Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 642475, 11 pages
http://dx.doi.org/10.1155/2014/642475

Research Article

Comparative Analysis for Robust Penalized Spline Smoothing Methods

Bin Wang, Wenzhong Shi, and Zelang Miao

Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Correspondence should be addressed to Wenzhong Shi; lsgiwzshi@gmail.com

Received 2 May 2014; Accepted 17 June 2014; Published 16 July 2014

Academic Editor: Yoshinori Hayafuji

Smoothing noisy data is commonly encountered in engineering domain, and currently robust penalized regression spline models are perceived to be the most promising methods for coping with this issue, due to their flexibilities in capturing the nonlinear trends in the data and effectively alleviating the disturbance from the outliers. Against such a background, this paper conducts a thoroughly comparative analysis of two popular robust smoothing techniques, the M-type estimator and S-estimation for penalized regression splines, both of which are reelaborated starting from their origins, with their derivation process reformulated and the corresponding algorithms reorganized under a unified framework. Performances of these two estimators are thoroughly evaluated from the aspects of fitting accuracy, robustness, and execution time upon the MATLAB platform. Elaborately comparative experiments demonstrate that robust penalized spline smoothing methods possess the capability of resistance to the noise effect compared with the nonrobust penalized LS spline regression method. Furthermore, the M-estimator exerts stable performance only for the observations with moderate perturbation error, whereas the S-estimator behaves fairly well even for heavily contaminated observations, but consuming more execution time. These findings can be served as guidance to the selection of appropriate approach for smoothing the noisy data.

1. Introduction

Penalized spline smoothing technique has already been perceived as a popular nonparametric smoothing approach for smoothing noisy data in engineering domain [1–3] during the past 20 years due to its ease of fitting, flexible choice of the inner knots, and smoothing parameter. The ideas of employing the combination of regression splines, with substantially smaller number of inner knots than the sample size, and the introduction of the penalized parameter, which compromises the fitting accuracy and the smoothness of the regression curve, can be traced back to at least O'Sullivan [4] who utilized a cubic B-spline basis for estimation in ill-posed inverse problems.

During the early stage, Kelly and Rice [5] and Besse et al. [6] approximated the smoothing splines by the hybrid splines equipped with the inner knots equal to the data samples and a penalty parameter for determining the smoothing amount of regression curve, which was usually calculated according to the (generalized) cross-validation criterion. Other criterions included the Mallows' C_p criterion [7] and the Akaike information criterion [8, 9]. Besides, Eilers and Marx [10] proposed flexible criteria for choosing an optimal penalty parameter of dealing with the B-splines coefficients. Wand [11] additionally derived a quick and simple formula approximation to the optimal penalty parameter for the penalized spline regression. Moreover, Ruppert and Carroll [12] investigated the spatially varying penalties and Ruppert [13] recommended the selection of designed inner knots. More discussion and examples on the penalized regression spline models can be referred to the excellent monograph written by Ruppert et al. [14].

The subsequent explorations on penalized spline smoothing remained a hot area arousing great interests from both theory researchers and practice engineers for several years. For example, the theoretical aspects and practical applications of penalized spline smoothing were fully discussed in a dissertation composed by Krivobokova [15] from financial mathematics. Crainiceanu et al. [16] investigated the implementation of nonparametric Bayesian analysis for penalized

spline regression using WinBUGS, in which the Markov Chain Monte Carlo (MCMC) mixing operator was substantially improved by employing the low rank thin plate splines basis. Besides, Marra and Radice [2] presented an accessible overview of generalized additive models based on the penalized regression splines and explored their potential applications into medical research. Generally, the penalized regression spline model is referred to the estimation of univariate smooth functions based on noisy data. However, some researchers also put forward its extension into higher dimension. Typical models include the thin plate regression splines [17] which are low rank smoothers obtained by truncating the basis of standard thin plate splines thereby requiring less computation to fit to large data sets and the penalized bivariate splines [18] with application on triangulations for irregular regions.

The regression curve for penalized spline is established by minimizing the sum of squared residuals subject to a bound on the size of the spline coefficients. Accordingly, a closed-form formula can be derived from the penalized minimization problem. Apparently, such penalized least squares regression splines cannot be resistant to the disturbance from the outliers. A simple and direct idea is to replace the squared residuals by some slower-increasing loss function similarly as employed in M-regression estimator [19, 20] for alleviating the outlier effects from the atypical observations. Early research regarding the M-type smoothing technique started from Huber et al. [21] and Cox's [22] work on cubic regression splines. Another robust model can be the S-estimation [23] which possesses high-breakdown point and minimizes the residual scale in an extremely robust manner. A fast algorithm has been created by Salibian-Barrera and Yohai [24] for computing the S-regression estimates. These two estimators have been elaborately compared in the linear regression by ÇetIn and Toka [25] with the results indicating that the S-estimator is more capable of resisting the outlier disturbance compared with the M-estimator but with lower efficiency.

Up until almost a decade ago, however, little work has been delivered taking the penalized spline regression into the category of robust estimation. Oh et al. [26] suggested a robust smoothing approach implemented to the period analysis of variable stars by simply imposing Huber's favorite loss function upon the generalized cross-validation criterion. And Lee and Oh [27] proposed an iterative algorithm for calculating the M-type estimator of penalized regression splines via introducing the empirical pseudodata. Afterwards, Finger [28] confirmed that such M-type estimator compromised between the very robust estimators and the very efficient least square type estimators for the penalized spline regressions when investigating the performance of different estimation techniques in crop insurance applications. Meanwhile, by replacing the least squares estimation with a suitable S-estimator, Tharmaratnam et al. [29] put forward an S-estimation for penalized regression spline as well, which proves to be equivalent to a weighted penalized least squares regression and behaves well even for heavily contaminated observations.

Against such a background, this paper performs a thoroughly comparative analysis mainly based on these two popular robust smoothing techniques, the M-type estimator and S-estimation for penalized regression splines, both of which are reelaborated starting from their origins, with their derivation process reformulated and the corresponding algorithms reorganized under a unified framework. The rest of the paper is organized as follows. Section 2 reelaborates the M-type estimator and S-estimation for penalized regression splines, starting from the least squares estimator and reorganizes the corresponding algorithms within a unified framework. Section 3 describes the comparative experiments upon both simulated synthetic data and one real weather balloon data set. Concluding remarks are included in the end of Section 4.

2. Robust Penalized Regression Splines

The following discussion is devoted to the reelaboration of penalized regression splines starting from the origin, with their derivation processes reformulated and the corresponding practical algorithms reorganized under a unified framework.

The idea of the penalized regression spline begins with the following model:

$$y_i \sim N\left(m\left(x_i\right), \sigma_\varepsilon^2\right), \quad i = 1, \ldots, n, \tag{1}$$

where $m(x)$ is supposed to be a smooth but unknown regression function which needs to be estimated based on the sample observations (x_i, y_i), $i = 1, \ldots, n$ and σ_ε^2 denotes the constant variance of the random deviation error between the response variable and the regression function $m(x)$.

It is a flexible concept as regards the penalized spline smoothing since different basis function may correspond to different penalized spline smoothing regression function. A common selection is the truncated polynomial bases, but other choices can also be explored straightway similarly

$$m\left(x; \boldsymbol{\beta}\right) = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{j=1}^{K} \beta_{p+j}\left(x - \kappa_j\right)_+^p, \tag{2}$$

where $(x - \kappa_j)_+ = \max(x - \kappa_j, 0)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p+K})^T$ denotes a vector of regression coefficients, $\kappa_j, \ldots, \kappa_K$ is the specified inner knots, and p indicates the exponential order for truncated power basis.

2.1. Penalized Least Squares Regression Splines. Given a group of sample observations $(x_1, y_1), \ldots, (x_n, y_n)$, an increasingly popular way for obtaining the estimation of $m(\cdot)$ is via transforming it into the category of a least squares problem, where we need to seek out the member from the class $m(x; \boldsymbol{\beta})$ which minimizes the sum of squared residuals. To prevent overfitting, a constraint on the spline coefficients is always imposed; that is,

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{p+K+1}} \sum_{i=1}^{n} \left(y_i - m\left(x_i, \boldsymbol{\beta}\right)\right)^2, \quad \text{subject to} \quad \sum_{j=1}^{K} \beta_{p+j}^2 \le C, \tag{3}$$

for some constant $C > 0$ as discussed in Ruppert et al. [14]. Thus, by employing the Lagrange multiplier, the coefficients of the penalized least squares regression spline $\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$ are equivalent to the minimizer of

$$\sum_{i=1}^{n}\left(y_i - m\left(x_i;\boldsymbol{\beta}\right)\right)^2 + \lambda\sum_{j=1}^{K}\beta_{p+j}^2, \tag{4}$$

for some smoothing (penalty) parameter $\lambda \geq 0$ with respect to $\boldsymbol{\beta}$, which compromises the bias and variance tradeoff of the fitted regression spline curve.

Introduce $F(x) = (1, x, \ldots, x^p, (x-\kappa_1)_+^p, \ldots, (x-k_K)_+^p)^T$, the spline design matrix

$$\begin{aligned}\mathbf{X} &= \left(F(x_1)^T, \ldots, F(x_n)^T\right)^T \\ &= \begin{pmatrix} 1 & x_1 & \cdots & x_1^p & (x_1-\kappa_1)_+^p & \cdots & (x_1-\kappa_K)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^p & (x_n-\kappa_1)_+^p & \cdots & (x_n-\kappa_K)_+^p \end{pmatrix},\end{aligned} \tag{5}$$

the vector of responses $\mathbf{y} = (y_1, \ldots, y_n)^T$, and the diagonal matrix $D = \mathrm{diag}(0_{(p+1)\times 1}, 1_{K\times 1})$ which indicates that only the spline coefficients are penalized. Thus, the closed-form minimizer for the objective function (4) can arrive by direct calculations; namely,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} = \left(\mathbf{X}^T\mathbf{X}+\lambda D\right)^{-1}\mathbf{X}^T\mathbf{y}. \tag{6}$$

Consequently, the corresponding estimation vector $\widehat{\mathbf{m}} = (\widehat{m}(x_1), \ldots, \widehat{m}(x_n))^T$ generated by the penalized least square regression spline can be represented as

$$\widehat{\mathbf{m}}_{\mathrm{LS}} = \mathbf{X}\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda D\right)^{-1}\mathbf{X}^T\mathbf{y}. \tag{7}$$

For figuring out the penalty parameter λ in (7), the (generalized) cross-validation technique is employed hereafter, which is computed by leave-one-out of the residual from sum of squares so as to avoid overfitting in the regression spline. The cross-validation criterion is given by

$$\mathrm{CV}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{m}_{-i}(x_i;\lambda)\right)^2, \tag{8}$$

where $\widehat{m}_{-i}(x_i;\lambda)$ is the regression spline estimated by leaving out the ith observation point. The optimized penalty parameter λ can be calculated by minimizing $\mathrm{CV}(\lambda)$ over $\lambda \geq 0$. For the linear smoothing matrix, $H_\lambda = \mathbf{X}(\mathbf{X}^T\mathbf{X} + \lambda D)^{-1}\mathbf{X}^T$, and it can be verified [14] that

$$\widehat{m}_{-i}\left(x_i;\lambda\right) = \sum_{j\neq i}\left(\frac{H_\lambda^{ij}}{1-H_\lambda^{ij}}y_i\right). \tag{9}$$

Substitution of $\widehat{m}_{-i}(x_i;\lambda)$ in (8) with the formula (9) results in

$$\mathrm{CV}(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \widehat{m}\left(x_i;\lambda\right)}{1-H_\lambda^{ii}}\right)^2. \tag{10}$$

The generalized cross-validation criterion is proposed [30] simply by replacing H_λ^{ii} in (10) with their average $\mathrm{tr}(H_\lambda)/n$; namely,

$$\mathrm{GCV}(\lambda) = n\sum_{i=1}^{n}\left(\frac{y_i - \widehat{m}\left(x_i;\lambda\right)}{n-\mathrm{tr}(H_\lambda)}\right)^2. \tag{11}$$

2.2. M-Type Estimator for Penalized Regression Splines. The above estimate $\widehat{\mathbf{m}}_{\mathrm{LS}}$ of the penalized least squares regression splines may greatly suffer from various robustness issues especially when the data is contaminated with outliers. For alleviating this effect, a penalized regression estimator can be straightforwardly constructed by replacing the previous squared residual loss function with the following M-type criterion [27]:

$$\sum_{i=1}^{n}\rho\left(y_i - m\left(x_i;\boldsymbol{\beta}\right)\right)^2 + \lambda\sum_{j=1}^{K}\beta_{p+j}^2, \tag{12}$$

where ρ is even, nondecreasing in $[0, +\infty)$ and $\rho(0) = 0$, for which a common choice is taken of the Huber loss function with cutoff $c > 0$,

$$\rho_c(t) = \begin{cases} t^2 & |t| \leq c \\ 2c|t| - c^2 & |t| > c. \end{cases} \tag{13}$$

Apparently, the above Huber's function is a parabola in the vicinity of zero, and it increases linearly at a given level $|t| > c$. A default choice of the tuning constant $c = 1.345$ aims for a 95% asymptotic efficiency with respect to the standard normal distribution.

Denote the derivative of ρ_c as $\psi_c(s) = \max[-c, \min(c, s)]$. When a set of fitted residuals $r_i = y_i - m(x_i;\boldsymbol{\beta})$, $i = 1, \ldots, n$ from the penalized spline regression are given, a robust M-scale estimator [19, 20] can be employed for estimating the standard deviation σ_ε of these residuals as follows:

$$\sum_{i=1}^{n}\psi_c\left(\frac{y_i - m\left(x_i;\boldsymbol{\beta}\right)}{\widehat{\sigma}_\varepsilon}\right) = 0. \tag{14}$$

A common choice for a measure of scale with a high finite sample breakdown point can be the value of ω determined by

$$P\left(\left|r - r_{0.5}\right| < \omega\right) = \frac{1}{2}, \tag{15}$$

where $r_{0.5}$ is the population median. Besides, the standard estimation of ω usually employs the median absolute deviation (MAD) statistic, which is defined as

$$\mathrm{MAD} = \mathrm{Median}\left\{\left|r_1 - M_r\right|, \ldots, \left|r_n - M_r\right|\right\}, \tag{16}$$

where M_r is the usual sample median of the fitted residuals and MAD is actually the sample median of the n values $|r_1 - M_r|, \ldots, |r_n - M_r|$, with its finite sample breakdown point approximately being 0.5 [20]. Suppose the fitted residuals are randomly sampled from a normal distribution; note that

Step 0. Input an initial curve estimate $\widehat{\mathbf{m}}_p^{(0)}$, termination tolerance ε, and the maximum iteration number $\text{Iter}_{\max}$. Meanwhile, set $k = 0$ and conduct the following loop iterations.
Step 1. Estimate $\widehat{\sigma}_\varepsilon$ by using MADN in (17);
Step 2. Produce the empirical pseudo data according to (19);
Step 3. Calculate the penalized least-square estimator $\widehat{\mathbf{m}}_p^{(k+1)}$ defined in (20) for the pseudo data, where the penalty parameter λ is chosen according to the generalized cross-validation (GCV) criterion formulated by (11) in the foregoing subsection.
Step 4. If $\|\widehat{\mathbf{m}}_p^{(k+1)} - \widehat{\mathbf{m}}_p^{(k)}\| < \varepsilon\|\widehat{\mathbf{m}}_p^{(k)}\|$ or $k = \text{Iter}_{\max}$, terminate and output the robust M-type penalized spline estimate $\widehat{\mathbf{m}}_M = \widehat{\mathbf{m}}_p^{(k+1)}$; else set $k = k + 1$ and continue Step 1.

Algorithm 1: M-type iterative algorithm for penalized regression splines.

MAD does not estimate the corresponding standard deviation, $\widehat{\sigma}_\varepsilon$, but rather estimates $Z_{0.75}\widehat{\sigma}_\varepsilon$ (with $Z_{0.75}$ being the 0.75 quantile of the standard normal distribution). To put MAD in a more familiar context, it typically needs to be rescaled so as to estimate $\widehat{\sigma}_\varepsilon$ especially when the residuals are sampled from a normal distribution. In particular,

$$\text{MADN} = \frac{\text{MAD}}{z_{0.75}} \approx 1.4836\,\text{MAD}. \tag{17}$$

By utilizing the aforementioned spline basis functions, an M-type penalized spline estimator using the standardized residuals can be computed as $\widehat{\mathbf{m}}_M = \mathbf{X}\widehat{\boldsymbol{\beta}}_M$, with its regression coefficients being as follows:

$$\widehat{\boldsymbol{\beta}}_M = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_c\left(\frac{y_i - m(x_i;\boldsymbol{\beta})}{\widehat{\sigma}_\varepsilon}\right) + \lambda\boldsymbol{\beta}^T D\boldsymbol{\beta}. \tag{18}$$

Due to the nonlinear nature of ρ_c and ψ_c, it is nontrivial to seek the minimizer of (18) while satisfying (14). Under such circumstances, an M-type iterative algorithm for calculating the penalized regression splines is proposed [27] by introducing the empirical pseudodata

$$\mathbf{z} = \widehat{\mathbf{m}}_p + \frac{\psi_c\left(\mathbf{y} - \widehat{\mathbf{m}}_p\right)}{2}, \tag{19}$$

where $\widehat{\mathbf{m}}_p$ is the penalized least square estimator corresponding to the above pseudodata

$$\widehat{\mathbf{m}}_p = \mathbf{X}\left(\mathbf{X}^T\mathbf{X} + \lambda D\right)^{-1}\mathbf{X}^T\mathbf{z}. \tag{20}$$

As proved in Theorem 1 of Lee and Oh's [27] work, the penalized least square estimator $\widehat{\mathbf{m}}_p$ converges to $\widehat{\mathbf{m}}_M$ asymptotically, which implies an M-type iterative algorithm consisting of the steps in Algorithm 1.

Regarding the M-type estimator for the penalized regression splines (see Algorithm 1), the initial curve estimate $\widehat{\mathbf{m}}_p^{(0)}$ in Step 0 is suggested [27], choosing the nonrobust penalized least square regression spline presented by (7).

2.3. S-Estimation for Penalized Regression Splines. The robustness properties of the aforementioned M-type estimator for penalized regression splines are not completely satisfactory since they have low breakdown point [31], which was clarified by Maronna and Yohai [32] for the general M-estimators when considering both the regression and scale, simultaneously.

Another more robust and relatively latest approach for smoothing the noisy data is the S-estimation for penalized regression splines [29], whose core idea is utilizing a flexible S-estimator to replace the traditional least squares estimator. The coefficient vector of the penalized S-regression splines $\widehat{\boldsymbol{\beta}}_S$ is defined as follows:

$$\widehat{\boldsymbol{\beta}}_S = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{n\widehat{\sigma}_\varepsilon^2(\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^T D\boldsymbol{\beta}\right\}, \tag{21}$$

in which, for each $\boldsymbol{\beta}$, $\widehat{\sigma}_\varepsilon^2(\boldsymbol{\beta})$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - F(x_i)^T\boldsymbol{\beta}}{\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})}\right) = E_\Phi(\rho(Z)), \tag{22}$$

where Φ is cumulative distribution function of the standard norm distribution, $Z \sim N(0,1)$, and the loss function is taken from Tukey's bisquare family [33]

$$\rho_d(u) = \begin{cases} 3\left(\dfrac{u}{d}\right)^2 - 3\left(\dfrac{u}{d}\right)^4 + \left(\dfrac{u}{d}\right)^6 & |u| \le d \\ 1 & |u| > d. \end{cases} \tag{23}$$

It is evidently known from theoretical assumption that each fitted residual vector should follow a normal distribution with mean of zero; that is, $\mathbf{r}_i \sim N(0,\sigma_\varepsilon^2)$, $i = 1,\ldots,n$, however its median statistic, Median($\mathbf{r}$), in which $\mathbf{r} = (r_1,\ldots,r_n)^T$, may greatly depart from zero to a large extent in the general practice. Considering this, we ameliorate the estimation of $\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})$ by using MADN in (17) with the addition of the absolute value of residual median; namely,

$$\widehat{\sigma}_\varepsilon(\boldsymbol{\beta}) = \text{MADN} + \text{Median}(\mathbf{r}(\boldsymbol{\beta})). \tag{24}$$

The penalized S-estimator for the regression spline model is computed as $\widehat{\mathbf{m}}_S = \mathbf{X}\widehat{\boldsymbol{\beta}}_S$, with its coefficients vector $\widehat{\boldsymbol{\beta}}_S$ satisfying both (21) and (22), simultaneously. The following formula derivations indicate that such penalized S-estimator can

actually be equivalent to a weighted penalized least squares regression.

Before proceeding with the derivation, it is forewarned that the symbol $\widehat{\boldsymbol{\beta}}_S$ is simply written as $\boldsymbol{\beta}$ just for clarity. Firstly, when (21) reaches its minimum, the following equation should be satisfied:

$$n\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})\nabla\widehat{\sigma}_\varepsilon(\boldsymbol{\beta}) + \lambda D\boldsymbol{\beta} = 0. \tag{25}$$

On the other hand, taking the derivative of the M-scale function in (22) with respect to $\boldsymbol{\beta}$ leads to

$$\sum_{i=1}^{n}\rho'\left(\frac{r_i(\boldsymbol{\beta})}{\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})}\right)\left(\frac{-F(x_i)^T\widehat{\sigma}_\varepsilon(\boldsymbol{\beta}) - r_i(\boldsymbol{\beta})\nabla\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})^T}{\widehat{\sigma}_\varepsilon^2(\boldsymbol{\beta})}\right) = 0, \tag{26}$$

in which $r_i(\boldsymbol{\beta}) = y_i - F(x_i)^T\boldsymbol{\beta}$, $\widetilde{r}_i(\boldsymbol{\beta}) = r_i(\boldsymbol{\beta})/\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})$, and $\Lambda(\boldsymbol{\beta}) = \operatorname{diag}(\rho'(\widetilde{r}_i(\boldsymbol{\beta}))/\widetilde{r}_i(\boldsymbol{\beta}))$, $i = 1, \ldots, n$.

Deformation of the above formula produces

$$\begin{aligned}\nabla\widehat{\sigma}_\varepsilon(\boldsymbol{\beta}) &= -\frac{\sum_{i=1}^{n}\rho'(r_i(\boldsymbol{\beta})/\widehat{\sigma}_\varepsilon(\boldsymbol{\beta}))F(x_i)}{\sum_{i=1}^{n}\rho'(r_i(\boldsymbol{\beta})/\widehat{\sigma}_\varepsilon(\boldsymbol{\beta}))(r_i(\boldsymbol{\beta})/\widehat{\sigma}_\varepsilon(\boldsymbol{\beta}))} \\ &= -\left(\frac{\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})\mathbf{X}^T\Lambda(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta})}{\mathbf{r}(\boldsymbol{\beta})^T\Lambda(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta})}\right).\end{aligned} \tag{27}$$

Substituting $\nabla\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})$ of (27) into (25),

$$-\frac{n\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})^2}{r(\boldsymbol{\beta})^T\Lambda(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta})}\mathbf{X}^T\Lambda(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta}) + \lambda D\boldsymbol{\beta} = 0. \tag{28}$$

Recall that $\mathbf{r}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ and import new symbol $\tau(\boldsymbol{\beta}) = n\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})^2/\mathbf{r}(\boldsymbol{\beta})^T\Lambda(\boldsymbol{\beta})\mathbf{r}(\boldsymbol{\beta})$; we have

$$-\tau(\boldsymbol{\beta})\mathbf{X}^T\Lambda(\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda D\boldsymbol{\beta} = 0. \tag{29}$$

And, thus, a further simple deformation arrives at the coefficient vector $\widehat{\boldsymbol{\beta}}_S$ with an iteration form

$$\boldsymbol{\beta} = \left(\mathbf{X}^T\Lambda(\boldsymbol{\beta})\mathbf{X} + \frac{\lambda}{\tau(\boldsymbol{\beta})}D\right)^{-1}\mathbf{X}^T\Lambda(\boldsymbol{\beta})\mathbf{y}, \tag{30}$$

which forms the basis of constructing an iterative algorithm for penalized S-regression spline.

The above formula derivations verify that $\widehat{\boldsymbol{\beta}}_S$ is actually the solution to the following weighted penalized least squares regression problem:

$$\min_{\boldsymbol{\beta}}\left\|\Lambda(\boldsymbol{\beta})^{1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\|^2 + \frac{\lambda}{\tau(\boldsymbol{\beta})}\boldsymbol{\beta}^T D\boldsymbol{\beta}, \tag{31}$$

And such representation indicates that the GCV criterion in (11) can be employed straightforwardly with response variable $\widetilde{\mathbf{y}} = \Lambda(\boldsymbol{\beta})^{1/2}\mathbf{y}$, weighted design matrix $\widetilde{\mathbf{X}} = \Lambda(\boldsymbol{\beta})^{1/2}\mathbf{X}$, and the penalty term $\lambda/\tau(\boldsymbol{\beta})$. Meanwhile, due to some weight elements being zero, the penalty parameter λ is thus calculated based on the following Regularized GCV criterion:

$$\operatorname{RGCV}(\lambda) = n_w\frac{\left\|\Lambda(\boldsymbol{\beta})^{1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\|^2}{\left(n_w - \operatorname{tr}\left(\widetilde{H}_\lambda\right)\right)^2}, \tag{32}$$

where $\widetilde{H}_\lambda = \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}} + (\lambda/\tau(\boldsymbol{\beta}))D)^{-1}\widetilde{\mathbf{X}}^T = \Lambda(\boldsymbol{\beta})^{1/2}\mathbf{X}(\widetilde{\mathbf{X}}^T\Lambda(\boldsymbol{\beta})\mathbf{X} + (\lambda/\tau(\boldsymbol{\beta}))D)^{-1}\mathbf{X}^T\Lambda(\boldsymbol{\beta})^{1/2}$ and n_w denotes the number of nonzero elements in weight matrix $\Lambda(\boldsymbol{\beta})$.

Consequently, based on the foregoing in-depth discussions, the S-estimator iterative algorithm for penalized regression splines can be constructed as in Algorithm 2.

In particular, it should be pointed out that since the employed loss function ρ_d is bounded giving rise to the nonconvexity of $\widehat{\sigma}_\varepsilon(\boldsymbol{\beta})$, the objective function in (21) may have several critical points, each of which only corresponds to one local minima. Therefore, the convergent critical point of the iterations derived from the above algorithm completely depends on the prescribed initial regression coefficients vector, $\widehat{\boldsymbol{\beta}}^{(0)}$, in Step 0 of Algorithm 2.

One possible solution to cope with this issue proposed by Tharmaratnam et al. [29] is based on the random sampling theory, which can be briefly described as, firstly, creating J random subsamples with a proper size, for example, one fifth of the sample number; namely, $n_s = \max(p + K + 1, \lfloor n/5 \rfloor)$, from the original sample observations. Each subsample brings about an initial candidate of the coefficient vector, for instance, assigned by least squares fitting for the penalized regression splines or the M-type estimator with a small amount of additional computation cost, with the initial regression coefficient vector of the jth subsample being denoted by $\widehat{\boldsymbol{\beta}}_j^{(0)}$, staring from which the final iterative convergent vector produced through Algorithm 2 is being written as $\widehat{\boldsymbol{\beta}}_j^{\star}$, $j = 1, \ldots, J$. Eventually, the global optimum is taken among all these J convergent vectors to be the one corresponding to the minimum objective function in (21); namely,

$$\widehat{\boldsymbol{\beta}}_S^{\star} = \operatorname*{argmin}_{1\le j\le J}\left\{n\widehat{\sigma}_\varepsilon^2\left(\widehat{\boldsymbol{\beta}}_j^{\star}\right) + \lambda\widehat{\boldsymbol{\beta}}_j^{\star T}D\widehat{\boldsymbol{\beta}}_j^{\star}\right\}. \tag{33}$$

Therefore, the robust S-estimator for penalized regression splines can be computed as $\widehat{\mathbf{m}}_S = \mathbf{X}\widehat{\boldsymbol{\beta}}_S^{\star}$.

3. Experimental Performance Evaluation

This section is devoted to the comparative experiments upon both simulated synthetic data and one real weather balloon data set for the foregoing explored penalized spline smoothing methods, including the nonrobust penalized least square regression spline (LS) presented by (7) and two robust smoothing approaches, the M-type estimator [27] and S-estimation [29], for penalized regression splines as described in Algorithms 1 and 2.

Step 0. Input an initial regression coefficients $\widehat{\boldsymbol{\beta}}^{(0)}$, the termination tolerance ε, and the maximum iteration number $\text{Iter}_{\max}$. Meanwhile, set $k = 0$ and enter into the following loop iterations.

Step 1. Estimate $\widehat{\sigma}_\varepsilon$ based on (24), that is

$$\widehat{\sigma}_\varepsilon\left(\widehat{\boldsymbol{\beta}}^{(k)}\right) = \text{MADN} + \text{Median}\left(\mathbf{r}\left(\widehat{\boldsymbol{\beta}}^{(k)}\right)\right).$$

Step 2. Compute the intermediate variables $\tau(\widehat{\boldsymbol{\beta}}^{(k)})$ and $\Lambda(\widehat{\boldsymbol{\beta}}^{(k)})$.

Step 3. Calculate the coefficient vector $\widehat{\boldsymbol{\beta}}^{(k+1)}$ according to (30) as

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left(\mathbf{X}^T\Lambda\left(\widehat{\boldsymbol{\beta}}^{(k)}\right)\mathbf{X} + \frac{\lambda}{\tau(\widehat{\boldsymbol{\beta}}^{(k)})}D\right)^{-1}\mathbf{X}^T\Lambda\left(\widehat{\beta}^{(k)}\right)\mathbf{y}.$$

Hereinto, the penalty parameter λ is figured out based on the regularized GCV criterion (32).

Step 4. If $\|\widehat{\boldsymbol{\beta}}^{(k+1)} - \widehat{\boldsymbol{\beta}}^{(k)}\| < \varepsilon\|\widehat{\boldsymbol{\beta}}^{(k)}\|$ or $k = \text{Iter}_{\max}$, terminate and output the convergent coefficient vector for the penalized regression spline of S-estimator $\widehat{\boldsymbol{\beta}}^{\star} = \widehat{\boldsymbol{\beta}}^{(k+1)}$; else accumulate the counter by setting $k = k + 1$ and repeat Step 1.

Algorithm 2: S-estimator iterative algorithm for penalized regression splines.

3.1. Configuration and Criteria. It should be incidentally mentioned that the exponential order $p = 3$ for truncated power basis is always employed for all these three methods, and the knot locations are selected in accordance with the recommended choice by Ruppert et al. [14], $\kappa_k = ((k+1)/(K+2))$th sample quantile of the Unique($\mathbf{x}$), $k = 1, \ldots, K$; together with a simple choice of the knot number $K = \min(35, (1/4) \times \text{Number of Unique}(\mathbf{x}))$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$. Besides, the corresponding penalty parameter λ is determined according to the generalized cross-validation (GCV) criterion for the LS fitting and M-estimation methods and the regularized GCV criterion for the S-estimation method. Meanwhile, the termination tolerance $\varepsilon = 10^{-6}$ and the maximum iteration number of 100 are assigned for the initial input parameters in Algorithms 1 and 2 and $J = 5$ subsamples are randomly generated for exploring the ultimate optimum.

Meanwhile, the hardware equipment and software environment established for the experiments below are Dell XPS 8700 Desktop (CPU: Intel Core i7-4770 @ 3.40 GHz), MATAB R2014a installed upon Windows 7 SP1 Ultimate (64 Bit).

3.2. Synthetic Data. The simulated study is designed the same as in Lee and Oh [27] with the testing function

$$y_i = \sin\left(2\pi\left(1 - x_i\right)^2\right) + 0.5\varepsilon_i, \quad i = 1, \ldots, n. \tag{34}$$

The designed samples are generated with their horizontal axes being uniformly extracted from the open interval $(0, 1)$ and the corresponding error distribution ε employing nine possible scenarios with the probability distribution function (pdf) specified if necessary, including (1) uniform distribution, $U(-1, 1)$, (2) standard normal distribution, $N(0, 1)$, (3) logistic distribution with its pdf as $e^x/(1 + e^x)^2$, (4) double exponential distribution with its pdf as $((1/2)e^{-|x|})$, (5) contaminated normal distribution I, $0.95N(0, 1) + 0.05N(0, 10)$, that is the error constituted with 95% from $N(0, 1)$ and 5% from $N(0, 10)$, (6) contaminated normal distribution II, $0.9N(0, 1) + 0.1N(0, 10)$, (7) slash distribution, defined from two statistically independent random variables following (1) and (2), that is, $N(0, 1)/U(0, 1)$, (8) Cauchy distribution with the pdf as $1/\pi(1 + x^2)$, whose simulated outliers can be generated from uniform distribution by inverse transformation technique [34], and (9) an asymmetric normal mixture distribution, $0.9N(0, 1) + 0.1N(30, 1)$. For the sake of reducing the simulation variability, the sample size is taken as a fixed value, $n = 200$. It should be mentioned incidentally that the above (1)~(8) error scenarios are symmetrically distributed and provided with the ascending order according to the heaviness of their tails, and scenario (9) denotes the standard normal distribution mixed with 10% asymmetric outliers from $N(30, 1)$.

Figure 1 presents the spline regression curves from the penalized LS fitting, M-estimator, and S-estimator for one time simulated scattered samples in nine possible scenarios, each of which with a subfigure panel demonstrating the comparative regression results for these three penalized spline smoothing methods.

For proceeding further quantitative evaluation of the comparative analysis for these three penalized spline smoothing methods, the statistical indicator of average squared error (ASE) is utilized here to qualify the goodness of fitting between the regression spline $\widehat{m}_j(x)$ for the jth simulation replicate from each of the above outlier scenarios and the true simulated function $m(x) = \sin(2\pi(1 - x)^2)$

$$\text{ASE}_j = \frac{1}{n}\sum_{i=1}^{n}\left(m\left(x_i\right) - \widehat{m}_j(x_i)\right)^2, \quad j = 1, \ldots, J, \tag{35}$$

with the specified replicate number $J = 10^3$ for each distribution.

Performances of the penalized spline smoothing methods are thoroughly evaluated from the aspects of the fitting

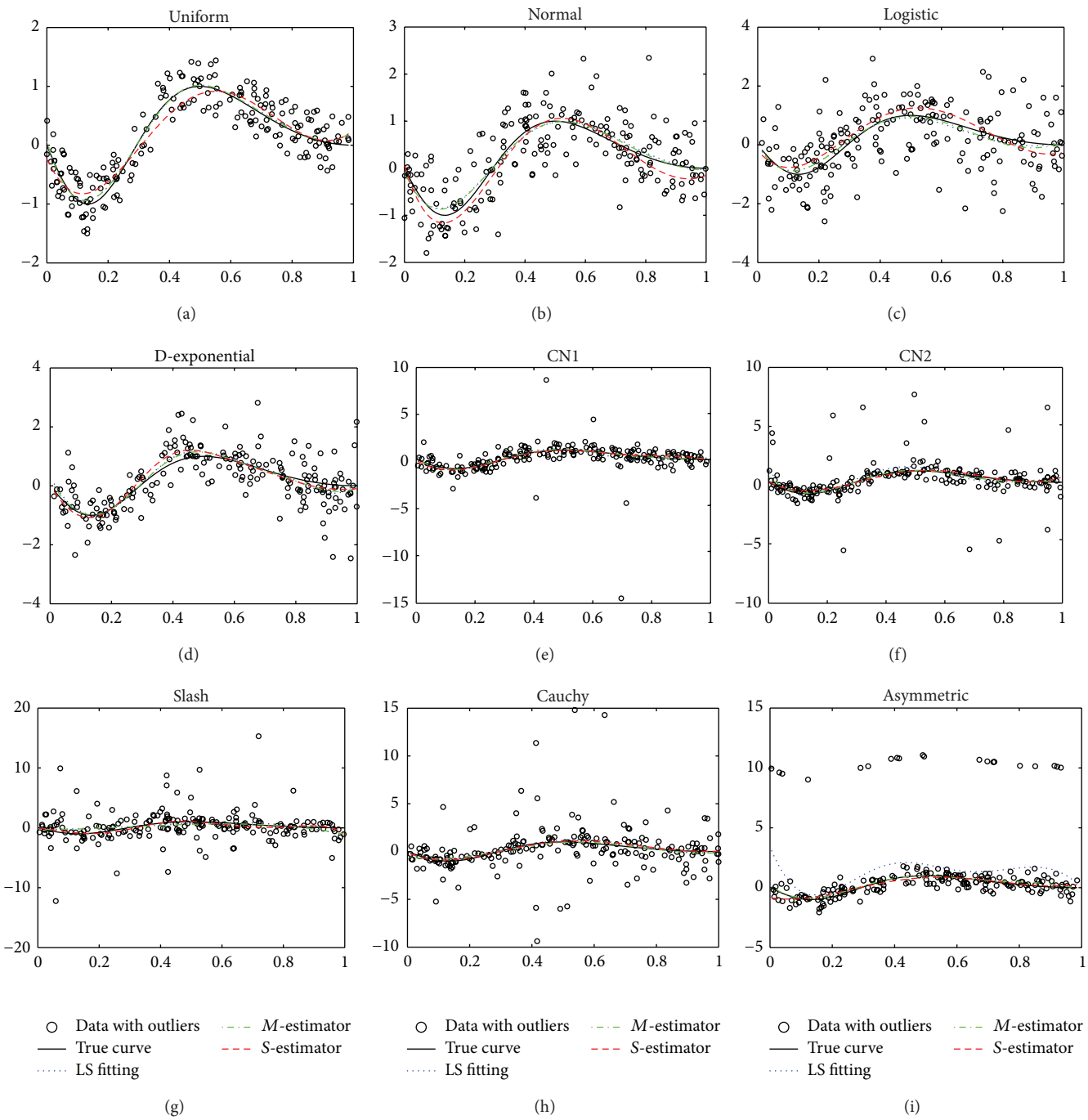

Figure 1: Spline regression curves (with true testing function in black solid line) from the penalized LS fitting (blue dotted), M-estimator (green dot-dashed), and S-estimator (red dashed) for the simulated samples (black circle-dot) in 9 possible scenarios.

accuracy, robustness, and execution time. In the following, Figures 2 and 3 illustrate the box plots of both ASEs and code running times using (a) penalized LS-estimation, (b) penalized M-estimation, and (c) penalized S-estimation for the spline smoothing methods, as well as their median value comparisons of these two indicators for the simulated samples from each of the 9 distribution scenarios.

From the ASE comparisons in Figure 2, it can be found that the penalized LS-estimation behaves well for the symmetric error scenarios with small tail weights, such as the simple uniform and standard normal distribution. However, it cannot cope with the complex conditions such as the three latter error distributions possessing heavy tails, thereby binging about large ASEs which are specially rescaled to

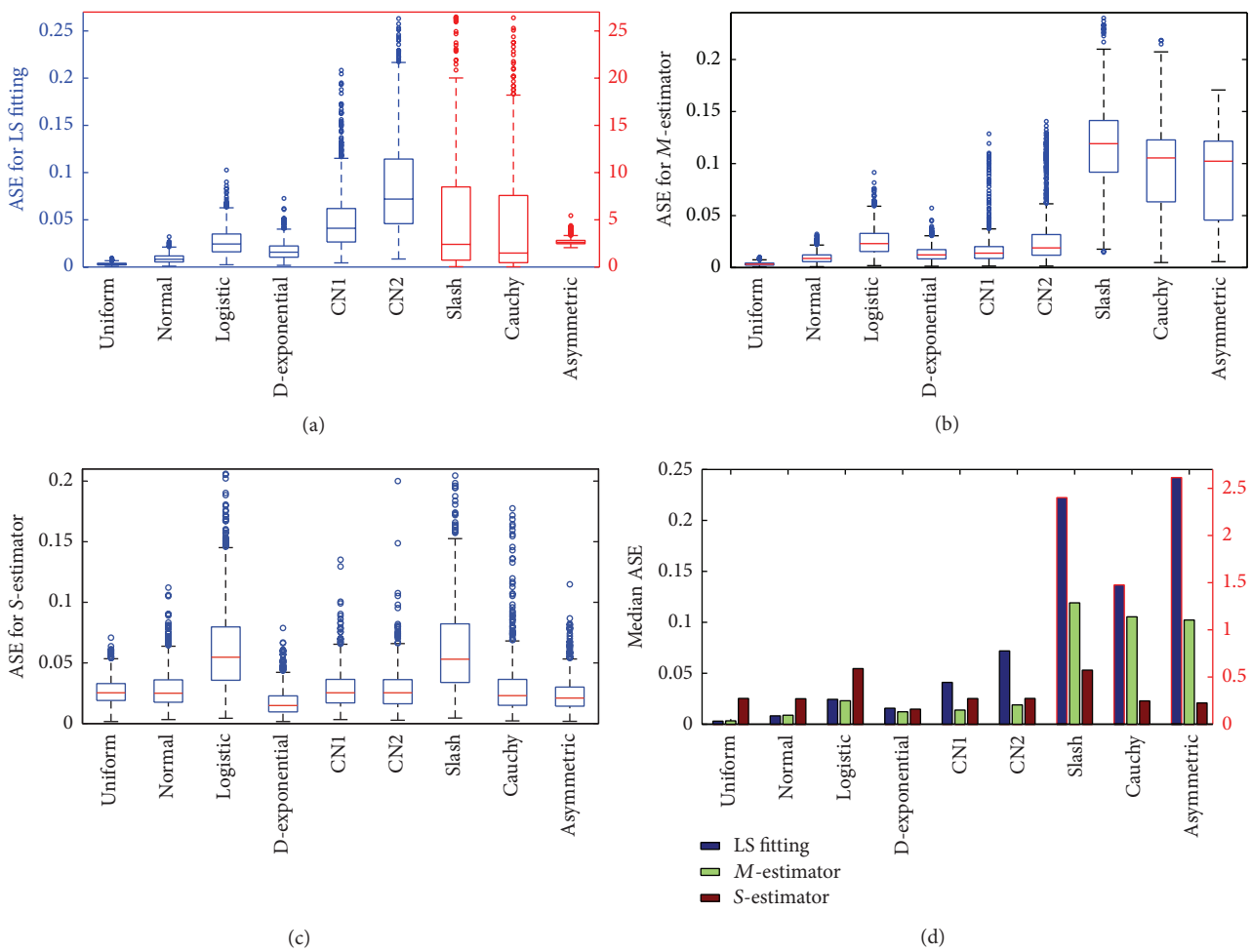

Figure 2: Box plots of ASEs using (a) penalized LS-estimation, (b) penalized M-estimation, (c) penalized S-estimation, and (d) the comparison of Median ASE for the simulated samples from each of the 9 distribution scenarios with replicate number $J = 10^3$.

the additive vertical coordinate on the right so as to be accommodated with the ASEs results for other error sources in the same window (refer to panels (a) and (d) in Figure 2). In comparison, the robust penalized M-type estimation presents satisfactory regression fittings for the distributions with moderate tail weights like (2)~(5) scenarios, which are located at the middle in each panel. Meanwhile, the robust S-estimator provides the comparatively meritorious smoothing results especially for the complicated circumstances like slash, Cauchy, or the asymmetric distributions with heavy tails.

In terms of the robustness, LS-estimator simply does not have the qualifications to be mentioned from this aspect. Instead, the M-estimator demonstrates stable performance for the designed samples with moderate perturbation error and the S-estimator behaves dominantly stable for all kinds of error source scenarios.

Unfortunately, as for the execution efficiency, the most time-elapsed one is the S-estimator, followed by the M-estimator, with the least consuming estimator being the LS-estimator since it has been treated as the initial curve estimate for the two former approaches. Moreover, although both S-estimator and M-estimator are composed of similar iterative process, the random sampling theory is further involved in the former model for exploring the ultimate solution for a nonconvex optimization problem.

3.3. Balloon Data. This section is devoted to the comparative experiments of the aforementioned penalized spline smoothing methods upon a set of weather balloon data [35]. The data consist of 4984 observations collected from a balloon at about 30 kilometers altitude, with the outliers caused due to the fact that the balloon slowly rotates, causing the ropes from which the measuring instrument is suspended to cut off the direct radiation from the sun.

As illustrated above, Figure 4 displays these balloon observational data together with the regression fitting curves by the penalized LS method, the robust penalized M-estimation method, and the robust S-estimation method. Besides, it can also be explicitly discovered by contrast that the non-robust LS regression curve suffers from the disturbance of the outliers, and this phenomenon seems even more significantly around the position of horizontal axis being 0.8, while the other two robust splines pass through approximately the majority of the observations. Meanwhile,

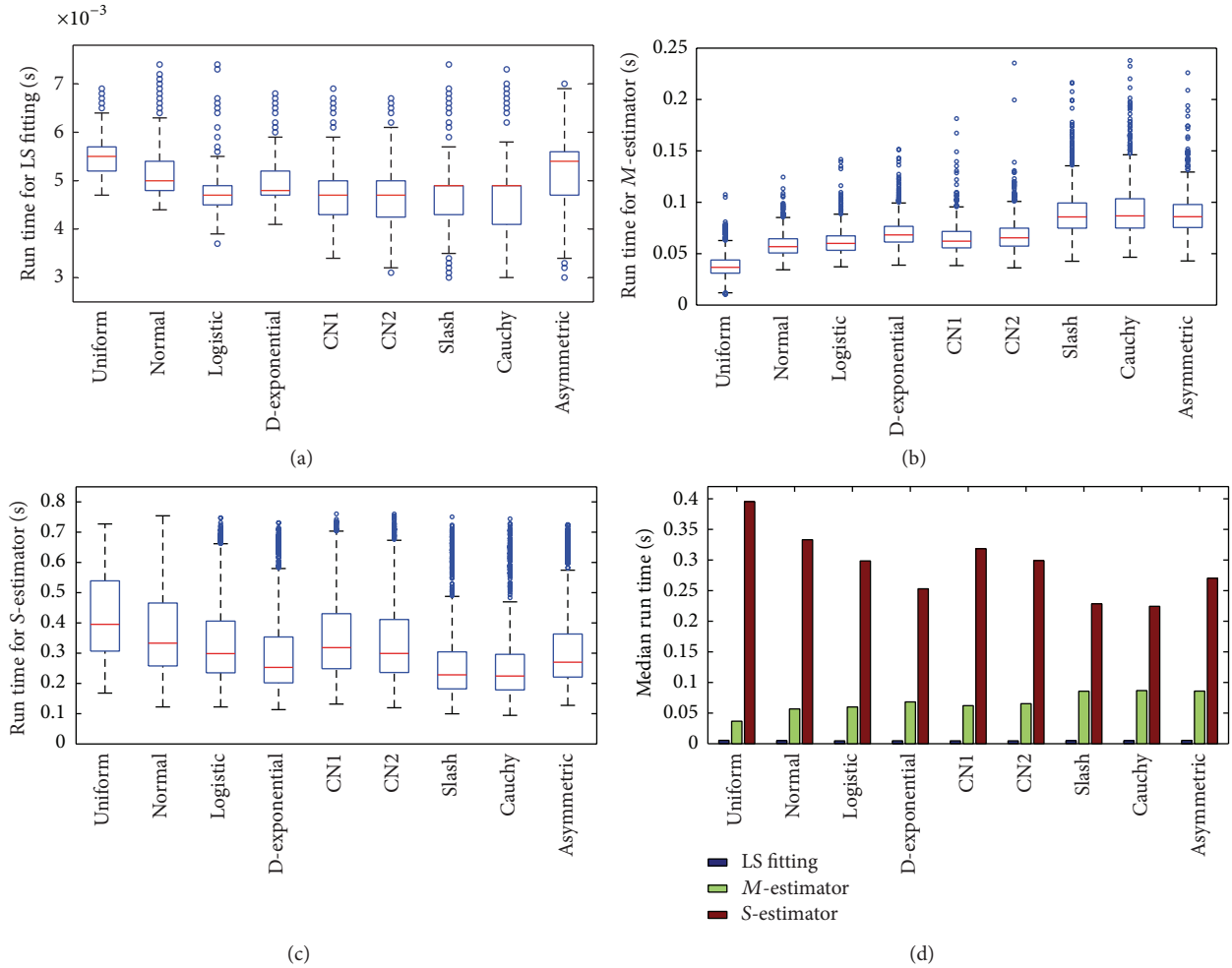

FIGURE 3: Box plots of run times using (a) penalized LS-estimation, (b) penalized M-estimation, (c) penalized S-estimation, and (d) the comparison of median run time for the simulated samples from each of the 9 distribution scenarios with replicate number $J = 10^3$.

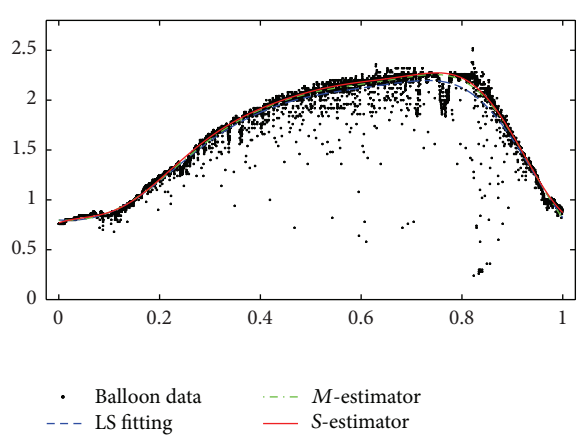

FIGURE 4: Regression fitted curves for the balloon data using penalized spline smoothing models including the penalized LS fitting (blue dashed), M-estimator (green dot-dashed), and S-estimator (red solid line).

the S-estimator seems to behave slightly more superior to the M-estimator since the former accords with the data trend better especially at the ridge and two terminal nodes.

At the same time, under the platform specified in Section 3.1, the algorithm execution times of this weather balloon data set with 4984 observations are 2.3 s, 35.1 s, and 58.9 s for the penalized LS regression, the penalized M-type estimator, and the penalized S-estimator, respectively.

4. Conclusions

This paper conducts a comprehensively comparative analysis of the current two popular robust penalized spline smoothing methods, the penalized M-type estimator and the penalized S-estimation, both of which are reelaborated starting from their origins, with their derivation process reformulated and the ultimate algorithms reorganized under a unified framework.

Experiments consist of firstly one simulated synthetic data set with the outliers in 9 possible scenarios and then a practical weather balloon data set for the aforementioned penalized regression spline, whose performances are thoroughly evaluated from the aspects of fitting accuracy, robustness, and execution time.

Conclusions from these experiments are that robust penalized spline smoothing methods can be resistant to the

influence of outliers compared with the nonrobust penalized LS spline fitting method. Further, the M-estimator can exert stable performance only for the observed samples with moderate perturbation error while the S-estimator behaves dominantly well even for the cases with heavy tail weights or higher percentage of contaminations, but necessary with more executive time. Therefore, taking both the solving accuracy and implementation efficiency of the fitted regression spline into account, M-type estimator is generally preferred, whereas if only the fitting credibility is of concern, then the S-estimator naturally deserves to become the first choice.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

Mr. Bin WANG is the beneficiary of a doctoral grant from the **AXA Research Fund**. Besides, the authors acknowledge the funding support from the Ministry of Science and Technology of China (Project nos. 2012BAJ15B04 and 2012AA12A305) and the National Natural Science Foundation (Project no. 41331175).

References

[1] J. H. Friedman and B. W. Silverman, "Flexible parsimonious smoothing and additive modeling," *Technometrics*, vol. 31, no. 1, pp. 3–21, 1989.

[2] G. Marra and R. Radice, "Penalised regression splines: theory and application to medical research," *Statistical Methods in Medical Research*, vol. 19, no. 2, pp. 107–125, 2010.

[3] D. Ruppert and R. J. Carroll, "Spatially-adaptive penalties for spline fitting," *Australian and New Zealand Journal of Statistics*, vol. 42, no. 2, pp. 205–223, 2000.

[4] F. O'Sullivan, "A statistical perspective on ill-posed inverse problems," *Statistical Science*, vol. 1, no. 4, pp. 502–518, 1986.

[5] C. Kelly and J. Rice, "Monotone smoothing with application to dose-response curves and the assessment of synergism," *Biometrics*, vol. 46, no. 4, pp. 1071–1085, 1990.

[6] P. C. Besse, H. Cardot, and F. Ferraty, "Simultaneous nonparametric regressions of unbalanced longitudinal data," *Computational Statistics & Data Analysis*, vol. 24, no. 3, pp. 255–270, 1997.

[7] C. L. Mallows, "Some comments on Cp.," *Technometrics*, vol. 15, no. 4, pp. 661–675, 1973.

[8] H. Akaike, "Maximum likelihood identification of Gaussian autoregressive moving average models," *Biometrika*, vol. 60, no. 2, pp. 255–265, 1973.

[9] C. M. Hurvich, J. S. Simonoff, and C. Tsai, "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion," *Journal of the Royal Statistical Society B*, vol. 60, no. 2, pp. 271–293, 1998.

[10] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B-splines and penalties," *Statistical Science.*, vol. 11, no. 2, pp. 89–102, 1996.

[11] M. P. Wand, "On the optimal amount of smoothing in penalised spline regression," *Biometrika*, vol. 86, no. 4, pp. 936–940, 1999.

[12] D. Ruppert and R. J. Carroll, "Spatially-adaptive penalties for spline fitting," *Australian & New Zealand Journal of Statistics*, vol. 42, no. 2, pp. 205–223, 2000.

[13] D. Ruppert, "Selecting the number of knots for penalized splines," *Journal of Computational and Graphical Statistics*, vol. 11, no. 4, pp. 735–757, 2002.

[14] D. Ruppert, M. P. Wand, and R. J. Carroll, *Semiparametric Regression*, Cambridge University Press, Cambridge, Mass, USA, 2003.

[15] T. Krivobokova, *Theoretical and Practical Aspects of Penalized Spline Smoothing*, Bielefeld University, 2006.

[16] C. M. Crainiceanu, D. Ruppert, and M. P. Wand, "Bayesian analysis for penalized spline regression using WinBUGS," *Journal of Statistical Software*, vol. 14, no. 14, 2005.

[17] S. N. Wood, "Thin plate regression splines," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 65, no. 1, pp. 95–114, 2003.

[18] L. Zhou and H. Pan, "Smoothing noisy data for irregular regions using penalized bivariate splines on triangulations," *Computational Statistics*, vol. 29, no. 1-2, pp. 263–281, 2014.

[19] P. J. Huber, "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[20] R. R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*, Academic Press, New York, NY, USA, 2012.

[21] P. J. Huber, E. Ronchetti, and MyiLibrary, *Robust Statistics*, Wiley Online Library, 1981.

[22] D. D. Cox, "Asymptotics for M-type smoothing splines," *The Annals of Statistics*, vol. 11, no. 2, pp. 530–551, 1983.

[23] P. Rousseeuw and V. Yohai, "Robust regression by means of S-estimators," in *Robust and Nonlinear Time Series Analysis*, vol. 26 of *Lecture Notes in Statistics*, pp. 256–272, Springer, New York, NY, USA, 1984.

[24] M. Salibian-Barrera and V. J. Yohai, "A fast algorithm for S-regression estimates," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 414–427, 2006.

[25] M. ÇetIn and O. Toka, "The comparing of S-estimator and M-estimators in linear regression," *Gazi University Journal of Science*, vol. 24, no. 4, pp. 747–752, 2011.

[26] H. Oh, D. Nychka, T. Brown, and P. Charbonneau, "Period analysis of variable stars by robust smoothing," *Journal of the Royal Statistical Society C: Applied Statistics*, vol. 53, no. 1, pp. 15–30, 2004.

[27] T. C. M. Lee and H. Oh, "Robust penalized regression spline fitting with application to additive mixed modeling," *Computational Statistics*, vol. 22, no. 1, pp. 159–171, 2007.

[28] R. Finger, "Investigating the performance of different estimation techniques for crop yield data analysis in crop insurance applications," *Agricultural Economics*, vol. 44, no. 2, pp. 217–230, 2013.

[29] K. Tharmaratnam, G. Claeskens, C. Croux, and M. Salibián-Barrera, "S-estimation for penalized regression splines," *Journal of Computational and Graphical Statistics*, vol. 19, no. 3, pp. 609–625, 2010.

[30] P. Craven and G. Wahba, "Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation," *Numerische Mathematik*, vol. 31, no. 4, pp. 377–403, 1979.

[31] D. L. Donoho and P. J. Huber, "The notion of breakdown point," in *A Festschrift for Erich L. Lehmann*, pp. 157–184, 1983.

[32] R. A. Maronna and V. J. Yohai, "The breakdown point of simultaneous general M estimates of regression and scale," *Journal of the American Statistical Association*, vol. 86, no. 415, pp. 699–703, 1991.

[33] A. E. Beaton and J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics*, vol. 16, no. 2, pp. 147–192, 1974.

[34] S. M. Ross, *A First Course in Probability*, Pearson Prentice Hall, Saddle River, NJ, USA, 8th edition, 2010.

[35] L. Davies and U. Gather, "The identification of multiple outliers," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 782–801, 1993.