The Annals of Statistics
1990, Vol. 18, No. 1, 63–89

SEMIPARAMETRIC COMPARISON OF REGRESSION CURVES

BY W. HÄRDLE[1] AND J. S. MARRON[1,2]

Universität Bonn and Universität Bonn and University of North Carolina at Chapel Hill

The comparison of nonparametric regression curves is considered. It is assumed that there are parametric (possibly nonlinear) transformations of the axes which map one curve into the other. Estimation and testing of the parameters in the transformations are studied. The rate of convergence is $n^{-1/2}$ although the nonparametric components of the model typically have a rate slower than that. A statistic is provided for testing the validity of a given completely parametric model.

1. Introduction. An important case of regression analysis is the comparison of regression curves from related samples. Even when there is no reasonable parametric model for each regression curve a way of quantifying differences across individual curves is often desirable. A well-known example is the study of child growth curves, where individual curves certainly seem to require nonparametric estimation techniques [Gasser, Müller, Köhler, Molinari and Prader (1984)] but may have a simple relationship between them. Another example appears in Figures 1(a) and 1(b), which show acceleration data from a study on automobile side impacts [Kallieris, Mattern and Härdle (1986)].

The curves give the impression that they are noisy versions of similar regression curves, where the main difference is that the time axis is shifted and there is a vertical rescaling. A parametric model that could be deduced from a physical or biomechanical theory is not available here; see Eppinger, Marcus and Morgan (1984), so a nonparametric smoothing technique seems to be a reasonable way to estimate the acceleration curves for inference regarding this data set. The problem of comparison of the two curves could be modeled parametrically because, to a large extent, the difference between them seems to be quantified by two parameters, horizontal shift and vertical scale. Hence, a comparison of nonparametric regression curves in a parametric framework is desirable for studying data sets of this type.

The main objective of this paper is the analysis of general semiparametric models where nonparametric curves are related in a parametric way. The case that is treated in detail is where there are two curves which are the same up to a transformation of the horizontal axis and a transformation of the vertical axis, and these transformations are indexed by some parameters. The techniques of this paper are adaptable to other semiparametric models such as multiplicative

Received January 1987; revised March 1989.
[1]Research supported by Deutsche Forschungsgemeinschaft, SFB 303.
[2]Research partially supported by NSF Grants DMS-8400602 and DMS-8701201.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62G99.
Key words and phrases. Semiparametric regression, nonparametric smoothing, parametric comparison, kernel estimators.

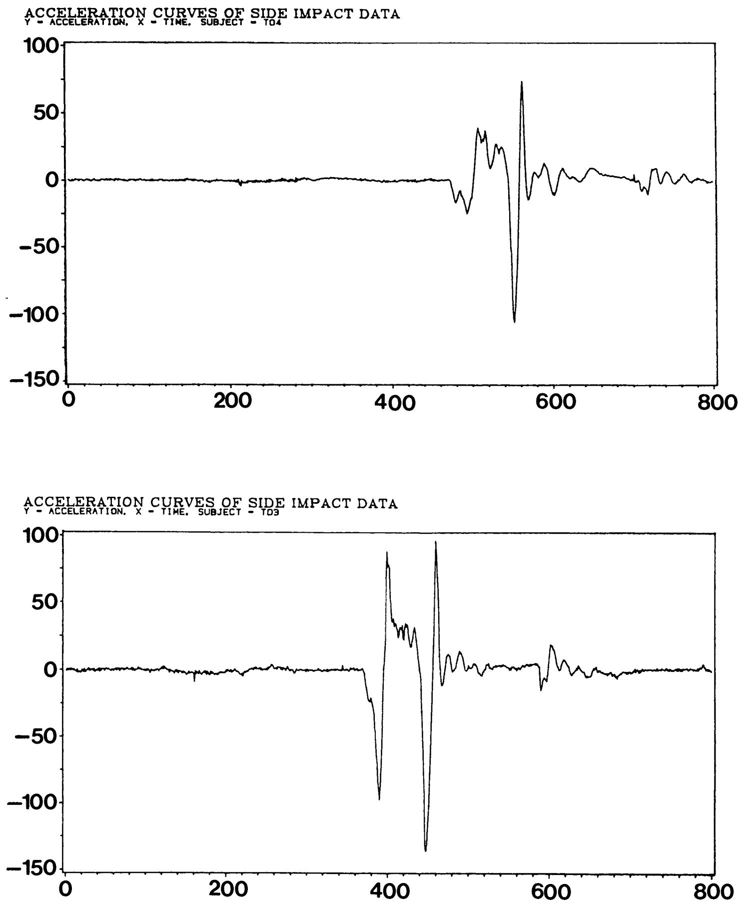

FIG. 1. *Two impact acceleration curves from Kallieris, Mattern and Härdle* (1986).

or additive combination of a nonparametric regression curve with a parametric "modulation" function. An additional benefit of the theory developed in this paper is that, with no extra work, a statistic is provided for testing the validity of a given completely parametric model. This test quantifies the idea of checking a parametric model by comparing the parametric fit to a nonparametric regression curve.

Section 2 contains a mathematical formulation of these ideas, together with a proposal for estimating the parameters. This parameter estimate is seen to be consistent under very mild conditions in Section 3. Asymptotic normality, with the rate of convergence typical to parametric problems, is established under somewhat stronger conditions in Section 4. Section 5 gives test statistics, together with their asymptotic null distributions, for testing whether some parameters can be eliminated from the model and also for testing whether a given semiparametric model is in fact appropriate.

2. Parametric comparison of nonparametric regression curves. The observations $(x_1, Y_1), \ldots, (x_n, Y_n)$, of the first curve are assumed to come from the nonparametric regression model,

$$Y_i = m_1(x_i) + \varepsilon_i, \qquad i = 1, \ldots, n.$$

The observation errors ε_i are assumed to be independent, mean 0, with common variance σ^2. The design points x_i are taken to be equally spaced on the unit interval $x_i = i/n$. Suppose the data from the second curve are $(x_1', Y_1'), \ldots, (x_n', Y_n')$, from the nonparametric regression model,

$$Y_i' = m_2(x_i') + \varepsilon_i',$$

where the ε_i' have common variance σ'^2, are independent of the ε_i and otherwise have the same stochastic structure as the ε_i, and where $x_i' = i/n$. While x_i' is the same as x_i, these are distinguished for the sake of clarity later in the paper.

The parametric nature of the curve comparison problem is modeled by

$$m_2(x') = S_{\theta_0}^{-1} m_1\left(T_{\theta_0}^{-1} x'\right), \tag{2.1}$$

where T_θ and S_θ are invertible transformations (e.g., shifts and scalings of the axes) indexed by the parameter $\theta \in \Theta \subseteq \mathbb{R}^d$, and where θ_0 is the true value of the parameter. Such a model for linear transformations S_θ and T_θ has been called "shape invariant" by Lawton, Sylvestre and Maggio (1972). A good estimate of θ_0 will be provided by a value of θ for which the curve $m_1(x)$ is closely approximated by

$$M(x, \theta) = S_\theta m_2(T_\theta x).$$

The effectiveness of each value of θ is assessed by the loss function,

$$L(\theta) = \int \left[m_1(x) - M(x, \theta)\right]^2 w(x)\, dx,$$

where w is a nonnegative weight function. Note that $M(x, \theta_0) = m_1(x)$, so θ_0 minimizes $L(\theta)$. The unknown regression functions m_1 and m_2 are estimated by kernel smoothers,

$$\hat{m}_1(x) = n^{-1} \sum_{i=1}^{n} K_h(x - x_i) Y_i,$$

$$\hat{m}_2(x) = n^{-1} \sum_{i=1}^{n} K_{h'}(x' - x_i') Y_i',$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$, for a kernel function K which integrates to 1. See Priestley and Chao (1972) and Collomb (1981, 1985) for properties of this estimator. Define the estimate $\hat{\theta}$ of θ_0, to be an argument which minimizes

$$\hat{L}(\theta) = \int \left[\hat{m}_1(x) - \hat{M}(x, \theta)\right]^2 w(x)\, dx,$$

where $\hat{M}(x, \theta) = S_\theta \hat{m}_2(T_\theta x)$. Since $\hat{L}(\theta)$ is a continuous and nonnegative function, there are no difficulties concerning the existence or measurability of $\hat{\theta}$. The weight function $w(x)$ is used to eliminate boundary effects and to restrict attention to a region where both $\hat{m}_1$ and $\hat{M}(x, \theta)$ provide reasonable estimates. This is illustrated by the following example.

Figure 2 is concerned with the specific setting

$$m_1(x) = (x - 0.4)^2,$$
$$m_2(x') = (x' - 0.5)^2 - 0.2.$$

This fits in the above framework by defining:

$$S_\theta(x) = x + \theta^{(2)},$$
$$T_\theta(x) = x + \theta^{(1)},$$

and letting

$$\theta_0 = \left(\theta_0^{(1)}, \theta_0^{(2)}\right) = (0.1, 0.2).$$

Figure 2(a) shows two sets of 100 simulated observations, where the (x_i, Y_i) are represented by squares, where the (x_i', Y_i') are represented by stars and where the errors are Gaussian with mean 0 and variance 0.0004. As a simple method of nullifying boundary effects we consider estimating $m_1(x)$ on the subinterval $x \in [\eta, 1 - \eta]$ (the choice of η is discussed below) and $m_2(x')$ on the subinterval $x' \in [\eta, 1 - \eta]$. For more complicated but also more efficient means of handling boundary effects see Gasser, Müller and Mammitzsch (1985) and Rice (1984a). To keep the focus on the main points under discussion here we do not incorporate this type of improvement. This second restriction corresponds to, for each θ, estimating

$$M(x, \theta) = S_\theta m_2(T_\theta x) = \left(x + \theta^{(1)} - 0.5\right)^2 - 0.2 + \theta^{(2)} \tag{2.2}$$

on the subinterval $x \in [\eta - \theta^{(1)}, (1 - \eta) - \theta^{(1)}]$. Hence, for $\theta^{(1)} > 0$, w should be 0 outside the interval $[\eta - \theta^{(1)}, (1 - \eta) - \theta^{(1)}]$. We do not take w to be the indicator of this interval because the minimizer of $\hat{L}(\theta)$ will then have some bias towards larger values of $\theta^{(1)}$ and we suspect that the minimum will be harder to compute. Figures 2(b) and 2(c) contain the same data as Figure 2(a), except that the (x_i', Y_i') have been replaced by $(x_i' + 0.106, Y_i' + 0.196)$ and $(x_i' + 0.2, Y_i' + 0.2)$, respectively. Observe that from these figures it is quite apparent that $\theta_0^{(1)} \in [0, 0.2]$. Hence, we can restrict Θ to only include θ with $\theta^{(1)} \in [0, 0.2]$, and take $w(x)$ to be the indicator of $[\eta, 0.8 - \eta]$.

In the general case, we assume that there is an interval $[a, b] \subseteq [0, 1]$ where boundary effects are eliminated and then define

$$\begin{aligned} w(x) &= \prod_{\theta \in \Theta} 1_{[a, b]}(T_\theta x) \\ &= 1_{\bigcap_{\theta \in \Theta} T_\theta^{-1}([a, b])}(x). \end{aligned}$$

Note that for $\hat{\theta}$ to be a reasonable estimate, this requires that Θ be rather small. This assumption does not seem too restrictive because these methods will only be applied after the experimenter has looked at some preliminary curve estimates. Such a previewing procedure does not cause any additional effort if an interactive graphical data analysis program is available. Hence, it is assumed that the

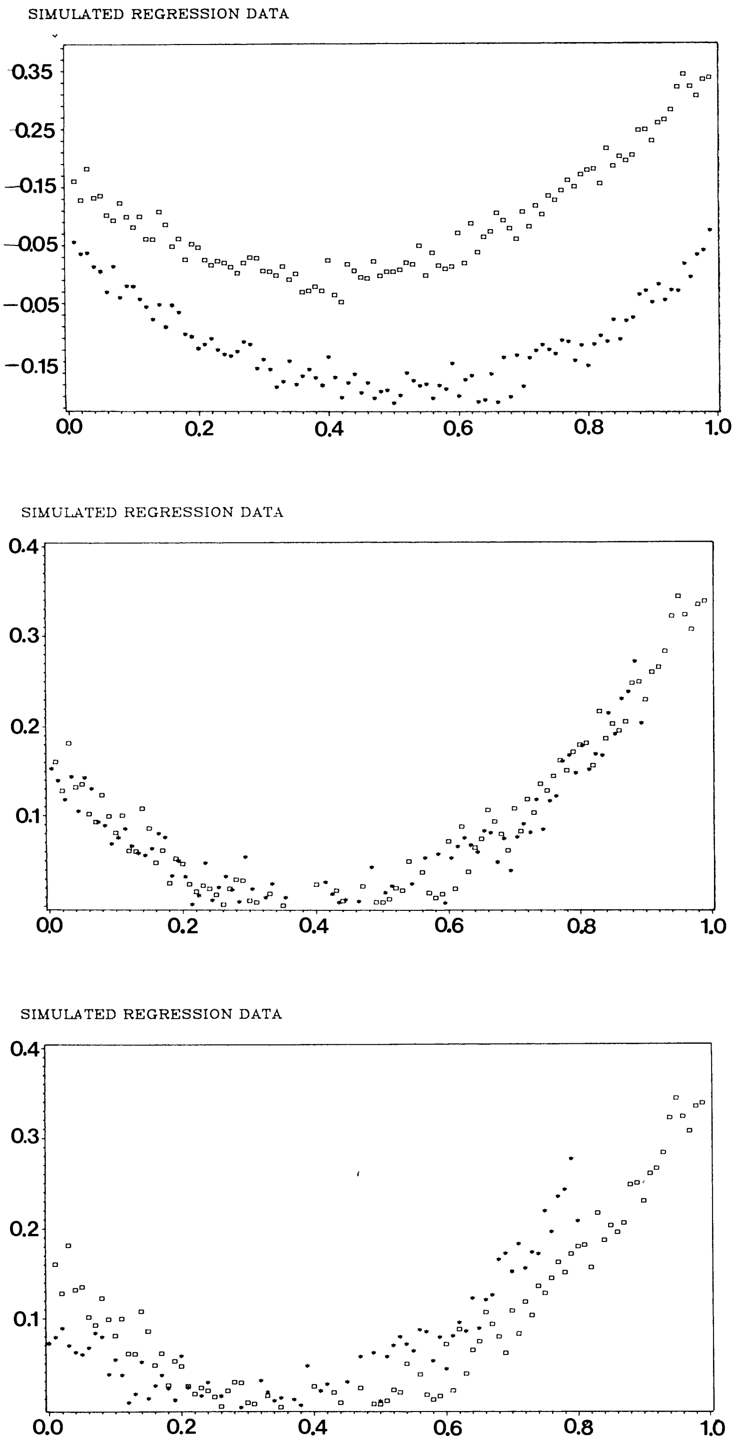

FIG. 2. *Simulated data.*

TABLE 1
Parameter estimates for the simulated regression data for different values of h and η. Reported are values of χ_1^2-statistics for $H_0^{(1)}$, $H_0^{(2)}$ and χ_2^2 for $H_0^{(3)}$

supp(w)	h	$\hat{\theta}^{(1)}$	$\hat{\theta}^{(2)}$	$H_0^{(1)}$ (With estimated covariance)	$H_0^{(2)}$ (With estimated covariance)	$H_0^{(3)}$ (With estimated covariance)	$H_0^{(1)}$ (With exact covariance)	$H_0^{(2)}$ (With exact covariance)	$H_0^{(3)}$ (With exact covariance)
(0.0, 0.8)	0.02	0.106	0.200	87.8	1328	1458	59	1000	1059
(0.1, 0.7)	0.02	0.114	0.198	36.2	974	1040	25	735	760
(0.2, 0.6)	0.02	0.114	0.198	10.2	668	675	7.5	490	497
(0.0, 0.8)	0.04	0.106	0.198	91.6	1356	1491	59	980	1040
(0.1, 0.7)	0.04	0.106	0.196	32.7	999	1040	25	720	745
(0.2, 0.6)	0.04	0.106	0.194	10.7	669	676	7.5	470	477
(0.1, 0.3)	0.10	0.106	0.194	70.3	946	1059	59	940	1000
(0.1, 0.7)	0.10	0.116	0.196	28.0	723	779	30	720	750
(0.2, 0.6)	0.10	0.120	0.192	9.8	474	497	9.5	460	470

experimenter has a good approximate idea of the value of θ_0. It is merely an assumption to the effect that the design of the experiment is appropriate for the type of inference to be done.

The first four columns of Table 1 show how the estimates $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$, which have been found by a gridsearch (Figure 4 gives an intuitive feeling for the type of grid that we used), depend on the support restriction η and the bandwidth h for the above simulated data set appearing in Figure 2. The remaining columns will be discussed in Section 5.

Observe that the parameter estimates are not very sensitive to the support restrictions as expressed by the cutoff parameter η. Also varying the bandwidth does not affect the estimates too much. Under the above assumptions, for the final estimation of the underlying curve $m_1(x)$, the two data sets can be pooled by using

$$\tfrac{1}{2}\hat{m}_1(x) + \tfrac{1}{2}\hat{M}(x, \hat{\theta}).$$

This will only be an effective estimate of $m_1(x)$ if the assumption of the curves being the same is correct, but even the assumption is not quite correct, this still provides a reasonable estimate of the "average curve." More than two regression curves can be analyzed by using preliminary estimates to choose one curve that seems to lie in the center and calling that m_1, then comparing the other curves to that. However, it should be kept in mind that this is only an example, so it is not possible to make general conclusions. Furthermore, it has been deliberately chosen so that the method may be expected to work well.

Alternative ways of formulating the semiparametric comparison model are to assume that $M(x, \theta) = m_1(x) + S_\theta(x)$, or $M(x, \theta) = m_1(x)S_\theta(x)$, where $S_\theta(x)$ is a "modulation" function which is assumed to be known up to the parameter $\theta \in \Theta$. The general ideas of this paper apply in this case; however, details of the proofs will be different. It appears that these forms should be substantially easier to analyze. There are some recent papers on a model of the first form; see Engle, Granger, Rice and Weiss (1986), Green (1985), Rice (1986) and Speckman (1986). A semiparametric model of the form (2.1) but with random parameters has also been investigated by Kneip and Gasser (1988). For an access to related work in the time series context, see Cameron and Hannan (1979), Cameron (1983) and Cameron and Thompson (1985). See He (1988) for another method of parameter estimation in a model similar to ours (but more specialized) in the interesting case of random design points.

3. Consistency of the parameter estimate. In this section, precise conditions are given for the convergence of $\hat{\theta}$ to θ_0 as the sample size grows. The most important assumption is that the loss function $L(\theta)$ be locally convex near θ_0 in the sense that: Given $\varepsilon > 0$, there is a $D(\varepsilon) > 0$, so that $|\theta - \theta_0| > \varepsilon$ implies

$$L(\theta) - L(\theta_0) > D(\varepsilon). \tag{3.1}$$

This condition ensures the identifiability of the parameters. An example of when this condition fails to hold is when $m(x)$ is constant and T_θ is a horizontal shift. The remaining assumptions ensure consistency of the regression estimates. To allow for use of an automatically chosen (and hence random) bandwidth, see Rice (1984b) and Härdle and Marron (1985a), and also to show that consistency of $\hat{\theta}$ is not dependent on the particular choice of the bandwidths, we establish consistency uniformly over h, h' in the interval

$$B_n = [n^{-1+\delta}, n^{-\delta}],$$

where $\delta > 0$ is arbitrary. The kernel function K, in addition to integrating to 1, is assumed to be compactly supported and Hölder continuous, i.e., there exist constants $\alpha, \beta > 0$ such that $|K(u) - K(v)| \leq \alpha|u - v|^\beta$. The regression functions $m_1(x)$ and $m_2(x)$ are assumed to be Hölder continuous. The transformations S_θ and T_θ are assumed to be smooth in the sense that:

$$\sup_{\theta \in \Theta} \sup_{x \in [0,1]} |S'_\theta(x)| < \infty, \tag{3.2}$$

$$\sup_{\theta \in \Theta} \sup_{x \in [0,1]} |(T_\theta^{-1})'(x)| < \infty. \tag{3.3}$$

Note that (3.2) and (3.3) are not any restriction at all if S_θ and T_θ are linear. The following theorem is proved in Section 6.

THEOREM 1. *Under the above assumptions $\hat{\theta}$ is consistent for θ_0, uniformly over h, $h' \in B_n$, in the sense that*

$$\sup_{h, h' \in B_n} |\hat{\theta} - \theta_0| \to 0 \quad a.s.$$

4. Asymptotic normality. In this section the rate of the convergence in Section 3 is studied by giving conditions for asymptotic normality of $n^{1/2}(\hat{\theta} - \theta_0)$. Since the nonparametric estimators $\hat{m}_1$ and $\hat{m}_2$ have a rate of convergence slower than $n^{1/2}$, some care must be taken to obtain the rate of convergence $n^{1/2}$

for the $\hat{\theta} - \theta_0$ limiting distribution. To this end we assume that

$$T_\theta x = \theta^{(1)} + \theta^{(2)} x, \tag{4.1}$$

$$S_\theta \text{ only depends on } \theta^{(3)}, \dots, \theta^{(d)}, \tag{4.2}$$

and that $\hat{m}_1$ and $\hat{m}_2$ employ the same amount of smoothing in the sense that

$$h' = \theta^{(2)} h. \tag{4.3}$$

Assumption (4.3) seems quite restrictive at first glance; however, an inspection of the proofs reveals that it is in fact necessary for $n^{1/2}$ convergence of the parameter estimates. A simple way of implementing this in practice is to choose the bandwidth for only $\hat{m}_1$, say by cross validation, and then using a preliminary estimate of $\theta^{(2)}$ to get an improved $\hat{\theta}^{(2)}$ and iterating. More efficient methods would pool the information from the two curves, as discussed in Marron and Rudemo (1988) and Marron and Schmitz (1988). This is complicated in the present situation because the smoothing parameter selection is confounded with the estimation of θ, but a promising possibility to be investigated is to choose both h and $\hat{\theta}$ to be the joint minimizers of the sum of $\hat{L}(\theta)$ and the cross-validation score functions for the two curves.

As in Section 3, a critical assumption concerns the identifiability of θ_0. Assume that

$$H(\theta_0) \text{ is positive definite,} \tag{4.4}$$

where $H(\theta)$ is the $d \times d$ matrix whose l, l'th entry is

$$\int M_l(x, \theta) M_{l'}(x, \theta) w(x)\, dx,$$

using the notation $M_l(x, \theta) = (\partial / \partial \theta^{(l)}) M(x, \theta)$. Under the assumptions of this section, it can be shown that (4.4) implies (3.1). To gain some insight into this, consider the case $S_\theta(x) = \theta^{(3)} + \theta^{(4)} x$, where

$$M_1(x, \theta) = \theta^{(4)} m_2'(\theta^{(1)} + \theta^{(2)} x),$$

$$M_2(x, \theta) = \theta^{(4)} m_2'(\theta^{(1)} + \theta^{(2)} x) x,$$

$$M_3(x, \theta) = 1,$$

$$M_4(x, \theta) = m_2(\theta^{(1)} + \theta^{(2)} x).$$

Observe that (4.4) is then essentially requiring that the functions 1, $m(x)$, $m'(x)$ and $xm'(x)$ be linearly independent in $L^2(w)$. To facilitate Taylor expansion arguments, it is assumed that $S_\theta(x)$ is smooth in the sense that the following functions are uniformly continuous and bounded uniformly over $x \in \operatorname{supp}(w)$

over $\theta \in \Theta$, and over $l, l' = 1, \dots, d$:

$$
\begin{aligned}
S'_\theta(x) &= \frac{\partial}{\partial x} S_\theta(x), \\
S_{\theta,l}(x) &= \frac{\partial}{\partial \theta^{(l)}} S_\theta(x), \\
S_{\theta,l,l'}(x) &= \frac{\partial}{\partial \theta^{(l')}} S_{\theta,l}(x), \qquad (4.5) \\
S'_{\theta,l}(x) &= \frac{\partial}{\partial x} S_{\theta,l}(x), \\
S'_{\theta,l,l'}(x) &= \frac{\partial}{\partial x} S_{\theta,l,l'}(x).
\end{aligned}
$$

This assumption is trivial if S_θ is linear. Also to facilitate expansions, assume

(4.6) $m_1''(x)$ exists and is uniformly continuous.

A consequence of (4.5), (4.6) and the linearity of T_θ is that

$$M_{l,l'}(x,\theta) = \frac{\partial}{\partial \theta^{(l')}} M_l(x,\theta) \qquad (4.7)$$

is uniformly continuous and bounded uniformly over $x \in \operatorname{supp}(w)$, $\theta \in \Theta$, and $l, l' = 1, \dots, d$. Also assume that K is a compactly supported probability density with Hölder continuous second derivative, and that $E\varepsilon_i^k < \infty$, for $k = 1, 2, \dots$, uniformly over $i = 1, \dots, n$. The final requirement is that the bandwidth h is taken to be an automatically selected bandwidth $\hat{h}$, as discussed in Rice (1984b), Härdle and Marron (1985a, b) and Härdle, Hall and Marron (1988) have shown that under the above assumptions

$$\hat{h} = h_0 + O_p(n^{-3/10}),$$

where $h_0 = c_0 n^{-1/5}$, for a constant c_0. Hence, if B_n^* is defined by

$$B_n^* = \left[h_0 - n^{-3/10+\alpha}, h_0 + n^{-3/10+\alpha}\right],$$

for some $\alpha \in (0, 1/10)$, then $P[\hat{h} \in B_n^*] \to 1$. Note that $\hat{h}$ is chosen only from the data $Y_1, \dots, Y_n$. This allows assumption (4.3) to be satisfied in a simple fashion. See the discussion there for other possibilities.

THEOREM 2. *Under the above assumptions*

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to_{\mathscr{L}} N\left(0, H^{-1}(\theta_0)\Sigma H^{-1}(\theta_0)\right),$$

where the l, l'th entry of Σ is

$$4\int\left[\sigma^2 + \sigma'^2\left(S'_{\theta_0}\left(m_2(T_{\theta_0}x)\right)\right)^2\right] M_l(x,\theta_0) M_{l'}(x,\theta_0) w(x)\,dx.$$

The proof of Theorem 2 is in Section 7. To add insight into this theorem, consider the special case of the example given in Section 2. Note that $T_\theta x =$

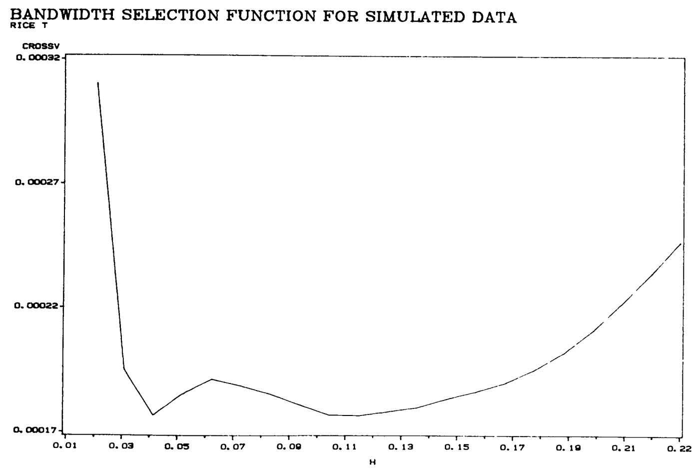

FIG. 3. *Bandwidth selection function based on the Epanečnikov kernel and weight function on* (0.1, 0.7).

$\theta^{(1)} + x$ and $S_\theta x = \theta^{(2)} + x$, so an obvious modification of the notation of this section will be made. In particular, from (2.2),

$$M_1(x,\theta) = 2(x + \theta^{(1)} - 0.5),$$
$$M_2(x,\theta) = 1,$$
$$S'_{\theta_0}(x) = 1,$$

and so

$$H(\theta_0) = \begin{pmatrix} \frac{8}{3}(0.4-\eta)^3 & 0 \\ 0 & 2(0.4-\eta) \end{pmatrix}.$$

Thus, $\sqrt{n}(\hat{\theta} - \theta)$ has asymptotic covariance matrix

$$\frac{\sigma^2}{(0.4-\eta)^3}\begin{pmatrix} 3 & 0 \\ 0 & 4(0.4-\eta)^2 \end{pmatrix}.$$

The bandwidth selection function computed for (x_i, Y_i), with w supported on $[0.1, 0.7]$ and the Epanečnikov kernel had a global minimum at $h = 0.04$ (Figure 3) but had a pronounced local minimum. In this simulated example we used

$$T(h) = n^{-1}\sum_{i=1}^{n}[Y_i - \hat{m}_1(x_i)]^2 w(x_i)/[1 - 2n^{-1}h^{-1}K(0)]$$

as a bandwidth selector. See Härdle, Hall and Marron (1988) for a more complete discussion of the issues of bandwidth selection. The negative loss function for this bandwidth is shown in Figure 4. Note that Figure 4 shows that the loss function is more sensitive to changes in $\theta^{(2)}$ than to changes in $\theta^{(1)}$. This is reflected intuitively by thinking about vertical and horizontal shifts in Figure 2(a).

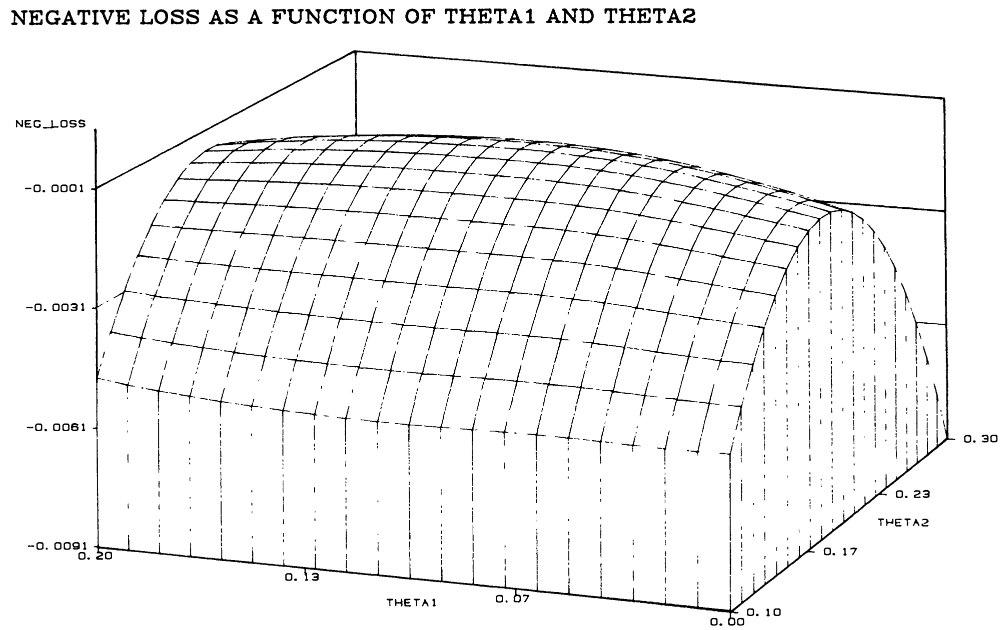

FIG. 4. *Negative loss as a function of* $\theta^{(1)}$ *and* $\theta^{(2)}$. $m(X) = (X - 0.4)^2$. *Errors* $N(0, 0.0004)$. *Weight function on* $(0.1, 0.7)$.

While the vertical shift is obvious, we find it much more difficult to justify a horizontal shift just by "eye inspection." Statistically, this can be quantified by $\mathrm{var}(\hat{\theta}^{(1)}) \approx 0.0444$, $\mathrm{var}(\hat{\theta}^{(2)}) \approx 0.0053$ (where these are the entries in the asymptotic covariance matrix given in Theorem 2). $L(\theta)$ is minimized at $\hat{\theta} = (\hat{\theta}^{(1)}, \hat{\theta}^{(2)}) = (0.106, 0.196)$ which is the shift used in the construction of Figure 2(b). An intuitive understanding of $\hat{\theta}$ can also be gained from Figure 5, which shows $\hat{m}_1(x)$ (solid line) and $\hat{M}(x, \hat{\theta})$ (dashed line). Note that either a horizontal or a vertical shift in the relative position of these curves will increase the integrated (over $[0.1, 0.7]$) squared difference between these.

A look at Figure 1 indicates that the shift-scale model, $T_\theta = \theta^{(1)} + x$, $S_\theta = \theta^{(4)}x$ (using notation consistent with this section) should be appropriate for the automobile side impact data. After transforming the X-values into the unit interval, the bandwidth $\hat{h} = 0.012$ was obtained by cross validation over the interval $[0.1, 0.7]$ for the data set shown in Figure 1(b), which we took to be $\{(x_i, Y_i)\}_{i=1}^n$ with $n = 800$. The negative loss function $\hat{L}(\theta)$ is shown in Figure 6, which for its form is called the "Sidney Opera.". As expected from a comparison of Figures 1(a) and 1(b), the choice of $\theta^{(1)}$ is more critical than that of $\theta^{(4)}$. The "side ridges" in the negative loss correspond to values of $\theta^{(1)}$, where there is a matching of "first peaks" to "second peaks." The loss function was minimized at $\hat{\theta} = (\hat{\theta}^{(1)}, \hat{\theta}^{(4)}) = (0.13, 1.45)$. Figure 7 shows how $\hat{m}_1(x)$ (solid curve) compares with $\hat{M}(x, \hat{\theta})$ (dashed curve).

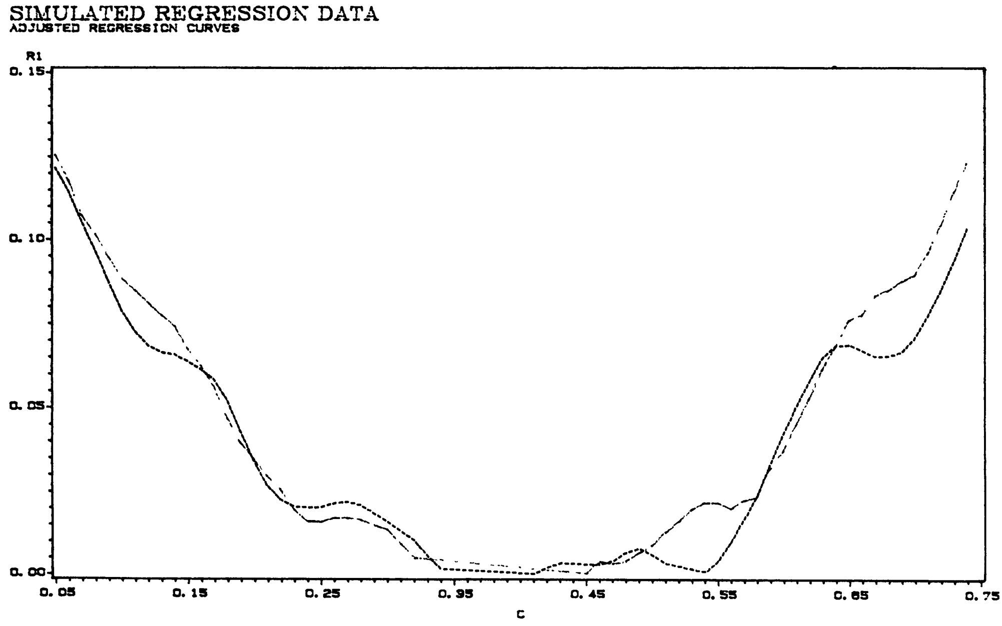

FIG. 5. *Adjusted regression curves for the simulated data.* $m(X) = (X - 0.4)^2$. *Errors* $N(0, 0.0004)$. $\theta^{(1)} = 0.106$; $\theta^{(2)} = 0.196$.

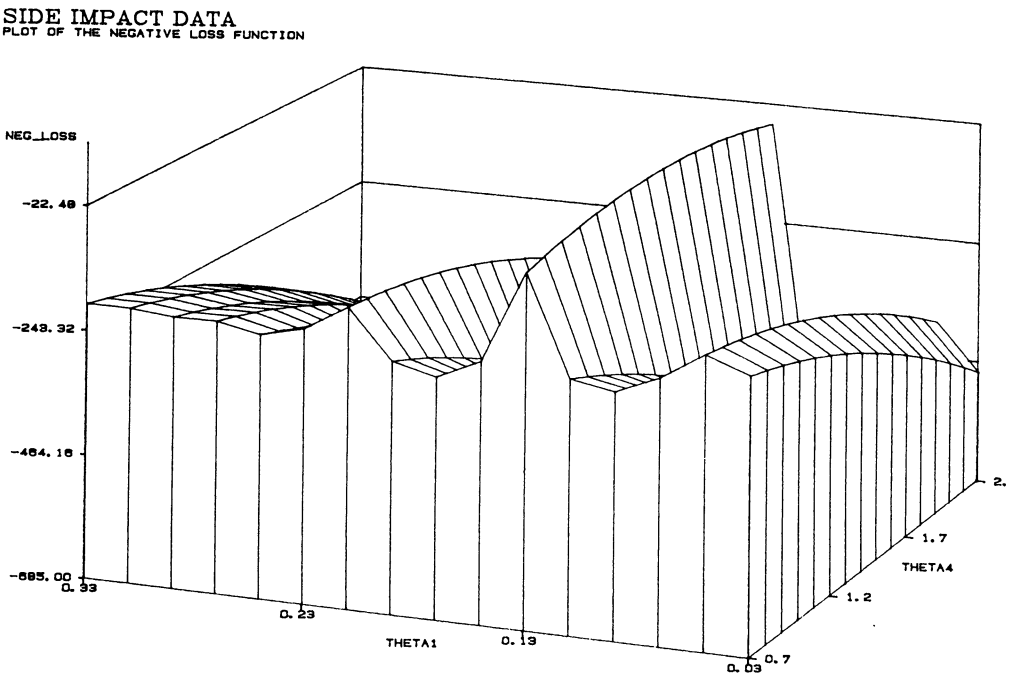

FIG. 6. *"Sidney Opera" negative loss function for the side impact data. Weight function on* (0.1, 0.7).

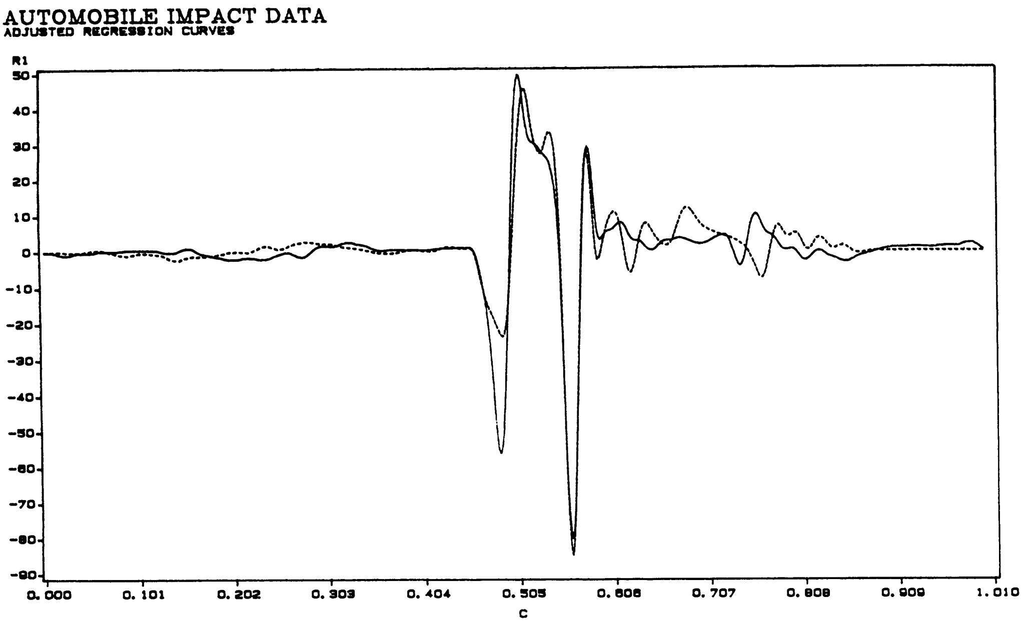

FIG. 7. *Adjusted regression curves for the automobile side impact data. Weight function on* (0.1, 0.8). $\theta^{(1)} = 0.13$; $\theta^{(4)} = 1.45$.

5. Hypothesis testing. There are two important hypotheses to test in this semiparametric model. First, can the parametric part of the model be reduced? (For example: Can a horizontal shift and scale be reasonably replaced by just a shift? Is an apparent vertical shift really significant?) Second, is the semiparametric model of this paper appropriate for a particular data set? [That is: Is $m_2(x)$ really a simple transformation of $m_1(x)$?] To formulate the first hypothesis, suppose there is a $\theta^* \in \Theta$ so that

$$m_1(x) = M(x, \theta^*).$$

For example, components of θ^* corresponding to the types of shifts discussed earlier are 0 and to the scaling are 1. A general way to formulate the hypothesis is

$$H_0\colon A(\theta_0 - \theta^*) = 0,$$

for an $r \times d$ matrix A of rank r. A reasonable basis for a hypothesis test is $A(\hat{\theta} - \theta^*)$, which has an asymptotic $N(0, \Sigma^*)$ distribution under H_0, where $\Sigma^* = AH(\theta_0)^{-1}\Sigma H(\theta_0)^{-1}A^T$. This suggests rejecting H_0 when $(\hat{\theta} - \theta^*)^T A^T \widehat{\Sigma}^{*-1} A(\hat{\theta} - \theta^*)$ is larger than the 95th percentile of the χ_r^2 distribution, where $\widehat{\Sigma}^*$ is a consistent estimate of Σ^*. These ideas can be illustrated in the simulated data example of Section 2 which is depicted in Figure 2(a).

Consider the hypotheses:

$$H_0^{(1)}: \theta^{(1)} = 0,$$

$$H_0^{(2)}: \theta^{(2)} = 0,$$

$$H_0^{(3)}: \theta^{(1)} = \theta^{(2)} = 0.$$

Table 1 shows the observed test statistics for the simulated regression data. To give some feel for the effect of estimating Σ^* by $\hat{\Sigma}^*$, two types of test statistics are shown, the first type using the exact value Σ^* and the second type using the estimate $\hat{\Sigma}^*$. The effect of various choices for w and h is also illustrated in Table 1.

Note that the observed values of the test statistics are relatively independent of the bandwidth, but depend quite heavily on the choice of η. It is not surprising that the values decrease with increasing η because larger η means less of the data are used, so the tests will lose power. This effect is most notable for $H_0^{(1)}$, which is easily understood by covering observations near the boundary in Figure 2(a). Note that in all cases the results here are highly significant. This is to be expected, except in the case $H_0^{(1)}$ with $\eta = 0.2$. The fact that the test proposed in this section is quite powerful in this example may be seen by covering the intervals $[0.0, 0.2]$ and $[0.6, 1.0]$ in Figure 2(a). We recommend taking η as small as possible. A means of doing this is to first start with some preliminary guess at η, use this to get a preliminary $\hat{h}$, then take a final η which just barely eliminates the boundary effects for this $\hat{h}$.

For the automobile impact data, using the notation of Section 4, we tested

$$H_0^{(1)}: \theta^{(1)} = 0,$$

$$H_0^{(2)}: \theta^{(4)} = 1,$$

$$H_0^{(3)}: \theta^{(1)} = \theta^{(4)} = 1.$$

The observed test statistics are presented in Table 2, which has a layout similar to Table 1. In contrast to Table 1, this time the observed values of the test statistics are relatively independent of η [not surprising since essentially all of the useful information is contained in the center of Figures 1(a) and (b)], but vary a lot with h. The reason that the tests lose power for larger values of h is that when $\hat{m}_1$ and $\hat{m}_2$ are oversmoothed, the distinctive peaks in Figures 1(a) and 1(b) are greatly diminished. As expected from the pictures, $H_0^{(2)}$ suffers the most from this effect, although we can still reject this hypothesis at the level 0.05, when $h = 0.012$ (selected by cross validation).

For testing the second hypothesis, that the model is correct, an obvious statistic is $\hat{L}(\hat{\theta})$, which should be small if the model is correct, but large otherwise. The asymptotic distribution of $\hat{L}(\hat{\theta})$ is summarized in

Theorem 3. *Under the assumptions of Section* 4,

$$nh_0^{1/2}\left(\hat{L}(\hat{\theta}) - n^{-1}h_0^{-1}C_\mu\right) \to_{\mathscr{L}} N\left(0, C_\sigma^2\right),$$

TABLE 2
χ^2-*statistics for different bandwidths and support restrictions for impact data*

supp(w)	h	$H_0^{(1)}$	$H_0^{(2)}$	$H_0^{(3)}$
		(With estimated covariance)		
(0.0, 0.8)	0.005	248	13.4	2180
(0.1, 0.7)	0.005	246	13.3	2150
(0.2, 0.6)	0.005	245	13.0	2130
(0.0, 0.8)	0.012	80.9	4.36	232
(0.1, 0.7)	0.012	80.3	4.32	229
(0.2, 0.6)	0.012	80.0	4.25	227
(0.0, 0.8)	0.040	41.9	2.26	62.3
(0.1, 0.7)	0.040	41.6	2.24	61.4
(0.2, 0.6)	0.040	41.4	2.20	60.9

where

$$C_\mu = \left(\int K^2\right)\left(\int\left[\frac{\sigma^2}{\theta_0^{(2)}} + \sigma'^2\left(S'_{\theta_0}(m_2(x))\right)^2\right]w\left(T_{\theta_0}^{-1}(x)\right)dx\right),$$

$$C_\sigma^2 = 2\theta_0^{(2)}\left(\int(K*K)^2\right)\left(\int\left[\frac{\sigma^2}{\theta_0^{(2)}} + \sigma'^2\left(S'_{\theta_0}(m_2(x))\right)^2\right]^2 w\left(T_{\theta_0}^{-1}(x)\right)dx\right).$$

The proof of Theorem 3 is in Section 8. It follows from Theorem 3 that a reasonable test, of the hypothesis that m_2 is indeed a parametric shift of m, will reject when

$$\hat{L}(\hat{\theta}) > (n\hat{h})^{-1}\hat{C}_\mu + n^{-1}\hat{h}^{-1/2}\hat{C}_\sigma z_{1-\alpha},$$

where z_{1-d} is the $(1-\alpha)$th quantile of the standard normal distribution, and where the estimates

$$\hat{C}_\mu = \left(\int K^2\right)\left(\int\left[\frac{\hat{\sigma}^2}{\hat{\theta}^{(2)}} + \hat{\sigma}'^2\left(S'_{\hat{\theta}}(\hat{m}_2(x))\right)^2\right]w\left(T_{\hat{\theta}}^{-1}(x)\right)dx\right),$$

$$\hat{C}_\sigma^2 = 2\hat{\theta}^{(2)}\left(\int(K*K)^2\right)\left(\int\left[\frac{\hat{\sigma}^2}{\hat{\theta}^{(2)}} + \hat{\sigma}'^2\left(S'_{\hat{\theta}}(\hat{m}_2(x))\right)\right]^2 w\left(T_{\hat{\theta}}^{-1}(x)\right)dx\right),$$

$$\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n\left(Y_i - \hat{m}_1(x_i)\right)^2, \qquad \hat{\sigma}'^2 = n^{-1}\sum_{i=1}^n\left(Y_i' - \hat{m}_2(x_i')\right)^2$$

have been used. The observed test statistics for the side impact data set are listed in Table 3 for a weight function concentrated on (0.1, 0.7). The shift-scale model that we proposed achieved a p-value of 0.02, whereas all the other studied submodels had p-values less than 0.001. Figure 7 provides an intuitive feeling for the power involved in this test. Note that while $\hat{\theta}$ clearly provides an informative choice of the parameters, it is also clear that the curves are certainly not the same. The fact that, at least in this example, the parameter estimation method

TABLE 3
The test statistic from Theorem 3 for correctness of the model.
$\text{supp}(w) = (0.1, 0.7)$

	$\hat{\theta}^{(1)}$	$\hat{\theta}^{(4)}$	Test statistic	p
Shift and scale model	0.13	1.45	2.01	0.0220
Shift model, only	0.03	1.00	30.02	< 0.0010
Scale model, only	0.00	0.10	21.18	< 0.0010
none	0.00	1.00	345.00	< 0.0010

of this paper provides good estimates of the amount of shift and scale, even when the underlying curves are not identical, seems to greatly enhance its potential applicability.

6. Proof of Theorem 1. To simplify notation, let $\sup_h$ mean $\sup_{h, h' \in B_n}$. Given $\varepsilon > 0$,

$$P\Big[\sup_h |\hat{\theta} - \theta_0| > \varepsilon\Big] \le P\Big[\sup_h \big(L(\hat{\theta}) - L(\theta_0)\big) > D(\varepsilon)\Big]$$
$$\le P\Big[\sup_h \big(L(\hat{\theta}) - \hat{L}(\hat{\theta}) + \hat{L}(\theta_0) - L(\theta_0)\big) > D(\varepsilon)\Big]$$
$$\le P\Big[\sup_h |L(\hat{\theta}) - \hat{L}(\hat{\theta})| > \frac{D(\varepsilon)}{2}\Big]$$
$$+ P\Big[\sup_h |\hat{L}(\theta_0) - L(\theta_0)| > \frac{D(\varepsilon)}{2}\Big].$$

Hence, Theorem 1 follows from: Given $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P\Big[\sup_\theta \sup_h |\hat{L}(\theta) - L(\theta)| > \varepsilon\Big] < \infty, \tag{6.1}$$

where $\sup_\theta$ means $\sup_{\theta \in \Theta}$.

To prove (6.1), note that by rearranging terms, by adding and subtracting $2\hat{m}_1(x)M(x, \theta)$ and by the triangle inequality,

$$|\hat{L}(\theta) - L(\theta)|$$
$$\le \int \Big[|(\hat{m}_1 - m_1)(\hat{m}_1 + m_1)| + 2|\hat{m}_1(M - \hat{M})| + 2|M(m_1 - \hat{m}_1)|$$
$$+ |(\hat{M} - M)(\hat{M} + M)| \Big] w\, dx.$$

Hence, by the Schwarz inequality, (6.1) follows from: Given $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P\Big[\sup_h \int (\hat{m}_1 - m_1)^2 w\, dx > \varepsilon\Big] < \infty, \tag{6.2}$$

$$\sum_{n=1}^{\infty} P\Big[\sup_\theta \sup_h \int (\hat{M} - M)^2 w\, dx > \varepsilon\Big] < \infty, \tag{6.3}$$

together with

$$\int m_1^2 w\,dx < \infty, \tag{6.4}$$

$$\sup_\theta \int M^2 w\,dx < \infty. \tag{6.5}$$

To prove (6.2), note that for $B_n' \subseteq B_n$,

$$P\left[\sup_h \int (\hat{m}_h - m_1)^2 w\,dx > \varepsilon\right]$$
$$\leq P\left[\sup_{h \in B_n'} \int (\hat{m}_h - m_1)^2 w\,dx > \frac{\varepsilon}{2}\right]$$
$$+ P\left[\sup_{h \in B_n} \inf_{h_1 \in B_n'} \left| \int (\hat{m}_h - m_1)^2 w\,dx - \int (\hat{m}_{h_1} - m_1)^2 w\,dx \right| > \frac{\varepsilon}{2}\right].$$

By Hölder continuity of m_1 and K, B_n' can be chosen so that the second term is 0, for n sufficiently large, and so that $\#(B_n') \leq n^{\xi}$, some $\xi > 0$. Hence, by Theorem 1 of Marron and Härdle (1986), (6.2) follows from

$$\sup_{h \in B_n'} E \int (\hat{m}_h - m_1)^2 w\,dx \to 0,$$

which is easily established by the methods of Rosenblatt (1971).

To prove (6.3), note that

$$\int [\hat{M}(x, \theta) - M(x, \theta)]^2 w(x)\,dx$$
$$= \int [S_\theta \hat{m}_2(T_\theta x) - S_\theta m_2(T_\theta x)]^2 w(x)\,dx$$
$$= \int_0^1 [S_\theta'(\xi)(\hat{m}_2(u) - m_2(u))]^2 w(T_\theta^{-1}(u))(T_\theta^{-1})'(u)\,du.$$

Hence, (6.3) follows from (3.2), (3.3) and the methods used to establish (6.2).

Note that (6.4) is a consequence of the Hölder continuity of $m_1(x)$. To prove (6.5), use Hölder continuity of $m_2(x)$ and an argument of the type used on (6.3). This completes the proof of Theorem 1. □

7. Proof of Theorem 2. Let $\nabla \hat{L}(\theta)$ denote the d-dimensional vector of partial derivatives, $\hat{L}_l(\theta) = (\partial/\partial\theta^{(l)})\hat{L}(\theta)$. Note that

$$0 = \nabla \hat{L}(\hat{\theta}) = \nabla \hat{L}(\theta_0) + \hat{H}(\hat{\xi}_n)(\hat{\theta} - \theta_0), \tag{7.1}$$

where $\hat{H}(\theta)$ is the Hessian matrix, whose components are $\hat{L}_{l,l'}(\theta) = (\partial/\partial\theta^{(l')})\hat{L}_l(\theta)$, and where $\hat{\xi}_n$ lies on a line segment connecting $\hat{\theta}$ and θ_0. Theorem 2 is a consequence of (7.1), together with the following lemmas.

LEMMA 2.1.

$$\sqrt{n}\, \nabla \hat{L}(\theta_0) \to_{\mathscr{L}} N(0, \Sigma).$$

LEMMA 2.2.

$$\hat{H}(\hat{\xi}_n) \to_p H(\theta_0).$$

To prove Lemma 2.1, note that the lth component of $\nabla\hat{L}(\theta_0)$ is

$$\int 2[\hat{m}_1(x) - \hat{M}(x,\theta_0)](-\hat{M}_l(x,\theta_0))w(x)\,dx.$$

For the rest of this section, let $\sup_h$ mean $\sup_{h \in B_n^*}$. Lemma 2.1 follows from Lemmas 2.1.1 through 2.1.4.

LEMMA 2.1.1. *For $l = 1, \dots, d$,*

$$\sup_h \left| \int [\hat{m}_1(x) - \hat{M}(x,\theta_0)](M_l(x,\theta_0) - \hat{M}_l(x,\theta_0))w(x)\,dx \right| = o_p(n^{-1/2}).$$

LEMMA 2.1.2. *For $l = 1, \dots, d$,*

$$\sup_h \left| \int [E\hat{m}_1(x) - E\hat{M}(x,\theta_0)] M_l(x,\theta_0)w(x)\,dx \right| = o(n^{-1/2}).$$

LEMMA 2.1.3. *For $l = 1, \dots, d$,*

$$\sup_h |Z_l(h) - Z_l(h_0)| = o_p(n^{-1/2}),$$

where

$$Z_l(h) = \int [\hat{m}_1(x) - E\hat{m}_1(x) + E\hat{M}(x,\theta_0) - \hat{M}(x,\theta_0)] M_l(x,\theta_0)w(x)\,dx.$$

LEMMA 2.1.4.

$$n^{1/2} 2Z(h_0) \to_{\mathscr{L}} N(0,\Sigma),$$

where $Z(h_0)$ is the vector whose components are the $Z_l(h_0)$.

To prove Lemma 2.1.1, note first that $m_1(x) = M(x,\theta_0)$. Hence, by the Schwarz inequality, it is enough to show

$$\sup_h \int [\hat{m}_1(x) - m_1(x)]^2 w(x)\,dx = o_p(n^{-7/10}), \tag{7.2}$$

$$\sup_h \int [\hat{M}(x,\theta_0) - M(x,\theta_0)]^2 w(x)\,dx = o_p(n^{-7/10}), \tag{7.3}$$

$$\sup_h \int [\hat{M}_l(x,\theta_0) - M_l(x,\theta_0)]^2 w(x)\,dx = o_p(n^{-3/10}). \tag{7.4}$$

The proofs of (7.2) and (7.3) use the same methods as were used on (6.2) and (6.3), together with the fact that, under the present stronger assumptions,

$$\sup_{h \in B_n'} E\int [\hat{m}_h - m_1]^2 w\,dx = O_p(n^{-4/5}).$$

To verify (7.4) in the case of $l \geq 3$, note that

$$\hat{M}_l(x,\theta_0) - M_l(x,\theta_0) = S_{\theta_0,l}\big(\hat{m}_2(T_{\theta_0}x)\big) - S_{\theta_0,l}\big(m_2(T_{\theta_0}x)\big) = S'_{\theta_0,l}(\xi_n)\big(\hat{m}_2(T_{\theta_0}x) - m_2(T_{\theta_0}x)\big).$$

Hence, the methods used on (7.3) together with (4.5) may be applied. To establish (7.4) when $l = 1$, write

$$\begin{aligned}\hat{M}_1(x,\theta_0) - M_1(x,\theta_0) &= S'_{\theta_0}\big(\hat{m}_2(T_{\theta_0}x)\big)\hat{m}'_2(T_{\theta_0}x) - S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)m'_2(T_{\theta_0}x)\\ &= \Big(S'_{\theta_0}\big(\hat{m}_2(T_{\theta_0}x)\big) - S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\Big)\hat{m}'_2(T_{\theta_0}x)\\ &\quad + S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\big(\hat{m}'_2(T_{\theta_0}x) - m'_2(T_{\theta_0}x)\big).\end{aligned}$$

Now use the Schwarz inequality and the above methods applied to estimation of m'_2 instead of m_2, together with assumption (4.5).

To finish the proof of (7.4), note that

$$\begin{aligned}M_2(x,\theta_0) &= S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)m'_2(T_{\theta_0}x)x,\\ \hat{M}_2(x,\theta_0) &= S'_{\theta_0}\big(\hat{m}_2(T_{\theta_0}x)\big)\\ &\quad\times \frac{\partial}{\partial\theta^{(2)}}\left[n^{-1}\sum_i \frac{1}{\theta^{(2)}h}K\left(\frac{\theta^{(2)}x + \theta_1^{(1)} - x'_i}{\theta^{(2)}h}\right)Y'_i\right]\Bigg|_{\theta=\theta_0}\\ &= S'_{\theta_0}\big(\hat{m}_2(T_{\theta_0}x)\big)n^{-1}\sum_i U(x,x'_i)Y'_i,\end{aligned}$$

where

$$U(x,x'_i) = \frac{-1}{\theta_0^{(2)^2}h}K\left(\frac{\theta_0^{(2)}x + \theta_0^{(1)} - x'_i}{\theta_0^{(2)}h}\right) - \frac{\theta_0^{(1)} - x'_i}{\theta_0^{(2)^3}h^2}K'\left(\frac{\theta_0^{(2)}x + \theta_0^{(1)} - x'_i}{\theta_0^{(2)}h}\right).$$

But, uniformly over $h \in B_n^*$ and over $x \in \operatorname{supp}(w)$,

$$\begin{aligned}n^{-1}\sum_i U(x,x'_i)m_2(x'_i) &= \int U(x,x')m_2(x')\,dx' + O(n^{-4/5})\\ &= \int\left[\frac{-1}{\theta_0^{(2)}h}K\left(\frac{x-u}{h}\right) + \frac{u}{\theta_0^{(2)}h^2}K'\left(\frac{x-u}{h}\right)\right]\\ &\quad\times m_2\big(\theta_0^{(1)} + \theta_0^{(2)}u\big)\,du + O(n^{-4/5})\\ &= -\frac{1}{h}\int\left(\frac{d}{du}\left[K\left(\frac{x-u}{h}\right)u\right]\right)\\ &\quad\times\frac{1}{\theta_0^{(2)}}m\big(\theta_0^{(1)} + \theta_0^{(2)}u\big)\,du + O(n^{-4/5})\\ &= \int\frac{1}{h}K\left(\frac{x-u}{h}\right)um'_2\big(\theta_0^{(1)} + \theta_0^{(2)}u\big)\,du + O(n^{-4/5})\\ &= xm'_2(T_{\theta_0}x) + O(n^{-1/5}).\end{aligned}\tag{7.5}$$

Thus,

$$\hat{M}_2(x,\theta_0) - M_2(x,\theta_0) = \mathrm{I} + \mathrm{II} + \mathrm{III},$$

where

$$\mathrm{I} = \left[S'_{\theta_0}\left(\hat{m}_2(T_{\theta_0}x)\right) - S'_{\theta_0}\left(m_2(T_{\theta_0}x)\right)\right] n^{-1} \sum_i U(x, x'_i) Y'_i,$$

$$\mathrm{II} = S'_{\theta_0}\left(m_2(T_{\theta_0}x)\right) n^{-1} \sum_i U(x, x'_i) \varepsilon'_i,$$

$$\mathrm{III} = S'_{\theta_0}\left(m_2(T_{\theta_0}x)\right) O_p(n^{-1/5}).$$

The $l = 2$ case of (7.4) now follows from the Schwarz inequality and the methods used on the other cases. This finishes the proof of Lemma 2.1.1.

To prove Lemma 2.1.2, note that uniformly over $h \in B_n^*$ and $x \in \operatorname{supp}(w)$,

$$\begin{aligned} E\hat{M}(x, \theta_0) &= E\Big[S_{\theta_0}\left(E\hat{m}_2(T_{\theta_0}x)\right) + \left(\hat{m}_2(T_{\theta_0}x) - E\hat{m}_2(T_{\theta_0}x)\right) \\ &\qquad \times S'_{\theta_0}\left(E\hat{m}_2(T_{\theta_0}x)\right) + \tfrac{1}{2}\left(\hat{m}_2(T_{\theta_0}x) - E\hat{m}_2(T_{\theta_0}x)\right)^2 S''_{\theta_0}(\xi_n)\Big] \\ &= S_{\theta_0}\left(E\hat{m}_2(T_{\theta_0}x)\right) + O(n^{-4/5}) \end{aligned} \tag{7.6}$$

and

$$\begin{aligned} E\hat{m}_2(T_{\theta_0}x) &= n^{-1} \sum_i K_{h'}(T_{\theta_0}x - x'_i) S_{\theta_0}^{-1} m_1\left(T_{\theta_0}^{-1} x'_i\right) \\ &= \int K_{h'}(T_{\theta_0}x - u') S_{\theta_0}^{-1} m_1\left(T_{\theta_0}^{-1}u'\right) du' + O(n^{-4/5}) \\ &= \int K(u) S_{\theta_0}^{-1} m_1(x - hu)\, du + O(n^{-4/5}) \\ &= \int K(u) \bigg[S_{\theta_0}^{-1}\left(\int K(z) m_1(x - hz)\, dz\right) \\ &\quad + \left(m_1(x - hu) - \int K(z) m_1(x - hz)\, dz\right) S'_{\theta_0}\left(\int K(z) m_1(x - hz)\, dz\right) \\ &\quad + \tfrac{1}{2}\left(m_1(x - hu) - \int K(z) m_1(x - hz)\, dz\right)^2 S''_{\theta_0}(\xi_n)\bigg]\, du + O(n^{-4/5}) \\ &= S_{\theta_0}^{-1}\left(\int K(z) m_1(x - hz)\, dz\right) + O(n^{-4/5}), \end{aligned} \tag{7.7}$$

from which it follows that $E\hat{M}(x, \theta_0) = \int K(z) m_1(x - hz)\, dz + O(n^{-4/5})$. Lemma 2.1.2 now follows from $E\hat{m}_1(x) = \int K(u) m_1(x - hu)\, du + O(n^{-4/5})$, and assumption (4.6).

To prove Lemma 2.1.3, note that

$$\int [\hat{m}_1(x) - E\hat{m}_1(x)] M_l(x, \theta_0) w(x)\, dx = n^{-1} \sum_i V_i(h) \varepsilon_i,$$

where

$$\begin{aligned} V_i(h) &= \int K_h(x - x_i) M_l(x, \theta_0) w(x)\, dx \\ &= \int K(u) M_l(x_i + hu) w(x_i + hu)\, du. \end{aligned}$$

Thus, by uniform continuity of M_l,

$$\sup_h \left| n^{-1} \sum_i (V_i(h) - V_i(h_0))\varepsilon_i \right| = o_p(n^{-1/2}).$$

The $\hat{M}$ part can be handled by similar methods together with the linearization technique of (7.6) and (7.7). This completes the proof of Lemma 2.1.3.

To prove Lemma 2.1.4, note that for $l = 1, \dots, d$,

$$Z_l(h_0) = n^{-1} \sum_i (A_{il}\varepsilon_i + B_{il}\varepsilon_i'),$$

where

$$\begin{aligned} A_{il} &= \int K_{h_0}(x - x_i) M_l(x, \theta_0) w(x)\, dx \\ &= M_l(x_i, \theta_0) w(x_i) + o_p(1), \\ B_{il} &= S_{\theta_0}'\big(E\hat{m}_2(T_{\theta_0} x_i)\big) M_l(x_i, \theta_0) w(x_i) + o_p(1). \end{aligned}$$

Using the Cramer–Wold device, a central limit theorem for $Z(h_0)$ can be established by showing asymptotic normality of each linear combination

$$n^{1/2} \sum_l c_l 2 Z_l(h_0) = 2n^{-1/2} \sum_i \left(\sum_l c_l (A_{il}\varepsilon_i + B_{il}\varepsilon_i') \right),$$

where $\Sigma_l c_l^2 > 0$. Since this is a sum of independent mean zero random variables with third moments, by Liapounov's version of the array-type central limit theorem [see Chung (1974), Theorem 7.1.2, for example], we need only check that the variance tends to a constant. But

$$\begin{aligned} &\operatorname{var}\left(2n^{-1/2} \sum_i \left(\left(\sum_l c_l A_{il} \right) \varepsilon_i + \left(\sum_l c_l B_{il} \right) \varepsilon_i' \right) \right) \\ &\quad = 4n^{-1} \sum_i \left(\left(\sum_l c_l A_{il} \right)^2 \sigma^2 + \left(\sum_l c_l A_{il} \right)^2 \sigma'^2 \right) \\ &\quad = 4 \int \Bigg[\left(\sum_l c_l M_l(x, \theta_0) \right)^2 \sigma^2 \\ &\qquad + \left(\sum_l c_l \big(S_{\theta_0}' E\hat{m}_2(T_{\theta_0} x) \big) M_l(x, \theta_0) \right)^2 \sigma'^2 \Bigg] w(x)\, dx + o(1), \end{aligned}$$

which is positive by assumption (4.4). Similarly, for $l, l' = 1, \dots, d$,

$$\begin{aligned} &\operatorname{cov}\big(n^{1/2} 2 Z_l(h_0), n^{1/2} 2 Z_{l'}(h_0) \big) \\ &\quad = 4 \int \Big[\sigma^2 + \sigma'^2 \big(S_{\theta_0}'\big(E\hat{m}_2(T_{\theta_0} x) \big) \big)^2 \Big] M_l(x, \theta_0) M_{l'}(x, \theta_0) w(x)\, dx + o(1). \end{aligned}$$

This completes the proof of Lemma 2.1.4 and hence also the proof of Lemma 2.1.

To prove Lemma 2.2, note that for $l, l' = 1, \dots, d$,

$$\hat{L}_{l,l'}(\hat{\xi}_n) = \int 2\big[\hat{M}_l(x, \hat{\xi}_n)\hat{M}_{l'}(x, \hat{\xi}_n) - (\hat{m}_1(x) - \hat{M}(x, \hat{\xi}_n))\hat{M}_{l,l'}(x, \hat{\xi}_n)\big] w(x)\,dx,$$

$$L_{l,l'}(\theta_0) = \int 2\big[M_l(x, \theta_0)M_l(x, \theta_0) - (m_1(x) - M(x, \theta_0))M_{l,l'}(x, \theta_0)\big] w(x)\,dx.$$

Thus, by appropriate adding and subtracting, by the Schwarz inequality, and by (4.7), it is enough to show:

$$\sup_h \int [\hat{m}_1(x) - m_1(x)]^2 w(x)\,dx \to_p 0, \tag{7.8}$$

$$\sup_h \int [\hat{M}(x, \hat{\xi}_n) - M(x, \hat{\xi}_n)]^2 w(x)\,dx \to_p 0, \tag{7.9}$$

$$\int [M(x, \hat{\xi}_n) - M(x, \theta_0)]^2 w(x)\,dx \to_p 0, \tag{7.10}$$

$$\sup_h \int [\hat{M}_l(x, \hat{\xi}_n) - M_l(x, \hat{\xi}_n)]^2 w(x)\,dx \to_p 0, \tag{7.11}$$

$$\int [M_l(x, \hat{\xi}_n) - M_l(x, \theta_0)]^2 w(x)\,dx \to_p 0, \tag{7.12}$$

$$\sup_h \int [\hat{M}_{l,l'}(x, \hat{\xi}_n) - M_{l,l'}(x, \hat{\xi}_n)]^2 w(x)\,dx \to_p 0. \tag{7.13}$$

Note that (7.8) and (7.9) are immediate corollaries of (6.2) and (6.3). Equations (7.10) and (7.12) are consequences of the uniform continuity assumption (4.7). Verification of (7.11) requires only a straightforward extension of the methods used on (7.4) to the case of $\hat{\xi}_n \to \theta_0$. To prove (7.13), the same general techniques as used on (7.4) apply. The only difference is that verification of

$$E\hat{M}_{l,l'}(x, \theta) \to M_{l,l'}(x, \theta) \tag{7.14}$$

requires more calculation in some cases. Note that for $l, l' \geq 3$,

$$L_{l,l'}(\theta) = S_{\theta,l,l'}(m_2(T_\theta x)),$$
$$\hat{L}_{l,l'}(\theta) = S_{\theta,l,l'}(\hat{m}_2(T_\theta x)),$$
$$L_{l,1}(\theta) = S'_{\theta,l}(m_2(T_\theta x))m'_2(T_\theta x),$$
$$\hat{L}_{l,1}(\theta) = S'_{\theta,l}(\hat{m}_2(T_\theta x))\hat{m}'_2(T_\theta x),$$
$$L_{l,2}(\theta) = S'_{\theta,l}(m_2(T_\theta x))m'_2(T_\theta x)x,$$
$$\hat{L}_{l,2}(\theta) = S'_{\theta,l}(\hat{m}_2(T_\theta x))\frac{\partial}{\partial\theta_2}\hat{m}_2(T_\theta x),$$
$$L_{11}(\theta) = S'_\theta(m_2(T_\theta x))m''_2(T_\theta x) + S''_\theta(m_2(T_\theta x))(m'_2(T_\theta x))^2,$$
$$\hat{L}_{11}(\theta) = S'_\theta(\hat{m}_2(T_\theta x))\hat{m}''_2(T_\theta x) + S''_\theta(\hat{m}_2(T_\theta x))(\hat{m}'_2(T_\theta x))^2,$$

so these cases can be handled as was done for (7.4). The hard cases arise because

$$L_{1,2}(\theta_0) = S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)m''_2(T_{\theta_0}x)x + S''_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\big(m'_2(T_{\theta_0}x)\big)^2x,$$

$$\hat{L}_{1,2}(\theta_0) = S'_{\theta_0}\big(\hat{m}_2(T_{\theta_0}x)\big)\left[\frac{\partial}{\partial\theta^{(2)}}\hat{m}'_2(T_\theta x)\right]_{\theta=\theta_0} + S''_{\theta_0}\big(\hat{m}_2(T_{\theta_0}x)\big)\left[\frac{\partial}{\partial\theta^{(2)}}\hat{m}_2(T_\theta x)\right]_{\theta=\theta_0}\hat{m}'_2(T_{\theta_0}x),$$

$$L_{2,2}(\theta_0) = S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\left[\frac{\partial^2}{\partial\theta^{(2)^2}}m_2(T_\theta x)\right]_{\theta=\theta_0} + S''_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\left[\frac{\partial}{\partial\theta^{(2)}}m_2(T_\theta x)\right]_{\theta=\theta_0},$$

$$\hat{L}_{2,2}(\theta_0) = S'_{\theta_0}\big(\hat{m}_2(T_{\theta_0}x)\big)\left[\frac{\partial^2}{\partial\theta^{(2)^2}}\hat{m}_2(T_\theta x)\right]_{\theta=\theta_0} + S''_{\theta_0}\big(\hat{m}_2(T_{\theta_0}x)\big)\left[\frac{\partial}{\partial\theta^{(2)}}\hat{m}_2(T_\theta x)\right]_{\theta=\theta_0}.$$

In view of the work done for (7.4), it remains to show that

$$E\frac{\partial}{\partial\theta^{(2)}}\hat{m}_2(T_\theta x)\bigg|_{\theta=\theta_0} \to \frac{\partial}{\partial\theta^{(2)}}m'_2(T_\theta x)\bigg|_{\theta=\theta_0} = m''_2(T_{\theta_0}x)x,$$

$$E\frac{\partial^2}{\partial\theta^{(2)^2}}\hat{m}_2(T_\theta x)\bigg|_{\theta=\theta_0} \to \frac{\partial^2}{\partial\theta^{(2)^2}}m_2(T_\theta x)\bigg|_{\theta=\theta_0} = m''_2(T_{\theta_0}x)x^2.$$

To check these, observe that, as in (7.4),

$$\begin{aligned}
E\frac{\partial}{\partial\theta^{(2)}}\hat{m}'_2(T_\theta x)\bigg|_{\theta=\theta_0} &= \int\Bigg[\frac{-2}{\theta_0^{(2)^3}h^2}K'\left(\frac{\theta_0^{(2)}x + \theta_0^{(1)} - x'}{\theta_0^{(2)}h}\right) \\
&\qquad + \frac{\big(x' - \theta_0^{(1)}\big)}{\theta_0^{(2)^4}h^3}K''\left(\frac{\theta_0^{(2)}x + \theta_0^{(1)} - x'}{\theta_0^{(2)}h}\right)\Bigg]m_2(x')\,dx' + o(1) \\
&= \int\frac{1}{\theta_0^{(2)^2}h}\frac{d^2}{du^2}\left[K\left(\frac{x-u}{h}\right)u\right]m_2\big(\theta_0^{(1)} + \theta_0^{(2)}u\big)\,du + o(1) \\
&= \int K_h(x-u)um''_2\big(\theta_0^{(1)} + \theta_0^{(2)}u\big)\,du + o(1) \\
&= xm''_2(T_{\theta_0}x) + o(1),
\end{aligned}$$

$$E\frac{\partial^2}{\partial\theta^{(2)^2}}\hat{m}_2(T_\theta x)\Big|_{\theta=\theta_0} = \int\Bigg[\frac{2}{\theta_0^{(2)^3}h}K\left(\frac{\theta_0^{(2)}x+\theta_0^{(1)}-x'}{\theta_0^{(2)}h}\right)$$

$$-\frac{2(x'-\theta_0^{(1)})}{\theta_0^{(2)^4}h^2}K'\left(\frac{\theta_0^{(2)}x+\theta_0^{(1)}-x'}{\theta_0^{(2)}h}\right)$$

$$+\frac{(x'-\theta_0^{(1)})^2}{\theta_0^{(2)^5}h^3}K''\left(\frac{\theta_0^{(2)}x+\theta_0^{(1)}-x'}{\theta_0^{(2)}h}\right)\Bigg]m_2(x')\,dx' + o(1)$$

$$= \int\frac{1}{h\theta_0^{(2)^2}}\left[\frac{d^2}{du^2}K\left(\frac{x-u}{h}\right)u^2\right]m_2\left(\theta_0^{(1)}+\theta_0^{(2)}u\right)du + o(1)$$

$$= x^2 m_2''(T_{\theta_0}x) + o(1).$$

This completes the proofs of (7.13), Lemma 2.2 and Theorem 2. □

8. Proof of Theorem 3. Since the technical details of this proof follow closely those of the proof of Theorem 2, only an outline is given. Note first that

$$\hat{L}(\theta_0) = \hat{L}(\hat{\theta}) + (\theta_0-\hat{\theta})\nabla\hat{L}(\hat{\theta}) + \tfrac{1}{2}(\theta_0-\hat{\theta})^T\hat{H}(\hat{\xi}_n)(\theta_0-\hat{\theta}),$$

where $\hat{\xi}_n$ is between θ_0 and $\hat{\theta}$. Now since the second term on the right side is 0, it is enough to show that

$$(\theta_0-\hat{\theta})^T\hat{H}(\hat{\xi}_n)(\theta_0-\hat{\theta}) = O_p(n^{-1}), \tag{8.1}$$

$$nh_0^{1/2}\left(\hat{L}(\theta_0) - n^{-1}h_0^{-1}C_\mu\right) \to_{\mathscr{L}} N(0, C_\sigma^2). \tag{8.2}$$

Theorem 2, together with the methods of Section 7, make (8.1) easy to verify. To check (8.2), note that

$$\hat{L}(\theta_0) = \int\left[(\hat{m}_1(x)-E\hat{m}_1(x)) - (\hat{M}(x,\theta_0)-E\hat{M}(x,\theta_0))\right]^2 w(x)\,dx$$

$$+ O_p(n^{-1})$$

$$= \int\Bigg[n^{-1}\sum_i K_h(x-x_i)\varepsilon_i - n^{-1}$$

$$\sum_i K_{h'}(T_{\theta_0}x - x_i')\varepsilon_i'\left(S_{\theta_0}'\left(m_2(T_{\theta_0}x)\right)\right)\Bigg]^2 w(x)\,dx$$

$$+ O_p(n^{-1})$$

$$= n^{-2}\sum_i\sum_{i'}\left[A_{ii'}\varepsilon_i,\varepsilon_{i'} - B_{ii'}\varepsilon_i\varepsilon_{i'}' - B_{i'i}\varepsilon_i'\varepsilon_{i'} + C_{ii'}\varepsilon_i'\varepsilon_{i'}'\right] + O_p(n^{-1}),$$

where

$$A_{ii'} = \int K_h(x - x_i)K_h(x - x_{i'})w(x)\,dx,$$

$$B_{ii'} = \int K_h(x - x_i)K_{h'}(T_{\theta_0}x - x'_{i'})\big(S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\big)w(x)\,dx,$$

$$C_{ii'} = \int K_{h'}(T_{\theta_0}x - x'_i)K_{h'}(T_{\theta_0}x - x'_{i'})\big(S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\big)^2 w(x)\,dx.$$

Hence,

$$\begin{aligned} E\big(\hat{L}(\theta_0)\big) &= n^{-2}\sum_i \big[A_{ii}\sigma^2 + C_{ii}\sigma'^2\big] + O(n^{-1}) \\ &= n^{-1}\sigma^2 \iint K_h(x-u)^2 w(x)\,dx\,du \\ &\quad + n^{-1}\sigma'^2 \iint K_{h'}(T_{\theta_0}x - x')^2\big(S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\big)^2 w(x)\,dx\,dx' \\ &\quad + O(n^{-1}) \\ &= (nh)^{-1}\sigma^2\Big(\int K^2\Big)\Big(\int w\Big) + (nh)^{-1}\sigma'^2\Big(\int K^2\Big)\Big(\int \big(S'_{\theta_0}(m_2(x))\big)^2 w\Big) \\ &\quad + O(n^{-1}). \end{aligned}$$

To understand the variance structure of $\hat{L}(\theta_0)$, note first that

$$\begin{aligned} &\operatorname{var}\Big(n^{-2}\sum_{i\neq i'}\sum A_{ii'}\varepsilon_i\varepsilon_{i'}\Big) \\ &\quad = n^{-4}\sum_{i\neq i'}\sum \big(A_{ii'}^2 + A_{ii'}A_{i'i}\big)\sigma^4 \\ &\quad = n^{-2}\sigma^4 \iint 2\Big[\int K_h(x-u_1)K_h(x-u_2)w(x)\,dx\Big]^2 du_1\,du_2 \\ &\qquad + o(n^{-2}) \\ &\quad = n^{-2}h^{-1}2\sigma^4\Big(\int K * K^2\Big)\Big(\int w\Big) + o(n^{-2}h^{-1}), \end{aligned}$$

where $*$ denotes convolution, that

$$\begin{aligned} &\operatorname{var}\Big(2n^{-2}\sum_{i\neq i'}\sum B_{ii'}\varepsilon_i\varepsilon'_{i'}\Big) \\ &\quad = 4n^{-4}\sum_{i\neq i'}\sum B_{i'i}^2\sigma^2\sigma'^2 \\ &\quad = n^{-2}4\sigma^2\sigma'^2 \iint \Big[\int K_h(x-u_1)K_{h'}(T_{\theta_0}x - u_2) \\ &\qquad\qquad \big(S'_{\theta_0}\big(m_2(T_{\theta_0}x)\big)\big)w(x)\,dx\Big]^2 du_1\,du_2 \\ &\qquad + o(n^{-1}) \\ &\quad = n^{-2}h^{-1}4\sigma^2\sigma'^2\Big(\int K * K^2\Big)\Big(\int \big[S'_{\theta_0}(m_2(x))\big]^2 w\big(T_{\theta_0}^{-1}x\big)\,dx\Big) + o(n^{-2}h^{-1}), \end{aligned}$$

and that

$$\operatorname{var}\Big(n^{-2}\sum_{i\neq i'}\sum C_{ii'}\varepsilon_i'\varepsilon_{i'}'\Big)$$
$$= n^{-2}\sigma'^4 2\int\int\Big[\int K_{h'}(T_{\theta_0}x-u_1)K_{h'}(T_{\theta_0}x-u_2)$$
$$S_{\theta_0}'\big(m_2(T_{\theta_0}x)\big)\big)^2 w(x)\,dx\Big]^2 du_1\,du_2$$
$$+\,o(n^{-1})$$
$$= n^{-2}h^{-1}2\sigma'^4\Big(\int (K*K)^2\Big)\Big(\theta_0^{(2)}\int\big[S_{\theta_0}'(m_2(x))\big]^4 w\big(T_{\theta_0}^{-1}x\big)\,dx\Big)$$
$$+\,O(n^{-2}h^{-1}).$$

But

$$\operatorname{var}\Big(n^{-2}\sum_i A_{ii}\varepsilon_i^2\Big) = n^{-4}\sum_i A_{ii}^2\operatorname{var}(\varepsilon_i^2) = O(n^{-3}h^{-2}),$$
$$\operatorname{var}\Big(n^{-2}\sum_i B_{ii}\varepsilon_i\varepsilon_i'\Big) = O(n^{-3}h^{-2}),$$
$$\operatorname{var}\Big(n^{-2}\sum_i C_{ii}\varepsilon_i'^2\Big) = O(n^{-3}h^{-2}),$$
$$\operatorname{cov}\Big(n^{-2}\sum_i\sum_{i'}A_{ii'}\varepsilon_i\varepsilon_{i'},\,2n^{-2}\sum_i\sum_{i'}B_{ii'}\varepsilon_i\varepsilon_{i'}'\Big) = 0,$$
$$\operatorname{cov}\Big(n^{-2}\sum_i\sum_{i'}A_{ii'}\varepsilon_i\varepsilon_{i'},\,n^{-2}\sum_i\sum_{i'}C_{ii'}\varepsilon_i'\varepsilon_{i'}'\Big) = 0,$$
$$\operatorname{cov}\Big(2n^{-2}\sum_i\sum_{i'}B_{ii'}\varepsilon_i\varepsilon_{i'},\,n^{-2}\sum_i\sum_{i'}C_{ii'}\varepsilon_i'\varepsilon_{i'}'\Big) = 0.$$

Hence,

$$\operatorname{var}\big(\hat{L}(\theta_0)\big) = n^{-2}h^{-1}C_\sigma^2 + o(n^{-2}h^{-1}).$$

To verify the asymptotic normality, first obtain it for $n^{-2}\Sigma_i\Sigma_{i'}A_{ii'}\varepsilon_i\varepsilon_{i'}$ and $n^{-2}\Sigma_i\Sigma_{i'}C_{ii'}\varepsilon_i'\varepsilon_{i'}'$ using Theorem 1 of Whittle (1964), with his r taken to be $n^{1/10}$, and for $2n^{-2}\Sigma_i\Sigma_{i'}B_{ii'}\varepsilon_i\varepsilon_{i'}'$ by an ordinary central limit theorem for arrays. An application of the Cramer–Wold device then gives (8.2). This completes the proof of Theorem 3. □

Acknowledgment. The authors are grateful to Raymond Carroll for interesting discussions and for suggesting the method of parameter estimation.

REFERENCES

CAMERON, M. A. (1983). The comparison of time series recorders. *Technometrics* **25** 9–22.

CAMERON, M. A. and HANNAN, E. J. (1979). Transient signals. *Biometrika* **66** 243–258.

CAMERON, M. A. and THOMPSON, P. J. (1985). Measuring attenuation. In *Handbook of Statistics* (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) **5** 363–387. Elsevier, Amsterdam.

CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic, New York.

COLLOMB, G. (1981). Estimation non paramétrique de la régression: Revué bibliographique. *Internat. Statist. Rev.* **49** 75–93.

COLLOMB, G. (1985). Non-parametric regression: An up-to-date bibliography. *Math. Operationsforsch. Ser. Statist.* **16** 309–324.

COLLOMB, G. and HÄRDLE, W. (1986). Strong uniform convergence rates in robust non-parametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic Process Appl.* **23** 77–89.

ENGLE, R. F., GRANGER, C. W. I., RICE, I. and WEISS, A. (1986). Semi-parametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81** 310–320.

EPPINGER, R. H., MARCUS, J. H. and MORGAN, D. H. (1984). Development of dummy and injury index for NHTSA's thoracic side impact protection research program. Government/Industry Meeting and Exposition, Washington D.C.

GASSER, T., MÜLLER, H. G., KÖHLER, W., MOLINARI, L. and PRADER, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.* **12** 210–229.

GASSER, T., MÜLLER, H. G. and MAMMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252.

GREEN, P. J. (1985). Linear models for field trials smoothing and cross-validation. *Biometrika* **72** 527–537.

HÄRDLE, W., HALL, P. and MARRON, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.* **83** 86–101.

HÄRDLE, W. and MARRON, J. S. (1985a). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13** 1465–1481.

HÄRDLE, W. and MARRON, J. S. (1985b). Asymptotic nonequivalence of some bandwidth selectors in nonparametric regression. *Biometrika* **72** 481–484.

HE, Q. (1988). On estimating the parametric part of the shape-invariant model. Unpublished manuscript.

KALLIERIS, D., MATTERN, R. and HÄRDLE, W. (1986). Belastungsgrenze und Verletzungsmechanik des angegurteten PKW-Insassen beim 90-Grad-Seitenaufprall. Forschungsvereinigung Automobiltechnik Schriftreihe 37, Frankfurt/M.

KNEIP, A. and GASSER, T. (1988). Convergence and consistency results for self modeling nonlinear regression. *Ann. Statist.* **16** 82–112.

LAWTON, W. H., SYLVESTRE, E. A. and MAGGIO, M. S. (1972). Self modeling nonlinear regression. *Technometrics* **14** 513–532.

MARRON, J. S. and HÄRDLE, W. (1986). Random approximation to an error criterion of nonparametric statistics. *J. Multivariate Anal.* **20** 91–113.

MARRON, J. S. and RUDEMO, M. (1988). Pooling smoothing information in nonparametric regression. Unpublished manuscript.

MARRON, J. S. and SCHMITZ, H. P. (1988). Simultaneous estimation of several size distributions of income. Discussion paper A-186. SFB303, Universität Bonn.

PRIESTLEY, M. B. and CHAO, M. T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B* **34** 385–392.

RICE, J. (1984a). Boundary modification for kernel regression. *Comm. Statist. A—Theory Methods* **13** 893–900.

RICE, J. (1984b). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12** 1215–1230.

RICE, J. (1986). Convergence rates for partially splined models. *Statist. Probab. Lett.* **4** 203–208.

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

SPECKMAN, P. (1986). Kernel smoothing in partial linear models. Unpublished manuscript.

WATSON, G. S. (1964). Smooth regression analysis. *Sanknyā Ser. A* **26** 359–372.

WHITTLE, P. (1964). On the convergence to normality of quadratic forms in independent variables. *Theory Probab. Appl.* **9** 103–108.

WIRTSCHAFTSTHEORIE II
UNIVERSITÄT BONN
ADENAUERALLEE 24-26
D-5300 BONN 1
WEST GERMANY

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL
CHAPEL HILL, NORTH CAROLINA 27514