한국통계학회논문집 제2권 2호
1995년 12월 pp. 266-276

Diagnostic for Smoothing Parameter Estimate in Nonparametric Regression Model

In-Suk Lee and Won-Tae Jung[1)]

Abstract

We have considered the study of local influence for smoothing parameter estimates in nonparametric regression model. Practically, generalized cross validation(GCV) does not work well in the presence of data perturbation. Thus we have proposed local influence measures for GCV estimates and examined effects of diagnostic by above measures.

1. Introduction

Smoothing splines are a type of nonparametric regression for estimators which the diagnostic problem has received a good bit of attention. Some early comments concerning diagnostic methods for these estimators can be found in Wold(1974). Diagnostic methods for smoothing splines have received increasing attention in recent years.

Current diagnostic methods for smoothing spline are mostly of the case-deletion variety, including parallels of residuals and Cook's distance measure. More detailed treatments of diagnostic methods for smoothing splines with examples can be found in Eubank(1984, 1985, 1988), Silverman(1985), and Eubank and Thomas(1993).

By the way, the decision of the measure of smoothness for diagnostic and estimation of regression function by smoothing splines is required. The measure is called the smoothing parameter. The smoothing parameter acts as a tuning constant to balance the competing aims of fidelity to the data and smoothness. Small values of smoothing parameter produce wiggly estimates and, in the extreme case smoothing parameter is zero, a spline which interpolates the data. Large values yield smoother estimates, with smoothing parameter corresponding to polynomial regression.

Selecting a value for the smoothing parameter is a crucial part of the fitting process, and automatic procedures for selecting such tuning constants based on the data are often preferred. Li(1985), Hall and Titterington(1987), and Hutchinson(1989) have employed generalized cross-validation(GCV) to select smoothing parameter in problems of estimation.

1 Dept. of Statistics, Kyungpook National Univ. in Taegu, KOREA.

The GCV estimate works well in the uniform design knots. However, Hall and Titterington showed that GCV estimators of smoothing parameter can be slightly underestimated in both linear ridge-regression and nonparametric regression. And results in Andrews(1989) and Eubank and Thomas(1993) suggest caution in the routine use of GCV for smoothing parameter selection in the presence of outliers and influential observations.

Thus, we need to use the proper smoothing parameter to produce satisfactory results. That is, diagnostics for estimates of smoothing parameter are needed. But, no diagnostics for smoothing parameter have been studied.

Hence, the objective of this paper is to present diagnostics for influence on an important aspect of a fitted smoothing parameter by GCV under some perturbation schemes.

In Section 2, we introduce the spline model and derive the local influence method for smoothing parameter estimates. In Section 3, as application for modifying influential observations by diagnostic measures, we practically detect influential observations through a example.

2. Model and Local Influence Measures

Consider the situation where responses $y_1, \cdots, y_n$ are observed corresponding to values $t_1, \cdots, t_n$ of an independent variable which, for convenience, are assumed to satisfy $a \le t_1 < \cdots < t_n \le b$.

The y_j and t_j are related by the model,

$$y_j = \mu(t_j) + \varepsilon_j ,$$

where μ is a some smooth regression function, the errors ε_j are uncorrelated with mean zero and constant variances σ^2. By smooth, we mean that μ belongs to the set $W_2^m[a,b]$ of functions g that, for some fixed m, have m-1 absolutely continuous derivatives and square-integrable mth derivative $g^{(m)} \in [a,b]$.

And then, a popular estimator $\hat{\lambda}$ based on the assumptions of above is the minimizer over $g \in W_2^m$ of

$$\frac{1}{n}\sum_{1}^{n}\{y_j - g(t_j)\}^2 + \lambda \int_a^b \{g^{(m)}(t)\}^2 dt , \quad \lambda > 0. \tag{1}$$

The GCV choice $\hat{\lambda}$ which minimizes

$$GCV(\lambda) = \frac{\underline{e}_\lambda{}^T \underline{e}_\lambda}{\{tr(I - H(\lambda))\}^2}$$

where $H(\lambda)$ is the hat matrix that transforms the data vector $\underline{y}$ into the vector of smoothing spline fitted values, and $\underline{e}_\lambda = (I-H(\lambda))\underline{y}$ is the vector of residuals.

2.1 Local Influence Measures

Now, by the local-influence method of Cook(1986) and Lawrance(1991), we drive diagnostics which is identify observation that have a disproportionately large impact on the determination of the GCV estimator $\hat{\lambda}$. Let Ω be some open set of allowable perturbations. Suppose that the perturbation is $\underline{w}$ and the null perturbation is $\underline{w}_0$.

In order to find direction of large local change, our first step is to approximate the actual surface with its tangent plane at $\hat{\lambda}(\underline{w}_0)$ and find the direction of maximum slope d_{max} on this tangent plane.

It is easy to show that the direction of maximum slope is

$$d_{max} \approx \frac{\partial \hat{\lambda}(\underline{\omega})}{\partial \underline{\omega}^T}$$

evaluated at $\underline{\omega}_0$.

The direction vector d_{max} tells us how to perturb the data and the model to produce the greatest local change. Thus itself is the influence diagnostic measure, and the largest absolute components of d_{max} identify locally influential cases.

Hence we have the following theorem.

Theorem 1. Let $GCV(\lambda, \underline{w})$ is the perturbed generalized cross valida- tion function by $\underline{w}$. Then the direction of maximum slope is given by

$$d_{max} \approx -\frac{\partial^2 GCV(\lambda, \underline{w})}{\partial \underline{w}^T \partial \lambda}$$

evaluated at $\hat{\lambda}$ and $\underline{w}_0$.

2.1.1 Additive Perturbation Case

The perturbation scheme consists of adding small perturbations to the responses, so that the vector of modified responses is

$$\underline{y}_w = \underline{y} + \underline{w},$$

here, $\underline{w}_0 = (0,0,\cdots,0)$ represents no modification of the data.

Additive perturbations of the responses have been used by Emerson, Hoaglin and Kempthorne(1984), Thomas and Cook(1989), and Lawrance(1991). Under this perturbation scheme, the penalized least squares criterion (1) becomes

$$\min_{g\in W_2^m}\left[\frac{1}{n}\sum_1^n\{y_j+w_j-g(t_j)\}^2+\lambda\int_a^b\{g^{(m)}(t)\}^2dt\right],\quad \lambda>0,$$

and GCV choose $\hat{\lambda}$ to minimize

$$GCV(\lambda,\underline{w})=\frac{(\underline{y}+\underline{w})^T(I-H(\lambda))^T(I-H(\lambda))(\underline{y}+\underline{w})}{\{tr[I-H(\lambda)]\}^2}.$$

Thus,

$$d_{\max}(\underline{y}+\underline{w})\approx-\frac{\partial^2 GCV(\lambda,\underline{w})}{\partial\,\underline{w}^T\partial\lambda},\text{at }\hat{\lambda},\quad \underline{w}_0$$

$$\approx-\left[\frac{tr[H(\hat{\lambda})(I-H(\hat{\lambda}))}{tr(I-H(\hat{\lambda}))}\times I-H(\hat{\lambda})\right](I-H(\hat{\lambda}))^2\underline{y}.$$

It is the local influence measure under additive perturbation scheme.

2.1.2 Multiplicative Perturbation Case

The vector of modified responses is $\underline{y}_w=\underline{w}\otimes\underline{y}$, as multiplicated data form, where $\otimes$ is Hadamard product. Here, $\underline{w}_0=(1,1,\cdots,1)$ repres- ents no modification of the data.
Multiplicative perturbation of the responses is similar to the case-weight perturbations form, and have been used by Pregibon(1981), Cook(1986), and Lawrance(1991). Under this perturbation scheme, the penalized least squares criterion (1) becomes

$$\min_{g\in W_2^m}\left[\frac{1}{n}\sum_1^n\{w_j\cdot y_j-g(t_j)\}^2+\lambda\int_a^b\{g^{(m)}(t)\}^2dt\right],\quad \lambda>0,$$

and then GCV choose $\hat{\lambda}$ to minimize

$$GCV(\lambda,\underline{w})=\frac{(\underline{w}\otimes\underline{y})^T(I-H(\lambda))^T(I-H(\lambda))(\underline{w}\otimes\underline{y})}{\{tr[I-H(\lambda)]\}^2}.$$

Thus,

$$d_{\max}(\underline{y}\otimes\underline{w})\approx-\frac{\partial^2 GCV(\lambda,\underline{w})}{\partial\,\underline{w}^T\partial\lambda},\text{at }\hat{\lambda},\quad \underline{w}_0$$

$$\approx-D(y)\left[\frac{tr[H(\hat{\lambda})(I-H(\hat{\lambda}))}{tr(I-H(\hat{\lambda}))}\times I-H(\hat{\lambda})\right](I-H(\hat{\lambda}))^2\underline{y}$$

$$\approx \underline{y} \otimes d_{\max}(\underline{y}+\underline{w}),$$

where $D(y)=diag(y_1,\cdots,y_n)$.

It is the local influence measure under multiplicative perturbation scheme.

3. Application for Modifying Influential Observations

We take the true regression function by

$$\mu(t)=2\sum_{j=1}^{4}\{a_j\cos(2\pi jt)+b_j\sin(2\pi jt)\}, \quad 0\le t\le 1,$$

with $\underline{a}^T=(-0.5,0.5,2.5,1.0)$ and $\underline{b}^T=(2.5,1.0,0.5,0.5)$.

This regression function is discussed in Hall and Titterington(1987). In our numerical work, this minimization was achieved using a Golden Section search.

3.1 Additive Perturbation Case

Suppose data structure is given by

$$y_i^*=y_i+w_i,$$

where y_i^* is observed value but contaminated by amount w_i, y_i is the correct value and

$$w_i=\begin{cases} w_i & \text{for contaminted } y_i^* \\ 0 & \text{for correct } y_i^*. \end{cases}$$

Box-plot and Stem-and-leaf plot offer a rough bound for largest components of $d_{\max}(\underline{y}+\underline{w})$.

For example, suppose that ith and jth components of $d_{\max}(\underline{y}+\underline{w})$ have negative and positive large values, respectively. Then we may suspect amounts w_i, w_j and correct values are given by,

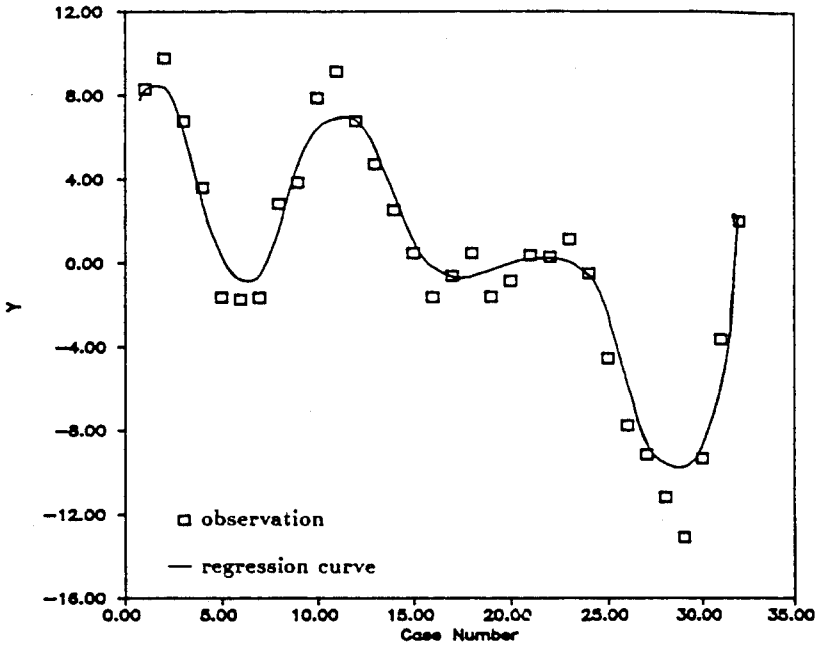
$$y_i^*=y_i+w_i,$$

$$y_j^*=y_j-w_j,$$

and

$$y_k=y_k^* \quad \text{for } k\neq i,j.$$

Figure 1. An Index Plot of Original Data



In Figure 1, the original data are plotted against case number rather than t to facilitate comparison with the diagnostics. Using $d_{max}(\underline{y}+\underline{w})$, we can take the cases (6,7) having the largest values, in Figure 2 and 3.

And we can find the case 1 with largest Cook's distance is not the same as the groups highlighted by the $d_{max}(\underline{y}+\underline{w})$ diagnostic, in Figure 4 and 5.

In this case, we can obtain smoothing parameter estimate, $\hat{\lambda}_a = 1.22 \times 10^{-7}$ from modified under the additive scheme and $\hat{\lambda}_0 = 7.6 \times 10^{-8}$ from original data by GCV.

Figure 2. An Index Plot of $d_{max}(\underline{y}+\underline{w})$

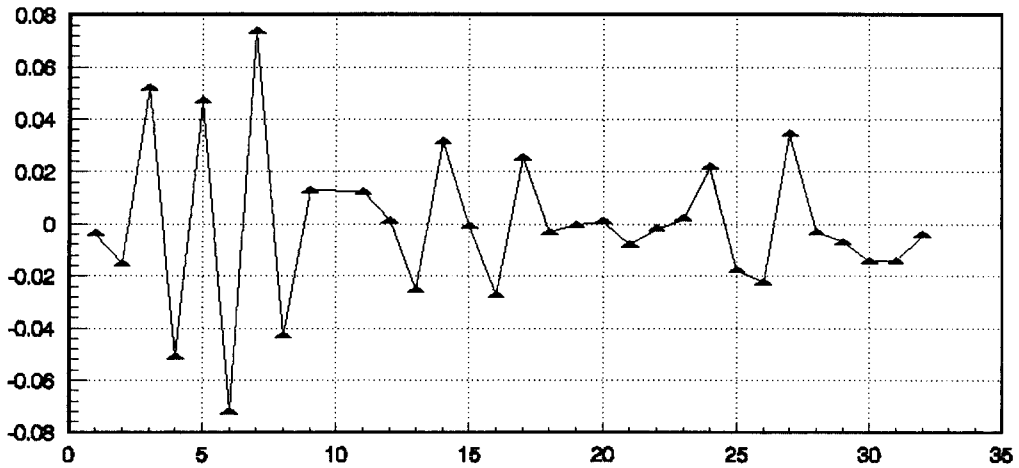


Figure 3. Box-Plot for $d_{max}(\underline{y}+\underline{w})$

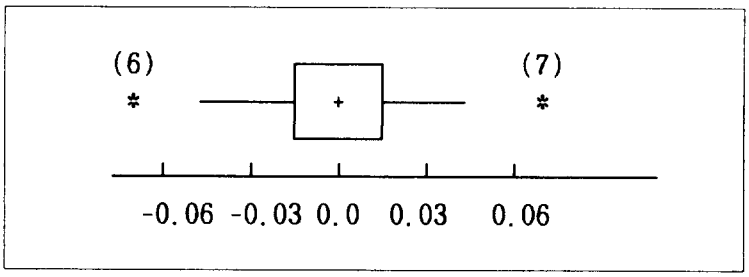


Figure 4. An Index Plot of Cook's Distance

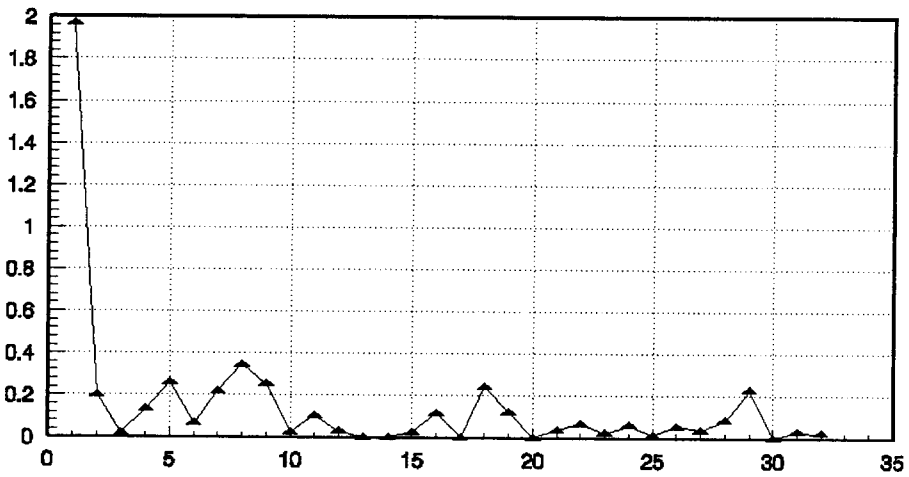
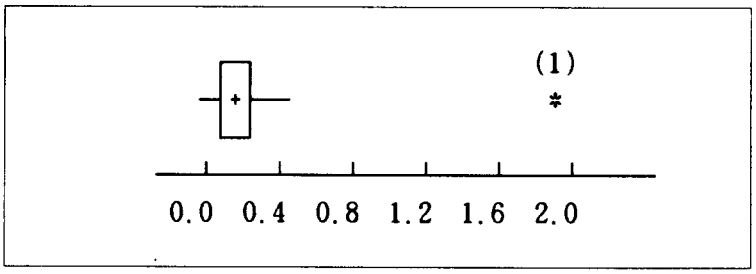


Figure 5. Box-Plot for Cook's Distance



3.2 Multiplicative Perturbation case

Suppose data structure is given by

$$y_i^* = y_i \cdot w_i,$$

where

$$w_i = \begin{cases} w_i & \text{for contaminted } y_i^* \\ 1 & \text{for correct } y_i^*. \end{cases}$$

For example, suppose that ith and jth components of $|d_{max}(\underline{w} \otimes \underline{y})|$ have largest values. Then we also suspect amounts w_i, w_j and correct values are given by

$$y_i^* = y_i / w_i,$$

$$y_j^* = y_j / w_j,$$

and $\quad y_k = y_k^* \quad$ for $k \neq i, j$.

In this scheme, we can get the cases (3,27) having the largest absolute values. In this case, we can obtain the smoothing parameter estimate, $\hat{\lambda}_m = 1.2 \times 10^{-7}$ from modified data under multiplicative scheme by using GCV, in Figure 6 and 7.

Figure 6. An Index Plot of $d_{max}(\underline{w} \otimes \underline{y})$

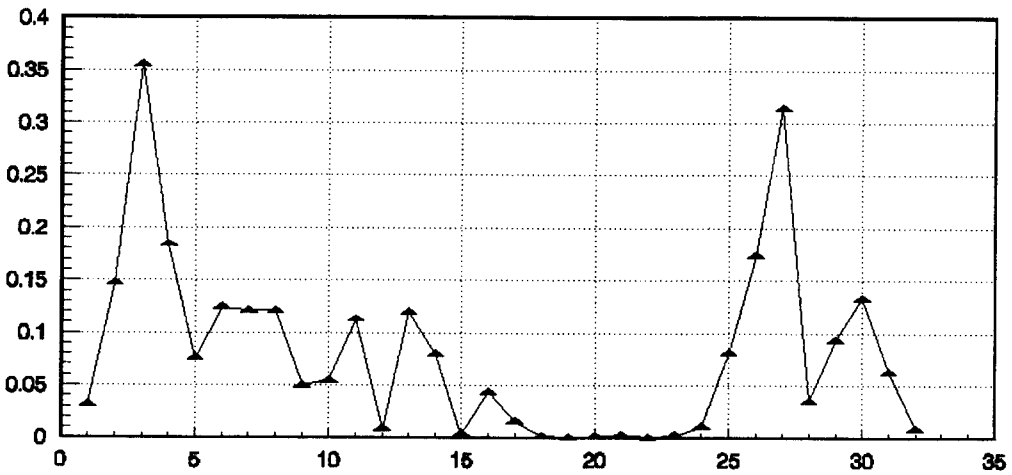
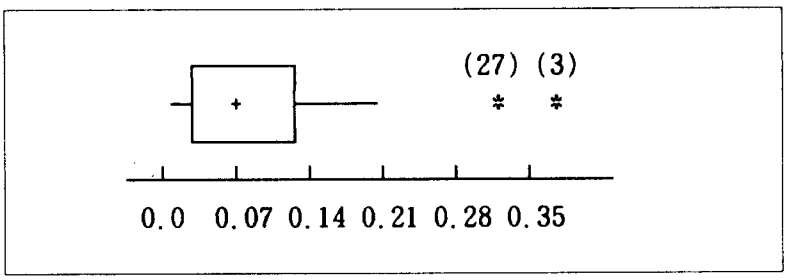


Figure 7: Box-Plot for $d_{max}(\underline{w} \otimes \underline{y})$



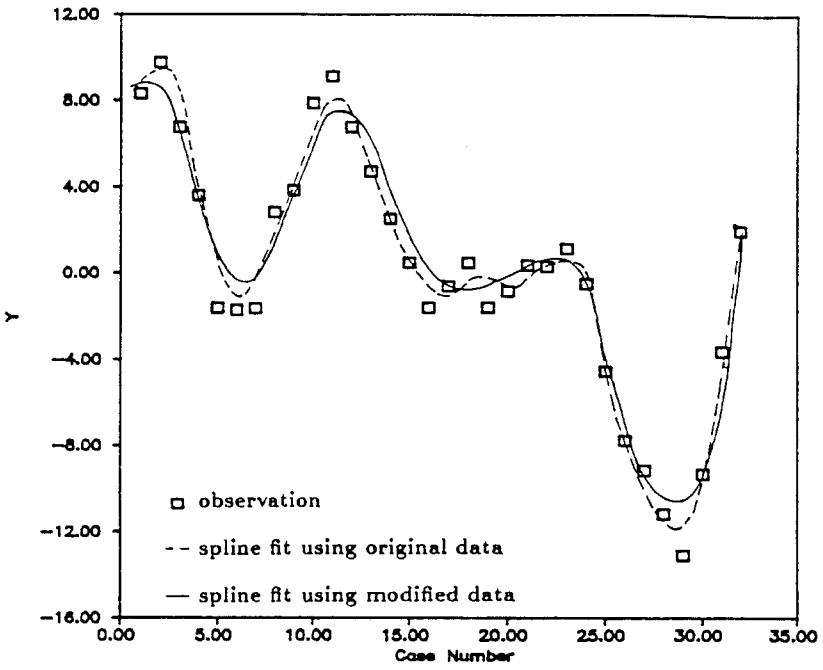
3.3 Summary

Table 1 show the summary for each diagnostic measure. In this example, we used the error's variance with $\sigma^2 = 0.5$. As we can see, the estimates of variance behind diagnostic procedures are closer to the true variance than the estimate of variance using original data. Figure 8 represents spline fits using the original($\hat{\lambda}_0 = 7.6 \times 10^{-8}$) and modified data ($\hat{\lambda}_m = 1.2 \times 10^{-7}$), respectively. And we can observe that smoothing spline fits under the diagnostic procedures are less wiggle.

Table 1. Summary for each Diagnostic Measure with $\sigma^2 = 0.5$

	Measure	$\hat{\lambda}$	$\hat{\sigma}^2$
Original		7.3×10^{-8}	0.412
Modified Data	Cook's Distance	7.10×10^{-8}	0.410
	Additive Scheme	1.22×10^{-7}	0.513
	Multipl. Scheme	1.20×10^{-7}	0.510

Figure 8. Spline Fits using the Original and Modified data, respectively



4. Remark and Discussion

In this paper, we can observe the following facts from results.

1) $d_{max}(\underline{w} \otimes \underline{y})$ is proportional to $d_{max}(\underline{w} + \underline{y})$ through $\underline{y}$.

2) Adjustment of an amount w_i in multiplicative perturbation case is easier to handle than that of additive case.

3) Diagnostic measure, Cook's distance in a single case deleted, can not find the influential observations in GCV estimate.

Reference

[1] Cook, R.D. (1986). Assessment of Local Influence (with Disscusion), *Journal of the Royal Statistical Society*, B, Vol. 48, 133-169.

[2] Craven, P. and Wahba, G. (1979). Smoothing Noisy Data with Spline Functions; Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation, *Numerical Mathematics*, Vol. 31, 377-403.

[3] Emersion, J.D., Hoaglin, D.C., and Kempthorne, P.J. (1984). Leverage in Least Squares Additive-Plus-Multiplicative Fits for Two-Way Tables, *Journal of the American Statistical Association*, Vol. 79, 329-335.

[4] Eubank, R.L. (1984). The Hat Matrix for Smoothing Splines, *Statistics and Probability Letters*, Vol. 2, 9-14.

[5] Eubank, R.L. (1985). Diagnostics for Smoothing Splines, *Journal of the Royal Statistical Society*, B, Vol. 47, 332-341.

[6] Eubank, R.L. and Gunst, R.F. (1986). Diagnostics for Penalized Least Squares Estimators, *Statistics and Probability Letters,* Vol. 4, 265-272.

[7] Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression* Marcel Decker, New York.

[8] Eubank, R.L. and Thomas, W. (1993). Dectecting Heteroscedasticity in Nonparametric Regression, *Journal of the Royal Statistical Society*, Vol. 55, 145-155.

[9] Hall, P. and Tittrington, D.M. (1987). Common Structure of Techniques for Automatic Smoothing Parameters in Regression Problems, *Journal of the Royal Statistical Society*, B, Vol. 49, 184-198.

[10] Hutchinson M.F.(1990). A Stocastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines,*Communicatoins in Statistics - Simulation*, Vol. 19, 433-450.

[11] Lawrance, A.J. (1991). *Directions in Robust Statistics and Diagnostics,* Springer-Verlag, New York.

[12] Li, K.C. (1985). From Stein's Unbiased Risk Estimates to the Method of Generalized Cross-Validation, *Annals of Statististics,* Vol. 13, 1352-1377.

[13] Pregibon, D. (1981). Logistic Regression Diagnostics, *Annals of Statistics*, Vol. 9, 705-724.

[14] Silverman, B.W. (1985). Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting, *Journal of the Royal Statistical Society*, B, Vol. 47, 1-52.

[15] Thomas, W. and Cook, R.D. (1989). Assessing Influence on regression Coefficients in Generalized Linear Models, *Biometrika*, Vol. 76, 741-749.