The Annals of Statistics
1991, Vol. 19, No. 2, 724–740

ESTIMATING A SMOOTH MONOTONE REGRESSION FUNCTION[1]

BY ENNO MAMMEN

Universität Heidelberg

The problem of estimating a smooth monotone regression function m will be studied. We will consider the estimator m_{SI} consisting of a smoothing step (application of a kernel estimator based on a kernel K) and of a isotonisation step (application of the pool adjacent violator algorithm). The estimator m_{SI} will be compared with the estimator m_{IS} where these two steps are interchanged. A higher order stochastic expansion of these estimators will be given which show that m_{SI} and m_{IS} are asymptotically first order equivalent and that m_{IS} has a smaller mean squared error than m_{SI} if and only if the kernel function of the kernel estimator is not too smooth.

0. Introduction. The problem of estimating a smooth monotone regression function m will be studied. Two estimators m_{SI} and m_{IS} are compared. m_{SI} consists of two steps: (i) smoothing of the data by a kernel estimator; (ii) isotonisation of the data by the pool adjacent violator algorithm. The estimator m_{IS} is constructed by interchanging these two steps. Estimates similar to m_{IS} or to m_{SI}, respectively, have been studied for instance by Cheng and Lin (1981), Wright (1982), Friedman and Tibshirani (1984), Mukerjee (1988) and in the context of estimation of a monotone density or failure rate function by Barlow and van Zwet (1969, 1970). For an application, see also Hildenbrand and Hildenbrand (1985).

We consider the asymptotic stochastic behavior of these estimators at a fixed point x_0, where the function m is assumed to be strictly monotone and smooth. If the bandwidth of the kernel estimator is chosen in the optimal order $n^{-1/5}$, the usual kernel estimator m_S is monotone with probability tending to 1 and therefore equal to m_{SI} (Theorem 1). As a kernel estimator, however, m_S is an estimator whose construction is only motivated by the smoothness of m, whereas m_{IS} is a modification of m_S taking care of the information that m is monotone. It will be shown that $m_{SI}(x_0)$ and $m_{IS}(x_0)$ are of order $n^{-2/5}$ and that they are asymptotically equivalent in first order. But $m_{SI}(x_0) - m_{IS}(x_0)$ is of the only slightly lower order $n^{-8/15} = n^{-2/5}n^{-2/15}$ (Theorem 2). In simulations reported in Section 5, we will see that the difference between m_{SI} and m_{IS} cannot be neglected for moderate sample sizes. This shows that for a satisfactory comparison of m_{SI} and m_{IS}, an asymptotic higher order analysis is necessary. In Theorem 3, we will give stochastic higher order expansions of $m_{SI}(x_0)$ and $m_{IS}(x_0)$. These expansions

Received October 1987; revised September 1989.
[1]Work supported by the Deutsche Forschungsgemeinschaft.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62J02, 62E20.
Key words and phrases. Nonparametric regression, isotonic regression, kernel estimator.

entail that $m_{IS}(x_0)$ has always a smaller variance and a larger bias than $m_{SI}(x_0)$. Furthermore, this result implies that mainly it depends on the kernel function K of the chosen kernel estimator if one should prefer the estimator m_{SI} or m_{IS}: If the bandwidth of the kernel estimator m_S is chosen such that the mean squared error is asymptotically minimized, then $m_{IS}(x_0)$ has asymptotically a smaller mean squared error than $m_{SI}(x_0)$ if and only if $\int K^2(t)\,dt[\int t^2 K(t)\,dt \int K'(t)^2\,dt]^{-1}$ is smaller than a universal constant.

1. Assumptions. For simplification, we will assume that $x_0 = 0$ and that the design points x_i are at equal distance: $x_i = i/n$ $(i = 0, \pm 1, \dots, \pm n)$. The model is

$$y_i = m(x_i) + \varepsilon_i \qquad (-n \le i \le n), \tag{1.1}$$

where

(1.2) the random variables ε_i are i.i.d. with $E\varepsilon_i = 0$ for $-n \le i \le n$. The Laplace transform $E\exp(t\varepsilon_i)$ is assumed to exist for $|t|$ small enough.

Furthermore, the regression function m: $[-1, 1] \to \mathbf{R}$ is assumed to be sufficiently smooth and monotone:

(1.3) m is two times continuously differentiable.

(1.4) $m'(x) \ge 0$ for $x \in [-1, 1]$ and $m'(0) > 0$.

For the kernel K: $\mathbf{R} \to \mathbf{R}$ which is used in the construction of the estimates m_S, m_{IS}, m_{SI} we assume

(1.5) K is continuous.

(1.6) K is a symmetric probability density function vanishing outside of a compact set. Furthermore, outside a finite set of points K is two times continuously differentiable with bounded second derivative.

2. Construction of the estimates. For a sequence of bandwidths h_n, the estimates m_S, m_I, m_{SI} and m_{IS} are defined as follows (for simplification we do not indicate in the notation that these estimates depend on n.) In the case of equidistant design points, the kernel estimator may be defined as

$$m_S(x) = \frac{1}{nh_n} \sum_{i=-n}^{n} K\left(\frac{x - x_i}{h_n}\right) Y_i. \tag{2.1}$$

m_{SI} is defined as the $L_2([-1, 1])$-projection of m_S onto the monotone functions

$$\int (m_{SI}(x) - m_S(x))^2\,dx = \inf_{g \text{ monotone}} \int (g(x) - m_S(x))^2\,dx. \tag{2.2}$$

The estimate m_{SI} is a slight modification of an estimate introduced by

Friedman and Tibshirani (1984). A similar estimate has been proposed by Wright (1982) [see also Barlow and van Zwet (1969, 1970)]. The estimate m_{IS} is constructed by interchanging the smoothing step and the isotonisation step

$$m_{IS}(x) = \frac{1}{nh_n} \sum_{i=-n}^{n} K\left(\frac{x - x_i}{h_n}\right) Y_i^*, \tag{2.3}$$

where (Y_i^*) is the least-squares projection of (Y_i) onto the monotone tuples $\{(Z_i): Z_{-n} \le \cdots \le Z_n\}$ (isotonic least squares regression):

$$\sum_{i=-n}^{n} \left(Y_i^* - Y_i\right)^2 = \inf_{Z_i \text{ monotone}} \sum_{i=-n}^{n} (Z_i - Y_i)^2. \tag{2.4}$$

The linear interpolation of (Y_i^*) will be called m_I. The estimator m_{IS} has been proposed in Mukerjee (1988). The asymptotic distribution of the process $(Y_i^*: -n \le i \le n)$ is given by Groeneboom (1985, 1989). For a further discussion of isotonic regression, we refer to Barlow, Bartholomew, Bremmer and Brunk (1972). We want only mention that $\sum_{i \le k} Y_i^*$ is the greatest convex minorant of $\sum_{i \le k} Y_i$ and that

$$Y_i^* = \min_{v \ge i} \max_{u \le i} \frac{1}{v - u} \sum_{j=u}^{v} Y_j. \tag{2.5}$$

Furthermore, (2.2) implies that $\int_{-1}^{x} m_{SI}(t)\, dt$ is the greatest convex minorant of $\int_{-1}^{x} m_S(t)\, dt$ and that

$$m_{SI}(x) = \inf_{v \ge x} \sup_{u \le x} \frac{1}{v - u} \int_u^v m_S(t)\, dt. \tag{2.6}$$

3. Results. It is well known that for two times continuously differentiable regression functions optimal choices of the bandwidth h_n of the kernel estimator m_S are of order $n^{-1/5}$. Without loss of generality, we assume

$$h_n = n^{-1/5}. \tag{3.1}$$

For this choice of bandwidth the derivative of m_S is a consistent estimate of m'. This will show the statement of the following theorem.

THEOREM 1. *Assume* (1.1), ..., (1.6), (3.1). *Then* $m_{SI}(0) = m_S(0)$ *with probability tending to* 1.

In the next theorem, we show that $m_{SI}(x_0)$ and $m_{IS}(x_0)$ are asymptotically equivalent in first order but that $m_{SI}(x_0) - m_{IS}(x_0)$ is of the only slightly lower order $n^{-8/15} = n^{-2/5} n^{-2/15}$. The proof of Theorem 2 (and Theorem 3) will be based on the following representations: by Theorem 1 we get, with a probability tending to 1

$$m_{SI}(0) = n^{1/5}\int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s)\,dP_n(s),$$

$$m_{IS}(0) - m_{SI}(0) = n^{1/5}\int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s)\,d(P_n^c - P_n)(s)$$

$$= -n^{2/5}\int_{-n^{-1/5}}^{n^{-1/5}} K'(n^{1/5}s)(P_n^c(s) - P_n(s))\,ds$$

where P_n is the partial sum process, defined by $P_n(s) = n^{-1}\Sigma_{i \le ns} Y_i$, $s \in [-1, 1]$, and where P_n^c is the convex minorant of P_n. Results in Prakasa Rao (1969) and Kiefer and Wolfowitz (1976) suggest that the distance between P_n and P_n^c is of order $O_P(n^{-2/3})$ at fixed points. Moreover, the results in Groeneboom (1985, 1989) suggest that, for fixed $M > 0$ and $s_0 \in (-1, 1)$, the process $\{n^{2/3}[P_n^c(s_0 + n^{-1/3}s) - P_n(s_0 + n^{-1/3}s)]: s \in [-M, M]\}$ converges in distribution to a nondegenerate limiting process. If this process is sufficiently mixing, this would give that

$$n^{8/15}(m_{IS}(0) - m_{SI}(0))$$
$$= -n^{-1/15}\int_{-n^{2/15}}^{n^{2/15}} K'(n^{-2/15}s)\,n^{2/3}\big(P_n^c(sn^{-1/3}) - P_n(sn^{-1/3})\big)\,ds$$

has a nondegenerate limiting distribution.

THEOREM 2. *Assume* (1.1), ..., (1.6), (3.1). *Then*

$$m_{IS}(0) = m_S(0) + O_P(n^{-8/15}). \tag{3.2}$$

In the next theorems, we will give a stochastic higher order expansion of m_{SI} and m_{IS}. This expansion will be used in the next section for an asymptotic comparison of the bias and variance of m_{SI} and m_{IS}.

THEOREM 3. *Assume* (1.1), ..., (1.6), (3.1). *Then there exist independent random variables* $U_{1,n}$ *and* $U_{2,n}$ *with* $EU_{1,n} = 0$ *and* $EU_{2,n} = 0$ *such that for some universal positive constants* c_1, c_2, *and* c_3, *the following hold*:

$$m_{SI}(0) = m(0) + \beta_n + U_{1,n} + o_P(n^{-2/3}), \tag{3.3}$$

$$m_{IS}(0) = m(0) + \beta_n + \delta_n + (1 - \varepsilon_n)U_{1,n} + U_{2,n} + o_P(n^{-2/3}), \tag{3.4}$$

where

$$\beta_n = Em_S(0) - m(0) = \tfrac{1}{2}m''(0)\int t^2K(t)\,dt\,n^{-2/5} + o(n^{-2/5}),$$

$$\delta_n = c_3\sigma^{4/3}m''(0)m'(0)^{-4/3}n^{-2/3}, \tag{3.5}$$

$$\varepsilon_n = c_2\sigma^{4/3}m'(0)^{-4/3}\int (K'(t))^2\,dt\left(\int K^2(t)\,dt\right)^{-1} n^{-4/15}.$$

Furthermore, $n^{2/5}U_{1,n}$ *and* $n^{8/15}U_{2,n}$ *are asymptotically normal with variances* $\sigma^2 \int K^2(t)\,dt$ *and* $c_1\sigma^{10/3}(m'(0))^{-4/3}\int (K'(t))^2\,dt$*, respectively.*

The consequences of Theorem 3 will be discussed in the next section. Theorem 1 and 3 do not hold if K is not continuous. Consider for instance the rectangle kernel $K_R(x) = \frac{1}{2}1(|x| \le 1)$. Then $m_S(0) = \frac{1}{2}n^{1/5}(P_n(n^{-1/5}) - P_n(-n^{-1/5}))$ and $m_{IS}(0) = \frac{1}{2}n^{1/5}(P_n^c(n^{-1/5}) - P_n^c(-n^{-1/5}))$. This gives $m_{IS}(0) - m_S(0) = O_P(n^{-7/15})$ because of $P_n^c(s) - P_n(s) = O_P(n^{-2/3})$.

THEOREM 4. *Assume* (1.1), ..., (1.4), (1.6), (3.1). *K is assumed to be continuous only outside a finite set of points and to make at these points jumps* $\Delta_i \neq 0$ $(i = 1, \dots, 2I)$. *Then*

$$m_{SI}(0) = m_S(0) + o_P(n^{-8/15}), \tag{3.6}$$

$$m_{IS}(0) = m_S(0) + O_P(n^{-7/15}). \tag{3.7}$$

There exist universal positive constants c_4 *and* c_5 *and independent random variables* $U_{1,n}$ *and* $(E_n, V_{2,n})$ *with* $EU_{1,n} = 0$, $EV_{2,n} = 0$, $E_n = O_P(n^{-2/15})$ *and* $E(E_n) = c_5\sigma^{2/3}m'(0)^{-2/3}\sum_{i=1}^{2I}\Delta_i^2(\int K^2(t)\,dt)^{-1}n^{-2/15}$ *such that the following hold*:

$$m_{SI}(0) = m(0) + \beta_n + U_{1,n} + o_P(n^{-8/15}), \tag{3.8}$$

$$m_{IS}(0) = m(0) + \beta_n + (1 - E_n)U_{1,n} + V_{2,n} + o_P(n^{-8/15}), \tag{3.9}$$

where $\beta_n = Em_S(0) - m(0) = \frac{1}{2}m''(0)\int t^2K(t)\,dt\,n^{-2/5} + o(n^{-2/5})$. *Furthermore,* $n^{2/5}U_{1,n}$ *is asymptotically normal with variances* $\sigma^2\int K^2(t)\,dt$ *and* $n^{7/15}V_{2,n}$ *has a nondegenerate limiting distribution with variance* $c_4\sigma^{8/3}m'(0)^{-2/3}\sum\Delta_i^2$.

4. Interpretation of the results.

REMARK 1. Theorem 3 can be used to calculate the second order asymptotic variance (as. var.) and bias (as. bias) of $m_{IS}(0)$ and $m_{SI}(0)$ for the kernel $K_h(t) = (1/h)K(t/h)$—this would correspond to the use of the kernel K with bandwidth $h_n = hn^{-1/5}$ in the case of n observations [see (3.1)].

$$\text{as. var.}(m_{SI}(0)) = \sigma^2\frac{1}{h}\int K^2(t)\,dt\,n^{-4/5},$$

$$\text{as. var.}(m_{IS}(0)) = \sigma^2\frac{1}{h}\int K^2(t)\,dt\,n^{-4/5} + (c_1 - 2c_2)\sigma^{10/3}m'(0)^{-4/3}\frac{1}{h^3}\int (K'(t))^2\,dt\,n^{-16/15},$$

$$\text{as. bias}(m_{SI}(0))^2 = \beta_n^2,$$

$$\text{as. bias}(m_{IS}(0))^2 = \beta_n^2 + c_3\sigma^{4/3}m''(0)^2m'(0)^{-4/3}h^2\int t^2K(t)\,dt\,n^{-16/15}.$$

The simulations of the next section suggest that $c_1 < 2c_2$. Therefore, isotonising the observations leads to a variance reduction and a larger bias. Furthermore $m_{IS}(0)$ has a smaller asymptotic mean squared error than $m_{SI}(0)$ if and only if

$$\frac{h^5 \int t^2 K(t)\, dt\, m''(0)^2}{\sigma^2 \int K'(t)^2\, dt} < \frac{2c_2 - c_1}{c_3}.$$

In the special case where h is chosen such that the asymptotic mean squared error of $m_S(0)$ and $m_{SI}(0)$ is minimized [roughly this may be the case if the integrated mean squared error is nearly minimized and $m''(s)$ varies not too much], this is equivalent to

$$\frac{\int K^2(t)\, dt}{\int t^2 K(t)\, dt \int K'(t)^2\, dt} < \frac{2c_2 - c_1}{c_3}.$$

Note that the left-hand side depends only on the chosen kernel K. This term is large for smooth kernel K. This suggests not to isotonise the observations (i.e., to use m_{IS}) if K is smooth (see also Section 5).

REMARK 2. If K is discontinuous and fulfills the conditions of Theorem 4, then $m_{IS}(0)$ and $m_{SI}(0)$ have the following second order asymptotic variance and bias if the kernel $K_h(t) = (1/h)K(t/h)$ is used (i.e., $h_n = hn^{-1/5}$):

$$\begin{aligned}
\text{as. var.}(m_{SI}(0)) &= \sigma^2 \frac{1}{h} \int K^2(t)\, dt\, n^{-4/5},\\
\text{as. var.}(m_{IS}(0)) &= \sigma^2 \frac{1}{h} \int K^2(t)\, dt\, n^{-4/5}\\
&\qquad + (c_4 - 2c_5)\sigma^{8/3} m'(0)^{-2/3} \sum \Delta_i^2 n^{-14/15},\\
\text{as. bias}(m_{SI}(0))^2 &= \text{as. bias}(m_{IS}(0))^2 + o(n^{-14/15}) = \beta_n^2 + o(n^{-4/5}).
\end{aligned}$$

The simulations of the next section suggest that $c_4 < 2c_5$. Therefore isotonising the observations leads to a variance reduction. Furthermore $m_{IS}(0)$ has always a smaller asymptotic mean squared error than $m_{SI}(0)$. For small sample sizes this should be carefully interpreted because of the small differences of the orders of convergence.

REMARK 3. Theorems 3 and 4 are examples that higher order stochastic expansions should be carefully interpreted. In Theorem 3, δ_n and $\varepsilon_n U_{1,n}$ are of smaller order than $U_{2,n}$. Therefore, $m_{IS}(0) = \beta_n + m(0) + U_{1,n} + U_{2,n} + o_P(n^{-8/15})$ with $U_{1,n}$ and $U_{2,n}$ independent. But nevertheless, $m_{SI}(0) = \beta_n + m(0) + U_{1,n} + o_P(n^{-8/15})$ has a larger asymptotic variance than $m_{IS}(0)$.

REMARK 4. If m is increasing but $m'(0) = 0$, then $m_S(0)$ is no more asymptotically equivalent to $m_{SI}(0)$. Then the random bandwidth of m_I (i.e., the distance between neighboring wedges of S_n^c) is significantly larger (at least of stochastic order $n^{-1/7}$) than $n^{-1/5}$ [see Wright (1981), Leurgans (1982)]. Simulations suggest that then isotonisation of the observations leads to a bias reduction and a variance reduction [compared with $m_S(0)$ or $m_{SI}(0)$].

REMARK 5. The assumptions that the design points are equally spaced and that K has compact support can be weakened. Furthermore (Y_i^*) and m_{IS} can also be weighted least squares estimators.

REMARK 6. For the validity of the theorem it is not necessary to assume that m is monotone on the full interval $[-1, 1]$. For instance, it suffices to assume that m is monotone in a neighborhood of 0 and bounded away from $m(0)$ outside this neighborhood.

REMARK 7. An analogous result for the estimation of (locally) isotone densities can be derived in a straightforward way.

5. Some simulations. We have carried out 100 simulations for the following regression functions:

$$m_1(x) = \exp(x/2) \qquad (0 \le x \le 1)$$

$$m_2(x) = \exp(x) \qquad (0 \le x \le 1)$$

$$m_3(x) = \exp(2x) \qquad (0 \le x \le 1)$$

$$m_4(x) = \text{const.} \qquad (0 \le x \le 1)$$

The n random variables ε_i are assumed to have a Gaussian distribution $N(0, 1)$. The design points are taken equispaced on $[0, 1]$ $x_i := i/n$. n is taken as 200. For the triangle kernel $K_T(u) := (1 - |u|)^+$, the mean squared error $\text{MSE}(m_*) = E(m_*(x_0) - m(x_0))^2$ of $m_* = m_{IS}, m_{SI}, m_S$ and m_I has been estimated for $x_0 = \frac{1}{2}$. Furthermore, the proportion $I_{SI/IS}$ of cases has been evaluated where the squared error of m_{SI} was less than the squared error of m_{IS}. The results are summarized in Table 1. For every row of Table 1, the same 100 Monte Carlo simulations have been used. The differences of the mean squared error of m_I for the same regression function are due to (pseudo) random fluctuations. The bandwidth h_n for the smoothing step of m_{SI} and m_{IS} which minimizes asymptotically the integrated mean squared error (IMSE) of the kernel estimator m_S are larger than 0.5 for m_1 and m_2 and equal to 0.22 for m_3.

Furthermore, we have carried out 5000 simulations to estimate the constants in Theorem 3 and Theorem 4. For the estimation of c_1, c_2 and c_3 in

TABLE 1
Monte Carlo estimates of the mean squared error at $x_0 = \frac{1}{2}$ for m_{SI}, m_{IS}, m_S and m_I

	h_n	MSE(m_{SI})	MSE(m_{IS})	$I_{SI/IS}$	MSE(m_S)	MSE(m_I)
m_1	0.357	0.0094	0.0074	0.32	0.0097	0.0145
m_1	0.45	0.0068	0.0063	0.45	0.0068	0.0122
m_2	0.181	0.0147	0.0106	0.32	0.0161	0.0247
m_2	0.363	0.0088	0.0079	0.48	0.0088	0.0202
m_2	0.45	0.0101	0.0099	0.56	0.0101	0.0262
m_3	0.078	0.0332	0.0304	0.46	0.0347	0.0564
m_3	0.156	0.0215	0.0204	0.49	0.0215	0.0522
m_3	0.45	0.0431	0.0454	0.86	0.0431	0.0558
m_4	0.45	0.0072	0.0055	0.38	0.0073	0.0078

Theorem 3, we have estimated the expectation and the covariance matrix of $(m_{SI}(x_0), m_{IS}(x_0))$, where $m(x) = x^2$ $(0 \le x \le 1)$ is estimated at $x_0 = 0.5$ by $n = 60$ observations. We have used the quartic kernel $K_Q(x) = \frac{15}{16}(1 - x^2)^2$ with bandwidth $h_n = 0.4$ and the Gaussian kernel $K_G(x) = \phi(x)$ with bandwidth $h_n = 0.15$. The variance of the observations has been chosen as $\sigma^2 = (0.2)^2$ or equal to $(0.6)^2$. The results are summarized in Table 2. The last line of Table 2 can be compared with $\int K^2\, dt [\int t^2 K\, dt \int (K')^2\, dt]^{-1}$ for different kernels K. One gets 2.0 (for the triangle kernel K_T), 2.3 (for the quartic kernel K_Q) and 2.0 (for the Gaussian kernel K_G). This suggests to use m_{SI} and not m_{IS} for these Kernels (see Remark 1). But note that $m_{\int t}$ does not behave better in the case of undersmoothing (see Table 1). For the estimation of c_4 and c_5 in Theorem 4, we have estimated the expectation and the covariance matrix of $(m_{SI}(x_0), m_{IS}(x_0))$ for the regression function $m(x) = x$ $(0 \le x \le 1)$ at $x_0 = 0.5$. We have used the rectangle kernel $K_R(x) = \frac{1}{2}1(|x| \le 1)$ with bandwidth $h_n = 0.4$ for $n = 60$ observations. The variance of the observations has been chosen as $\sigma^2 = (0.2)^2$ or $= (0.6)^2$. The results are summarized in Table 3. These simulations suggest that $c_4 < 2c_5$ and that $m_{IS}(0)$ has always a smaller asymptotic mean squared error than $m_{SI}(0)$ if a discontinuous kernel is used (see Remark 2).

TABLE 2
Monte Carlo estimates of the constants c_1, c_2 and c_3 in Theorem 3

	$\sigma = 0.2$, $K = K_Q$	$\sigma = 0.2$, $K = K_G$	$\sigma = 0.6$, $K = K_Q$	$\sigma = 0.6$, $K = K_G$
c_1	0.11	0.11	0.09	0.08
c_2	0.27	0.31	0.19	0.15
c_3	0.50	0.48	0.42	0.34
$2c_2 - c_1$	0.43	0.51	0.29	0.22
$(2c_2 - c_1)/c_3$	0.86	1.07	0.68	0.62

TABLE 3
Monte Carlo estimates of the constants c_4 and c_5 in Theorem 4

	$\sigma = 0.2$ $K = K_R$	$\sigma = 0.4$ $K = K_R$
c_4	0.14	0.11
c_5	0.21	0.16
$2c_5 - c_4$	0.28	0.21

6. Proof of the theorems.

PROOF OF THEOREM 1. Choose $\varepsilon > 0$ small enough. We will show

$$\inf_{-\varepsilon \le x \le \varepsilon} m'_S(x) > 0 \text{ with probability tending to } 1. \tag{6.1}$$

$$\sup_{x \le -\varepsilon} m_S(x) \le \inf_{x \ge \varepsilon} m_S(x) \text{ with probability tending to } 1. \tag{6.2}$$

(6.1) and (6.2) would imply the statement of Theorem 1. But (6.1) and (6.2) follow from

$$\sup_{-1 \le x \le 1} \left| m_S^{(k)}(x) - m^{(k)}(x) \right| = o_P(1) \quad \text{for } k = 0, 1. \tag{6.3}$$

(6.3) follows from Lemma 5.2 in Müller and Stadtmüller (1987) for two times continuously differentiable kernel K [see also Chapter 11 in Müller (1988)]. For kernels K fulfilling (1.5) and (1.6) statement (6.3) can be proved along the lines of the proof of this lemma. □

PROOF OF THEOREM 2. Theorem 2 follows directly from Theorem 3. □

PROOF OF THEOREM 3. For the proof we will use strong approximations of the partial sum process $j \to \sum_{i \le j} Y_i$. By Komlós, Major and Tusnády (1975), there exists a sequence of two-sided Brownian motions W_n, starting at $W_n(0) = 0$, constructed on the same probability space as the ε_i's and a constant C with

$$\limsup_n \frac{1}{\log n} \sup_{1 \le k \le n} \left(\sqrt{n}\, W_n(k/n) - \sum_{i=1}^{k} \varepsilon \right) \le C, \tag{6.4}$$

$$\limsup_n \frac{1}{\log n} \sup_{1 \le k \le n} \left(\sqrt{n}\, W_n(-k/n) - \sum_{i=0}^{k-1} \varepsilon \right) \le C. \tag{6.5}$$

Put

$$S_n(s) = \int_0^s m(s)\, ds + \frac{1}{\sqrt{n}} W_n(s) \quad \text{for } -1 \le s \le 1, \tag{6.6}$$

where $\int_0^s \cdots$ is defined as $-\int_s^0 \cdots$ for $s < 0$. Then

$$(6.7) \qquad \sup_{-n \le j, k \le n} \left| \frac{1}{n} \sum_{i=j+1}^{k} Y_i - \left(S_n\left(\frac{k}{n}\right) - S_n\left(\frac{j}{n}\right) \right) \right| = O_P\left(\frac{\log n}{n} \right),$$

$$(6.8) \qquad \sup_{-n \le j, k \le n} \left| \frac{1}{n} \sum_{i=j+1}^{k} Y_i^* - \left(S_n^c\left(\frac{k}{n}\right) - S_n^c\left(\frac{j}{n}\right) \right) \right| = O_P\left(\frac{\log n}{n} \right),$$

where f^c denotes the greatest convex minorant of a function f. By partial summation one can easily deduce from (6.7) and (6.8)

$$(6.9) \qquad m_S(0) = n^{1/5} \int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s) S_n(ds) + O_P(n^{-4/5} \log n),$$

$$(6.10) \qquad m_{IS}(0) = n^{1/5} \int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s) S_n^c(ds) + O_P(n^{-4/5} \log n).$$

Now let

$$(6.11) \qquad T_n(s) = S_n(s) - \frac{1}{\sqrt{n}} g_n(s) X_n,$$

where

$$(6.12) \qquad X_n = n^{1/10} \int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s) W_n(ds)$$

and

$$(6.13) \qquad g_n(s) = \left(n^{1/10} \int_{-n^{-1/5}}^{s} K(n^{1/5}t)\, dt \right) \left(n^{1/5} \int_{-n^{-1/5}}^{n^{-1/5}} K^2(n^{1/5}t)\, dt \right)^{-1}.$$

Then $T_n(\cdot)$ and X_n are independent because of

$$E(X_n(T_n(s) - ET_n(s))) = \frac{1}{\sqrt{n}} E(X_n(W_n(s) - g_n(s) X_n)) = 0.$$

Define

$$(6.14) \qquad U_{1,n} = n^{-2/5} X_n,$$

$$(6.15) \qquad U_{2,n}' = n^{1/5} \int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s)(T_n^c(ds) - m(s)\, ds),$$

$$(6.16) \qquad E_n = \frac{1}{\sqrt{n}} n^{1/5} \int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s)(g_n'(s) - g_n^*(s))\, ds,$$

where $g_n^*(s) = (V(s) - U(s)) \int_{U(s)}^{V(s)} g_n(s)\, ds$ and $U(s) < V(s)$ are the two wedges of S_n^c (i.e., points, where the slope of S_n^c changes) nearest to s (g_n^* is the derivative of the linear interpolation of g_n restricted to the wedges of S_n^c). Theorem 1 yields

$$(6.17) \qquad m_{SI}(0) = m(0) + \beta_n + U_{1,n} + o_P(n^{-2/3}).$$

Furthermore, on the set A_n = {the wedges of S_n^c and T_n^c coincide}, one gets [see (6.10), (6.11) and (6.15)]

$$m_{IS}(0) = m(0) + \beta_n + U'_{2,n} - E_n X_n$$
$$(6.18) \qquad + \frac{1}{\sqrt{n}} n^{1/5} \int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s) g'_n(s)\, ds\, X_n + o_P(n^{-2/3})$$
$$= m(0) + \beta_n + \left(1 - n^{2/5}E_n\right)U_{1,n} + U'_{2,n} + o_P(n^{-2/3}),$$

where in the last equation (6.19) has been used

$$(6.19) \qquad \frac{1}{\sqrt{n}} n^{1/5} \int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}s) g'_n(s)\, ds = \frac{1}{\sqrt{n}} n^{1/10}.$$

For the theorem it remains to prove

$$(6.20) \qquad P(A_n) \to 1,$$

$$(6.21) \qquad n^{2/5}E_n - \varepsilon_n = o_P(n^{-4/15}),$$

(6.22) $n^{8/15}U'_{2,n}$ is asymptotically Gaussian with expectation $n^{8/15}\delta_n$ and with variance $c_1\sigma^{10/3}(m'(0))^{-4/3}\int(K'(t))^2\, dt$.

Proof of (6.20). $g_{n,1}$ and $g_{n,2}$ are defined by

$$(6.23) \qquad g_{n,i}(s) = n^{1/10}\int_{-n^{-1/5}}^{s} K_i(n^{1/5}t)\, dt \Big/ \int_{-1}^{+1} K^2(t)\, dt \quad \text{for } |s| \le 2n^{-1/5},$$

where $K_1(s) = \int_{-1}^{s}(K'(t))^+\, dt$ and $K_2(s) = \int_{-1}^{s}(K'(t))^-\, dt$. $g_{n,1}$ and $g_{n,2}$ are convex functions and $g_n = g_{n,1} - g_{n,2}$. Now let for α small enough $a_n = n^{\alpha-1/2}$ and $S_{n,+} = S_n + a_n g_{n,1} + a_n g_{n,2}$ and $S_{n,-} = S_n - a_n g_{n,1} - a_n g_{n,2}$. Then with probability tending to 1, $X_n < n^\alpha$ and therefore $(S_{n,+})^c$ has more wedges than S_n^c and T_n^c and furthermore $(S_{n,-})^c$ has less wedges than S_n^c and T_n^c. So it remains to show that every wedge of $(S_{n,+})^c$ is a wedge of $(S_{n,-})^c$ (with probability tending to 1). This follows if $S_{n,-}$ restricted to the set of wedges of $(S_{n,+})^c$ is convex (with probability tending to 1). We will show this by proving:

(i) All changes of slopes of $(S_{n,+})^c$ in $(-n^{-1/5}, n^{-1/5})$ are greater than $b_n = n^{-7/15-\alpha}$ (with probability tending to 1).

(ii) All changes of slopes of 2 $a_n(g_{n,1} + g_{n,2}) = S_{n,+} - S_{n,-}$ [restricted to the set of wedges of $(S_{n,+})^c$] are smaller than b_n (with probability tending to 1).

For the proof of (i), note that

(6.24) With probability tending to 1, $(S_{n,+})^c$ has less than $n^{\alpha/2}n^{1/3-1/5}$ wedges in $[-2n^{-1/5}, 2n^{-1/5}]$.

Furthermore, from the first equation on page 103 of Groeneboom (1989), one can show by integrating a bound of the right side:

(6.25) P(the next change of slope after s is smaller than $bn^{-1/3}$) $\le$ const. b.

Therefore

$$P(\text{all changes of slopes of } (S_{n,+})^c \text{ in } (-n^{-1/5}, n^{-1/5}) \text{ are greater than } b_n = n^{-7/15-\alpha})$$
$$> (1 - \text{const.}\, n^{-7/15-\alpha+1/3})^{n^{\alpha/2+1/3-1/5}} \to 1.$$

For the proof of (ii), note that

$$g''_{n,i} = O(n^{3/10}). \tag{6.26}$$

Furthermore, from the proof of Lemma 6.2 in Prakasa Rao (1969) it can be seen that $P\,((W(s) + s^2)^c$ has no wedge in $[0, c]) \le \text{const.} \exp(-c^3/32)$. This implies

(6.27) The maximal distance between two neighboring wedges of $(S_{n,+})^c$ is of order $O_P(n^{-1/3}\log(n))$.

(6.26) and (6.27) imply that the maximal change of the slope of $2a_n(g_{n,1} + g_{n,2})$ [restricted to the set of wedges of $(S_{n,+})^c$] is of order

$$O_P(a_n n^{3/10-1/3}\log(n)) = O_P(b_n)$$

for α small enough.

Proof of (6.21) and (6.22). First we prove the asymptotic normality of

$$U'_{2,n} = m_{IS}(0) - m_{SI}(0) + E_n X_n + o_P(n^{-2/3})$$
$$= U''_{2,n} + o_P(n^{-2/3}) = \tilde{U}_{2,n} + O_P(n^{-8/15}),$$

where

$$\tilde{U}_{2,n} = n^{1/5}\int_{-n^{-1/5}}^{n^{-1/5}} K(n^{1/5}t)(S_n^c - S_n)(dt)$$
$$= -n^{2/5}\int_{-n^{-1/5}}^{n^{-1/5}} K'(n^{1/5}t)(S_n^c(t) - S_n(t))\,dt$$

and

$$U''_{2,n} = \tilde{U}_{2,n} + \varepsilon_n U_{1,n}.$$

Define

$$t_i = -n^{-1/5} + in^{-4/15}(\log n)^2$$
$$\text{for } i = 0, \dots, I = \left[2n^{-1/5+4/15}(\log n)^{-2}\right] + 1.$$

Put

$$\tilde{Z}_i = n^{14/15}\int_{t_{i-1}}^{t_i} K'(n^{1/5}t)(S_n(t) - S_n^c(t))\,dt.$$

Then one gets

$$n^{8/15}U'_{2,n} = \sum_{i=1}^{I} \tilde{Z}_i + o_p(1).$$

Now define

$$Z_i = n^{14/15} \int_{t_{i-1}}^{t_i} K'(n^{1/5}t)\big(S_n(t) - \tilde{S}_n^c(t)\big)\, dt,$$

where for $s \in (t_{i-1}, t_i)$, the function $\tilde{S}_n^c(s)$ denotes the greatest convex minorant of $S_n(s)$ restricted to (t_{i-1}, t_i). Now note that

$$0 \le \tilde{S}_n^c(s) - S_n^c(s) \le S_n(s) - S_n^c(s)$$

and

$$\sup_{-n^{-1/5} \le s \le n^{-1/5}} S_n(s) - S_n^c(s) = o_p((\log n) n^{-2/3})$$

[see Kiefer and Wolfowitz (1976)].

This and (6.27) imply that

$$\sup_{1 \le i \le I} \big|Z_i - \tilde{Z}_i\big| = o_P\big((\log n)^2 n^{-1/15}\big)$$

and that

$$n^{8/15} U'_{2,n} = \sum_{i=1}^{I} Z_i + o_P(1).$$

For the proof of asymptotic normality of $n^{8/15}U'_{2,n}$, first note that the Z_i's are independent. Furthermore, bounds of the variance and the fourth central moment can be calculated by an application of the following two inequalities [suppose $m'(0) = 1$]

$$E\big(M(s)^k \big| \tilde{U}(s) = u, \tilde{V}(s) = v\big) \le \text{const.}(k, \varepsilon) \exp\Big(\big(\tfrac{1}{2} + \varepsilon\big)\big[n^{1/3}(v - u)\big]^3\Big), \tag{6.28}$$

$$\sup_s E \exp\Big(\big(\tfrac{2}{3} - \varepsilon\big)\big[n^{1/3}(\tilde{V}(s) - \tilde{U}(s))\big]^3\Big) \le \text{const.}(\varepsilon), \tag{6.29}$$

where $\varepsilon > 0$ and where $\tilde{U}(s) < \tilde{V}(s)$ are the two wedges of $\tilde{S}_n^c$ nearest to s and where $M(s) = \sup[n^{2/3}(S_n(t) - \tilde{S}_n^c(t))\colon \tilde{U}(s) \le t \le \tilde{V}(s)]$.

(6.28) follows from the following upper bound of the conditional density of $M(s)$

$$P\big(M(s) \in dt \big| \tilde{U}(s) = u, \tilde{V}(s) = u + n^{-1/3}\delta\big) \le \text{const.} \sum_{k \ge 1} \exp\big(\tfrac{1}{2}\delta^3/k^2\big)\sigma_k^{-1}\varphi\big((t - 2\delta\sigma_k^2)/\sigma_k\big)(t/\sigma_k)^3\, dt,$$

where $\sigma_k^2 = \frac{1}{4}\delta/k^2$.

This bound can be derived from formula (11.10) in Billingsley (1968) by an application of the Cameron–Martin–Girsanov formula [see also the proof of Lemma 2.1 in Groeneboom (1989)]. The proof of (6.29) can be based on an upper bound of the right-hand side of the first equation on page 103 of Groeneboom (1989) [see also (6.25)].

Now one can show by application of (6.28) and (6.29) that for $s_1 < s_2 < s_3 < s_4 \in (t_i, t_{i+1})$

$$E\Delta_n(s_1)\Delta_n(s_2) \le \text{const.}\, n^{-4/3} \exp\left(-\text{const.}\, n^{1/3}(s_2 - s_1)\right),$$

$$E\Delta_n(s_1)\Delta_n(s_2)\Delta_n(s_3)\Delta_n(s_4) \le \text{const.}\, n^{-8/3} \exp\left(-\text{const.}\, n^{1/3}(s_2 - s_1 + s_4 - s_3)\right),$$

where $\Delta_n(s) = S_n(s) - \tilde{S}_n^c(s) - E(S_n(s) - \tilde{S}_n^c(s))$. The first inequality can be proved by introducing two new processes $S_{n,1}$ and $S_{n,2}$ with the same distribution as S_n and such that $S_{n,1}(s) = S_n(s)$ for $s \le s_m = (s_1 + s_2)/2$ and $S_{n,2}(s) = S_n(s)$ for $s \ge s_m$ and such that the processes $S_n, (S_{n,1}(s) - S_n(s_m)$: $s > s_m)$ and $(S_{n,2}(s) - S_n(s_m)$: $s < s_m)$ are independent. Now define $\Delta_n^*(s_1)$ and $\Delta_n^*(s_2)$ as $\Delta_n(s_1)$ and $\Delta_n(s_2)$, but with using $S_{n,1}$ (or $S_{n,2}$, resp.) instead of S_n. Then the first inequality follows by estimating $E\Delta_n(s_1)\Delta_n(s_2) - \Delta_n^*(s_1)\Delta_n^*(s_2) = E\Delta_n(s_1)\Delta_n(s_2)$. The second inequality follows similarly. These two inequalities give bounds for the moments of Z_i which show the asymptotic normality of $U'_{2,n}$. The mean and variance of $U''_{2,n}$ can be calculated using that

$$EU''_{2,n} = n^{-2/5} \int_{-n^{-1/5}}^{n^{-1/5}} K'(n^{1/5}t) E(S_n(t) - S_n^c(t))\, dt + o(n^{-2/3}),$$

$$\begin{aligned} \operatorname{var} U''_{2,n} &= E\left(U''_{2,n}\right)^2 + O(n^{-4/3}), \\ &= n^{4/5} \int_{-n^{-1/5}}^{n^{-1/5}} \int_{-n^{-1/5}}^{n^{-1/5}} K'(n^{1/5}s) K'(n^{1/5}t) \\ &\qquad \times E((S_n^c - S_n)(s)(S_n^c - S_n)(t))\, ds\, dt + O(n^{-4/3}). \end{aligned}$$

The proof of (6.21) is straightforward [see (6.29)]. □

Proof of Theorem 4. First we prove (3.6). Note that $m_S(x)$ is a weighted average of a kernel estimate [with a kernel fulfilling (1.5) and (1.6)] and of kernel estimates with rectangle kernels (with different bandwidths). Therefore, because of Theorem 1, we can assume without loss of generality that K is the rectangle kernel $K(x) = \frac{1}{2}1(|x| \le 1)$. Choose $0 < \delta < \frac{1}{15}$ and put $m_S^*(x) := m_S(x^*)$, where x^* is the element of $\gamma_n \mathbf{Z}$ lying next to x for $\gamma_n = n^{-8/15-\delta}$. First note

$$\sup_{-1 \le x \le 1} |Em_S^*(x) - Em_S(x)| = o(n^{-8/15})$$

and

$$\sup_{k-j \le n^{7/15-\delta}} n^{-4/5} \sum_{i=j}^{k} \varepsilon_i = O_P\left(n^{-17/30-\delta/2}\sqrt{\log(n)}\right) = o_P(n^{-8/15}).$$

This implies

$$\sup_{-1 \le x \le 1} |m_S^*(x) - m_S(x)| = o_P(n^{-8/15}). \tag{6.30}$$

We will show

$$(6.31)\quad \sup_{x\le 0} m_S^*(x) \le m_S^*(0) \le \inf_{x\ge 0} m_S^*(x) \quad \text{with probability tending to 1.}$$

This implies

$$(6.32)\qquad P(m_S^*(0) = m_{SI}^*(0)) \to 1 \quad \text{for } n\to\infty,$$

where $m_{SI}^*(x) = \inf_{v\ge x}\sup_{u\le x}(v-u)^{-1}\int_u^v m_S^*(t)\,dt$. Using (6.30), one gets

$$(6.33)\qquad m_{SI}^*(0) = m_{SI}(0) + o_P(n^{-8/15}).$$

This shows (3.6). It remains to show (6.31). First note that for ε small enough and n large enough,

$$n^{1/5}\inf_{|x|>n^{-1/5}}|Em_S^*(x) - Em_S^*(0)| > \varepsilon$$

and

$$\inf_{\gamma_n<|x|<n^{-1/5}}\frac{|Em_S^*(x) - Em_S^*(0)|}{|x|} > \varepsilon.$$

Therefore it suffices to show

$$(6.34)\quad n^{1/5}\sup_{|x|>n^{-1/5}}|m_S^*(x) - m_S^*(0) - (Em_S^*(x) - Em_S^*(0))| = o_P(1)$$

and

$$(6.35)\quad \sup_{\gamma_n<|x|<n^{-1/5}}\frac{|m_S^*(x) - m_S^*(0) - (Em_S^*(x) - Em_S^*(0))|}{|x|} = o_P(1).$$

Proof of (6.35). Choose x with $\gamma_n < |x| < n^{-1/5}$. Then there exist a set I (with at most $2xn$ elements) and $s_i \in \{-1,1\}$ (for $i\in I$) with

$$(6.36)\quad \begin{aligned} &P\left(\frac{m_S^*(x) - m_S^*(0) - (Em_S^*(x) - Em_S^*(0))}{x} > C\right)\\ &= P\left(n^{-4/5}\sum_{i\in I}\frac{s_i\varepsilon_i}{x} > C\right)\\ &\le \exp(-tC)\left\{E\exp\left(\frac{t\varepsilon_i}{xn^{4/5}}\right)\right\}^{\#\{i:\, s_i=1\}}\\ &\quad\times\left\{E\exp\left(-\frac{t\varepsilon_i}{xn^{4/5}}\right)\right\}^{\#\{i:\, s_i=-1\}}, \end{aligned}$$

where $t = \rho x n^{3/5}$ (with a ρ chosen small enough). Now note that for a constant C_0, because of (1.2), $E\exp(u\varepsilon_i) \le 1 + C_0u^2$ for $|u|$ small enough and that $t/(xn^{4/5}) \to 0$. Therefore the last term of (6.36) can be bounded by

$$\le \exp\left(-tC + \#IC_0\left\{\frac{t}{xn^{4/5}}\right\}^2\right)$$
$$\le \exp\left(-xn^{3/5}\{C\rho - C_0 2\rho^2\}\right)$$
$$\le \exp\left(-n^{1/15-\delta}\{C\rho - C_0 2\rho^2\}\right).$$

With the same arguments one can show

$$(6.37)\qquad P\left(\left|\frac{m_S^*(x) - m_S^*(0) - (Em_S^*(x) - Em_S^*(0))}{x}\right| > C\right)$$
$$\le 2\exp\left(-n^{1/15-\delta}\{C\rho - C_0 2\rho^2\}\right).$$

(6.37) implies (6.35) because $\gamma_n \mathbf{Z} \cap \{x: \gamma_n < |x| < n^{-1/5}\}$ contains at most $2n^{3/5}$ elements. (6.34) can be shown with similar arguments.

(3.7), (3.8) and (3.9) can be proved with similar arguments as in the proof of Theorem 3. □

Acknowledgments. This paper was motivated by a discussion with Stefan Luckhaus and Wolfgang Härdle. We would like to thank M. C. Jones for pointing out some errors in the discussion of the results and a referee for very helpful comments and for bringing the paper Kiefer and Wolfowitz (1976) to our attention.

REFERENCES

Barlow, R. E., Bartholomew, D. J., Bremmer, J. M. and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

Barlow, R. E. and van Zwet, W. R. (1969). Asymptotic properties of isotonic estimators for the generalized failure rate function. Part II: Asymptotic distributions. Operations Research Center Report ORC 69-10, Univ. California.

Barlow, R. E. and van Zwet, W. R. (1970). Asymptotic properties of isotonic estimators for the generalized failure rate function. Part I: Strong consistency. In *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.) 159–173. Cambridge Univ. Press.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Cheng, K. F. and Lin, P. E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. Verw. Gebiete* **57** 223–233.

Friedman, J. and Tibshirani, R. (1984). The monotone smoothing of scatter plots. *Technometrics* **26** 243–350.

Groeneboom, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) 539–555. Wadsworth, Belmont, Calif.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and airy functions. *Probab. Theory Related Fields* **81** 79–109.

Hildenbrand, K. and Hildenbrand, W. (1985). On the mean income effect: A data analysis of the U.K. family expenditure survey. In *Contributions to Mathematical Economics* (W. Hildenbrand and A. Mas-Colell, eds.) 247–268. North-Holland, Amsterdam.

Kiefer, J. and Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete* **34** 73–85.

KOMLÓS, J., MAJOR P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.

LEURGANS, S. (1982). Asymptotic distributions of slope-of-greatest-convex-minorant estimators. *Ann. Statist.* **10** 287–296.

MUKERJEE, H. (1988). Monotone nonparametric regression. *Ann. Statist.* **16** 741–750.

MÜLLER, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data. Lecture Notes in Statist.* **46**. Springer, New York.

MÜLLER, H.-G. and STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15** 610–625.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448.

WRIGHT, F. T. (1982). Monotone regression estimates for grouped observations. *Ann. Statist.* **10** 278–286.

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT HEIDELBERG
IM NEUENHEIMER FELD 294
6900 HEIDELBERG
GERMANY