Statistica Sinica **7**(1997), 1155-1169

M-TYPE SMOOTHING SPLINES IN NONPARAMETRIC AND SEMIPARAMETRIC REGRESSION MODELS

Jiti Gao and Peide Shi

Queensland University of Technology and Beijing University

Abstract: Consider the regression model $Y_i = g(t_i) + e_i$ for $i = 1, \ldots, n$. Here $-\infty < Y_i, e_i < \infty$, $t_i \in T \subset R^d$, $g \in H$, and H is a specified class of continuous functions from T to R. Based on a finite series expansion $\tilde{g}_n$ of g, an M-estimate $\hat{g}_n$ of g is constructed, and the asymptotic normality of the estimate is investigated. Meanwhile, a test statistic for testing $H_0 : g(\cdot) = g_0(\cdot)$ (a known function) is discussed. We also consider M-estimates for semiparametric regression models and show that they are consistent and asymptotically normal.

Key words and phrases: Asymptotic normality, M-estimation, nonparametotic regression model, semiparametric regression model, spline smoothing technique.

1. Introduction

Consider the model given by

$$Y_i = g(t_i) + e_i, \ i = 1, 2, \ldots, \tag{1}$$

where $t_i = (t_{i1}, \ldots, t_{id})^\tau \in T \subset R^d$ are fixed design points, $g \in H$ is an unknown function over T, and e_i are i.i.d. random errors.

In recent years, nonparametric regression analysis has become an increasingly popular tool for data smoothing. Many of the commonly used estimates of nonparametric regression functions including kernel estimates (Härdle (1990)) and smoothing splines regression estimates (Eubank (1988) and Wahba (1990)) are based on the least squares (LS) method. However, it is well-known that the estimates constructed by using the LS method are sensitive to outliers in the observations, and that the error distribution may be heavy-tailed, deteriorating their performance. Hence a more robust procedure for the model (1) needs to be investigated. In the discussion to Stone's (1977) paper, Brillinger raised the point that a nonlinear M-type estimate of the nonparametric regression curve might be worthwhile to study in order to achieve desirable robustness properties. An important paper on kernel M-estimation of the regression function is Härdle and Gasser (1984). See also Härdle (1984), Härdle and Tsybakov (1988), and Hall and Jones (1990). Härdle and Gasser (1984) showed that these estimates have many of the advantages typically associated with robust inferences.

Other related results about M-estimates for linear models and nonparametric regression models havn been given by a number of authors. See, for example, Yohai and Maronna (1979), Huber (1981), Cox (1983), Portnoy (1984, 1985), Welsh (1989), Chen et al. (1990), Pollard (1991), Davis et al. (1992), Bai, et al. (1992, 1993), and Shi (1992). More recently, Shi and Li (1995), based on the unknown regression approximated by B-spline function, discussed some optimal convergence rates of M-estimates.

In this paper, our objective is to replace the traditional choice $\rho(u) = u^2$ with another choice that is less sensitive to extreme values and, therefore, provides an estimate which is more resistant to the influence of outliers. By using the finite series expansion, we construct M-estimates and establish the asymptotic normality of these estimates for nonparametric and semiparametric regression models. The results given below improve and generalize the related results of Portnoy (1985), Andrews (1991), and Gao et al. (1994).

The organization of this paper is as follows: Section 2.1 investigates the asymptotic normality of M-estimates for nonparametric regression models. Section 2.2 considers the asymptotic normality of M-estimates for semiparametric regression models. Section 3 proposes a computational algorithm. Assumptions and proofs for the results stated in Section 2 are given in Appendix.

2. Main Results

2.1. Asymptotic normality in nonparametric regression models

Consider the regression model given by (1). The objective is to estimate various functions of $g(\cdot)$, such as $g(t)$ and its first derivative $g'(t)$ for arbitrary $t \in T$. The approach taken here is to approximate $g(\cdot)$ by the finite series expansion $\sum_{k=1}^{q} z_k(\cdot)\gamma_{0k}$, where $\{z_k(\cdot); k = 1, 2, \ldots\}$ is a prespecified family of functions from T to R, $\gamma_0 = (\gamma_{01}, \ldots, \gamma_{0q})^\tau$ is an unknown parameter vector, and $q = q_n$ is the number of summands in the series expansion when the sample size is n. In this paper, q_n is taken to be nonrandom. Some results for random data-dependent value of q_n are given in Andrews (1989). The results given below are stated so that they can apply to any family $\{z_k(\cdot); k = 1, 2, \ldots\}$ that satisfies certain properties. Families of particular interest include polynomial, trigonometric, and B-spline functions.

To define the M-estimates we introduce some notation. Let

$$\begin{aligned} Z(\cdot) = Z_q(\cdot) = (z_1(\cdot), \ldots, z_q(\cdot))^\tau,\ & Z = Z_q = (Z(t_1), \ldots, Z(t_n))^\tau, \\ Y = (Y_1, \ldots, Y_n)^\tau, & \\ \text{and} \qquad e = (e_1, \ldots, e_n)^\tau. & \end{aligned} \tag{2}$$

Now, based on Assumption A.1 below and the model $Y_i = Z(t_i)^\tau \gamma_0 + e_i$, we define an M-estimate $\hat{\gamma}_n = \hat{\gamma}_n(q)$ of γ_0 by minimizing

$$\sum_{i=1}^{n} \rho(Y_i - Z(t_i)^\tau \hat{\gamma}_n), \tag{3}$$

where $\rho(\cdot)$ satisfies Assumption A.2 below.

The corresponding M-estimate $\hat{g}_n(\cdot)$ of $g(\cdot)$ is

$$\hat{g}_n(\cdot) = Z(\cdot)^\tau \hat{\gamma}_n. \tag{4}$$

Next, we describe the functions of $g(\cdot)$ that we wish to estimate. Let $G_n = G_n(g)$ denote the estimand, where $G_n(\cdot)$ is a function from H to R^m. The function $G_n(\cdot)$ is defined so as to include some important examples. (See Andrews (1991) for more details.)

The corresponding M-estimate of $G_n(\cdot)$ that we consider is $G_n(\hat{g}_n)$. Since the functional $G_n(\cdot)$ is assumed below to be linear, we have

$$\hat{G}_n = \hat{G}_n(q) = G_n(\hat{g}_n) = G_n(Z(\cdot)^\tau \hat{\gamma}_n(q)) = {L_n}^\tau \hat{\gamma}_n(q), \tag{5}$$

where $L_n(k) = G_n(z_k(\cdot)) \in R^m$ for any k and $L_n = (L_n(1), \ldots, L_n(q))^\tau \in R^{q\times m}$. Now, we give the main result of this section.

Theorem 2.1. *Suppose that Assumptions* A.1 *through* A.4 *below hold. Then as* $n \to \infty$

$$A_n^{-1/2}(\hat{G}_n - G_n) \to_D N(0, A^{-2}\sigma^2), \tag{6}$$

where A is defined by Assumption A.2(ii) *and $A_n = L_n^\tau (Z^\tau Z)^+ L_n$ is nonsingular and therefore $A_n^{-1/2}$ is well-defined under Assumption* A.4.

Remark 2.1. (i) Under very mild assumptions, we obtain the asymptotic normality of M-estimates of a class of functionals. This result generalizes Theorem 1(b) of Andrews (1991).
(ii) Although this section takes the regressor $\{t_i\}$ to be nonrandom, Theorem 2.1 also holds for the random design $\{T_i\}$ if the assumptions of the Theorem hold conditional on $\{T_i\} = \{t_i\}$ with probability 1.

Remark 2.2. Another important problem is testing a null hypothesis $H_0 : g(\cdot) = g_0(\cdot)$, where g_0 is a known function. The hypothesis tests for $H_0 : g(\cdot) = g_0(\cdot)$ have been discussed by some authors. See, for example, Eubank and Hart (1992), Whang and Andrews (1993), Azzalini and Bowman (1993), and Gao (1995a). The details will be given in Section 2.2.

2.2. Asymptotic normality in semiparametric regression models

Consider the regression model given by

$$Y_i = x_i^\tau \beta_0 + g(t_i) + e_i,\ i = 1, 2, \ldots, \tag{7}$$

where $x_i = (x_{i1}, \ldots, x_{ip})^\tau$ ($\subset R^p$, $p \geq 1$) and $t_i = (t_{i1}, \ldots, t_{id})^\tau$ ($\in T \subset R^d$) are known and nonrandom design points, $\beta_0 = (\beta_{01}, \ldots, \beta_{0p})^\tau$ is a vector of unknown parameters, $g \in H$ is an unknown function, e_i are i.i.d. random errors, and p is a finite integer or $p = p_n \to \infty$ as $n \to \infty$. The model (7) belongs to a class of partly linear regression models, which was discussed by some authors. See the recent review by Gao et al. (1994) for more details.

In this section, we investigate the asymptotic normality of the M-estimate $\hat{\beta}_n$ of β_0 based on $g(\cdot)$ approximated by a finite series expansion.

Now, based on Assumption A.6 below and the model $Y_i = x_i^\tau \beta_0 + Z(t_i)^\tau \gamma_0 + e_i$, we define the M-estimates $\hat{\beta}_n = \hat{\beta}_n(q)$ and $\hat{g}_n(\cdot) = Z(\cdot)^\tau \hat{\gamma}_n$ of β_0 and $g(\cdot)$ by minimizing

$$\sum_{i=1}^{n} \rho(Y_i - x_i^\tau \hat{\beta}_n - Z(t_i)^\tau \hat{\gamma}_n), \tag{8}$$

where ρ satisfies Assumption A.2 below.

In the following, we give the main results of this section.

Theorem 2.2. (i) *Suppose that Assumptions* A.2 *and* A.6 *through* A.9 *hold. Let $p \geq 1$ be a finite integer. Then as $n \to \infty$*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to_D N(0, A^{-2} B_0^{-1} \sigma^2), \tag{9}$$

where A *and* B_0 *are defined as in Assumptions* A.2(ii) *and* A.7(i).
(ii) *Suppose that Assumptions* A.2, A.6, *and* A.8 *through* A.11 *hold. Let $p = p_n \to \infty$ as $n \to \infty$ and* $\mathrm{Var}\,(\psi(e_1)^2) < \infty$. *Then*

$$(2p)^{-1/2} \sigma^{-2} (A^2 (\hat{\beta}_n - \beta_0)^\tau X^\tau X (\hat{\beta}_n - \beta_0) - p\sigma^2) \to_D N(0, 1). \tag{10}$$

(iii) *Suppose that Assumptions* A.2, A.10, *and* A.11 *hold. Let $g(\cdot) \equiv 0$ in* (7), *$p = p_n \to \infty$ as $n \to \infty$ and* $\mathrm{Var}\,(\psi(e_1)^2) < \infty$. *Then* (10) *holds.*

Remark 2.3. (i) Theorem 2.2(i) shows that the M-estimate $\hat{\beta}_n$ of β_0 can achieve the optimal convergence rate $n^{-1/2}$ with the smallest possible variance. This result covers some related results for the model (7). (See, Gao et al. (1994) for more details.)
(ii) Simulation studies suggest that when the random errors are normally distributed, the M-estimates are as good as LS estimates; however, when the random errors are drawn from a symmetrically contaminated normal distribution, the M-estimates are superior to LS estimates, and when the radom errors are distributed as Cauchy distribution, the M-estimates seem acceptable but the LS estimates behave poorly. The details of a Monte Carlo study will be given in Section 3 below.

(iii) Theorem 2.2(ii) generalizes Theorem 2.2(i) to the case where the dimension number is large enough. Theorem 2.2(iii) gives the asymptotic normality of the M-estimate $\hat{\beta}_n$ of β_0 for the case where (7) is a simple linear model. The conclusion of Theorem 2.2(iii) is the same as Theorem 3.3 of Portnoy (1985) but the conditions of Theorem 2.2(iii) are weaker than those of Portnoy.

In the following, we give the average squared error (ASE) for the nonparametric regression function estimate, which also provides a reasonable selection for q.

Theorem 2.3. *Suppose that either the conditions of Theorem* 2.2(i) *or the conditions of Theorem* 2.2(ii) *except* $\mathrm{Var}\,(\psi(e_1)^2)<\infty$ *hold. Then for large enough* n

$$ASE(q) = \frac{1}{n}\sum_{i=1}^{n}(\hat{g}_n(t_i) - g(t_i))^2 = O_p(\frac{q}{n}\sigma^2) + o_p(\frac{q}{n}\sigma^2). \tag{11}$$

Remark 2.4. (i) It is clear from the results of Stone (1982) that $q = q_n = O(n^{1/3})$ in (11) is optimal, so $ASE(q)$ can achieve the optimal rate $O(n^{-2/3})$.
(ii) By observing the conditions of Theorems 2.1 through 2.3, we know that the number of the approximation terms q_n affects the asymptotic properties of the proposed estimates. As described in the Algorithm and Remark A.1(ii) below, the optimal q_n is proportional to the reasonable candidate $n^{1/(2r+1)}$ if $g(\cdot)$ has r-order derivative. In fact, some optimal selective procedures for q_n have been proposed by using the GCV criterion. (See, Gao (1995a) and Section 9.3.4 of Hastie and Tibshirani (1990) for more details.)
(iii) The above Theorems 2.1 and 2.2(i) show that under some mild conditions there exist truncation sequences q_n such that the proposed estimates $\hat{G}_n$ and $\hat{\beta}_n$ are asymptotically normal with zero means. Here the Assumptions A.1 and A.6 below are two key smoothness conditions on the regression function $g(\cdot)$, which guarantee that the asymptotic biases can be negligible. Normally, the asymptotically efficient estimate in the kernel case always has a nontrivial bias. (See Härdle (1990) and Hall and Jones (1990).) As a matter of fact, based on q_n selected by using the GCV, the $\hat{G}_n(q_n)$ and $\hat{\beta}_n(q_n)$ in (5) and (8) can be modified to asymptotically efficient estimates.

Another important problem considered in this section is testing the null hypothesis $H_0 : g(\cdot) = g_0(\cdot)$ (a known function). Like Azzalini and Bowman (1993), we can construct a test statistic for the null hypothesis $H_0 : g(\cdot) = g_0(\cdot)$. But here $q = q_n \to \infty$ as $n \to \infty$, so we need to modify the F-statistic used by Azzalini and Bowman (1993). Based on Assumption A.6, the null hypothesis $H_0 : g(\cdot) = g_0(\cdot)$ is equivalent to $H_0' : \gamma_0 = \tilde{\gamma}_0$ (known vector). This suggests using a test statistic of the form

$$F_{1n} = (2q_n)^{-1/2}\sigma^{-2}(A^2(\hat{\gamma}_n - \gamma_0)^\tau Z^\tau Z(\hat{\gamma}_n - \gamma_0) - q\sigma^2), \tag{12}$$

where σ^2 is defined by Assumption A.2(i) below. If σ^2 is unknown, it can be replaced by a consistent estimate $\hat{\sigma}_n^2$ below without affecting Theorem 2.4 below. Here $\hat{\sigma}_n^2$ is defined by

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \psi^2(Y_i - x_i^\tau \hat{\beta}_n - \hat{g}_n(t_i)). \tag{13}$$

Next, we give the last result of this section.

Theorem 2.4. *Suppose that the conditions of Theorem* 2.3*, Assumption* A.5*, and* Var $(\psi(e_1)^2) < \infty$ *hold. Then under* $H_0 : g(\cdot) = g_0(\cdot)$

$$\tilde{F}_{1n}^2 \to_D \chi^2(1) \tag{14}$$

as $n \to \infty$*, where* $\tilde{F}_{1n} = (2q_n)^{-1/2}\sigma^{-2}(A^2(\hat{\gamma}_n - \tilde{\gamma}_0)^\tau Z^\tau Z(\hat{\gamma}_n - \tilde{\gamma}_0) - q\sigma^2)$*. Furthermore, if* $H_1 : g(\cdot) \neq g_0(\cdot)$ *holds, then* $\tilde{F}_{1n} \to_p \infty$.

Remark 2.5. (i) Theorem 2.4 states that $\tilde{F}_{1n}^2$ has an asymptotic standard χ^2 distribution under H_0. In general, H_0 should be rejected if $\tilde{F}_{1n}^2$ exceeds some approximate upper-tail critical value, F_0^2, of the standard χ^2 distribution.
(ii) Theorem 2.4 generalizes the related results of Eubank and Spiegelman (1990) and Gao (1995a). When $\rho(u) = u^2$ in (8), two different test statistics for testing the H_0 were proposed by them.

The proofs of Theorems 2.1 through 2.4 are given in the Appendix.

3. Computational Aspects

In this section, we present some procedures for estimating the parametric components and the unknown smooth function for the model (7). The simulations are based on $g(\cdot)$ approximated by B-spline functioans (see Schumaker (1981) or Nurnberger (1989)). In this situation some related conditions of Theorem 2.2(i) hold. The measures for the estimates $\hat{\beta}_n$ and $\hat{g}_n$ are taken respectively to be

$$\| \hat{\beta}_n - \beta_0 \|^2$$

and

$$\frac{1}{n}\sum_{i=1}^{n} (\hat{g}_n(t_i) - g(t_i))^2. \tag{15}$$

Theoretically speaking, it can be assumed that the knots of B-splines are known and the sample size tends to infinity with n. But dealing with real data, one needs to consider the problem of knot placement and deletion for the finite sample case.

3.1. The algorithm

In practical application of the proposed procedure, the first problem is to determine where the knots of B-spline functions are to be placed and how many knots are to be used.

For a non-negative integer m, let $0 \le s_1^* \le \cdots \le s_m^* \le 1$ and h be a function defined on $[0,1]$. Let $[s_1^*, \ldots, s_m^*]h$ be the divided difference of h defined by

$$[s_i^*]h = h(s_1^*), \tag{16}$$

$$[s_1^*, \ldots, s_m^*]h = \frac{[s^*{}_2, \ldots, s^*{}_m]h - [s^*{}_1, \ldots, s^*{}_{m-1}]h}{s^*{}_m - s^*{}_1},\ s_m^* \neq s_1^* \tag{17}$$

$$[s^*{}_1, \ldots, s^*{}_m]h = \frac{d^{m-1}}{dt^{m-1}} \frac{h(s_1^*)}{(m-1)!},\ s_m^* = s_1^* \in (0,1) \tag{18}$$

$$[s_1^*, \ldots, s_m^*]h = \frac{d_+^{m-1}}{dt^{m-1}} \frac{h(s_1^*)}{(m-1)!},\ s_m^* = s_1^* = 0 \tag{19}$$

and

$$[s_1^*, \ldots, s_m^*]h = \frac{d_-^{m-1}}{dt^{m-1}} \frac{h(s_1^*)}{(m-1)!},\ s_m^* = s_1^* = 1, \tag{20}$$

where $\frac{d_-^{m-1}}{dt^{m-1}}$ and $\frac{d_+^{m-1}}{dt^{m-1}}$ denote the mth left and right derivative operators respectively.

Let $d = 1$ and $T = [0,1]$ in (7). For nonnegative integers r and $K_n = [n^{1/(2r+1)}]$ ($[x]$ denotes the largest integer part of x), let $v_k = k/K_n,\ k = 0, 1, \ldots, K_n$. Let again $q = r + K_n$ and

$$0 = s_1 = \cdots = s_{r+1}, s_{r+2} = v_1, \ldots, s_q = v_{K_n-1}, s_{q+1} = \cdots = s_{r+q+1} = 1,$$

$$z_j(\cdot) = (-1)^{r+1}(s_{j+r+1} - s_j)[s_j, \ldots, s_{j+r+1}](\cdot - s)_+^r,\ j = 1, \ldots, q, \tag{21}$$

and

$$Z(t) = (z_1(t), \ldots, z_q(t))^\tau. \tag{22}$$

Let $m = K_n - 1$ in (16)-(20); for given r, $Z(\cdot)$ is determined by $t_1^*, \ldots, t_m^*$. Hence, we can use the notation $Z_{t_1^*,\ldots,t_m^*}(\cdot)$ instead of $Z(\cdot)$ in this section. Let again

$$S_{t_1^*,\ldots,t_m^*}(\theta) = \frac{1}{n} \sum_{i=1}^n (Y_i - x_i^\tau \beta - Z_{t_1^*,\ldots,t_m^*}(t_i)^\tau \gamma)^2 (1 - q_n/n)^{-2}, \tag{23}$$

where $\theta = (\beta^\tau, \gamma^\tau)^\tau$.

Recall from (8) that $\hat\theta_{t_1^*,\ldots,t_m^*} = (\hat\beta^\tau_{t_1^*,\ldots,t_m^*}, \hat\gamma^\tau_{t_1^*,\ldots,t_m^*})^\tau$ is a vector satisfying

$$\sum_{i=1}^n \rho(Y_i - x_i^\tau \hat\beta_{t_1^*,\ldots,t_m^*} - Z_{t_1^*,\ldots,t_m^*}(t_i)^\tau \hat\gamma_{t_1^*,\ldots,t_m^*}) = \min!, \tag{24}$$

for $m = 1, 2, \ldots$

Let $V_i = (x_i^\tau, Z_{t_1^*,\ldots,t_m^*}(t_i)^\tau)^\tau$, $1 \le i \le n$. For an initial vector $\hat{\theta}^{(1)}$, the well-known Huber iterated procedure is as follows:

$$\hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} + (\sum_{i=1}^n V_i V_i^\tau)^+ \sum_{i=1}^n \psi_c(Y_i - V_i^\tau \hat{\theta}^{(j)}) V_i \tag{25}$$

for $j = 1, 2, \ldots$ At some step j_0, if $|\hat{\theta}^{j_0+1} - \hat{\theta}_0^j| < 0.0001$, then the above iterative procedure is terminated and $\hat{\theta}^{j_0+1} = ((\hat{\beta}^{j_0+1})^\tau, (\hat{\gamma}^{j_0+1})^\tau)^\tau$ is taken to be the desired estimate of $\hat{\theta}_{t_1^*,\ldots,t_m^*} = ((\hat{\beta}_{t_1^*,\ldots,t_m^*})^\tau, (\hat{\gamma}_{t_1^*,\ldots,t_m^*})^\tau)^\tau$. The LS estimate is obtained directly from the model $Y_i = V_i^\tau \theta_0 + e_i$.

For implementing the proposed method, we adopt an automatic and stepwise strategy for knot placement and deletion. In each step, $S_{t_1^*,\ldots,t_m^*}(\hat{\theta})$ is minimized. The first knot t_1^* is placed at the position for which the following equation is satisfied:

$$S_{t_1^*}(\hat{\theta}_{t_1^*}) = \inf_{t_1 \in [0,1]} S_{t_1}(\hat{\theta}_{t_1}). \tag{26}$$

Suppose that when $t_1^*, \ldots, t_{m-1}^*$ have been found, the additional knot t_m^* is placed at the position satisfying

$$S_{t_1^*,\ldots,t_m^*}(\hat{\theta}_{t_1^*,\ldots,t_m^*}) = \inf_{t_m \in [0,1]} S_{t_1^*,\ldots,t_{m-1}^*,t_m}(\hat{\theta}_{t_1^*,\ldots,t_{m-1}^*,t_m})$$

and

$$S_{t_1^*,\ldots,t_m^*}(\hat{\theta}_{t_1^*,\ldots,t_m^*}) < S_{t_1^*,\ldots,t_{m-1}^*}(\hat{\theta}_{t_1^*,\ldots,t_{m-1}^*}). \tag{27}$$

Here $S_{t_1^*,\ldots,t_m^*}(\hat{\theta}_{t_1^*,\ldots,t_m^*})$ is called the score for the knot set $\{t_1^*, \ldots, t_m^*\}$.

If the last inequality is not satisfied, then the knot placement procedure is terminated. The current knots $t_1^*, \ldots, t_m^*$ are taken to be the input of the knot deletion procedure in which the leave-one-out technique is adopted. Assume that the current knots selected are $t_1^*, \ldots, t_j^*$ and the j scores obtained by leaving one knot out are greater than that of the $j+1$ knots $t_1^*, \ldots, t_{j+1}^*$, then the procedure is terminated and the current knots are taken to be the optimal knots selected. Otherwise, delete the knot that is excluded from the knot set on which the minimum score is attained and repeat the above knot deletion procedure for the j knots left.

3.2. Simulation conditions

In the following experiment, we investigate Huber's M-estimates and the LS estimates of β_0 and $g(\cdot)$.

Now, we use Huber's ψ_c-function with $c = 1.5$ and set $p = 3$, $\beta_0 = (1, 2, 3)^\tau$, and $g(t) = 2\cos(3t\pi)$. x_i and t_i are independently drawn from multivariate normal $N(0, I)$ and $U(0, 1)$ respectively. The random errors e_i are independently taken from one of the following distributions:

1. Normal $NOR = NOR(0,1)$;
2. Symmetric Contaminated Normal
   $SCN(0,9) = 0.85NOR(0,1)\ +0.15NOR(0,9^2)$;
3. Cauchy $CAU(0,1)$.

The data in each case of the error distributions consist of $N = 250$ replications of samples of sizes $n = 30$, 50, and 100.

Here, we observe from the above (21) and (22), Theorem 4.27 in Chapter 2 of Nürnberger (1989), and Lemma 5.1 of Shi and Li (1995) that Assumption A.2 holds and Assumptions A.5-A.9 below hold with probability 1.

3.3. Simulation results

In the following, by using the measure (15), we make the comparsions for three kinds of the error distributions, the normal, the symmetric contaminated normal, and Cauchy distributions.

Table 1. The averages of the biases and MSEs of the 250 ME (LSE)

Distribution	n	$R_{BL}[M]$	$MSE[M]$	$R_{BL}[LS]$	$MSE[LS]$
NOR	30	0.22508	0.35938	0.22339	0.35163
NOR	50	0.26303	0.73738	0.49011	1.49514
NOR	100	0.15509	0.41754	0.34259	0.91502
SCN	30	0.44065	1.54111	0.74766	0.90464
SCN	50	0.26303	0.73738	0.49011	1.49514
SCN	100	0.15509	0.41754	0.34259	0.91502
CAU	30	0.54571	2.55320	2.43413	157.46580
CAU	50	0.38473	1.52363	3.11801	215.42201
CAU	100	0.22866	0.91875	4.86162	1117.10426

Table 2. The medians of the biases and the MSEs of the 250 ME (LSE)

Distribution	n	$R_{BL}[M]$	$MSE[M]$	$R_{BL}[LS]$	$MSE[LS]$
NOR	30	0.21510	0.33540	0.21071	0.33066
NOR	50	0.15504	0.21807	0.14558	0.21423
NOR	100	0.09643	0.12249	0.09480	0.11966
SCN	30	0.39357	1.16801	0.66563	2.36988
SCN	50	0.23915	0.46207	0.42071	1.15080
SCN	100	0.14516	0.35872	0.32459	0.66959
CAU	30	0.46656	1.84178	1.06328	6.00122
CAU	50	0.35434	1.19524	1.20970	5.63482
CAU	100	0.21638	0.72569	1.08377	5.51905

The averages of the biases ($R_{BL}[M]$ and $R_{BL}[LS]$) of the M-estimates and the LS estimates of the parametric components and the averages of the mean squared errors ($MSE[M]$ and $MSE[LS]$) of the B-spline M-estimates and the LS estimates for the unknown smooth function are listed in Table 1. The corresponding medians are given in Table 2.

In this section, we only conducted simulations based on $g(\cdot)$ being approximated by the B-splines. In fact, another simulation can be given based on the $g(\cdot)$ approximated by the family of trigonometric functions. (See Gao and Liang (1995b) for details.)

Acknowledgements

The authors wish to thank the Editor, Associate Editors and referees for their constructive suggestions which have greatly improved this paper. The computations referred to in this paper were carried out on computing equipment supplied to the School of Mathematics at the Queensland University of Technology under the Digital Equipment Agreement ERP No. 2057. This research was supported by the Queensland University of Technology Research Council. The research of the second author was supported by the National Natural Sciences Foundation of China.

A. Appendix

A.1. Assumptions

Assumption A.1. For $q = q_n \geq 1$, $q_n \to \infty$ as $n \to \infty$ and $\{z_k(\cdot); k = 1, 2, \ldots\}$ given above, there exists an unknown vector $\gamma_0 = (\gamma_{01}, \ldots, \gamma_{0q})^\tau$ such that for $n \to \infty$

$$Q_n^{1/2} \Big\| \sum_{k=1}^{q} z_k(\cdot)\gamma_{0k} - g(\cdot) \Big\|_{L,T} \to 0, \tag{28}$$

where $\| f \|_{L,T} = \sum_{r:|r|\leq L} \sup_{t\in T} |D^r f(t)|$, $D^r f(t) = \frac{\partial^{|r|}}{\partial t_1^{r_1} \cdots \partial t_d^{r_d}} f(t)$, $|r| = \sum_{i=1}^{d} r_d$, and $Q_n : I_+ \to I_+$ denotes the inverse function of the truncation sequence q_n. It is defined such that $Q_n \geq n$ for any $n \geq 1$. In particular, for any $q \geq 1$, let

$$Q_n = \min\{n \in I_+ : q_k > q \text{ for any } k > n\}.$$

If $q_n \to \infty$ as $n \to \infty$, then Q_n is well-defined on I_+. For example, if $q_n = [n^r]$ for any $n \geq 1$ and some $r \in (0, 1)$, (here $[x]$ denotes the integer part of x), then $Q_n = [q^{1/r}] + a_q$ for any $q \geq 1$, where $a_q = 0$ if $q^{1/r} \in I_+$ and $a_q = 1$ otherwise.

Assumption A.2. (i) $\rho(\cdot)$ is a convex function over $\mathbf{R}$, with left and right derivaties ψ_- and ψ_+, and ψ is a function in $\mathbf{R}$ such that $\psi_- \leq \psi \leq \psi_+$;
(ii) $E\psi(e_1) = 0$ and $E\psi(e_1)^2 = \sigma^2 < \infty$;
(iii) there exist some constants $A \neq 0$, B, and $C > 0$ such that as $|x| \to 0$ $E\psi(e_1 + x) = Ax + Bx^2 + o(|x|^2)$ and $E[\psi(e_1 + x) - \psi(e_1)]^2 \leq C|x|$.

Assumption A.3. (i) $\max_{1\le i\le n} p_{ii} = O(q/n) \to 0$ as $n \to \infty$, where $\{p_{ii}\}$ denotes the ith diagonal element of the $n\times n$ projection matrix $P = Z(Z^\tau Z)^+ Z^\tau$ and $(\cdot)^+$ denotes the Moore-Penrose inverse;
(ii) either Z is full column rank q_n for n large or $G_n(g) = g(t_j)$ for some regressor $\{t_j\}$ that is observed for n large and $z_k(t_j) \neq 0$ for some k;
(iii) $q^{3/2} \max_{1\le i\le n} \parallel z_{in} \parallel \to 0$ as $n \to \infty$, where $\{z_{in}\}$ satisfies $z_{in}^\tau z_{in} = Z(t_i)^\tau (Z^\tau Z)^+ Z(t_i)$ and $\parallel \cdot \parallel$ denotes the Euclidean norm.

Assumption A.4. (i) $G_n(\cdot)$ is a uniformly bounded sequence of linear functional on H. That is, $G_n(\cdot)$ is linear and for some $C < \infty$ and $L > 0$, we have $\parallel G_n(h) \parallel \le C \parallel h \parallel_{L,T}$ for a ll $n \ge 1$ and all $h \in H$;
(ii) $\liminf_{n\to\infty} \lambda_{\min}(L_n^\tau L_n) > 0$, where $\lambda_{\min}(B)$ denotes the smallest eigenvalue of B;
(iii) $\liminf_{n\to\infty} q_n \lambda_{\min}(L_n^\tau (Z^\tau Z/n)^+ L_n) > 0$.

The following Assumption A.5 is sufficient for establishing the asymptotic normality of a generalized quadratic form.

Assumption A.5. (i) There exists a sequence of real numbers j_n ($j_n \to \infty$ as $n \to \infty$) such that for $n \to \infty$

$$j_n q_n^{-1} \max_{1\le i\le n} \sum_{j=1,\neq i}^{n} p_{ij}^2 \to 0; \tag{29}$$

(ii) $q_n^{-1} \max_{1\le i\le n} \lambda_i(P) \to 0$, where $\{\lambda_i(P)\}$ denotes the ith eigenvalues of P.

Assumption A.6. For $q = q_n \ge 1$, $q = q_n \to \infty$ as $n \to \infty$ and $\{z_k(\cdot), k = 1, 2, \ldots\}$ given above, there exists an unknown parameter vector $\gamma_0 = (\gamma_{01}, \ldots, \gamma_{0q})^\tau \in \Theta_2$ such that for n large enough $\sup_{t\in T} |\sum_{i=1}^{q} z_k(t)\gamma_{0k} - g(t)| = o(n^{-1/2})$.

Assumption A.7. (i) As p is a finite integer, there exists a positive definite matrix B_0 with the order $p \times p$ such that

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} x_i x_i^\tau = B_0. \tag{30}$$

(ii) Assume that X and Z satisfy orthogonality condition

$$X^\tau Z = 0, \tag{31}$$

where $X = (x_1, \ldots, x_n)^\tau$.

Assumption A.8. (i) Z is of full column rank q for n large enough, and $\liminf_{n\to\infty} q_n \lambda_{\min}(Z^\tau Z/n) > 0$;
(ii) $\max_{1\le i\le n} p_{ii} = O(q/n) \to 0$ as $n \to \infty$, where $\{p_{ij}\}$ denotes the ith row element and jth rank element of the $n \times n$ projection matrix $P = Z(Z^\tau Z)^+ Z^\tau$.

The following Assumption A.9 is needed for the case where p is a finite integer or $p = p_n \to \infty$ as $n \to \infty$.

Assumption A.9. $r_n^3 \max_{1\le i\le n}\{x_i^\tau (X^\tau X)^+ x_i + Z(t_i)^\tau (Z^\tau Z)^+ Z(t_i)\} \to 0$ as $n \to \infty$, where $r_n = \max(p, q_n)$.

As $p = p_n \to \infty$, we need the following assumption.

Assumption A.10.

(i) X is of full column rank p for n large enough, and $\liminf_{n\to\infty}\lambda_{\min}(X^\tau X/n) > 0$;

(ii) $\max_{1\le i\le n} r_{ii} = O(p/n) \to 0$ as $n \to \infty$, where $\{r_{ij}\}$ denotes the ith row elememt and jth rank element of the $n \times n$ projection matrix $R = X(X^\tau X)^+ X^\tau$;

(iii) $p_n^3 \max_{1\le i\le n}[x_i^\tau (X^\tau X)^+ x_i] \to 0$ as $n \to \infty$.

The following Assumption A.11 is sufficient for establishing the asymptotic normality of a generalized quadratic form.

Assumption A.11. (i) There exists a sequence of real numbers k_n ($k_n \to \infty$ as $n \to \infty$) such that for $n \to \infty$

$$k_n p_n^{-1} \max_{1\le i\le n} \sum_{j=1,\neq i}^{n} r_{ij}^2 \to 0;$$

(ii) $p_n^{-1} \max_{1\le i\le n} \lambda_i(R) \to 0$, where $\lambda_i(R)$ denotes the eigenvalues of R.

Some important remarks for the Assumptions A.1-A.11 above can be found in Gao (1995c).

A.2. Lemmas

Lemma A.1. *Under either the conditions of Theorem* 2.2(i) *or those of Theorem* 2.2(ii) *except* $\mathrm{Var}\,(\psi(e_1)^2) < \infty$, *we have*

$$\lim_{n\to\infty} r_n^3 v_i^\tau (V^\tau V)^+ v_i = 0, \tag{32}$$

where $v_i = (x_i^\tau, Z(t_i)^\tau)^\tau$, $V = (v_1, \ldots, v_n)^\tau$, $(V^\tau V)^+ = (v^{ij})_{1\le i,j\le 2}$, $v^{11} = (X^\tau X)^+$, $v^{22} = (Z^\tau Z)^+$, *and* $v^{12} = v^{21} = 0$.

Lemma A.2. (i) *Under the conditions of Theorem* 2.1, *we have as* $n \to \infty$

$$\hat\gamma_n - \gamma_0 = A^{-1}(Z^\tau Z)^+ \sum_{i=1}^{n} Z(t_i)\psi(e_i) + o_p((\lambda_Z)^{-1/2}), \tag{33}$$

where $\lambda_A = \lambda_{\min}(A^\tau A)$ *throughout this paper.*

(ii) *Under either the conditions of Theorem* 3.1(i)*or those of Theorem* 3.1(ii) *except* Var $(\psi(e_1)^2) < \infty$, *we have as* $n \to \infty$

$$\hat{\theta}_n - \theta_0 = A^{-1}(V^\tau V)^+ \sum_{i=1}^n v_i \psi(e_i) + o_p((\lambda_V)^{-1/2}), \tag{34}$$

where $\hat{\theta}_n = (\hat{\beta}_n^\tau, \hat{\gamma}_n^\tau)^\tau$ *and* $\theta_0 = (\beta_0^\tau, \gamma_0^\tau)^\tau$.

The proofs of Lemmas A.1–A.2 can also be found in Gao (1995c). Here we omit the details.

A.3. Proofs of Theorems

In the following, we only give an outline of the proofs. The details can also be found in Gao (1995c).

A.3.1. Proof of Theorem 2.1

By Assumptions A.1, A.3, A.4, and Lemma A.2(i), we have

$$\begin{aligned}&\hat{G}_n - G_n = \hat{G}_n - \tilde{G}_n + \tilde{G}_n - G_n \\ &= L_n^\tau [(Z^\tau Z)^+ A^{-1} \sum_{i=1}^n Z(t_i)\psi(e_i) + o_p((\lambda_Z)^{-1/2})] + G_n(\tilde{g}_n - g),\end{aligned} \tag{35}$$

where $\tilde{g}_n = Z\gamma_0$.

Now, by checking the Lindeberg condition and using Assumptions A.1, A.3, and A.4, we can get the proof of Theorem 2.1. In fact, the proof of Theorem 2.1 can also be obtained by using similar reasoning as in (A.1) through (A.10) of Andrews (1991).

A.3.2. Proofs of Theorems 2.2 through 2.4

(i) By the definition of $(V^\tau V)^+$ (see Lemma A.1) and Lemma A.2(ii), we have

$$\hat{\beta}_n - \beta_0 = (X^\tau X)^{-1} X^\tau \Psi(e) + o_p((\lambda_X)^{-1/2}) \tag{36}$$

and

$$\hat{\gamma}_n - \gamma_0 = (Z^\tau Z)^{-1} Z^\tau \Psi(e) + o_p((\lambda_Z)^{-1/2}), \tag{37}$$

where $\Psi(e) = (\psi(e_1), \ldots, \psi(e_n))^\tau$.

Thus, using Assumption A.7 and checking the Lindeberg condition again, we obtain the proof of Theorem 2.2(i) immediately.

(ii) Theorem 2.2(iii) is a special case of Theorem 2.2(ii). Its proof is trivial. The proof of Theorem 2.2(ii) follows similarly from that of Theorem 2.2 of Gao (1995a) by using (36) and Assumption A.10(i).

(iii) The proof of Theorem 2.3 follows from that of Theorem 2.6 of Gao and Liang (1995b) by using (37) and Assumption A.8(i).
(iv) The proof of Theorem 2.4 follows similarly from that of Theorem 2.2(ii) by using Assumption A.5. We have now finished all proofs.

References

Andrews, D. W. K. (1989). Asymptotic normality of series estimates of nonparametric and semiparametric regression models. Cowles Foundation Discussion paper No. 874R. Yale University.

Andrews, D. W. K. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* **59**, 307-345.

Azzalini, A. and Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc. Ser. B* **55**, 549-557.

Bai, Z. D., Rao, C. R. and Wu, Y. H. (1992). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statist. Sinica* **2**, 237-254.

Bai, Z. D., Rao, C. R. and Zhao, L. C. (1993). MANOVA type tests under a convex discrepancy function for the standard multivariate linear model. *J. Statist. Plann. Inference* **36**, 77-90.

Chen, X. R., Bai, Z. D., Zhao, L. C. and Wu, Y. H. (1990). Asymptotic normality of minimum L_1-norm estimates in linear models. *Science in China Ser. A* **33**, 1311-1328.

Cox, D. D. (1983). Asymptotics for M-type smoothing splines. *Ann. Statist.* **11**, 530-551.

Davis, R. A., Knight, K. and Liu, J. (1992). M-estimation for autoregressions with infinite variance. *Stochastic Process. Appl.* **40**, 145-180.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* Marcel-Dekker, New York.

Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression via order selection criterion. *Ann. Statist.* **20**, 1412-1425.

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85**, 387-392.

Gao, J. T. (1992). *A Large Sample Theory in Semiparametric Regression Models.* Ph.D. thesis, Graduate School at the University of Science and Technology of China, Hefei, Anhui, P. R. China.

Gao, J. T., Hong, S. Y., Liang, H. and Shi, P. D. (1994). Survey of Chinese work on semiparametric regression models. *Chinese J. Appl. Probab. Statist.* **10**, 96-104.

Gao, J. T. (1995a). Parametric and nonparametric tests in semiparametric regression model. *Comm. Statist.-Theory Methods* **26**, 783-800.

Gao, J. T. and Liang, H. (1995b). Statistical inference in semiparametric single-index and partially nonlinear regression models. *Ann. Inst. Statist. Math.* **49** (in press).

Gao, J. T. (1995c). M-type spline smoothing estimation in nonparametric and semiparametric regression models. Technical report, Department of Statistics, University of Auckland, New Zealand.

Hall, P. and Jones, M. (1990). Adaptive M-estimation in nonparametric regression. *Ann. Statist.* **18**, 1712-1728.

Härdle, W. and Gasser, T. (1984). Robust nonparametric function fitting. *J. Roy. Statist. Soc. Ser. B* **46**, 42-51.

Härdle, W. (1984). Robust regression function estimation. *J. Multivarate Anal.* **14**, 169-180.

Härdle, W. and Tsybakov, A. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.* **16**, 120-135.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, New York.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman and Hall.

Huber, P. (1981). *Robust Statistics.* John Wiley, New York.

Nürnberger, G. (1989). *Approximation by Spline Functions.* Springer, New York.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econom. Theory* **7**, 186-199.

Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I: Consistency. *Ann. Statist.* **12**, 1298-1309.

Portnoy, S. (1985). Asymptotic behavior of M-estimatots of p regression parameters when p^2/n is large. II: Normal approximation. *Ann. Statist.* **13**, 1403-1417. (correction: *Ann. Statist.* **19**, No. 4, p.2282).

Schumaker, L. (1981). *Spline Functions.* John Wiley, New York.

Shi, P. D. (1992).*Asymptotic Behavior of Piecewise Polynomialand B-Spline M-estimates in Regression.* Ph. D. thesis, Institute of Systems Science, Academia Sinica, P. R. China.

Shi, P. D. and Li, G. Y. (1995). Global convergence rates of B-spline M-estimators in nonparametric regression. *Statist. Sinica* **5**, 303-318.

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5**, 595-645.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040-1053.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, Philadelphia.

Welsh, A. H. (1989). On M-processes and M-estimation. *Ann. Statist.* **17**, 337-361. (correction: *Ann. Statist.* **18**, No. 3, p.1500).

Whang, Y. J. and Andrews, D. W. K. (1993). Tests of specification for parametric and semiparametric models. *J. Econometrics* **57**, 277-318.

Yohai, V. J. and Maronna, R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist.* **7**, 258-268.

School of Mathematics, The Queensland University of Technology, Gardens Point, GPO Box 2434, Brisbane, Qld 4001, Australia.

E-mail: j.gao@fsc.qut.edu.au

Department of Probability and Statistics, Beijing University, Beijing 100852, China.

E-mail: p.shi@sunrise.pku.edu.cn

(Received July 1995; accepted October 1996)