The Korean Communications in Statistics
Vol. 13 No. 2, 2006, pp. 379-387

On Convex Combination of Local Constant Regression[1)]

Jungwon Mun[2)] and Choongrak Kim[3)]

Abstract

Local polynomial regression is widely used because of good properties such as such as the adaptation to various types of designs, the absence of boundary effects and minimax efficiency. Choi and Hall (1998) proposed an estimator of regression function using a convex combination idea. They showed that a convex combination of three local linear estimators produces an estimator which has the same order of bias as a local cubic smoother. In this paper we suggest another estimator of regression function based on a convex combination of five local constant estimates. It turned out that this estimator has the same order of bias as a local cubic smoother.

Keywords : Bandwidth; Bias; Kernel; Local polynomial regression.

1. Introduction

Local polynomial fitting has good properties in both theoretical and practical sense. This method has lots of good properties, such as the adaptation to various types of designs, the absence of boundary effects and minimax efficiency. Local constant smoothers (e.g. Nadaraya (1964) and Gasser-Muller (1984)) and the local linear smoothers have their conditional bias of size h^2 where h is a bandwidth to be estimated. Gasser-Muller estimator has the same conditional bias as local linear smoother, but it has larger variance (the same variance in fixed design). There are many papers which show the advantages of local linear smoothing, e.g. Fan (1993), Hastie and Loader (1993), Cleveland and Loader (1996), Fan and Gijbels (1996).

In local polynomial fitting, we put larger weight in a neighborhood of x with kernel function K, to estimate the regression function $m(x)$. Provided the $(p+1)^{th}$ derivative of $m(\cdot)$ at the point x exists, we can approximate the unknown regression function $m(\cdot)$ locally by a polynomial of order p. To reduce the bias

1) This work was supported by Pusan National University Research Grant.
2) Department of Statistics, University of Wisconsin, Madison, WI. 53705, USA.
3) Professor, Department of Statistics, Pusan National University, Pusan 609-735, Korea. Correspondence : crkim@pusan.ac.kr

of local polynomial estimator of $m(x)$, we can take p as large. For large p, however, we encounter with two serious problems. First, the variance increases as p increases. Of course the asymptotic order of the leading term is independent of p, but the constant of the leading term is an increasing function of p. Secondly, when the data are sparse, calculation of a local polynomial smoother involves inversion of $(p+1)\times(p+1)$ matrix, with potential numerical problems resulting from near-singularity. So we usually take $p=0$ or 1, in other words, local constant and local linear approaches are preferred to local quadratic or local cubic to avoid potential numerical problems for the sparse design. For the sparse design, we might encounter with numerical problems when p is large. For example, the determinant arising from the estimation of regression function is close to zero as argued by Choi and Hall (1998).

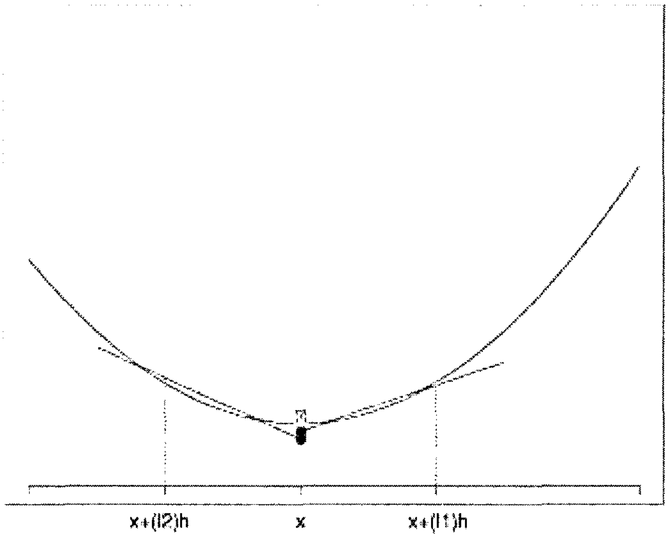
Choi and Hall (1998) proposed an estimator of $m(x)$ using a convex combination idea. To be more specific, they showed that a convex combination of three local linear estimators produces an estimator which has the same order of bias as a local cubic smoother, .i.e. it can reduce the order of bias up to h^4.

In this paper we suggest another estimator of $m(x)$ based on a convex combination of five local constant estimates. It turned out that this estimator has the same order of bias as a local cubic smoother. Section 2 gives basic concepts and approaches of convex combinations. A suggested estimator and its asymptotic bias are derived in Section 3. Finally concluding remarks and further research area are given in Section 4.

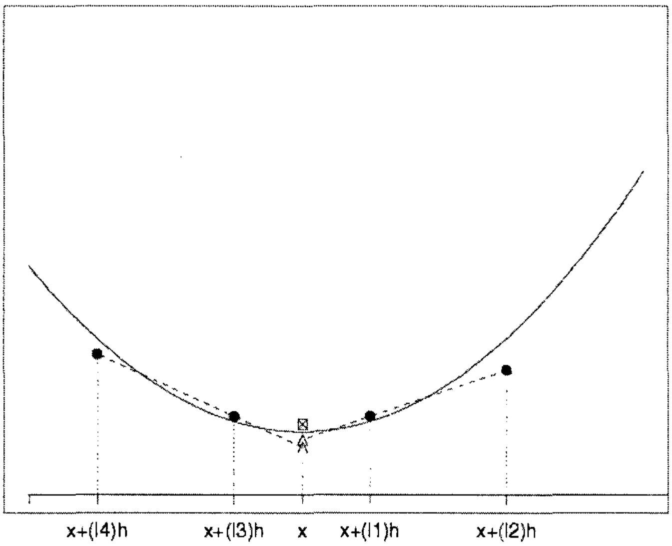
2. Basic Concepts and Approaches

Choi and Hall (1998) suggested an estimator, $\tilde{m}_L$ that is a convex combination of three local linear smoothers. i.e. $\hat{m}_L(x|x)$, $\hat{m}_L(x|x+lh)$ and $\hat{m}_L(x|x-lh)$. See <Figure 1>. With proper weight λ and l, their estimator reduces the bias by two orders of magnitude. We will discuss this estimator in detail in Section 3.

If the data points near the target point x are very sparse then even the local linear estimator can also be unstable. So, it can be more stable if we deal with the local constant smoother rather than the local linear smoother. First, we start from five local constant estimators. As shown in <Figure 2>, we get two estimators using the line passing through a pair of local constant smoothers at $x+l_ih$ and $x+2l_ih(i=1,2)$. Now, we call the line determined by two local constant smoothers "Local Constant Line".



<Figure 1> A curve indicates true regression function and solid lines indicate local linear line at target points $x+l_1h$ and $x+l_2h$. Two circular points denote skewed estimators and rectangular point denotes local linear estimator evaluated at x. Convex combination is done for three points.



<Figure 2> A curve indicates true regression function and dashed lines indicate local constant line which passes through two local constant smoothers at target points $x+l_ih$ and $x+2l_ih(i=1,2)$. Two triangular points denote skewed estimators and rectangular point denotes local linear estimator evaluated at x. Convex combination is done for three points.

3. Estimators and Properties

Assume that pairs of random variables (X_1, Y_1), $\cdots$, (X_n, Y_n) are distributed independently and identically. We wish to estimate the regression mean, $m(x)$. Local constant regression and local linear regression use the least squares method through data pairs in the neighborhood of x. In the case of local linear, we find the minimizer of

$$S(a,b) = \sum_{i=1}^{n} (Y_i - a - b(X_i - x))^2 K_h(X_i - x),$$

where $K_h(u) = h^{-1}K(u/h)$, K is a nonnegative and symmetric kernel function, $\int K(u)du = 1$, and h is a bandwidth. The minimizing pair of $(\hat{a}(x), \hat{b}(x))$ is

$$\hat{a}(x) = \frac{r_0(x)s_2(x) - r_1(x)s_1(x)}{s_0(x)s_2(x) - s_1(x)^2}, \quad \hat{b}(x) = \frac{r_1(x)s_0(x) - r_0(x)s_1(x)}{s_0(x)s_2(x) - s_1(x)^2},$$

where

$$r_j(x) = \sum_i (X_i - x)^j K_h(X_i - x) Y_i,$$

$$s_j(x) = \sum_i (X_i - x)^j K_h(X_i - x), \quad j = 0,1,2.$$

Then the estimator of the line is

$$\hat{m}(u|x) = \hat{a}(x) + \hat{b}(x)(u - x).$$

The estimator, $\tilde{m}_L(x)$, suggested by Choi and Hall (1998) is

$$\tilde{m}_L(x) = \lambda_1 \hat{m}_L(x|x + l_1 h) + \hat{m}_L$$

where λ_1, $\lambda_2 > 0$ are weights, $l_1 > 0$, $l_2 < 0$. If we take $\lambda_1 = \lambda_2 = \lambda$ and $l_1 = -l_2 = l_L$, say, then, it enhances not only the symmetrical structure of $\tilde{m}(x)$, but also reduces the conditional bias. It turned out that the constant l is $l_L(\lambda) = \{(1+\lambda)\kappa_2/(2\lambda+1)\}^{1/2}$, where $\kappa_2 = \int u^2 K(u)du$. See the appendix of Choi and Hall (1998) for proof. In fact they showed the following result.

Theorem 1. (Bias of $\tilde{m}_L(x)$, Choi and Hall (1998)) Assume that $m(x)$ has four bounded continuous derivatives in a neighborhood of x that has three bounded continuous derivatives there $f(x) > 0$; that the kernel K is nonnegative, bounded, symmetric with $\int K(u)du = 1$; and that $h = h(n) \rightarrow 0$ and $nh \rightarrow \infty$. Take $l_1 = -l_2$ and $l_L = -l_L(\lambda)$. Then,

$$Bias(\tilde{m}_L(x)) = B_L(x)h^4 + o_p\{h^4 + (nh)^{-1/2}\},$$

where

$$B_L(x) = \frac{1}{16}\left[2(\kappa_2^2 - \kappa_4)\left\{2m''(x)\frac{f''(x)}{f(x)} + 4m'''(x)\frac{f'(x)}{f(x)} + m^{(iv)}(x)\right\} - m^{(iv)}(x)\frac{\kappa_2^2}{\lambda}\right]$$

The proof of Theorem 1 and comments are suggested in Choi and Hall (1998). In local constant smoother case, we wish to minimize

$$S(a) = \sum_{i=1}^{n}(Y_i - a)^2 K_h(X_i - x),$$

and the minimizer is

$$\hat{a}(x) = \frac{\sum K_h(X_i - x)Y_i}{\sum K_h(X_i - x)} \equiv \hat{m}_C(x),$$

which is Nadaraya-Watson estimator of $m(x)$. This estimator has conditional bias of size h^2, i.e.,

$$Bias(\widehat{m_C}(x)) = \frac{h_2}{2}\kappa_2\left(m''(x) + \frac{2m'(x)f'(x)}{f(x)}\right) + o_p\{h^2 + (nh)^{-1/2}\}.$$

We suggest an estimator based on a convex combination of three estimators. One is the Nadaraya-Watson estimator $\hat{m}_C(x)$ and the others come from the local constant line passing two points $(x + l_i h,\ \hat{m}_C(x + l_i h))$ and $(x + 2l_i h, \hat{m}_C(x + 2l_i h))$, $L_i(z)$, say, i.e.,

$$L_i(z) = \frac{\hat{m}_C(x + 2l_i h) - \hat{m}_C(x + l_i h)}{l_i h}(z - x - l_i h) + \hat{m}_C(x + l_i h)$$

Then, the convex combination of $L_i(x)$, $(i = 1,2)$, evaluation of the local constant line at $z = x$, and the estimator $\hat{m}_C(x)$ enables us to produce an estimator, $\tilde{m}_C(x)$,

$$\tilde{m}_C(x) = \frac{\lambda_1 L_1(x) + \hat{m}_C(x) + \lambda_2 L_2(x)}{\lambda_1 + 1 + \lambda_2},$$

where $\lambda > 0$ is weight , $l_1 < 0$, $l_2 > 0$. In order to enhance the symmetrical structure and reduce the conditional bias, take $\lambda_1 = \lambda_2 = \lambda$ and $l_1 = -l_2 = l_C$, say, and it will turn out in Theorem 2 that

$$l_C(\lambda) = \sqrt{\left(\frac{2\lambda + 1}{2\lambda}\right)\kappa_2}.$$

Now, the bias of the proposed estimator $\tilde{m}_C(x)$ is given in the following theorem, and see Appendix for proof.

Theorem 2. (Bias of $\tilde{m}_C(x)$) Assume that $m(x)$ has four bounded continuous derivatives in a neighborhood of x that has three bounded continuous derivatives there and $f(x) > 0$; that the kernel K is nonnegative, bounded, symmetric and

$\int K(u)du = 1$. Take $\lambda_1 = \lambda_2 = \lambda$, $l_1 = -l_2$ and $l_C = l_C(\lambda)$. Then,

$$Bias(\widetilde{m}_C(x) = B_C(x)h^4 + o_p\{h^4 + (nh)^{-1/2}\},$$

where

$$\begin{aligned} B_C(x) = & \frac{m'(x)}{2}\left(\frac{f'''(x)}{3f(x)}\kappa_4 - \frac{f'(x)f''(x)}{f(x)^2}\kappa_2^2\right) + \frac{m''(x)}{4}\frac{f''(x)}{f(x)}(\kappa_4 - \kappa_2^2) \\ & + \frac{m'''(x)}{3!}\frac{f'(x)}{f(x)}\kappa_4 + \frac{m^{(iv)}(x)}{4!} - \frac{\kappa_2}{2\lambda m''(x)}\left(\frac{f'(x)}{f(x)}m'(x) + \frac{m''(x)}{2}\right) \\ & + \left[4\kappa_2\left\{m'(x)\left(\frac{f'''(x)}{2f(x)} - \frac{3f'(x)f''(x)}{2f(x)^2} + (\frac{f'(x)}{f(x)})^3\right) + m''(x)\left(\frac{f'''(x)}{f(x)}\right.\right.\right. \\ & \left.\left. - (\frac{f'(x)}{f(x)})^2\right) + \frac{m'''(x)}{2}\frac{f'(x)}{f(x)} + \frac{m^{(iv)}(x)}{4}\right\} + \frac{7(2\lambda+1)\kappa_2}{2\lambda}\frac{m^{(iv)}(x)}{2m} \\ & \left.\times\left(\frac{f'(x)}{f(x)}m'(x) + \frac{m''(x)}{2}\right)\right]. \end{aligned}$$

We note that the order of the leading term of the bias of the suggested estimator has the same as the order of Choi and Hall (1988)'s estimator even though the constant terms are different. Also, it is clear that the orders of the leading term of the asymptotic variance for each estimator remains unchanged.

4. Concluding Remarks

The local constant and the local linear smoothers have bias of size h^2. By the convex combination of three local linear estimators, Choi and Hall (1998) showed that it can reduce the bias to h^4, bias of a local cubic smoother. Similarly, we showed that the convex combination of three local constant estimators produce an estimator that has the same order of bias as a local cubic smoother at the expense of stronger assumptions such as the existence of $m^{(iv)}$. The suggested estimator is more useful than that of Choi and Hall(1998) especially when the data are very sparse.

As a further research, it will be useful to compute the variance of $\widetilde{m}_C(x)$ even though it requires a very tedious algebra. Evaluation of λ for various kinds of kernel function is worth pursuing. In this case, we choose one of three possible criteria; Minimizing the mean squared error, variance , or squared bias. Also it is worth comparing the leading terms of bias of $\widetilde{m}_L(x)$ and $\widetilde{m}_C(x)$ for various types of $f(x)$ and $m(x)$ with kernel $K(x)$.

Appendix

To prove the Theorem 2 , we require weaker symmetry conditions than those imposed on K. That is we need only $\kappa_1=\kappa_3=0$, κ_5 doesn't have to be zero. Note that

$$\hat{m}(x)=\frac{K_h(X_i-x)Y_i}{K_h(X_i-x)},\ E[\hat{m}(x)|X]=\frac{K_h(X_i-x)E(Y_i|X)}{K_h(X_i-x)}.$$

and

$$E(Y_i|X)=m(x)+m'(x)(X_i-x)+\frac{m''(x)}{2}(X_i-x)^2+\frac{m'''(x)}{3!}(X_i-x)^3 + \frac{m^{(iv)}(x)}{4!}(X_i-x)^4+o\{(X_i-x)^4\}.$$

Expansion of the expectation of $\hat{m}(x)$ results in

$$E(\hat{m}(x)|X)=m(x)+m'(x)\frac{S_1}{S_0}+\frac{m''(x)}{2}\frac{S_2}{S_0}+\frac{m'''(x)}{3!}\frac{S_3}{S_0}+\frac{m^{(iv)}(x)}{4!}\frac{S_4}{S_0} + \frac{m^{(v)}(x)}{5!}\frac{S_5}{S_0}+R(x)\ ,$$

where $S_j=\sum K_h(X_i-x)(X_i-x)^j$, $j=0,1,2,\cdots$, and $R(x)$ is the remainder term. Since $S_j=E(S_j)+O_p(\sqrt{Var(S_j)})$ by tightness, it can be easily shown that, for example, see Fan and Gijbels (1996, pp. 101),

$$S_j=nh^j\Big\{f(x)\kappa_j+f'(x)\kappa_{j+1}h+\frac{f''(x)}{2!}\kappa_{j+2}h^2+\frac{f'''(x)}{3!}\kappa_{j+3}h^3 + \frac{f^{(iv)}(x)}{4!}\kappa_{j+4}h^4+o(h^4)+O_p(\frac{1}{\sqrt{nh}})\Big\}.$$

Then we can derive the following result.

$$\begin{aligned}E(\hat{m}(x)|X)=m(x)&+\left[m'(x)\frac{f'(x)}{f(x)}\kappa_2+m''(x)\frac{\kappa_2}{2}\right]h^2\\&+\left[\frac{m'(x)}{2}\left(\frac{f'''(x)}{3f(x)}\kappa_4-\frac{f'(x)f''(x)}{f(x)^2}\kappa_2^2\right)+\frac{m''(x)}{4}\frac{f''(x)}{f(x)}(\kappa_4-\kappa_2^2)\right.\\&\left.+\frac{m'''(x)}{3!}\frac{f'(x)}{f(x)}\kappa_4+\frac{m^{(iv)}(x)}{4!}\kappa_4\right]h^4\\&+o(h^4)+O_p(\frac{1}{\sqrt{nh}}).\end{aligned}$$

By expanding this result with Taylor series, we prove that, for any fixed l,

$$\begin{aligned}E(\hat{m}_C(x+lh)|X)&=m(x)+lhm'(x)\\&+\left[\kappa_2m'(x)\frac{f'(x)}{f(x)}+m''(x)\frac{\kappa_2}{2}+l^2\frac{m''(x)}{2}\right]h^2\\&+\left[l\kappa_2\Big\{m'(x)\Big(\frac{f''(x)}{f(x)}-(\frac{f'(x)}{f(x)})^2\Big)+m''(x)\frac{f'(x)}{f(x)}+\frac{m'''(x)}{2}\Big\}+l^3\frac{m'''(x)}{3!}\right]h^3\\&+\left[\frac{m'(x)}{2}\left(\frac{f'''(x)}{3f(x)}\kappa_4-\frac{f'(x)f''(x)}{f(x)^2}\kappa_2^2\right)+\frac{m''(x)}{4}\frac{f''(x)}{f(x)}(\kappa_4-\kappa_2^2)\right.\end{aligned}$$

$$
+\frac{m'''(x)}{3!}\frac{f'(x)}{f(x)}\kappa_4+\frac{m^{(iv)}(x)}{4!}\kappa_4+l^2\kappa_2\Bigg\{m'(x)\Bigg(\frac{f'''(x)}{2f(x)}-\frac{3f'(x)f''(x)}{2(f(x))^2}
$$

$$
+(\frac{f'(x)}{f(x)})^3\Bigg)+m''(x)\Bigg(\frac{f''(x)}{f(x)}-(\frac{f'(x)}{f(x)})^2\Bigg)+\frac{m'''(x)}{2}\frac{f'(x)}{f(x)}+\frac{m^{(iv)}(x)}{4}\Bigg\}
$$

$$
+l^4\frac{m^{(iv)}(x)}{4!}\Bigg]h^4+o(h^4)+O_p(\frac{1}{\sqrt{nh}}).
$$

Similarly,

$$
E(\hat{m}_C(x+2lh)|X)=m(x)+2lhm'(x)
$$

$$
+\left[\kappa_2 m'(x)\frac{f'(x)}{f(x)}+m''(x)\frac{\kappa_2}{2}+2l^2\frac{m''(x)}{2}\right]h^2
$$

$$
+\left[2l\kappa_2\Bigg\{m'(x)\Bigg(\frac{f''(x)}{f(x)}-(\frac{f'(x)}{f(x)})^2\Bigg)+m''(x)\frac{f'(x)}{f(x)}+\frac{m'''(x)}{2}\Bigg\}+l^3\frac{4m'''(x)}{3}\right]h^3
$$

$$
+\Bigg[\frac{m'(x)}{2}\Bigg(\frac{f'''(x)}{3f(x)}\kappa_4-\frac{f'(x)f''(x)}{f(x)^2}\kappa_2^2\Bigg)+\frac{m''(x)}{4}\frac{f''(x)}{f(x)}(\kappa_4-\kappa_2^2)
$$

$$
+\frac{m'''(x)}{3!}\frac{f'(x)}{f(x)}\kappa_4+\frac{m^{(iv)}(x)}{4!}\kappa_4+4l^2\kappa_2\Bigg\{m'(x)\Bigg(\frac{f'''(x)}{2f(x)}-\frac{3f'(x)f''(x)}{2(f(x))^2}
$$

$$
+(\frac{f'(x)}{f(x)})^3\Bigg)+m''(x)\Bigg(\frac{f''(x)}{f(x)}-(\frac{f'(x)}{f(x)})^2\Bigg)+\frac{m'''(x)}{2}\frac{f'(x)}{f(x)}+\frac{m^{(iv)}(x)}{4}\Bigg\}
$$

$$
+l^4\frac{m^{(iv)}(x)}{3}\Bigg]h^4+o(h^4)+O_p(\frac{1}{\sqrt{nh}})
$$

It is easy to make a line passing through $(x+l_ih,\ \hat{m}(x+l_ih))$ and $(x+2l_ih,\ \hat{m}(x+2l_ih))$ say $L_i(z)$. Expectation of estimator based on this line is

$$
E[L_i(x)]=m(x)+\Bigg\{\frac{f'(x)}{f(x)}m'(x)\kappa_2+\frac{m''(x)}{2}\kappa_2-l_i^2m''(x)\Bigg\}h^2+l_i^3m'''(x)h^3
$$

$$
+\Bigg\{-2\kappa_2l_i^2m'(x)\Bigg(\frac{f'''(x)}{2f(x)}-\frac{3f'(x)f'''(x)}{2f(x)^2}+(\frac{f'(x)}{f(x)})^3\Bigg)
$$

$$
+m''(x)\Bigg(\frac{f''(x)}{f(x)}-(\frac{f'(x)}{f(x)})^2\Bigg)+\frac{m'''(x)}{2}\frac{f'(x)}{f(x)}+\frac{m^{(iv)}(x)}{4}
$$

$$
+\frac{m'(x)}{2}\Bigg(\frac{f'''(x)}{3f(x)}\kappa_4-\frac{f'(x)f''(x)}{f(x)^2}\kappa_2^2\Bigg)+\frac{m'''(x)}{4}\frac{f''(x)}{f(x)}(\kappa_4-\kappa_2^2)
$$

$$
+\frac{m'''(x)}{3!}\frac{f'(x)}{f(x)}\kappa_4+\frac{m^{(iv)}(x)}{4!}\kappa_4-\frac{7}{2}l_i^4m^{(iv)}(x)\Bigg\}h^4
$$

$$
+o(h^4)+O_p(\frac{1}{\sqrt{nh}}).
$$

Now, consider the conditions that the terms h^2 and h^3 vanish. Those conditions are

$$
\kappa_2(\lambda_1+1+\lambda_2)-(\lambda_1l_1^2+\lambda_2l_2^2)=0
$$

$$
\lambda_1l_1^3+\lambda_2l_2^3=0.
$$

The second equation suggests that if we take $\lambda_1 = \lambda_2 = \lambda$, then $l_1 = -l_2 = l_C$, follows consequently. Then the first equation implies

$$l_C(\lambda) = \sqrt{\left(\frac{2\lambda+1}{2\lambda}\right)\kappa_2}\,.$$

Finally, conditional bias expansion in Theorem 2 follows directly.

References

[1] Choi, E. and Hall, P. (1998). On bias reduction in local linear smoothing. *Biometrika*, Vol. 85, 333-345.

[2] Cleveland, W.S. and Loader ,C. (1996). Smoothing by local regression: principles and methods. In *Statistical Theory and Computational Aspects of Smoothing*, Ed. W.Hardle and M.G.Schimek 10-49 Heidelberg ; Physica-Verlag.

[3] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, Vol. 21, 196-216.

[4] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall , London.

[5] Fan, J., Gijbels, I., Hu, T. and Huang, L. (1996). A sutdy of bandwidth selection for local polynomial regression. *Statistica Sinica*, Vol. 6, 113-127.

[6] Gasser, T and Miller, H.G. (1984). Estimating regression functions and their derivatives by the kernel mothod. *Scandinavian Journal of Statistics*, Vol. 11, 171-185.

[7] Hastie, T. and Loader, C. (1993). Local regression: automatic kernel carpentry (with Comments). *Statistical Science,* Vol. 8, 120-143.

[8] Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability and its Applications,* Vol. 9, 141-142.

[Received April 2006, Accepted June 2006]