ increases. This situation corresponds over smoothing.

4. CONCLUSION

This study has been mainly motivated by the increased research activity in applied and methodological aspects of the nonparametric regression approach. We presented the three most common nonparametric regression models, which are called spline, Bayesian spline and penalized Bayesian spline, discussing advantages and disadvantages of them representations. In addition, we proposed a new smoothing parameter using the information content of normal distribution for penalized Bayesian spline regression model. The data application included in this study concerned ratios of export to import data set given in Turkish Statistical Institute (TUIK). Application is used to compare the performance of the regression models to that of the splines and different penalty terms.

When we compared the spline and Bayesian spline regression models, both models were shown similar characteristics. The coefficients of β and b parameter vectors were very similar and the coefficients of determination of two models were obtained same. But, the standard errors of parameter estimations of Bayesian spline regression were smaller than spline regression models. For this reason, we conclude that Bayesian spline regression model parameter estimation is more reliable then spline regression model. AIC is often used model selection and variable selection in Bayesian analysis. To investigate this further we computed the value of the AIC for the spline regression model and for the Bayesian spline regression model. The results show that the spline regression model provides a better fit to the data in terms of lower AIC.

We also compared penalized Bayesian spline models using different penalty terms. The different models on the same data set have been set up using different value of λ. From the results, observe that the absolute value of the coefficients of basis functions decrease as the penalty term $1/\lambda$ increase. Also, the coefficient of determination of the model gradually diminishes. Another point is that if $1/\lambda$ is large, then the effect of the knots diminishes and the model approaches to the least-squares line. The selection of the smooth parameter λ is of great importance in penalized Bayesian spline regression. The small value of λ corresponds over smoothing. The large value of corresponds under smoothing.

In this study, we proposed a new smoothing parameter using the information content of normal distribution. Under the assumption of the coefficients of basis functions are normally distributed, the new smoothing parameter (λ^*) is defined as the ratio of the information content of normal distribution. According to results, small changes in λ^* have made drastic changes in smoothing of the model. So, we conclude that λ^* is more sensitive than traditional smoothing parameter (λ). If the amount of information contained of the distribution of basis functions increases, the value of λ^* decreases. It corresponds under smoothing. If the information contained of the distribution of error term decreases, the value of $1/\lambda^*$ increases. This situation corresponds over smoothing. We conclude that the proposed smoothing parameter provides a better insight into the different levels of penalization terms that imposed the smoothing for spline curve. This can be useful for prior distribution inflection within a Bayesian inference framework.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Berry S, Carroll RJ, Ruppert D. Bayesian smoothing and regression splines for measurement error problems. Journal of the American Statistical Association. 2002;97:160–169.
2. Ruppert D. Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics. 2002;11:735-757.
3. Eubank R. Spline smoothing and nonparametric regression. Marcel Dekker, New York, USA; 1988.
4. Green P, Silverman. Nonparametric regression and generalized linear models. Chapman and Hall, London, UK; 1994.
5. Wahba G. Spline models for observational data. Society for Industrial and Applied Mathematics, Philadelphia PA, USA; 1990.
6. Friedman J. Multivariate adaptive regression splines (with discussion). The Annals of Statistics. 1991;19:1–141.
7. Hansen M, Kooperberg C. Spline adaptation in extended linear models (with discussion). Statistical Science. 2002;17:2–51.

8. Hastie T, Tibshirani R. Generalized additive models. Chapman and Hall, London, UK; 1990.
9. Eilers P, Marx B. Flexible smoothing with B-splines and Penalties. Statistical Science. 1996;11(2):89–121.
10. Ruppert D, Wand M, Carroll R. Semiparametric regression. Cambridge University Press, Cambridge, UK; 2003.
11. Robinson PM. Root-*n* consistent semiparametric regression. Econometrica (The Econometric Society). 1988;56(4):931–954.
12. Li Qi, Racine Jeffrey S. Nonparametric econometrics: Theory and Practice. Princeton University Press; 2007.
13. Racine JS, Qui L. A partially linear kernel estimator for categorical data. Unpublished Manuscript, Mcmaster University; 2007.
14. Ichimura H. Semiparametric Least Squares (SLS) and weighted SLS estimation of single index models. Journal of Econometrics. 1993;58:71–120.
15. Klein RW, Spady RH. An efficient semiparametric estimator for binary response models. Econometrica (The Econometric Society); 1993;61(2):387–421.
16. Hastie T, Tibshirani R. Varying-coefficient models. Journal of the Royal Statistical Society, Series B. 1993;55:757–796.
17. McCune B. Non-parametric habitat models with automatic interactions. Journal of Vegetation Science. 2006;17(6):819–830.
18. Grundel R, Pavlovic NB. Response of bird species densities to habitat structure and fire history along a Midwestern open–forest Gradient. The Condor. 2007;109(4):734–749.
19. Segal MR. Tree-structured methods for longitudinal data. Journal of the American Statistical Association. 1992;87(418):407–418.
20. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis (2nd edition), USA: Chapman and Hall; 2004.
21. Lawrence CM, David RC. Spline regression models, Sage Publications. 2001;1-60.
22. Akaike Hirotugu. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974;19(6):716–723.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=697&id=22&aid=6384