Available online at www.sciencedirect.com
SCIENCE DIRECT®

Statistics & Probability Letters 67 (2004) 133–140

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Improved smoothing spline regression by combining estimates of different smoothness

Thomas C.M. Lee*

Department of Statistics, Colorado State University, Fort Collins 80523-1877, CO, USA

Received October 2003

Abstract

This paper studies nonparametric regression using smoothing splines. It proposes a method that combines smoothing spline estimates of different smoothness to form a final improved estimate. This new method is straightforward to implement, computationally inexpensive, and gives reliable performances in simulations.

Keywords: Combining estimates; Nonparametric regression; Plug-in methods; Local risk estimation; Smoothing parameter; Smoothing splines

1. Introduction

This paper considers the problem of nonparametric curve estimation using smoothing splines. Suppose observed are n pairs of measurements $\{x_i, y_i\}_{i=1}^n$ satisfying the model

$$y_i = f(x_i) + \varepsilon_i, \quad a < x_1 < \cdots < x_n < b, \quad \varepsilon_i \sim \text{iid N}(0, \sigma^2),$$

where $f(x)$ is an unknown function of interest. A cubic smoothing spline estimate $\hat{f}_\lambda$ for f is defined as the minimizer of the penalized criterion

$$\frac{1}{n}\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda\int_a^b \{f''(x)\}^2\,\mathrm{d}x.$$

* Tel.: +1-970-491-5269; fax: +1-970-491-7895.
E-mail address: tlee@stat.colostate.edu (T.C.M. Lee).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2004.01.003

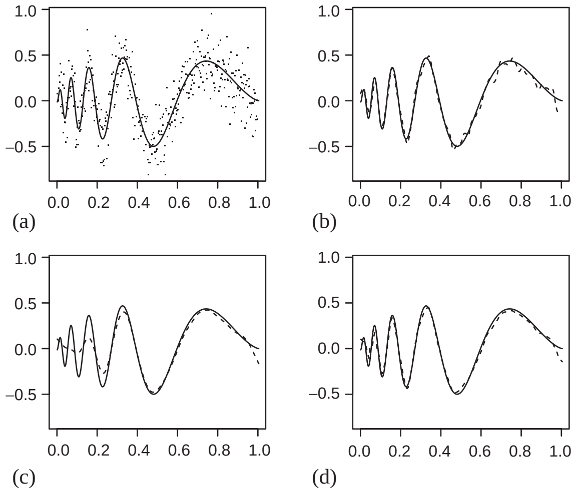

Fig. 1. (a) True regression function with noisy observations; (b) true regression function (solid line) and smoothing spline estimate (broken line) computed with a small λ; (c) similar to (b) but the smoothing spline estimate was computed using a large λ; (d) similar to (b) but the curve estimate was obtained by the proposed method.

In the above, λ is a positive constant known as the smoothing parameter. It controls the trade-off between the bias and the variance of $\hat{f}_\lambda$. For general references on smoothing splines, see, for examples, Eubank (1988), Green and Silverman (1994) and Wahba (1990). It is widely known that λ has a crucial effect on the quality of $\hat{f}_\lambda$. Popular automatic methods for choosing λ include cross-validation, generalized cross-validation, Mallows' C_p criterion and the Akaike Information Criterion. Descriptions on these methods can be found for examples in Hurvich et al. (1998), Lee (2003b) and references given therein. In addition, Cantoni and Ronchetti (2001) study the robust fitting of smoothing splines.

When the unknown function f consists of different spatial structures, it is often difficult to obtain a single $\hat{f}_\lambda$ that estimates f uniformly well across its entire domain. It is because slow-varying structures require a relatively larger λ to stabilize the variance while fast-changing structures require a relatively smaller λ to reduce the bias. Fig. 1 illustrates this point. A noisy data set, together with the corresponding f, are displayed in Fig. 1(a). Fig. 1(b) displays a $\hat{f}_\lambda$ computed with a small λ. Notice that the curve estimate captures the fast-changing structures in the left portion of f reasonably well, but undersmooths the remaining slow-varying structures. Fig. 1(c) displays a $\hat{f}_\lambda$ computed with a large λ, which demonstrates the opposite effects. Thus it is strongly desirable to develop automatic methods for combining various $\hat{f}_\lambda$'s (obtained from different λ's) together to form an improved final estimate for f. The aim of this paper is to propose such a method. Fig. 1(d) displays the final curve estimate obtained by this method.

The rest of this paper is organized as follows. The proposed method is presented in Section 2. Section 3 studies the empirical performance of the proposed method via a simulation study, while conclusions are offered in Section 4.

2. The proposed method

2.1. General idea

We first provide a brief description of the proposed method. Suppose at each design point x_i a set of different smoothing spline estimates $\hat{f}_{\lambda_1}(x_i), \ldots, \hat{f}_{\lambda_m}(x_i)$ are computed. The proposed method aims to choose the $\hat{f}_\lambda(x_i)$ that minimizes the following *local* risk calculated at x_i:

$$R_\lambda(x_i) = E\{f(x_i) - \hat{f}_\lambda(x_i)\}^2.$$

Of course, $R_\lambda(x_i)$ is an unknown quantity and hence cannot be practically minimized. To overcome this problem an estimator for $R_\lambda(x_i)$ is first constructed and $\hat{f}_\lambda(x_i)$ is then chosen as the minimizer of the resulting estimator. This procedure is repeated for all x_i's and upon completion a final combined estimate for f is obtained.

2.2. Local risk estimation

This subsection presents our method for estimating $R_\lambda(x_i)$, and we need additional notation to proceed. Let $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, $\boldsymbol{f} = (f(x_1), \ldots, f(x_n))^{\mathrm{T}}$ and $\hat{\boldsymbol{f}}_\lambda = (\hat{f}_\lambda(x_1), \ldots, \hat{f}_\lambda(x_n))^{\mathrm{T}}$. Further, let S_λ be the "hat" matrix that maps $\boldsymbol{y}$ into $\hat{\boldsymbol{f}}_\lambda : \hat{\boldsymbol{f}}_\lambda = S_\lambda \boldsymbol{y}$. One can show that $S_\lambda = (I + \lambda K)^{-1}$, where I is the identity matrix and K is a matrix depending only on $x_1, \ldots, x_n$ (see, e.g., Green and Silverman 1994, Chapter 2) Denote the ith element of the vector $S_\lambda \boldsymbol{f}$ as $(S_\lambda \boldsymbol{f})(x_i)$ and the ith diagonal element of the square matrix $S_\lambda S_\lambda^{\mathrm{T}}$ as $s_\lambda(x_i)$. The trace of a matrix A is denoted as $\mathrm{tr}(A)$.

Straightforward calculation gives the following bias-variance decomposition for $R_\lambda(x_i)$:

$$R_\lambda(x_i) = \{(S_\lambda \boldsymbol{f})(x_i) - f(x_i)\}^2 + \sigma^2 s_\lambda(x_i). \tag{1}$$

We suggest estimating $R_\lambda(x_i)$ by replacing the unknown quantities $\boldsymbol{f}$ and σ^2 in (1) with *pilot estimates* (see below). Denote the subsequent estimator as $\hat{R}_\lambda(x_i)$. Then, for each i, $f(x_i)$ can be estimated by the $\hat{f}_\lambda(x_i)$ that minimizes $\hat{R}_\lambda(x_i)$.

For the pilot estimate of $\boldsymbol{f}$, we use a smoothing spline estimate $\hat{f}_{\lambda_p}$ where the pilot smoothing parameter λ_p is chosen by the $\mathrm{AIC_c}$ method proposed by Hurvich et al. (1998). This $\mathrm{AIC_c}$ method performed very well in the simulation study conducted by Lee (2003b). For the pilot estimate of σ^2, we use

$$\hat{\sigma}^2_{\lambda_p} = \frac{\sum_{i=1}^n \{y_i - \hat{f}_{\lambda_p}(x_i)\}^2}{\mathrm{tr}(1 - S_{\lambda_p})}. \tag{2}$$

Reasons for using $\mathrm{tr}(1 - S_{\lambda_p})$ as the normalizing constant in $\hat{\sigma}^2_{\lambda_p}$ can be found for example in Green and Silverman (1994, Section 3.4). By replacing the unknown quantities in Expression (1) for $R_\lambda(x_i)$ with these pilots, our estimator for $R_\lambda(x_i)$ admits the expression

$$\hat{R}_\lambda(x_i) = \{(S_\lambda \hat{\boldsymbol{f}}_{\lambda_p})(x_i) - \hat{f}_{\lambda_p}(x_i)\}^2 + \hat{\sigma}^2_{\lambda_p} s_\lambda(x_i), \tag{3}$$

where $(S_\lambda \hat{\boldsymbol{f}}_{\lambda_p})(x_i)$ is the ith element of the vector $S_\lambda \hat{\boldsymbol{f}}_{\lambda_p}$.

The above idea of local risk estimation has been used by Lee (2003a) in the context of penalized spline regression with spatially varying penalties.

2.3. Practical implementation

The proposed method can be implemented with the following steps:

(1) For a set of pre-selected smoothing parameters $\lambda_1 < \cdots < \lambda_m$, compute the corresponding set of smoothing spline estimates: $\mathscr{F} = \{\hat{f}_{\lambda_1}, \ldots, \hat{f}_{\lambda_m}\}$.
(2) Calculate the following $\mathrm{AIC_C}$ score for each member in $\mathscr{F}$:

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

(3) The $\hat{f}_\lambda$ that gives the smallest $\mathrm{AIC_C}$ score is the (global) $\mathrm{AIC_C}$ smoothing spline estimate for f. Denote this $\hat{f}_\lambda$ as $\hat{f}_{\lambda_p}$, and calculate $\hat{\sigma}^2_{\lambda_p}$ using (2).
(4) Substitute the pilots $\hat{f}_{\lambda_p}$ and $\hat{\sigma}^2_{\lambda_p}$ into expression (1) and obtain $\hat{R}_\lambda(x_i)$.
(5) For each x_i obtain the $\hat{f}_\lambda(x_i)$ from $\mathscr{F}$ that minimizes $\hat{R}_\lambda(x_i)$. This minimizer is the final estimate for $f(x_i)$.

In the above the most computationally demanding calculations are in Step 1. Notice that these calculations are also required for obtaining many traditional (global) smoothing spline estimates (e.g., by cross-validation or $\mathrm{AIC_C}$). Therefore the computational time required for the proposed method is not much longer than those traditional methods, as the additional calculations required in Steps 3–5 are relatively minor.

3. Simulation study

This section reports the results of a simulation study that was conducted to evaluate the performances of the proposed methods together with two other smoothing spline methods existing in the literature. These two methods are the $\mathrm{AIC_C}$ method mentioned above and the RECP method studied in Lee (2003b). These two methods gave overall the best performances in the simulation study conducted by Lee (2003b). Notice that these two methods are global in nature.

The experimental setup adopted here was essentially the same as Lee (2003b). This setup, originally due to Professor Steve Marron, was designed to study the effects of changing the (i) noise level, (ii) design density, (iii) degree of spatial variation and (iv) noise variance function in an independent and effective fashion. The idea is as follows. Totally four sets of numerical experiments are to be performed. For each set of experiments, only one of the above four experimental factors (e.g., noise level) is changed while the remaining three are being kept unchanged. Within each set of experiments, the factor under consideration is changed six times, and hence there are altogether 24 different configurations. In this way the hope is that patterns can be more easily detected. The number of replications for each of the 24 configurations were 200. For completeness, the setup specification is listed in Table 1.

Table 1
Specification of the simulation setup

Factor	Generic form	Particular choices
Noise level	$y_{ij} = f(x_i) + \sigma_j \varepsilon_i$	$\sigma_j = 0.02 + 0.04(j-1)^2,\ n = 200$
Design density	$y_{ij} = f(X_{ji}) + \sigma \varepsilon_i$	$\sigma = 0.1,\ X_{ji} = F_j^{-1}(X_i),\ n = 200$
Spatial variation	$y_{ij} = f_j(x_i) + \sigma \varepsilon_i$	$\sigma = 0.2, f_j(x) = \sqrt{x(1-x)}\sin[\frac{2\pi\{1+2^{(9-4j)/5}\}}{x+2^{(9-4j)/5}}],\ n = 400$
Variance function	$y_{ij} = f(x_i) + \sqrt{v_j(x_i)}\varepsilon_i$	$v_j(x) = [0.15\{1 + 0.4(2j-7)(x-0.5)\}]^2,\ n = 200$

$j = 1, \ldots, 6$; $x_i = \frac{i-0.5}{n}$; $\varepsilon_i \sim$ iid $N(0,1)$; $f(x) = 1.5\phi(\frac{x-0.35}{0.15}) - \phi(\frac{x-0.8}{0.04})$
$\phi(u) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-u^2}{2})$; $X_i \sim$ iid Uniform[0, 1]; F_j is the Beta$(\frac{j+4}{5}, \frac{11-j}{5})$ c.d.f.

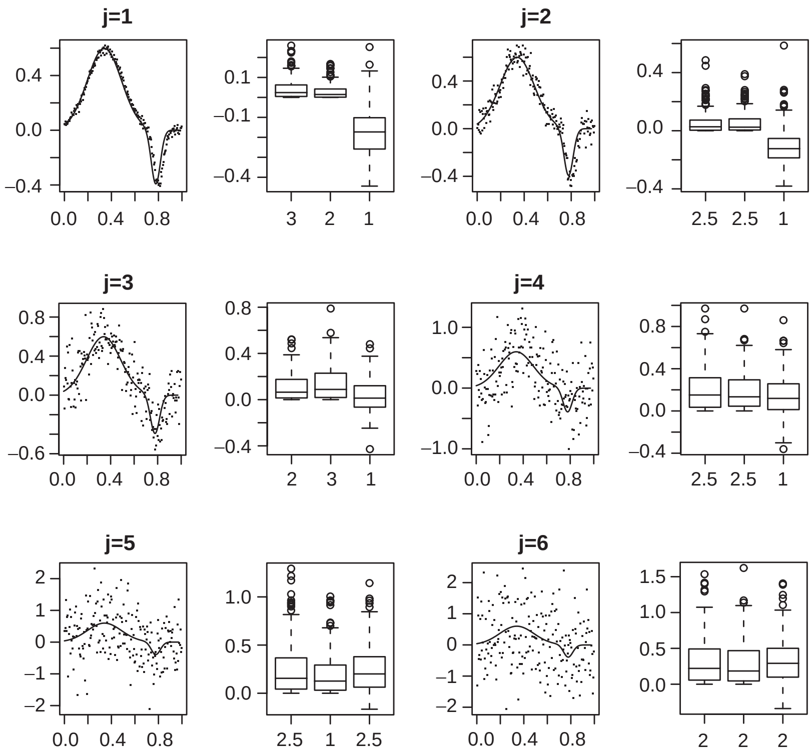

Fig. 2. Simulation results correspond to the noise level factor. In each pair of panels the left plot displays the true regression function together with one typical simulated data set. The right plot displays the boxplots of the $\log_e$ MSE values for, from left to right, AIC_C, RECP and the proposed method. The numbers below the boxplots are the paired Wilcoxon test rankings.

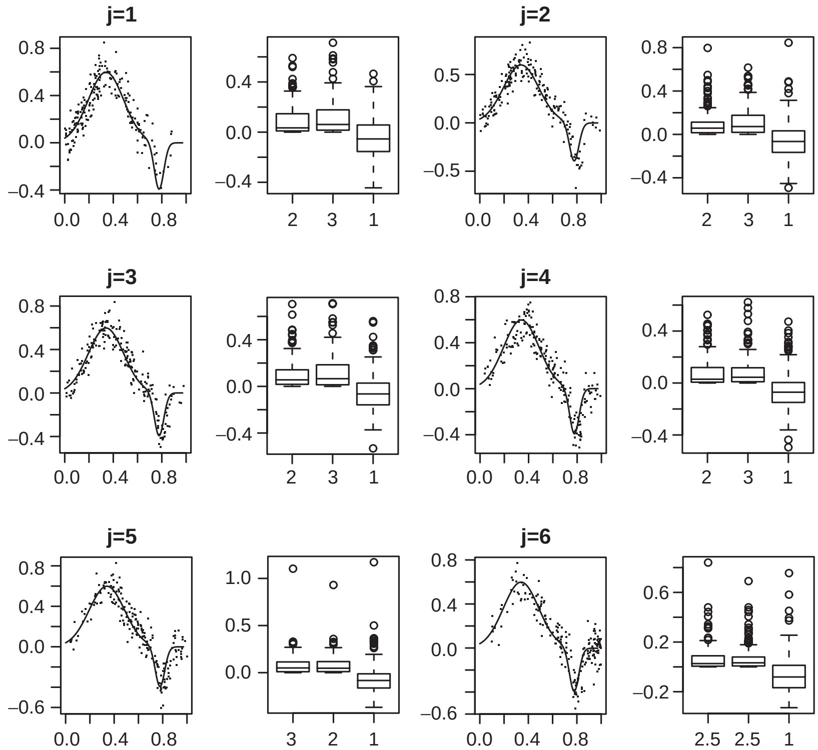

Fig. 3. Similar to Fig. 2 but for the design density factor.

For each simulated data set, we used the mean-squared-errors (MSE) to evaluate the quality of any curve estimate $\hat{f}$:

$$\mathrm{MSE} = \sum_{i=1}^{n} \{ f(x_i) - \hat{f}(x_i) \}^2.$$

Boxplots of the $\log_e$ MSE values for the 24 different configurations are given in Figs. 2–5.

Paired Wilcoxon tests were also applied to test if the difference between the median MSE values of any two methods is significant or not. The significance level used was $\frac{5}{3}\% = 1.67\%$. The methods were also ranked in the following manner. If the median MSE value of a method is significantly less than the remaining two, it will be assigned a rank 1. If the median MSE value of a method is significantly larger than one but less than the remaining one, it will be assigned a rank 2, and similarly for rank 3. Methods having non-significantly different median values will share the same averaged rank. The resulting rankings are also given in Figs. 2–5, and the averaged rankings are tabulated in Table 2.

The overall Wilcoxon test rankings for $\mathrm{AIC_C}$, RECP and the proposed method are, respectively, 2.29, 2.48 and 1.23. Therefore there seems to be some evidence to support that the proposed method is superior. In fact, out of the 24 different simulation configurations examined, the proposed method scored rank 1 for 19 configurations and shared the best rank for 3 of the 5 remaining configurations.

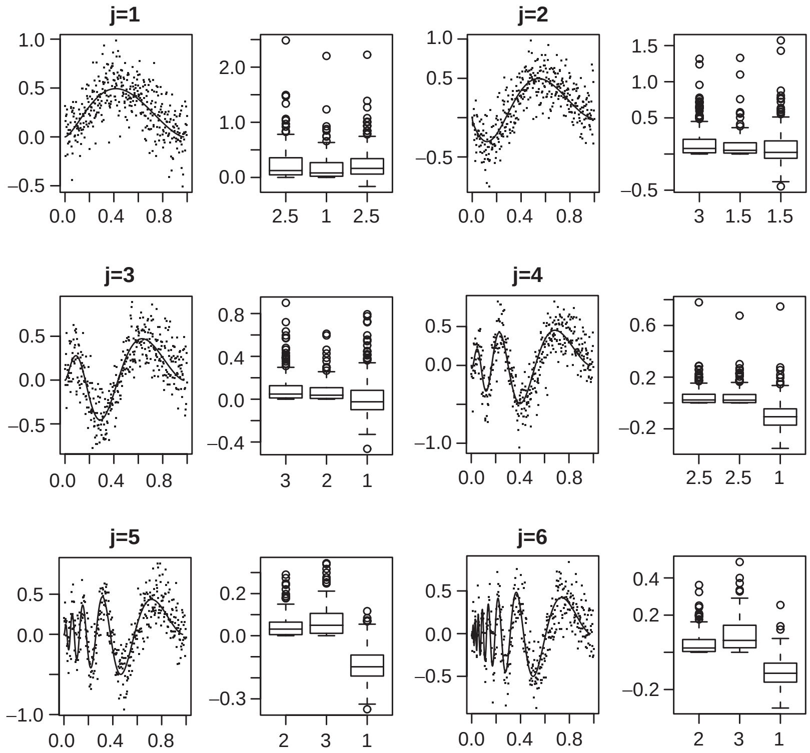

Fig. 4. Similar to Fig. 2 but for the spatial variation factor.

Table 2
Averaged Wilcoxon test rankings for the three smoothing methods

	Noise level	Design density	Spatial variation	Variance function	Overall
AIC_C	2.42	2.25	2.50	2.00	2.29
RECP	2.17	2.75	2.17	2.83	2.48
Proposed	1.42	1.00	1.33	1.17	1.23

4. Conclusion

In this paper a new method for performing smoothing spline regression is proposed. This new method combines smoothing spline estimates of different smoothness together to produce a final curve estimate. The proposed method is straightforward to implement, computationally inexpensive, and possesses superior empirical properties.

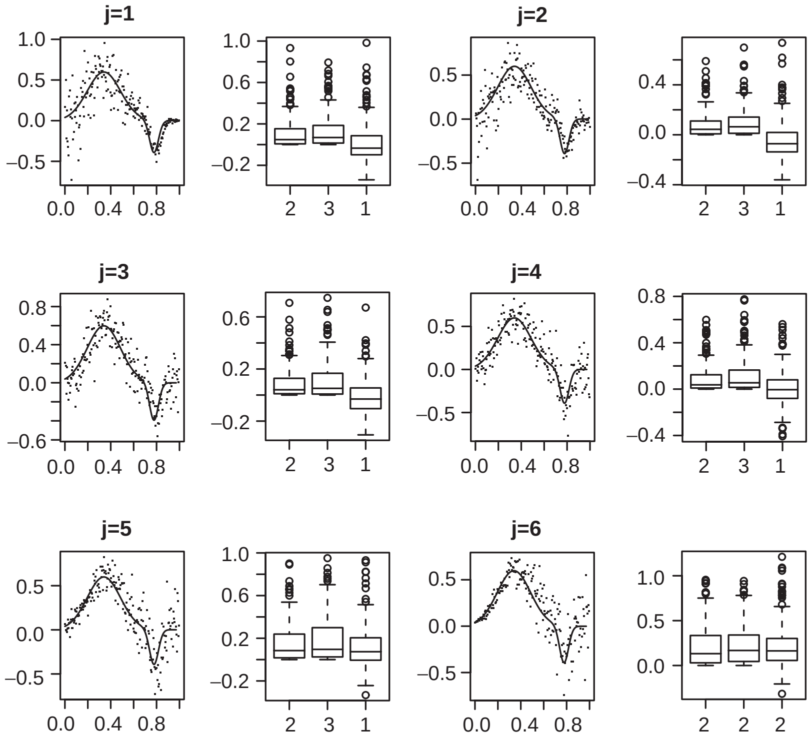

Fig. 5. Similar to Fig. 2 but for the variance function factor.

Acknowledgements

This research was supported in part by a grant from the U.S. National Science Foundation.

References

Cantoni, E., Ronchetti, E., 2001. Resistant selection of the smoothing parameter for smoothing splines. Statis. Comput. 11, 141–146.

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

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

Hurvich, C.M., Simonoff, J.S., Tsai, C.-L., 1998. Smoothing parameter selection in non-parametric regression using an improved Akaike information criterion. J. Roy. Statist. Soc. Ser. B 60, 271–293.

Lee, T.C.M., 2003a. On spline regression with spatially-adaptive penalties. Unpublished manuscript.

Lee, T.C.M., 2003b. Smoothing parameter selection for smoothing splines: a simulation study. Comput. Statist. Data Anal. 42, 139–148.

Wahba, G., 1990. Spline models for observational data. CBMS-NSF, Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.