The Annals of Statistics
2000, Vol. 28, No. 4, 1054–1082

TESTING MONOTONICITY OF REGRESSION

BY SUBHASHIS GHOSAL, ARUSHARKA SEN AND AAD W. VAN DER VAART

Free University, Amsterdam,
University of Hyderabad and Free University, Amsterdam

We consider the problem of testing monotonicity of the regression function in a nonparametric regression model. We introduce test statistics that are functionals of a certain natural U-process. We study the limiting distribution of these test statistics through strong approximation methods and the extreme value theory for Gaussian processes. We show that the tests are consistent against general alternatives.

1. Introduction. Consider two real valued random variables X and Y related by the regression model

$$Y = m(X) + \varepsilon, \tag{1.1}$$

where $E\varepsilon = 0$. Occasionally, one assumes that $m(\cdot)$ has a known functional form containing a few unknown parameters and estimates the parameters by the method of least squares. In the absence of specific assumptions, $m(\cdot)$ can still be estimated by various non-parametric methods such as kernel methods only under certain smoothness conditions. In some situations, one expects some qualitative structure in the regression function such as monotonicity. For example, if X stands for income and Y for expenditure on some specific item such as food or housing, one would naturally expect that $m(\cdot)$ is an increasing function. A similar situation also arises in bio-medical or nutrition studies where the response variable may be expected to be an increasing function of the dose of a drug or the level of a nutrient, at least over some range of interest. Methods of estimation of regression function under a monotonicity restriction (or the so called isotonic regression) have been widely discussed in the literature; see, Barlow, Bartholomew, Bremner and Brunk (1972), Robertson, Wright and Dykstra (1988), Hanson, Pledger and Wright (1973), Wright (1981, 1982), Ramsay (1988, 1998), Mukerjee (1988), Mammen (1991) and the references therein. Somewhat analogous percentile regression estimates have also been considered; see Cryer, Robertson, Wright and Casady (1972), Casady (1976) and Wright (1984).

The problem of testing monotonicity of the regression function is relatively less addressed in the literature. A test for monotonicity may be viewed as a composite goodness of fit test. In the parametric framework, Robertson and Wegman (1978) and Robertson, Wright and Dykstra (1988) consider the discrete version of the problem and discuss the likelihood ratio test for the pres-

Received February 1999; revised June 2000.
AMS 1991 *subject classifications*. Primary 62G08, 62G10; secondary 62G20.
Key words and phrases. Empirical process, extreme values, Gaussian process, monotone regression, strong approximation, U-process.

ence of order in the means of the normal populations associated with the different values of a non-stochastic covariate. They also discuss the more general case of exponential families of distributions and approximations of the distribution of the likelihood ratio statistics. In the non-parametric set up, Schlee (1982) discusses a test based on an estimate of the derivative of the regression function. Some very recent works consider the problem of testing monotonicity. Bowman, Jones and Gijbels (1998) construct a test using the idea of the critical bandwidth, originally introduced by Silverman (1981) for testing unimodality of a density. The essential idea behind this test is to fit a local linear regression and look at the smallest value of the bandwidth for which the estimate becomes monotone. This critical bandwidth is then used as a test statistic and the P-value is computed by the bootstrap method. Hall and Heckman (2000) develop a test for monotonicity by calibrating for linear regression. In their approach, Hall and Heckman (2000) test the positivity of the slope of the fitted least square linear regression line over each small block of observations. They also use the bootstrap method and calculate the P-value. Dümbgen and Spokoiny (1998) consider the white noise model and construct tests for qualitative hypotheses such as monotonicity or convexity, by considering the supremum, over all bandwidths, of the L_∞-distance between a kernel estimate and the null hypothesis. Although, the test of Dümbgen and Spokoiny (1998) has a natural appeal, it requires the computation of kernel estimates for all bandwidths, and may be difficult to implement. Moreover, their procedure does not carry over immediately to the regression case. If for a given sample, the bandwidth is made arbitrarily small, the kernel estimate would be highly variable, and in fact will not be defined at many points if the kernel is compactly supported. In the context of densities, Woodroofe and Sun (1999) test uniformity against monotonicity.

In this article, we consider testing monotonicity of the regression function and construct tests with a given asymptotic level. Our test statistics are suitable functionals of a stochastic process which may be viewed as a local version of Kendall's tau statistic and have simple natural interpretations. The process involved is a degree-two U-process, as in Nolan and Pollard (1987). We determine the critical regions of the tests by computing the limiting distributions of the test statistics with the help of the empirical process approximation of the U-process defined by the Hájek projection, strong approximation of the empirical process by a Gaussian process and the extreme value theory for stationary Gaussian processes. The test is therefore straightforward to implement. We neither need to compute kernel regression estimates for different bandwidths nor need to do bootstrap simulations.

The paper is organized as follows. In Section 2, we introduce two different types of test statistics. We also formally describe the model and the hypothesis and explain the notation and regularity conditions in this section. In Section 3, we investigate the asymptotic behavior of the U-process and establish the Gaussian process approximation. In Section 4, we discuss the limiting distribution of the first test statistics using the extreme value theory for stationary Gaussian processes and the results of Section 3. In Section 5, we show that this test is consistent against all fixed alternatives and also against alternatives that approach the null sufficiently slowly. The second test statistic is studied in Section 6. In Section 7, a small simulation study is carried out to investigate the behavior of the first statistic for finite sample sizes. Technical proofs are presented in Section 8 and the Appendix.

2. The test statistics. Suppose that we have n independent and identically distributed (i.i.d.) observations (X_i, Y_i), $i = 1, \dots, n$, where $Y_i = m(X_i) + \varepsilon_i$ and X_i and ε_i are independent random variables with distribution functions F and G respectively. For identifiability, we shall assume that $E\varepsilon = 0$, though the assumption is otherwise unnecessary. Let $T = [a, b]$ be a compact interval which is the region of our interest in the domain of definition of the regression function $m(\cdot)$. Consider testing the hypothesis

$$H_0 : \ m(\cdot) \text{ is an increasing function on } T.$$

We assume that F has a density f which is continuous and positive on T and G has a continuous density g. We also assume that the support of F is a compact interval containing T. From a practical point of view, this is normally a satisfactory assumption. Further, it will turn out that our test statistics are not affected by observations outside an open interval containing T, and therefore the condition may be assumed without any essential loss of generality. Also, we suppose that the function $m(\cdot)$ is continuously differentiable. The hypothesis H_0 can now be written as

$$H_0 : \ m'(t) \geq 0 \quad \text{for all } t \in T.$$

Intuitively, for a given $t \in T$, $m'(t) \geq 0$ if X and Y are concordant for X-values close to t. We recall that the degree of concordance may be estimated from the sample by the Kendall tau statistic. Since we have to restrict our attention to X-values close to t, it is therefore natural to use a locally weighted version of Kendall's tau, where more weight is attached to X-values close to t.

Let $k(\cdot)$ be a nonnegative, symmetric, continuous kernel supported in $[-1, 1]$ and twice continuously differentiable in $(-1, 1)$. Let h_n be a positive sequence converging to 0 and will be referred to as the bandwidth in what follows. Put $k_n(x) = h_n^{-1} k(h_n^{-1} x)$. With

$$\operatorname{sign}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -1, & \text{if } x < 0, \end{cases}$$

we consider

$$U_n(t) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \operatorname{sign}(Y_j - Y_i) \operatorname{sign}(X_i - X_j) \times k_n(X_i - t) k_n(X_j - t) \tag{2.1}$$

as a measure of discordance between X and Y when X is close to t. Ideally, if $m'(t) \geq 0$, $U_n(t)$ should be, apart from random fluctuations, less than or equal to 0. Therefore the U-process $\{U_n(t) : t \in T\}$ should mostly lie below the level

0 if H_0 is true. Depending on the way we measure the degree of violation, we may construct different natural test statistics.

We observe that the process $U_n(\cdot)$ is invariant under monotone transformation of Y. Therefore, any test based on $U_n(\cdot)$ will work even if only some monotone transformation of Y satisfies (1.1), that is, the test for monotonicity of $m(\cdot)$ will work for the bigger model

$$\phi(Y) = m(X) + \varepsilon, \tag{2.2}$$

where ϕ is a monotone function.

First, we note that $EU_n(t) \le 0$ for large n if $m'(t) < 0$. To see this, let $\tilde{G}$ stand for the distribution of $-\varepsilon$. Note that

$$\begin{aligned}
EU_n(t) &= \int\int\int\int \operatorname{sign}(m(x_2) + \varepsilon_2 - m(x_1) - \varepsilon_1)\operatorname{sign}(x_1 - x_2) \\
&\qquad\qquad \times k_n(x_1 - t)k_n(x_2 - t)g(\varepsilon_1)g(\varepsilon_2)f(x_1)f(x_2)d\varepsilon_1 d\varepsilon_2 dx_1 dx_2 \\
&= \int\int \Big[2\int \tilde{G}(m(x_2) - m(x_1) - \varepsilon_1)g(\varepsilon_1)d\varepsilon_1 - 1\Big]\operatorname{sign}(x_1 - x_2) \qquad (2.3) \\
&\qquad \times k_n(x_1 - t)k_n(x_2 - t)f(x_1)f(x_2)dx_1 dx_2 \\
&= \int\int [2G^*(m(t + h_n v) - m(t + h_n u)) - 1]\operatorname{sign}(U - v) \\
&\qquad \times k(u)k(v)f(t + h_n u)f(t + h_n v)du\,dv,
\end{aligned}$$

where $G^*(\varepsilon) = \int \tilde{G}(\varepsilon - \varepsilon')g(\varepsilon')d\varepsilon'$, the convolution of $\tilde{G}$ with G, or alternatively, the distribution of $\varepsilon_1 - \varepsilon_2$. The density g^* of G^* exists, is continuous and $g^*(0) = \int g^2(\varepsilon)d\varepsilon$. Therefore,

$$\frac{1}{h_n}[2G^*(m(t + h_n v) - m(t + h_n u)) - 1] \to 2g^*(0)m'(t)(v - u). \tag{2.4}$$

Applying a dominated convergence argument, we therefore conclude that

$$\begin{aligned}
&\frac{1}{h_n} EU_n(t) \\
&\qquad \to -2\left(\int g^2(\varepsilon)d\varepsilon\right) m'(t) \left(\int\int |u - v| k(u)k(v)du\,dv\right) f^2(t)
\end{aligned} \tag{2.5}$$

and so the limit is less than or equal to 0 if and only if $m'(t) \ge 0$. This partly justifies the intuition that $U_n(t)$ should be less than or equal to 0 under H_0, at least in expectation.

For each $t \in T$, let $c_n(t) = c_n(t; X_1, \ldots, X_n)$ be a positive random variable, possibly depending on $X_1, \ldots, X_n$, but not on $Y_1, \ldots, Y_n$. Further assume that, as a process in t, $c_n(\cdot)$ has continuous sample paths. Then a general class of test statistics is given by

$$S_n = \psi_n\left(\frac{U_n(\cdot)}{c_n(\cdot)}\right), \tag{2.6}$$

where $\psi_n(\cdot)$ is a positive functional on $C(T)$, the space of continuous functions on T. [By a positive functional $\psi(\cdot)$ on $C(T)$, we mean that $\psi(\phi_1) \ge \psi(\phi_2)$

whenever $\phi_1 \geq \phi_2$ pointwise.] In order to implement the testing procedure, however, we need to choose $\psi_n(\cdot)$ and $c_n(\cdot)$ judiciously such that S_n has a reasonably simple asymptotic null distribution. We shall shortly return to this issue.

First observe that any test statistic of the form (2.6) satisfies the following important inequality: Let $P^n_{m,F,G}$ denote the joint distribution of $(X_1, Y_1), \dots, (X_n, Y_n)$, where X has distribution F, ε has distribution G and the regression function in (1.1) is $m(\cdot)$. Let $P^n_{0,F,G}$ denote the same when $m(\cdot) \equiv 0$. Then for all $m \in H_0$,

$$P^n_{m,F,G}\{S_n \geq s\} \leq P^n_{0,F,G}\{S_n \geq s\}, \quad s \in \mathbb{R}. \tag{2.7}$$

This means that the Type I error probability is maximized in H_0 when $m(\cdot) \equiv 0$, and therefore, for the purpose of obtaining the limiting null distribution, it suffices to look at the case $m(\cdot) \equiv 0$ only.

To prove (2.7), set

$$\begin{aligned} U^0_n(t) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} & \operatorname{sign}(\varepsilon_j - \varepsilon_i)\operatorname{sign}(X_i - X_j) \\ & \times k_n(X_i - t)k_n(X_j - t). \end{aligned} \tag{2.8}$$

Note that when $m(\cdot) \equiv 0$, $U^0_n(\cdot) \equiv U_n(\cdot)$, and the distribution of $U^0_n(\cdot)$ does not depend on $m(\cdot)$.

If $m(\cdot)$ is increasing, then clearly

$$\begin{aligned} &\operatorname{sign}(Y_j - Y_i)\operatorname{sign}(X_i - X_j) \\ &\qquad = \operatorname{sign}(m(X_j) + \varepsilon_j - m(X_i) - \varepsilon_i)\operatorname{sign}(X_i - X_j) \\ &\qquad \leq \operatorname{sign}(\varepsilon_j - \varepsilon_i)\operatorname{sign}(X_i - X_j), \end{aligned} \tag{2.9}$$

and therefore, for all $t \in T$, $U_n(t) \leq U^0_n(t)$. By the nature of S_n, (2.7) now follows easily.

It may be noted that not every statistic of the form (2.6) is reasonable. The statistic should only look for the occurrence of large values in the process and ignore the small values. For example, if the functional is an integral, then cancellation between large and small values will tend to convince us that monotonicity is not violated and hence will result in an undesirable test statistic. Now, largeness of a process may be measured in different ways, for example, by the largest value of the process or the duration of the excursions of the process above a pre-determined level. These lead us to the following two choices of $\psi_n(\cdot)$:

(i) $\psi_n(\phi) = \psi(\phi) = \sup\{\phi(t) : t \in T\}$,
(ii) $\psi_n(\phi) = \operatorname{meas}\{t \in T : \phi(t) > u_n\} = \int_T \{t : \phi(t) > u_n\} dt$

for some suitable sequence u_n.

(As we did here, we shall use the same notation for indicators and the corresponding sets.) Admittedly, there could be many other natural choices of $\psi_n(\cdot)$. A more general class containing (ii) is mentioned at the end of Section 6. The first statistic is expected to be more sensitive toward violations of monotonicity in terms of magnitude (measured by the absolute value of the derivative of the regression function) while the second statistic can be expected to perform well against alternatives which violate monotonicity on long intervals.

We choose $c_n(t)$ such that the variability of $U_n(t)/c_n(t)$ is approximately the same for different t's when $m(\cdot) \equiv 0$. From a statistical point of view, this works as a transformation bringing in homoscedasticity and so making comparisons of the values of the process at different t's more meaningful. Technically, this will turn out to be a great advantage, as we shall see that the process $U_n(t)/c_n(t)$ can then be approximated by a stationary Gaussian process, for which the limiting distribution of functionals of the form (i) and (ii) are already well investigated in the literature. A similar device is employed by Bickel and Rosenblatt (1973) in their study of global functionals of density estimators.

We now make the notion "the variability of $U_n(t))/(c_n(t)$ is approximately the same over different t" more precise. When $m(\cdot) \equiv 0$, the Hájek projection of $U_n(t)$ [the linear approximation to $U_n(t)$ first found by Hoeffding (1948)] is given by

$$\hat{U}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \psi_{n,t}(X_i, \varepsilon_i), \tag{2.10}$$

where

$$\psi_{n,t}(x, \varepsilon) = 2(1 - 2G(\varepsilon)) \int \operatorname{sign}(x - w) k_n(w - t)\, dF(w) k_n(x - t). \tag{2.11}$$

Note that $E\hat{U}_n(t) = 0$ and the variance of $\hat{U}_n(t)$ is given by $\sigma_n^2(t)/n$, where

$$\begin{aligned} \sigma_n^2(t) &= 4\int (1 - 2G(\varepsilon))^2 dG(\varepsilon) \int \left(\int \operatorname{sign}(x - w) k_n(w - t)\, dF(w)\right)^2 \\ &\qquad\qquad \times k_n^2(x - t)\, dF(x) \\ &= \tfrac{4}{3} \int\int\int \operatorname{sign}(x - w_1)\operatorname{sign}(x - w_2) k_n(w_1 - t) k_n(w_2 - t) \\ &\qquad\qquad \times k_n^2(x - t)\, dF(w_1)\, dF(w_2)\, dF(x). \end{aligned} \tag{2.12}$$

One may thus like to choose $c_n(t) = \sigma_n(t)/\sqrt{n}$. However, $\sigma_n(t)$ involves the unknown F which must be estimated. Our final choice of $c_n(t)$ will be $\hat{\sigma}_n(t)/\sqrt{n}$ where $\hat{\sigma}_n^2(t)$ is the U-statistic for $\sigma_n^2(t)$, that is,

$$\begin{aligned} \hat{\sigma}_n^2(t) = \frac{4}{3n(n-1)(n-2)} \sum_{\substack{1 \le i,j,k \le n \\ i \ne j \ne k}} &\operatorname{sign}(X_i - X_j)\operatorname{sign}(X_i - X_k) \\ &\times k_n(X_j - t) k_n(X_k - t) k_n^2(X_i - t). \end{aligned} \tag{2.13}$$

Put

$$S_{1,n} = \sup\left\{\frac{\sqrt{n}\,U_n(t)}{\hat{\sigma}_n(t)} : t \in T\right\} \tag{2.14}$$

and

$$S_{2,n} = \operatorname{meas}\left\{t \in T : \frac{\sqrt{n}\,U_n(t)}{\hat{\sigma}_n(t)} > u_n\right\}. \tag{2.15}$$

We shall refer to $S_{1,n}$ and $S_{2,n}$ respectively as the supremum statistic and the time spent statistic. The respective tests are given by

$$\text{Reject } H_0 \text{ at level } \alpha \text{ if } S_{i,n} > \tau_{i,n,\alpha}, \qquad i = 1, 2, \tag{2.16}$$

where

$$\lim_{n\to\infty} P^n_{0,F,G}\{S_{i,n} > \tau_{i,n,\alpha}\} = \alpha, \qquad i = 1, 2. \tag{2.17}$$

To approximate the critical values, we need the limiting distributions of $S_{1,n}$ and $S_{2,n}$. In the following sections, we study these limiting distributions.

3. Gaussian process approximation. Since our test statistics are functionals of the process

$$\left\{\frac{\sqrt{n}\,U_n(t)}{\hat{\sigma}_n(t)} : t \in T\right\},$$

the asymptotic properties of the test statistics may be obtained from the corresponding properties of the process. In this section, we show that this process may be approximated by a stationary Gaussian process with continuous sample paths.

The technique is based on, among other things, strong approximation and should be interpreted in the sense that we can obtain a rich enough probability space and copies of the original observations on it so that the corresponding copy of the U-process has the desired approximation property. Since ultimately we are interested in distribution of the process only, we can regard the new probability space as our given one.

We first need to introduce some additional notations. Let $K(\cdot)$ stand for the distribution function corresponding to the density $k(\cdot)$, that is, $K(x) = \int_{z\le x} k(z)dz$. Set

$$q(x) = \int \operatorname{sign}(x - w)k(w)dw = 2K(x) - 1 \tag{3.1}$$

and

$$\rho(s) = \frac{\int q(z)q(z-s)k(z)k(z-s)dz}{\int q^2(z)k^2(z)dz}. \tag{3.2}$$

Also set $T_n = [0, (b-a)/h_n]$. It is assumed throughout this section that $m(\cdot) \equiv 0$.

THEOREM 3.1. *Let the bandwidths satisfy the conditions*

$$h_n\sqrt{\log n}\to 0 \quad \text{and} \quad nh_n^2/(\log n)^2\to\infty. \tag{3.3}$$

Then there exists a sequence of stationary Gaussian processes ξ_n, *defined on the same sample space and indexed by* T_n, *such that* $\xi_n(\cdot)$ *has continuous sample paths*,

$$E\xi_n(s)=0, \qquad E(\xi_n(s_1)\xi_n(s_2))=\rho(s_1-s_2) \quad s, s_1, s_2\in T_n,$$

and

$$\begin{aligned}\sup_{t\in T}\left|\sqrt{n}\frac{U_n(t)}{\hat\sigma_n(t)}-\xi_n(h_n^{-1}(t-a))\right| \\ = O_p(n^{-1/4}h_n^{-1/2}\sqrt{\log n}+h_n\sqrt{\log n}).\end{aligned} \tag{3.4}$$

The proof of Theorem 3.1 is long and involves several steps:

1. approximation of the U-process $U_n(\cdot)$ by $\hat U_n(\cdot)$;
2. strong approximation of the empirical process $\sqrt{n}\,\hat U_n(\cdot)$ by a Gaussian process $G_n(\cdot)$, say;
3. uniform approximation of $\hat\sigma_n(t)$ by $\sigma_n(t)$;
4. approximation of the scaled Gaussian process $\frac{G_n(t)}{\sigma_n(t)}$ by a stationary Gaussian process $\xi_n(t)$.

Below, we state these approximations as Lemmas 3.1–3.4. The proofs of these lemmas are deferred to Section 7. The condition (3.3) is assumed throughout.

LEMMA 3.1.

$$\sup_{t\in T}|U_n(t)-\hat U_n(t)|=O_p(n^{-1}h_n^{-3/2}).$$

LEMMA 3.2. *There exists a sequence of Gaussian processes* $G_n(\cdot)$, *indexed by* T, *with continuous sample paths and with*

$$EG_n(t)=0, \quad E(G_n(t_1)G_n(t_2))=E(\psi_{n,t_1}(X,\varepsilon)\psi_{n,t_2}(X,\varepsilon)), \quad t, t_1, t_2\in T,$$

where X *has distribution* F, ε *has distribution* G *and* $\psi_{n,t}$ *is as defined in* (2.11), *such that*

$$\sup_{t\in T}|\sqrt{n}\hat U_n(t)-G_n(t)|=O\left(n^{-1/4}h_n^{-1}\sqrt{\log n}\right) \qquad a.s. \tag{3.5}$$

LEMMA 3.3. *The following assertions hold*:

(a) $\lim_{n\to\infty} h_n\sigma_n^2(t)=\frac{4}{3}f^3(t)\int q^2(x)k^2(x)dx$ *uniformly in* t;
(b) $\liminf_{n\to\infty} h_n \inf_{t\in T}\sigma_n^2(t)>0$;
(c) $\sup_{t\in T}|\sigma_n^2(t)-\hat\sigma_n^2(t)|=O_p(n^{-1/2}h_n^{-2})$.

LEMMA 3.4. *For the sequence of Gaussian processes $\{G_n(t) : t \in T\}$ obtained in Lemma 3.2, there corresponds a sequence of stationary Gaussian processes $\{\xi_n(s) : s \in T_n\}$ with continuous sample paths such that*

$$E\xi_n(s) = 0, \quad E(\xi_n(s_1)\xi_n(s_2)) = \rho(s_1 - s_2), \quad s, s_1, s_2 \in T_n, \tag{3.6}$$

where $\rho(\cdot)$ is defined by (3.2) *and*

$$\sup_{t\in T}\left|\frac{G_n(t)}{\sigma_n(t)} - \xi_n(h_n^{-1}(t-a))\right| = O_p(h_n\sqrt{\log h_n^{-1}}). \tag{3.7}$$

PROOF OF THEOREM 3.1. We have

$$\begin{aligned}\sup_{t\in T}\left|\sqrt{n}\frac{U_n(t)}{\hat\sigma_n(t)} - \xi_n(h_n^{-1}(t-a))\right| &\le \sqrt{n}\sup_{t\in T}\left|\frac{U_n(t)}{\hat\sigma_n(t)} - \frac{U_n(t)}{\sigma_n(t)}\right| \\ &\quad + \sqrt{n}\sup_{t\in T}\left|\frac{U_n(t)}{\sigma_n(t)} - \frac{\hat U_n(t)}{\sigma_n(t)}\right| \\ &\quad + \sup_{t\in T}\left|\sqrt{n}\frac{\hat U_n(t)}{\sigma_n(t)} - \frac{G_n(t)}{\sigma_n(t)}\right| \\ &\quad + \sup_{t\in T}\left|\frac{G_n(t)}{\sigma_n(t)} - \xi_n(h_n^{-1}(t-a))\right|.\end{aligned} \tag{3.8}$$

The last term on the right hand side (RHS) of (3.8) is $O_p(h_n\sqrt{\log h_n^{-1}})$ by Lemma 3.4. For the third term, note that, by applications of Lemma 3.2 and Lemma 3.3(b),

$$\begin{aligned}\sup_{t\in T}\left|\sqrt{n}\frac{\hat U_n(t)}{\sigma_n(t)} - \frac{G_n(t)}{\sigma_n(t)}\right| &\le \frac{\sup_{t\in T}|\sqrt{n}\,\hat U_n(t) - G_n(t)|}{\inf_{t\in T}\sigma_n(t)} \\ &= O_p(n^{-1/4}h_n^{-1}\sqrt{\log n})O(h_n^{1/2}) \\ &= O_p(n^{-1/4}h_n^{-1/2}\sqrt{\log n}).\end{aligned} \tag{3.9}$$

The second term on the RHS of (3.8) is dominated by

$$\frac{\sup_{t\in T}\sqrt{n}|U_n(t) - \hat U_n(t)|}{\inf_{t\in T}\sigma_n(t)} = O_p(n^{-1/2}h_n^{-1}) \tag{3.10}$$

by virtue of Lemma 3.1 and Lemma 3.3(b).

It therefore follows that

$$\begin{aligned}&\sqrt{n}\sup_{t\in T}\left|\frac{U_n(t)}{\sigma_n(t)}\right| - \sup_{t\in T}\left|\xi_n(h_n^{-1}(t-a))\right| \\ &\quad = O_p\left(h_n\sqrt{\log h_n^{-1}} + n^{-1/4}h_n^{-1/2}\sqrt{\log n}\right).\end{aligned} \tag{3.11}$$

Now $\sup_{t\in T}|\xi_n(h_n^{-1}(t-a))| = \sup_{s\in T_n}|\xi_n(s)|$, and since $\rho'(0) = 0$ in view of the symmetry of $k(\cdot)$, a Taylor series expansion yields that

$$E(\xi_n(s_1) - \xi_n(s_2))^2 = 2(1 - \rho(|s_1 - s_2|)) \le C_1|s_1 - s_2|^2 \tag{3.12}$$

for some constant C_1, independent of n. Also, because of unit variance, the L_2-diameter of the process is clearly bounded by 2. By Corollary 2.2.8 of van der Vaart and Wellner (1996), it then easily follows that for some constants C_2, C_3,

$$E\left(\sup_{s\in T_n}|\xi_n(s)|\right) \le E|\xi_n(0)| + C_2\int_0^2 \sqrt{\log\frac{C_3 h_n^{-1}}{\varepsilon}}\,d\varepsilon, \tag{3.13}$$

whence it follows that

$$\sup_{t\in T}|\xi_n(h_n^{-1}(t-a))| = O_p\left(\sqrt{\log h_n^{-1}}\right). \tag{3.14}$$

This shows that

$$\sqrt{n}\sup_{t\in T}\left|\frac{U_n(t)}{\sigma_n(t)}\right| = O_p\left(\sqrt{\log h_n^{-1}}\right). \tag{3.15}$$

Now, for the first term on the RHS of (3.8), we have

$$\sqrt{n}\sup_{t\in T}\left|\frac{U_n(t)}{\hat\sigma_n(t)} - \frac{U_n(t)}{\sigma_n(t)}\right| \le \sqrt{n}\sup_{t\in T}\left|\frac{U_n(t)}{\sigma_n(t)}\right| \sup_{t\in T}\left|\frac{\sigma_n(t)}{\hat\sigma_n(t)} - 1\right|. \tag{3.16}$$

Since $|x-1| \le |x^2-1|$ for $x \ge 0$,

$$\sup_{t\in T}\left|\frac{\hat\sigma_n(t)}{\sigma_n(t)} - 1\right| \le \sup_{t\in T}\left|\frac{\hat\sigma_n^2(t)}{\sigma_n^2(t)} - 1\right| \le \frac{\sup_{t\in T}|\hat\sigma_n^2(t)-\sigma_n^2(t)|}{\inf_{t\in T}\sigma_n^2(t)} = O_p(n^{-1/2}h_n^{-1}) \tag{3.17}$$

as a consequence of Lemma 3.3(b) and (c). Therefore,

$$\sup_{t\in T}\left|\frac{\sigma_n(t)}{\hat\sigma_n(t)} - 1\right| \le \frac{\sup_{t\in T}\left|\frac{\hat\sigma_n(t)}{\sigma_n(t)} - 1\right|}{1 - \sup_{t\in T}\left|\frac{\hat\sigma_n(t)}{\sigma_n(t)} - 1\right|} = O_p(n^{-1/2}h_n^{-1}). \tag{3.18}$$

Note that by the assumption on h_n, it also follows that $\log h_n^{-1} = O(\log n)$. Thus the first term on the RHS of (3.8) is $O_p(n^{-1/2}h_n^{-1}\sqrt{\log n})$. We thus obtain (3.4) by combining all the above assertions. □

4. Distribution of the supremum statistic. From Theorem 3.1, we obtain that

$$S_{1,n} = \sup_{s\in T_n}\xi_n(s) + O_p(\delta_n), \tag{4.1}$$

where

$$\delta_n = h_n\sqrt{\log n} + n^{-1/4}h_n^{-1/2}\sqrt{\log n}. \tag{4.2}$$

Therefore if for some positive a_n and real number b_n, we have

$$a_n\left(\sup_{s\in T_n}\xi_n(s) - b_n\right) \to_d Z \tag{4.3}$$

for some random variable Z, then we also have

$$a_n(S_{1,n} - b_n) \to_d Z, \tag{4.4}$$

provided that $a_n \delta_n \to 0$. The study of (4.3) involves extreme value theory of stationary Gaussian processes. Note that since we are only interested in distributions and since the covariance function for the process $\xi_n(\cdot)$ is also free from n, we may assume that all the $\xi_n(\cdot)$'s are the same Gaussian process $\xi(\cdot)$, say. To be more precise, we may assume that there is a stationary Gaussian process $\xi(t)$, $t \geq 0$, with continuous sample paths satisfying

$$E\xi(t) = 0, \quad E(\xi(t_1)\xi(t_2)) = \rho(t_1 - t_2), \qquad t, t_1, t_2 \geq 0,$$

such that ξ_n is the restriction of ξ to T_n. This brings us to the standard set up for the extremal theory of stationary Gaussian processes.

Recall that $T_n = [0, (b-a)/h_n]$. We have the following theorem.

THEOREM 4.1. *For any x, we have*

$$\lim_{n\to\infty} P\left\{a_n\left(\sup_{t\in T_n} \xi(t) - b_n\right) \leq x\right\} = \exp(-e^{-x}), \tag{4.5}$$

where

$$a_n = \sqrt{2\log((b-a)/h_n)}, \tag{4.6}$$

$$b_n = \sqrt{2\log((b-a)/h_n)} + \frac{\log \frac{\lambda^{1/2}}{2\pi}}{\sqrt{2\log((b-a)/h_n)}} \tag{4.7}$$

and

$$\lambda = -\frac{6\int(2K(x)-1)k^2(x)k'(x)dx + \int(2K(x)-1)^2 k(x)k''(x)dx}{\int(2K(x)-1)^2 k^2(x)dx}. \tag{4.8}$$

We first observe the fact that $\xi(t)$ is a mean square differentiable process, that is,

$$\rho(t) = 1 - \frac{\lambda t^2}{2} + o(t^2) \qquad \text{as } t \to 0 \tag{4.9}$$

with λ as defined in (4.8). Since $\rho(t) = 0$ for $t > 2$ (as $k(\cdot)$ has support in $[-1, 1]$), Theorem 4.1 is a direct consequence of Theorem 8.2.7 of Leadbetter, Lindgren and Rootzén (1983).

From Theorem 4.1, we easily obtain the asymptotic distribution of $S_{1,n}$.

THEOREM 4.2. *If*

$$h_n \log n \to 0 \quad \textit{and} \quad nh_n^2/(\log n)^4 \to \infty, \tag{4.10}$$

then

$$\lim_{n\to\infty} P\left\{a_n(S_{1,n} - b_n) \leq x\right\} = \exp(-e^{-x}), \tag{4.11}$$

so for any $0 < \alpha < 1$, *the test based on the critical region*

$$S_{1,n} \geq b_n + \frac{1}{a_n} \log\left(\frac{1}{\log(1-\alpha)^{-1}}\right) = \sqrt{2\log\frac{b-a}{h_n}} - \frac{\log\log(1-\alpha)^{-1} - \log\frac{\lambda^{1/2}}{2\pi}}{\sqrt{2\log\frac{b-a}{h_n}}} \tag{4.12}$$

has asymptotic level α.

We note that the condition (4.10) is needed to ensure that $a_n\delta_n \to 0$ and is therefore slightly stronger than (3.3), which only means that $\delta_n \to 0$. Bandwidth sequences such as $n^{-\beta}$, $\beta < 1/2$, and $(\log n)^{-\gamma}$, $\gamma > 1$, satisfy (4.10) and so the test given by (4.12) has asymptotic level α. Therefore any sequence that can be bounded above and below, up to constant multiples, by these two sequences respectively, also satisfies the required condition.

In this context, it may also be observed that a test for the constancy of the regression function on T against general alternatives is given by the statistic $\sup\{\sqrt{n}|U_n(t)|/\hat{\sigma}_n(t) : t \in T\}$. This statistic, centered by b_n and then multiplied by a_n, where a_n and b_n are defined by (4.6) and (4.7) respectively, converges in law to the distribution $\exp[-2e^{-x}]$; see Corollary 11.1.6 of Leadbetter, Lindgren and Rootzén (1983).

5. Consistency and rate of separation. In this section, we prove that the test specified by the supremum statistic is consistent against general alternatives. Also for certain alternatives that approach the null in a specific manner, we show that the power tends to one.

THEOREM 5.1. *If, in the model* (1.1), *the regression function is decreasing at some point, that is,* $m'(t) < 0$ *for some* $t \in [a, b]$, *then under* $m(\cdot)$, *the probability of the event that* (4.12) *happens, tends to* 1 *provided that* $nh_n^3/\log h_n^{-1} \to \infty$. *In other words, the test specified by* $S_{1,n}$ *is consistent at any level.*

In Theorem 5.1, the condition (4.10) is not used although the test with critical region (4.12) may not have asymptotic level α if (4.10) is not satisfied. The condition $nh_n^3/\log h_n^{-1} \to \infty$, equivalently $nh_n^3/\log n \to \infty$, however, implies the second part of (4.10). Bandwidth sequences such as $n^{-\beta}$, $\beta < 1/3$ and $(\log n)^{-\gamma}$, $\gamma > 1$ yield tests that are both level α and consistent.

PROOF OF THEOREM 5.1. Obviously, $S_{1,n} \geq \sqrt{n}U_n(t)/\hat{\sigma}_n(t)$ for any $t \in [a, b]$, in particular, for the t for which $m'(t) < 0$. Now, by Lemma 3.3, with probability tending to one, $\hat{\sigma}_n(t) \sim \sigma_n(t)$, where $\sim$ means that the ratio of the two sides tends to 1. We first show that the behavior of $U_n(t)$ is essentially the same as $EU_n(t)$. For this, we estimate $\mathrm{var}(U_n(t))$. Let

$$H_{n,t}((x_1, \varepsilon_1), (x_2, \varepsilon_2)) = \mathrm{sign}(m(x_2) + \varepsilon_2 - m(x_1) - \varepsilon_1) \times \mathrm{sign}(x_1 - x_2)k_n(x_1 - t)k_n(x_2 - t) \tag{5.1}$$

stand for the kernel for $U_n(t)$ and let

$$\tilde{H}_{n,t}(x,\varepsilon) = 2\int (1 - 2G(m(w) - m(x) - \varepsilon))\operatorname{sign}(x - w) \times k_n(w - t)\,dF(w)\,k_n(x - t), \tag{5.2}$$

denote the kernel for the Hájek projection. We note that

$$EH^2_{n,t}((X_1,\varepsilon_1),(X_2,\varepsilon_2)) \le \left(\int k_n^2(x-t)\,dF(x)\right)^2 = O(h_n^{-2}) \tag{5.3}$$

and

$$E\tilde{H}^2_{n,t}(X,\varepsilon) \le 4\left(\int k_n(w-t)\,dF(w)\right)^2 \int k_n^2(x-t)\,dF(x) = O(h_n^{-1}). \tag{5.4}$$

Therefore

$$\begin{aligned}
\operatorname{var}(U_n(t)) &= \frac{4}{n^2(n-1)^2}\Biggl\{\sum_{i<j}\operatorname{var}(H_{n,t}((X_i,\varepsilon_i),(X_j,\varepsilon_j))) \\
&\qquad + 6\sum_{i<j<k}\operatorname{cov}(H_{n,t}((X_i,\varepsilon_i),(X_j,\varepsilon_j)), \\
&\qquad\qquad H_{n,t}((X_i,\varepsilon_i),(X_k,\varepsilon_k)))\Biggr\} \\
&\le \frac{1}{n^2(n-1)^2}\left\{n(n-1)O(h_n^{-2}) + n(n-1)(n-2)O(h_n^{-1})\right\} \\
&= O(n^{-1}h_n^{-1}).
\end{aligned} \tag{5.5}$$

Therefore

$$\sqrt{nh_n}(U_n(t) - EU_n(t)) = O_p(1) \tag{5.6}$$

by Chebyshev's inequality, and hence from (2.5), we obtain that

$$h_n^{-1}U_n(t) \to_p -2\left(\int g^2(\varepsilon)\,d\varepsilon\right) m'(t)\left(\int\int |U - v| k(u)k(v)\,du\,dv\right) f^2(t), \tag{5.7}$$

which is positive, by assumption. Since by Lemma 3.3, $\hat{\sigma}_n(t)/\sigma_n(t) \to_p 1$ and $h_n^{1/2}\sigma_n(t)$ tends to a positive limit, it follows that $S_{1,n}$ is of the order $n^{1/2}h_n^{3/2}$, which exceeds the order of b_n, the dominant term in (4.12), under the condition $nh_n^3/\log h_n^{-1} \to \infty$. This completes the proof. □

To compute the limiting power function, we have to compute the limiting distribution of the test statistics under a sequence of alternatives that approaches the null in some appropriate sense. While the approximation of the U-process by the Hájek projection process and the strong approximation of the Hájek projection by a suitable (non-stationary) Gaussian process may be established to be valid under a sequence of alternatives by essentially the same arguments, it however seems difficult to study the limiting distribution of the supremum functional of that sequence of Gaussian processes. Nevertheless, we can still show that for certain alternatives that approach the null sufficiently slowly, the power tends to one.

Set $\tau_g^2 = \int g^2(\varepsilon)d\varepsilon$, $\tau_k^2 = \int\int |\mathrm{U}-v|k(u)k(v)dudv$ and $\eta = \inf\{f(t) : t \in T\}$.

THEOREM 5.2. *Assume $nh_n^3/\log h_n^{-1} \to \infty$ and let the sequence of alternative regression functions $m_n(\cdot)$ be continuously differentiable such that there exists t_n for which*

$$m_n'(t) \le -M\lambda_n \quad on \ [t_n - \delta_n, t_n + \delta_n].$$

Then

$$\lim_{n\to\infty} P^n_{m_n,F,G}\left\{S_{1,n} \ge b_n + \frac{1}{a_n}\log\left(\frac{1}{\log(1-\alpha)^{-1}}\right)\right\} = 1 \tag{5.8}$$

if either

$$\delta_n \ge h_n,$$
$$\lambda_n \ge n^{-1/2}h_n^{-3/2}\sqrt{\log((b-a)/h_n)},$$
$$M > \tau_g^{-2}\tau_k^{-2}\sqrt{\frac{2}{3\eta}\int(2K(x)-1)^2k^2(x)dx},$$

or

$$\delta_n = o(h_n),$$
$$m_n'(t) \le 0 \ on \ [t_n - h_n, t_n + h_n],$$
$$\lambda_n \ge n^{-1/2}h_n^{-3/2}\delta_n^{-3}\sqrt{\log((b-a)/h_n)},$$
$$M > \frac{1}{4}\tau_g^{-2}(k(0))^{-2}\sqrt{\frac{3}{2\eta}\int(2K(x)-1)^2k^2(x)dx}.$$

For instance, we might choose $h_n = n^{-\alpha}$ for some $\alpha < 1/3$. Then the power tends to one along any sequence of alternatives as in Theorem 5.2 with $\delta_n \ge n^{-\alpha}$, $\lambda_n \ge n^{-(1-3\alpha)/2}\sqrt{\log n}$ and a sufficiently large constant M. Note that $(1-3\alpha)/2 > 0$ for $\alpha < 1/3$. In particular for $\alpha = 1/6$, which corresponds to the order of optimal bandwidth for the estimation of the derivative of the regression function, we obtain $\delta_n \ge n^{-1/6}$ and $\lambda_n \ge n^{-1/4}\sqrt{\log n}$.

PROOF OF THEOREM 5.2. We proceed as in the proof of Theorem 5.1. As (5.3) and (5.4) are valid for any regression function, (5.6) also holds for the sequence m_n. By slight extensions of the argument used in the derivation of (2.5), in the first case, one can show that

$$EU_n(t_n) \ge -2h_n\tau_k^2\tau_g^2 f^2(t_n)\sup\{m_n'(t) : |t-t_n| \le h_n\}(1+o(1)). \tag{5.9}$$

Therefore, by the assumed conditions,

$$\liminf_{n\to\infty} \sqrt{n}EU_n(t_n)/b_n\sigma_n(t_n) > 1,$$

from which, it is easy to conclude that $S_{1,n}$ exceeds the RHS of (4.12), with probability tending to 1. The proof for the second case also follows from a similar computation. We omit the details. □

6. The time spent statistic. In this section, we consider the time spent statistic $S_{2,n}$, which is also known as the excursion time or the sojourn time in the literature. As the following theorem shows, with the choice

$$u_n = \sqrt{2\log\left(\frac{\sqrt{\lambda}(b-a)}{2\sqrt{\pi}h_n}\right)}$$

in (2.15), where λ is as defined in (4.8), a compound Poisson limit is obtained.

We apply Theorem 4.1 of Berman (1980) [see also Cramér and Leadbetter (1967), Chapter 12]. Let $U(t)$, $t \geq 0$, be a Gaussian process with stationary increments satisfying

$$U(0) = 0, \quad EU(t) = 0, \quad E(U(t) - U(s))^2 = 2(t-s)^2 \tag{6.1}$$

Let $J(x)$ be the function defined by

$$J(x) = \int_0^\infty P(\text{meas}\{s : U(s) - s^2 + y > 0\} > x)e^{-y}dy. \tag{6.2}$$

Let Z stand for a random variable whose Laplace transform is given by

$$\Omega(s) = \exp\left[-\int_0^\infty (1 - e^{-sx})dJ'(x)\right]. \tag{6.3}$$

In other words, Z has a compound Poisson distribution with $-J'(\cdot)$ as its compounding distribution.

THEOREM 6.1. *Under condition* (4.10),

$$h_n^{-1}\sqrt{\lambda \log((b-a)/h_n)}S_{2,n} \to_d Z, \tag{6.4}$$

where Z is as above.

PROOF. The basic idea behind the proof is to relate $S_{2,n}$ with similar functionals of the approximating Gaussian process $\xi_n(\cdot)$, Note that by Theorem 3.1, with large probability,

$$\begin{aligned} &h_n \,\text{meas}\{s \in T_n : \xi_n(s) \geq u_n + c\delta_n\} \\ &\quad\leq S_{2,n} \\ &\quad\leq h_n \,\text{meas}\{s \in T_n : \xi_n(s) \geq u_n - c\delta_n\}, \end{aligned} \tag{6.5}$$

where δ_n is as in (4.2) and c is a large constant. The probability of the event (6.5) can be made arbitrarily close to one for all sufficiently large n by choosing c sufficiently large.

We now show that the two extreme sides of (6.5) have the same weak limit Z. With $\alpha = 2$, $t = (b-a)/h_n$, $v = \sqrt{\lambda \log((b-a)/h_n)}$ and $u = u_n \pm c\delta_n$, observe that if $u_n\delta_n \to 0$, or equivalently, if (4.10) holds, then

$$u \sim \sqrt{2\log t},$$

$$P(u < \xi(0) < bu) \sim P(\xi(0) > u) \sim \frac{1}{u\sqrt{2\pi}}e^{-u^2/2} \sim \frac{1}{tv}, \quad b > 1$$

and

$$u^2(1-\rho(1/v)) \to 1;$$

in the last relation we have used the expansion (4.9).

The last three relations are sufficient for an application of Theorem 4.1 of Berman (1980). Therefore

$$\sqrt{\lambda \log((b-a)/h_n)}\,\mathrm{meas}\{s \in T_n : \xi_n(s) > u_n \pm c\delta_n\} \to_d Z. \tag{6.6}$$

Combining (6.5) and (6.6), the result follows. □

The test based on $S_{2,n}$ is also consistent against general alternatives.

THEOREM 6.2. *If* $nh_n^3/\log h_n^{-1} \to \infty$ *and* $m(\cdot)$ *is a regression function with* $m'(t_0) < 0$ *for some* $t_0 \in T$, *then*

$$P^n_{m,F,G}\Big\{h_n^{-1}\sqrt{\lambda \log((b-a)/h_n)}S_{2,n} > \tau_\alpha\Big\} \to 1, \tag{6.7}$$

where τ_α *is the* $(1-\alpha)$*th quantile of* Z.

PROOF. By the assumed continuity of $m'(t)$, there is an $\eta > 0$ such that $m'(t) < -\eta$ for all $t \in I_n = (t_0 - h_n, t_0 + h_n)$. We claim that

$$\sup_{t\in I_n} |U_n(t) - EU_n(t)| = O_p(n^{-1/2}h_n^{-1/2}). \tag{6.8}$$

Then, since clearly (2.5) holds uniformly (by considering $t_n \to t$), for all $t \in I_n$, $\sqrt{n}EU_n(t)/\hat\sigma_n(t)$ is uniformly bounded below by a multiple of $\sqrt{n}h_n^{3/2}$ on I_n. By (6.8), $\sqrt{n}|U_n(t)-EU_n(t)|/\hat\sigma_n(t)$ is uniformly bounded above by a constant, and is therefore of smaller order than $\sqrt{n}EU_n(t)/\hat\sigma_n(t)$ on I_n under the condition $nh_n^3 \to \infty$. The latter exceeds u_n if $nh_n^3/\log h_n^{-1} \to \infty$. Therefore, with a large probability, $S_{2,n}$ is bounded below by the length of the interval I_n, and so

$$h_n^{-1}\sqrt{\lambda \log((b-a)/h_n)}S_{2,n} \to_p \infty. \tag{6.9}$$

This shows that the power tends to one.

To prove that (6.8) holds, we use Theorem A.2 of the Appendix. The notations of Section 8 and the Appendix are followed. We can estimate the entropy numbers used in Theorem A.2 by the same arguments as in the proof Lemma 3.1. Specifically, the kernels $H_{n,t}((x_1,\varepsilon_1),(x_2,\varepsilon_2))$ defined in (5.1) of U_n, where $t \in I_n$, form a VC-class of functions [see, e.g., Section 2.6 of van der Vaart and Wellner (1996) for the definition of a VC-class] with envelope function

$$H_n = h_n^{-2}\{|x_1 - t_0| < 2h_n\}\{|x_2 - t_0| < 2h_n\}.$$

This shows that

$$\sup_Q N(\varepsilon\|H_n\|_{Q,2}, \{H_{n,t} : t \in I_n\}, L_2(Q)) \lesssim \varepsilon^{-V}$$

for some V and hence

$$\frac{1}{n}E\left(\int_0^\infty \log N(\varepsilon, \{H_{n,t} : t \in I_n\}, L_2(U_2^n))d\varepsilon\right)$$
$$\lesssim \frac{1}{n}\sqrt{E\left(h_n^{-2}\{|X_1 - t_0| < 2h_n\}\{|X_2 - t_0| < 2h_n\}\right)^2}\int_0^\infty \log\frac{1}{\varepsilon}d\varepsilon$$
$$= O(n^{-1}h_n^{-1}).$$

Next, by Lemma A.2, the projections

$$\bar{H}_{n,t}(x_1, \varepsilon_1) = \int\int H_{n,t}((x_1, \varepsilon_1), (x_2, \varepsilon_2))dF(x_2)dG(\varepsilon_2)$$

satisfy

$$\sup_Q N(\varepsilon\|\bar{H}_n\|_{Q,2}, \{\bar{H}_{n,t} : t \in I_n\}, L_2(Q)) \lesssim \varepsilon^{-2V}.$$

Thus

$$\frac{1}{\sqrt{n}}E\left(\int_0^\infty \sqrt{\log N(\varepsilon, \{\bar{H}_{n,t} : t \in I_n\}, L_2(U_1^n))}d\varepsilon\right)$$
$$\lesssim \frac{1}{\sqrt{n}}\sqrt{E\left(h_n^{-2}\{|X_1 - t_0| < 2h_n\}\int\{|X_2 - t_0| < 2h_n\}dF(x_2)\right)^2}\int_0^\infty \log\frac{1}{\varepsilon}d\varepsilon$$
$$= O(n^{-1/2}h_n^{-1/2}).$$

Finally Theorem A.2 gives (6.8). □

In the above, the particular choice of u_n was made in order to obtain a limiting distribution using the extreme value theory for stationary Gaussian processes. Smaller choices, such as $u_n = 0$ look very reasonable. However, we are not aware of the limit theory for such functionals of a stationary Gaussian process.

A more general class of test statistic may be considered by looking at

$$\int_T \left(\frac{\sqrt{n}U_n(t)}{\hat{\sigma}_n(t)} - u_n\right)_+^m dt,$$

where the "+" stands for the positive part and m is a nonnegative integer. The case $m = 0$ corresponds to $S_{2,n}$. For $m \geq 1$, the functional is Lipschitz continuous, and hence its asymptotic distribution may be found easily from that of $\int_{T_n}(\xi_n(s) - u_n)_+^m ds$, once the latter is available. Although it seems that limit theorems for such functionals of a stationary Gaussian process are not available in the literature yet, expressions for the mean and the variance are available; see (10.8.7) and (10.8.8) of Cramér and Leadbetter (1967). The case of $m = 1$ and $u_n = 0$ is particularly interesting, signifying the total area under the graph of the process $\sqrt{n}U_n(t)/\hat{\sigma}_n(t)$ above the level 0. The expressions for the mean and the variance also simplify in this case.

TABLE 1
Estimated power of the supremum test statistic

Sample size	**band-width**	**Critical value**	**Proportion of rejections for** $m_1(x)$	$m_2(x)$	$m_3(x)$	$m_4(x)$
50	0.2287	3.0341	0.027	0.429	0.107	0.026
100	0.1991	3.0510	0.030	0.735	0.509	0.080
200	0.1733	0.0725	0.034	0.964	0.945	0.757
500	0.1443	3.1062	0.038	1.000	1.000	1.000

7. A simulation study. In this section, we report the results of a small simulation study to get some idea about the behaviour of the error probabilities of the supremum statistic for finite sample sizes. We generated the independent variable X from the uniform distribution on $[0, 1]$ and considered four different regression functions: $m_1(x) = 0$, $m_2(x) = x(1-x)$, $m_3(x) = x + 0.415e^{-50x^2}$ and

$$m_4(x) = \begin{cases} 10(x-0.5)^3 - e^{-100(x-0.25)^2}, & \text{if } x < 0.5, \\ 0.1(x-0.5) - e^{-100(x-0.25)^2}, & \text{otherwise.} \end{cases}$$

Errors are generated from a normal distribution with mean 0 and standard deviation 0.1. The first regression function m_1 gives an idea about the level while m_2 is a standard non-monotone function. A function of the type m_3 was used by both Bowman, Jones and Gijbels (1998) and Hall and Heckman (1999). A function like m_4, having a sharp dip at $x = 0.25$ and a relatively flat portion on $x > 0.5$, was used by Hall and Heckman (2000) to demonstrate that the power of the test of Bowman, Jones and Gijbels (1998) does not tend to one. In all the cases, we take $T = [0.05, 0.95]$ where monotonicity is tested. We show results for sample sizes $n = 50, 100, 200$ and 500. We use the kernel $k(x) = 0.75(1-x^2)$ for $-1 < x < 1$ and 0 otherwise. The bandwidth $h = 0.5\,n^{-1/5}$ was used for sample size n. In each case, 1000 replications were generated to estimate the probability of rejection of the null hypothesis. We used a C program and ran on a Sun SPARC station with time as the seed for the random numbers. We did not study the dependence of the test on the bandwidth.

We observe that at $m_2(x)$ and $m_3(x)$, the test has a reasonably good power for $n = 100$ and very high power for $n \geq 200$. The alternative $m_4(x)$ is almost a monotone function except in a small neighborhood of $x = 0.25$ and so the power is small for moderate sample size like $n = 100$. The power does pick up however for bigger sample sizes. Actually, a smaller bandwidth does a better job for this function. The actual level [i.e., the power at $m_1(x)$] is however smaller than the asymptotic level 0.05. It is known that the convergence in (4.5) is slow, so perhaps it is not totally unexpected. The approximation to the distribution of the supremum of a Gaussian process may possibly be improved by considering more terms in its asymptotic expansion as done in Konakov and Piterbarg (1984). One may therefore expect to get better approximations to the actual critical value of the test statistic by replacing the asymptotic distribution by asymptotic expansions. However, since the level turns out to be on the conservative side, we do not pursue this possibility here.

8. Proof of Lemmas 3.1–3.4. Since $m(\cdot) \equiv 0$, the kernel $H_{n,t}$ in (5.1) simplifies to

$$H_{n,t}((x_1,\varepsilon_1),(x_2,\varepsilon_2)) = \operatorname{sign}(\varepsilon_2-\varepsilon_1)\operatorname{sign}(x_1-x_2)k_n(x_1-t)k_n(x_2-t), \tag{8.1}$$

and the kernel for the Hájek projection is given by (2.11). We also note that $EH_{n,t} = E\psi_{n,t} = 0$, where $\psi_{n,t}$ is given by (2.11).

Note that $U_n^{(2)}(t) = U_n(t) - \hat{U}_n(t)$ is a degree two degenerate U-process with kernel

$$\begin{aligned} &\varphi_{n,t}((x_1,\varepsilon_1),(x_2,\varepsilon_2)) \\ &\quad = \operatorname{sign}(\varepsilon_2-\varepsilon_1)\operatorname{sign}(x_1-x_2)k_n(x_1-t)k_n(x_2-t) \\ &\qquad - (1-2G(\varepsilon_1))\int \operatorname{sign}(x_1-w)k_n(w-t)\,dF(w)k_n(x_1-t) \\ &\qquad - (1-2G(\varepsilon_2))\int \operatorname{sign}(x_2-w)k_n(w-t)\,dF(w)k_n(x_2-t). \end{aligned} \tag{8.2}$$

For a metric space M with a distance d on it, let $N(\varepsilon, S, M)$, $\varepsilon > 0$, denote the ε-covering number of $S \subset M$, that is, the smallest integer m such that m balls of radius ε in M covers S. For a function f, $\|f\|_{Q,r}$ will stand for its $L_r(Q)$-norm $(\int |f|^r dQ)^{1/r}$ and $\|f\|_\infty$ will denote its supremum norm $\sup\{|f(x)| : x \in \mathbb{R}\}$. From here onward, $\lesssim$ will stand for an inequality up to a constant multiple.

PROOF OF LEMMA 3.1. We use Theorem A.1 in the Appendix applied with $r = 2$ and the functions $\mathscr{F} = \{f_t : t \in T\}$, where

$$f_t((x_1,\varepsilon_1),(x_2,\varepsilon_2)) = \operatorname{sign}(x_1-x_2)\operatorname{sign}(\varepsilon_1-\varepsilon_2)k_n(x_1-t)k_n(x_2-t).$$

These are contained in the product of the classes

$$\begin{aligned} \mathscr{F}_1 &= \left\{ k\left(\frac{x_1-t}{h_n}\right) : t \in T \right\}, \\ \mathscr{F}_2 &= \left\{ k\left(\frac{x_2-t}{h_n}\right) : t \in T \right\}, \\ \mathscr{F}_3 &= \left\{ \frac{1}{h_n^2}\operatorname{sign}(x_1-x_2)\operatorname{sign}(\varepsilon_1-\varepsilon_2)\{|x_1-x_2| < 2h_n\} \right\}, \end{aligned}$$

with envelopes $\|k\|_\infty$, $\|k\|_\infty$ and $h_n^{-2}\{|x_1-x_2| < 2h_n\}$ respectively.

Since $k(\cdot)$ is of bounded variation, the class of functions $x \mapsto k((x-t)/h_n)$ is the difference of two classes of the type $x \mapsto \varphi((x-t)/h_n)$ for φ monotone, and hence is the difference of two VC-classes of index less than or equal to 2 by Lemma 2.6.16 of van der Vaart and Wellner (1996). It then follows, by

Theorem 2.6.7 of van der Vaart and Wellner (1996) and the fact that the 2ε-covering numbers of the sum of the two classes are bounded by the product of the ε-covering numbers of the two classes, that

$$\sup\left\{N\left(\varepsilon,\left\{k\left(\frac{\cdot-t}{h_n}\right):t\in T\right\},L_2(Q)\right):\ Q\text{ is a probability}\right\}\lesssim\varepsilon^{-4}.$$

By Lemma A.1,

$$\sup\{N(\varepsilon\|h_n^{-2}\{|x_1-x_2|<2h_n\}\|_{Q,2},\mathscr{F}_1\mathscr{F}_2\mathscr{F}_3,L_2(Q)):\\ Q\text{ is a probability}\}\lesssim\varepsilon^{-8}.$$

By Theorem A.1 and the discussion following Theorem A.1 and Theorem A.2, it follows that

$$\begin{aligned}&nE\left(\sup_{t\in T}|U_2^nf_t-2U_1^n\Pi_1f_t|\right)\\&\quad\lesssim E\int_0^\infty\log N(\varepsilon,\mathscr{F},L_2(U_m^n))d\varepsilon\\&\quad\lesssim\int_0^1\log(\varepsilon^{-8})d\varepsilon\sqrt{E\left(\frac{1}{h_n^2}\{|X_1-X_2|<2h_n\}\right)^2}\\&\quad\lesssim h_n^{-3/2}.\end{aligned}$$

The result now follows because $U_n(t)-\hat{U}_n(t)=U_2^nf_t-2U_1^n\Pi_1f_t$. □

PROOF OF LEMMA 3.2. We use the local strong invariance principle—Theorem 1.1 of Rio (1994). Put $u=2(1-2G(\varepsilon))$ and

$$\phi_{n,t}(x,u)=\psi_{n,t}(x,\varepsilon)=u\int\operatorname{sign}(x-w)k_n(w-t)f(w)dwk_n(x-t),$$

and consider the process $\hat{U}_{n,t}$ as function of $(X_1,u_1),\dots,(X_n,u_n)$. Formally, Rio's theorem requires that X and u take values in $[0,1]$ and have a positive density there which is also continuous in the unit square. Since our F is supported on compact intervals where the density is positive and continuous inside the intervals and u has a uniform distribution, we may achieve this by simple affine transformations. Moreover, applying the same affine transformation to t that applies to X, we can also preserve the structure of a location family for the kernel $k(\cdot)$. Finally, an inverse transformation brings back everything to our original domain. So we may and do assume that Rio's invariance applies to the original domain.

Now, since integration over x absorbs a factor h_n^{-1}, it is easy to see that

$$\int\int\left(\left|\frac{\partial\phi_{n,t}}{\partial x}(x,u)\right|+\left|\frac{\partial\phi_{n,t}}{\partial u}(x,u)\right|\right)dxdu=O(h_n^{-1})\tag{8.3}$$

uniformly in t. This implies that $\{h_n\phi_{n,t} : t \in T\}$ satisfies the UBV condition of Rio (1994). To check the LUBV condition there, let C be a square with sides of length a. Then for some constant A,

$$\int\int_{(x,u)\in C}\left|\frac{\partial\phi_{n,t}}{\partial x}(x,u)\right|dxdu + \int\int_{(x,u)\in C}\left|\frac{\partial\phi_{n,t}}{\partial u}(x,u)\right|dxdu \leq Ah_n^{-1}\left\{h_n^{-1}\min(ah_n, a^2) + \min(ah_n, a^2)\right\}, \tag{8.4}$$

which is bounded by $Ah_n^{-1}a$. This shows that the LUBV condition holds for the class $\{h_n\phi_{n,t} : t \in T\}$ [see the proof of Lemma 4.1 of Rio (1994)].

We now show that

$$\sup\left\{N\left(\varepsilon, \left\{h_n\phi_{n,t}(\cdot,\cdot) : t \in T\right\}, L_1(Q)\right) : \ Q \text{ is a probability}\right\} \lesssim \varepsilon^{-V}, \tag{8.5}$$

for some $V > 0$.

First we observe that $h_n\phi_{n,t}(x,u)$ is bounded by a constant uniformly in $t \in T$, and so can be enveloped by a constant function. The function $h_n\phi_{n,t}(x,u)$ is obtained by averaging

$$\eta_t(x, w, u) = h_n u \operatorname{sign}(x-w)k_n(x-t)k_n(w-t)$$

over w with respect to the probability distribution F. The family $\{k((\cdot - t)/h_n) : t \in T\}$ is uniformly bounded and is a difference of two VC-classes of index 2 and hence by Theorem 2.6.7 of van der Vaart and Wellner (1996),

$$\sup\left\{N\left(\varepsilon, \left\{k\left(\frac{\cdot - t}{h_n}\right) : t \in T\right\}, L_1(Q)\right) : \ Q \text{ is a probability}\right\} \lesssim \varepsilon^{-2}.$$

The class $\{\eta_t : t \in T\}$ has envelope a constant multiple of $h_n^{-1}\{|x-w| < 2h_n\}$. Therefore by Lemma A.1,

$$\sup\left\{N\left(\varepsilon h_n^{-1}\|\{|x-w| < 2h_n\}\|_{Q,1}, \{\eta_t : t \in T\}, L_1(Q)\right) : Q \text{ is a probability}\right\} \lesssim \varepsilon^{-4}.$$

An application of Lemma A.2 now proves (8.5).

Therefore, by Rio's theorem (and switching back to the original variable ε), it easily follows that there exists a sequence of centered Gaussian Processes G_n with the stated covariance satisfying (3.5). Further, the processes has continuous sample paths, where continuity is with respect to the L_2-metric on the class of functions $\{\psi_{n,t} : t \in T\}$. Now, as is easily seen, the mapping $t \mapsto \psi_{n,t}$ is continuous from T to $L_2(P)$, and so the processes $G_n(\cdot)$ has continuous sample paths with respect to the usual metric on T. □

REMARK 8.1. Letting $\hat{U}_{n,h}$ stand for $\hat{U}_n$ when the bandwidth is h instead of h_n, one may actually make a stronger claim: If h_n satisfies (3.3), then for any $0 < \beta_1 \leq \beta_2 < \infty$, there exist centered Gaussian processes

$G_{n,h}$ with continuous sample paths and covariance $E(G_{n,h}(t_1)G_{n,h}(t_2)) = E(\tilde{\psi}_{h,t_1}(X,\varepsilon)\tilde{\psi}_{h,t_2}(X,\varepsilon))$ such that

$$\sup_{t\in T}\ \sup_{\beta_1 h_n\le h\le\beta_2 h_n} |\sqrt{n}\hat{U}_{n,h}(t) - G_{n,h}(t)| = O(n^{-1/4}h_n^{-1}\sqrt{\log n}), \tag{8.6}$$

where $\tilde{\psi}_{h,t}$ is obtained by replacing h_n by h in (2.11). This additional flexibility allows us to handle some data-driven choices of bandwidths as well. We, however, do not pursue this possibility.

REMARK 8.2. Because Rio's construction of the approximating Gaussian process is based on a more basic construction of Komlós, Major and Tusnády's (1975), one may actually assert the following: On a suitable probability space, there exists a copy $(X_1^*,\varepsilon_1^*),\dots,(X_n^*,\varepsilon_n^*)$ of the original random variables $(X_1,\varepsilon_1),\dots,(X_n,\varepsilon_n)$ such that the copy of $\hat{U}_n(t)$ admitting the strong approximation is obtained by replacing $(X_1,\varepsilon_1),\dots,(X_n,\varepsilon_n)$ by $(X_1^*,\varepsilon_1^*),\dots,(X_n^*,\varepsilon_n^*)$ in (2.10). Therefore, we can also copy $U_n(t)$ on this new probability space so that the conclusion of Lemma 3.1 is respected.

REMARK 8.3. Let $\mathscr{S}_n$ denote the class of functions $\{g_{n,t} : t \in T\}$, where $g_{n,t}(x,\varepsilon) = \psi_{n,t}(x,\varepsilon)/\sigma_n(t)$. Let $\mathscr{S}_n'$ stand for the class of functions $\{\tilde{g}_{n,t} : t \in T\}$, where

$$\tilde{g}_{n,t}(x,\varepsilon) = \frac{\sqrt{3}(1-2G(\varepsilon))\int \operatorname{sign}(x-w)k_n(w-t)\,dw\,k_n(x-t)}{\left(\int\left(\int \operatorname{sign}(z-w)k_n(w-t)\,dw\right)^2 k_n^2(z-t)\,dz\right)^{1/2} f^{1/2}(x)}. \tag{8.7}$$

The Gaussian process $\{G_n(t)/\sigma_n(t) : t \in T\}$ may be thought of as a Brownian bridge on $\mathscr{S}_n$. We can actually increase the domain of definition to $\mathscr{S}_n \cup \mathscr{S}_n'$. To do that, we might well have started with the index set $\mathscr{S}_n \cup \mathscr{S}_n'$ in Rio's invariance principle. Arguments similar to those used in the proof of Lemma 3.2 will also show that $\mathscr{S}_n'$ has all the properties necessary for an application of Theorem 1.1 of Rio (1994), and so the same is true for $\mathscr{S}_n \cup \mathscr{S}_n'$. Therefore, we have a Brownian bridge, say $\{B_n(g) : g \in \mathscr{S}_n \cup \mathscr{S}_n'\}$ [i.e., $EB_n(g) = 0$, $E(B_n(g_1)B_n(g_2)) = \operatorname{cov}(g_1,g_2)$ for all $g, g_1, g_2 \in \mathscr{S}_n \cup \mathscr{S}_n'$] with continuous sample paths with respect to the L_2-metric such that $G_n(t) = \sigma_n(t)B_n(\psi_{n,t})$ satisfies the conclusion of Lemma 3.2. This extension will be used in the proof of Lemma 3.4 to get hold of the claimed stationary Gaussian process on the same probability space.

PROOF OF LEMMA 3.3. Since T is compact, part (a) amounts to showing that for any convergent sequence $t_n \to t$, we have

$$h_n\sigma_n^2(t_n) \to \tfrac{4}{3}f^3(t)\int q^2(x)k^2(x)\,dx.$$

This follows by replacing t by t_n in (2.12) and applying the dominated convergence theorem. Since the limit on the RHS of (a) is always positive, part (b) also follows.

Now to prove part (c). Set

$$\tilde{f}_t(x, y, z) = \tfrac{4}{3}\operatorname{sign}(x-y)\operatorname{sign}(x-z)k_n^2(x-t)k_n(y-t)k_n(z-t)$$

and let f_t be $\tilde{f}_t$ symmetrized in (x, y, z). Then $\hat{\sigma}_n^2(t) = U_3^n f_t$ and $\sigma_n^2(t) = Pf_t$. We shall show that

$$E(\sup\{|\hat{\sigma}_n^2(t) - \sigma_n^2(t)| : t \in T\}) \lesssim n^{-1/2}h_n^{-2}.$$

As argued in the proof of Lemma 3.1, we have

$$\sup_Q N\left(\varepsilon, \left\{k^2\left(\frac{\cdot - t}{h_n}\right) : t \in T\right\}, L_2(Q)\right) \lesssim \varepsilon^{-4}$$

and similarly for the functions $y \mapsto k((y-t)/h_n)$ and $z \mapsto k((z-t)/h_n)$. The functions $\tilde{f}_t$, $t \in T$, are contained in the product of the classes

$$\mathscr{F}_1 = \left\{k^2\left(\frac{x-t}{h_n}\right) : t \in T\right\},$$

$$\mathscr{F}_2 = \left\{k\left(\frac{y-t}{h_n}\right) : t \in T\right\},$$

$$\mathscr{F}_3 = \left\{k\left(\frac{z-t}{h_n}\right) : t \in T\right\},$$

$$\mathscr{F}_4 = \left\{\frac{4}{3h_n^4}\operatorname{sign}(x-y)\operatorname{sign}(x-z)\{\|(x, y, z) - D\| < 2h_n\}\right\},$$

where D stands for the diagonal in $\mathbb{R}^3$. These classes have envelopes $\|k^2\|_\infty$, $\|k\|_\infty$, $\|k\|_\infty$ and $\frac{4}{3h_n^4}\{\|(x, y, z) - D\| < 2h_n\}$. By Lemma A.1,

$$\sup_Q N\left(\varepsilon\frac{4}{3h_n^4}\|\{\|(x, y, z) - D\| < 2h_n\}\|_{Q,2}, \mathscr{F}_1\mathscr{F}_2\mathscr{F}_3\mathscr{F}_4, L_2(Q)\right) \lesssim \varepsilon^{-12}.$$

By Lemma A.2, the projections $\Pi_1\tilde{f}_t$ also have polynomial covering numbers relative to the envelope a multiple of

$$x \mapsto h_n^{-4}\int\int\{\|(x, y, z) - D\| < 2h_n\}dP(y)dP(z) \lesssim h_n^{-2}.$$

The projections $\Pi_2\tilde{f}_t$ similarly have polynomial covering numbers relative to the envelope a multiple of

$$(x, y) \mapsto h_n^{-4}\int\{\|(x, y, z) - D\| < 2h_n\}dP(z) \lesssim h_n^{-3}\{|x-y| < 2h_n\}.$$

It follows by Theorem A.2 that

$$\begin{aligned} &E(\sup\{|\hat{\sigma}_n^2(t) - \sigma_n^2(t)| : t \in T\}) \\ &\quad\lesssim n^{-1/2}\sqrt{(h_n^{-2})^2} + n^{-1}\sqrt{E(h_n^{-3}\{|x-y| < 2h_n\})^2} \\ &\qquad + n^{-3/2}\sqrt{E(h_n^{-4}\{\|(x, y, z) - D\| < 2h_n\})^2} \\ &\quad\lesssim \frac{1}{\sqrt{n}h_n^2} + \frac{1}{nh_n^{5/2}} + \frac{1}{n^{3/2}h_n^3}, \end{aligned}$$

which is $O(n^{-1/2}h_n^{-2})$ under (3.3). $\square$

Proof of Lemma 3.4. Put $\tilde{\xi}_n(t) = B_n(\tilde{g}_{n,t})$, where B_n is the Brownian bridge explained in Remark 8.3 and $\tilde{g}_{n,t}(\cdot)$ is as defined in (8.7). We note that $\gamma_n(t) = G_n(t)/\sigma_n(t) - \tilde{\xi}_n(t)$ is also a mean zero Gaussian process with

$$E(\gamma_n(t_1)\gamma_n(t_2)) = E(g_{n,t_1} - \tilde{g}_{n,t_1})(g_{n,t_2} - \tilde{g}_{n,t_2}). \tag{8.8}$$

First, we claim that

$$\sup_{t\in T} E(\gamma_n(t))^2 = O(h_n^2), \tag{8.9}$$

that is, the L_2-diameter of $\gamma_n(\cdot)$ is $O(h_n)$. Note that

$$\begin{aligned}
&\int\int (g_{n,t}(x,\varepsilon) - \tilde{g}_{n,t}(x,\varepsilon))^2 dG(\varepsilon)dF(x)\\
&= \int \left[\frac{\int \operatorname{sign}(x-w)k_n(w-t)f(w)dw k_n(x-t)f^{1/2}(x)}{\left\{\int \left(\int \operatorname{sign}(z-w)k_n(w-t)f(w)dw\right)^2 k_n^2(z-t)f(z)dz\right\}^{1/2}} \right.\\
&\qquad \left. - \frac{\int \operatorname{sign}(x-w)k_n(w-t)dw k_n(x-t)}{\left\{\int \left(\int \operatorname{sign}(z-w)k_n(w-t)dw\right)^2 k_n^2(z-t)dz\right\}^{1/2}} \right]^2 dx,
\end{aligned} \tag{8.10}$$

where, since F has a compact support, the integral is actually over a compact interval. To prove (8.9), it suffices to show that the RHS of (8.10) is $O(h_n^2)$ uniformly in t. Writing $x = t + h_n u$, $z = t + h_n v$ and $w = t + h_n s$, we can write the RHS of (8.10) as

$$\begin{aligned}
&\int \left[\frac{\int \operatorname{sign}(u-s)k(s)f(t+h_n s)ds k(u)f^{1/2}(t+h_n u)}{\left\{\int \left(\int \operatorname{sign}(v-s)k(s)f(t+h_n s)ds\right)^2 k^2(v)f(t+h_n v)dv\right\}^{1/2}} \right.\\
&\qquad \left. - \frac{\int \operatorname{sign}(u-s)k(s)f(t)ds k(u)f^{1/2}(t)}{\left\{\int \left(\int \operatorname{sign}(v-s)k(s)f(t)ds\right)^2 k^2(v)f(t)dv\right\}^{1/2}} \right]^2 du.
\end{aligned}$$

Since $k(\cdot)$ has also a compact support and f is Lipschitz continuous, the denominators in the two terms inside the integral in the last display differ by $O(h_n)$. Similarly the numerators also differ by $O(h_n)$ uniformly in u, and so do the ratios. Since u ranges only in $[-1, 1]$, it follows that the integral is also $O(h_n^2)$. This proves (8.9).

Next, we claim that there exists a constant C such that for all $t_1, t_2 \in T$,

$$E(g_{n,t_1} - \tilde{g}_{n,t_1} - g_{n,t_2} + \tilde{g}_{n,t_2})^2 \le C h_n^{-2}|t_1 - t_2|^2. \tag{8.11}$$

After a change of variable $x = t_1 + h_n u$, we may rewrite the LHS of (8.11) as

$$\int \left[\frac{\int \operatorname{sign}(u-s)k(s)f(t_1+h_n s)ds k(u)f^{1/2}(t_1+h_n u)}{\left\{\int \left(\int \operatorname{sign}(v-s)k(s)f(t_1+h_n s)ds\right)^2 k^2(v)f(t_1+h_n v)dv\right\}^{1/2}} \right.$$

$$
\begin{aligned}
&- \int \frac{\int \operatorname{sign}(u-s)k(s)ds k(u)}{\left\{\int \left(\int \operatorname{sign}(v-s)k(s)ds\right)^2 k^2(v)dv\right\}^{1/2}} \\
&- \int \frac{\int \operatorname{sign}(u-s)k(s+\frac{t_1-t_2}{h_n})f(t_1+h_ns)ds k(u+\frac{t_2-t_1}{h_n})f^{1/2}(t_1+h_nu)}{\left\{\int \left(\int \operatorname{sign}(v-s)k(s)f(t_2+h_ns)ds\right)^2 k^2(v)f(t_2+h_nv)dv\right\}^{1/2}} \\
&\qquad\qquad + \left. \int \frac{\int \operatorname{sign}(u-s)k(s+\frac{t_1-t_2}{h_n})ds k(u+\frac{t_2-t_1}{h_n})}{\left\{\int \left(\int \operatorname{sign}(v-s)k(s)ds\right)^2 k^2(v)dv\right\}^{1/2}} \right]^2 du.
\end{aligned}
$$

We pair the first term with the third and the second with the fourth. The third term is similar to the first term except that some of the arguments inside the functions $k(\cdot)$ and $f(\cdot)$ differ by $h_n^{-1}(t_1 - t_2)$ or $(t_1 - t_2)$. By the Lipschitz continuity, the terms differ by $O(h_n^{-1}|t_1 - t_2|)$ uniformly in u. Similarly, the fourth term differs from the second by $O(h_n^{-1}|t_1 - t_2|)$ uniformly in u. Thus the integral is $O(h_n^{-2}|t_1 - t_2|^2)$ and (8.11) is obtained.

Therefore,

$$
N(\varepsilon, \{g_{n,t} - \tilde{g}_{n,t} : t \in T\}, L_2(P)) \le N\left(\frac{h_n\varepsilon}{\sqrt{C}}, T, |\cdot|\right) \lesssim \frac{1}{h_n\varepsilon}.
$$

Applying Corollary 2.2.8 of van der Vaart and Wellner (1996), we obtain

$$
\begin{aligned}
E\left(\sup_{t\in T}|\gamma_n(t)|\right) &\lesssim \int_0^\infty \sqrt{\log N(\varepsilon, \{g_{n,t} - \tilde{g}_{n,t} : t \in T\}, L_2(P))}d\varepsilon \\
&\lesssim \int_0^{O(h_n)} \sqrt{\log \frac{1}{h_n\varepsilon}}d\varepsilon \\
&= O(h_n\sqrt{\log h_n^{-1}}).
\end{aligned}
$$

Now choose $\xi_n(s) = \tilde{\xi}_n(a + h_ns)$, $s \in T_n$, as the desired process. □

APPENDIX

Suppose we have i.i.d. observations $X_1, X_2, \dots, X_n$ that take values in a sample space $\mathcal{X}$ and have probability law $\mathbb{P}$. Let $f : \mathcal{X}^m \to \mathbb{R}$ be symmetric in its arguments and let $U_m^n f$ be the U-statistic with kernel f based on $(X_1, X_2, \dots, X_n)$. Let $\Pi_c f(X_1, \dots, X_c)$ stand for the projection of $f(X_1, \dots, X_m)$ onto the space of all functions of $(X_1, \dots, X_c)$ that are orthogonal to every function of less than c arguments. Then

$$
U_c^n \Pi_c f = \frac{1}{\binom{n}{c}} \sum_{(i_1,\dots,i_c)\in\mathcal{J}_c^n} \Pi_c f(X_{i_1}, \dots, X_{i_c}),
$$

where $\mathcal{J}_c^n$ denotes the set of all combinations of c numbers from $\{1, \dots, n\}$.

Note that

$$\begin{aligned}\Pi_c f(X_1,\dots,X_c) = {} & E(f(X_1,\dots,X_m)|X_1,\dots,X_c)\\ & - \sum_{\{i_1,\dots,i_{c-1}\}\in\mathscr{J}_{c-1}^m} E(f(X_1,\dots,X_m)|X_{i_1},\dots,X_{i_{c-1}})\\ & + \sum_{\{i_1,\dots,i_{c-2}\}\in\mathscr{J}_{c-2}^m} E(f(X_1,\dots,X_m)|X_{i_1},\dots,X_{i_{c-2}})\\ & - \cdots\\ & +(-1)^{c-1}\sum_{\{i_1\}\in\mathscr{J}_1^m} E(f(X_1,\dots,X_m)|X_{i_1})\\ & +(-1)^c E(f(X_1,\dots,X_m))\end{aligned}$$

and $U_c^n\Pi_c f$ is the projection of $U_m^n f$ onto the set of all functions that are sums of functions of at most c variables from $X_1,\dots,X_n$ that are orthogonal to all functions of less than c variables. Also,

$$f(X_1,\dots,X_m) = \sum_{c=0}^{m}\sum_{A\in\mathscr{J}_c^m}\Pi_c f(X_i : i\in A)$$

and

$$U_m^n f = \sum_{c=0}^{m}\binom{m}{c}U_c^n\Pi_c f.$$

We can also consider U_m^n as the random discrete measure putting mass $\binom{n}{m}^{-1}$ on each of the points $(X_{i_1},\dots,X_{i_m})\in\mathfrak{X}^m$, $\{i_1,\dots,i_m\}\in\mathscr{J}_m^n$.

Let $\mathscr{F}$ be a class of functions. The following two theorems are implicit in Arcones and Giné (1993). See also Arcones and Giné (1995).

THEOREM A.1. *There exists a constant C depending only on m such that*

$$\begin{aligned}E\left(\sup\left\{\left|U_m^n f - \sum_{c=0}^{r-1}\binom{m}{c}U_c^n\Pi_c f\right| : f\in\mathscr{F}\right\}\right)\\ \le Cn^{-r/2}E\int_0^\infty(\log N(\varepsilon,\mathscr{F},L_2(U_m^n)))^{r/2}d\varepsilon.\end{aligned}$$

THEOREM A.2. *There exists a constant C depending only on m such that*

$$\begin{aligned}E\left(\sup\{|U_m^n f - EU_m^n f| : f\in\mathscr{F}\}\right)\\ \le C\sum_{c=1}^{m}n^{-c/2}E\int_0^\infty(\log N(\varepsilon,\Pi_c\mathscr{F},L_2(U_c^n)))^{c/2}d\varepsilon.\end{aligned}$$

In each of the integrals, the covering number $N(\varepsilon,\Pi_c\mathscr{F},L_2(U_c^n))$ may also be replaced by $N(\varepsilon,\mathscr{F},L_2(U_m^n))$.

Let F be an envelope for the class $\mathscr{F}$. Then the integrals above can be further bounded by means of the inequality

$$\int_0^\infty (\log N(\varepsilon, \mathscr{F}, L_2(U_m^n))^{c/2} d\varepsilon$$
$$\leq \|F\|_{U_m^n,2} \int_0^1 (\sup\{\log N(\varepsilon\|F\|_{Q,2}, \mathscr{F}, L_2(Q)) : Q \text{ discrete}\})^{c/2} d\varepsilon,$$

where $\|F\|_{Q,2}$ stands for the $L_2(Q)$-norm of F. Further,

$$E\|F\|_{U_m^n,2} \leq \sqrt{E(U_m^n F)^2} \leq \sqrt{EF^2(X_1, \ldots, X_m)}.$$

The following Lemma is implicit in Pollard [(1990), Chapter 5] and van der Vaart and Wellner [(1996), Section 2.10.3].

LEMMA A.1. *Let $\mathscr{F}$ and $\mathscr{G}$ be classes of functions with envelopes F and G respectively. If $\mathscr{F}\mathscr{G}$ stands for the class of pointwise products of functions from $\mathscr{F}$ and $\mathscr{G}$, then for $1 \leq r < \infty$,*

$$\begin{aligned} &\sup\nolimits_Q N(2\varepsilon\|FG\|_{Q,r}, \mathscr{F}\mathscr{G}, L_r(Q)) \\ &\quad \leq \sup\nolimits_Q N(\varepsilon\|F\|_{Q,r}, \mathscr{F}, L_r(Q)) \sup\nolimits_Q N(\varepsilon\|G\|_{Q,r}, \mathscr{G}, L_r(Q)), \end{aligned} \tag{A.1}$$

where the supremum is over all discrete probability measures.

The functions $\Pi_c f$ arise from f by taking linear combinations of functions obtained by integrating out variables from f. To control the entropies of the classes $\Pi_c \mathscr{F}$, the following Lemma may be useful. It is in the spirit of Lemma 5 in Sherman (1994). However, since we are unable to find it in its present form in the literature, its short proof is also included.

Let $\mathscr{F}$ be a class of functions $f : \mathfrak{X} \times \mathfrak{Y} \to \mathbb{R}$ with envelope F and R a fixed probability measure on $\mathfrak{Y}$. For a given $f \in \mathscr{F}$, let $\bar{f} : \mathfrak{X} \to \mathbb{R}$ be the function

$$\bar{f}(x) = \int f(x, y) dR(y).$$

Set $\bar{\mathscr{F}} = \{\bar{f} : f \in \mathscr{F}\}$.

Note that $\bar{F}$ is an envelope of $\bar{\mathscr{F}}$.

LEMMA A.2. *For any $r, s \geq 1$,*

$$\sup_Q N(2\varepsilon\|\bar{F}\|_{Q,r}, \bar{\mathscr{F}}, L_r(Q)) \lesssim \sup_Q N(\varepsilon^r\|F\|_{Q\times R,s}, \mathscr{F}, L_s(Q \times R)),$$

where the supremum ranges over all probability measures.

PROOF. By Jensen's inequality,

$$E_Q|\bar{f} - \bar{g}| \leq E_{Q\times R}|f - g|.$$

Hence $N(\varepsilon, \bar{\mathscr{F}}, L_1(Q)) \leq N(\varepsilon, \mathscr{F}, L_1(Q \times R))$ for every $\varepsilon > 0$ and Q. Because $\|\bar{F}\|_{Q,1} = \|F\|_{Q\times R,1}$, we obtain

$$\sup_Q N(\varepsilon\|\bar{F}\|_{Q,1}, \bar{\mathscr{F}}, L_1(Q)) \leq \sup_Q N(\varepsilon\|F\|_{Q\times R,1}, \mathscr{F}, L_1(Q \times R)).$$

The RHS does not decrease if the L_1-norm is replaced by the L_s-norm ($s \geq 1$), by Problem 2.10.4 of van der Vaart and Wellner (1996). By the second part of the proof of Theorem 2.6.7 of van der Vaart and Wellner (1996), we also have that

$$\sup_Q N(2\varepsilon\|G\|_{Q,r}, \mathscr{G}, L_r(Q)) \leq \sup_Q N(\varepsilon^r\|G\|_{Q,1}, \mathscr{G}, L_1(Q))$$

for any class of functions $\mathscr{G}$. The result follows. $\square$

It follows, for instance, if $\mathscr{F}$ is a VC-class of index V, then the class $\bar{\mathscr{F}}$ has polynomial covering numbers relative to the envelope $\bar{F}$ in that

$$\sup_Q N(\varepsilon\|\bar{F}\|_{r,Q}, \bar{\mathscr{F}}, L_r(Q)) \lesssim \varepsilon^{-r(V-1)}.$$

Acknowledgments. We are grateful to the referees and the associate editor for several helpful remarks and for pointing out some important references. We would also like to thank Professors Adrian Bowman, Peter Hall and Michael Woodroofe for giving us the copies of their papers.

REFERENCES

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542.

ARCONES, M. A. and GINÉ, E. (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Stochastic Process Appl.* **58** 217–245.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions*. Wiley, London.

BERMAN, S. M. (1980). A compound Poisson limit for stationary sums, and sojourns of Gaussian processes. *Ann. Probab.* **8** 511–538.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

BOWMAN, A. W., JONES, M. C. and GIJBELS, I. (1998). Testing monotonicity of regression. *J. Comput. Graph. Statist.* **7** 489–500.

CASADY, R. J. and CRYER, J. D. (1976). Monotone percentile regression. *Ann. Statist.* **4** 532–541.

CRAMÉR, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes.* Wiley, New York.

CRYER, J. D., ROBERTSON, T., WRIGHT, F. T. and CASADY, R. J. (1972). Monotone median regression. *Ann. Math. Statist.* **43** 1459–1469.

DÜMBGEN, L. and SPOKOINY, V. G. (1998). Optimal nonparametric testing of qualitative hypotheses. Preprint.

HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39.

HANSON, D. L., PLEDGER, G. and WRIGHT, F. T. (1973). On consistency in monotonic regression. *Ann. Statist.* **1** 401–421.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

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

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's and the sample df. I. *Z. Warsch. Verw. Gebiete* **32** 111–131.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

MAMMEN, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.* **19** 724–740.

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

NOLAN, D. and POLLARD, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15** 780–799.

RAMSAY, J. O. (1988). Monotone regression splines in action (with discussion). *Statist. Sci.* **3** 425–461.

RAMSAY, J. O. (1998). Estimating smooth monotone functions. *J. Roy. Statist. Soc. Ser. B* **60** 365–375.

RIO, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21–45.

ROBERTSON, T. and WEGMAN, E. J. (1978). Likelihood ratio tests for order restrictions in exponential families. *Ann. Statist.* **6** 485–505.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, Chichester.

SCHLEE, W. (1982). Nonparametric tests of the monotony and convexity of regression. In *Nonparametric Statistical Inference II* (B. V. Gnedenko, M. L. Puri and I. Vincze, eds.) 823–836. North-Holland, Amsterdam.

SHERMAN, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22** 439–459.

SILVERMAN, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B* **43** 97–99.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

VAN ZWET, W. R. (1984). A Berry-Essen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete* **66** 425–440.

WOODROOFE, M. and SUN, J. (1999). Testing uniformity versus a monotone density. *Ann. Statist.* **27** 338–360.

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

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

WRIGHT, F. T. (1984). The asymptotic behavior of monotone percentile regression estimates. *Canad. J. Statist.* **12** 229–236.

S. GHOSAL
A. W. VAN DER VAART
DIVISION OF MATHEMATICS AND COMPUTER SCIENCE
FREE UNIVERSITY
DE BOELELAAN 1081A
1081 HV AMSTERDAM
THE NETHERLANDS

A. SEN
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF HYDERABAD
HYDERABAD 500046
INDIA