The Annals of Statistics
1993, Vol. 21, No. 4, 1926–1947

COMPARING NONPARAMETRIC VERSUS PARAMETRIC REGRESSION FITS[1]

BY W. HÄRDLE AND E. MAMMEN

Humboldt-Universität zu Berlin

In general, there will be visible differences between a parametric and a nonparametric curve estimate. It is therefore quite natural to compare these in order to decide whether the parametric model could be justified. An asymptotic quantification is the distribution of the integrated squared difference between these curves. We show that the standard way of bootstrapping this statistic fails. We use and analyse a different form of bootstrapping for this task. We call this method the wild bootstrap and apply it to fitting Engel curves in expenditure data analysis.

1. Motivation. The appropriateness of parametric modelling of regression data may be judged by comparison with a nonparametric smoothing estimator. For this purpose one may use a squared deviation measure between the two fits. The integrated squared deviation can be used as a test statistic for testing the parametric model where the critical value is determined by the asymptotic distribution of this statistic. The convergence to the asymptotic normal distribution is quite slow so that it seems more appropriate not to use the asymptotic critical values. A way of computing critical values could possibly be based on resampling from the entire set of observations. It is shown here that this method of bootstrapping fails. Instead we propose in this setting a new variant of the bootstrap [due to Wu (1986)] which we call the *wild bootstrap*. The classical bootstrap—to resample from the entire data—does not work. The bootstrapped statistic has not the same limit behaviour. It will be shown that the wild bootstrap works. The fact that classic resampling does not work here is theoretically appealing and can be understood via a Hoeffding decomposition. Indeed the quadratic term of the proposed test statistic dominates asymptotically the linear term.

It is surprising that although the nonparametric approach in modelling regression relationships has received a lot of attention recently [see Collomb, (1981)], there are only a few theoretical results on how to compare parametric with nonparametric fits. In practical studies the importance of comparing parametric with nonparametric curves has been pointed out in the analysis of growth curves by Gasser, Köhler, Müller, Lago, Molinari and Prader (1985). Another example stems from the analysis of the income distribution of

Received May 1990; revised January 1993.

[1]Research supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 123, "Stochastische Mathematische Modelle" and Sonderforschungsbereich 303, "Information und Koordination Wirtschaftlicher Aktivitäten", CORE, Université Catholique de Louvain and CentER, Katholieke Universiteit Brabant.

AMS 1991 *subject classifications*. Primary 62G07; secondary 62G09.

Key words and phrases. Kernel estimate, bootstrap, wild bootstrap, goodness-of-fit test.

British households (*Family Expenditure Survey*, 1968–1983). Hildenbrand and Hildenbrand (1986) found that the widely used lognormal fit for the income distribution was not able to model the seemingly bimodal distribution.

Theoretical results in this direction are offered by Yanagimoto and Yanagimoto (1987), Cleveland and Devlin (1988), Neuhaus (1986, 1988), Cox, Koh, Wahba and Yandell (1988), Azzalini, Bowman and Härdle (1989), Cox and Koh (1989), Munson and Jernigan (1989), Eubank and Spiegelman (1990), Härdle and Marron (1990), le Cessie and van Houwelingen (1991), Staniswalis and Severini (1991). LaRiccia (1991) used the idea of comparing a parametric model for the quantile function against a nonparametric alternative for testing a composite goodness-of-fit null hypothesis. An asymptotic χ^2 distribution was derived.

In Section 2 we derive the asymptotic distribution of the squared deviation between the parametric and the nonparametric fit. The asymptotic distribution could be estimated with the "plug-in" method, although the involved functionals seem to be rather complicated. Section 3 is devoted to the question of how to bootstrap in this setting. Section 4 gives several simulations and an application to Engel curve estimation. The proofs are given in Section 5.

2. How far is the nonparametric from the parametric model? We consider the following model. Given are n i.i.d. observations $\{(X_i, Y_i)\}_{i=1}^n$ ($X_i \in \mathbb{R}^d$, $Y_i \in \mathbb{R}$) with unknown regression function $m(\cdot) = E(Y_i|X_i = \cdot)$. We write also $Y_i = m(X_i) + \varepsilon_i$ with $E(\varepsilon_i|X_i) = 0$. We do not assume that the ε_i are conditional i.i.d. as in Eubank and Spiegelman (1990) and Härdle and Marron (1990). In particular, this contains the case of conditional heteroscedasticity. We are interested in the following testing problem. We wish to test the parametric model $\{m_\theta: \theta \in \Theta\}$ against the nonparametric alternative which only assumes that $m(\cdot)$ is "smooth." A natural approach is to plot a parametric regression estimator $m_{\hat\theta}$ and a kernel estimator $\hat m_h$ with bandwidth $h = h_n$ and kernel K [Nadaraya (1964); Watson (1964)]:

$$\hat m_h(\cdot) = \frac{\sum_{i=1}^n K_h(\cdot - X_i) Y_i}{\sum_{i=1}^n K_h(\cdot - X_i)},$$

$$K_h(\cdot) = h^{-d} K(\cdot / h).$$

For simplification of notation the dependence of $\hat\theta$ and h on n will be dropped.

The question arises if visible differences between $m_{\hat\theta}$ and $\hat m_h$ can be explained by stochastic fluctuations or if they suggest to use nonparametric instead of parametric methods. One way to proceed is to measure the distance between $m_{\hat\theta}$ and $\hat m_h$ and to use this distance as test statistic for testing the parametric model. Here we study the L_2-distance between the nonparametric and parametric fits. The use of this distance is motivated by mathematical convenience. Certainly from a more data analytic point of view distances would be more satisfactory which reflect similarities in the shape of the regression functions, but nevertheless we will restrict ourselves to the treatment of the weighted L_2-distance $\int (\hat m_h - m_{\hat\theta})^2 \pi$ where π is a weight function.

Let $\mathscr{K}_{h,n}$ denote the (random) smoothing operator

$$\mathscr{K}_{h,n}g(\cdot) = \frac{\sum_{i=1}^n K_h(\cdot - X_i)g(X_i)}{\sum_{i=1}^n K_h(\cdot - X_i)}.$$

Because of $E(\hat{m}_h(\cdot)|X_1, \dots, X_n) = \mathscr{K}_{h,n}m(\cdot)$ we consider the following modification of the squared deviation between $\hat{m}_h$ and $m_{\hat{\theta}}$:

$$T_n = nh^{d/2} \int \left(\hat{m}_h(x) - \mathscr{K}_{h,n}m_{\hat{\theta}}(x)\right)^2 \pi(x)\, dx.$$

In this definition $\hat{m}_h$ is compared with the parametric "estimate" $\mathscr{K}_{h,n}m_{\hat{\theta}}$ of the conditional expectation of $\hat{m}_h$.

We propose to use T_n as a test statistic to test the parametric hypothesis,

$$m \in \{m_\theta: \theta \in \Theta\}.$$

On the hypothesis, T_n is asymptotically equivalent to the sum of a constant and a purely quadratic form (see the proof of Proposition 1). On the contrary, the test statistic $\int(\hat{m}_h - m_{\hat{\theta}})^2\pi$ contains asymptotically also a linear term. This linear term makes this test only sensitive against certain "smooth" deviations from the hypothesis (see Proposition 2).

Alternative definitions of a test statistic are possible. The integral may be replaced by a sum (e.g., over the design points). Furthermore the integrand may be multiplied by a power of $\hat{f}_h(\cdot) = n^{-1}\sum_{i=1}^n K_h(\cdot - X_i)$. Under our assumptions the asymptotic arguments for these modifications are applicable but constants might have to be changed.

For an approximate calculation of critical values we determine the asymptotic distribution of T_n for a parametric $m = m_{\theta_0}$. Furthermore for a comparison of T_n with other goodness-of-fit tests we calculate the asymptotic power of T_n if m (possibly depending on n) lies in the alternative: say $m(x) = m_n(x) = m_{\theta_0}(x) + c_n\Delta_n(x)$ for certain sequences c_n and Δ_n. It is most appropriate to choose c_n such that the asymptotic power of T_n is bounded away from 1 and from the level. We will show that for instance for "regular" constant $\Delta_n(x) = \Delta(x)$ this will be the case for $c_n = n^{-1/2}h^{-d/4}$.

3. Assumptions. We make the following assumptions on the stochastic nature of the observations and the parametric estimator of the regression function.

(A1). With probability 1 X_i lies in a compact set (w.l.o.g. $[0,1]^d$). The marginal density $f(\cdot)$ of X_i is bounded away from zero.

(A2). $m(\cdot)$ and $f(\cdot)$ are twice continuously differentiable. π is continuously differentiable.

(A3). $\Delta_n(\cdot)$ is bounded (uniformly in x and n) and $c_n = n^{-1/2}h^{-d/4}$. This contains the parametric case in particular because $\Delta_n \equiv 0$ is possible.

(A4). $\sigma^2(\cdot) = \operatorname{var}(Y_i|X_i = x)$ is bounded away from 0 and from ∞.

(A5). $E\exp(t\varepsilon_i)$ is uniformly bounded in i and n for $|t|$ small enough [where $\varepsilon_i = Y_i - m(X_i)$].

Before stating our assumptions for the parametric model let us consider the special case of a k-dimensional linear model,

$$m_\theta(\cdot) = \theta_1 g_1(x) + \cdots + \theta_k g_k(x) = \langle \theta, g(\cdot)\rangle,$$

where g is a $\mathbb{R}^k$-valued function. With a smooth weight function w the weighted least squares estimator $\hat{\theta}_n = \hat{\theta}$ is defined by

$$\hat{\theta} = \arg\min_\theta \sum_{i=1}^n w(X_i)(Y_i - m_\theta(X_i))^2.$$

In the linear model $\hat{\theta}$ can easily be calculated:

$$\hat{\theta} = \left(\sum_{i=1}^n w(X_i)g(X_i)g(X_i)^T\right)^{-1} \sum_{i=1}^n w(X_i)g(X_i)Y_i.$$

Consider now a regression function $m = m_n$ which may lie in the hypotheses or in the alternative. We want to write m as $m(\cdot) = m_{\theta_0}(\cdot) + c_n\Delta_n(\cdot)$ for some θ_0 (which may also depend on n) and Δ_n. θ_0 and $\Delta_n(\cdot)$ may be chosen as follows:

$$\theta_0 = \arg\min_\theta \int w(x)(m(x) - m_\theta(x))^2\,dx,$$

$$\Delta_n(\cdot) = \frac{1}{c_n}(m(\cdot) - m_{\theta_0}(\cdot)).$$

With this choice of m_{θ_0} and Δ_n, Δ_n is orthogonal to $\{m_\theta(x)\colon \theta \in \Theta\}$ in the following sense:

$$\int w(x)f(x)\Delta_n(x)g_j(x)\,dx = 0, \qquad j = 1,\dots,k.$$

This implies that the expectation of $\hat{\theta}$ is approximately θ_0 as can be seen by the following stochastic expansion of $\hat{\theta}$:

$$\begin{aligned}\hat{\theta} &= \theta_0 + \left(\int w(x)f(x)g(x)g(x)^T\,dx\right)^{-1}\\ &\quad\times\left\{\frac{1}{n}\sum_{i=1}^n w(X_i)g(X_i)\varepsilon_i + c_n\int w(x)f(x)\Delta_n(x)g(x)\,dx\right\} + O_p\left(\frac{c_n}{\sqrt{n}}\right)\\ &= \theta_0 + \frac{1}{n}\sum_{i=1}^n h(X_i)\varepsilon_i + O_p\left(\frac{c_n}{\sqrt{n}}\right),\end{aligned}$$

where

$$h(\cdot) = \left(\int w(x) f(x) g(x) g(x)^T \, dx\right)^{-1} w(\cdot) g(\cdot).$$

If g, h are bounded functions, then especially this implies the following:

(P1). $m_{\hat{\theta}}(\cdot) - m_{\theta_0}(\cdot) = (1/n)\Sigma_{i=1}^n < g(\cdot),\ h(X_i) > \varepsilon_i + o_p((n \log n)^{-1/2})$ (uniformly in x), where g and h are bounded functions taking values in $\mathbb{R}^k$ for some k.

For the parametric model we will assume in this section only (P1). By linearization it can be shown that (P1) holds also for weighted least squares estimators $\hat{\theta}$ in nonlinear models if $m(\cdot)$ and $w(\cdot)$ are "smooth" and Δ_n and θ_0 are chosen similarly with $g_j(\cdot) = (\partial/\partial\theta_j) m_{\theta_0}(\cdot)$.

For the kernel K we make the following assumptions.

(K1). The kernel K is a symmetric, twice continuously differentiable function with compact support, furthermore $\int K(u)\, du = 1$.

(K2). The bandwidth h fulfills $h = h_n \sim n^{-1/(d+4)}$.

In particular (K2) is fulfilled for every choice of the bandwidth h which is asymptotically optimal for the class of twice continuously differentiable regression functions. For simplicity of notation we do not consider bandwidths which are asymptotically optimal for other smoothness classes.

4. The asymptotic behavior of T_n.

In the following proposition we will approximate the distribution of T_n by a Gaussian distribution with mean which converges to infinity (also for the null hypothesis). We will measure the distance between these distributions by the following modification of the Mallows distance

$$d(\mu, \nu) = \inf_{X, Y} \left(E\|X - Y\|^2 \wedge 1 : \mathscr{L}(X) = \mu, \mathscr{L}(Y) = \nu\right).$$

Convergence in this metric is equivalent to weak convergence.

PROPOSITION 1. *Assume* (*A*1)–(*A*5), (*P*1), (*K*1) *and* (*K*2). *Then*

$$d\left(\mathscr{L}(T_n), N\left(b_h + \int (\mathscr{K}_h \Delta_n)^2 \pi, V\right)\right) \to 0,$$

where

$$b_h = h^{-d/2} K^{(2)}(0) \int \frac{\sigma^2(x)\pi(x)}{f(x)}\, dx$$

$$V = 2h^d \int \frac{\sigma^2(x)\sigma^2(y)\pi(x)\pi(y)}{f(x) f(y)} \left(K_h^{(2)}(x - y)\right)^2 dx\, dy.$$

$\mathscr{K}_h$ denotes the following smoothing operator:

$$\mathscr{K}_h g(\cdot) = \int K_h(\cdot - t) g(t)\, dt.$$

$K_h^{(j)}$ denotes the j-times convolution product of K_h. If $\sigma^2(\cdot)$ is continuous, V can be chosen as

$$V = 2K^{(4)}(0) \int \frac{[\sigma^2(x)]^2 \pi(x)^2}{f^2(x)}\, dx.$$

Because of the slow order of convergence, we do not recommend using Proposition 1 for the approximate calculation of critical values (see also the simulations in Section 6). We will use Proposition 1 to study consistency of different bootstrap procedures in the next section. The proposition gives also a rough impression on the power of T_n. It shows that for $d = 1$ the power of the goodness of fit test based on T_n is asymptotically constant on regions of the form $\{m_{\theta_0} + n^{-9/20}\Delta\colon \int(\mathscr{K}_h\Delta)^2\pi = \text{const.}\}$. This can be compared with the behavior of other goodness of fit tests. The accuracy of the parametric model may also be checked by testing against a higher dimensional parametric model or by test statistics which are asymptotically of Cramér–von Mises type or of Kolmogorov–Smirnov type. These tests have nontrivial power on points contiguous to the parametric model (i.e., $m = m_{\theta_0} + n^{-1/2}\Delta$) but they are of more parametric nature—in the sense that they look into certain one-dimensional directions [Durbin and Knott (1972) and Milbrodt and Strasser (1990)]. The nonparametric behaviour of T_n (nearly the same power for all deviations of fixed weighted L_2-norm) must be paid by the larger distance ($n^{-9/20}$ instead of $n^{-1/2}$) which can be detected by the test.

We expect that the proposition holds true also for data adapted bandwidth $\hat{h}$ as long as $\hat{h}/h \to_p 1$ but we do not investigate this further. In a related context (two-sample tests based on kernel density estimates) this has been shown by tightness arguments; see Chapter 3 of Mammen (1992).

Tests based on the statistic $\int(\hat{m}_h - m_{\hat{\theta}})^2\pi$ behave quite differently. This can be seen from Proposition 2. This test is an asymptotic linear test. It is sensitive to deviations from the hypothesis in one direction of order $n^{-1/2}$.

PROPOSITION 2. *Suppose* (A1), (A2), (A4), (A5), (P1), (K1), (K2) *and* (A3′) *$\Delta_n(\cdot)$ is bounded (uniformly in x and n) and $c_n = n^{-1/2}$.*

Furthermore, suppose that $\sigma^2(\cdot)$ is continuous and for simplicity, that the dimension d is 1. *Then*

$$d\Bigg(\mathscr{L}\bigg(nh^{1/2}\int(\hat{m}_h - m_{\hat{\theta}})^2\pi\bigg),$$

$$N\bigg(b'_h + n^{1/2}h^{5/2}c_K\int\rho(x)\Delta_n(x)\pi(x)\, dx, V'\bigg)\Bigg) \to 0,$$

where

$$\rho(x) = m''(x) + \frac{2f'(x)}{f(x)} m'(x),$$

$$c_K = \int u^2 K(u)\, du,$$

$$b'_h = b_h + \tfrac{1}{4} c_K^2 n h^{9/2} \int \rho^2(x) \pi(x)\, dx,$$

$$V' = V + c_K^2 n h^5 \int \Big[\rho(x)\pi(x) - h(x) f(x) \int \rho(u) g(u) \pi(u)\, du \Big]^2 \sigma^2(x) f^{-1}(x)\, dx.$$

5. How to bootstrap. The proof of Proposition 1 is based on a stochastic expansion with error terms of order $n^{-1/10}$. Therefore the theorem can only give a rough idea of the stochastic behaviour of T_n if the sample is small. In the simulations given in the next section we will see that the normal approximation does not work very well for moderate sample sizes. We will study in this section bootstrap methods as an alternative to asymptotics. We consider three different possibilities of bootstrapping:

1. the naive resampling method;
2. the adjusted residual bootstrap;
3. the wild bootstrap.

We show that only the third type of bootstrap will work. The *naive bootstrap* consists of simple resampling from the original observations. That is the bootstrap sample $\{(X_i^*, Y_i^*)\}_{i=1}^n$ is drawn (with replacement) out of the set $\{(X_i, Y_i)\}_{i=1}^n$. Then create $T^{*,N}$ like T_n by the squared deviation between the parametric fit $m_{\hat\theta^*}$ and the nonparametric fit $\hat m_h^*$ (both computed from the bootstrap sample $\{(X_i^*, Y_i^*)\}_{i=1}^n$),

$$T^{*,N} = n h^{d/2} \int \big(\hat m_h^*(x) - \mathscr{K}_{h,n} m_{\hat\theta^*}(x) \big)^2 \pi(x)\, dx.$$

The conditional distribution $\mathscr{L}^*(T^{*,N}) = \mathscr{L}(T^{*,N} | \{(X_i, Y_i)\}_{i=1}^n)$ can be approximated by Monte Carlo simulations. From this Monte Carlo approximation define the $(1-\alpha)$ quantile $\hat t_\alpha^N$ and reject the parametric hypothesis if $T_n > \hat t_\alpha^N$. We call this the naive bootstrap because this resampling does not correctly reflect the stochastic structure of our model as we will see below.

The bootstrap estimate should suffice two conditions. If m lies in the parametric model, $\mathscr{L}^*(T^*)$ should consistently estimate the distribution of T_n. On the alternative $\mathscr{L}^*(T^*)$ should approximate a distribution of T_n under the null hypothesis. This is important for a good power performance. For the second aim we consider the following modification. The bootstrap with adjusted residuals is defined by resampling from the observations $\{(X_i, Y_i -$

$\hat{m}_h(X_i) + m_{\hat{\theta}}(X_i))\}_{i=1}^n$. $T^{*,A}$ might now be created like T_n by the squared deviance between the parametric fit and the nonparametric fit. As above the conditional distribution $\mathscr{L}^*(T^{*,A}) = \mathscr{L}(T^{*,A}|\{(X_i, Y_i)\}_{i=1}^n)$ can be approximated by Monte Carlo simulations. From $\mathscr{L}^*(T^{*,A})$ define the $(1-\alpha)$ quantile $\hat{t}_\alpha^A$ and reject the parametric hypothesis if $T_n > \hat{t}_\alpha^A$. We call this the *adjusted residual bootstrap*. In the following theorem we show that on the null hypothesis both $\mathscr{L}^*(T^{*,A})$ and $\mathscr{L}^*(T^{*,N})$ have variance larger than that of $\mathscr{L}(T_n)$. This implies that naive bootstrap and adjusted residual bootstrap do not work. Both procedures lead to very conservative tests.

THEOREM 1. *Assume* (A1)–(A5), (K1), (K2) *and* $\Delta_n = 0$. (*This implies that m lies on the hypothesis, i.e.,* $m = m_{\theta_0}$ *for a* θ_0.) *Define* $\hat{\theta}^{*,N}$ *and* $\hat{\theta}^{*,A}$ *but with the bootstrap data* $\{(X_i^*, Y_i^*)\}_{i=1}^n$ *instead of* $\{(X_i, Y_i)\}_{i=1}^n$. *Assume for* $\hat{\theta}^* = \hat{\theta}^{*,N}$ (*or* $= \hat{\theta}^{*,A}$, *resp.*)

$$(P1') \qquad m_{\hat{\theta}^*}(\cdot) - m_{\theta_0}(\cdot) = \frac{1}{n}\sum_{i=1}^n \langle g(\cdot), h(X_i)\rangle \varepsilon_i^* + o_p\left(\frac{1}{\sqrt{n \log n}}\right),$$

where $\varepsilon_i^* = Y_i^* - m(X_i^*)$. *Assume also that the variance function* $\sigma^2(\cdot)$ *is continuous. Then the bootstrap estimates of the variance of* T_n *converge as follows*:

$$\operatorname{var}^*(T^{*,N}) - 3\operatorname{var}(T_n) \to_p 0,$$

$$\operatorname{var}^*(T^{*,A}) - \operatorname{var}(T_n) - 4\int \frac{[\sigma^2(x)]^2 \pi^2(x)}{f^2(x)}\,dx \int \left(K^{(2)}(t) - K^{(3)}(t)\right)^2 dt \to_p 0.$$

As above it can easily be seen that (P1′) is fulfilled for weighted least squares estimators $\hat{\theta}$ in linear models. Furthermore it can be shown that (P1′) holds for nonlinear models under standard regularity conditions. Theorem 1 shows that the above proposed bootstrap procedures (also after a bias correction) are inconsistent. At first sight this result seems to be surprising and against the intuition of the bootstrap. The deeper reason lies in the fact that under the bootstrap distribution the regression function is *not* the conditional expectation of the observation. Under our assumptions we get almost surely $E^*(Y_i^*|X_i^*) = Y_i^*$, which is typically different from $m_{\hat{\theta}}(X_i^*)$. Here E^* denotes the conditional expectation $E(\cdot\,|\{(X_i, Y_i)\}_{i=1}^n)$. Wu (1986) has pointed out inconsistency for bootstrap estimates of the least-squares estimator in linear models for nonconstant conditional variance function. In our case the bootstrap will also break down, even for homoscedastic errors. As an alternative we recommend the *wild bootstrap* which is related to proposals of Wu (1986) [see also Beran (1986), Liu (1988) and Mammen (1993)].

This approach does not mimic the i.i.d. structure of (X_i, Y_i). It is rather constructed so that

$$E^*(Y_i^*|X_i^*) = m_{\hat{\theta}}(X_i^*).$$

For this purpose define

$$\tilde{\varepsilon}_i = Y_i - \hat{m}_h(X_i).$$

Since we are going to use this *single residual* $\tilde{\varepsilon}_i$ to estimate the conditional distribution $\mathscr{L}(Y_i - m(X_i)|X_i)$ by an $\hat{F}_i$ we are calling it the *wild bootstrap*. More precisely define an arbitrary distribution $\hat{F}_i$ such that

$$E_{\hat{F}_i} Z = 0,$$

$$E_{\hat{F}_i} Z^2 = (\tilde{\varepsilon}_i)^2,$$

$$E_{\hat{F}_i} Z^3 = (\tilde{\varepsilon}_i)^3.$$

We use a two-point distribution which is uniquely determined by these requirements. For other constructions see Liu (1988).

Now construct independent $\varepsilon_i^* \sim \hat{F}_i$ and use $(X_i,\ Y_i^* = m_{\hat{\theta}}(X_i) + \varepsilon_i^*)$ as bootstrap observations. Then create $T^{*,W}$ like T_n by the squared deviation between the parametric fit and the nonparametric fit. From the Monte Carlo approximation of $\mathscr{L}^*(T^{*,W})$ construct the $(1-\alpha)$ quantile $\hat{t}_\alpha^W$ and reject the parametric hypothesis if $T_n > \hat{t}_\alpha^W$. In the following theorem we show that this procedure works. On the null hypothesis it estimates consistently the distribution of T_n. On the alternative the wild bootstrap estimate converges to a distribution under the null hypothesis. Note that the wild bootstrap mimics correctly the conditional expectation $E^*(Y_i^*) = m_{\hat{\theta}}(X_i)$.

THEOREM 2. *Assume (A1)–(A5), (P1), (K1) and (K2). Furthermore suppose for the parametric estimator $\hat{\theta}^*$ (based on the bootstrap sample)*

$$(P1'') \quad m_{\hat{\theta}^*}(\cdot) - m_{\hat{\theta}}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \langle g(x), h(X_i)\rangle \varepsilon_i^* + o_p\left(\frac{1}{\sqrt{n \log n}}\right).$$

Then

$$d\big(\mathscr{L}^*(T^{*,W}), N(b_h, V)\big) \to_p 0,$$

where b_h and V are defined in Theorem 1.

Like (P1′) condition (P1″) is also fulfilled under standard regularity conditions. Note that the proposed wild bootstrap procedure requires rather weak assumptions on the conditional distributions of the errors. If one had prior information on the error structure, one could modify the resampling scheme. One such prior information is smoothness of the variance function $\sigma^2(x)$. In this case we propose to define the variance of $\hat{F}_i$ as $\hat{\sigma}^2(X_i)$, where $\hat{\sigma}^2(\cdot)$ is a nonparametric estimator of $\sigma^2(\cdot)$, see Carroll (1982). For another bootstrap procedure where the conditional error distributions vary smoothly with x, see

Cao-Abad and Gonzales-Manteiga (1990). For the special case where the errors are conditionally i.i.d. one could bootstrap from the entire set of (scaled) residuals. We expect that these schemes would work better under the additional model assumptions.

6. Simulations and applications. We have checked the validity of our asymptotic results in a Monte Carlo experiment. In a first simulation we generated $\{X_i\}_{i=1}^n$, $n = 100$, uniformly in $[0,1]$ and $Y_i = m(X_i) + \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2)$, $\sigma = 0.1$, independent of X_i. The regression function has been put $m(x) \equiv 0$. For construction of the smooth kernel we have used the quartic kernel

$$K(u) = \tfrac{15}{16}(1-u^2)^2 I(|u| \le 1).$$

The integral of the test statistic T_n has been numerically approximated. The bootstrap resampling was performed $B = 100$ times for each sample, that is, the Monte Carlo approximation to $\mathscr{L}^*(T^{*,W})$ has been performed by B repetitions of the Wild Bootstrap algorithm. In order to study the distribution of T_n the whole sampling mechanism procedure was carried out $M = 1000$ times.

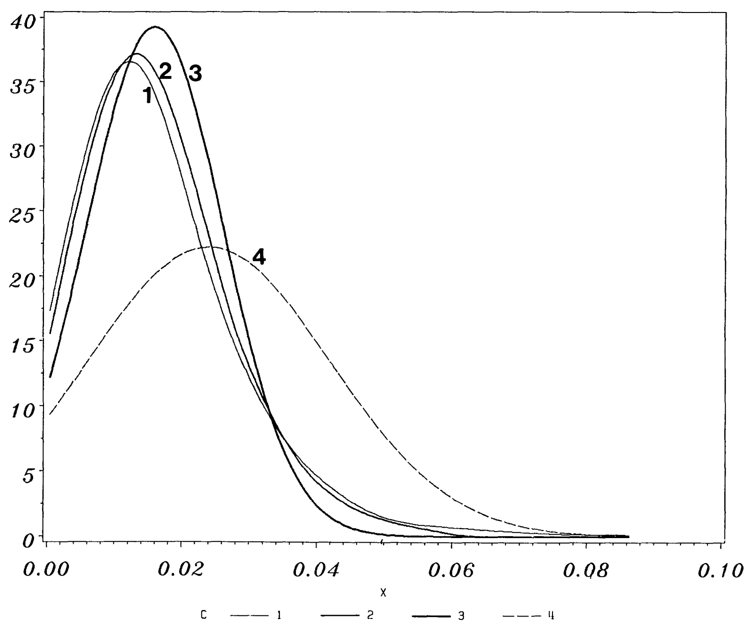

FIG. 1. *Monte Carlo density of T_n, bootstrap density estimate and normal density. The thin line 1 is the Monte Carlo density of T_n; line 2 is the kernel density of T_n^* from ONE bootstrap sample. Line 3 is the normal density with asymptotic mean and variance, given in Proposition 1; line 4 is the normal density approximation with estimated mean and variance. The parametric model consists of constant functions.*

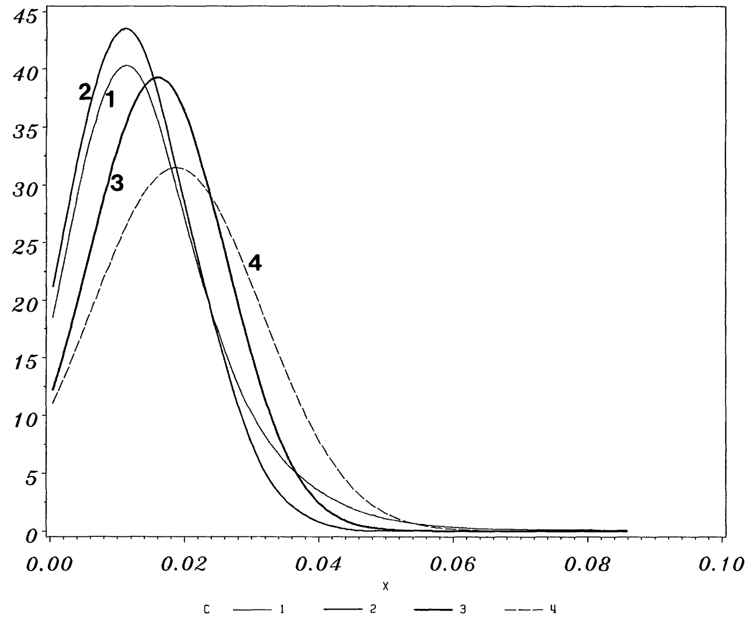

FIG. 2. *Monte Carlo density of T_n, bootstrap density estimate and normal density. The thin line* 1 *is the Monte Carlo density of T_n; line* 2 *is the kernel density of T_n^* from ONE bootstrap sample. Line* 3 *is the normal density with asymptotic mean and variance, given in Proposition* 1; *line* 4 *is the normal density approximation with estimated mean and variance. The parametric model consists of linear functions.*

At first we consider the parametric model of polynomials of degree k, $k = 0, 1, 2, 3$. The true regression curve $m(\cdot)$ is in this model class for each k. For the kernel estimator the bandwidth $h = 0.2$ has been chosen. In Figures 1–4 we present four curves for each k.

The thin line 1 denotes the Monte Carlo kernel estimate of the density of T_n from the M runs. The medium thin line 2 is the kernel density of one bootstrap sample out of the M runs (taken at random). The thick line 3 corresponds to the normal theory density as given in Proposition 1 based on the true b_h and V. The dashed line 4 finally shows the normal theory density based on estimated b_h and V. The quantities b_h and V have been estimated by the so-called "plug-in" method, that is, consistent estimators $\hat{f}_h(\cdot)$ and $\hat{\sigma}^2(\cdot)$ (also based on the kernel technique) have been used. The stability (under variation of the smoothing parameter) of these estimators was satisfactory and did not affect Figures 1–4.

In all four cases the wild bootstrap estimates the distribution of the T_n distance quite well. The normal approximation with estimated b_h and V is totally misleading. The normal densities have considerable mass on the negative axis. The inaccuracy of the normal approximations increases with the dimension of the parametric model. To study the power of our bootstrap test

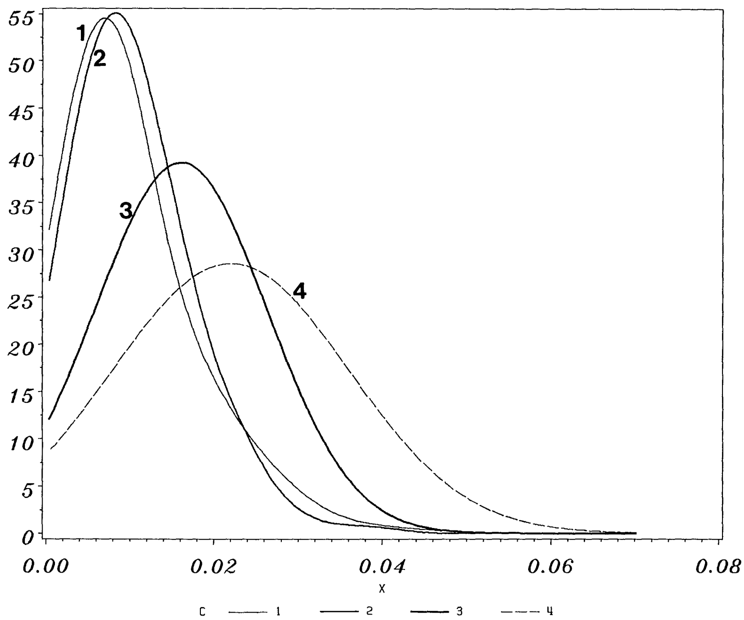

FIG. 3. *Monte Carlo density of* T_n, *bootstrap density estimate and normal density. The thin line* 1 *is the Monte Carlo density of* T_n; *line* 2 *is the kernel density of* T_n^* *from ONE bootstrap sample. Line* 3 *is the normal density with asymptotic mean and variance, given in Proposition* 1; *line* 4 *is the normal density approximation with estimated mean and variance. The parametric model consists of quadratic polynomials.*

we have chosen the parametric model,

$$m_\theta(x) = \theta_1 + \theta_2 x + \theta_3 x^2$$

and for different c the regression function

$$m(x) = 2x - x^2 + c\left(x - \tfrac{1}{4}\right)\left(x - \tfrac{1}{2}\right)\left(x - \tfrac{3}{4}\right).$$

Monte Carlo estimates of the power are summarized in Table 1 for different c and different bandwidth h.

The bandwidth has an influence on the level. We have on purpose selected a range of bandwidth wider than the fluctuation of crossvalidated bandwidth which was on average 0.23. For bandwidths around this value the level was held at 0.05; the more extreme smoothing parameters led to an under- and over-estimation of the level respectively.

In case of using a data adaptative bandwidth $\hat{h}$ the randomness of $\hat{h}$ might also affect the level. To capture this in the bootstrap resampling one could also use a data adaptive bandwidth $\hat{h}^*$, based on the bootstrap sample, in every bootstrap loop. For a discussion of this procedure in a related context see Chapter 3 in Mammen (1992). In our simulations the bandwidth h is fixed and nonrandom.

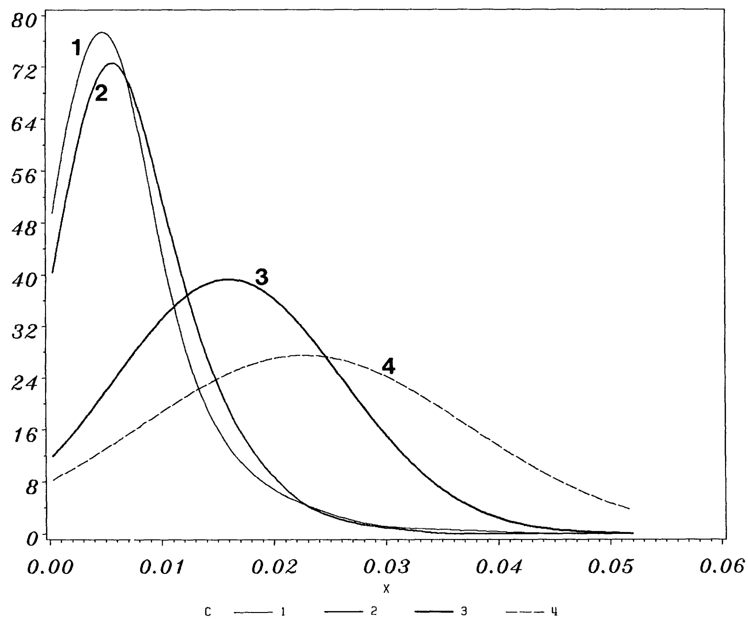

FIG. 4. *Monte Carlo density of T_n, bootstrap density estimate and normal density. The thin line 1 is the Monte Carlo density of T_n; line 2 is the kernel density of T_n^* from ONE bootstrap sample. Line 3 is the normal density with asymptotic mean and variance, given in Proposition 1; line 4 is the normal density approximation with estimated mean and variance. The parametric model consists of cubic polynomials.*

Figure 5 shows a linear fit and a working fit and a nonparametric smoothing estimate for the Engel curve for food as a function of total expenditure. The Engel curve is the mean expenditure curve for a certain good. Theoretical economists and econometricians are interested in the form of this regression curve since this form has consequences on theoretical and social questions, see Engel (1895). For the particularly chosen parametric form we refer to Leser (1963). The data came from the *Family Expenditure Survey* (1968–1983). The data used for this figure was from 1969; the X and the Y data have been rescaled by the mean of X. The quartic kernel has been used with a bandwidth of $h = 0.2$. The bootstrap test rejected the linear regression model for all

TABLE 1
Monte Carlo estimates of the power for the regression function $m(x) = 2x - x^2 + c(x - 1/4)(x - 1/2)(x - 3/4)$. The level has been chosen to be 0.05

h, c	0.0	0.5	1.0	2.0
0.10	0.105	0.157	0.325	0.784
0.20	0.054	0.120	0.252	0.795
0.25	0.053	0.099	0.263	0.765
0.30	0.039	0.078	0.225	0.714

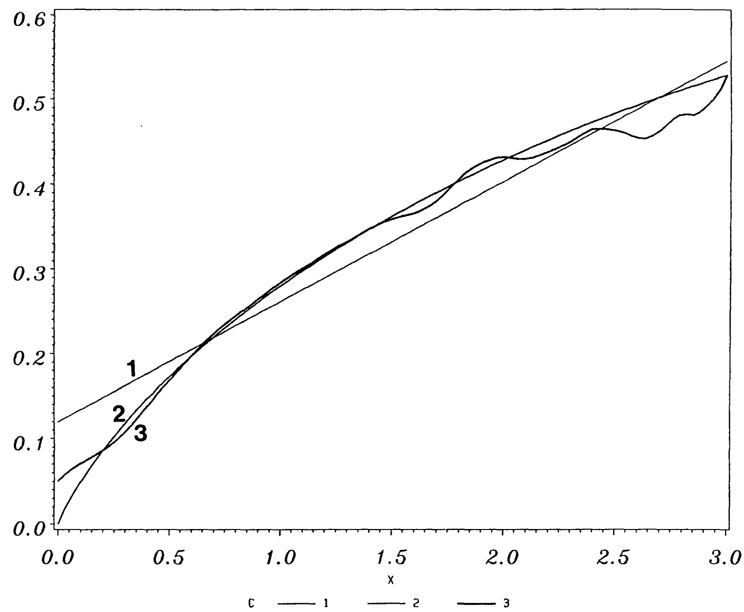

FIG. 5. *Working, linear and kernel smoother fit for a food expenditure Engel curve. The linear fit has label* 1, *the Working curve has label* 2 *and the nonparametric kernel smoother has label* 3.

considered bandwidths for both variables. For food the Working curve has been rejected only for some small bandwidths. This is of course due to the fact of an inflated value of the statistic T_n. The crossvalidated bandwidth for this data set lies at $h \approx 0.2$. For a picture of the crossvalidation function see Härdle [(1990), Section 5]. A summary of the bootstrap estimates of the observed critical values is given in Table 2.

TABLE 2
Observed critical values for two parametric fits for the Family Expenditure data set. The number of bootstrap simulations is 100

	Working	**Working**	**Linear**	**Linear**
h	**Fuel**	**Food**	**Fuel**	**Food**
0.05	0.0	0.0	0.0	0.0
0.10	0.0	0.0	0.0	0.0
0.15	0.02	0.08	0.0	0.0
0.20	0.05	0.38	0.0	0.0
0.25	0.08	0.57	0.0	0.0
0.30	0.06	0.62	0.0	0.0
0.35	0.05	0.55	0.0	0.0
0.40	0.06	0.55	0.0	0.0
0.45	0.07	0.54	0.0	0.0
0.50	0.08	0.51	0.0	0.0

7. Proofs. W.l.o.g. we will give the proofs only for $d = 1$ and $\pi(x) \equiv 1$.

PROOF OF PROPOSITION 1. First note that

$$\begin{aligned} \hat{f}_h(x) &= \frac{1}{n} \sum_{i=1}^{n} K_h(X_i - x) \\ &= f(x) + O_p\left(n^{-2/5}\sqrt{\log n}\right) \qquad \text{(uniformly in } x\text{)}, \end{aligned} \tag{7.1}$$

[see Collomb and Härdle (1986)] and that

$$\hat{m}_h(x) = m(x) + O_p\left(n^{-2/5}\sqrt{\log n}\right) \qquad \text{(uniformly in } x\text{)}, \tag{7.2}$$

[see Mack and Silverman (1982)] for $d = 1$. For $d > 1$ one shows for every $\eta > 0$ that

$$\hat{m}_h(x) = m(x) + O_p(n^{-2/(4+d)+\eta}) \qquad \text{(uniformly in } x\text{)}.$$

This can be proved as in Härdle [(1990), Section 4] calculating the moments of $\hat{m}_h \hat{f}_h$ using (A5) and the Lipschitz continuity of the kernel.

First we show for $q > 0$,

$$P\left(n^{-1} \sum_{n=1}^{n} K_h(X_i - x)(Y_i - m(X_i)) \geq C n^{-2/(4+d)}\sqrt{\log n}\right) = O(n^{-q})$$

for C large enough. This follows by the following simple application of the Markov inequality. Choose $k > 2 + d/2$. Then for a constant $c > 0$ it holds with $A_n = n^{2/(4+d)-1}\sqrt{\log n}\, K_h(X_1 - x)\varepsilon_1$,

$$\begin{aligned} &P\left(n^{-1} \sum_{i=1}^{n} K_h(X_i - x)(Y_i - m(X_i)) \geq C n^{-2/(4+d)}\sqrt{\log n}\right) \\ &\quad \leq \left\{E \exp\left(\left[n^{2/(4+d)}\sqrt{\log n}\right]\left[n^{-1}K_h(X_1 - x)\varepsilon_i\right]\right)\right\}^n \exp(-C(\log n)) \\ &\quad \leq \left[1 + E[A_n]^2 + \cdots + E[A_n]^k[1 + \exp(A_n)]\right]^n n^C \\ &\quad \leq [1 + c(\log n)/n + o((\log n)/n)]^n n^{-C} \\ &\quad \leq \exp(c \log n) + o(\log n) n^{-C} \\ &\quad \leq n^{c - C + o(1)} = O(n^{-q}). \end{aligned}$$

Similarly one can treat $n^{-1}\sum_{i=1}^{n} K_h(X_i - x)(m(X_i) - Y_i)$, $\hat{f}_h(x) - f(x)$ and $f(x) - \hat{f}_h(x)$. Now note that $\hat{f}_h(x)$, f and $\partial/\partial x[n^{-1}\sum_{i=1}^{n} K_h(X_i - x)m(X_i)]$ are bounded by deterministic constants with polynomial growth and that with (A5)

$$\sup_x \left[\partial/\partial x\left[n^{-1} \sum_{i=1}^{n} K_h(X_i - x) Y_i\right]\right] = O_p(h^{-2} \log n).$$

Equations (7.1) and (7.2) give

$$T_n = n\sqrt{h}\int_0^1 (\hat{m}_h(x) - \mathscr{K}_{h,n} m_{\hat{\theta}}(x))^2\, dx$$

$$= n\sqrt{h}\int_0^1 (\hat{m}_h(x) - \mathscr{K}_{h,n} m_{\hat{\theta}}(x))^2 \left(\frac{\hat{f}_h(x)}{f(x)}\right)^2 dx + o_p(1)$$

$$= n\sqrt{h}\int_0^1 \frac{((1/n)\Sigma_{i=1}^n K_h(X_i - x)(m(X_i) + \varepsilon_i - m_{\hat{\theta}}(X_i)))^2}{f^2(x)}\, dx + o_p(1).$$

Now apply (P1) and $m(\cdot) = m_{\theta_0}(\cdot) + n^{-1/2}h^{1/4}\Delta_n(\cdot)$: Then one gets

$$T_n = n\sqrt{h}\int_0^1 (U_{n,1}(x) + U_{n,2}(x) + U_{n,3}(x))^2\, dx + o_p(1),$$

where

$$U_{n,1}(x) = \frac{(1/n)\Sigma_{i=1}^n K_h(X_i - x) n^{-1/2} h^{-1/4}\Delta_n(X_i)}{f(x)},$$

$$U_{n,2}(x) = \frac{(1/n)\Sigma_{i=1}^n K_h(X_i - x)\varepsilon_i}{f(x)},$$

$$U_{n,3}(x) = \frac{-(1/n)\Sigma_{i=1}^n K_h(X_i - x)(1/n)\Sigma_{j=1}^n g(X_i)^T h(X_j)\varepsilon_j}{f(x)}.$$

By straightforward calculations one gets

$$En\sqrt{h}\int_0^1 U_{n,3}^2(x)\, dx \to 0,$$

$$En^2 h\left[\int_0^1 U_{n,i}(x) U_{n,j}(x)\, dx\right]^2 \to 0$$

for $1 \le i < j \le 3$.

This implies

$$T_n = n\sqrt{h}\int_0^1 U_{n,1}^2(x) + U_{n,2}^2(x)\, dx + o_p(1)$$

$$= T_{n,1} + T_{n,2} + T_{n,3} + o_p(1),$$

where

$$T_{n,1} = n\sqrt{h}\int_0^1 [U_{n,1}^2(x)]\, dx$$

$$T_{n,2} = \frac{\sqrt{h}}{n}\int_0^1 \frac{\Sigma_{i=1}^n K_h(X_i - x)^2 \varepsilon_i^2}{f^2(x)}\, dx$$

$$T_{n,3} = \frac{\sqrt{h}}{h}\int_0^1 \frac{\Sigma_{i\neq j} K_h(X_i - x) K_h(X_j - x)\varepsilon_i\varepsilon_j}{f^2(x)}\, dx.$$

We will show

$$T_{n,1} = \int_0^1 \mathscr{K}_h \Delta_n(x)^2 \, dx + o_p(1), \tag{7.3}$$

$$T_{n,2} = b_h + o_p(1), \tag{7.4}$$

$$\mathscr{L}(T_{n,3}) \quad \Rightarrow \quad N(0, V) \qquad \text{(weakly)}. \tag{7.5}$$

But (7.3)–(7.5) entail the statement of Proposition 1. It remains to show (7.3)–(7.5).

PROOF OF (7.3). Arguing as above one sees that

$$\begin{aligned} T_{n,1} &= n\sqrt{h} \int_0^1 \frac{\left((1/n)\sum_{i=1}^n K_h(X_i - x) n^{-1/2} h^{-1/4} \Delta_n(X_i)\right)^2}{f^2(x)} \, dx \\ &= \int_0^1 \frac{\left((1/n)\sum_{i=1}^n K_h(X_i - x) \Delta_n(X_i)\right)^2}{\hat{f}^2(x)} \, dx + o_p(1) \\ &= \int_0^1 \mathscr{K}_h \Delta_n(x)^2 \, dx + o_p(1). \end{aligned}$$

PROOF OF (7.4). First note

$$\begin{aligned} ET_{n,2} &= E\sqrt{h} \int_0^1 \frac{K_h(X_1 - x)^2}{f^2(x)} \sigma^2(x) \, dx \\ &= \int_0^1 \int_0^1 \sqrt{h} \, \frac{K_h(u - x)^2}{f^2(x)} f(u) \sigma^2(x) \, dx \\ &= b_h + o(1). \end{aligned}$$

Because of $\operatorname{var}(T_{n,2} | X_1, \dots, X_n) = O_p(h^{-3} n^{-1}) = o_p(1)$ this implies (7.4).

PROOF OF (7.5). Put

$$W_{ijn} = \begin{cases} \sqrt{h}\, n \displaystyle\int_0^1 \frac{K_h(X_i - x) K_h(X_j - x)}{f^2(x)} \, dx \, \varepsilon_i \varepsilon_j, & \text{if } i \neq j, \\ 0, & \text{otherwise}. \end{cases}$$

Then

$$T_{n,3} = \sum_{i,j} W_{ijn}.$$

According to Theorem 2.1 in de Jong (1987) for (7.5) it suffices to prove

$$(7.6)\qquad \operatorname{var}(T_{n,3}) \to V,$$

$$(7.7)\qquad \max_{1\le i\le n} \sum_{j=1}^{n} \operatorname{var}(W_{ijn}) \Big/ \operatorname{var}(T_{n,3}) \to 0,$$

$$(7.8)\qquad ET_{n,3}^4/(\operatorname{var}(T_{n,3}))^2 \to 3.$$

The proof of (7.6) is straightforward. Statement (7.7) follows from

$$\operatorname{var}(W_{ijn}) = O\left(\frac{1}{n^2h}\right).$$

For the proof of (7.8) note that

$$\begin{aligned} ET_{n,3}^4 &= 12\sum\nolimits^{\neq} EW_{ijn}^2 W_{kln}^2 + 8\sum\nolimits^{\neq} EW_{ijn}^4 + 48\sum\nolimits^{\neq}_{i,j,k,l} EW_{ijn}W_{jkn}W_{kln}W_{lin} \\ &\quad + 192\sum\nolimits^{\neq} EW_{ijn}W_{ikn}^2 W_{jkn} \\ &= 3\operatorname{var}(T_{n,3})^2 + o(1), \end{aligned}$$

(here $\Sigma^{\neq}$ denotes summation over only all pairwise different indices) because of

$$EW_{12n}^4 = O\left(\frac{1}{n^4h^2}\right) = o\left(\frac{1}{n^2}\right),$$

$$\begin{aligned} &EW_{12n}W_{23n}W_{34n}W_{41n} \\ &\quad = \frac{h^2}{n^4}\int \frac{K_h(u_1-x_1)K_h(u_2-x_1)K_h(u_2-x_2)\cdots K_h(u_1-x_4)}{f^2(x_1)\cdots f^2(x_4)} \\ &\qquad\qquad \times f(u_1)\cdots f(u_4)\,dx_1\cdots dx_4\,du_1\cdots du_4\cdot O(1) \\ &\quad = O\left(\frac{h^2}{n^4}\right)\int K_h(u_1-u_2)K_h(u_2-u_3)K_h(u_3-u_4) \\ &\qquad \times K_h(u_4-u_1)\,du_1\cdots du_4 \\ &\quad = O\left(\frac{h^2}{n^4}\right)\int (K_h^{(2)}(u))^2\,du = O\left(\frac{h}{n^4}\right) = o\left(\frac{1}{n^4}\right) \end{aligned}$$

and

$$\begin{aligned} EW_{12n}W_{23n}^2W_{31n} &= O\left(\frac{h^2}{n^4}\right)\int K_h(u)^2 K_h^{(2)}(u)\,du \\ &= O\left(\frac{1}{n^4}\right) = o\left(\frac{1}{n^3}\right), \end{aligned}$$

$$\operatorname{var}(T_{n,3}) = 2\sum\nolimits^{\neq} EW_{ijn}^2. \qquad \square$$

We will give only indications of the proofs of Proposition 2 and Theorems 1 and 2.

PROOF OF PROPOSITION 2. First note that

$$nh^{1/2}\int(\hat{m}_h - m_{\hat{\theta}})^2\pi = nh^{1/2}\int(V_{n,1}(x) + \cdots + V_{n,4}(x))^2\pi(x)\,dx,$$

where

$$\begin{aligned} V_{n,1}(x) &= \hat{m}_h(x) - \mathscr{K}_{h,n}m(x),\\ V_{n,2}(x) &= \mathscr{K}_{h,n}m(x) - m(x),\\ V_{n,3} &= m(x) - m_\theta(x),\\ V_{n,4} &= m_\theta(x) - m_{\hat{\theta}}(x). \end{aligned}$$

Now $nh^{1/2}\int V_{n,1}(x)^2\pi(x)\,dx$ can be expanded as $nh^{1/2}\int U_{n,1}(x)^2\pi(x)\,dx$ in the proof of Proposition 1. The other terms can be treated similarly. For instance, one gets

$$\begin{aligned} 2nh^{1/2}\int V_{n,2}(x)V_{n,3}(x)\pi(x)\,dx &= nh^{1/2}c_K c_n \int h^2\rho(x)\Delta(x)\pi(x)\,dx + o_p(1)\\ &= n^{1/2}h^{5/2}c_K\int\rho(x)\Delta(x)\pi(x)\,dx + o_p(1). \quad \square \end{aligned}$$

PROOF OF THEOREM 1. As in the proof of Proposition 1 (but by a little bit finer arguments) one shows first for $T^* = T^{*,N}$ (or $T^* = T^{*,A}$) that (note that $\Delta_n = 0$)

$$T^* = \frac{\sqrt{h}}{n}\sum_{1\le i,j\le n}\frac{K^{(2)}(X_i^* - X_j^*)\eta_i^*\eta_j^*}{f(X_i^*)f(X_j^*)} + \Gamma$$

for a random variable Γ with

$$E^*\Gamma^2 = o_p(1).$$

Here E^* denotes the conditional expectation given $\{(X_i, Y_i)\}_{i=1}^n$. $\{(X_i^*, \eta_i^*)\}$ are drawn (with replacement) out of the set $\{(X_i, \eta_i)\}$ where

$$\eta_i = \varepsilon_i - \frac{1}{n}\sum_{j=1}^n g^T(X_i)h(X_j)\varepsilon_j \qquad \text{if } T^* = T^{*,N},$$

$$\eta_i = \varepsilon_i + (m - \mathscr{K}_{h,n}m)(X_i) - \frac{1}{n}\sum_{j=1}^n \frac{K_h(X_i - X_j)\varepsilon_j}{\hat{f}_h(X_j)} \qquad \text{if } T^* = T^{*,A}.$$

Define

$$A_{ij} = \frac{K_h(X_i^* - X_j^*)\eta_i^*\eta_j^*}{f(X_i^*)f(X_j^*)}.$$

Put $a = E^*(A_{ij})$ for $i \neq j$.

Then one gets with $a = O_p(n^{-1}h^{-1})$:

$$\begin{aligned}
\operatorname{var}^*(T^*) &= \frac{h}{n^2}\operatorname{var}^*\Bigl(\sum_{i,j} A_{ij}\Bigr) + o_p(1)\\
&= \frac{h}{n^2}\operatorname{var}^*\Bigl(\sum_{i\neq j} A_{ij}\Bigr) + o_p(1)\\
&= \frac{h}{n^2}E^*\Bigl(\sum_{i\neq j} A_{ij} - a\Bigr)^2 + o_p(1)\\
&= \frac{h}{n^2}E^*2\sum_{i\neq j}(A_{ij}-a)^2\\
&\quad + \frac{h}{n^2}E^*4\sum_{i\neq j\neq k\neq i}(A_{ij}-a)(A_{jk}-a) + o_p(1)\\
&= 2hE^*A_{12}^2 + 4hnE^*A_{12}A_{23} + o_p(1)\\
&= 2\frac{h}{n^2}\sum_{i,j}\frac{K_h(X_i-X_j)^2}{f^2(X_i)f^2(X_j)}\eta_i^2\eta_j^2\\
&\quad + 4\frac{h}{n^2}\sum_{i,j,k}\frac{K_h(X_i-X_j)K_h(X_j-X_k)}{f(X_i)f^2(X_j)f(X_k)}\eta_i\eta_j^2\eta_k\\
&\quad + o_p(1).
\end{aligned}$$

This gives for $T^* = T^{*,N}$ by straightforward calculations

$$\operatorname{var}^*(T^{*,N}) = 6\frac{h}{n^2}\sum_{i,j}\frac{K_h(X_i-X_j)^2}{f^2(X_i)f^2(X_j)}\varepsilon_i^2\varepsilon_j^2 + o_p(1).$$

But the right-hand term converges in probability to its expectation. This proves the first statement of Theorem 1. The second statement follows by a very lengthy evaluation of the above approximation of $\operatorname{var}^*(T^{*,A})$. □

Proof of Theorem 2. The proof goes along the lines of Proposition 1. Especially (A5) entails $\sup_i \varepsilon_i^2 = O_p(\log n)$ and $E|\varepsilon_i|^8 <$ const. (uniformly in i and n). This can be used to prove the two conditions of the theorem of de Jong (1987). □

Acknowledgments. We would like to express our gratitude to the Deutsche Forschungsgemeinschaft (SFB 303, SFB 123), CentER and CORE for financial support. We also thank B. Turlach for assistance in programming and are grateful to an Associate Editor and a referee for number of remarks and suggestions.

REFERENCES

AZZALINI, A., BOWMAN, A. W. and HÄRDLE, W. (1989). On the use of nonparametric regression for model checking. *Biometrika* **76** 1–11.

BERAN, R. (1986). Comment on "Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14** 1295–1298.

CARROLL, R. J. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10** 1224–1233.

CAO-ABAD, R. and GONZALEZ-MANTEIGA, W. (1990). On bootstrapping regression curves. Unpublished manuscript.

CLEVELAND, W. S. and DEVLIN, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* **83** 596–610.

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

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

COX, D., KOH, E., WAHBA, G. and YANDELL, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16** 113–119.

COX, D. and KOH, E. (1989). A smoothing spline based test of model adequacy in polynomial regression. *Ann. Inst. Statist. Math.* **41** 383–400.

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277.

DUCPETIAUX, E. (1855). *Budgets économiques des classes ouvrières en Belgique*. Brussels.

DURBIN, J. and KNOTT, M. (1972). Components of Cramer–von Mises statistics I. *J. Roy. Statist. Soc. Ser. B* **34** 290–307.

ENGEL, E. (1895). Die Produktions- und Consumptionsverhältnisse des Königsreichs Sachsen. Abdruck in *Bulletin de l'Institut International de Statistique* **9** 1–54.

EUBANK, R. and SPIEGELMAN, C. (1990). Testing the goodness-of-fit of linear models via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85** 387–392.

FAMILY EXPENDITURE SURVEY, ANNUAL BASE TAPES (1968–1983). Department of Employment, Statistics Division, Her Majesty's Stationery Office, London 1968–1983. (The data utilized in this paper were made available by the ESRC Data Archive at the University of Essex.)

GASSER, T., KÖHLER, W., MÜLLER, H. G., LAGO, R., MOLINARI, L. and PRADER, A. (1985). Human height growth: Correlational and multivariate structure of velocity and acceleration. *Annals of Human Biology* **12** 501–515.

HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Econometric Society Monograph Series **19**, Cambridge Univ. Press.

HÄRDLE, W. and MARRON, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63–89.

HILDRETH, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49** 598–619.

HILDENBRAND, K. and HILDENBRAND, W. (1986). 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.) North-Holland, Amsterdam.

KONAKOV, V. D. and PITERBARG, V. I. (1984). On the convergence rate of maximal deviation distribution for kernel regression estimators. *J. Multivariate Anal.* **15** 279–294.

LARICCIA, V. N. (1991). Smooth goodness-of-fit tests: A quantile function approach. *J. Amer. Statist. Assoc.* **86** 427–431.

LE CESSIE, S. and VAN HOUWELINGEN, J. C. (1991). A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics* **47** 1267–1282.

LESER, C. E. (1963). Forms of Engel functions. *Econometrica* **31** 694–703.

LIU, R. (1988). Bootstrap procedures under some non i.i.d. models. *Ann Statist.* **16** 1696–1708.

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61** 405–415.

MAMMEN, E. (1992). When Does Bootstrap Work: Asymptotic Results and Simulations. *Lecture Notes in Statist.* **77** Springer, Berlin.

MAMMEN, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21** 255–285.

MILBRODT, H. and STRASSER, H. (1990). On the asymptotic power of the two-sided Kolmogorov–Smirnov test. *J. Statist. Plann. Inference* **26** 1–23.

MUNSON, P. J. and JERNIGAN, R. W. (1989). A cubic spline extension of the Durbin–Watson test. *Biometrika* **76** 39–47.

NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* **10** 186–190.

NEUHAUS, G. (1986). A class of quadratic goodness of fit tests. Preprint.

NEUHAUS, G. (1988). Addendum to: Local asymptotics for linear rank statistics with estimated score functions. *Annals of Statist.* **16** 1342–1343.

PREECE, M. A. and BAINES, M. J. (1978). A new family of mathematical models describing the human growth curve. *Annals of Human Biology* **5** 1–24.

STANISWALIS, J. G. and SEVERINI, T. A. (1991). Diagnostics for assessing regression models. *J. Amer. Statist. Assoc.* **86** 684–692.

TUKEY, J. W. (1961). Curves as parameters and touch estimation. *Proc. Fourth Berkeley Symp. Math Statist. Probab.* 681–694. Univ. Califonia Press, Berkeley.

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

WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1350.

YANAGIMOTO, T. and YANAGIMOTO, M. (1987). The use of marginal likelihood for a diagnostic test for the goodness of fit of the simple linear regression model. *Technometrics* **29** 95–101.

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR STATISTIK UND ÖKONOMETRIE
WIRTSCHAFTSWISSENSCHAFTLICHE FAKULTÄT
SPANDAUER STRASSE 1,
D-10178 BERLIN
GERMANY

HUMBOLDT-UNIVERSITÄT ZU BERLIN
FB MATHEMATIK
INSTITUT FÜR STOCHASTIK
UNTER DEN LINDEN 6
D-10099 BERLIN
GERMANY