应用概率统计 第二十三卷
第二期 2007 年 5 月

Chinese Journal of Applied Probability
and Statistics Vol. 23 No. 2 May 2007

A Note on the Optimal Smoothing in Partial Linear Models with Penalized Least Squares Estimator

XU WANGLI

(*Department of Statistics, Renmin University of China, Beijing, 100872*)

LI ZAIXING

(*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100080*)

Abstract

Partially linear models are assumed to be linearly related to one or more variable, but the relation to an additional variable or variables is not assumed to be easily parameterized. One primary approach to estimate the parameter and nonparametric part is the method of penalized least squares method, generalized cross-validation (GCV) approach is a popular method for selecting the smoothing parameters. However, the optimality of GCV in the partial linear model with penalized least squares has not been proved. In this article, we provide the support for using GCV through its optimality of the smoothing parameter. Simulation studies are employed to investigate the empirical performance of generalized cross-validation and that of cross-validation for comparison in the context.

Keywords: Generalized cross-validation, partially linear model, penalized least squares, smoothing parameter.

AMS Subject Classification: 62G05.

§ 1. Introduction

Inference in partially linear model has received considerable attention in the last decade, the specific model considered here is

$$y_i = g(t_i) + x_i^\tau \beta + \epsilon_i, \tag{1}$$

where, the regression function $g(t)$ is assumed to be a unknown smooth function; $\beta = (\beta_1, \beta_2, \cdots, \beta_p)^\tau$ is a $p \times 1$ vector of regression coefficient associated with covariates $x_i = (x_{i1}, x_{i2}, \cdots, x_{ip})^\tau$; ϵ_i are normal distribution with zero-mean and σ^2 variance and are independent of each other; furthermore, the conditional expectation of ϵ given (T, X) equals zero. This models, also called the semiparametric model, was proposed in Wahba (1984a) and Engle, Granger, Rice and Weiss (1986).

There have been several approaches to estimate β and g. One primary approach is the method of penalized least squares introduced by Engle, Granger, Rice and Weiss (1986) and Wahba (1984a, b) among others. Estimates are obtained by minimizing the quantity over β and g

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - g(t_i) - x_i^\tau \beta)^2 + \lambda J(g), \tag{2}$$

Received 2005. 4. 12. Revised 2005. 10. 18.

where the quadratic function $J(g)$ quantifies the roughness of g. If t is one dimensional in the bounded domain $D = [0\ 1]$, for example, we can take

$$J(g) = \int_0^1 \{g^{(m)}(t)\}^2 \mathrm{d}t,$$

which gives the cubic smoothing spline for $m = 2$. The smoothing parameter λ chosen by the statistician for a suitable fit controls the trade-off between the goodness-of-fit and the smoothness of g. Solving this minimization problem produces simultaneous estimates of β and g. Because equation (2) is the extension of the equation defining smoothing spline to the partial linear model, this estimator of g has been called a 'partial smoothing spline' by Wahba (1984b).

Penalized least squares are attractive for several reasons. The principle of adding a penalty term $J(g)$ to a sum of squares or more generally to a log-likelihood applies to a wide variety of linear and non-linear problems, see O'Sullivan (1986). There is also a Bayesian interpretation to the method as in, among others, Shiller (1984), Green, Jennison and Seheult (1985) and Eubank (1986). Most importantly, these researchers report that the method simply seems to work well.

From a practical point of view, however, we are more concerned with asymptotic properties when the smoothing parameter λ is chosen by some data-driven methods, such as the generalized cross-validation (GCV) proposed by Craven and Wahba (1979). To my knowledge, the optimality of GCV in the partial linear model with penalized least squares has not been proved. As a result, it is of interest to study the optimality of GCV methods since there is no reference for this.

The purpose of this article is the selection of smoothing parameters, i.e. λ, through GCV method, which is shown to yield optimal smoothing. The proofs for the optimality of GCV may be easy to extend to the proof of the optimality of some other data-driven method, for example, cross-validation (CV).

§2. Penalized Least Squares Methods

In this article, the model terms g and β in (1) will be estimated using the penalized least squares method through the minimization of (2). The purpose of this article is to select the smoothing parameters λ by standard generalized cross-validation method of Craven and Wahba (1979).

Consider the minimization of (2) for g in a q-dimensional space span $\{\xi_1, \cdots, \xi_q\}$. Functions in this space can be expressed as

$$g(t) = \sum_{j=1}^{q} c_j \xi_j(t) = \xi^T(t)c. \tag{3}$$

Plugging (3) into (2), one minimizes

$$(Y - Rc - X\beta)^T(Y - Rc - X\beta) + n\lambda c^T Q c \tag{4}$$

with respect to c and β, where R is a $n \times q$ with the (i,j)th entry $\xi_j(t_i)$, and Q is a $q \times q$ with the (j,k)th entry $J(\xi_i, \xi_k)$. Differentiating (4) with respect to c and β and setting the derivatives to 0, one has

$$\begin{pmatrix} R^\tau R + n\lambda Q & R^\tau X \\ X^\tau R & X^\tau X \end{pmatrix} \begin{pmatrix} c \\ \beta \end{pmatrix} = \begin{pmatrix} R^\tau Y \\ X^\tau Y \end{pmatrix}. \tag{5}$$

Assume that the matrix of $\begin{pmatrix} R^\tau R + n\lambda Q & R^\tau X \\ X^\tau R & X^\tau X \end{pmatrix}$ is invertible, the solution of (5) is then given by

$$\begin{pmatrix} \hat{c} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} R^\tau R + n\lambda Q & R^\tau X \\ X^\tau R & X^\tau X \end{pmatrix}^{-1} \begin{pmatrix} R^\tau \\ X^\tau \end{pmatrix} Y = A(\lambda) Y,$$

and simple algebra derives

$$A(\lambda) = \tilde{A}(\lambda) + (I - \tilde{A}(\lambda))X(X^\tau(I - \tilde{A}(\lambda))X)^{-1}X^\tau(I - \tilde{A}(\lambda)),$$
$$\hat{\beta} = (X^\tau(I - \tilde{A}(\lambda))X)^{-1}X^\tau(I - \tilde{A}(\lambda))Y, \qquad \hat{g} = \tilde{A}(\lambda)(y - X\hat{\beta}),$$

where $\tilde{A}(\lambda) = R(R^\tau R + n\lambda Q)^{-1}R^\tau$ is the smoothing matrix when the linear part $X\beta$ are absent.

With the standard formulation of penalized least squares regression, the minimization of (4) is in a so-called reproducing kernel Hilbert space, among others, see Gu (2002) in Chapter 2, $\mathcal{H} \subseteq \{g : J(g) < \infty\}$ in which $J(g)$ is a square semi norm, and the solution exists in the space $\mathcal{N}_J \oplus \text{span}\{R_J(t_i, \cdot), i = 1, 2, \cdots, n\}$, where $\mathcal{N}_J = \{g : J(g) = 0\}$ is the null space of $J(g)$ and $R_J(\cdot, \cdot)$ is the so-called reproducing kernel in $\mathcal{H} \ominus \mathcal{N}_j$. The expression of the solution is

$$g(t) = \sum_{i=1}^{m} d_i \phi_i(t) + \sum_{i=1}^{n} \tilde{c}_i R_J(t_i, t), \tag{6}$$

where $\{\phi_v\}_{v=1}^m$ is a basis of $\mathcal{N}_J$. It follows that $R = (S, \tilde{Q})$, where S is a $n \times m$ with the (i,j)th entry $\phi_j(t_i)$ and $\tilde{Q}$ is a $n \times n$ with the (i,j)th entry $R_J(t_i, t_j)$. From the property of reproducing kernel, we know that $J(R_J(t_i, \cdot), R_J(t_j, \cdot)) = R_J(t_i, t_j)$, so $Q = \text{diag}(0, \tilde{Q})$. See, e.g., Wahba (1990) and Gu (2002).

The formulation of (3) and (4) also covers general penalized regression spline so long as (5) is solvable. A sufficient condition for both R and X is to be of full column rank.

§ 3. Optimality of Generalized Cross-Validation

Generalized cross-validation offers a way to estimate appropriate values of parameters λ, it, to be minimized as a function of λ, is given by

$$G_n(\lambda) = \frac{n^{-1}Y^\tau(I - A)^2 Y}{(n^{-1}\text{tr}(I - A))^2}, \tag{7}$$

where the arguments λ are dropped from the notation of the smoothing matrix A. It will be shown in this section that the minimizers of $G_n(\lambda)$ yield optimal smoothing asymptotically in the

sense to be specified later. Numerical verification of the asymptotic analysis will be showed in this section. Generalized cross-validation was proposed by Craven and Wahba (1979) for independent data, with the asymptotic optimality established by Li (1986) in that setting, and by Härdle and Marron (1985) in the context of nonparametric regression with kernel estimator, see also Speckman (1985) and Härdle, Hall and Marron (1988) and the references therein.

Our aim in this paper is to select the optimal $\widehat{\lambda}_{\text{MASE}}$ which minimizes the following average mean squared error (AMSE)

$$L_n(\lambda) = \frac{1}{n}(\mathsf{E}(Y) - \widehat{Y})^\tau(\mathsf{E}(Y) - \widehat{Y}), \tag{8}$$

which is a natural loss. Simple algebra yields

$$L_n(\lambda) = \frac{1}{n}(g + X\beta)^\tau(I - A)^2(g + X\beta) - \frac{2}{n}(g + X\beta)^\tau(I - A)A\epsilon + \frac{1}{n}\epsilon^\tau A^2\epsilon.$$

Taking expectation with respect to ϵ, the risk is

$$R_n(\lambda) = \mathsf{E}(L_n(\lambda)) = \frac{1}{n}(g + X\beta)^\tau(I - A)^2(g + X\beta) + \frac{\sigma^2}{n}\text{tr}(A^2).$$

Let $\widehat{\lambda}_G$ and $\widehat{\lambda}_{\text{MASE}}$ be the minimizers of $G_n(\lambda)$ and $L_n(\lambda)$ respectively. The data-driven smoothing parameter $\widehat{\lambda}_G$ is called ***asymptotically optimal*** if

$$L_n(\widehat{\lambda}_G)/L_n(\widehat{\lambda}_{\text{MASE}}) \to 1 \text{ in probability.}$$

Define

$$U_n(\lambda) = \frac{1}{n}Y^\tau(I - A)^2Y + \frac{2}{n}\sigma^2\text{tr}(A).$$

It is easily derive that

$$U_n(\lambda) - L_n(\lambda) - \frac{1}{n}\epsilon^\tau\epsilon = \frac{2}{n}(g + X\beta)^\tau(I - A)\epsilon - \frac{2}{n}(\epsilon^\tau A\epsilon - \sigma^2\text{tr}A).$$

We shall establish the optimality under the following conditions.

C.1 As $n \to \infty$, $nR_n(\lambda) \to \infty$.

C.2 As $n \to \infty$, $(n^{-1}\text{tr}(A))^2/[n^{-1}\text{tr}(A^\tau A)] \to 0$.

In the absence of linear part, that is: model (1) change into a nonparametric model $y_i = g(t_i) + \epsilon_i$ $(1 \le i \le n)$, the risk is seen to be $n^{-1}g^\tau(I - \widetilde{A})^2g + \sigma^2n^{-1}\text{tr}(\widetilde{A}^2)$ in this nonparametric setting, and condition C.1 typically holds since $n^{-1}g^\tau(I - \widetilde{A})^2g = O(\lambda^s)$ for some $s \in [1, 2]$, and $\text{tr}(\widetilde{A}^2) = O(\lambda^{-1/r})$ as $\lambda \to 0$ and $n\lambda^{1/r} \to \infty$ for some $r > 1$ for univariate smoothing splines. See, e.g., Craven and Wahba (1979), Wahba (1985), Gu (2002) and Gu (2005).In the partially linear model, the following Lemma 2 will show the reasonableness of C.1. As to Condition C.2, in the absence of linear part of model (1), that is, $A = \widetilde{A}$, it generally holds since $\text{tr}\widetilde{A}(\lambda) = O(\lambda^{-1/r})$ as $\lambda \to 0$ and $n\lambda^{1/r} \to \infty$, of the same order as $\text{tr}\widetilde{A}^2(\lambda)$. See, e.g., Craven and Wahba (1979), Wahba (1985), Li (1986), Gu (2002) and Gu (2005). In the partially linear model, the following Lemma 3 illustrate the Condition C.2 clearly holds.

Lemma 1 As $n \to \infty$, $n\lambda^{1/r} \to \infty$ and $\lambda \to 0$,

$$\mathrm{tr}(A) = \mathrm{tr}(\widetilde{A})(1 + o_p(1)) \qquad \text{and} \qquad \mathrm{tr}(A^2) = \mathrm{tr}(\widetilde{A}^2)(1 + o_p(1)).$$

proof of Lemma 1 This proof is similar as the Lemma 2 of Speckman (1988). #

Lemma 2 As $n \to \infty$, $n\lambda^{1/r} \to \infty$ and $\lambda \to 0$, then $R_n(\lambda) = O(\lambda^s + n^{-1}\lambda^{-1/r})$.

proof of Lemma 2 Because

$$\begin{aligned} R_n(\lambda) &= \frac{1}{n}\mathsf{E}\|g + X\beta - \widehat{g} - X\widehat{\beta}\|^2 \\ &= \frac{1}{n}\mathsf{E}\|(g + X\beta) - \widetilde{A}(\lambda)(g + X\beta + \epsilon - X\widehat{\beta}) - X\widehat{\beta}\|^2 \\ &= \frac{1}{n}\|(I - \widetilde{A})(g - X\mathrm{bias}(\widehat{\beta}))\|^2 \ \text{(squared bias term)} \\ &\quad + \frac{\sigma^2}{n}\{\mathrm{tr}(\widetilde{A}^2) + 2X^\tau(I - \widetilde{A})\widetilde{A}(I - \widetilde{A})X(X^\tau(I - \widetilde{A})X)^{-1} \\ &\quad + X^\tau(I - \widetilde{A})^2 X\mathsf{Var}(\widehat{\beta})/\sigma^2\} \ \text{(variance term)}, \end{aligned}$$

we have that the two main terms in squared bias term are

$$\frac{1}{n}\|(I - \widetilde{A})g\|^2 = O(\lambda^s), \qquad \frac{1}{n}\|(I - \widetilde{A})X\|^2\mathrm{bias}^2(\widehat{\beta}) = O(\lambda^s),$$

and it can be shown that the variance term is dominated by $\sigma^2 n^{-1}\mathrm{tr}(\widetilde{A}^2) = O(n^{-1}\lambda^{-1/r})$. #

Lemma 3 As $n \to \infty$,

$$\frac{(n^{-1}\mathrm{tr}(A))^2}{n^{-1}\mathrm{tr}(A^2)} \to 0.$$

Proof of Lemma 3 Given Lemma 1 and $\mathrm{tr}\widetilde{A}(\lambda) = O(\lambda^{-1/r})$ as $\lambda \to 0$ and $n\lambda^{1/r} \to \infty$, of the same order as $\mathrm{tr}\widetilde{A}^2(\lambda)$, the lemma 3 follows. #

Theorem 1 Under the Conditions C.1, as $n \to \infty$, one has

$$U_n(\lambda) - L_n(\lambda) - \frac{1}{n}\epsilon^\tau\epsilon = o_p(L_n(\lambda)).$$

Proof of Theorem 1 For the proof of theorem 1, it suffices to show that

$$\begin{aligned} &n^{-1}(\epsilon^\tau A\epsilon - \sigma^2\mathrm{tr}A) = o_p(R_n(\lambda)), \\ &n^{-1}(g + X\beta)^\tau(I - A)\epsilon = o_p(R_n(\lambda)), \\ &U_n(\lambda) - R_n(\lambda) = o_p(R_n(\lambda)). \end{aligned}$$

Note that

$$\mathsf{Var}(n^{-1}\epsilon^\tau A\epsilon) = n^{-2}2\sigma^4\mathrm{tr}(A^2),$$

and from Lemma 1 and Condition C.1, it can easily derived

$$\mathsf{Var}(n^{-1}\epsilon^\tau A\epsilon) = n^{-1}O(R_n(\lambda)) = o(R_n(\lambda)^2).$$

From the following expression

$$\begin{aligned}\mathrm{Var}\,(n^{-1}(g+X\beta)^\tau(I-A)\epsilon) &= n^{-2}\sigma^2(g+X\beta)^\tau(I-A)^2(g+X\beta)\\ &= n^{-2}\sigma^2 g^\tau(I-A)^2 g,\end{aligned}$$

and that the term $(I-A)g$ is dominated by $(I-\widetilde{A})g$, it is easy to see that $\mathrm{Var}\,(n^{-1}(g+X\beta)^\tau(I-A)\epsilon) = o(R_n(\lambda)^2)$. Likewise

$$\mathrm{Var}\,(L_n(\lambda)) = o(R_n(\lambda)^2). \qquad \#$$

Theorem 2 Under the Conditions C.1 and C.2, as $n \to \infty$, one has

$$G_n(\lambda) - L_n(\lambda) - \frac{1}{n}\epsilon^\tau\epsilon = o_p(L_n(\lambda)).$$

Proof of theorem 2 Write $\mu = n^{-1}\mathrm{tr}A(\lambda)$ and $\widetilde{\sigma}^2 = n^{-1}\epsilon^\tau\epsilon$. Straightforward algebra yields

$$\begin{aligned}G_n(\lambda) - L_n(\lambda) - \frac{1}{n}\epsilon^\tau\epsilon &= \frac{1}{(1-\mu)^2}\{U_n(\lambda) - 2\sigma^2\mu - (L_n(\lambda)+\widetilde{\sigma}^2)(1-\mu)^2\}\\ &= \frac{U_n(\lambda) - L_n(\lambda) - \widetilde{\sigma}^2}{(1-\mu)^2} + \frac{(2-\mu)\mu L_n(\lambda)}{(1-\mu)^2} - \frac{\mu^2\widetilde{\sigma}^2}{(1-\mu)^2} + \frac{2\mu(\widetilde{\sigma}^2-\sigma^2)}{(1-\mu)^2}.\end{aligned}$$

Note that $n^{-1}\mathrm{tr}(A^2) < 1$, so Condition C.2 implies that $\mu \to 0$. The first term is $o_p(L_n(\lambda))$ by Theorem 1. The second term is $o_p(L_n(\lambda))$ since $\mu \to \infty$. By Condition C.2, $\mu^2 = o_p(L_n(\lambda))$, so the third term is $o_p(L_n(\lambda))$. Combining this with $\widetilde{\sigma}^2 - \sigma^2 = O_p(n^{-1/2}) = o_p(L_n^{1/2}(\lambda))$, one obtains $o_p(L_n(\lambda))$ for the fourth term. $\#$

Furthermore, assume that

C.3 For any zero-mean random variable series $\zeta_n(\lambda)$ satisfying $\zeta_n(\lambda) = o_p(R_n(\lambda))$ and $\mathrm{Var}\,(\zeta_n(\lambda)) = o^*(R_n^2(\lambda))$, we have $\zeta_n(\lambda) = o_p^*(R_n(\lambda))$, where $o_p^*(\cdot)$ indicates $o_p(\cdot)$ holds uniformly in a neighborhood Λ_n of λ_0 to which $\widehat{\lambda}_{\mathrm{MASE}}$ converges.

When Conditions C.1 and C.2 hold uniformly in Λ_n, together with condition C.3, the minimizer of $G_n(\lambda)$ and $U_n(\lambda)$ approximate the minimizer of $L_n(\lambda)$, see Gu (2002, § 3) and Gu and Ma (2005). In fact, in the case of optimal bandwidth selection for semiparametric model, Condition C.3 is reasonable, see Lemma 4.1 in Hong (1999). As a result, the selection of smoothing parameters through $G_n(\lambda)$, i.e., the GCV method is shown to yield optimally. From $U_n(\lambda)$, we can also select the optimal smoothing parameters if σ^2 is given. All together, we have,

$$L_n(\widehat{\lambda}_G)/L_n(\widehat{\lambda}_{\mathrm{MASE}}) \to 1 \text{ in probability.}$$

As a result, the selection of smoothing parameters through the generalize cross-validation method is shown to yield optimal smoothing.

Remark 1 Speckman (1988) and Denby (1986) have proposed an estimator of β and g, that is,

$$\widetilde{\beta} = (X^\tau(I-\widetilde{A}(\lambda))^2X)^{-1}X^\tau(I-\widetilde{A}(\lambda))^2Y, \qquad \widetilde{g} = \widetilde{A}(\lambda)(y - X\widetilde{\beta}),$$

which can have a faster convergence rate than $\widehat{\beta}$. The proofs of asymptotic optimal for λ based on generalized cross-validation is similar.

Remark 2 The cross-validation is another general procedure that can be applied to estimate smoothing parameters. To be specific, let Y^{-i} be the $n-1$ dimension vector with the ith observation, y_i, removed from the original response vector Y, and $\widehat{g}^{-i}$ and $\widehat{\beta}^{-i}$ be the estimators of g and β in (1) based on Y^{-i} observations respectively. A cross-validation estimate of λ is the minimizer of

$$C_n(\lambda) = n^{-1}\sum_{i=1}^{n}(y_i - \widehat{g}^{-i}(t_i) - x_i^T\widehat{\beta}^{-i})^2, \tag{9}$$

with respect to λ, which can be rewritten as

$$C_n(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \widehat{g}(t_i) - x_i^T\widehat{\beta}}{1-a_{ii}}\right)^2, \tag{10}$$

where a_{ii} are diagonal components of matrix A. In view of the asymptotic equivalence of these two smooth parameters selectors, one would expect their performances to be about the same, at least for large n. Indeed, it is shown in Härdle, Hall and Marron (1988) that both of these are asymptotically optimal for nonparametric curve estimators in the specific context of kernel regression estimation. In this paper, the results of smoothing parameters selection based on these two methods are shown in simulation part to demonstrate the performance of GCV and CV.

§4. Empirical Performance

In this section, we outline some Monte Carlo experiments designed to investigate the performance of the proposed generalized cross-validation and that of cross-validation for comparison in the context. In the simulation we conducted, the underlying model is

$$y = 2x_1 + 3x_2 + 4(t^2 - 1) + \epsilon,$$

where t is uniformly distributed on $[0,1]$, x_1, x_2 and ϵ are all normal distribution with zero mean and one variance. One hundred replicates of sample of size $n = 20$ and $n = 50$ are generated from the above model. Cubic smoothing splines, see Gu (2002, p33–p34) for selection R and Q, are calculated with λ minimizing $L_n(\lambda)$, $U_n(\lambda)$, $G_n(\lambda)$ and $C_n(\lambda)$. The average mean square error $L_n(\lambda)$ is calculated for all the one hundred replicates estimates, from which the optimal λ is located. For purpose of convenience, the minimizers of $L_n(\lambda)$, $U_n(\lambda)$, $G_n(\lambda)$ and $C_n(\lambda)$ are denoted by λ_m, λ_u, λ_g and λ_c respectively. The losses $L_n(\lambda_u)$, $L_n(\lambda_g)$ and $L_n(\lambda_c)$ are plotted against $L_n(\lambda_m)$, for all replicates in Figure 1, where a point on the line illustrates a perfect selection by the empirical method. Figure 1(a) and (c) picture $L_n(\lambda_m)$ versus $L_n(\lambda_u)$ with the sample size $n = 20$ and $n = 50$ respectively. $L_n(\lambda_m)$ against $L_n(\lambda_g)$ and against $L_n(\lambda_c)$ with sample size $n = 20$ are plotted in Figure 1(b), and with sample size $n = 50$ in Figure 1(d). From the Figures, generally speaking, all

of the methods seem to perform well most of time even when the sample size is 20, with occasional failures. Also, the empirical performance with larger sample size $n = 50$ is better than that with sample size $n = 20$. From Figure 1(b) and (d), it seems that CV method performs better than GCV method when sample size equal to 20, whereas, in the case of $n = 50$ the results based on GCV method is better.

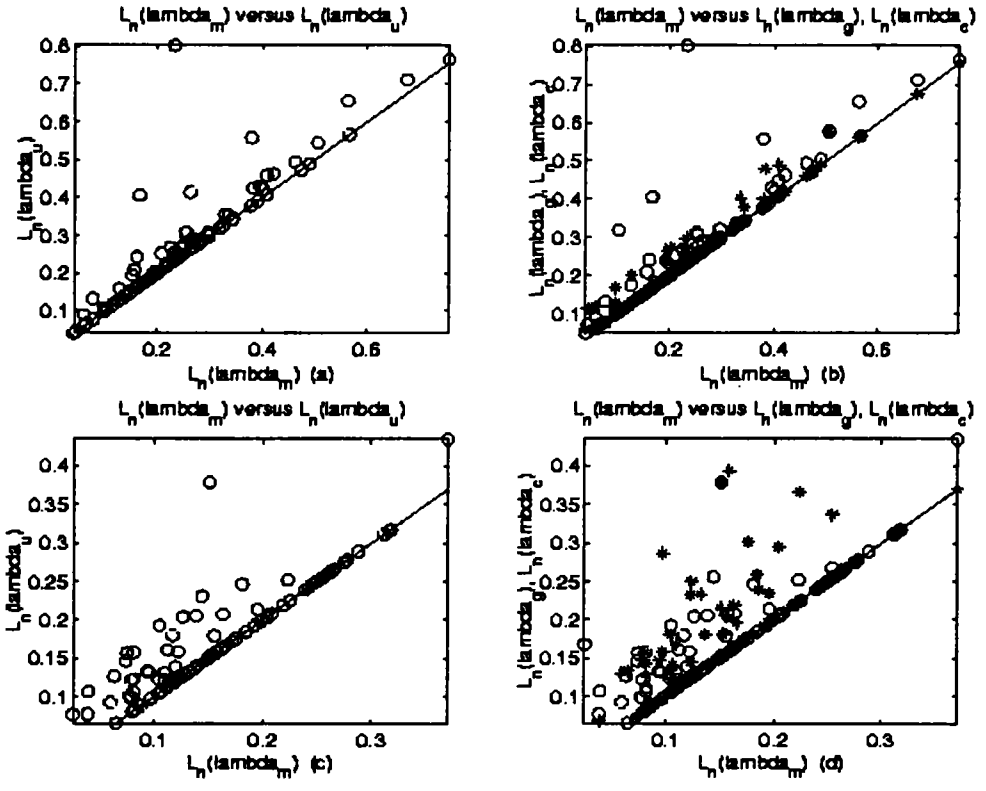

Figure 1 Simulation for partial linear models

(a): 'o'-$L_n(\lambda_m)$ versus $L_n(\lambda_u)$ with sample size $n = 20$;

(b): 'o'-$L_n(\lambda_m)$ versus $L_n(\lambda_g)$; '$*$'-$L_n(\lambda_m)$ versus $L_n(\lambda_c)$ with sample size $n = 20$;

(c): 'o'-$L_n(\lambda_m)$ versus $L_n(\lambda_u)$ with sample size $n = 50$;

(d): 'o'-$L_n(\lambda_m)$ versus $L_n(\lambda_g)$; '$*$'-$L_n(\lambda_m)$ versus $L_n(\lambda_c)$ with sample size $n = 50$.

References

[1] Craven, P., Wahba, G., Smoothing noisy data with spline functions, *Numer. Math.*, **31**(1979), 377–403.

[2] Densby, L., *Smooth Regression Function*, Tech. Report 26, AT&T Bell Laboratories, Murray Hill, NJ, 1986.

[3] Engle, R.F., Granger, C.W., Rice, J. and Weiss, A., Semiparametric estimate of the relation between weather and electricity sales, *J. Amer. Statist. Assoc.*, **81**(1986), 310–320.

[4] Eubank, R.L., A note on smoothness priors and nonlinear regression, *J. Amer. Statist. Assoc.*, **81**(1986), 514–517.

[5] Gu, C., *Smoothing Spline ANOVA Models*, New York: Springer-Verlag, 2002.

[6] Gu, C. and Ma, P., Optimal smoothing in nonparametric mixed-effect models, *Ann. Statist.*, **33**(2005), 1357–1379.

[7] Green, P., Jennison, C. and Seheult, A., Analysis of field experiments by least squares smoothing, *J. Roy. Statist. Soc. Ser. B*, **47**(1985), 299–315.

[8] Härdle, W., Hall, P., Marron, J.S., How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.*, **83**(1988), 86–101.

[9] Härdle, W., Marron, J.S., Optimal bandwidth selection in nonparametric regression, *Ann. Statist.*, **13**(1985), 1465–1481.

[10] Hong, S.Y., Automatic bandwidth choice in a semiparametric regression model, *Statist. Sinica*, **9**(1999), 775–794.

[11] Li, K.C., Asymptotic optimality of C_L and generalized cross-validation in the ridge regression with application to spline smoothing, *Ann. Statist.*, **14**(1986), 1101–1112.

[12] O'Sullivan, F., Ill-posed inverse problems, *Statist. Sci.*, **1**(1986), 502–527.

[13] Shiller, R.J., Smoothness priors and nonlinear regression, *J. Amer. Statist. Assoc.*, **79**(1984), 609–615.

[14] Speckman, P., Kernel smoothing in partial linear models, *J. Roy. Statist. Soc. Ser. B*, **50**(1988), 413–436.

[15] Wahba, G., Cross validated spline methods for the estimation of multivariate functions from data on functionals, *In Statistics: An Appraisal, Proc. 50th Anniversary Conf. Iowa State Statistical Laboratory* (Edited by David, H.A. and David, H.T.), 205–235, Iowa State University Press, Ames, 1984a.

[16] Wahba, G., Partial spline models for the semiparametric estimation of functions of several variables, *In Analysis for Time Series, Japan-US Joint Sem.*, 319–329, Tokyo: Institute of Statistical Mathematics, 1984b.

[17] Wahba, G., A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, *Ann. Statist.*, **13**(1985), 1378–1402.

[18] Wahba, G., *Spline Models for Observational Data*, Volume 59 of CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia: SIAM, 1990.

对部分线性模型用惩罚最小二乘估计时最优光滑化的注记

许王莉　　李再兴

(中国人民大学统计学院，北京，100872)　(中国科学院数学与系统科学研究院，北京，100080)

部分线性模型也就是响应变量关于一个或者多个协变量是线性的，但对于其他的协变量是非线性的关系．对于部分线性模型中的参数和非参数部分的估计方法，惩罚最小二乘估计是重要的估计方法之一．对于这种估计方法，广义交叉验证法提供了一种确定光滑参数的方法．但是，在部分线性模型中，用广义交叉验证法确定光滑参数的最优性还没有被证明．本文证明了利用惩罚最小二乘估计对于部分线性模型估计时，用广义交叉验证法选择光滑参数的最优性．通过模拟验证了本文中所提出的用广义交叉验证法选择光滑参数具有很好的效果，同时，本文在模拟部分比较了广义交叉验证和最小二乘交叉验证的优劣．

关 键 词：广义交叉验证，部分线性模型，惩罚最小二乘，光滑参数．

学 科 分 类 号：O212.