Semiparametric penalized spline regression

Takuma Yoshida[1] and Kanta Naito[2*]
[1]*Graduate School of Science and Engineering* and [2]*Department of Mathematics, Shimane University*

Abstract

In this paper, we propose a new semiparametric regression estimator by using a hybrid technique of parametric approach and nonparametric penalized spline method. The main shape of true regression function is captured by the parametric part, and its residual is consistently estimated by nonparametric part. Asymptotic theory for the proposed semiparametric estimator is developed, which shows that its behavior is depending on the asymptotics for nonparametric penalized spline estimator as well as the discrepancy between the true regression function and the parametric part. As a naturally associated application of asymptotics, some criteria for the selection of parametric models are addressed. Numerical experiments show that the proposed estimator performs better than kernel-based existing semiparametric estimator and fully nonparametric estimator, and the proposed criteria work well for choosing a reasonable parametric model.

key words: Asymptotic theory; Bias reduction; B-spline; Parametric model; Penalized spline; Semiparametric regression

1 Introduction

There have been several nonparametric smoothing techniques in regression problem, such as lowess, kernel smoothing, spline smoothing, wavelet, series method and so on. The nonparametric estimators have consistency in general, which is an advantage of nonparametric approach. Hence if the nonparametric estimator is used, we can expect that the true regression can be captured as sample size increases. However because the form of nonparametric estimator is sometime complicated, the interpretation of estimated structure might not be clear.

On the other hand, in the parametric regression problem with the true regression function controlled by a finite dimensional parameter vector, the estimated structure is easy to understand, although the estimator does not have consistency. Therefore each of parametric method and nonparametric method has advantage and disadvantage. This motivates us to consider a fine hybrid of parametric and nonparametric methods for regression problem, and we in fact introduce a semiparametric regression method so that the estimator has the advantage of both parametric and nonparametric approaches.

The semiparametric method in this paper consists of two steps. In the first step we utilize an appropriate parametric estimator. And then, in the second step, we apply a certain nonparametric smoother to the residual data associated with the parametric estimator in the first step.

*E-mail: `naito@riko.shimane-u.ac.jp`

The parametric estimator in the first step and the nonparametric smoother in the second step will be combined into the proposed semiparametric estimator.

Similar semiparametric approaches for smoothing have been developed by many authors. Hjort and Glad (1995) and Naito (2004) discussed similar methods in density estimation literature. Glad (1998) and Naito (2002) addressed the semiparametric regression method, Martins et al. (2008) introduced general decomposition including additive and multiplicative correction in regression. And recently Fan et al. (2009) discussed the semiparametric approach in the framework of generalized linear model. Note that all works mentioned above used kernel smoothing in the second step estimation. Our proposal is to utilize the penalized spline method for residual smoothing in the second step. Throughout the paper, the fully nonparametric penalized spline estimator is designated by NPSE, and the semiparametric penalized spline estimator including two steps manipulations mentioned above is denoted by SPSE.

2 Semiparametric penalized spline estimator

Consider the relationship of dataset $\{(x_i, y_i) : i = 1, \cdots, n\}$ as the regression model

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \cdots, n,$$

where the explanatory x_i is generated from density $q(x)$ which supports $[0, 1]$, $f(x) = E[Y|X = x]$ is an unknown regression function and the errors ε_i are assumed to be uncorrelated with $E[\varepsilon_i|X_i = x_i] = 0$ and $V[\varepsilon_i|X_i = x_i] = \sigma^2(x_i) < \infty$. Let $f(x|\boldsymbol{\beta}), \boldsymbol{\beta} \in B \subseteq \mathbb{R}^m$ be a parametric model. We now construct the semiparametric estimator of $f(x)$. First we obtain an appropriate estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ via a suitable method of estimation. Then $f(x)$ can be written as

$$f(x) = f(x|\hat{\boldsymbol{\beta}}) + f(x|\hat{\boldsymbol{\beta}})^\gamma r_\gamma(x, \hat{\boldsymbol{\beta}}), \tag{1}$$

where $r_\gamma(x, \boldsymbol{\beta}) = \{f(x) - f(x|\boldsymbol{\beta})\}/f(x|\boldsymbol{\beta})^\gamma$ for some $\gamma \in \{0, 1\}$. When $\gamma = 0$, this decomposition becomes $f(x) = f(x|\hat{\boldsymbol{\beta}}) + \{f(x) - f(x|\hat{\boldsymbol{\beta}})\}$ which is called additive correction. When $\gamma = 1$, on the other hand, we have multiplicative correction $f(x) = f(x|\hat{\boldsymbol{\beta}})\{f(x)/f(x|\hat{\boldsymbol{\beta}})\}$. By using the parameter γ, we can treat additive and multiplicative correction systematically (see, Fan et al. (2009)). In the second step, $r_\gamma(x, \hat{\boldsymbol{\beta}})$ is estimated by applying nonparametric technique to $\{(x_i, \{y_i - f(x_i|\hat{\boldsymbol{\beta}})\}/f(x_i|\hat{\boldsymbol{\beta}})^\gamma) : i = 1, \cdots, n\}$. The SPSE is obtained as

$$\hat{f}(x, \gamma) = f(x|\hat{\boldsymbol{\beta}}) + f(x|\hat{\boldsymbol{\beta}})^\gamma \hat{r}_\gamma(x, \hat{\boldsymbol{\beta}}), \tag{2}$$

where $\hat{r}_\gamma(x, \hat{\boldsymbol{\beta}})$ is a nonparametric estimator of $r_\gamma(x, \hat{\boldsymbol{\beta}})$.

We adopt the penalized spline to estimate $r_\gamma(x, \hat{\boldsymbol{\beta}})$. Let $\{B^{[p]}_{-p+1}(x), \cdots, B^{[p]}_{K_n}(x)\}$ be a marginal B-spline basis of degree p with equally spaced knots $\kappa_k = k/K_n (k = -p+1, \cdots, K_n+p)$. Then we consider the B-spline model

$$\sum_{k=-p+1}^{K_n} B^{[p]}_k(x) b_k$$

as an approximation to $r_\gamma(x, \hat{\boldsymbol{\beta}})$, where b_k's are unknown parameters. The definition and fundamental property of B-spline basis are detailed in de Boor (2001). Let $\boldsymbol{R}_\gamma$ be the n-vector with ith element $\{y_i - f(x_i|\hat{\boldsymbol{\beta}})\}/f(x_i|\hat{\boldsymbol{\beta}})^\gamma$ and let $Z = (B^{[p]}_{-p+j}(x_i))_{ij}$, $\boldsymbol{b} = (b_{-p+1} \ \cdots \ b_{Kn})'$. The penalized spline estimator $\hat{\boldsymbol{b}} = (\hat{b}_{-p+1} \ \cdots \ \hat{b}_{K_n})'$ of $\boldsymbol{b}$ is defined as the minimizer of

$$(\boldsymbol{R}_\gamma - Z\boldsymbol{b})'(\boldsymbol{R}_\gamma - Z\boldsymbol{b}) + \lambda_n \boldsymbol{b}' Q_m \boldsymbol{b},$$

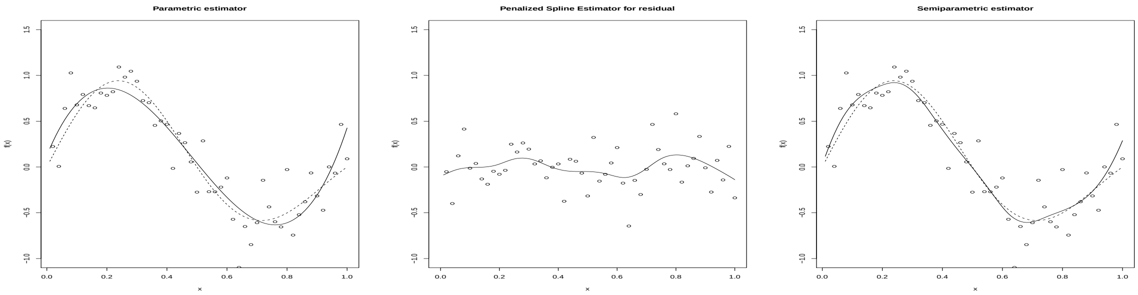

Figure 1: Plot of true $f(x)$, parametric estimator $f(x|\hat{\boldsymbol{\beta}})$, penalized spline estimator of $\hat{r}_0(x, \hat{\boldsymbol{\beta}})$ and SPSE $\hat{f}(x, 0)$ for one random sample.

where λ_n is the smoothing parameter and Q_m is mth difference matrix. The estimator of $r_\gamma(x, \hat{\boldsymbol{\beta}})$ is defined as

$$\hat{r}_\gamma(x, \hat{\boldsymbol{\beta}}) = \sum_{k=-p+1}^{K_n} B_k^{[p]}(x)\hat{b}_k = \boldsymbol{B}(x)'(Z'Z + \lambda_n Q_m)^{-1} Z' \boldsymbol{R}_\gamma, \tag{3}$$

where $\boldsymbol{B}(x) = (B_{-p+1}^{[p]}(x) \ \cdots \ B_{K_n}^{[p]}(x))'$.

In Figure 1, an example of SPSE is drawn. In the left panel, the true function $f(x) = \exp[-x^2]\sin(2\pi x)$ and the least square estimator $f(x|\hat{\boldsymbol{\beta}})$ of $f(x|\boldsymbol{\beta}) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ are drawn as dashed line and solid line, respectively. In the middle panel, the residual of $f(x|\hat{\boldsymbol{\beta}})$ and penalized spline estimator of $r_0(x, \hat{\boldsymbol{\beta}})$ are drawn. In the right panel, true function (dashed) and SPSE (solid) given as (2) are drawn. As the interpretation of $\hat{f}(x)$ of this example, the parametric part captures main shape of $f(x)$ and nonparametric part explains details which could not be captured by the $f(x|\hat{\boldsymbol{\beta}})$. Similarly, we can construct the SPSE with multiplicative correction.

3 Asymptotic Result

Asymptotics for NPSE has been developed by Claeskens et al. (2009). By using their results, we show the asymptotic bias, variance and asymptotic distribution of SPSE. We give some assumptions for asymptotics of SPSE.

Assumption

1. There exists $a > 0$ such that $a < f(x|\boldsymbol{\beta})$ for all $x \in [0, 1]$, $\boldsymbol{\beta} \in B$.
2. $\sup_{z\in[0,1]}\{q(z)\} < \infty$.
3. $|\partial f(x|\boldsymbol{\beta})/\partial\beta_i| < \infty$, for $x \in [0, 1]$, $\boldsymbol{\beta} \in B$, $i = 1, \cdots, m$.
4. $|\partial^2 f(x|\boldsymbol{\beta})/\partial\beta_i\partial\beta_j| < \infty$, for $x \in [0, 1]$, $\boldsymbol{\beta} \in B$, $i, j = 1, \cdots, m$.
5. $|d^i f(x)/dx^i| < \infty$, for $x \in [0, 1]$, $i = 1, \cdots, p + 1$.
6. $K_n = o(n^{1/2})$ and $\lambda_n = o(nK_n^{-1})$.

Define $G(q) = (g_{ij})_{ij}$, where

$$g_{ij} = \int_0^1 B_{-p+i}^{[p]}(u) B_{-p+j}^{[p]}(u) q(u) du$$

and $G(\sigma, \beta, \gamma, q) = (g_{\sigma,ij})_{ij}$, where

$$g_{\sigma,ij} = \int_0^1 B_{-p+i}^{[p]}(u) B_{-p+j}^{[p]}(u) \frac{\sigma^2(u) q(u)}{f(u|\boldsymbol{\beta})^{2\gamma}} du.$$

Let $\boldsymbol{b}^*(\boldsymbol{\beta}, \gamma)$ be a best L_∞ approximation to $(f(x) - f(x|\boldsymbol{\beta}))/f(x|\boldsymbol{\beta})^\gamma$. It means that $\boldsymbol{b}^*(\boldsymbol{\beta}, \gamma)$ satisfies

$$\sup_{x \in (0,1)} \left| \frac{f(x) - f(x|\boldsymbol{\beta})}{f(x|\boldsymbol{\beta})^\gamma} + b_{a1}(x|\boldsymbol{\beta}, \gamma) - \boldsymbol{B}(x)' \boldsymbol{b}^*(\boldsymbol{\beta}, \gamma) \right| = o(K_n^{-(p+1)}),$$

where

$$b_{a1}(x|\boldsymbol{\beta}, \gamma) = -\left(\frac{f(x) - f(x|\boldsymbol{\beta})}{f(x|\boldsymbol{\beta})^\gamma} \right)^{(p+1)} \frac{1}{K_n^{p+1}(p+1)!} \sum_{j=1}^{K_n} I(\kappa_{j-1} \le x < \kappa_j) B_{p+1}\left(\frac{x - \kappa_{j-1}}{K_n^{-1}} \right),$$

$I(a < x < b)$ is indicator function of the interval (a, b) and $B_p(x)$ is pth Bernoulli polynomial. For random variable U_n, $E[U_n|\boldsymbol{X}_n]$ and $V[U_n|\boldsymbol{X}_n]$ are the conditional expectation and variance of U_n given $(X_1, \cdots, X_n) = (x_1, \cdots, x_n)$.

Theorem 1. *Let $f \in C^{p+1}, f(\cdot|\boldsymbol{\beta}_0) \in C^{p+1}$. Then under the assumption, for a fixed $x \in (0,1)$,*

$$\begin{aligned} E[\hat{f}(x,\gamma)|\boldsymbol{X}_n] &= f(x) + b_a(x|\boldsymbol{\beta}_0, \gamma) + b_\lambda(x|\boldsymbol{\beta}_0, \gamma) + O_P(n^{-1}) + o_P(K_n^{-(p+1)}) + o_P(\lambda_n K_n n^{-1}), \\ V[\hat{f}(x,\gamma)|\boldsymbol{X}_n] &= \frac{f(x|\boldsymbol{\beta}_0)^{2\gamma}}{n} \boldsymbol{B}(x)' G(q)^{-1} G(\sigma, \beta_0, \gamma, q) G(q)^{-1} \boldsymbol{B}(x) + o_P(K_n n^{-1}), \end{aligned}$$

where

$$\begin{aligned} b_a(x|\boldsymbol{\beta}_0, \gamma) &= -\frac{f(x|\boldsymbol{\beta}_0) r_\gamma^{(p+1)}(x|\boldsymbol{\beta}_0)}{K_n^{p+1}(p+1)!} \sum_{j=1}^{K_n} I(\kappa_{j-1} \le x < \kappa_j) B_{p+1}\left(\frac{x - \kappa_{j-1}}{K_n^{-1}} \right), \\ b_\lambda(x|\boldsymbol{\beta}_0, \gamma) &= -\frac{\lambda_n}{n} f(x|\boldsymbol{\beta}_0)^\gamma \boldsymbol{B}(x)' G(q)^{-1} Q_m \boldsymbol{b}^*(\boldsymbol{\beta}_0, \gamma). \end{aligned}$$

Theorem 1 and Lyapunov's theorem yield the asymptotic distribution of SPSE.

Theorem 2. *Suppose that $E[|\varepsilon_i|^{2+\delta}|X_i = x_i] < C$ for some $\delta \ge 2$ and Assumption is satisfied. Then under $K_n = O(n^{1/(2p+1)})$ and $\lambda_n = O(n^{p/(2p+1)})$,*

$$\frac{\hat{f}(x,\gamma) - f(x) - b_a(x|\boldsymbol{\beta}_0, \gamma) - b_\lambda(x|\boldsymbol{\beta}_0, \gamma)}{\sqrt{V[\hat{f}(x,\gamma)|\boldsymbol{X}_n]}} \xrightarrow{D} N(0,1),$$

where $b_a(x|\boldsymbol{\beta}_0, \gamma)$ and $b_\lambda(x|\boldsymbol{\beta}_0, \gamma)$ are those given in Theorem 1.

Theorems 1 and 2 are obtained from one parametric model.

4 Parametric model selection

In this section, we describe how to choose one parametric model. If the true regression function satisfies $f \in \{f(\cdot|\boldsymbol{\beta})|\boldsymbol{\beta} \in B \subseteq \mathbb{R}^m\}$, the bias of SPSE is reduced. So we determine the initial parametric model from bias reduction point of view. More specifically, our purpose is to choose a parametric model so that the asymptotic bias of SPSE becomes smaller than that of NPSE:

$$|b_a(x|\boldsymbol{\beta}_0, \gamma)| < |b_a(x)| \quad \text{and} \quad |b_\lambda(x|\boldsymbol{\beta}_0, \gamma)| < |b_\lambda(x)|, \text{ for all } x \in (0,1), \tag{4}$$

where $b_a(x)$ and $b_\lambda(x)$ are the asymptotic bias of NPSE. If $f(x|\boldsymbol{\beta})$ is constant, $b_a(x|\boldsymbol{\beta}_0, \gamma)$ and $b_\lambda(x|\boldsymbol{\beta}_0, \gamma)$ are equivalent to $b_a(x)$ and $b_\lambda(x)$, respectively. When the same K_n and λ_n are used in both SPSE and NPSE, (4) can be rewritten as $L_a(x, \gamma) > 0$ and $L_\lambda(x, \gamma) > 0$ for all $x \in (0,1)$, where

$$\begin{aligned} L_a(x,\gamma) &= |f^{(p+1)}(x)| - \left| f(x|\boldsymbol{\beta}_0)^\gamma \left(\frac{f(x) - f(x|\boldsymbol{\beta}_0)}{f(x|\boldsymbol{\beta}_0)^\gamma} \right)^{(p+1)} \right|, \\ L_\lambda(x,\gamma) &= |\boldsymbol{B}(x)' G(q)^{-1} Q_m \boldsymbol{b}_f^*| - |f(x|\boldsymbol{\beta}_0)^\gamma \boldsymbol{B}(x)' G(q)^{-1} Q_m \boldsymbol{b}^*(\boldsymbol{\beta}_0, \gamma)| \end{aligned}$$

and $\boldsymbol{b}_f^*$ is a best L_∞ approximation to $f(x)$. As a pilot estimator of f and its $(p+1)$th derivative, we can use local polynomial estimator $\hat{f}$ with degree $p+2$. Then the estimator of $L_a(x, \gamma)$ and $L_\lambda(x, \gamma)$ can be obtained as

$$\hat{L}_a(x,\gamma) = |\hat{f}^{(p+1)}(x)| - \left| f(x|\hat{\boldsymbol{\beta}})^\gamma \left(\frac{\hat{f}(x) - f(x|\hat{\boldsymbol{\beta}})}{f(x|\hat{\boldsymbol{\beta}})^\gamma} \right)^{(p+1)} \right|$$

and by using empirical form,

$$\hat{L}_\lambda(x,\gamma) = |\boldsymbol{B}(x)'\Lambda^{-1} Q_m (Z'Z)^{-1} Z' \hat{\boldsymbol{f}}| - |f(x|\hat{\boldsymbol{\beta}})^\gamma \boldsymbol{B}(x)' \Lambda^{-1} Q_m (Z'Z)^{-1} Z' \hat{\boldsymbol{r}}_\gamma|,$$

where $\hat{\boldsymbol{f}} = (\hat{f}(x_1) \ \cdots \ \hat{f}(x_n))'$ and $\hat{\boldsymbol{r}}_\gamma$ is n vector with ith component $\{\hat{f}(x_i) - f(x_i|\hat{\boldsymbol{\beta}})\}/f(x_i|\hat{\boldsymbol{\beta}})^\gamma$. Here, we use the fact

$$\lambda_n f(x|\hat{\boldsymbol{\beta}})^\gamma \boldsymbol{B}(x)' \Lambda^{-1} Q_m (Z'Z)^{-1} Z' \hat{\boldsymbol{r}}_\gamma = b_\lambda(x|\boldsymbol{\beta}_0, \gamma) + o_P(\lambda_n K_n n^{-1}),$$

which is detailed in the proof of Theorem 2 (a) of Claeskens et al. (2009). We would decide one parametric model by relative evaluation. Let

$$C_{a\cap\lambda}(f(\cdot|\boldsymbol{\beta})) = \# \left\{ z_j \in (0,1) \middle| \hat{L}_a(z_j, \gamma) > 0, \hat{L}_\lambda(z_j, \gamma) > 0, j = 1, \cdots, J \right\},$$

for a given parametric model $f(\cdot|\boldsymbol{\beta})$ and some finite grid points $\{z_j\}_1^J$ on $(0,1)$. Here, for a set A, $\#A$ is the cardinality of A. Once we prepare a class of candidate parametric models $\{f_k = f_k(\cdot|\boldsymbol{\beta}_k); k = 1, \cdots, K\}$, then we choose a parametric model satisfying

$$f(x|\boldsymbol{\beta}) = \underset{f_k}{\operatorname{argmax}} \left\{ C_{a\cap\lambda}(f(\cdot|\boldsymbol{\beta}_k)) \right\}. \tag{5}$$

In summary, for each parametric model f_k, we calculate $\hat{L}_a$, $\hat{L}_\lambda$ and $C_{a\cap\lambda}(f(\cdot|\boldsymbol{\beta}_k))$. By using the parametric model which satisfies (5), we construct SPSE. If we can choose a good parametric model and a good $\hat{\boldsymbol{\beta}}$, the SPSE would have better behavior than NPSE.

5 Simulation

We will report the results of numerical study to confirming the behavior of SPSE compared to NPSE on finite sample. Accuracy of the model selection criteria discussed in Section 4 are also investigated, of which results will be addressed at the symposium.

References

[1] Claeskens,G., Krivobokova,T. and Opsomer,J.D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika.* **96**, 529-544.

[2] de Boor,C. (2001). *A Practical Guide to Splines.* Springer-Verlag.

[3] Fan,J., Wu,Y. and Feng,Y. (2009). Local quasi-likelihood with a parametric guide. *Ann. Statist.* **37** 4153-4183.

[4] Glad,I.K. (1998). Parametrically guided non-parametric regression. *Scand.J.Statist.* **25** 649-668.

[5] Hjort,N.L. and GLAD,I.K. (1995). Nonparametric density estimation with a parametric start. *Ann. Statist.* **23** 882-904.

[6] Martins-Filho,C., Mishra,S. and Ullah,A. (2008). A class of improved parametrically guided nonparametric regression estimators. *Econometric Rev.* **27** 542-573.

[7] Naito,K. (2002). Semiparametric regression with multiplicative adjustment. *Communications in Statistics, Theory and Methods* **31** 2289-2309.

[8] Naito,K. (2004). Semiparametric density estimation by local L_2-fitting. *Ann. Statist.* **32** 1162-1191.