The Annals of Statistics
1993, Vol. 21, No. 1, 45–60

ASYMPTOTICALLY OPTIMAL TESTS FOR CONDITIONAL DISTRIBUTIONS

By M. Falk and F. Marohn

Katholische Universität Eichstätt

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent replicates of the random vector $(X, Y) \in \mathbb{R}^{d+m}$, where X is $\mathbb{R}^d$-valued and Y is $\mathbb{R}^m$-valued. We assume that the conditional distribution $P(Y \in \cdot | X = x) = Q_\vartheta(\cdot)$ of Y given $X = x$ is a member of a parametric family, where the parameter space Θ is an open subset of $\mathbb{R}^k$ with $0 \in \Theta$. Under suitable regularity conditions we establish upper bounds for the power functions of asymptotic level-α tests for the problem $\vartheta = 0$ against a sequence of contiguous alternatives, as well as asymptotically optimal tests which attain these bounds. Since the testing problem involves the joint density of (X, Y) as an infinite dimensional nuisance parameter, its solution is not standard. A Monte Carlo simulation exemplifies the influence of this nuisance parameter. As a main tool we establish local asymptotic normality (LAN) of certain Poisson point processes which approximately describe our initial sample.

0. Introduction. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent replicates of the random vector (X, Y), where X is $\mathbb{R}^d$-valued and Y is $\mathbb{R}^m$-valued. The main topic of classical regression analysis is the estimation of the conditional mean

$$m(x) = E(Y | X = x)$$

of Y given $X = x$ that is of particular interest in applied statistics [see, e.g., Eubank (1988) and the literature cited therein]. Only in recent years the estimation of a broader class of conditional quantities such as the conditional median has received increasing attention due to the robustness against outliers of their corresponding empirical counterparts [Härdle, Janssen and Serfling (1988), Truong (1989), Jones and Hall (1990), Bhattacharya and Gangopadhyay (1990), Manteiga (1990) and Chaudhuri (1991) among others].

While the *estimation* of conditional quantities has been playing a preeminent role in regression analysis, conditional testing problems do not seem to be deeply developed. By conditional testing problems we do not mean the problem whether a specific parameter of the underlying conditional distribution $Q(\cdot | x) = P(Y \in \cdot | X = x)$ of Y given $X = x$ such as the mean $m(x)$ or the median coincides with the hypothetical one, but we are rather interested in the problem whether the underlying conditional distribution $Q(\cdot | x)$ itself coincides with the hypothetical one. We assume that $Q(\cdot | x)$ is a member of a parametric family, where the parameter space Θ is an open subset of $\mathbb{R}^k$ with $0 \in \mathbb{R}^k$, and

Received November 1990; revised March 1992.
AMS 1991 *subject classifications*. Primary 62F03; secondary 62F05.
Key words and phrases. Conditional distribution, optimal tests, contiguous alternatives, LAN, empirical point process, Poisson point process, Monte Carlo simulation.

we will investigate the simple *conditional testing problem*

$$Q(\cdot|x) = Q_0(\cdot|x) \quad \text{against} \quad Q(\cdot|x) = Q_\vartheta(\cdot|x),$$

where $\vartheta \neq 0$.

Statistical inference on conditional quantities naturally focuses on those observations Y_i among the sample $Y_1, \dots, Y_n$ whose corresponding X values are close to the given x: Nearest neighbor, kernel and recursive partition estimators of $m(x)$ are based on local averages, the conditional median of Y given $X = x$ is computed from a local sample.

Since we observe data Y_i whose X_i values are only *close* to x in a way specified below, say $V_1, \dots, V_{K(n)}$, our set of data $V_1, \dots, V_{K(n)}$, on which we will base statistical inference, is usually *not* generated according to our target conditional distribution $Q(\cdot|x)$ of Y given $X = x$ but to some distribution which is close to $Q(\cdot|x)$. This error is determined by the joint density f of (X, Y) which is therefore some kind of infinite dimensional nuisance parameter.

Bounds for the error which one commits if the V_i are replaced by their ideal counterparts W_i being independently generated according to $Q(\cdot|x)$, were established by Falk and Reiss (1992b). This approach, by which one may study the fairly general problem of evaluating functional parameters $T(Q(\cdot|x_1), \dots, Q(\cdot|x_p))$ is as follows:

We consider only those observations Y_i among the sample $(X_1, Y_1), \dots, (X_n, Y_n)$, whose corresponding X values lie in a small cube in $\mathbb{R}^d$ with center x, that is,

$$X_i \in S_n := \left[x - a_n^{1/d}/2,\ x + a_n^{1/d}/2\right],$$

where $a_n = (a_{n,1}, \dots, a_{n,d}) \in (0, \infty)^d$ converges to zero as n increases. The operations $a_n^{1/d}/2$ are meant componentwise.

Speaking in terms of empirical point processes, we observe

$$N_n(B) := \sum_{i=1}^{n} \varepsilon_{Y_i}(B)\varepsilon_{X_i}(S_n), \qquad B \in \mathbb{B}^m,$$

where $\varepsilon_z(\cdot)$ denotes the Dirac measure with mass one at z and $\mathbb{B}^m$ is the Borel σ-algebra of $\mathbb{R}^m$. As follows from Lemma 1 in Falk and Reiss (1992a), we can write

$$N_n(B) = \sum_{i=1}^{K(n)} \varepsilon_{V_i}(B),$$

where $K(n) := N_n(\mathbb{R}^m) = \sum_{i=1}^n \varepsilon_{X_i}(S_n)$ is the number of X_i in S_n, $V_1, \dots, V_{K(n)}$ denote those Y_i whose X values fall into S_n, arranged in the original order of their outcome, and $K(n)$, V_1, $V_2, \dots$ are independent random variables (rvs).

Note that $K(n)$ is a Binomial rv with parameters n and

$$p(n) = P\{X \in S_n\} \sim \operatorname{vol}(S_n) g(x),$$

where $\operatorname{vol}(S_n) = \prod_{j=1}^d a_{n,j}^{1/d}$ is the volume of the cube S_n, and $g(x)$ denotes the marginal density of X at x which we assume to exist near x and to be positive at x. Moreover, the distribution of V_i is the conditional distribution of Y given

$X \in S_n$, denoted by $Q(\cdot|S_n)$, that is,

$$P\{V_i \in B\} = P(Y \in B | X \in S_n) = \frac{P\{Y \in B,\, X \in S_n\}}{P\{X \in S_n\}} = Q(B|S_n).$$

The Poisson approximation of the Binomial distribution therefore suggests the approximation of the empirical point process $N_n(\cdot)$ by the Poisson point process

$$N_n^*(\cdot) = \sum_{i=1}^{\tau(n)} \varepsilon_{W_i}(\cdot),$$

where $\tau(n)$ is a Poisson rv with parameter $n\,\mathrm{vol}(S_n)g(x)$, $W_1, W_2, \dots$ are independent rvs with common distribution $Q(\cdot|x)$ and $\tau(n)$, $W_1, W_2, \dots$ are independent.

In Falk and Reiss (1992b) bounds for the Hellinger distance between N_n and N_n^* were established and consequently, within these error bounds those observations Y_i, whose X values fall into the cube S_n, can jointly be handled like the ideal rvs W_i, and the number of those like the independent Poisson rv $\tau(n)$. By this approach one can therefore reduce conditional statistical problems to unconditional ones.

The size of our local data set from which we will deduce statistical inference is $N_n(\mathbb{R}^m) = K(n)$ which has expectation $np(n)$ being of order $n\,\mathrm{vol}(S_n)$. The adequate rate at which the alternatives ϑ_n for the sample size n have to converge to zero is therefore

$$\delta_n := (n\,\mathrm{vol}(S_n))^{-1/2}, \qquad n \in \mathbb{N}.$$

With this choice we will investigate in this paper the following three problems associated with the simple testing problem

$$Q_0(\cdot|x) \quad \text{against} \quad Q_{\vartheta\delta_n}(\cdot|x).$$

1. Find a semiparametric model of possible distributions of (X, Y) with conditional distribution of Y given $X = x$ being an element of $\{Q_\vartheta(\cdot|x)\colon \vartheta \in \Theta\}$, such that the Poisson process approximation described above holds uniformly on it. The joint distributions P of (X, Y) are (infinite dimensional) nuisance parameters within our approach.
2. Establish a minimum asymptotic upper bound $\beta_P(\vartheta)$ such that for any test sequence φ_n of asymptotic level α based on N_n, that is, $\limsup_{n\to\infty} E_P(\varphi_n(N_n)) \le \alpha$ with P such that $\vartheta = 0$, we have along alternatives P_n with $\vartheta_n = \vartheta\delta_n$,
$$\limsup_{n\to\infty} E_{P_n}(\varphi_n(N_n)) \le \beta_P(\vartheta).$$
3. Find an asymptotically optimal test sequence φ_n^* of (asymptotic) level α whose corresponding power functions attain this bound:
$$\lim_{n\to\infty} E_{P_n}(\varphi_n^*(N_n)) = \beta_P(\vartheta).$$

Suppose that for $\vartheta \in \Theta$ the probability measure $Q_\vartheta(\cdot|x)$ is absolutely continuous with respect to $Q_0(\cdot|x)$. The ad hoc test statistic based on N_n for testing a particular value $\vartheta \neq 0$ against the null hypothesis $\vartheta = 0$ is

$$\varphi(N_n) = 1_{(u,\infty)}\left(\sum_{i=1}^{K(n)} \log \frac{dQ_\vartheta(\cdot|x)}{dQ_0(\cdot|x)}(V_i)\right)$$
$$= 1_{(u,\infty)}\left(\sum_{i=1}^{n} \left(\log \frac{dQ_\vartheta(\cdot|x)}{dQ_0(\cdot|x)}(Y_i)\right) 1_{S_n}(X_i)\right)$$

with some level determining critical value u, which is suggested by the Neyman–Pearson lemma. Notice, however, that the distribution of the iid random variables $V_1, V_2, \dots$ is neither exactly $Q_\vartheta(\cdot|x)$ nor $Q_0(\cdot|x)$, but it is only close to one of these. This error is in addition intertwined with the marginal density $g(x)$ of X at x, which determines the asymptotic behavior of the sample size $K(n) = N_n(\mathbb{R}^m)$.

In view of this it becomes obvious that the conditional testing problem described above is actually a semiparametric one and the (asymptotic) properties of $\varphi(N_n)$ cannot be judged immediately but have to be investigated in more detail. The results in this paper show that $\varphi(N_n)$—being essentially $\varphi^*_{n,\mathrm{opt}}(N_n)$ in Theorem 1.7—is in fact asymptotically optimal for particular sequences ϑ_n of alternatives iff the corresponding sequence of marginal densities $g_n(x)$ can be neglected in a proper sense; if this sequence cannot be neglected, then $\varphi(N_n)$ loses its asymptotic optimality along ϑ_n. Our investigations will be carried out within the framework of LAN theory [see Le Cam (1986), Strasser (1985) and Ibragimov and Has'minskii (1981, 1991)]. For a general theory on semiparametric problems we refer to Pfanzagl (1990) and the literature cited therein.

By $\langle \cdot , \cdot \rangle$ we denote the usual inner product of the Euclidean space and by $\| \; \|$ the norm induced by $\langle \cdot , \cdot \rangle$. We denote by $\mathscr{L}(N_n)$ the distribution of N_n with (X, Y) and so on. By $H(\cdot, \cdot)$ we denote the Hellinger distance between two distributions on the same space.

1. Model assumptions and main results. We suppose that the rv (X, Y) has a density f on a strip $[x - \varepsilon_0, x + \varepsilon_0] \times \mathbb{R}^m$ $(\subset \mathbb{R}^{d+m})$, which we decompose as

$$f(z,y) = g(z)q(y|z), \qquad z \in [x - \varepsilon_0, x + \varepsilon_0], y \in \mathbb{R}^m,$$

where g denotes the marginal density of X and $q(\cdot|z)$ the conditional density of Y given $X = z$.

We require (g, q) to be a member of the following class of smooth functions

$$(\mathscr{G}, \mathscr{Q}) := (\mathscr{G}, \mathscr{Q})(C_1, C_2, C_3)$$
$$:= \Big\{ (g,q) \colon g \colon [x - \varepsilon_0, x + \varepsilon_0] \to [0,\infty),\ q(\cdot|\cdot) \colon \mathbb{R}^m$$
$$\times [x - \varepsilon_0, x + \varepsilon_0] \to [0,\infty)$$

such that $0 < g(x) \le C_1$ and for any $\varepsilon \in (0, \varepsilon_0]$,

$$|(g(x+\varepsilon)q(y|x+\varepsilon)/(g(x)q(y|x)))^{1/2} - (1 + \langle \varepsilon, h_{(g,q)}(y)\rangle)|$$
$$\le C_2\|\varepsilon\|^2 r_{(g,q)}(y)$$

for some functions $h_{(g,q)}\colon \mathbb{R}^m \to \mathbb{R}^d$, $r_{(g,q)}\colon \mathbb{R}^m \to \mathbb{R}$ satisfying

$$\int \left(\|h_{(g,q)}(y)\|^4 + r^4_{(g,q)}(y)\right) q(y|x)\,dy \le C_3 \Big\},$$

where C_1, C_2, C_3 are fixed positive constants. The function $h_{(g,q)}$ in the linear approximation above, which is suggested by Taylor's formula, reflects the dependence between the conditional distributions of Y given $X = x$ and X near x.

Denote by q_ϑ the Lebesgue density of $Q_\vartheta(\cdot) = Q_\vartheta(\cdot|x)$, $\vartheta \in \Theta$, where Θ is an open subset of $\mathbb{R}^k$ with $0 \in \Theta$. The class of possible distributions P of (X, Y) which we consider is then characterized by

$$\mathscr{P}((\mathscr{G}, \mathscr{Q}), \Theta) := \{P|\mathbb{R}^{d+m}\colon\ P \text{ has density } g(z)q(y|z) \text{ on } [x-\varepsilon_0, x+\varepsilon_0] \times \mathbb{R}^m \text{ such that } (g,q) \in (\mathscr{G}, \mathscr{Q}) \text{ and } q(\cdot|x) \in \{q_\vartheta\colon \vartheta \in \Theta\}\}.$$

Note that $\mathscr{P}((\mathscr{G}, \mathscr{Q}), \Theta)$ forms a semiparametric family of distributions, with the vector $(g, q) \in (\mathscr{G}, \mathscr{Q})$ of marginal and conditional densities of (X, Y) over the interval $[x - \varepsilon_0, x + \varepsilon_0]$ being the nonparametric part (in which we are primarily not interested), and Θ being the k-dimensional parametric part (we are primarily interested in). As a consequence, we index expectations, distributions and so on by $E_{(g,q),\vartheta}$, $\mathscr{L}_{(g,q),\vartheta}$ and so on.

The main tool for the solution of problems 2 and 3 formulated above is the following Lemma 1.1, which is immediate from the proof of Theorem 2 in Falk and Reiss (1992b). By this result we can handle our data $V_1, \dots, V_{K(n)}$ within a certain error bound as being independently generated according to Q_ϑ, where the independent sample size is a Poisson rv with parameter $n\,\mathrm{vol}(S_n)g(x)$; in other words, we can handle the empirical point process N_n (which we observe) as the ideal Poisson process N_n^*. For this ideal situation we will serve problem 2 and 3 first (see Theorem 1.2 and Corollaries 1.3 and 1.4). These results will then carry over to our actual data N_n (see Theorem 1.7).

1.1 LEMMA. *We have*

$$\sup_{\mathscr{P}((\mathscr{G}, \mathscr{Q}), \Theta)} H\left(\mathscr{L}_{(g,q),\vartheta}(N_n), \mathscr{L}_{g(x),\vartheta}(N_n^*)\right)$$
$$= O\left(\mathrm{vol}(S_n) + (n\,\mathrm{vol}(S_n))^{1/2}\|a_n^{1/d}\|^2\right).$$

Notice that in the preceding result the distribution of the Poisson process $N_n^*(\cdot) = \sum_{i=1}^{\tau(n)} \varepsilon_{W_i}(\cdot)$ depends only on ϑ and the real parameter $g(x)$, with $n\,\mathrm{vol}(S_n)g(x)$ being the expectation of the Poisson variable $\tau(n)$.

By the preceding model approximation we can reduce the semiparametric problem $\mathscr{L}_{(g,q),\vartheta}(N_n)$ with unknown $(g,q) \in (\mathscr{G}, \mathscr{Q})$ and $\vartheta \in \Theta$ to the $(k+1)$-dimensional parametric problem

$$\mathscr{L}_{c,\vartheta}(N_n^*) = \mathscr{L}_{c,\vartheta}\left(\sum_{i=1}^{\tau(n)} \varepsilon_{W_i}\right),$$

where $\tau(n)$ is a Poisson variable with expectation $n\,\mathrm{vol}(S_n)c$, $c \in (0, C_1]$, $W_1, W_2, \ldots$ are iid rvs with distribution Q_ϑ and $\tau(n)$ and $W_1, W_2, \ldots$ are independent.

Note that a Binomial process approximation of N_n, where V_i is replaced by W_i but their number $K(n)$ being kept, does not improve the bound in Lemma 1.1 essentially. We may therefore benefit from the technical ease which we gain by utilizing the Poisson process approximation.

If Q_ϑ is absolutely continuous w.r.t. Q_0 we obtain from Theorem 3.1.1 in Reiss (1993) that $\mathscr{L}_{d,\vartheta\delta_n}(N_n^*)$ is absolutely continuous w.r.t. $\mathscr{L}_{c,0}(N_n^*)$ with density

$$L_{n,d,\vartheta}^*(\mu) := \frac{d\mathscr{L}_{d,\vartheta\delta_n}(N_n^*)}{d\mathscr{L}_{c,0}(N_n^*)}(\mu)$$

$$(1) \qquad = \exp\left(\sum_{i=1}^{\mu(\mathbb{R}^m)} \log \frac{dQ_{\vartheta\delta_n}}{dQ_0}(w_i) + \mu(\mathbb{R}^m)\log\left(\frac{d}{c}\right) + n\,\mathrm{vol}(S_n)(c-d)\right),$$

where $\mu = \sum_{i=1}^{\mu(\mathbb{R}^m)} \varepsilon_{w_i}$, $\mu(\mathbb{R}^m) < \infty$, is an atomization of a (finite) point measure μ on $\mathbb{R}^m$.

Fix $c > 0$. By the Neyman–Pearson lemma, the best test of level α based on N_n^* for the testing problem

$$(c, \vartheta) = (c, 0) \quad \text{against} \quad (c_n, \vartheta_n)$$

is $\varphi_n^*(N_n^*)$, where

$$\varphi_n^*(\mu) := 1_{(u_{n,\alpha},\infty)}\left(\log\left(\frac{d\mathscr{L}_{c_n,\vartheta_n}(N_n^*)}{d\mathscr{L}_{c,0}(N_n^*)}(\mu)\right)\right) + \gamma_n 1_{\{u_{n,\alpha}\}}\left(\log\left(\frac{d\mathscr{L}_{c_n,\vartheta_n}(N_n^*)}{d\mathscr{L}_{c,0}(N_n^*)}(\mu)\right)\right),$$

$\gamma_n \in [0,1]$ and $u_{n,\alpha}$, γ_n satisfy

$$E_{c,0}(\varphi_n^*(N_n^*)) = \alpha.$$

If we choose $(c_n, \vartheta_n) = (c + o(\delta_n), \vartheta\delta_n)$, then the remainder term

$$\log(L_{n,c_n,\vartheta}^*(N_n^*)) - \sum_{i=1}^{N_n^*(\mathbb{R}^m)} \log \frac{dQ_{\vartheta_n}}{dQ_0}(W_i)$$

vanishes asymptotically under $(c, 0)$ and (c_n, ϑ_n), and thus,

$$\varphi_{n,\mathrm{opt}}(\mu) := 1_{(u_{n,\alpha},\infty)}\left(\int \log \frac{dQ_{\vartheta_n}}{dQ_0}\, d\mu\right)$$

is asymptotically equivalent to φ_n^* (whenever the randomization can asymptotically be neglected), compare the proof of Theorem 1.2.

Notice that

$$\varphi_{n,\mathrm{opt}}(N_n^*) = 1_{(u_n,\infty)}\left(\sum_{i=1}^{\tau(n)} \log \frac{q_{\vartheta_n}}{q_0}(W_i)\right)$$

is the ad hoc statistic which one would use for testing $\vartheta = 0$ against ϑ_n based on N_n^*. We will show in the following that $\varphi_{n,\mathrm{opt}}(N_n^*)$ is in fact an asymptotically optimal level α test for this problem along the alternatives

$$(c_n, \vartheta_n) = (c + o(\delta_n), \vartheta\delta_n).$$

Since $\varphi_{n,\mathrm{opt}}(N_n^*)$ does not depend on c as shown below, it is asymptotically optimal along these alternatives *uniformly* in c.

If we allow however a slower rate of convergence of c_n, that is, if we consider

$$(c_n, \vartheta_n) = (c + O(\delta_n), \vartheta\delta_n),$$

then the nuisance parameter c_n becomes relevant and $\varphi_{n,\mathrm{opt}}(N_n^*)$ loses its asymptotic optimality along the alternatives (c_n, ϑ_n); see Corollary 1.3 and 1.4. By the bound for the model approximation established in Lemma 1.1, the considerations carry over to $\varphi_{n,\mathrm{opt}}$ applied to our real data set, that is, the empirical point process N_n.

In order to establish the limit of the power functions $E_{c_n,\vartheta_n}(\varphi_n^*(N_n^*))$, we require Hellinger differentiability of q_ϑ at zero:

(A) $$q_\vartheta^{1/2}(\cdot) = q_0^{1/2}(\cdot)\{1 + \langle \vartheta, v(\cdot)\rangle/2 + \|\vartheta\|\tilde{r}_\vartheta(\cdot)\},$$

with derivative $v = (v_1, \dots, v_k)$, $v_j \in \mathscr{L}_2(Q_0)$, $j = 1, \dots, k$, and $\|\tilde{r}_\vartheta\|_{L_2(Q_0)} = (\int \tilde{r}_\vartheta^2\, dQ_0)^{1/2} = o(\|\vartheta\|^0)$.

In the following we consider alternatives of the form

$$c_n(\eta) = c + \eta\delta_n + o(\delta_n), \qquad \eta \in \mathbb{R},$$
$$\vartheta_n = \vartheta\delta_n$$

and the corresponding sequence of binary experiments

$$E_n = \left(M(\mathbb{R}^m), \mathscr{M}(\mathbb{R}^m), \{\mathscr{L}_{c,0}(N_n^*), \mathscr{L}_{c_n(\eta),\vartheta_n}(N_n^*)\}\right).$$

By $M(\mathbb{R}^m)$ we denote the set of point-measures on $\mathbb{R}^m$ and $\mathscr{M}(\mathbb{R}^m)$ denotes the smallest σ-algebra such that for any $B \in \mathbb{B}^m$ the projection π_B: $M(\mathbb{R}^m) \to \{0, 1, 2, \dots\}$, $\pi_B(\mu) := \mu(B)$ is measurable.

1.2 THEOREM [LAN of $(E_n)_n$]. *Fix $c > 0$. Under condition (A) we have with $c_n(\eta) = c + \eta\delta_n + o(\delta_n)$ and $\vartheta_n = \vartheta\delta_n$*

$$\log\left(\frac{d\mathscr{L}_{c_n(\eta),\vartheta_n}(N_n^*)}{d\mathscr{L}_{c,0}(N_n^*)}(\cdot)\right) = \langle(\eta,\vartheta),(Z_n^{(1)}(\cdot),Z_n^{(2)}(\cdot))\rangle_{c,0,+} - \frac{1}{2}\|(\eta,\vartheta)\|_{c,0,+}^2 + R_n(\cdot) \tag{2}$$

with central sequence $(Z_n^{(1)}, Z_n^{(2)})$: $M(\mathbb{R}^m) \to \mathbb{R}^{k+1}$ given by

$$Z_n^{(1)}(\mu) = \delta_n(\mu(\mathbb{R}^m) - c\delta_n^{-2}),$$

$$Z_n^{(2)}(\mu) = (\delta_n\mu(\mathbb{R}^m))^{-1}\Gamma_0^{-1}\int v\,d\mu$$

and $R_n \to 0$ in probability under $(c, 0)$; the inner product $\langle\cdot,\cdot\rangle_{c,0,+}$ on $\mathbb{R}^{1+k}$ is defined by

$$\langle(a,s),(b,t)\rangle_{c,0,+} := ab/c + s'c\Gamma_0 t, \qquad a, b, \in \mathbb{R},\ s,t \in \mathbb{R}^k$$

and the matrix $\Gamma_0 = (\int v_i v_j\, dQ_0)_{i,j=1,\dots,k}$ is assumed to be positive definite.

The first coordinate $Z_n^{(1)}$ of the central sequence depends on the localization point c which cannot be eliminated or replaced by an adaptive estimator without affecting the asymptotics. If we replace for example c simply by $\mu(\mathbb{R}^m)/(n\operatorname{vol}(S_n))$, the term $Z_n^{(1)}$ vanishes.

Theorem 1.2 implies the following results, where $u_\alpha = \Phi^{-1}(1-\alpha)$ denotes the $(1-\alpha)$-quantile of the standard normal distribution function Φ.

1.3 COROLLARY. *The test sequence*

$$\varphi_{n,\eta,opt}^*(N_n^*) := 1_{(u_\alpha\|(\eta,\vartheta)\|_{c,0,+},\infty)}\left(\langle(\eta,\vartheta),(Z_n^{(1)}(N_n^*),Z_n^{(2)}(N_n^*))\rangle_{c,0,+}\right) \tag{3}$$

is asymptotically optimal for testing $(c, 0)$ against $(c_n(\eta), \vartheta_n)$ at level α based on N_n^*. We have

$$\lim_{n\to\infty} E_{c,0}(\varphi_{n,\eta,opt}^*(N_n^*)) = \alpha$$

and

$$\beta_c(\eta,\vartheta) := \lim_{n\to\infty} E_{c_n(\eta),\vartheta_n}(\varphi_{n,\eta,opt}^*(N_n^*)) = 1 - \Phi(u_\alpha - \|(\eta,\vartheta)\|_{c,0,+}).$$

1.4 COROLLARY. *For $\eta = 0$, that is, $c_n := c_n(0) = c + o(\delta_n)$, the sequence E_n converges weakly to the Gaussian–Shift on $(\{0,\vartheta\}, \langle\cdot,\cdot\rangle_{c,0})$, where*

$$\langle s,t\rangle_{c,0} := s'c\Gamma_0 t, \qquad s,t \in \mathbb{R}^k.$$

The central sequence is now $Z_n^{(2)}$ which is independent of c. Moreover, the test sequence

$$\varphi_{n,opt}^*(N_n^*) := 1_{(u_\alpha(\vartheta'\Gamma_0\vartheta)^{1/2},\infty)}\left(\vartheta'(N_n^*(\mathbb{R}^m))^{-1/2}\int v\,dN_n^*\right),$$

which is independent of c, *is asymptotically equivalent to* $\varphi^*_{n,0,opt}(N_n^*)$. *Consequently,* $\varphi^*_{n,opt}(N_n^*)$ *is asymptotically optimal for* $(c, 0)$ *against* (c_n, ϑ_n) *at level* α *uniformly for* $c > 0$. *In this case, the upper bound is* $\beta_c(0, \vartheta) = 1 - \Phi(u_\alpha - \|\vartheta\|_{c,0})$.

Proof. The asymptotical equivalence follows from the fact that $\delta_n^{-1} c^{1/2} \sim N_n^*(\mathbb{R}^m)^{1/2}$ under $(c, 0)$ and, by contiguity, also under (c_n, ϑ_n). □

Notice that in the case of a one-dimensional parameter space, that is, $k = 1$, the test sequence $\varphi^*_{n,\mathrm{opt}}(N_n^*)$ is independent of ϑ up to the sign of ϑ. Hence, $\varphi^*_{n,\mathrm{opt}}(N_n^*)$ is also optimal uniformly for $\vartheta > 0$ or $\vartheta < 0$. We do not know whether there exists a test sequence which is asymptotically optimal *uniformly* in c if $\eta \neq 0$.

In order to prove Theorem 1.2, we need the following auxiliary results which are of interest of their own.

1.5 Lemma. *Let* $(\Omega, \mathscr{A})$ *be a measurable space supporting a Poisson process* N_n^*. *Suppose that under* P_t, $t \in (-\varepsilon, \varepsilon)$, N_n^* *has the intensity measure* $\lambda_t(n) Q_t(\cdot)$, *where* $\lambda_t(n) \in (0, \infty)$ *and* Q_t *is a probability measure on* $\mathbb{R}^m$ *dominated by the m-dimensional Lebesgue measure. If the curve* $t \to Q_t$ *is Hellinger differentiable at* 0 *with derivative* v *and* $\lambda_0(n) \to \infty$ *as* $n \to \infty$, *then the following expansion holds with* $\delta_n = (\lambda_0(n))^{-1/2}$:

$$\int \log \frac{dQ_{\delta_n}}{dQ_0}\, dN_n^* = \delta_n \int v\, dN_n^* - \frac{1}{2}\int v^2\, dQ_0 + R_n(N_n^*),$$

where $P_0\{|R_n(N_n^*)| > \varepsilon\} = \mathscr{L}_0(N_n^*)\{|R_n| > \varepsilon\}$ *converges to zero for* $n \to \infty$ *and each* $\varepsilon > 0$.

Proof. Using conditioning techniques, the proof runs along the lines of the proof in the classical situation [see, e.g., Strasser (1985), Chapter 12]. Note that $\tau(n)\delta_n^2 \to 1$ in P_0 probability. □

The following result is immediate from Lemma 1 in Falk and Reiss (1992b) and the Cramér–Wold device.

1.6 Lemma. *Let* $N_n^* = \sum_{i=1}^{\tau(n)} \varepsilon_{X_i}$ *be a Poisson process* [*over some probability space* $(\Omega, \mathscr{A}, P)$] *with intensity measure* $EN_n^*(\cdot) = \lambda(n) Q(\cdot)$, *where* $\tau(n)$ *is a Poisson rv with* $E\tau(n) = \lambda(n) \to \infty$ *as* $n \to \infty$ *and* Q *denotes the distribution of the independent,* $\mathbb{R}^m$*-valued rvs* $X_1, X_2, \dots$ *being independent of* $\tau(n)$. *Let* $v_i \in L_2(Q)$, $\int v_i\, dQ = 0$, $i = 1, \dots, k$, *and* $\Gamma = (\int v_i v_j\, dQ)_{i,j=1,\dots,k}$. *Then*

$$(\lambda(n))^{-1/2} \int v\, dN_n^* \to_{\mathscr{D}} \mathscr{N}(0, \Gamma),$$

where $\to_{\mathscr{D}}$ *denotes convergence in distribution and* $\mathscr{N}(\zeta, \Sigma)$ *denotes the normal distribution on the Euclidean space with mean vector* ζ *and covariance matrix* Σ.

Next we will establish the proof of Theorem 1.2.

PROOF OF THEOREM 1.2. By condition (A) the curve $t \to Q_{t\vartheta}$ is differentiable at zero with tangent vector $v_\vartheta = \langle \vartheta, v\rangle$. With $\tilde{\vartheta} = c^{1/2}\vartheta$ and $\tilde{\delta}_n = c^{-1/2}\delta_n$ we obtain from Lemmas 1.5 and 1.6 the expansion

$$\int \log \frac{dQ_{\tilde{\vartheta}\tilde{\delta}_n}}{dQ_0}\, dN_n^* = \tilde{\delta}_n \int \langle \tilde{\vartheta}, v\rangle\, dN_n^* - \frac{1}{2}\int \langle \tilde{\vartheta}, v\rangle^2\, dQ_0 + R_{1,n}(N_n^*)$$

with

$$\Gamma_0^{-1}\tilde{\delta}_n \int v\, dN_n^* \to_{\mathscr{D}} \mathscr{N}(0, \Gamma_0^{-1}).$$

Since $\delta_n^2 N_n^*(\mathbb{R}^m) = \delta_n^2 \tau(n) \to c$ in probability we get

$$\int \log \frac{dQ_{\vartheta\delta_n}}{dQ_0}\, dN_n^* = \langle \vartheta, Z_n^{(2)}(N_n^*)\rangle_{c,0} - \frac{1}{2}\|\vartheta\|_{c,0}^2 + R_{2,n}(N_n^*), \tag{4}$$

where $Z_n^{(2)}(N_n^*)$ converges weakly under $(c,0)$ to the standard normal distribution on $(\mathbb{R}^k, \langle\,\cdot\,,\,\cdot\,\rangle_{c,0})$, that is,

$$Z_n^{(2)}(N_n^*) \to_{\mathscr{D}} \mathscr{N}(0, c^{-1}\Gamma_0^{-1}) \tag{5}$$

and $R_{2,n}(N_n^*) \to 0$ in probability under $(c,0)$ (for the definition of $\langle\,\cdot\,,\,\cdot\,\rangle_{c,0}$ see Corollary 1.4). Straightforward calculations show that the remainder term of the expansion (1)

$$\begin{aligned} R_{3,n}(N_n^*) &:= \log L_{n, c_n(\eta), \vartheta}^*(N_n^*) - \int \log \frac{dQ_{\vartheta\delta_n}}{dQ_0}\, dN_n^* \\ &= \tau(n)\log\left(\frac{c_n(\eta)}{c}\right) + n\,\mathrm{vol}(S_n)(c - c_n(\eta)) \end{aligned}$$

has the representation

$$\frac{\eta}{c}\delta_n\left(N_n^*(\mathbb{R}^m) - c\delta_n^2\right) - \frac{\eta^2}{2c} + R_{4,n}(N_n^*)$$

with $R_{4,n}(N_n^*) \to 0$ in probability under $(c,0)$. Then (4) and (5) imply the expansion (2).

Straightforward but lengthy calculations show that $(Z_n^{(1)}, Z_n^{(2)})$ converges weakly to the standard normal distribution on $(\mathbb{R}^{k+1}, \langle\,\cdot\,,\,\cdot\,\rangle_{c,0,+})$ under $(c,0)$, that is,

$$\mathscr{L}_{c,0}\left(Z_n^{(1)}(N_n^*), Z_n^{(2)}(N_n^*)\right) \to_{\mathscr{D}} \mathscr{N}(0,c) \times \mathscr{N}(0, c^{-1}\Gamma_0^{-1}),$$

where $\times$ denotes the product measure. The proof is complete. □

With the preceding notations, the following main result of this paper—which gives an answer to problems 2 and 3 mentioned in the introduction—is a straightforward consequence of Lemma 1.1 and Corollaries 1.3 and 1.4. The asymptotically optimal test sequence $\varphi^*_{n,\eta,\mathrm{opt}}$ defined in (3) turns out to be also an asymptotically optimal level α test for testing $Q_0(\cdot|x)$ against $Q_{\vartheta\delta_n}(\cdot|x)$ if applied to the empirical point process N_n.

1.7 THEOREM. *Consider the testing problem*

$$Q_0(\cdot|x) \quad \textit{against} \quad Q_{\vartheta\delta_n}(\cdot|x).$$

Let $(\varphi_n)_n$ *be a test sequence of asymptotic level* α *based on* N_n, *that is,*

$$\limsup_{n\to\infty} E_{(g,q),0}(\varphi_n(N_n)) \le \alpha$$

for any $(g,q)\in(\mathscr{G},\mathscr{Q})$ *with* $q(\cdot|x)=q_0(\cdot)$. *If* $\|a_n\|\to 0$, $n\,\mathrm{vol}(S_n)\|a_n^{1/d}\|^4\to 0$ *and* $n\,\mathrm{vol}(S_n)\to\infty$, *then under condition* (A) *we have for any sequence* $(g_n,q_n)\in(\mathscr{G},\mathscr{Q})$ *with* $g_n(x)=g(x)+\eta\delta_n+o(\delta_n)$ *and* $q_n(\cdot|x)=q_{\vartheta\delta_n}(\cdot)$:

(i) $\limsup_{n\to\infty} E_{(g_n,q_n),\vartheta\delta_n}(\varphi_n(N_n)) \le \beta_{g(x)}(\eta,\vartheta)$
$= 1-\Phi(u_\alpha - \|(\eta,\vartheta)\|_{g(x),0,+})$

(ii) $\lim_{n\to\infty} E_{(g,q),0}(\varphi^*_{n,\eta,opt}(N_n)) = \alpha$ *and*

$$\lim_{n\to\infty} E_{(g_n,q_n),\vartheta\delta_n}\big(\varphi^*_{n,\eta,opt}(N_n)\big) = 1-\Phi\big(u_\alpha - \|(\eta,\vartheta)\|_{g(x),0,+}\big),$$

that is, $(\varphi^*_{n,\eta,opt})_n$ *as defined in* (3) *yields an asymptotically optimal test sequence for* $(g(x),0)$ *against* $(g_n(x),\vartheta\delta_n)=(g(x)+\eta\delta_n+o(\delta_n),\vartheta\delta_n)$ *based on* N_n *which is of asymptotic level* α.

In the case $\eta=0$ *the test sequence* $\varphi^*_{n,\eta,opt}(N_n)$ *is asymptotically equivalent to*

$$\varphi^*_{n,opt}(N_n) = 1_{(u_\alpha(\vartheta'\Gamma_0\vartheta)^{1/2},\infty)}\left(\vartheta'(N_n(\mathbb{R}^m))^{-1/2}\int v\,dN_n\right),$$

which does not depend on $g_n(x)$, $g(x)$ *and which is therefore asymptotically optimal, uniformly in* (g,q), *for* $(g(x),0)$ *against* $(g_n(x),\vartheta\delta_n)$.

The preceding results show in particular that the test sequence $\varphi^*_{n,\mathrm{opt}}(N_n)$, which is asymptotically equivalent to the ad hoc test

$$1_{(u_{n,\alpha},\infty)}\left(\sum_{i=1}^{K(n)}\log\big(q_{\vartheta_n}(V_i)/q_0(V_i)\big)\right)$$

defined in the introduction, is an asymptotic level α test for $\vartheta=0$ for any $(g,q)\in(\mathscr{G},\mathscr{Q})$ with $q(\cdot|x)=q_0(\cdot)$. But it is asymptotically optimal along alternatives ϑ_n with $(g_n,q_n)\in(\mathscr{G},\mathscr{Q})$, $q_n(\cdot|x)=q_{\vartheta_n}(\cdot)$ if and only if $g_n(x)=g(x)+o(\delta_n)$, in which case the nuisance parameter $g(x)$ can be neglected.

REMARK. If we choose $a_{n,1}=\cdots=a_{n,d}=b_n$, then we obtain $\mathrm{vol}(S_n)=b_n$, $n\,\mathrm{vol}(S_n)\|a_n^{1/d}\|^4=O(nb_n^{(d+4)/d})$ and $\delta_n=(nb_n)^{-1/2}$. The choice $b_n=\varepsilon_n^2 n^{-d/(d+4)}$ with $\varepsilon_n\to 0$ results in δ_n of minimum order $O(\varepsilon_n^{-1}n^{-2/(d+4)})$. Note that this is up to ε_n^{-1} the optimal (local) accuracy of estimation of a twice continuously differentiable (i.e., nonparametric) mean regression curve [cf. Stone (1982), Millar (1982), Nussbaum (1985), Truong (1989) and Chaudhuri

(1991) for a corresponding result for quantile regression and the literature cited therein]. It is well known that in regular families of distributions there exists no test sequence which detects alternatives approaching the hypothesis at the optimal rate that estimators achieve. This explains the factor ε_n^{-1} in the above optimal rate.

A data based version of $\varphi^*_{n,\mathrm{opt}}(N_n)$ with (asymptotically optimal) binwidth a_n automatically chosen would clearly be desirable. Such adaptive selection techniques are well known in nonparametric curve estimation [see, e.g., the survey by Marron (1989)]. But, to the best of our knowledge, the derivation of corresponding (optimal) automatic selection rules for our particular testing situation, seems to be an open problem.

PROOF OF THEOREM 1.7. Since the total variation distance is bounded by the Hellinger distance [cf. Lemma 3.3.9 in Reiss (1989)], Lemma 1.1 implies uniformly for any $\vartheta \in \Theta$, $(g, q) \in (\mathscr{G}, \mathscr{Q})$ with $q(\cdot|x) = q_\vartheta(\cdot)$

$$\sup_{M \in \mathscr{M}(\mathbb{R}^m)} \left| \mathscr{L}_{(g,q),\vartheta}(N_n)(M) - \mathscr{L}_{c,\vartheta}(N_n^*)(M) \right| = o(1) \tag{6}$$

as $n \to \infty$ where $c = g(x)$. Hence,

$$\left| E_{(g,q),0}\big(\varphi^*_{n,\eta,\mathrm{opt}}(N_n)\big) - E_{c,0}\big(\varphi^*_{n,\eta,\mathrm{opt}}(N_n^*)\big) \right| \to 0$$

and

$$\left| E_{(g_n,q_n),\vartheta_n}\big(\varphi^*_{n,\eta,\mathrm{opt}}(N_n)\big) - E_{c_n,\vartheta_n}\big(\varphi^*_{n,\eta,\mathrm{opt}}(N_n^*)\big) \right| \to 0$$

as $n \to \infty$ with $c_n(\eta) = g_n(x) = g(x) + \eta\delta_n + o(\delta_n)$. Furthermore,

$$\begin{aligned}
&\liminf_{n\to\infty}\Big(E_{c_n(\eta),\vartheta_n}\big(\varphi^*_{n,\eta,\mathrm{opt}}(N_n^*)\big) - E_{(g_n,q_n),\vartheta_n}\big(\varphi_n(N_n)\big)\Big)\\
&\quad\ge \liminf_{n\to\infty}\Big(E_{c_n(\eta),\vartheta_n}\big(\varphi^*_{n,\eta,\mathrm{opt}}(N_n^*)\big) - E_{c_n(\eta),\vartheta_n}\big(\varphi_n(N_n^*)\big)\Big)\\
&\qquad + \liminf_{n\to\infty}\Big(E_{c_n(\eta),\vartheta_n}\big(\varphi_n(N_n^*)\big) - E_{(g_n,q_n),\vartheta_n}\big(\varphi_n(N_n)\big)\Big)\\
&\quad= \beta_c(\eta,\vartheta) - \limsup_{n\to\infty} E_{c_n(\eta),\vartheta_n}\big(\varphi_n(N_n^*)\big) + 0 \ge 0,
\end{aligned}$$

where the last inequality follows from (6) and $\limsup_{n\to\infty} E_{c,0}(\varphi_n(N_n^*)) \le \alpha$. □

2. A simulation study. In this section we briefly report some Monte Carlo simulations for the testing problem considered in this paper, which exemplify the influence of the nuisance parameter on the finite sample behavior of the asymptotically optimal procedures, derived in the previous section.

Consider $X \sim \mathscr{N}(0, \sigma^2)$, $Z \sim \mathscr{N}(0, 1)$, independent and both in $\mathbb{R}^1$ and define for $\vartheta \in \Theta := (-1, \infty)$ the vector

$$(X, Y) := (X, \rho X + (1 + \vartheta)Z),$$

where the parameter $\rho \in \mathbb{R}$ determines the dependence between X and Y.

Obviously, we have in this case with $x = 0$

$$Q_\vartheta(\cdot|0) = Q_\vartheta(\cdot) = P_\vartheta(Y \in \cdot|X = 0) = \mathscr{N}\big(0, (1+\vartheta)^2\big)(\cdot),$$

independent of ρ and $\sigma^2 > 0$, whereas $Y \sim \mathscr{N}(0, \rho^2\sigma^2 + (1+\vartheta)^2)$. Our testing problem is now $\vartheta = 0$ against $\vartheta \neq 0$.

The joint density of (X, Y) is given by

$$f(z,y) = \frac{1}{\sigma}\phi\left(\frac{z}{\sigma}\right)\frac{1}{1+\vartheta}\phi\left(\frac{y-\rho z}{1+\vartheta}\right) = g(z)q(y|z), \qquad z, y \in \mathbb{R},$$

where ϕ denotes the standard normal density, and the conditional density q_ϑ of Y given $X = 0$ is simply

$$q_\vartheta(y) = \frac{1}{1+\vartheta}\phi\left(\frac{y}{1+\vartheta}\right), \qquad y \in \mathbb{R}.$$

Notice that in this specific example the joint density f depends on the three parameters $\vartheta > -1$, $\rho \in \mathbb{R}$ and $\sigma > 0$ with ρ and σ being nuisance parameters of a different character: While σ essentially determines the expected sample size,

$$E(K(n)) = E\left(\sum_{i=1}^{n} 1_{[-a_n/2, a_n/2]}(X_i)\right) \sim n a_n (2\pi)^{-1/2}\sigma^{-1}$$

of our Y_i data with $X_i \in [-a_n/2, a_n/2]$, the structural parameter ρ roughly controls the joint distribution of the vector (X, Y).

Taylor expansion of the exponential function at zero implies the expansion

$$\left(\frac{g(\varepsilon)q(y|\varepsilon)}{g(0)q(y|0)}\right)^{1/2} - 1 = \varepsilon\frac{\rho}{2(1+\vartheta)^2}y + O\left(\varepsilon^2 \exp\left(\frac{|y\rho| + \rho^2}{(1+\vartheta)^2}\right)\right)$$

$$=: \varepsilon h_{(g,q)}(y) + O\big(\varepsilon^2 r_{(g,q)}(y)\big)$$

uniformly for y, $\rho \in \mathbb{R}$, $\vartheta > -1$, ε small and $1/\sigma \leq C_1$. Observe that $\int (h_{(g,q)}(y)^4 + r_{(g,q)}(y)^4) q(y|0)\, dy < \infty$.

Easy calculations show that the family $\{Q_\vartheta\} = \{\mathscr{N}(0, (1+\vartheta)^2)\}$ is Hellinger differentiable at $\vartheta = 0$ with derivative $v(z) = z^2 - 1$ and variance $\Gamma_0 = \int v^2(z) Q_0(dz) = \int (z^2-1)^2 \mathscr{N}(0,1)(dz) = 2$. Up to a normalizing factor, the central sequence $N_n(\mathbb{R})^{-1/2}\int v\, dN_n = K(n)^{-1/2}\int v\, dN_n$ becomes in this case

$$Z_n := K(n)^{-1/2}\sum_{i=1}^{n}\big(Y_i^2 - 1\big)1_{[-a_n/2, a_n/2]}(X_i),$$

which is approximately normal with mean zero and variance 2 under $(g, q) \in (\mathscr{G}, \mathscr{Q})$ with $q(\cdot|x) = q_0(\cdot)$.

According to Theorem 1.7, the asymptotically optimal test for testing $\vartheta = 0$ against $\vartheta_n = \vartheta\delta_n = \vartheta(n a_n)^{-1/2}$ uniformly for $\vartheta > 0$ along $(g_n, q_n) \in (\mathscr{G}, \mathscr{Q})$

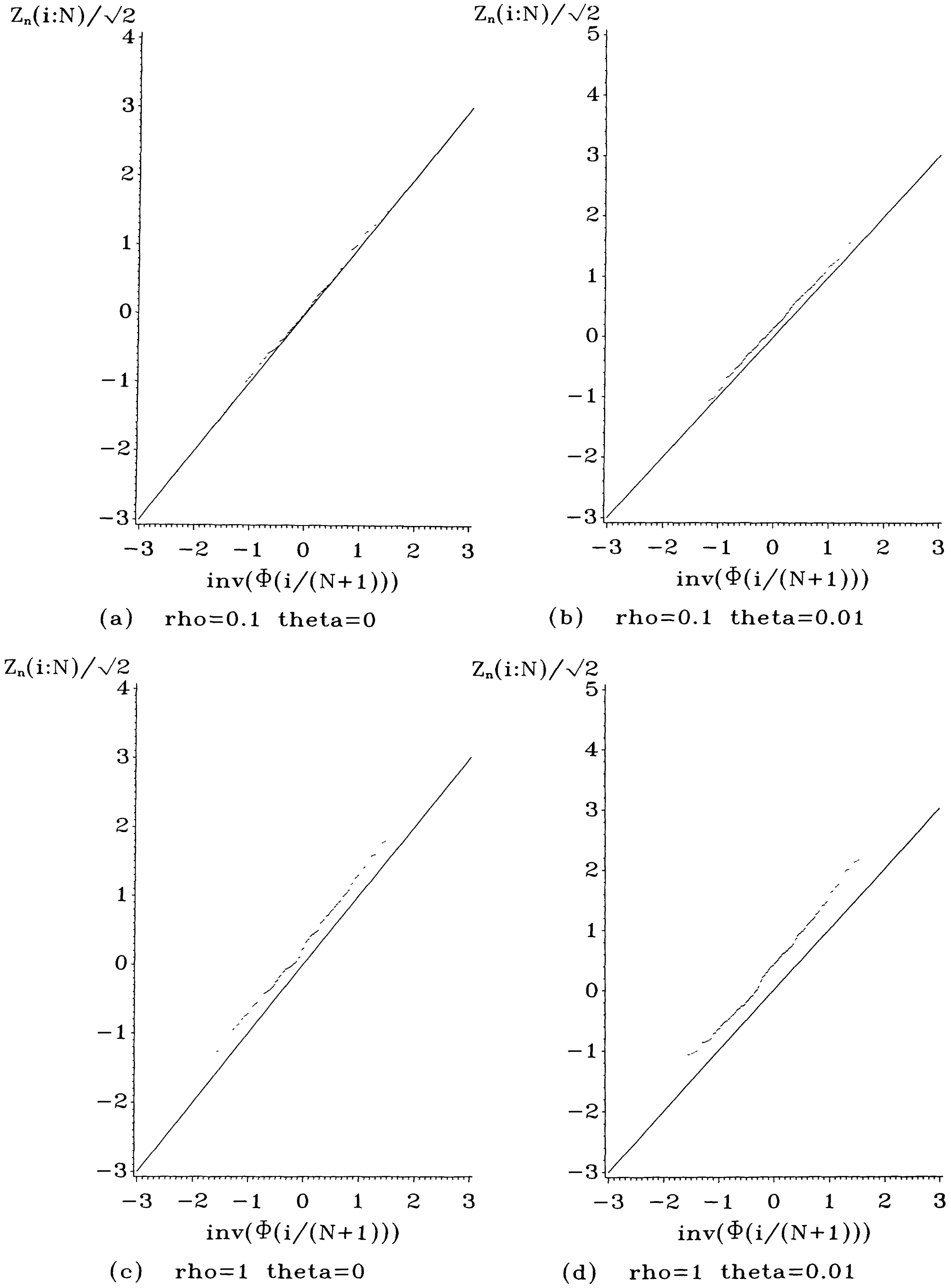
$Z_n(i:N)/\sqrt{2}$
inv(Φ(i/(N+1)))
(a) rho=0.1 theta=0
(b) rho=0.1 theta=0.01
(c) rho=1 theta=0
(d) rho=1 theta=0.01

FIG. 1.

with

$$g_n(0) = (2\pi)^{-1/2}\sigma_n^{-1} = g(0) + o(\delta_n) = (2\pi)^{-1/2}\sigma^{-1} + o(\delta_n)$$
$$\Leftrightarrow \quad \sigma_n = \sigma + o\big((na_n)^{-1/2}\big)$$

and

$$q_n(\cdot|x) = q_{\vartheta\delta_n}(\cdot)$$

is in this case

$$\begin{aligned}\varphi^*_{n,\,\mathrm{opt}}(N_n) &= 1_{(u_\alpha|\vartheta|2^{1/2},\,\infty)}(\vartheta Z_n)\\ &= 1_{(u_\alpha,\,\infty)}\big(2^{-1/2}\vartheta/|\vartheta|Z_n\big)\\ &= 1_{(u_\alpha,\,\infty)}\big(2^{-1/2}Z_n\big).\end{aligned}$$

We generated normal probability plots for N independent replicates $2^{-1/2}Z_n(1),\dots,2^{-1/2}Z_n(N)$ of $2^{-1/2}Z_n$ with different values of N and n, using the SAS 6.06 functions NORMAL for the generation of standard normal data, PROBIT for the inverse of Φ, and SASGRAPH for the graphical output. There is clearly a trade-off between the size of the bandwidth a_n, which ought to be small to give a good model approximation of N_n by N_n^*, and the random sample size $K(n) \sim na_n$, in which a_n should be large to make the distribution of Z_n nearly normal.

The following plots with $n = 400$, $N = 500$, $a_n = 4/(\log(\log(400))400^{1/5})$ ≈ 0.6741 and $\sigma^2 = 1$ exemplify our simulations. The null-hypothesis is $\vartheta = 0$ with $\rho = 0.1$ and $\rho = 1.0$; the alternatives are $\vartheta = 0.01$ with the same choices of ρ. (See Figure 1.)

The plots show the points $(\Phi^{-1}(i/(N+1)), 2^{-1/2}Z_n(i:N))$, $i = 1,\dots,N$ with $N = 500$ and $n = 400$, where $Z_n(1:N) \le \cdots \le Z_n(N:N)$ denote the order statistics pertaining to the N independent replicates $Z_n(1),\dots,Z_n(N)$ of Z_n. Deviations from the straight line, being the identity, visualize deviations of the distribution of $2^{-1/2}Z_n$ from the hypothetical standard normal one.

The first two plots show that specific behavior of Z_n which ought to be expected: an underlying alternative parameter $\vartheta = 0.01$ shifts the distribution of Z_n to the right; in both cases the structural parameter ρ is 0.1. If we however increase ρ to 1, then the distribution of Z_n is drastically shifted to the right, not only under the alternative $\vartheta = 0.01$ but also under the null-hypothesis $\vartheta = 0$ [plot (d) and (c)]. In both cases, $\varphi^*_{n,\,\mathrm{opt}}(N_n)$ would tend to reject the null-hypothesis.

Our simulations showed the general tendency that the distribution of Z_n is fairly robust against various choices of σ, but it is quite sensitive to the choice of the structural parameter ρ. This observation exemplifies the crucial role of the joint density $f(z,y) = g(z)q(y|z)$ of (X,Y) for z near x as an (usually infinite dimensional) nuisance parameter for small sample sizes n.

Acknowledgments. We would like to thank two anonymous referees for their constructive criticism from which the paper has benefited a lot. We are also grateful to Rainer Becker for his programming assistance.

REFERENCES

BHATTACHARYA, P. K. and GANGOPADHYAY, A. K. (1990). Kernel and nearest neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415.

CHAUDHURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777.

EUBANK, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Dekker, New York.

FALK, M. and REISS, R.-D. (1992a). Poisson approximation of empirical processes. *Statist. Probab. Lett.* **14** 39–48.

FALK, M. and REISS, R.-D. (1992b). Statistical inference for conditional curves: Poisson process approach. *Ann. Statist.* **20** 779–796.

HÄRDLE, W., JANSSEN, P. and SERFLING, R. J. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16** 1428–1449.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation. Application of Mathematics* **16**. Springer, New York.

IBRAGIMOV, I. A. and KHAS'MINSKII, R. Z. (1991). Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.* **19** 1681–1724.

JONES, M. C. and HALL, P. (1990). Mean squared error properties of kernel estimates of regression quantiles. *Statist. Probab. Lett.* **10** 283–289.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

MANTEIGA, W. G. (1990). Asymptotic normality of generalized functional estimators dependent on covariables. *J. Statist. Plann. Inference* **24** 377–390.

MARRON, J. S. (1989). Automatic smoothing parameter selection: a survey. In *Semiparametric and Nonparametric Economics* (A. Ullah, ed.) 65–86. Physica, Heidelberg.

MILLAR, P. W. (1982). Optimal estimation of a general regression function. *Ann. Statist.* **10** 717–740.

NUSSBAUM, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984–997.

PFANZAGL, J. (1990). *Estimation in Semiparametric Models. Lecture Notes in Statist.* **63**. Springer, New York.

REISS, R.-D. (1989). *Approximate Distributions of Order Statistics (With Applications to Nonparametric Statistics)*. Springer, New York.

REISS, R.-D. (1993). *A Course on Point Processes*. Springer, New York.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

STRASSER, H. (1985). *Mathematical Theory of Statistics. Studies in Mathematics* **7**. de Gruyter, Berlin.

TRUONG, Y. K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.* **17** 606–613.

MATHEMATISCH–GEOGRAPHISCHE FAKULTÄT
KATHOLISCHE UNIVERSITÄT EICHSTÄTT
OSTENSTRASSE 26-28
D-8078 EICHSTÄTT
GERMANY