The Annals of Statistics
1987, Vol. 15, No. 2, 491–512

LOCAL ASYMPTOTICS FOR LINEAR RANK STATISTICS WITH ESTIMATED SCORE FUNCTIONS

BY GEORG NEUHAUS

University of Hamburg

Modified versions of linear rank statistics $S_N(\varphi)$ with score function φ are studied for the two-sample testing problem of randomness. Depending on the unknown underlying alternatives, some score function $\varphi = b$ is known to be approximately optimal. Behnen, Neuhaus and Ruymgaart (1983) proposed estimating b by some estimator $\hat{b}_N$ obtained by the kernel method based on ranks and used the quadratic rank statistic $S_N(\hat{b}_N)$ for testing the hypothesis of randomness H_0 versus the omnibus alternative. In the present paper the behavior of the corresponding test as well as that of a variant adapted to stochastically larger alternatives is studied by means of local asymptotic results with bandwidth of the kernel fixed. It turns out that the present asymptotics fit finite sample Monte Carlo results much better than previous results do and is able to explain to a large extent the power behavior of the proposed tests. Critical values as well as recommendations for the use of the tests in practice are included.

1. Introduction and statistical background. In the papers of Behnen and Neuhaus (1983) (BN), Behnen, Neuhaus and Ruymgaart (1983) (BNR) and Behnen and Husková (1984) (BH), some attempts have been made to extend the range of sensitivity of linear rank tests to larger classes of alternatives. The present paper concludes these series to a certain extent and so it seems worthwhile to summarize the statistical background and some new aspects of the philosophy behind all these papers. As in the above three papers, for the sake of brevity let us concentrate on the two-sample problem for testing differences in location, although the two-sample problem for testing differences in dispersion, the one-sample symmetry problem and the bivariate independence problem and some other testing problems have been treated in some not yet published manuscripts by Behnen and the author, too.

Let $X_1, \dots, X_m, Y_1, \dots, Y_n$ be independent real random variables, the X's being distributed according to a continuous distribution function (df) F, shortly $X_i \sim F$, $\forall i$, and, similarly, $Y_j \sim G$, $\forall j$. Writing $X_{m+j} = Y_j$, $j = 1, \dots, n$, let R_i be the rank of X_i in the pooled sample $(X_1, \dots, X_N)$, $N = m + n$.

Under the (*null*) *hypothesis of randomness* H_0,

$$H_0\colon F = G, \tag{1.1}$$

the rank vector $R = (R_1, \dots, R_N)$ is uniformly distributed on the $N!$ permutations of $(1, \dots, N)$, independently of $F = G$, i.e., each test based on ranks is distribution free under H_0. Various alternatives of different generality have been

Received December 1984; revised August 1986.
AMS 1980 *subject classifications*. Primary 62E20; secondary 62G10.
Key words and phrases. Rank tests, estimated scores, omnibus and one-sided tests, contiguous alternatives, asymptotic distribution and power.

considered in the past, e.g., the *omnibus alternative*

$$H_1^0\colon F \neq G, \tag{1.2}$$

the *stochastically larger alternative*

$$H_1\colon F \leq G, \qquad F \neq G, \tag{1.3}$$

and the classical *shift alternative*

$$H_1(F_0)\colon (F,G) \in H_1, \qquad F, G \in \{F_0(\cdot - t)\colon t \in \mathbb{R}\}, \tag{1.4}$$

where F_0 is some known continuous df.

Most linear rank statistics for testing H_0 versus one of these alternatives may be written as

$$S_N(b) = \sum_{i=1}^{N} c_{Ni}\tilde{b}_N(R_i), \tag{1.5}$$

where b is some *score function* from $L_2(0,1) = \{f\colon (0,1) \to \mathbb{R}\colon \int_0^1 f^2\, d\lambda < \infty\}$, $\tilde{b}_N(i) = N\int_0^1 b(u)1((i-1)/N \leq u \leq i/N)\lambda(du)$, $1 \leq i \leq N$, and

$$c_{Ni} = \left(\frac{mn}{N}\right)^{1/2} \begin{cases} m^{-1}, & \text{as } 1 \leq i \leq m, \\ -n^{-1}, & \text{as } m+1 \leq i \leq N. \end{cases} \tag{1.6}$$

For the classical shift problem H_0 versus $H_1(F_0)$, where F_0 has an absolutely continuous density f_0 with derivative f_0', it is well known that the test rejecting for large values of $S_N(\varphi(\cdot, f_0))$ with

$$\varphi(u, f_0) = -\frac{f_0'}{f_0}(F_0^{-1}(u)), \qquad 0 < u < 1, \tag{1.7}$$

is optimal in a local asymptotic sense, see Hájek and Šidák (1967). The drawback of this optimality result is that it is true only for $H_1(F_0)$. For other shift alternatives $H_1(F_1)$ the asymptotic power of this test depends on the size of $c(f_0, f_1) = \langle\varphi(\cdot, f_0), \varphi(\cdot, f_1)\rangle(\|\varphi(\cdot, f_0)\|\,\|\varphi(\cdot, f_1\|)^{-1}$, i.e., on the cosine of the angle between $\varphi(\cdot, f_0)$ and $\varphi(\cdot, f_1)$ being maximal (optimal) for $F_0 = F_1$ but may be very small for $F_0 \neq F_1$. Here $\langle \cdot, \cdot \rangle$ denotes the scalar product in $L_2(0,1)$ and $\|\cdot\|$ the corresponding norm.

Since, in practice, the true shape of F_0 usually is unknown one is forced in this setting to make a choice of a more or less suitable F_0 in advance, and often the logistic shift model is chosen for the only reason that the most simple test, namely the Wilcoxon test, comes out. If one really could trust in the shift assumptions one seemingly should search for *data based* selection rules of an adequate F_0, resp. $\varphi(\cdot, f_0)$. In principle, this problem has been solved by Hájek and Šidák (1967), who construct a consistent sequence of estimators $\{\tilde{\varphi}_N\}$ for $\varphi(\cdot, f_0)$ based on the order statistics of $X_1, \dots, X_N$. Another method for a data-based selection rule of a suitable score function $\varphi(\cdot, f_0)$ under shift assumption has been proposed by Randles and Hogg (1973). Starting from a set of three score functions $\varphi(\cdot, f_i)$, $i = 1, 2, 3$, corresponding to light tails, medium tails and heavy tails, respectively, these authors choose one of the three score functions by a rule based on the order statistics of the pooled sample $X_1, \dots, X_N$.

The common feature of the approach of Hájek and Šidák as well as that of Randles and Hogg is that they fully exploit the assumption of *shift* alternatives. Only under this ideal model is the score function $\varphi(\cdot, f_0)$ optimal and it is only the shift assumption that allows one to use order statistics for estimating $\varphi(\cdot, f_0)$.

More realistic than shift alternatives are stochastically larger alternatives H_1: $F \le G$, $F \ne G$. Since $F \le G$ is equivalent to $F(x) = G(x - D(x))$, $\forall x$, for some *shift function* D: $\mathbb{R} \to [0, \infty)$, H_1 may be interpreted as shift alternatives, where the size of the shift may vary for different parts of the distribution. In such a general situation, the score function $\varphi(\cdot, f_0)$ loses its meaning and since H_0 and H_1 are invariant under bijective, increasing transformations of the observations, the order statistics will contain no useful information for deciding between H_0 and H_1.

For a better understanding of the structure of H_1^0 and H_1, we will replace the parameter (F, G) by an equivalent one. For that purpose write $H = \eta_N F + (1 - \eta_N)G$ with $\eta_N = m/N$, see Behnen and Neuhaus (1983). Then the df of $H(X_i)$ is $F \circ H^{-1}$ for $1 \le i \le m$, resp. $G \circ H^{-1}$ for $m + 1 \le i \le N$. The corresponding distributions are dominated by the Lebesgue measure λ on $(0,1)$ with $\bar{f} = dF \circ H^{-1}/d\lambda$ and $\bar{g} = dG \circ H^{-1}/d\lambda$. Since $\eta_N \bar{f} + (1 - \eta_N)\bar{g} = 1$ the score function $\bar{b} = \bar{f} - \bar{g}$ satisfies

$$(1.8)\qquad \bar{b} \in L_2^0(0,1) = \left\{ f \in L_2(0,1)\colon \int_0^1 f\,d\lambda = 0 \right\}, \qquad -\frac{N}{n} \le \bar{b} \le \frac{N}{m}.$$

We could use the well-defined pair $(\bar{b}, H)$ as a new parameter instead of (F, G), since (F, G) can be reconstructed from $(\bar{b}, H)$ by the equalities

$$(1.9)\qquad \frac{dF}{dH} = 1 + (1 - \eta_N)\bar{b} \circ H, \qquad \frac{dG}{dH} = 1 - \eta_N \bar{b} \circ H.$$

In fact, for discussing local asymptotic results, it is more convenient to replace $\bar{b}$ by its rescaled version $b = (mn/N)^{1/2}\bar{b}$. Therefore, we will use (b, H) as our new parameter. Whereas H is a nuisance parameter, the function b contains all information on the belonging of (F, G) to H_0 and H_1. This follows from the equivalences

$$(1.10)\qquad \begin{aligned} &F \le G \Leftrightarrow B(t) = \int_0^t b\,d\lambda \le 0, \qquad \forall t \in (0,1), \\ &F = G \Leftrightarrow b = 0, \end{aligned}$$

which themselves are entailed by

$$(1.11)\qquad (mn/N)^{1/2}(F - G) \circ H^{-1}(t) = \int_0^t b\,d\lambda = B(t), \qquad \forall t \in (0,1).$$

Writing $\mathscr{F}^c = \{F\colon F \text{ continuous df on } \mathbb{R}\}$ and

$$M_N = \left\{ b \in L_2^0(0,1)\colon -(Nm/n)^{1/2} \le b \le (Nn/m)^{1/2} \right\},$$

the new parameter (b, H) can take on each element in $M_N \times \mathscr{F}^c$ for fixed sample numbers m, n. Here and below we always assume that $m = m(N)$ and

$n = n(N)$. Adding an index N we may write H_{0N}: $b = 0$, H_{1N}: $b \neq 0$, $\int_0^{\cdot} b\, d\lambda \leq 0$ and H_{1N}^0: $b \neq 0$.

Similarly, as in the shift case for every $b \in L_2^0(0,1)$, $b \neq 0$, the sequence of rank tests $\{\psi_N^b\}$ with

$$\psi_N^b = 1\big(S_N(b) \geq \Phi^{-1}(1-\alpha)\|b\|\big), \tag{1.12}$$

where Φ is the standard normal df, $0 < \alpha < 1$, is optimal in a local asymptotic sense. More exactly, define for each sequence $\{H_N\} \subset \mathscr{F}^c$ local sequences $\{(F_N, G_N)\}$ by

$$\begin{aligned} F_N &= \int_0^{H_N} f_N\, d\lambda, \qquad f_N = 1 + c_{N1} b_N, \\ G_N &= \int_0^{H_N} g_N\, d\lambda, \qquad g_N = 1 + c_{NN} b_N, \end{aligned} \tag{1.13}$$

with $b_N \in M_N$, $\forall N$, and $b_N - b \to 0$ in $L_2(0,1)$. Let $Q_N^{b_N}$ denote the joint distribution of $X_1, \dots, X_N$ with $X_i \sim F_N$ as $1 \leq i \leq m$, $X_i \sim G_N$ as $m+1 \leq i \leq N$. Then the sequence of linear rank tests $\{\psi_N^b\}$ is asymptotically optimal at level α for testing H_0 versus $\{Q_N^{b_N}\}$. Seemingly, here we are in a similar situation as in the shift case: For every *direction* $b \in L_2^0(0,1)$, $b \neq 0$, the corresponding sequence $\{\psi_N^b\}$ is asymptotically optimal, but only for its "own" alternatives $\{Q_N^{b_N}\}$. For other directions $h \neq b$, the limiting power depends on the size of $c(h,b) = \langle h, b\rangle / \|h\|\,\|b\|$, i.e., on the angle between h and b. In practice, one is confronted with the same problem as in the shift case: Since b is unknown, one needs some rules for the choice of a suitable b. Whereas in the shift case, estimators $\tilde{\varphi}_N$ of $\varphi(\cdot, f_0)$ have been constructed, here in an analogous way we will search for estimators or approximations $\hat{b}_N$, say, of the unknown score function b, leading to (nonlinear) rank tests based on $S_N(\hat{b}_N)$. Yet the analogy is not as complete as it appears at first sight. Whereas under the shift assumptions, $\varphi(\cdot, f_0)$ is a *nuisance* parameter, which bears no information about the question whether H_0 or some alternative holds true, the score function b contains *all* information on that question. This difference has far-reaching consequences. Whereas the nuisance score function $\varphi(\cdot, f_0)$ can be estimated consistently resulting in an "adaptive" test, i.e., a test with the same asymptotic power as if f_0 would be known in advance, a consistent estimation of b under local alternatives (1.13) is, in principle, impossible since $\{Q_N^{b_N}\}$ is contiguous to H_0, implying that every sequence of estimators $\{\hat{b}_N\}$ estimating b under H_0, i.e., $b = 0$, would likewise tend to zero under $\{Q_N^{b_N}\}$ for $b \neq 0$. Therefore, under local alternatives (1.13) one has to content oneself with estimators $\hat{b}_N$ having weaker convergence properties than consistency. In fact, we will consider estimators $\hat{b}_N(t)$ converging under (1.13) *in distribution* towards a limiting distribution with (approximate) mean $b(t)$.

2. Approximating the score function by kernel estimators. Adopting the notation of Section 1, we will consider data-based approximations of the score function

$$(2.1)\qquad b = \left(\frac{mn}{N}\right)^{1/2} \frac{d}{d\lambda}(F \circ H^{-1} - G \circ H^{-1}) = \frac{dB}{d\lambda}.$$

Since b remains invariant under bijective, strictly increasing transformations of the observations and since the vector of ranks $(R_1, \dots, R_N)$ is maximal invariant under the group of such transformations the estimators will be based on $(R_1, \dots, R_N)$. If $\hat{F}_N$, resp. $\hat{G}_N$, denotes the empirical df of $X_1, \dots, X_m$, resp. $Y_1, \dots, Y_n$, and $\hat{H}_N = \eta_N \hat{F}_N + (1 - \eta_N)\hat{G}_N$ the empirical df of the pooled sample $X_1, \dots, X_N$, a natural estimator of B is $\hat{B}_N$

$$(2.2)\qquad \hat{B}_N(t) = \left(\frac{mn}{N}\right)^{1/2} (\hat{F}_N - \hat{G}_N) \circ \hat{H}_N^{-1}(t), \qquad 0 \le t \le 1.$$

For notational reasons it will be more convenient to consider the linearized version $\overline{W}_N$ of $\hat{B}_N$, i.e.,

$$(2.3)\qquad \begin{gathered} \overline{W}_N\left(\frac{i}{N}\right) = \hat{B}_N\left(\frac{i}{N}\right), \\ i = 0, 1, \dots, N,\ \overline{W}_N \text{ linear in all intervals } [(i-1)/N, i/N]. \end{gathered}$$

A preliminary "estimator" of b is a version $\bar{b}_N$ of the λ-a.s. existing derivative of $\overline{W}_N$:

$$(2.4)\quad \bar{b}_N(t) = N\left(\hat{B}_N\left(\frac{i}{N}\right) - \hat{B}_N\left(\frac{i-1}{N}\right)\right), \quad \text{for } \frac{i-1}{N} \le t < \frac{i}{N}, 1 \le i \le N.$$

In order to handle the one-sided testing problem H_0 versus H_1, we need the projection $\overline{W}_N^0 = \overline{W}_N \wedge 0$ of $\overline{W}_N$ onto the cone of nonpositive functions on $[0,1]$. Then $\bar{b}_N^0 = d\overline{W}_N^0/dt$ denotes a version of the λ-a.s. existing derivative of $\overline{W}_N^0$.

Using $\overline{W}_N$ and its derivative $\bar{b}_N$ the linear rank statistic $S_N(b)$ in (1.5) becomes a simple scalar product

$$(2.5)\qquad S_N(b) = \sum_{i=1}^{N} c_{Ni}\tilde{b}_N(R_i) = \int_0^1 b\, d\overline{W}_N = \langle b, \bar{b}_N \rangle.$$

If $\hat{b}_N$ is an estimator of b we will use $S_N(\hat{b}_N) = \langle \hat{b}_N, \bar{b}_N \rangle$ as a new test statistic.

2.1. *Definition of the test statistics and finite sample results.* In BNR kernel estimators $\hat{b}_N$ of b are defined in the following way. Let $K\colon \mathbb{R} \to \mathbb{R}$ be a *kernel* with

$$(2.6)\qquad K(x) = K(-x), \qquad \forall x \in \mathbb{R},\ K(x) = 0 \text{ as } |x| > 1 \text{ and } \int_{-1}^{1} K\, d\lambda = 1.$$

Usually K will be nonnegative, i.e., a density on $\mathbb{R}$. But, sometimes it will be convenient to admit also negative values for K, see Example 2.3 below. K is the basis for defining convolution kernels $K_a(\cdot,\cdot)$ by

$$K_a(s,t) = \frac{1}{a}\left\{K\left(\frac{s+t}{a}\right) + K\left(\frac{s-t}{a}\right) + K\left(\frac{s+t-2}{a}\right)\right\}, \tag{2.7}$$

$$0 \le s, t \le 1,$$

for some bandwidth $a \in (0,1]$. Usually a convolution kernel would be defined as $(1/a)K(s - t/a)$; the first and third term in the curled brackets are included for compensating "disappearing" mass of K near the boundary point $s = 0$, resp. $s = 1$. Under the additional assumption

$$\infty > \int_{-1}^{1} K^2\, d\lambda = \sigma^2, \quad \text{say}, \tag{2.8}$$

$K_a(\cdot,\cdot)$ is square integrable on $[0,1]^2$ w.r.t. $\lambda \otimes \lambda$; hence

$$\mathscr{K}_a g = \int_0^1 g(t) K_a(\cdot, t)\lambda(dt), \qquad g \in L_2(0,1), \tag{2.9}$$

defines a linear operator $\mathscr{K}_a$: $L_2(0,1) \to L_2(0,1)$ of the Hilbert–Schmidt type. According to the philosophy of kernel estimators, $\mathscr{K}_a \bar{b}_N$ estimates b and the resulting test statistic for testing H_0: $b = 0$ versus the omnibus alternative H_1^0: $b \neq 0$ becomes

$$\begin{aligned} S_N(\mathscr{K}_a \bar{b}_N) &= \langle \mathscr{K}_a \bar{b}_N, \bar{b}_N \rangle = \int_0^1\int_0^1 K_a(s,t)\bar{b}_N(s)\bar{b}_N(t)\lambda(ds)\lambda(dt) \\ &= \int_0^1\int_0^1 K_a(s,t)\overline{W}_N(ds)\overline{W}_N(dt) \\ &= \sum_{i=1}^N \sum_{j=1}^N c_{Ni}c_{Nj}a_N(R_i, R_j), \end{aligned} \tag{2.10}$$

with

$$a_N(i,j) = N^2 \int_{(i-1)/N}^{i/N} \int_{(j-1)/N}^{j/N} K_a(\cdot,\cdot)\, d\lambda\, d\lambda, \qquad 1 \le i, j \le N. \tag{2.11}$$

In fact, in BNR a slightly altered version was used with $a_N(i, j)$ replaced by the approximating quantities (if K is smooth)

$$K_a\left(\frac{i-1/2}{N}, \frac{j-1/2}{N}\right), \qquad 1 \le i, j \le N. \tag{2.12}$$

Similarly, for testing H_0: $b = 0$ versus the stochastically larger alternative H_1: $b \neq 0$, $\int_0^{\cdot} b\, d\lambda \le 0$, let us replace $\bar{b}_N$ by $\bar{b}_N^0$ (satisfying $\int_0^{\cdot} \bar{b}_N^0\, d\lambda = \overline{W}_N^0 \le 0$ by construction):

$$\begin{aligned} S_N(\mathscr{K}_a \bar{b}_N^0) &= \langle \mathscr{K}_a \bar{b}_N^0, \bar{b}_N \rangle \\ &= \sum_{i=1}^N \sum_{j=1}^N c_{Ni}c_{Nj}a_N^0(R_i, R_j), \end{aligned} \tag{2.13}$$

with the random quantities

$$(2.14)\quad a_N^0(i,j) = N^2 \int_{(i-1)/N}^{i/N} \int_{(j-1)/N}^{j/N} K_a(s,t) 1(\overline{W}_N(s) \le 0) \lambda(ds)\lambda(dt).$$

One may expect the test statistic (2.13) to be more specific to H_1 than the omnibus statistic $S_N(\mathscr{K}_a \bar{b}_N)$. Again for computational reasons $a_N^0(i,j)$ may be replaced by

$$(2.15)\qquad K_a\left(\frac{i-1/2}{N}, \frac{j-1/2}{N}\right)\delta_N(i),$$

with

$$(2.16)\qquad \delta_N(i) = N\int_{(i-1)/N}^{i/N} 1(\overline{W}_N(s) \le 0)\lambda(ds).$$

Apparently, $0 \le \delta_N(i) \le 1$ with $\delta_N(i) = 1$ if $\overline{W}_N((i-1)/N) \le 0$, $\overline{W}_N(i/N) \le 0$ and $\delta_N(i) = 0$ if $\overline{W}_N((i-1)/N) \ge 0$ and $\overline{W}_N(i/N) \ge 0$. δ_N is a "cutting rule" taking care of the property $\int_0^{\cdot} b\, d\lambda \le 0$. We will get a better understanding of the performance of $\langle \mathscr{K}_a \bar{b}_N, \bar{b}_N\rangle$, resp. $\langle \mathscr{K}_a \bar{b}_N^0, \bar{b}_N\rangle$, by computing the spectral representation of $\mathscr{K}_a$.

LEMMA 2.1. *Let $\psi_\kappa \in L_2^0(0,1)$ be given by*

$$(2.17)\qquad \psi_\kappa(t) = \sqrt{2}\cos \pi\kappa t, \qquad 0 < t < 1,\ \kappa \ge 1,$$

and write

$$(2.18)\qquad \lambda_\kappa = \lambda_\kappa(a) = \int_{-1}^{1} K(t)\cos(\pi\kappa a t)\lambda(dt), \qquad \kappa \ge 1, 0 < a \le 1.$$

Then

$$(2.19)\qquad \mathscr{K}_a b = \langle b, 1\rangle + \sum_{\kappa=1}^{\infty} \lambda_\kappa \langle \psi_\kappa, b\rangle \psi_\kappa, \qquad \forall b \in L_2(0,1),$$

where in (2.19) *the infinite series converges in* $L_2(0,1)$.

PROOF. Since $1, \psi_1, \psi_2, \dots$ form a complete orthonormal basis in $L_2(0,1)$ and because of $\mathscr{K}_a 1 = 1$, see (2.9), (2.7) and (2.6), we only have to show $\mathscr{K}_a \psi_\kappa = \lambda_\kappa \psi_\kappa$, $\forall \kappa \ge 1$. With $A = \pi\kappa$ a straightforward calculation yields

$$(2.20)\qquad \int_0^1 K_a(s,t)\sqrt{2}\cos At\,\lambda(dt) = \sqrt{2}\cos As \int_{-1}^{1} K(x)\cos Aax\,\lambda(dx). \qquad \square$$

The functions $1, \psi_1, \psi_2, \dots$ are the *eigenfunctions* of $\mathscr{K}_a$, while $1, \lambda_1, \lambda_2, \dots$ are the corresponding *eigenvalues*. The remarkable feature is that for *all* kernels K the set of eigenfunctions remains the same. Only the sequence $\{\lambda_\kappa(a)\}$ varies for different kernels and different bandwidths. Using (2.19) we obtain the finite sample (!) expansion

$$(2.21)\qquad S_N(\mathscr{K}_a \bar{b}_n) = \sum_{\kappa=1}^{\infty} \lambda_\kappa \langle \psi_\kappa, \bar{b}_N\rangle^2 = \sum_{\kappa=1}^{\infty} \lambda_\kappa \langle \psi_\kappa', \overline{W}_N\rangle^2,$$

with $\psi_\kappa'(t) = d\psi_\kappa/dt$ and $\langle \psi_\kappa', \overline{W}_N \rangle = -\langle \psi_\kappa, \bar{b}_N \rangle$ (partial integration). Similarly,

$$
\begin{aligned}
S_N(\mathscr{K}_a \bar{b}_N^0) &= \sum_{\kappa=1}^{\infty} \lambda_\kappa \langle \psi_\kappa, \bar{b}_N \rangle \langle \psi_\kappa, \bar{b}_N^0 \rangle \\
&= \sum_{\kappa=1}^{\infty} \lambda_\kappa \langle \psi_\kappa', \overline{W}_N \rangle \langle \psi_\kappa', \overline{W}_N^0 \rangle.
\end{aligned} \tag{2.22}
$$

From Section 1 we know that the linear rank statistic $\langle \psi_\kappa, \bar{b}_N \rangle$ leads to asymptotically optimal tests for alternatives (1.13) with direction $b = \psi_\kappa$. Therefore, the test based on $\langle \mathscr{K}_a \bar{b}_N, \bar{b}_N \rangle$, being a quadratic form in $\langle \psi_\kappa, \bar{b}_N \rangle$, $\kappa \geq 1$, will be sensitive to all directions b in the $L_2(0,1)$-span $[\psi_1, \psi_2, \dots] = L_2^0(0,1)$, of course at different rates for different ψ_κ according to the size of the weights λ_κ in (2.21). Actually, in most examples only the first few eigenvalues λ_κ will be significantly different from zero.

In general, the eigenvalues λ_κ may be negative, but, as we will see below, the test based on $\langle \mathscr{K}_a \bar{b}_N, \bar{b}_N \rangle$ will be asymptotically unbiased if and only if all λ_κ are nonnegative. Therefore, we shall consider mainly kernels K with $\lambda_\kappa(a) \geq 0$, $\forall \kappa \geq 1$.

For a given kernel K the shape of the sequence $\{\lambda_\kappa(a)\}$ is completely determined by the choice of the bandwidth "a." In general, a small "a" causes a slow decrease to zero of $\lambda_\kappa(a)$ for $\kappa = 1, 2, \dots$, whereas a big "a" entails a fast decrease, see Table 1 and (2.18). In principle, the same arguments apply to (2.22) though the single terms $\langle \psi_\kappa, \bar{b}_N^0 \rangle \langle \psi_\kappa, \bar{b}_N \rangle$ do not allow such a simple interpretation as for $\langle \psi_\kappa, \bar{b}_N \rangle^2$ in (2.21).

Having identified the omnibus statistic $S_N(\mathscr{K}_a \bar{b}_N)$ as a quadratic form in linear rank statistics one could ask why we have made the above lengthy development instead of starting immediately with arbitrary quadratic forms in linear rank statistic. Beside the fact that this connection was not clear in advance, the above considerations indicate how to adapt the omnibus test statistic to the one-sided case H_0 versus H_1, which would be impossible to guess when starting from quadratic forms directly.

Let us give some examples.

EXAMPLE 2.2 (Cramér–von Mises statistics). Consider the rather artificial kernel

$$K(t) = \left\{ \frac{2}{3} - \frac{|t|}{2} \left(1 - \frac{|t|}{2} \right) \right\} 1(|t| \leq 1).$$

Then (2.18) yields for $A = \pi \kappa a$

$$\lambda_\kappa(a) = \frac{A - \sin A}{A^3} + \frac{5}{6} \frac{\sin A}{A}. \tag{2.23}$$

Putting $a = 1$ yields $\lambda_\kappa = \lambda_\kappa(1) = 1/(\pi\kappa)^2$, $\forall\kappa$. Therefore,

$$S_N(\mathscr{K}_1\bar{b}_N) = \langle \mathscr{K}_1\bar{b}_N, \bar{b}_N\rangle = \sum_{\kappa=1}^{\infty} \lambda_\kappa(1)\langle \psi_\kappa', \overline{W}_N\rangle^2 \tag{2.24}$$
$$= \sum_{\kappa=1}^{\infty} \langle 2^{1/2}\sin \pi\kappa\,\cdot\,, \overline{W}_N\rangle^2 = \|\overline{W}_N\|^2$$

and, similarly,

$$S_N(\mathscr{K}_1\bar{b}_N^0) = \langle \overline{W}_N, \overline{W}_N^0\rangle = \|\overline{W}_N^0\|^2. \tag{2.25}$$

Replacing $\overline{W}_N, \overline{W}_N^0$ by its jump versions $\hat{B}_N$, $\hat{B}_N^0 = \hat{B}_N \wedge 0$ leads to

$$\|\hat{B}_N\|^2 = \frac{mn}{N}\int \left(\hat{F}_N - \hat{G}_N\right)^2 d\hat{H}_N, \tag{2.26}$$

i.e., the omnibus Cramér–von Mises statistic, resp.

$$\|\hat{B}_N^0\|^2 = \frac{mn}{N}\int \left\{\left(\hat{F}_N - \hat{G}_N\right)^-\right\}^2 d\hat{H}_N, \tag{2.27}$$

i.e., the one-sided Cramér–von Mises statistic. Thus, the Cramér–von Mises statistics are (essentially) linear rank statistics with estimated scores, based on the above kernel K with a very large bandwidth $a = 1$ corresponding to the fast decrease of $1/(\pi\kappa)^2$ for $\kappa \to \infty$.

If we allow the kernel to take on negative values the sequence of nonnull eigenvalues even may become finite.

EXAMPLE 2.3 (Dirichlet kernel). Let $K = K_r$ be given by

$$K_r(t) = \frac{\sin\left(r + \frac{1}{2}\right)\pi t}{2\sin(\pi t/2)} 1(|t| \le 1), \quad \text{for } r = 1, 2, \dots, \tag{2.28}$$

which is known as the *Dirichlet kernel of order r*. Here r^{-1} plays the role of the bandwidth, while the parameter "a" is fixed to $a = 1$. The eigenvalues $\lambda_\kappa = \lambda_\kappa(1)$ become

$$\lambda_\kappa = 1 \text{ as } \kappa = 1, \dots, r \quad \text{and} \quad \lambda_\kappa = 0 \text{ as } \kappa > r. \tag{2.29}$$

If $\mathscr{K}_r$ denotes the operator (2.9) corresponding to K_r, then (2.19) yields

$$\mathscr{K}_r b = \langle b, 1\rangle + \sum_{\kappa=1}^{r} \langle \psi_\kappa, b\rangle\psi_\kappa, \qquad b \in L_2(0,1), \tag{2.30}$$

which is the finite Fourier expansion of b in $L_2(0,1)$. If $V = [1, \psi_1, \dots, \psi_r]$ denotes the linear subspace of $L_2(0,1)$ spanned by $1, \psi_1, \dots, \psi_r$, $\mathscr{K}_r$ is simply the orthogonal projection on V, i.e., $\mathscr{K}_r = \Pi_V$. In this case,

$$S_N(\mathscr{K}_r\bar{b}_N) = \|\Pi_V\bar{b}_N\|^2 = \sum_{\kappa=1}^{r} \langle \psi_\kappa, \bar{b}_N\rangle^2, \tag{2.31}$$

resp.

$$S_N(\mathscr{K}_r \bar{b}_N^0) = \sum_{\kappa=1}^{r} \langle \psi_\kappa, \bar{b}_N \rangle \langle \psi_\kappa, \bar{b}_N^0 \rangle. \tag{2.32}$$

This example links the present method of constructing data based approximations of the score function b by kernel estimators to the following *projection method*: Take a suitable cone or linear subspace in $L_2^0(0,1)$ and use the projection $\Pi_V \bar{b}_N = \hat{b}_N$ as a data-based approximation for b, leading to $\langle \Pi_V \bar{b}_N, \bar{b}_N \rangle$ as a test statistic. For example, for testing H_0 versus H_1 one could choose some representative score functions $b_1, \dots, b_r$ from H_1 and put $V = \{\sum_{i=1}^r \lambda_i b_i \colon \lambda_i \geq 0; \forall i\}$. This method has been studied in a not yet published manuscript of Behnen and the author.

A less exotic kernel than the above ones is the following kernel of Parzen, which plays an important role in time series analysis as a so-called lag window; see, e.g., Koopmans (1974), Chapter 8. We will use this kernel in our Monte Carlo study below and recommend it for practical applications.

EXAMPLE 2.4 (Parzen-2 kernel). Let K be the density of $U = (U_1 + U_2 + U_3 + U_4)/4$, where U_j are i.i.d. with uniform distribution on $(-1,1)$, i.e.,

$$K(t) = \frac{4}{3} \begin{cases} 1 - 6t^2 + 6|t|^3, & \text{for } |t| \leq \frac{1}{2}, \\ 2(1-|t|)^3, & \text{for } \frac{1}{2} \leq |t| \leq 1, \\ 0, & \text{for } |t| > 1. \end{cases} \tag{2.33}$$

Then

$$\lambda_\kappa(a) = g(\pi\kappa a), \quad \text{with } g(x) = \left(\frac{\sin x/4}{x/4} \right)^4. \tag{2.34}$$

Since the characteristic function of U_j is $g_0(x) = \sin x/x$, $1 \leq j \leq 4$, (2.34) follows from (2.18) and

$$\int_{-1}^{1} \cos(xt) K(t)\, dt = \int_{-1}^{1} e^{itx} K(t)\, dt = Ee^{ixU} = \prod_{j=1}^{4} Ee^{i(x/4)U_j} = g_0^4\left(\frac{x}{4}\right).$$

2.2. *Asymptotic distributions of the test statistics under local alternatives.* In BNR kernel estimators of $d(F-G) \circ H^{-1}/d\lambda$ were introduced and consistency results were proved under fixed alternatives with bandwidth $a = a_N \to 0$ as well as asymptotic normality of the standardized omnibus statistic $S_N(\mathscr{K}_a \bar{b}_N)$. But, it turned out in Monte Carlo simulations that these asymptotic results did not fully explain the power properties of the test based on $S_N(\mathscr{K}_a \bar{b}_N)$ for finite samples. From here the need for a better asymptotic arose. Moreover, the results in BNR when specialized to H_0 degenerate to convergence towards 0. Therefore, using another standardization BH subsequently proved asymptotic normality of $S_N(\mathscr{K}_a \bar{b}_N)$ for $a_N \to 0$ under H_0. Extending these results to contiguous alternatives (1.13) shows the completely unsatisfactory feature that this sort of local theory predicts a power equal to the level α of the test, quite in contrast to the reality. The reality is that the test shows very good power behavior in Monte Carlo studies for the *same* situations, where the local theory with bandwidth $a_N \to 0$ predicts power α. Therefore, we will keep the bandwidth "a" fixed in the asymptotic theorems of the present paper and it will turn out that the power as well as the critical values computed from this sort of asymptotics fit much better the finite sample Monte Carlo results than using the asymptotics of BNR, resp. BH. With the help of the expansion (2.21) one easily recognizes why asymptotics with $a_N \to 0$ only poorly reflect finite sample properties: According to (2.18) $\lambda_\kappa(a) \to 1$ for $a \to 0$, $\forall \kappa \geq 1$. Consequently, the quadratic form $\sum_{\kappa=1}^{\infty} \lambda_\kappa \langle \psi_\kappa, \bar{b}_N \rangle^2$ takes into account more and more linear rank statistics $\langle \psi_\kappa, \bar{b}_N \rangle$ for $a \to 0$, i.e., its distribution will vary markedly for decreasing a_N and increasing N, so that the asymptotic distribution has only little connection with its finite sample counterpart. Having the expansions (2.21) and (2.22) at our disposal, it is very easy to get the limiting distributions of $S_N(\mathscr{K}_a \bar{b}_N)$ and $S_N(\mathscr{K}_a \bar{b}_N^0)$ under local alternatives (1.13). We use the following functional limit theorem [see Hájek and Šidák (1967), V.3.5]: Under alternatives (F_N, G_N) from (1.13) and

$$\eta_N = \frac{m}{N} \to \eta \in (0,1) \quad \text{as } N \to \infty, \tag{2.35}$$

one has convergence in distribution ($\to_{\mathscr{D}}$) of the rank processes $\overline{W}_N$ in the space $C = C[0,1]$ of continuous functions on $[0,1]$

$$\overline{W}_N \to_{\mathscr{D}} W_0 + \int_0^{\cdot} b\, d\lambda \quad \text{as } N \to \infty, \tag{2.36}$$

where W_0 is a Brownian bridge process on $[0,1]$.

In fact, Hájek and Šidák (1967) prove (2.36) only under H_0. Since tightness is preserved under contiguous alternatives (1.13) [use the $\varepsilon - \delta$ form of contiguity, cf. Strasser (1985), Lemma 18.6] one only has to prove convergence of the finite-dimensional distributions of $\overline{W}_N$ to the appropriate limit, which can be achieved by the Cramér–Wold device applied to the linear rank statistics $\hat{B}_N(t)$, t fixed.

Theorem 2.5. (a) *Under* $\{(F_N, G_N)\}$, *from* (1.13) *and assuming* (2.35) *as well as* [*for some fixed* $a \in (0,1]$, $\lambda_\kappa := \lambda_\kappa(a)$]

$$\sum_{\kappa=1}^{\infty} |\lambda_\kappa| < \infty, \tag{2.37}$$

it follows that

$$S_N(\mathscr{K}_a \bar{b}_N) = \langle \mathscr{K}_a \bar{b}_N, \bar{b}_N \rangle \to_{\mathscr{D}} \sum_{\kappa=1}^{\infty} \lambda_\kappa (\xi_\kappa + \rho_\kappa)^2 = X^2(a, b), \qquad \textit{say}, \tag{2.38}$$

with i.i.d. standard normal random variables ξ_κ, $\kappa \geq 1$,

$$\xi_\kappa = -\int_0^1 \psi_\kappa' W_0\, d\lambda, \qquad \rho_\kappa = \langle \psi_\kappa, b \rangle, \qquad \kappa \geq 1. \tag{2.39}$$

(b) *Moreover, if even*

$$\sum_{\kappa=1}^{\infty} |\lambda_\kappa|\kappa < \infty \tag{2.40}$$

holds true, then

$$S_N(\mathscr{K}_a \bar{b}_N^0) \to_{\mathscr{D}} \sum_{\kappa=1}^{\infty} \lambda_\kappa \Big\langle \psi_\kappa', W_0 + \int_0^{\cdot} b\,d\lambda \Big\rangle \Big\langle \psi_\kappa', \Big(W_0 + \int_0^{\cdot} b\,d\lambda\Big) \wedge 0 \Big\rangle. \tag{2.41}$$

PROOF. (a) Write $\mathscr{H}_k(f) = \sum_{\kappa=1}^{k} \lambda_\kappa \langle \psi_\kappa', f\rangle^2$ for $f \in C$, $k = 1, 2, \dots, \infty$. For finite k the function $\mathscr{H}_k$ is continuous on $(C, \|\cdot\|_\infty)$. Thus, (2.36) implies

$$\mathscr{H}_k(\overline{W}_N) \to_{\mathscr{D}} \mathscr{H}_k\Big(W_0 + \int_0^{\cdot} b\,d\lambda\Big), \quad \forall k \in \mathbb{N}. \tag{2.42}$$

Using (2.21) we have $S_N(\mathscr{K}_a \bar{b}_N) = \mathscr{H}_\infty(\overline{W}_N)$ and we want to extend (2.42) to $k = \infty$. We begin with a look at the right-hand side of (2.38). It is a well-known fact that for $\{g_\kappa\colon \kappa \geq 1\} \subset L_2(0,1)$ the stochastic integrals $\{\int_0^1 g_\kappa W_0\,d\lambda\}_{\kappa\geq 1}$ are centered, jointly normal with covariances

$$E\int_0^1 g_\kappa W_0\,d\lambda \int_0^1 g_\tau W_0\,d\lambda = \int_0^1 G_\kappa G_\tau\,d\lambda - \int_0^1 G_\kappa\,d\lambda \int_0^1 G_\tau\,d\lambda, \qquad \kappa, \tau \geq 1, \tag{2.43}$$

where G_κ denotes an integral of g_κ. Putting $G_\kappa = \psi_\kappa$, $g_\kappa = \psi_\kappa'$ one obtains that the rv's ξ_κ, $\kappa \geq 1$, are indeed i.i.d. standard normal. Since $-\int_0^1 \psi_\kappa'(\int_0^{\cdot} b\,d\lambda)\,d\lambda = \langle \psi_\kappa, b\rangle = \rho_\kappa$, it follows from the Khintchine–Kolmogorov convergence theorem [see, e.g., Chow and Teicher (1978), Theorem 5.1] that the right-hand side of (2.38) converges a.s. and in quadratic mean. Under H_0

$$\begin{aligned} E_{H_0}\big|\mathscr{H}_\infty(\overline{W}_N) - \mathscr{H}_k(\overline{W}_N)\big| &= E_{H_0}\Big|\sum_{\kappa=k+1}^{\infty} \lambda_\kappa \langle \psi_\kappa, \bar{b}_N\rangle^2\Big| \\ &\leq \sum_{\kappa=k+1}^{\infty} |\lambda_\kappa| E_{H_0}\langle \psi_\kappa, \bar{b}_N\rangle^2 \\ &= \frac{N}{N-1} \sum_{\kappa=k+1}^{\infty} |\lambda_\kappa| \left(\frac{1}{N}\sum_{i=1}^{N}\Big\{N\int_{(i-1)/N}^{i/N} \psi_\kappa\,d\lambda\Big\}^2\right) \\ &\leq 2\sum_{\kappa=k+1}^{\infty} |\lambda_\kappa|\, \|\psi_\kappa\|^2 = 2\sum_{\kappa=k+1}^{\infty} |\lambda_\kappa|. \end{aligned} \tag{2.44}$$

Now, (2.44) and Markov's inequality yield

$$\limsup_{N\to\infty} P_{H_0}\Big\{\big|\mathscr{H}_k(\overline{W}_N) - \mathscr{H}_\infty(\overline{W}_N)\big| \geq \varepsilon\Big\} \to 0, \quad \text{for } k \to \infty, \forall \varepsilon > 0. \tag{2.45}$$

Contiguity, $\{Q_N^{b_N}\} \lhd \{Q_N^0\}$, implies that (2.45) holds true under (F_N, G_N), too, see e.g., Strasser (1985), Lemma 18.6. Since $\mathscr{H}_k(W_0 + \int_0^{\cdot} b\,d\lambda) \to_{\mathscr{D}} \mathscr{H}_\infty(W_0 + \int_0^{\cdot} b\,d\lambda)$ for $k \to \infty$, Theorem 4.2 of Billingsley (1968) yields (2.42) for $k = \infty$.

(b) In order to show (2.41) we proceed as above and define

$$\mathscr{H}_k^0(f) = \sum_{\kappa=1}^{k} \lambda_\kappa \langle \psi_\kappa', f\rangle \langle \psi_\kappa', f^0\rangle$$

for $f \in C$, $f^0 = f \wedge 0$, $k = 1, 2, \ldots, \infty$. Since $\mathscr{H}_k^0$ is continuous on $(C, \|\cdot\|_\infty)$ for finite k, (2.42) holds true with $\mathscr{H}_k$ replaced by $\mathscr{H}_k^0$. The infinite series in (2.41) converges in $L_1(0,1)$, since

$$E\left|\left(\mathscr{H}_k^0 - \mathscr{H}_\infty^0\right)\left(W_0 + \int_0^{\cdot} b\,d\lambda\right)\right|$$

$$\text{(2.46)}\qquad \leq \sum_{\kappa=k+1}^{\infty} |\lambda_\kappa| \left\{E(\xi_\kappa + \rho_\kappa)^2 \|\psi_\kappa'\|^2 E\left\|W_0 + \int_0^{\cdot} b\,d\lambda\right\|^2\right\}^{1/2}$$

$$\leq \text{const.} \sum_{\kappa=k+1}^{\infty} |\lambda_\kappa|\kappa, \quad \text{since } \|\psi_\kappa'\| = \pi\kappa.$$

Instead of (2.44) we get, similarly,

$$\text{(2.47)}\qquad E_{H_0}\left|\mathscr{H}_\infty^0(\overline{W}_N) - \mathscr{H}_k^0(\overline{W}_N)\right| \leq \text{const.} \sum_{\kappa=k+1}^{\infty} |\lambda_\kappa|\kappa, \qquad \forall k \geq 1.$$

The remainder of the proof is the same as before. □

Apparently the above theorem applies to the Parzen-2 kernel (2.33) for every $a \in (0,1]$. Under H_0 the limiting distribution of $S_N(\mathscr{K}_a \bar{b}_N)$ is that of $\sum_{\kappa=1}^\infty \lambda_\kappa \xi_\kappa^2$, i.e., a weighted sum of i.i.d. χ_1^2 random variables. Nowadays there are standard methods for numerical computation of such distributions as well as the distribution of $X^2(a,b)$ under alternatives ($b \neq 0$). Table 2 ($N = \infty$) and Table 4 ($N = \infty$) are computed by the method of Davies (1980).

At times a method for numerical evaluation of the limiting distribution of $S_N(\mathscr{K}_a \bar{b}_N^0)$ is not available, even under H_0, since the random variables $\langle \psi_\kappa', W_0 \wedge 0\rangle$, $\kappa \geq 1$, are neither normally distributed nor independent. So we have to be content with simulation results in this case.

REMARK 2.6. Under the assumptions of Theorem 2.5, the test statistics

$$\text{(2.48)}\qquad S_N = \sum_{i=1}^N \sum_{j=1}^N c_{Ni} c_{Nj} K_a(\tilde{R}_i, \tilde{R}_j), \qquad \tilde{R}_i = \frac{R_i - \frac{1}{2}}{N}$$

[cf. (2.12)], resp.

$$\text{(2.49)}\qquad S_N^0 = \sum_{i=1}^N \sum_{j=1}^N c_{Ni} c_{Nj} K_a(\tilde{R}_i, \tilde{R}_j) \delta_N(R_i)$$

[cf. (2.15) and (2.16)], have the same limiting distributions as in (2.38), resp. (2.41). In order to *prove* the limiting results, define the empirical jump rank process $W_N(t) = \sum_{i=1}^N c_{Ni} 1(\tilde{R}_i \leq t)$, $0 \leq t \leq 1$. Then

$$\text{(2.50)}\qquad S_N = \sum_{\kappa=1}^\infty \lambda_\kappa \langle \psi_\kappa', W_N \rangle^2,$$

resp. [with $\bar{\delta}_N(s) = \delta_N([Ns] + 1)$, $[\cdot]$ denoting the integer part function]

$$(2.51)\qquad \begin{aligned} S_N^0 &= \sum_{\kappa=1}^{\infty} \lambda_\kappa\Big(\int_0^1 \psi_\kappa \bar{\delta}_N\, dW_N\Big)\Big(\int_0^1 \psi_\kappa\, dW_N\Big) \\ &= \sum_{\kappa=1}^{\infty} \lambda_\kappa \langle \psi_\kappa', W_N^0\rangle\langle \psi_\kappa', W_N\rangle, \end{aligned}$$

with $W_N^0 = W_N \wedge 0$. The last equality in (2.51) follows from

$$(2.52)\qquad \int_0^1 \psi_\kappa \bar{\delta}_N\, dW_N = \int_0^1 \psi_\kappa\, dW_N^0 = -\langle \psi_\kappa', W_N^0\rangle,$$

where the first equality in (2.52) follows from

$$(2.53)\quad \delta_N(i)\Big\{W_N\Big(\frac{R_i}{N}\Big) - W_N\Big(\frac{R_i - 1}{N}\Big)\Big\} = W_N^0\Big(\frac{R_i}{N}\Big) - W_N^0\Big(\frac{R_i - 1}{N}\Big), \qquad \forall i,$$

which itself is a consequence of the convenient definition of δ_N; the second equality in (2.52) follows from integration by parts. One notices that the only difference between the proof of Theorem 2.5 and the present one is the use of W_N instead of $\overline{W}_N$, resp. W_N^0 instead of $\overline{W}_N^0$. Since $\|W_N - \overline{W}_N\|_\infty = O(N^{-1/2})$ (2.42) holds with $\overline{W}_N$ replaced by W_N for $\mathscr{H}_k$ as well for $\mathscr{H}_k^0$. The remainder of the proof is essentially the same as before.

REMARK 2.7. If one is only interested in (omnibus) quadratic rank statistics

$$(2.54)\quad T_N = T_N(k_N) = \sum_{i=1}^{N}\sum_{j=1}^{N} c_{Ni}c_{Nj}k_N\big(\tilde{R}_i, \tilde{R}_j\big), \qquad \tilde{R}_i = \frac{R_i - \frac{1}{2}}{N},$$

with $k_N \in U_N^2 = \{k \in L_2(0,1)^2$: k constant on all intervals $[(i-1)/N, i/N) \times [(j-1)/N, j/N)$, $1 \le i, j \le N\}$ one can derive more general results than (2.38) under less stringent assumptions. Let us shortly discuss these results. Assume that for some $k \in L_2(0,1)^2$

$$(2.55)\qquad k_N \to k \quad \text{in } L_2(0,1)^2$$

and let k and all k_N satisfy

$$(2.56)\qquad k(s,t) = k(t,s), \qquad \forall (s,t) \in (0,1)^2,$$

and

$$(2.57)\qquad \int_0^1 k(\cdot, t)\lambda(dt) = 0[\lambda], \quad \int_0^1 k(s,\cdot)\lambda(ds) = 0[\lambda].$$

It is well known that for $k \in L_2(0,1)^2$ with (2.56) and (2.57), there exists a sequence $\{\lambda_k\}$ in $\mathbb{R}$ with $\sum_{\kappa=1}^{\infty}\lambda_\kappa^2 < \infty$ and an orthonormal system $\{\varphi_\kappa\}$ in $L_2^0(0,1)$ with

$$(2.58)\qquad k(\cdot,\cdot) = \sum_{\kappa=1}^{\infty} \lambda_\kappa \varphi_\kappa \otimes \varphi_\kappa,$$

where $\varphi_\kappa \otimes \varphi_\kappa(s,t) = \varphi_\kappa(s)\varphi_\kappa(t)$. The convergence in (2.58) is understood in the space $L_2(0,1)^2$; see Dunford and Schwartz (1963), Volume 2, XI, 8.56. Then, under the above assumptions and notation under (F_N, G_N), from (1.13) and (2.35) we have

$$T_N - E_{H_0}T_N \to_{\mathscr{D}} \sum_{\kappa=1}^{\infty} \lambda_\kappa\{(\xi_\kappa + \rho_\kappa)^2 - 1\}, \tag{2.59}$$

with $\rho_\kappa = \langle \varphi_\kappa, b\rangle$, $\kappa \geq 1$, and $\xi_1, \xi_2, \ldots$ i.i.d. standard normal rv's. Apparently, (2.59) entails (2.38).

The pattern of deriving the limiting distribution of quadratic forms similar to (2.54) is well known; see, e.g., Gregory (1977), Schach (1969) and especially Rothe (1976). Another way for deriving these results consists in comparing T_N in (2.54) with

$$\tilde{T}_N = \sum_{i=1}^{N} \sum_{j=1}^{N} c_{Ni}c_{Nj}k_N(U_i, U_j), \qquad U_i = H_N(X_i),$$

being essentially a U-statistic, for which the limiting distribution is well known; see Gregory (1977).

3. Some asymptotic power considerations for the omnibus tests. Under H_0 the limiting rv in Theorem 2.5 is $X^2(a) \equiv X^2(a,0) = \sum_{\kappa=1}^{\infty}\lambda_\kappa\xi_\kappa^2$. For $0 < \alpha < 1$ let $c(\alpha, a)$ fulfill $P\{X^2(a) > c(\alpha, a)\} = \alpha$. Then the sequence $\{\varphi_N^* = 1(S_N^* > c(\alpha, a))\}$, $S_N^* \equiv S_N(\mathscr{K}_a\bar{b}_N)$, has asymptotic level α. Applying Theorem 2.5 to alternatives (1.13) with $b = c\psi_\kappa$, c large, $\kappa \geq 1$, immediately shows that $\{\varphi_N^*\}$ is asymptotically unbiased iff $\lambda_\kappa \geq 0$, $\forall\kappa \geq 1$.

The only optimality result which seems to be possible for asymptotically unbiased tests $\{\varphi_N^*\}$ in general is asymptotic admissibility. In order to formulate this result we need some notions and results from the theory of experiments as described, e.g., in Strasser (1985): Recall that $Q_N^{b_N}$ is the joint distribution $\mathscr{L}(X_1, \ldots, X_N | F_N, G_N)$ for (F_N, G_N) from (1.13) with

$$b_N \in M_N = \{b \in L_2^0(0,1)\colon -(Nm/n)^{1/2} \leq b \leq (Nn/m)^{1/2}\}$$

and put $M = \bigcup_{N=2}^{\infty} M_N$. Then it is well known that

$$\frac{dQ_N^b}{dQ_N^0} = \exp\left\{X_N(b) - \frac{1}{2}\|b\|^2 + o_{Q_N^0}(1)\right\}, \qquad \forall b \in M, \tag{3.1}$$

with

$$X_N(b) = \sum_{i=1}^{N} c_{Ni} b \circ H_N(X_i) \to_{\mathscr{D}} \text{Normal}(0, \|b\|^2) \quad \text{under } Q_N^0. \tag{3.2}$$

According to Strasser (1985) the sequence of experiments

$$E_N(H_N) = (\mathbb{R}^N, \mathbb{B}^N, \{Q_N^b\colon b \in M_N\}), \qquad N \geq 2, \tag{3.3}$$

converges weakly to the Gaussian-shift experiment

$$E = (B, \text{Borel}(B), \{\mu_b\colon b \in M\}), \tag{3.4}$$

where $B = \{f \in C[0,1]\colon f(0) = f(1) = 0\}$ and $\mu_b = \mathscr{L}(W_0 + \int_0^{\cdot} b\, d\lambda)$.

Under the assumptions of Theorem 2.5 it follows from (2.42) with $k = \infty$ and $\mathscr{H}_\infty(\overline{W}_N) = S_N^*$,

$$\int \varphi_N^* \, dQ_N^b \to \int \varphi^* \, d\mu_b, \qquad \forall b \in M, \tag{3.5}$$

with $\varphi^*(f) = 1(\mathscr{H}_\infty(f) > c(\alpha, a))$, $f \in B$. Asymptotic admissibility means that each sequence $\{\varphi_N\}$ of tests with $\limsup \int \varphi_N \, dQ_N^0 \le \alpha$ and $\liminf \int \varphi_N \, dQ_N^b \ge \int \varphi^* \, d\mu_b$, $\forall b \in M$ is asymptotically equivalent to $\{\varphi_N^*\}$, i.e., $\int |\varphi_N - \varphi_N^*| \, dQ_N^b \to 0$, $\forall b \in M$. Following the lines of Example 82.23 of Strasser (1985), where the Kolmogorov–Smirnov test is considered, we obtain immediately the asymptotic admissibility of $\{\varphi_N^*\}$ by noticing that

$$\{f \in B: \mathscr{H}_\infty(f) \le c\} = \bigcap_{k=1}^{\infty} \left\{ f \in B: \sum_{\kappa=1}^{k} \lambda_\kappa \langle \psi_\kappa', f \rangle^2 \le c \right\} \tag{3.6}$$

is closed and convex in B if $\lambda_\kappa \ge 0$, $\forall \kappa$.

The above result shows that, e.g., the sequence of omnibus tests $\{\varphi_N^*\}$ based on the Parzen-2 kernel of Example 2.4 ($\lambda_\kappa(a) \ge 0$, $\forall \kappa \ge 0$, $\forall a \in (0,1]$!) cannot be improved for some direction "b" without diminishing the power for some other direction. In this sense the test $\{\varphi_N^*\}$ puts its power in an optimal way onto the *principal directions* $b = \psi_\kappa$, $\kappa \ge 1$, the influence of each direction ψ_κ being weighted by the eigenvalue $\lambda_\kappa(a)$. Since $S_N^* = \sum_{\kappa=1}^{\infty} \lambda_\kappa(a) \langle \psi_\kappa, \overline{b}_N \rangle^2$, the corresponding test, φ_N^*, is completely determined by the sequence $\{\lambda_\kappa(a)\}$.

For practical applications one has to choose a suitable kernel $K: \mathbb{R} \to \mathbb{R}$ and to choose a suitable bandwidth "a." Similarly, as in density estimation, the choice of the bandwidth is of much more importance for the power behavior of the test than the shape of the kernel K, as long as it is bellshaped. The following result is helpful for showing this fact.

THEOREM 3.1. *Let K be a nonnegative, bounded kernel fulfilling* (2.6) *and let $\sigma^2 = \int_{-1}^{1} K^2 \, d\lambda$ be finite. Then*

$$\sqrt{a} \sum_{\kappa=1}^{\infty} \lambda_\kappa(a) (\xi_\kappa^2 - 1) \to_{\mathscr{D}} \text{Normal}(0, 2\sigma^2) \quad \text{as } a \to 0. \tag{3.7}$$

According to (3.7) two kernels K_i, with $\sigma_i^2 = \int_{-1}^{1} K_i^2 \, d\lambda$, $i = 1, 2$, lead to the same limiting normal distribution if the respective bandwidths a_1, a_2 satisfy

$$\frac{\sigma_1^2}{a_1} = \frac{\sigma_2^2}{a_2}. \tag{3.8}$$

To give an example, let K_1 be the Parzen-2 kernel of Example 2.4 and let K_2: $\mathbb{R} \to \mathbb{R}$ be the quartic kernel

$$K_2(x) = \tfrac{15}{16}(1 - x^2)^2 1(|x| \le 1), \qquad x \in \mathbb{R}, \tag{3.9}$$

with eigenvalues

$$\lambda_\kappa(a) = g_2(\pi \kappa a), \qquad \kappa \ge 1, \tag{3.10}$$

TABLE 1

Some eigenvalues $\lambda_\kappa(a)$ corresponding to the Parzen-2 kernel. In parentheses the eigenvalues of the kernel K_2, see (3.9), with bandwidth $a \cdot 225/302$, see (3.11).

a \ κ	1	2	3	4	5	6	7	8
0.30	0.96	0.86	0.71	0.54	0.38	0.24	0.13	0.06
	(0.96	0.87	0.72	0.55	0.37	0.21	0.09	0.01)
0.40	0.94	0.77	0.54	0.33	0.16	0.01	0.00	0.00
	(0.94	0.77	0.55	0.32	0.13	0.01	−0.04	−0.04)
0.50	0.90	0.66	0.38	0.16	0.05	0.01	0.00	0.00
	(0.90	0.66	0.37	0.13	−0.01	−0.04	−0.02	0.00)

with $g_2(x) = \frac{15}{8}\{(24 - 8x^2)\sin x - 24x\cos x\}/x^5$. Since $\sigma_1^2 = 302/315$ and $\sigma_2^2 = 5/7$, equality (3.8) means

$$a_2 = \tfrac{225}{302} a_1. \tag{3.11}$$

In Table 1 some eigenvalues for K_1 and K_2 are computed for $a_1 = 0.5, 0.4, 0.3$, resp. [according to (3.11)] $a_2 = 0.37, 0.30, 0.22$. One recognizes that under (3.11) there is no essential difference between K_1 and K_2. Therefore, we choose a fixed kernel, namely the Parzen-2 kernel, and will discuss in the next section only the choice of the bandwidth "a."

PROOF OF THEOREM 3.1. Define the iterated kernel $K_a^{(i)}(\cdot,\cdot)$ of $K_a(\cdot,\cdot)$ by

$$K_a^{(i+1)}(s,t) = \int_0^1 K_a^{(i)}(s,u)K_a(u,t)\lambda(du), \qquad K_a^{(1)} = K_a,\ i \geq 1. \tag{3.12}$$

Then the following equality is well known for $\lambda_\kappa = \lambda_\kappa(a)$ [see, e.g., Dunford and Schwartz (1963), Volume 2, XI, 8.49],

$$1 + \sum_{\kappa=1}^{\infty} \lambda_\kappa^i = \int_0^1 K_a^{(i)}(s,s)\lambda(ds), \qquad \forall i \geq 2. \tag{3.13}$$

We show that all moments of the left-hand side of (3.7) tend to the corresponding moments of a normal distribution. Write

$$X_k^2(a) = \sum_{\kappa=1}^{k} \lambda_\kappa(\xi_\kappa^2 - 1), \qquad k \geq 1. \tag{3.14}$$

Then, for $p \geq 2$

$$\begin{aligned} E\big(X_k^2(a)\big)^p &= \sum_{i_1=1}^{k} \cdots \sum_{i_p=1}^{k} \lambda_{i_1} \cdots \lambda_{i_p} E\big(\xi_{i_1}^2 - 1\big) \cdots \big(\xi_{i_p}^2 - 1\big) \\ &= \sum_{r=1}^{p} \sum_{\mathscr{P}_r} \prod_{j=1}^{r} \left(\sum_{i_j=1}^{k} \lambda_{i_j}^{|I_j|} E\big(\xi_1^2 - 1\big)^{|I_j|} \right), \end{aligned} \tag{3.15}$$

where Σ over $\mathscr{P}_r$ means the summation over the set $\mathscr{P}_r$ of all partitions $\{I_1, \ldots, I_r\}$ of $\{1, \ldots, p\}$ with number of elements $|I_j| \geq 2$. (3.15) shows that for $p \geq 1$ the sequence $\{(X_k^2(a))^p \colon k \geq 1\}$ is uniformly integrable. Therefore, using (3.13), one has

$$\text{(3.16)} \quad E\big(X_\infty^2(a)\big)^p = \sum_{r=1}^{p} \sum_{\mathscr{P}_r} \prod_{j=1}^{r} \left(\int_0^1 K_a^{(|I_j|)}(s, s)\lambda(ds) - 1 \right) E\big(\xi_1^2 - 1\big)^{|I_j|}.$$

From (3.12) we get for $i \geq 1$ with $\|\cdot\|_\infty$ denoting sup–norm

$$\text{(3.17)} \quad \begin{aligned} \int_0^1 K_a^{(i+1)}(s, s)\lambda(ds) &\leq \|K_a(\cdot,\cdot)\|_\infty \int_0^1 \int_0^1 K_a^{(i)}(s, t)\lambda(ds)\lambda(dt) \\ &\leq \frac{3}{a}\|K\|_\infty, \end{aligned}$$

implying that each term $\prod_{j=1}^r (\int_0^1 K_a^{(|I_j|)}(s, s)\lambda(ds) - 1)$ has order $O(a^{-r})$. Because of $r \leq p/2$, one gets for $p = 2m - 1$, $m \in \mathbb{N}$,

$$\text{(3.18)} \quad E\big(X_k^2(a)\big)^p = o\big(a^{-p/2}\big),$$

resp. for $p = 2m$, $m \in \mathbb{N}$,

$$\text{(3.19)} \quad E\big(X_k^2(a)\big)^{2m} - |\mathscr{P}_m| \left(\int_0^1 K_a^{(2)}(s, s)\lambda(ds) - 1 \right)^m \left(E\big(\xi_1^2 - 1\big)^2 \right)^m = o\big(a^{-m}\big).$$

Because of $|\mathscr{P}_m| = (2m)!/(m!2^m)$, $E(\xi_1^2 - 1)^2 = 2$ and

$$\text{(3.20)} \quad \int_0^1 K_a^{(2)}(s, s)\lambda(ds) - 1 = \frac{\sigma^2}{a} - \frac{1}{2},$$

following from an orthogonal series expansion of $(1/a)K(\cdot/a)$ w.r.t. $1, \psi_1, \psi_2, \ldots$ and (3.13), we get

$$\text{(3.21)} \quad E\big(\sqrt{a}\, X^2(a)\big)^{2m} = \frac{(2m)!}{m!2^m}\big(2\sigma^2\big)^m + o(1).$$

Since $(2m)!/(m!2^m) = E\xi_1^{2m}$, $\forall m \in \mathbb{N}$, the theorem is proved. □

The above theorem is in accordance with the main result of BH, where asymptotic normality of $S_N(\mathscr{K}_{a_N} \bar{b}_N)$ towards Normal$(0, 2\sigma^2)$ is proved under the assumption $a_N \to 0$ as $N \to \infty$. Since this normal approximation becomes rather bad if only a few of the $\lambda_\kappa(a)$ are large, i.e., $\lambda_\kappa(a) \approx 1$, it is usually not suited in practice for computing critical values and, in fact, not needed since the exact distribution of the left-hand side may be computed by standard methods, see Section 2.

Remark. (3.7) holds true under alternatives, i.e., if ξ_κ is replaced by $(\xi_\kappa + \rho_\kappa)$, since the factor $\sqrt{a}$ suppresses the influence of ρ_κ as $a \to 0$. This shows again that local asymptotics with bandwidths $a_N \to 0$ cannot exhibit the real power situation.

In the present section we only considered omnibus tests. Parallel results for the test based on the one-sided statistic $S_N(\mathscr{K}_a \bar{b}_N^0)$ would be desirable but are at time unknown because of the complicated form of the limiting distribution in (2.41).

4. Numerical results and practical considerations. In practice, having fixed the kernel K one has to choose a suitable bandwidth "a." The smaller "a" is chosen, the more $\lambda_\kappa(a)$, $\kappa \geq 1$, will gain influence and the test will pay its attention to more and more alternative directions $b = \psi_\kappa$ paralleled by neglecting the power for each single direction. In order to balance the contradicting aims of broad sensitivity and high power, we have systematically computed the expansions $b = \sum_{\kappa=1}^{\infty} \rho_\kappa \psi_\kappa$, $\rho_\kappa = \langle b, \psi_\kappa \rangle$, for many directions

$$b = (mn/N)^{1/2} d(F - G) \circ H^{-1}/d\lambda,$$

see Section 1, with df's (F, G) taken, e.g., from normal-, logistic- and Cauchy-shift models. In all cases it turns out that the first three to four ρ_κ's dominate. This leads to rather large values of "a" for the Parzen-2 kernel, e.g., $a = 0.40$ or $a = 0.50$; see Table 1.

In order to demonstrate the power behavior of the omnibus test, based on S_N, see (2.48), resp. of the one-sided test, based on S_N^0, see (2.49), two different Monte

TABLE 2

Power (in%) of the omnibus test for the two-sample problem H_0 versus H_1^0 based on S_N [see (2.48)] with Parzen-2 kernel (2.33) for various bandwidths "a" under alternatives (f_N, g_N) from (1.13) with $b_N = 3\sqrt{2}\cos \pi\kappa\,\cdot$ for $\kappa = 1,2,3,4$ and $m = n$, $N = m + n = 20, 80$ (obtained by simulation) and $N = \infty$ (exact asymptotic power). The last two columns show the corresponding power of the rank test based on $|\sum_i c_{Ni}\varphi(\tilde{R}_i)|$ with $\varphi = id - 1/2$ (two-sided Wilcoxon test) resp. $\varphi = \cos(\pi\kappa\,\cdot)$ (two-sided rank test with asymptotically optimal score function). All tests are at level $\alpha = 0.1$.

	$N = 20$					**$N = 40$**				
κ \ a	**0.30**	**0.40**	**0.50**	**Wil.**	**opt.**	**0.30**	**0.40**	**0.50**	**Wil.**	**opt.**
1	89.7	92.7	94.1	96.4	96.7	84.4	87.5	89.6	93.8	94.0
2	83.7	86.0	85.0	11.9	88.7	81.1	82.3	82.0	12.4	89.8
3	71.4	68.4	58.1	11.9	71.9	73.2	68.3	59.5	12.7	82.5
4	55.2	42.3	25.4	9.8	54.2	58.9	44.7	28.2	10.9	72.9

	$N = 80$					**$N = \infty$**				
κ \ a	**0.30**	**0.40**	**0.50**	**Wil.**	**opt.**	**0.30**	**0.40**	**0.50**	**Wil.**	**opt.**
1	83.4	86.5	88.4	91.5	91.8	81.4	84.6	86.7	90.9	91.2
2	80.5	82.1	82.0	10.7	91.2	78.6	80.2	80.2	10.0	91.2
3	73.5	70.0	61.7	11.9	87.4	73.0	70.0	62.1	11.9	91.2
4	62.0	48.4	29.4	9.9	82.1	63.4	49.9	29.2	10.0	91.2

Carlo studies have been performed: In the omnibus case *principal alternatives* (1.13) with $f_N = 1 + c_{N1}3\psi_\kappa$, $g_N = 1 + c_{NN}3\psi_\kappa$ were used for various values of κ. Though the directions ψ_κ cannot be interpreted directly one should recall that they form a complete orthonormal system in the space $L_2^0(0,1)$ of *all* possible directions $b = \sum_{\kappa=1}^{\infty}\rho_\kappa\psi_\kappa$. If $b \in [\psi_1, \ldots, \psi_k]$, with fixed norm $\|b\| = d > 0$, and if $\lambda_1(a) \geq \cdots \geq \lambda_k(a) > 0$ it may be shown with the help of Proposition 2.1 in Neuhaus (1976) that the asymptotic power under these direction b is in between its maximal value for $b = d\psi_1$ and its minimal value for $b = d\psi_k$. As mentioned above the space $[\psi_1, \ldots, \psi_k]$ with $k = 3$ or $k = 4$ is large enough to cover most

TABLE 3

Power (in %) of the one-sided test for the two-sample problem H_0 versus H_1 based on S_N^0 [see (2.49)] with Parzen-2 kernel (2.33) for various bandwidths "a" under generalized shift alternatives (F, G) with $F(x) = G(x - D(x))$ and D from (4.1) resp. pure shift $D \equiv 1/2$ for G standard normal (N), logistic (L) and Cauchy (C) df's. The cases $m = n$ and $N = m + n = 20, 80$ are handled. The last two columns show the corresponding power of the Wilcoxon test (Wil.) resp. of the test based on $\sum_{i=1}^m \log(dF/dG)\circ G^{-1}(\tilde{R}_i)$ (opt.), which is approximately the rank-likelihood ratio statistic. All tests are at level $\alpha = 0.1$.

Shift	G	$a = 0.30$	0.40	0.50	Wil.	opt.
		$N = 20$				
Upper	N	26.7	27.0	27.7	23.5	32.1
	L	17.9	18.0	18.4	15.9	22.3
	C	17.6	17.7	18.2	15.8	20.8
Central	N	24.5	24.8	26.0	22.3	31.2
	L	17.5	17.6	18.3	16.5	21.5
	C	19.9	20.0	20.7	18.5	24.2
Lower	N	22.2	22.5	23.0	19.4	31.0
	L	16.2	16.1	16.4	14.6	20.8
	C	15.5	15.6	15.8	14.3	19.3
Pure	N	36.1	36.9	38.1	39.6	41.9
	L	23.4	23.6	24.3	24.2	26.1
	C	23.5	23.7	24.8	23.2	27.7
		$N = 80$				
shift	G	$a = 0.30$	0.40	0.50	Wil.	opt.
Upper	N	64.0	64.6	64.5	48.1	75.6
	L	33.0	33.7	33.5	28.1	47.1
	C	33.4	33.6	33.1	27.5	47.2
Central	N	57.4	57.7	56.8	45.6	69.1
	L	32.9	33.2	33.4	29.6	47.2
	C	40.9	41.5	41.2	35.0	54.7
Lower	N	51.0	52.3	52.2	37.7	68.2
	L	29.2	29.7	29.6	25.0	40.7
	C	26.8	27.0	26.6	23.6	38.8
Pure	N	75.5	76.4	77.3	82.0	83.2
	L	45.0	46.0	47.0	50.9	51.0
	C	46.8	48.5	49.5	48.0	58.5

interesting alternatives. E.g., for $a = 0.4$ one notices from Table 2 ($N = \infty$) that at the cost of about 6.6% of the maximal possible power 91.2%, which a two-sided test may achieve for alternatives (F_N, G_N) as above, the present omnibus test has asymptotic power at least 49.9% for all alternatives (1.13) with $f_N = 1 + c_{N1}b$, $g_N = 1 + c_{NN}b$, $\|b\| = 3$ and $b \in [\psi_1, \dots, \psi_4]$. Notice the poor power properties of the (two-sided) Wilcoxon test for $\kappa = 2, 3, 4$ in contrast to the new test. Furthermore, a comparison of the cases $N = 20, 40, 80, \infty$ shows the good agreement of finite sample Monte Carlo results with its asymptotic counterpart. Even for the case $m = n = 10$ the qualitative power properties of the new test are the same as for $N = \infty$.

In the one-sided case we introduce the notion of *generalized shift alternatives* with $F(x) = G(x - D(x))$ for some *shift function* $D \geq 0$; see Section 1. In order to get realistic departures from the ideal shift model ($D \equiv$ const.), we consider shift functions D shifting only the "upper part," the "central part" and the "lower part" of G. More formally, we choose

$$\text{(4.1)} \qquad \begin{aligned} &\text{upper shift:} && D(x) = G(4x)/2, \\ &\text{central shift:} && D(x) = 2G(2x)(1 - G(2x)), \\ &\text{lower shift:} && D(x) = (1 - G(4x))/2, \end{aligned}$$

with G the standard normal, logistic and Cauchy df's. The results are contained in Table 3. Notice that in Table 2 the alternatives vary for $N \to \infty$ in order to demonstrate how well the asymptotic results fit the finite sample results, while in Table 3 (F, G) are the same for $N = 20, 80$. Table 3 ($a = 0.4$) shows that the new test beats the Wilcoxon test for upper-, central-, and lower-alternatives.

TABLE 4

Critical values of the omnibus test, resp. one-sided test, for the two-sample problem H_0 versus H_1^0, resp. versus H_1, based on S_N [see (2.48)], resp. S_N^0 [see (2.49)], with Parzen-2 kernel (2.33) for various bandwidths "a" at various levels α for $N = 20, 40, 80$ (obtained by 10,000 Monte Carlo runs) and for $N = \infty$ (exact asymptotic value).

		omnibus test			one-sided test		
		bandwidth			bandwidth		
α	N	0.30	0.40	0.50	0.30	0.40	0.50
0.01	20	9.93	8.45	7.38	8.40	7.16	6.27
	40	10.96	9.06	7.88	8.83	7.34	6.31
	80	11.01	9.10	7.80	9.27	7.71	6.58
	∞	11.49	9.42	8.06	—	—	—
0.05	20	7.91	6.43	5.42	6.07	4.89	4.07
	40	8.22	6.53	5.42	6.17	4.85	4.01
	80	8.28	6.48	5.42	6.01	4.74	3.89
	∞	8.40	6.61	5.46	—	—	—
0.10	20	6.96	5.42	4.48	4.84	3.74	2.98
	40	6.98	5.44	4.41	4.91	3.75	3.01
	80	6.95	5.35	4.34	4.70	3.57	2.86
	∞	7.02	5.39	4.35	—	—	—

Only for *pure* normal- and logistic-shift the Wilcoxon test wins. These differences in power are, of course, rather small for small values of N (for small N all sensible rank tests are more or less identical) but increase for increasing N. A comparison with the best possible rank test in column "opt." shows that the increase in power of the new test is substantial. In conclusion, we can recommend the new tests in situations where there are doubts that all parts of the underlying distribution are shifted at the same rate. A good choice of the bandwidth for the Parzen-2 kernel is $a = 0.4$. Some critical values of the omnibus test, resp. one-sided test, are contained in Table 4.

Acknowledgments. The author is most grateful to W. Schütt for computing the tables of the present paper. Moreover, the author wishes to thank several referees for carefully reading the paper and for suggestions, which led to an improved version of the paper.

REFERENCES

BEHNEN, K. and HUSKOVÁ, M. (1984). A simple algorithm for the adaptation of scores and power behavior of the corresponding rank test. *Comm. Statist. A—Theory Methods* **13** 305–325.

BEHNEN, K. and NEUHAUS, G. (1983). Galton's test as a linear rank test with estimated scores and its local asymptotic efficiency. *Ann. Statist.* **11** 588–599.

BEHNEN, K., NEUHAUS, G. and RUYMGAART, F. (1983). Two sample rank estimators of optimal nonparametric score-functions and corresponding adaptive rank statistics. *Ann. Statist.* **11** 1175–1189.

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

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

DAVIES, R. B. (1980). The distribution of a linear combination of χ^2 random variables. *Appl. Statist.* **29** 323–333.

DUNFORD, N. and SCHWARTZ, J. T. (1963). *Linear Operators* **2**. Wiley, New York.

GREGORY, G. G. (1977). Large sample theory for *U*-statistics and tests of fit. *Ann. Statist.* **5** 110–123.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests.* Academic, New York.

KOOPMANS, L. H. (1974). *The Spectral Analysis of Time Series.* Academic, New York.

NEUHAUS, G. (1976). Asymptotic power properties of the Cramér–von Mises test under contiguous alternatives. *J. Multivariate Anal.* **6** 95–110.

RANDLES, R. H. and HOGG, R. V. (1973). Adaptive distribution free tests. *Comm. Statist.* **2** 337–356.

ROTHE, G. (1976). Die asymptotische Verteilung degenerierter quadratischer Rangstatistiken und verwandte funktionale Grenzwertsätze. Dissertation, Univ. of Dortmund.

SCHACH, S. (1969). The asymptotic distribution of some non-linear functions of the two-sample rank vector. *Ann. Math. Statist.* **40** 1011–1020.

STRASSER, H. (1985). *Mathematical Theory of Statistics.* De Gruyter, New York.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT HAMBURG
BUNDESSTRASSE 55
2000 HAMBURG 13
FEDERAL REPUBLIC OF GERMANY