A Power Comparison Between Nonparametric Regression Tests

Chunming Zhang
Department of Statistics
University of Wisconsin
1210 West Dayton Street
WI 53706-1685, USA
cmzhang@stat.wisc.edu

Holger Dette
Fakultät für Mathematik
Ruhr-Universität Bochum
Universitätsstraße 150
44780 Bochum, Germany
holger.dette@ruhr-uni-bochum.de

August 11, 2003

Abstract

In this paper, we consider three major types of nonparametric regression tests that are based on kernel and local polynomial smoothing techniques. Their asymptotic power comparisons are established systematically under the fixed and contiguous alternatives, and are also illustrated through non-asymptotic investigations and finite-sample simulation studies.

MSC: primary 62G10; secondary 62G05, 62G20

Keywords: Goodness-of-fit; Local alternative; Local polynomial regression; Power; Smoothing parameter

Running title: Comparing Nonparametric Regression Tests

1 Introduction

An important application of nonparametric estimation technique is to checking the agreement between a proposed parametric model with the observed data. Suppose we have a sequence of independent observations $\{(X_i, Y_i), i = 1, \dots, n\}$ coming from a population (X, Y), in which the unknown regression function $m(x) = E(Y|X = x)$ is assumed to be smooth. To justify the use of a parametric model, a specification test on the functional form of the regression is needed. Given a parametric function $g(x; \boldsymbol{\theta})$, the null and alternative hypotheses can be described as

$$H_0: \qquad P\{m(X) = g(X; \boldsymbol{\theta}_0)\} = 1, \text{ for some } \boldsymbol{\theta}_0 \in \Theta, \tag{1.1}$$

$$H_1: \qquad P\{m(X) = g(X; \boldsymbol{\theta})\} < 1, \text{ for any } \boldsymbol{\theta} \in \Theta, \tag{1.2}$$

where Θ denotes the parameter space.

In applications it is often the case that many nonparametric regression tests, based on a similar type of nonparametric estimation technique, are potentially applicable to the above model checking problem. A question that arises naturally in the use of the kernel and local polynomial smoothing methods, for example, is, for the same smoothing parameter, which test statistic will be more powerful? While a careful study of such question could lead to a better understanding and more efficient use of nonparametric tests, this issue has rarely been addressed in the literature and thus a comparison is clearly needed. In this paper, the relative performance of three major types of nonparametric tests will be compared from an asymptotic point of view and by means of a simulation study.

To facilitate presentation of these tests, we first introduce some notations. The pth degree local polynomial regression estimates, $\widehat{\boldsymbol{\beta}}(x) = (\widehat{\beta}_0(x), \dots, \widehat{\beta}_p(x))^T$, are defined to be the minimizer of

$$\sum_{i=1}^{n} \{Y_i - \beta_0 - (X_i - x)\beta_1 - \cdots - (X_i - x)\beta_p\}^2 K_h(X_i - x)$$

with respect to $(\beta_0, \dots, \beta_p)^T$, where $h > 0$ is known as the bandwidth, and K is called a kernel function, usually taken to be a symmetric probability density function, with $K_h(\cdot) = K(\cdot/h)/h$. Denote by $\widehat{m}_h(x) = \widehat{\beta}_0(x)$ the local polynomial estimate of $m(x)$; the kernel and local linear regression estimates correspond to those of degrees $p = 0$ and $p = 1$ respectively. Let $\widehat{\boldsymbol{\theta}}$ denote a consistent estimate of $\boldsymbol{\theta}_0$, and $\widehat{g}_h(x; \widehat{\boldsymbol{\theta}})$ a local polynomial regression estimate of $g(x; \widehat{\boldsymbol{\theta}})$, via locally regressing $g(X_i; \widehat{\boldsymbol{\theta}})$ on X_i, $i = 1, \dots, n$. Define $\widehat{f}_h$ the kernel density estimate of the probability density function, f, of the regressor variable X, according to $\widehat{f}_h(x) = n^{-1} \sum_{i=1}^{n} K_h(X_i - x)$. Refer to Silverman (1986) and Fan and Gijbels (1996) for more details of nonparametric smoothing methods. In the following, we use π_j, $j = 1, 2, 3$, to represent three positive weight functions.

Operationally, the Härdle and Mammen's test (1993), T_{1n}, to be defined in (2.1), is based on

$$\int \{\widehat{m}_h(x) - \widehat{g}_h(x; \widehat{\boldsymbol{\theta}})\}^2 \pi_1(x) dx, \tag{1.3}$$

a Riemannian approximation of which was studied simultaneously by González-Manteiga and Cao (1993). The Zheng's test (1996) employs the statistic,

$$T_{2n} = n^{-1} \sum\sum_{1 \le i \ne j \le n} K_h(X_i - X_j)\{Y_i - g(X_i; \widehat{\boldsymbol{\theta}})\}\{Y_j - g(X_j; \widehat{\boldsymbol{\theta}})\}\pi_2(X_i), \tag{1.4}$$

whereas Dette (1999) and Fan, Zhang & Zhang (2001) build their tests from the statistic,

$$T_{3n} = \sum_{i=1}^{n} \{Y_i - g(X_i; \widehat{\boldsymbol{\theta}})\}^2 \pi_3(X_i) - \sum_{i=1}^{n} \{Y_i - \widehat{m}_h(X_i)\}^2 \pi_3(X_i). \tag{1.5}$$

(For a modified version of T_{3n}, see also Azzalini and Bowman (1993).) These structurally distinct test procedures are necessarily proposed from different criterions that aim at detecting the departure of the alternative regression from the null hypothesis. Section 2 offers a preliminary comparison of these criterions. The asymptotic power comparison is made in Section 3, and is evaluated through finite-sample simulation studies in Section 4.

2 Preliminary comparison

To gain insight into the structures of the three tests, we start with some familiar properties of conditional expectation. Write $\varepsilon_0 = Y - g(X;\boldsymbol{\theta}_0)$. Then $E(\varepsilon_0|X) = m(X) - g(X;\boldsymbol{\theta}_0)$, which equals zero (with probability one) if and only if the null hypothesis holds. Since $\boldsymbol{\theta}_0$ is usually unknown, a sample analogue of ε_0 is given by $Y - g(X;\widehat{\boldsymbol{\theta}})$, and thus $E(\varepsilon_0|X)$ can be estimated by $\widehat{m}_h(X) - \widehat{g}_h(X;\widehat{\boldsymbol{\theta}})$, which serves as a smoothed version of $Y - g(X;\widehat{\boldsymbol{\theta}})$.

We now examine the criterion under which each of the three tests is constructed. We obtain the following observations.

(a) The test of Härdle and Mammen (1993) is based on $C_1 = E^2(\varepsilon_0|X)\pi_1(X)$. Since $E(C_1) = 0$ iff $P\{E(\varepsilon_0|X) = 0\} = 1$ iff H_0 holds, the further apart $E(C_1)$ deviates from zero, the stronger the evidence is against the null hypothesis. Clearly, the sample analogue of $E(C_1)$ is formed by T_{1n}/n, where

$$T_{1n} = \sum_{i=1}^{n}\{\widehat{m}_h(X_i) - \widehat{g}_h(X_i;\widehat{\boldsymbol{\theta}})\}^2\pi_1(X_i); \tag{2.1}$$

see Gonzáles-Manteiga and Cao (1993). The L_2-version of (2.1) leads to (1.3).

(b) Analogously, the test of Zheng (1996) is based on $C_2 = \varepsilon_0 E(\varepsilon_0|X)f(X)\pi_2(X)$. Again, $E(C_2) = E\{E^2(\varepsilon_0|X)f(X)\pi_2(X)\} = 0$ if and only if H_0 holds. The sample version of $E(C_2)$ is

$$n^{-1}\sum_{i=1}^{n}\{Y_i - g(X_i;\widehat{\boldsymbol{\theta}})\}\{\widehat{m}_h(X_i) - \widehat{g}_h(X_i;\widehat{\boldsymbol{\theta}})\}\widehat{f}_h(X_i)\pi_2(X_i), \tag{2.2}$$

which was introduced by Gozalo and Linton (2001) in the context of testing additivity and coincides (up to an additive term) with the statistic T_{2n}/n in (1.4), when the kernel regression method is used.

(c) By a similar analysis, the test statistic T_{3n} is based on $C_3 = [\varepsilon_0^2 - \{\varepsilon_0 - E(\varepsilon_0|X)\}^2]\pi_3(X)$. It is easily seen that

$$E(C_3) = E\{[2\varepsilon_0 E(\varepsilon_0|X) - E^2(\varepsilon_0|X)]\pi_3(X)\} = E\{E^2(\varepsilon_0|X)\pi_3(X)\}, \tag{2.3}$$

which is identical to zero if and only if the null hypothesis H_0 holds. The sample version of $E(C_3)$ is

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}(\{Y_i - g(X_i;\widehat{\boldsymbol{\theta}})\}^2 - [\{Y_i - g(X_i;\widehat{\boldsymbol{\theta}})\} - \{\widehat{m}_h(X_i) - \widehat{g}_h(X_i;\widehat{\boldsymbol{\theta}})\}]^2)\pi_3(X_i) \\ = \; & n^{-1}\sum_{i=1}^{n}[\{Y_i - g(X_i;\widehat{\boldsymbol{\theta}})\}^2 - \{Y_i - \widehat{m}_h(X_i) - g(X_i;\widehat{\boldsymbol{\theta}}) + \widehat{g}_h(X_i;\widehat{\boldsymbol{\theta}})\}^2]\pi_3(X_i). \end{aligned} \tag{2.4}$$

Note that (2.4) exactly agrees with T_{3n}/n, when $g(X_i;\widehat{\boldsymbol{\theta}}) = \widehat{g}_h(X_i;\widehat{\boldsymbol{\theta}})$, for $i = 1,\dots,n$. This condition is met, if, for instance, $g(x;\boldsymbol{\theta})$ is a qth degree polynomial function of x, and a pth ($p \geq q$) degree local polynomial regression method is used to obtain estimates $\widehat{g}_h(X_i;\widehat{\boldsymbol{\theta}})$.

To confine our discussion to the relevant testing procedures themselves, we shall deal with a situation, in which $\pi_1 \equiv \pi_2 f \equiv \pi_3 \equiv c$ for a constant $c > 0$. This leads to $C_1 = c[E^2(\varepsilon_0|X)]$,

$C_2 = c[\varepsilon_0 E(\varepsilon_0|X)]$, and $C_3 = c[\varepsilon_0^2 - \{\varepsilon_0 - E(\varepsilon_0|X)\}^2] = c[2\varepsilon_0 E(\varepsilon_0|X) - E^2(\varepsilon_0|X)]$. It is then straightforward to check that

$$E(C_1) = E(C_2) = E(C_3) = c[E\{E^2(\varepsilon_0|X)\}]. \tag{2.5}$$

However, the variabilities of C_j, $j = 1, 2, 3$, are no longer identical, as shown in Lemma 1.

Lemma 1 *If* $\pi_1 \equiv \pi_2 f \equiv \pi_3 \equiv c$, *then* $\mathrm{var}(C_1) \leq \mathrm{var}(C_2) \leq \mathrm{var}(C_3)$.

Proof. From (2.5), we only need to verify that $E(C_1^2) \leq E(C_2^2) \leq E(C_3^2)$. These inequalities follow directly from the identities,

$$\begin{aligned}
E(C_1^2) &= c^2\, E\{E^4(\varepsilon_0|X)\} = c^2\, E\{E^2(\varepsilon_0|X) E^2(\varepsilon_0|X)\}, \\
E(C_2^2) &= c^2\, E\{\varepsilon_0^2 E^2(\varepsilon_0|X)\} = c^2\, E\{E(\varepsilon_0^2|X) E^2(\varepsilon_0|X)\}, \\
E(C_3^2) &= c^2\, E\{4\varepsilon_0^2 E^2(\varepsilon_0|X) - 4\varepsilon_0 E^3(\varepsilon_0|X) + E^4(\varepsilon_0|X)\} \\
&= c^2\, E\{4E(\varepsilon_0^2|X) E^2(\varepsilon_0|X) - 3E^4(\varepsilon_0|X)\} \\
&= c^2\, E[\{4E(\varepsilon_0^2|X) - 3E^2(\varepsilon_0|X)\} E^2(\varepsilon_0|X)],
\end{aligned}$$

and the inequality, $E^2(\varepsilon_0|X) \leq E(\varepsilon_0^2|X)$. ■

Roughly speaking, since T_{jn}/n mimics C_j, $j = 1, 2, 3$, we would expect from the discussions **(a)**–**(c)** that the preceding mean and variance relations between C_j carry over to those between T_{jn}. An intuitive implication of Lemma 1 is that T_{1n} will be more powerful than T_{2n}, which has power greater than T_{3n}. More rigorous comparison with respect to powers will be addressed in the next section.

3 Asymptotic comparison

The asymptotic distributions of the three tests (defined in (2.1), (1.4), and (1.5) respectively), under the null hypothesis, the fixed alternative, and the local alternative, will be presented in Theorems 1–4. They summarize and extend the research results in the aforementioned literature, and thus we omit the technical details to save space. In each case, the local linear method is used for estimating the functional form of the regression function.

3.1 Asymptotic null distribution

Denote by $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{\mathcal{P}}$ convergence in distribution and in probability, respectively. The following Condition (A) is a listing of assumptions:

(A1) K is a symmetric probability density function, having a compact support.

(A2) K is continuously differentiable.

(A3) f has compact support Ω.

(A4) f is continuously differentiable on Ω, and $f > 0$ on Ω.

(A5) m has compact support Ω.

(A6) m is twice continuously differentiable.
(A7) The functions $\sigma^2(x) = \mathrm{var}(Y|X = x)$, and $\pi_j(x)$, $j = 1, 2, 3$, are continuous on Ω.
(A8) $g(x; \boldsymbol{\theta})$ is a twice continuously differentiable real function on Θ for each x.
(A9) $\widehat{\boldsymbol{\theta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$.

We first obtain in Theorem 1 the asymptotic normal distributions of the three tests under the null hypothesis.

Theorem 1 *Assume that Condition (A) and the null hypothesis H_0 hold. Then as $n \to \infty$, $h \to 0$ and $nh \to \infty$,*

$$\frac{h^{1/2}\{T_{1n} - h^{-1}b_1 - (nh^2)^{-1}c_1\}}{\sigma_1} \xrightarrow{\mathcal{L}} N(0,1), \tag{3.1}$$

$$\frac{h^{1/2}T_{2n}}{\sigma_2} \xrightarrow{\mathcal{L}} N(0,1). \tag{3.2}$$

For T_{3n}, suppose either (a) $n \to \infty$, $h \to 0$, $nh \to \infty$, and $nh^{9/2} \to 0$, if H_0 denotes a nonlinear parametric model, or (b) $n \to \infty$, $h \to 0$, $nh \to \infty$, if H_0 denotes a linear parametric model, then

$$\frac{h^{1/2}\{T_{3n} - h^{-1}b_3 - (nh^2)^{-1}c_3\}}{\sigma_3} \xrightarrow{\mathcal{L}} N(0,1), \tag{3.3}$$

where

$$\begin{aligned}
&b_1 = K * K(0)\int \sigma^2(x)\pi_1(x)dx, \quad b_3 = (2K - K * K)(0)\int \sigma^2(x)\pi_3(x)dx,\\
&c_1 = K^2(0)E\{\sigma^2(X)\pi_1(X)/f^2(X)\}, \quad c_3 = -K^2(0)E\{\sigma^2(X)\pi_3(X)/f^2(X)\},\\
&\sigma_1^2 = 2\|K * K\|_2^2 \int \{\sigma^2(x)\}^2\pi_1^2(x)dx, \quad \sigma_2^2 = 2\|K\|_2^2 \int \{\sigma^2(x)\}^2\pi_2^2(x)f^2(x)dx,\\
&\sigma_3^2 = 2\|2K - K * K\|_2^2 \int \{\sigma^2(x)\}^2\pi_3^2(x)dx,
\end{aligned}$$

the symbol $*$ denotes the convolution operator, and $\|\cdot\|_2$ denotes the L_2 norm with respect to Lebesgue measure.

3.2 Asymptotic distribution under fixed alternatives

Theorem 2 below demonstrates that each of the three tests is consistent against all fixed deviations from the parametric null model.

Theorem 2 *Assume that Condition (A) and the fixed alternative H_1 hold. Then as $n \to \infty$, $h \to 0$ and $nh \to \infty$,*

$$n^{-1}T_{1n} \xrightarrow{\mathcal{P}} \Delta_1 = E[\{m(X) - g(X; \boldsymbol{\theta}_0)\}^2\pi_1(X)] > 0, \tag{3.4}$$

$$n^{-1}T_{2n} \xrightarrow{\mathcal{P}} \Delta_2 = E[\{m(X) - g(X; \boldsymbol{\theta}_0)\}^2\pi_2(X)f(X)] > 0, \tag{3.5}$$

$$n^{-1}T_{3n} \xrightarrow{\mathcal{P}} \Delta_3 = E[\{m(X) - g(X; \boldsymbol{\theta}_0)\}^2\pi_3(X)] > 0. \tag{3.6}$$

The results of Theorem 2 indicate that under the fixed alternative H_1,

$$\begin{aligned}
\frac{h^{1/2}\{T_{1n} - h^{-1}b_1 - (nh^2)^{-1}c_1\}}{\sigma_1} &= \frac{nh^{1/2}[\Delta_1\{1+o_P(1)\} - (nh)^{-1}b_1 - (nh)^{-2}c_1]}{\sigma_1} \xrightarrow{\mathcal{P}} +\infty, \\
\frac{h^{1/2}T_{2n}}{\sigma_2} &= \frac{nh^{1/2}[\Delta_2\{1+o_P(1)\}]}{\sigma_2} \xrightarrow{\mathcal{P}} +\infty, \\
\frac{h^{1/2}\{T_{3n} - h^{-1}b_3 - (nh^2)^{-1}c_3\}}{\sigma_3} &= \frac{nh^{1/2}[\Delta_3\{1+o_P(1)\} - (nh)^{-1}b_3 - (nh)^{-2}c_3]}{\sigma_3} \xrightarrow{\mathcal{P}} +\infty.
\end{aligned}$$

Hence, $nh^{1/2}$ stands for the common rate at which the standardized test statistics of T_{jn}, $j = 1, 2, 3$, diverge to infinity in probability. Furthermore, the proportionality constants, assuming $\pi_1 \equiv \pi_2 f \equiv \pi_3$, i.e., $\Delta_1 = \Delta_2 = \Delta_3$, depend inversely on the asymptotic variances σ_j; Section 3.4 contains numerical information of σ_j. Our next result gives the asymptotic distribution under a fixed alternative and is a refinement of Theorems 2 and 4. The proofs are similar to those in Dette (1999) and therefore omitted.

Theorem 3 *If the assumptions of Theorem 2 are satisfied, then*

$$\frac{1}{\sqrt{n}}(T_{jn} - h^{-1}b_j - n\Delta_j) \xrightarrow{\mathcal{L}} N(0, \tau_j^2), \quad j = 1, 2, 3,$$

where b_1, b_3 are defined in Theorem 1, $b_2 = 0$, $\Delta_1, \Delta_2, \Delta_3$ are given in Theorem 2 and the asymptotic variances are obtained as

$$\begin{aligned}
\tau_1^2 &= 4E[\sigma^2(X)\{(\Delta\pi_1)(X) - P(\Delta\pi_1)(X)\}^2] + \mathrm{var}[(\Delta\pi_1)^2(X)], && (3.7) \\
\tau_2^2 &= 4E[\sigma^2(X)\{(\Delta\pi_2 f)(X) - P(\Delta\pi_2 f)(X)\}^2] && (3.8) \\
&\quad +4\,\mathrm{var}[(\Delta\pi_2)(X)\{(\Delta\pi_2 f)(X) - P(\Delta\pi_2 f)(X)\}], \\
\tau_3^2 &= 4E[\sigma^2(X)(\Delta\pi_3)^2(X)] + \mathrm{var}[(\Delta\pi_3)^2(X)]. && (3.9)
\end{aligned}$$

Here

$$P(m)(X) = g(X; \boldsymbol{\theta}_0^*)$$

with $\boldsymbol{\theta}_0^* = argmin_{\boldsymbol{\theta}} E[\{m(X) - g(X; \boldsymbol{\theta})\}^2]$ is the parameter corresponding to the best L_2-approximation of the function m by functions of the form $g(\cdot; \boldsymbol{\theta})$, $\Delta = m - P(m)$ is the difference between m and the best approximation.

3.3 Asymptotic distribution under contiguous alternatives

Consider a sequence of local alternatives, defined by

$$H_{1n}: \ m(x) = g(x; \boldsymbol{\theta}_0) + \delta_n \cdot l(x), \tag{3.10}$$

where $\delta_n = (nh^{1/2})^{-1/2}$ and $l(x)$ is twice continuously differentiable on Ω. The following theorem gives the asymptotic distributions of the three tests under the local alternatives H_{1n}.

Theorem 4 *Assume that Condition (A) and the local alternatives H_{1n} hold. Then under the same conditions on n and h as given in Theorem 1,*

$$\frac{h^{1/2}\{T_{1n} - h^{-1}b_1 - (nh^2)^{-1}c_1\}}{\sigma_1} \xrightarrow{\mathcal{L}} N(e_1/\sigma_1, 1), \tag{3.11}$$

$$\frac{h^{1/2}T_{2n}}{\sigma_2} \xrightarrow{\mathcal{L}} N(e_2/\sigma_2, 1), \tag{3.12}$$

$$\frac{h^{1/2}\{T_{3n} - h^{-1}b_3 - (nh^2)^{-1}c_3\}}{\sigma_3} \xrightarrow{\mathcal{L}} N(e_3/\sigma_3, 1), \tag{3.13}$$

where $e_1 = E\{l^2(X)\pi_1(X)\}$, $e_2 = E\{l^2(X)\pi_2(X)f(X)\}$, and $e_3 = E\{l^2(X)\pi_3(X)\}$.

The results of Theorems 1 and 4 indicate that, all three tests have some powers detecting local alternative approaching the null at rate $(nh^{1/2})^{-1/2}$. A simple calculation shows that the asymptotic power of T_{jn} against local misspecifications (3.10) equals

$$\Phi(e_j/\sigma_j - z_{1-\alpha}), \quad j = 1, 2, 3,$$

where Φ and $z_{1-\alpha}$ denote the cumulative distribution function and the $1-\alpha$ quantile of the standard normal distribution. Again assuming $\pi_1 = \pi_2 f = \pi_3$, the local power of T_{jn} is inversely proportional to σ_j.

3.4 Relative Efficiency

From a local asymptotic point of view, it follows from Theorems 1 and 4 that we only need to compare the magnitude of σ_j or, equivalently, the L_2-norm of a kernel-dependent function, which uniquely characterizes the underlying test procedure. The norm comparison is provided in Lemma 2. Its proof can be found in Dette and von Lieres und Wilkau (2001).

Lemma 2 *If K is a probability density function, then*

$$\|K * K\|_2 \leq \|K\|_2 \leq \|2K - K * K\|_2.$$

Thus, if we assume $\pi_1 \equiv \pi_2 f \equiv \pi_3$, ignore the bias and use an identical bandwidth h for all three tests, then a test based on the statistic T_{2n} is more powerful, from a local asymptotic point of view, than T_{3n}, but is less powerful than T_{1n}. The conclusion of this study lends further support to the interpretation in Section 2. On the other hand, this conclusion may not be substantiated in general, if different tests use distinct bandwidths, or if $\pi_1 \neq \pi_2 f \neq \pi_3$. Our results can be used to evaluate the performance of other nonparametric tests based on kernel and local polynomial smoothing.

In many practical applications, the multiweight kernel functions, of the form below, are commonly used:

$$\frac{1}{\mathrm{Beta}(1/2, \ell+1)}(1-t^2)^{\ell} I(|t| \leq 1), \quad \ell = 0,\ 1,\ \ldots\ .$$

Table 1 summarizes the values of the L_2-norms for the Epanechnikov kernel ($\ell = 1$), biweight kernel ($\ell = 2$), and triweight kernel ($\ell = 3$). We see that the differences in the asymptotic variances caused by the different test statistics can be substantial.

Table 1: L_2-norm Comparison of Kernel-Dependent Functions

Kernel	$\|K*K\|_2^2$	$\|K\|_2^2$	$\|2K-K*K\|_2^2$
Epanechnikov	0.4338	0.6000	0.8510
Biweight	0.5164	0.7143	1.0066
Triweight	0.5879	0.8159	1.1528

We now compare the different tests under fixed alternatives by an application of Theorem 3. It is easy to see that the power function of the test based on the statistic T_{jn} is approximately equal to

$$P(\text{“rejection”}) \approx \Phi\left(\sqrt{n}\frac{\Delta_j}{\tau_j} - \frac{\sigma_j}{\tau_j}\frac{z_{1-\alpha}}{\sqrt{nh}}\right), \quad j = 1, 2, 3, \tag{3.14}$$

and is therefore increasing with the expression Δ_j/τ_j. These ratios depend on certain features of the model in a complicated manner and therefore a comparison of the kernel-based tests is more difficult, even in the case $\pi_1 \equiv \pi_2 f \equiv \pi_3$. Nevertheless, the formula (3.14) can be used to obtain a qualitative impression of the tests with respect to particular alternatives and we will illustrate this in the following section.

4 Simulations

In this section, we present simulation studies to illustrate the extent to which the forgoing asymptotic results on power comparison are reflected in finite-samples. The independent data $\{(X_i, Y_i)_{i=1}^n\}$ are generated as follows from a nonparametric regression model:

$$Y = m(X) + \epsilon, \tag{4.1}$$

in which the regressor X is independent of the noise $\epsilon \sim N(0,1)$, and three choices of $m(x)$ are considered:

Example 1: $m(x) = 1 + 2(4x-2) + \theta(4x-2)^2, \quad \theta \geq 0.$

Example 2: $m(x) = 1 + \cos\{\theta(4x-2)\pi\}, \quad \theta \geq 0.$

Example 3: $m(x) = 1 + 2(4x-2) + \theta(4x-2)^2 + \cos\{\theta(4x-2)\pi\}, \quad \theta \geq 0.$

The null hypothesis we want to test is that a linear model is correct. In each example, $\theta = 0$ corresponds the null regression model from which data are simulated. At the nominal 5% significance level, the critical values of T_{jn} are estimated from the 95th sample percentiles of the corresponding test statistics among 400 random samples, and the powers are approximated by the rejection rates among 400 random samples. For simplicity, the weight functions are set to be $\pi_1 = \pi_2 = \pi_3 \equiv 1$.

[*Put Figures 1–3 about here*]

The power comparisons are displayed in Figures 1–3. Three types of design densities f are treated: $U(0,1)$ in Figure 1, $\text{Beta}(2,3)$ in Figure 2, and $N(0,\sigma)$ with $\sigma = 1/3$ in Figure 3. Sample sizes consist of $n = 100$ and $n = 400$. In each panel, the non-stochastic bandwidth h for the local linear method is chosen according to the empirical formula

$$h = 1.5 \times \text{std}(X) \times n^{-2/9};$$

for ease of computations, the Epanechnikov kernel is used.

We first inspect Figure 1, in which the uniform design, satisfying $\pi_1 = \pi_2 f = \pi_3$, is ideal for the comparison of powers under local alternatives. Example 1 demonstrates the test based on T_{1n} clearly outperforms the tests based on T_{2n} and T_{3n}; this observation agrees with our asymptotic study in Section 3. It is also seen in Examples 2–3 that the relative behaviours of the three tests alternate as the alternatives contain higher-frequency components, leading to less-smooth alternatives.

For Figures 2–3, we do not have $\pi_1 = \pi_2 f = \pi_3$, and the normal designs do not fullfill condition (A3). Therefore, the conclusions based on Figure 1 may not necessarily hold for Figures 2–3; see discussions in the paragraph following Lemma 2. However, if T_{2n} is modified from (2.2) to be

$$T_{2n} = \sum_{i=1}^{n} \{Y_i - g(X_i; \widehat{\boldsymbol{\theta}})\}\{\widehat{m}_h(X_i) - \widehat{g}_h(X_i; \widehat{\boldsymbol{\theta}})\}\pi_2(X_i),$$

where the local linear method is used for the term $\widehat{m}_h(X_i) - \widehat{g}_h(X_i; \widehat{\boldsymbol{\theta}})$, then our simulations reveal that the power plots of T_{jn}, $j = 1, 2, 3$, in Figures 2–3, are similar to those in Figure 1.

We finally illustrate the application of formula (3.14) to obtain a qualitative picture for the power function of the three tests. For this we consider the situation in Figure 1, where $X \sim U[0,1]$. In Figure 4 we display the approximation of the power function obtained by formula (3.14) for the three examples, where the sample size was chosen as $n = 100$. We find a good qualitative agreement with the picture obtained in the simulations. For example, we observe from Figure 4 that in Example 1 the power function is increasing, where the test based on T_{1n} yields the best results. This reflects the observations from the simulation study. Similarly, in Example 3 Figure 4 indicates some problems of the tests to detect alternatives with $\theta = 0.25$, which was also observed in our simulation study. The situation in Example 2 is not so clear, because the asymptotic theory suggests that the power for $n = 100$ is close to one, even for large values of θ. Nevertheless the oscillating behaviour of the power function of all tests in this example for larger values of θ is clearly reflected by the asymptotic result. These examples indicate that the asymptotic results under fixed alternatives are useful to obtain a qualitative picture for the power of the tests with respect to particular alternatives without performing an extensive simulation study.

[*Put Figure 4 about here*]

Acknowledgements. The second author is grateful to the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen) for financial support during this research.

References

Azzalini, A., Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Statist. Soc., Ser.* B, 55, 549-559.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* 27, 1012-1050.

Dette, H., von Lieres und Wilkau, C. (2001). Testing additivity by kernel-based methods. *Bernoulli* 7, 669-697.

Fan, J., and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*, London: Chapman and Hall.

Fan, J., Zhang, C.M. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* 29, 153-193.

Gozalo, P.L., Linton, O.B. (2001). Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics* 104, 1-48.

González-Manteiga, W., Cao, R. (1993). Testing the hypothesis of a general linear model using nonparametric regression estimation. *Test* 2, 161-188.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* 21, 1926-1947.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75, 263-289.

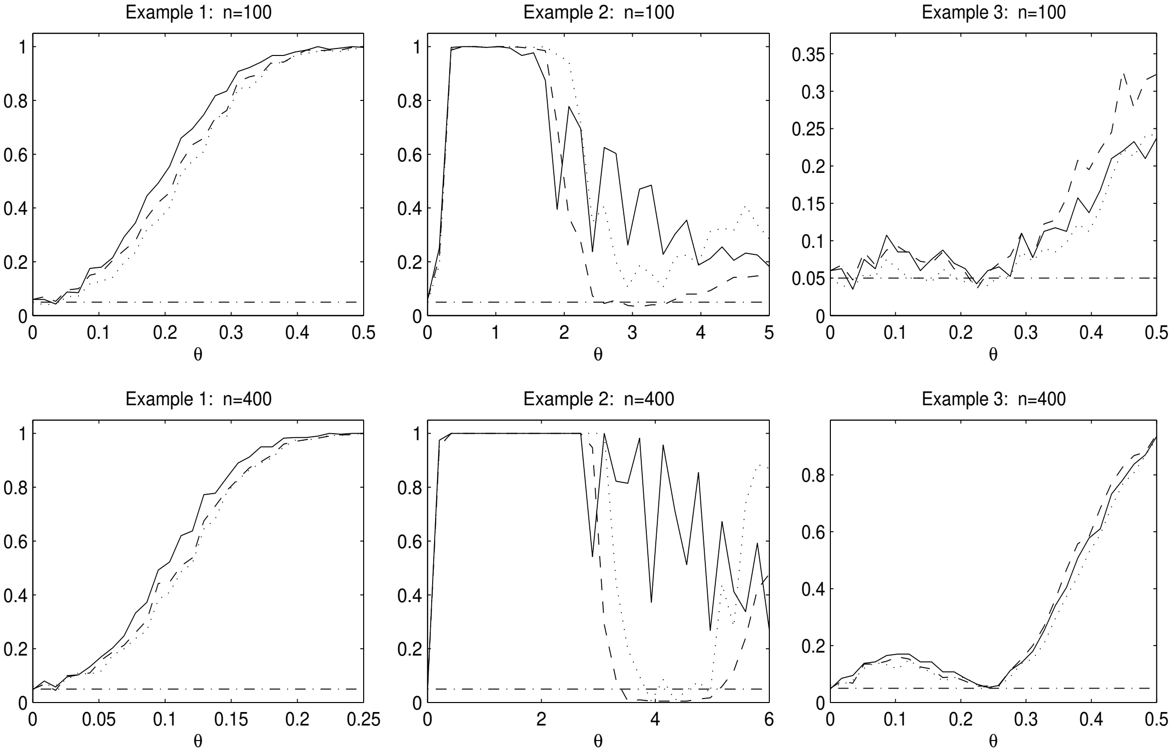

Figure 1: Simulated power curves of three test statistics for linearity, when observations of (X, Y) are simulated from (4.1) and $X \sim U(0, 1)$. Solid curve: the T_{1n} test; dashed curve: the T_{2n} test; dotted curve: the T_{3n} test. The bottom dash-dotted line denotes the nominal 5% significance level.

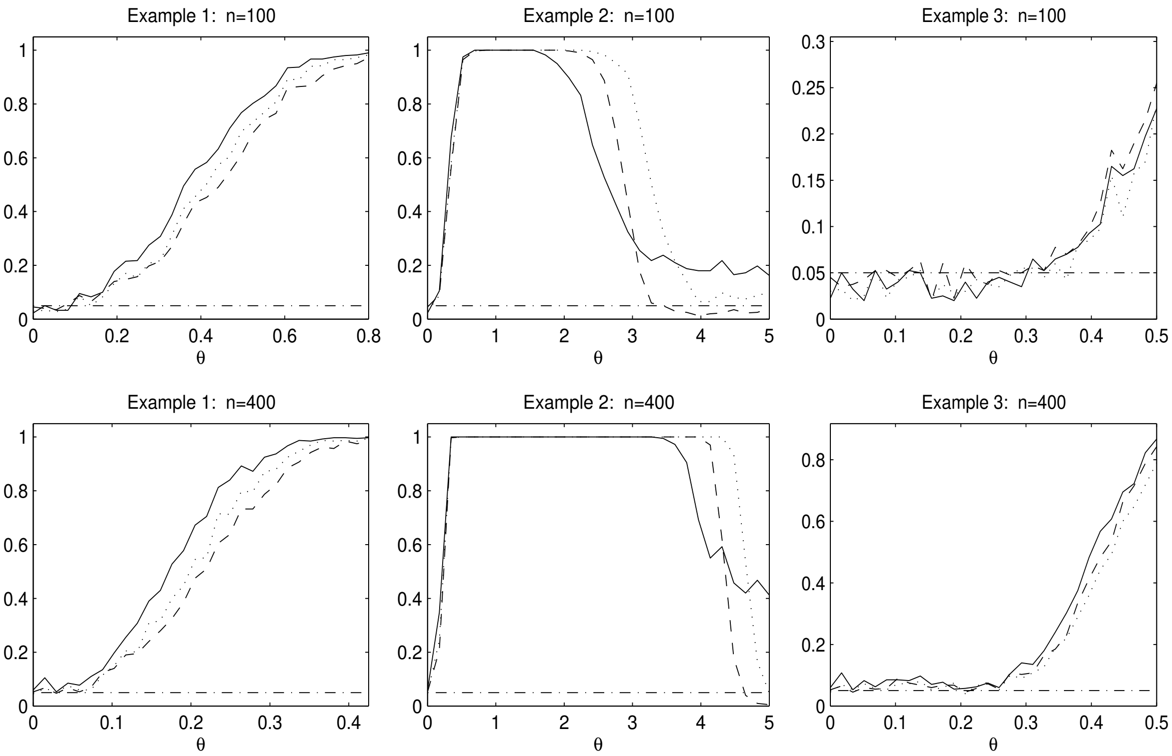

Figure 2: The captions are similar to those of Figure 1, except $X \sim \mathrm{Beta}(2, 3)$.

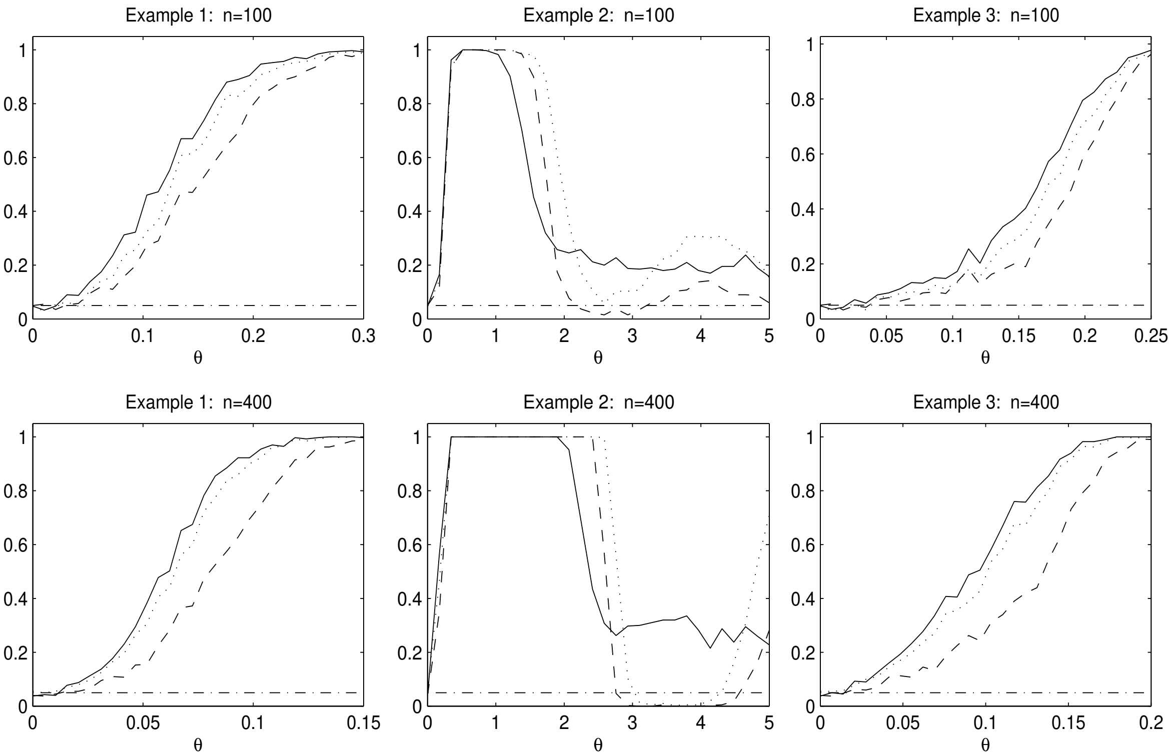

Figure 3: The captions are similar to those of Figure 1, except $X \sim N(0, 1/3)$.

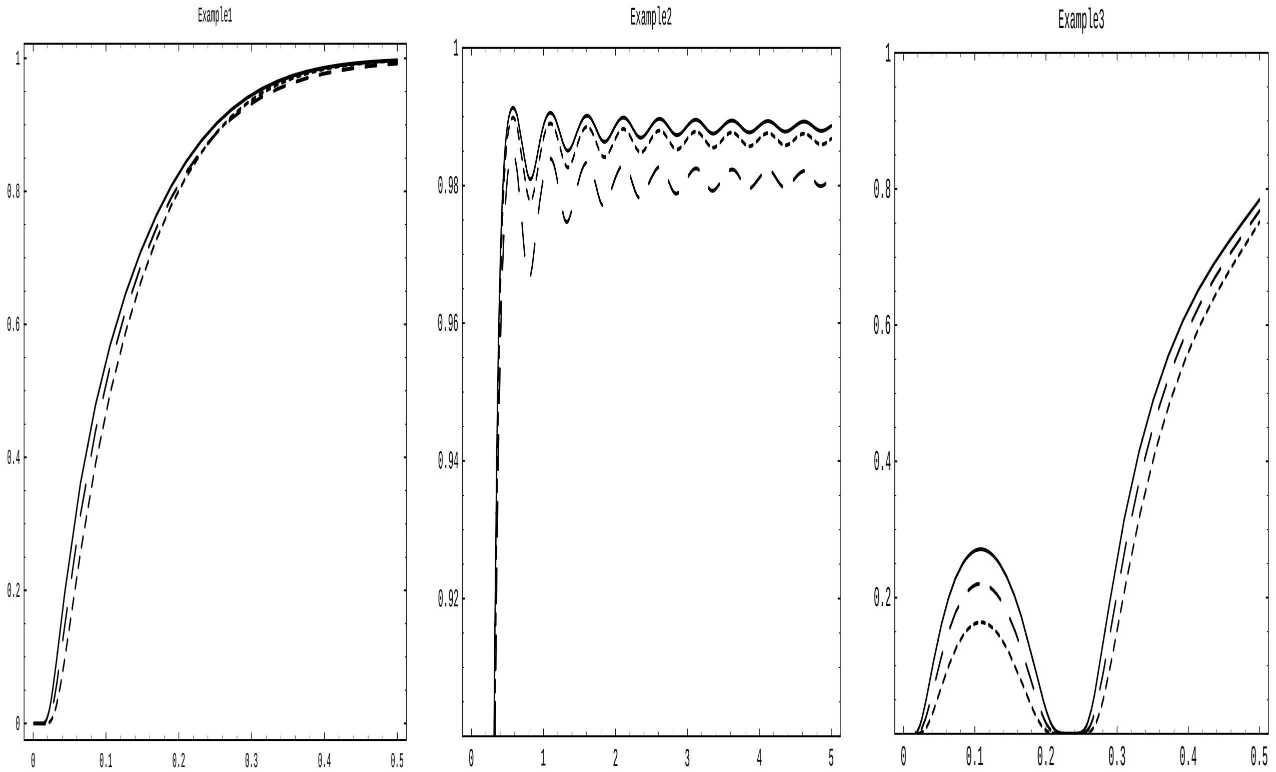

Figure 4: Qualitative behaviour of the power functions obtained from the approximation (3.14) in the situation considered in Figure 1.