

ON M -ESTIMATORS AND NORMAL QUANTILES

BY ANDRZEJ S. KOZEK

Macquarie University

This paper explores a class of robust estimators of normal quantiles filling the gap between maximum likelihood estimators and empirical quantiles. Our estimators are linear combinations of M -estimators. Their asymptotic variances can be arbitrarily close to variances of the maximum likelihood estimators. Compared with empirical quantiles, the new estimators offer considerable reduction of variance at near normal probability distributions.

1. Introduction and summary. In this paper we explore a class of robust and highly efficient estimators of normal quantiles. For $p \in (0, 1)$ the normal p -quantiles are of the form

$$(1) \quad q_p = \mu + \sigma \cdot q_{0,p},$$

where μ and σ are the mean and standard deviation of $N(\mu, \sigma^2)$, respectively, and $q_{0,p}$ denotes the p -quantile of the standard normal distribution. Relation (1) is not valid beyond the normal family of probability distributions. It is not valid even for approximately normal but nonnormal probability distributions. Consequently, robust estimation of μ and σ in (1) cannot produce robust estimators of quantiles even when the sample probability distribution is near normal. Robust estimators of normal quantiles should be able to capture changes in shape of probability distributions near the normal family. This is a very important property because, by the central limit theorem (CLT), probability distributions which are close to normal distributions, but which are not exactly normal, are frequently observed in practice.

Clearly, one can use empirical quantiles as estimators of q_p ; however, their asymptotic variance is much higher than the variance of the maximum likelihood estimator and higher than the variance of the robust estimators considered in this paper. Kernel-smoothed empirical quantiles [Nadaraya (1964), Parzen (1979) and Reiss (1989)] are also known to be superior to the standard empirical quantiles [cf. Azzalini (1981), Falk (1984), Mack (1987) and Falk and Reiss (1989)]. We achieve a convolution-type smoothing by considering properly chosen nonsymmetric M -functionals. In the M -functional approach, there is a scaling parameter σ which corresponds to a smoothing parameter h used in the kernel-type estimators. In the literature on robust estimation, the scaling parameter σ does not

Received July 2001; revised August 2002.

AMS 2000 subject classifications. Primary 62F35; secondary 62F12.

Key words and phrases. a -quantiles, asymptotics, M -estimator, M -functional, normal quantiles, robust.

vary with the sample size, while in the kernel smoothing, the window width $h = h_n$ converges to zero when the sample size increases to infinity. The M -functionals are more convenient to work with and our robust estimators of quantiles are linear combinations of M -estimators, with either constant or estimated coefficients.

In Section 2 we consider statistical functionals, which are either M -functionals or linear combinations of M -functionals. We show how these functionals are related to quantiles of smoothed probability distributions. To recover quantiles of the sample distribution we consider simple linear combinations of M -functionals (10) and (14) which coincide with normal quantiles for $F \sim N(\mu, \sigma^2)$. Following the standard approach, going back to von Mises (1947), we obtain estimators from functionals by evaluating them at the empirical distribution function. We deduce robustness of our estimators from the standard robustness theory, presented, for example, in Huber (1981).

In Section 3 we present the main results of the paper, given in Theorems 1 and 2. These theorems state that, under the normality assumption, the asymptotic variances (3) of the maximum likelihood quantile estimators provide a lower envelope for the asymptotic variances of our estimators of normal quantiles. A similar result holds true also in the case of a known variance [cf. Kozek (2001)].

In Section 4 we report results of simulations. We show that the Huber M -function can be safely used in our estimators and thereby replace the more sophisticated probit M -function. We also show that our estimators capture departures of quantiles from the normal probability distribution for near normal distributions.

In the Appendix we give short proofs of the results stated in the paper. More details are available in Kozek (2001).

The literature on quantiles, quantile-like functionals, M -functionals and on the corresponding estimators, is enormous. We cite here, chronologically, a few additional important references on topics relevant to the present paper: Pfanzagl (1976), Bassett and Koenker (1978), Kemperman (1987), Milasevic and Ducharme (1987), Newey and Powell (1987), Breckling and Chambers (1988), Antoch and Janssen (1989), Efron (1991), Kaigh and Cheng (1991), Chaudhuri (1992), Ralescu (1992), Jones (1994), Chaudhuri (1996), Cheng and Parzen (1997) and Koltchinskii (1997).

Throughout the paper we use the following notation. $N(\mu, \sigma^2)$ denotes the normal probability distribution with mean μ and standard deviation σ . The symbols Φ and ϕ denote the cumulative distribution function (c.d.f.) and the probability density function (p.d.f.) of the standard normal probability distribution, respectively. $Y \sim F$ means that a random variable (r.v.) Y has c.d.f. F . \mathcal{F} denotes the set of all c.d.f.'s on \mathbb{R} ; $q_p(F)$, or, equivalently, $q_p(Y)$, denotes the p -quantile of F given by

$$q_p(F) = q_p(Y) = \inf\{x : F(x) \geq p\},$$

where $Y \sim F$. $\text{Var}(G)$, where G is a c.d.f., denotes the variance of a r.v. $Z \sim G$. F_σ denotes a c.d.f. $F(\frac{\cdot}{\sigma})$, coinciding with the c.d.f. of $\sigma \cdot Y$, where $Y \sim F$ and $\sigma > 0$.

For any μ and $\sigma > 0$ we have

$$q_p(\sigma Y + \mu) = q_p(F_\sigma(\cdot - \mu)) = \sigma q_p(Y) + \mu$$

and in particular,

$$q_p(F_\sigma) = \sigma q_p(F).$$

Let us recall that if $Y \sim N(\mu, \sigma^2)$, with both μ and σ unknown, then the maximum likelihood estimator of the normal quantile q_p is given by

$$(2) \quad \hat{q}_p = \bar{Y}_n + \hat{\sigma} \cdot q_{0,p},$$

where $\hat{\sigma} = \sqrt{\sum(Y_i - \bar{Y})^2/n}$. The asymptotic variance of \hat{q}_p is given by

$$(3) \quad \lim_{n \rightarrow \infty} n \cdot \text{Var}(\hat{q}_p) = \sigma^2(1 + \frac{1}{2}q_{0,p}^2).$$

2. M-functionals. Let $Y \sim F$ be a sample random variable representing the data and let G be an arbitrary c.d.f. to be chosen by the analyst. Let Z denote a random variable independent of Y and such that $Z \sim G$. If $Y \sim N(\mu, \sigma^2)$, and $Z \sim N(0, s^2)$ then $V = Y - Z \sim N(\mu, \sigma^2 + s^2)$ and we have the following relation between quantiles of Y and quantiles of V :

$$(4) \quad q_p(Y) = q_{1/2}(V) + \sqrt{\frac{\sigma^2}{\sigma^2 + s^2}}(q_p(V) - q_{1/2}(V)).$$

One of the advantages of using quantiles of V instead of quantiles of F is that the distribution of the random variable V can be considered as a “smoothed” distribution of Y . Quantiles of kernel-smoothed empirical distributions are known to have better asymptotic performance than the sample quantiles [cf. Azzalini (1981), Falk (1984), Mack (1987) and Falk and Reiss (1989)]. Moreover, smoothed empirical processes converge to a Brownian bridge [cf. Yukich (1992) and van der Vaart (1994)], as do the standard empirical processes. Therefore, even if we were to choose a nonnormal distribution for Z , we can still obtain robust and efficient estimators.

The convolution-type smoothing effect and the distribution of V can also be obtained using M -functionals and M -estimators. Let M be a convex function on \mathbb{R} with a bounded right-hand side derivative $\psi(x)$ such that

$$(5) \quad -\infty < -\alpha = \lim_{x \rightarrow -\infty} \psi(x) < 0 < \lim_{x \rightarrow \infty} \psi(x) = \beta < \infty.$$

Let

$$(6) \quad \mathbb{M}(\theta, F) = E_F [M(Y - \theta) - M(Y)] = \int [M(y - \theta) - M(y)]F(dy),$$

where $\theta \in \mathbb{R}$. Any function $\mathcal{Q}(F)$ from the set of all distribution functions \mathcal{F} into \mathbb{R} minimizing $\mathbb{M}(\theta, F)$ is called an M -functional. Let G be a c.d.f. given by

$$(7) \quad G(z) = \frac{1}{\alpha + \beta}(\psi(z) + \alpha).$$

LEMMA 1 [Kozek and Pawlak (2002), Lemma A.1]. *The M-functional $\mathcal{Q}(F)$ minimizing $\mathbb{M}(\theta, F)$ coincides with a p -quantile of $V = Y - Z$, where Z is a random variable independent of Y with c.d.f. G given by (7) and*

$$(8) \quad p = \frac{\beta}{\alpha + \beta}.$$

Without loss of generality we can set values $\alpha = 2(1 - p)$ and $\beta = 2p$ in (7) and (8). Then we get the following:

LEMMA 2. *Let $Y \sim F$, $Z \sim G$ and let Z and Y be independent. Given $p \in (0, 1)$, the M-function meeting conditions (5), (7) and (8) for which the corresponding M-functional $\mathcal{Q}(F)$ minimizing $\mathbb{M}(\theta, F)$ coincides with a p -quantile of $V = Y - Z$, is given by*

$$(9) \quad \begin{aligned} M_{p,G}(y) &= M_G(y) + (2p - 1)y, \\ M_G(y) &= \int_0^y (2G(z) - 1) dz. \end{aligned}$$

As we shall consider simultaneously several M-functionals based on M-functions given by (9) it will be convenient to use notation $\mathcal{Q}_{p,G}(F)$ indicating the actual values of p and G , instead of the general symbol $\mathcal{Q}(F)$.

EXAMPLE 1. Let $G(z) = \Phi_s(z) = \Phi(\frac{z}{s})$, where $\Phi(z)$ is the c.d.f. of the standard normal distribution $N(0, 1)$. Then $Z \sim N(0, s^2)$. The corresponding M-function is called a *probit M-function* with parameter s and is given by

$$M_{\Phi_s}(y) = \int_0^y \left(2\Phi\left(\frac{z}{s}\right) - 1 \right) dz = y \left(2\Phi\left(\frac{y}{s}\right) - 1 \right) + 2s \left(\phi\left(\frac{y}{s}\right) - \phi(0) \right).$$

For every $p \in (0, 1)$ the M-functional $\mathcal{Q}_{p,\Phi_s}(F)$ coincides with the corresponding p -quantile of $U = Y - Z$. In particular, if $Y \sim N(\mu, \sigma^2)$ then $V = Y - Z \sim N(\mu, \sigma^2 + s^2)$, $\mathcal{Q}_{p,\Phi_s}(F)$ equals $q_p(V) = \mu + \sqrt{\sigma^2 + s^2}q_{0,p}$ and we have $|q_p - \mu| > \sigma \cdot |q_{0,p}|$. It is clear that the difference between $q_p(V)$ and $q_p(F) = \mu + \sigma \cdot q_{0,p}$, the p -quantile of $N(\mu, \sigma^2)$, may be arbitrarily large for large s .

EXAMPLE 2. Let $G(z) = U_s(z) = U(\frac{z}{s})$, where $U(z)$ is the c.d.f. of the uniform probability distribution on $[-1, 1]$. Then Z is uniformly distributed on the interval $[-s, s]$. The corresponding M-function is called a *Huber M-function* with parameter s and is given by

$$M_{U_s}(y) = \int_0^y \left(2U\left(\frac{z}{s}\right) - 1 \right) dz = \begin{cases} \frac{1}{2s}y^2, & \text{if } y \in [-s, s], \\ |y| - \frac{s}{2}, & \text{otherwise.} \end{cases}$$

The M -functional $\mathcal{Q}_{p,U_s}(F)$ coincides with the p -quantile of $U = Y - Z$. If $Y \sim N(\mu, \sigma^2)$ then no closed formula for $q_p(U)$ is available; however, Anderson's lemma [Anderson (1955)] implies that $|q_p - \mu| > \sigma \cdot |q_{0,p}|$.

EXAMPLE 3. Let $G(z)$ be the c.d.f. of a degenerate probability distribution concentrated with probability 1 at 0 and given by

$$G(z) = A(z) = \begin{cases} 0, & \text{if } z \leq 0, \\ 1, & \text{otherwise.} \end{cases}$$

Then $Z = 0$ with probability 1 and the corresponding M -function is the absolute value function given by

$$M_A(y) = |y|.$$

For $M_{p,A}$ given by (9) and the absolute value function, the corresponding M -functional $\mathcal{Q}_{p,A}(F)$ equals $q_p(F)$, the p -quantile of Y . If $Y \sim N(\mu, \sigma^2)$ then $q_p(F) = \mu + \sigma \cdot q_{0,p}$.

Let us note that condition (5) implies weak continuity of M -functionals, whenever they are unique. Using similar arguments as in Huber [(1981), page 48], we obtain the following.

PROPOSITION 1. *Let $M(y) = M_{p,G}(y)$ be given by (9). If F or G is strictly increasing on \mathbb{R} , then the M -functional $\mathcal{Q}_{p,G}(F)$ minimizing $\mathbb{M}(\theta, F)$ is continuous at F in the topology of weak convergence. In particular, the probit functional $\mathcal{Q}_{p,\Phi_s}(F)$ is continuous at every F and the Huber functional $\mathcal{Q}_{p,U_s}(F)$ is continuous at every $F(\cdot) = \Phi(\frac{\cdot - \mu}{\sigma})$ for $\mu \in \mathbb{R}$ and $\sigma > 0$.*

By evaluating M -functionals $\mathcal{Q}_{p,G}(F)$ at the empirical c.d.f. \hat{F}_n , we obtain natural estimators $\mathcal{Q}_{p,G}(\hat{F}_n)$, called M -estimators or empirical M -functionals.

Since the Glivenko–Cantelli theorem implies that the empirical c.d.f. \hat{F}_n converges with probability 1 and uniformly to the sample c.d.f. F , it does so, also, in the weak topology. Hence, we have the corollary.

COROLLARY 1. *The empirical probit M -functional $\mathcal{Q}_{p,\Phi_s}(\hat{F}_n)$ is strongly consistent for $\mathcal{Q}_{p,\Phi_s}(F)$ at every $F \in \mathcal{F}$, that is,*

$$\lim_{n \rightarrow \infty} \mathcal{Q}_{p,\Phi_s}(\hat{F}_n) = \mathcal{Q}_{p,\Phi_s}(F)$$

holds with probability 1 for every $F \in \mathcal{F}$.

COROLLARY 2. *The empirical Huber M -functional $\mathcal{Q}_{p,U_s}(\hat{F}_n)$ is strongly consistent for $\mathcal{Q}_{p,U_s}(F)$ at every $F(\cdot) = \Phi(\frac{\cdot - \mu}{\sigma})$, $\mu \in \mathbb{R}$ and $\sigma > 0$.*

By the Hampel theorem [Huber (1981), page 41], the continuity of $\mathcal{Q}(F)$ is equivalent to the robustness of the corresponding M -estimators. Hence we have the corollary.

COROLLARY 3. *For any c.d.f. G the empirical M -functional $\mathcal{Q}_{p,G}(\hat{F}_n)$ is robust at the normal family.*

It is clear that any linear combination of M -estimators is also robust at the normal family.

Following the pattern given by (4) we define $\mathcal{Q}_{p,G}^a(F)$, an approximate quantile functional or *a-quantile functional*, for short, given by

$$(10) \quad \mathcal{Q}_{p,G}^a(F) = \mathcal{Q}_{1/2,G}(F) + S_0(F) \cdot (\mathcal{Q}_{p,G}(F) - \mathcal{Q}_{1/2,G}(F)).$$

The shrinking factor $S_0(F)$ is given by

$$(11) \quad S_0(F) = \sqrt{\frac{S(F)^2}{S(F)^2 + \text{Var}(G)}},$$

where $S(F)$ is a scale functional which is continuous with respect to weak convergence of the c.d.f.'s on \mathbb{R} , satisfies $S(\Phi_s) = s$ and is used to replace $\sigma = \sqrt{\text{Var}(F)}$.

PROPOSITION 2. *Let $S(F)$ be a continuous scale functional such that $S(\Phi_s) = s$ for $s > 0$ and let the M -function used in (10) be a probit M -function. If $F(\cdot) = \Phi(\frac{\cdot - \mu}{\sigma})$ then the a -quantile functional $\mathcal{Q}_{p,\Phi_s}^a(F)$, given by (10), coincides with a p -quantile of F . Moreover, the probit M -functions are the only M -functions having this property.*

By using the Huber M -function, corresponding to a uniform distribution of Z , one can achieve for $p \in [0.05, 0.95]$ good approximations of normal quantiles by use of $\mathcal{Q}_{p,U_s}^a(F)$. This is a nice and rather unexpected feature of Huber M -functions, because the uniform distribution differs significantly from the normal one.

Table 1 contains maximal discrepancies between the standard normal quantiles and their approximations $\mathcal{Q}_{p,U_s}^a(F)$, given by (10), for $p \in [p_{\min}, 1 - p_{\min}]$ and for several values of s . It shows that, for probability distributions F close to the standard normal distribution, one can achieve, for $s \leq 1.5$, very good approximations of the normal quantile function $q_{0,p}$ by using a Huber M -function in (6). The limitation to $s \leq 1.5$ shows that the ratio of variances of F and G should not exceed 1.5. Such a balance is of no importance in the case of the probit M -functions because on the family of normal distributions the functional $\mathcal{Q}_{p,\Phi_s}^a(F)$ coincides with normal quantiles for any $s > 0$.

TABLE 1
Discrepancies between $q_{0,p}$, standard normal quantiles, and $\mathcal{Q}_{p,U_s}^a(\Phi)$, their a -quantiles approximation. $\mathcal{Q}_{p,U_s}^a(\Phi)$ is given by (10) with $G = U_s$ and corresponds to the Huber M -function: $[\max\{|q_{0,p} - \mathcal{Q}_{p,U_s}^a(\Phi)| : p \in [p_{\min}, 1 - p_{\min}]\}]$

	$s = 1$	$s = 1.5$	$s = 2.0$	$s = 2.5$	$s = 3.0$
$p_{\min} = 0.01$	0.021	0.066	0.127	0.188	0.243
$p_{\min} = 0.05$	0.000069	0.004	0.013	0.025	0.037

The proper balance between variances of F and G can be automatically retained by using scaling, a procedure recommended by both the theory and practice of M -estimators [cf. Huber (1981) or Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. By replacing $\mathbb{M}(\theta, F)$ with

$$(12) \quad \mathbb{M}_{sc}(\theta, \sigma, F) = \int \left[M\left(\frac{y - \theta}{\sigma}\right) - M\left(\frac{y}{\sigma}\right) \right] F(dy),$$

we obtain scaled M -functionals. Given $\sigma > 0$, any function $\mathcal{Q}_{sc}(\sigma, F)$ from \mathcal{F} into \mathbb{R} minimizing $\mathbb{M}_{sc}(\theta, \sigma, F)$ will be called a *scaled M -functional*. Let us note that if, for a given c.d.f. G and $p \in (0, 1)$, the M -function is defined by (9) then

$$\mathcal{Q}_{sc}(\sigma, F) = \mathcal{Q}_{p,G_\sigma}(F).$$

We have the following interpretation of scaled M -functionals.

LEMMA 3. *Let the M -function be given by (9) and random variables Y and Z be independent and such that $Y \sim F$ and $Z \sim G$. Then the scaled functional $\mathcal{Q}_{p,G_\sigma}(F)$ coincides with a p -quantile of $Y - \sigma \cdot Z$, that is,*

$$\mathcal{Q}_{p,G_\sigma}(F) = q_p(Y - \sigma Z).$$

Now, consider the case where $Y \sim F$ is $N(\mu, \sigma^2)$, $Z \sim N(0, s^2)$ and Y and Z are independent. Then $W = Y - \sigma Z \sim N(\mu, \sigma^2(1 + s^2))$ and we have the following relation between quantiles of Y and quantiles of W :

$$q_p(F) = q_{1/2}(W) + \sqrt{\frac{1}{1 + s^2}}(q_p(W) - q_{1/2}(W)),$$

which can be rewritten using the scaled M -functionals notation

$$(13) \quad q_p(F) = \mathcal{Q}_{1/2,\Phi_{\sigma s}}(F) + \sqrt{\frac{1}{1 + s^2}}(\mathcal{Q}_{p,\Phi_{\sigma s}}(F) - \mathcal{Q}_{1/2,\Phi_{\sigma s}}(F)).$$

Clearly, by replacing Φ_s with another c.d.f. G we lose the equality. Table 1 shows, however, that in the case of the Huber M -functions M_{p,U_s} and when the ratio between variances of Y and Z does not exceed 1.5, one obtains very good

approximations of $q_p(F)$. In the case of a -quantiles (10), these variances are σ^2 and s^2 , respectively, and it is not possible to control their ratio, as σ is not known and has to be estimated. The present case is more convenient because

$$\text{Var}(Y) = \sigma^2 \quad \text{and} \quad \text{Var}(\sigma \cdot Z) = \sigma^2 \cdot s^2$$

so that the ratio of these variances equals $1/s^2$. Hence, the quality of the approximation depends only on the choice of s and does not depend on the unknown parameter σ .

We define a *scaled a -quantile* functional $\mathcal{Q}_{p,G_\sigma}^{\text{sc}}(F)$ by replacing Φ_s in (13) with a general distribution function G ,

$$\begin{aligned} \mathcal{Q}_{p,G_\sigma}^{\text{sc}}(F) &= \mathcal{Q}_{1/2,G_\sigma}(F) + S_1 \cdot (\mathcal{Q}_{p,G_\sigma}(F) - \mathcal{Q}_{1/2,G_\sigma}(F)), \\ (14) \quad S_1 &= \sqrt{1/(1 + \text{Var}(G))}, \end{aligned}$$

where S_1 is a constant shrinking factor, depending only on the choice of G . Let us note that in (14) we use simultaneously $\mathcal{Q}_{1/2,G_\sigma}(F)$, $\mathcal{Q}_{p,G_\sigma}(F)$ and σ . It requires minimization of the following scaled convex functionals:

$$\begin{aligned} \mathbb{M}_{\text{sc}}(\theta_1, \sigma, F) &= \int \left[M_{1/2,G} \left(\frac{y - \theta_1}{\sigma} \right) - M_{1/2,G} \left(\frac{y}{\sigma} \right) \right] F(dy), \\ (15) \quad \mathbb{M}_{\text{sc}}(\theta_2, \sigma, F) &= \int \left[M_{p,G} \left(\frac{y - \theta_2}{\sigma} \right) - M_{p,G} \left(\frac{y}{\sigma} \right) \right] F(dy), \end{aligned}$$

under the condition

$$(16) \quad E_F \chi \left(\frac{Y - \theta_1}{\sigma} \right) = 0.$$

We assume that the function $\chi(y)$ is symmetric, differentiable, bounded and bowl-shaped.

An argument similar to that of Proposition 2 implies the following property of the scaled functionals $\mathcal{Q}_{p,G_\sigma}^{\text{sc}}(F)$.

PROPOSITION 3. *Let $F(y) = \Phi(\frac{y-\mu}{\sigma})$ for some μ and for $\sigma > 0$. Assume that the scale functional $S_{\text{sc}}(\cdot)$ solving (16) coincides at F with the standard deviation σ . Then, for the probit M -function the scaled a -quantiles given by (14) and the quantiles of F coincide.*

3. Asymptotic properties of estimators. Conditions for consistency and asymptotic normality of M -functionals are well known [cf. Huber (1967, 1981)] and we shall not recall them here explicitly. We shall consider estimators of quantiles which are empirical a -quantiles and empirical scaled a -quantiles, given by (10) and (14), respectively. They are linear combinations of M -estimators. Under the normality assumption we explore in more detail the surprisingly good behavior of their asymptotic variances.

Using the asymptotic normality of M -estimators we first get the asymptotic distributions of our estimators in the case of a general sample c.d.f. F and a general c.d.f. G . Next, in the case of a normal sample c.d.f. F and probit M -function M_{p,Φ_s} , we use the resulting limit distributions to show that the asymptotic maximum likelihood variances (3) form lower envelopes, as $s \rightarrow \infty$, for variances of the estimators under consideration. This is the main result of the present paper, stated in Theorems 1 and 2. Let us note that when $s \rightarrow 0$ then our estimators converge to empirical quantiles. Hence, our class of robust estimators of normal quantiles fills the gap between empirical quantiles and maximum likelihood estimators.

We shall use continuous bowl-shaped functions $\chi(y)$, which are differentiable at all but a finite number of points, such that

- (17a) $y\chi'(y) \geq 0$ for all y where $\chi'(y)$ exists,
- (17b) $y\chi'(y) > 0$ on a nondegenerate interval, (c_1, c_2) ,
 $c_1, c_2 \in [-\infty, \infty]$,
- (17c) $\lim_{|y| \rightarrow \infty} \chi(y) = c_3 > 0$,
- (17d) $\chi(0) = c_4 < 0$.

In the case of the a -quantile functional (10) we need to minimize

$$(18) \quad \mathbb{M}_{p_i}(\theta_i, F) = \int [M_{p_i,G}(y - \theta_i) - M_{p_i,G}(y)] F(dy)$$

for $i = 1, 2$, with $p_1 = \frac{1}{2}$ and $p_2 = p$, respectively, under the condition

$$(19) \quad E\chi\left(\frac{Y - \theta_1}{\sigma}\right) = 0,$$

where $\chi(y)$ is a function meeting conditions (17a)–(17d). Let $\{\mathcal{Q}_{1/2,G}(F), \mathcal{Q}_{p,G}(F), S(F)\}$ be a solution to this problem. The estimator of the a -quantile functional (10) is given by

$$(20) \quad \hat{\mathcal{Q}}_{p,G}^a = \mathcal{Q}_{p,G}^a(\hat{F}_n) = \mathcal{Q}_{1/2,G}(\hat{F}_n) + \hat{S}_0 \cdot (\mathcal{Q}_{p,G}(\hat{F}_n) - \mathcal{Q}_{1/2,G}(\hat{F}_n)),$$

where $\mathcal{Q}_{1/2,G}(\hat{F}_n)$ and $\mathcal{Q}_{p,G}(\hat{F}_n)$ minimize $\mathbb{M}_{p_i}(\theta_i, F)$ for $i = 1, 2$, respectively, condition (19) is met with $\hat{S} = S(\hat{F}_n)$ and

$$(21) \quad \hat{S}_0 = \sqrt{\frac{\hat{S}^2}{\hat{S}^2 + \text{Var}(G)}}.$$

Let s_p^2 be given by

$$(22) \quad s_p^2 = \mathbf{b}S_p\mathbf{b}',$$

where

$$(23) \quad \mathbf{b} = [1 - S_0(F), S_0(F), S_0(F)' \cdot (\mathcal{Q}_{p,G}(F) - \mathcal{Q}_{1/2,G}(F))],$$

$S_0(F)'$ is the derivative of $S_0(x)$ given by (11) at $x = S(F)$, and S_p is the covariance matrix of $\{\mathcal{Q}_{1/2,G}(\hat{F}_n), \mathcal{Q}_{p,G}(\hat{F}_n), \hat{S}\}$ given by the sandwich theorem [Huber (1981), Chapter 6, Corollary 3.2].

PROPOSITION 4. *Let $M_{p,G}$ be an M -function given by (9) with $\text{Var}(G) < \infty$ and let $\chi(y)$ meet conditions (17a)–(17d). Let Y_1, \dots, Y_n be i.i.d. r.v.'s with a c.d.f. F . If s_p^2 given by (22) is finite and positive then $\hat{\mathcal{Q}}_{p,G}^a$ is asymptotically normal, that is,*

$$\lim_{n \rightarrow \infty} \Pr \left(\sqrt{n} \frac{\hat{\mathcal{Q}}_{p,G}^a - \mathcal{Q}_{p,G}^a(F)}{s_p} < t \right) = \Phi(t).$$

The derived asymptotic distribution is needed to prove the following theorem. Theorem 1 shows that, at the normal family, estimators $\hat{\mathcal{Q}}_{p,\phi_s}^a$ may have variances arbitrarily close to the variances of the maximum likelihood estimator.

THEOREM 1. *Let $F(\cdot) = \Phi(\frac{\cdot - \mu}{\sigma})$ for some $\mu \in R$ and $\sigma > 0$ and let $M_{p_i,G}$ in (18) be probit M -functions $M_{1/2,\phi_s}$ and M_{p,ϕ_s} , respectively, with parameter s . Moreover, let the function $\chi(y)$ in (19) be given by*

$$(24) \quad \chi_s(y) = \min(s^2, y^2) - \beta(s),$$

where $\beta(s) = \int \min(s^2, y^2) \Phi(dy)$. Then, for every $p \in (0, 1)$, we have

$$\lim_{s \rightarrow \infty} s_p^2(s) = \sigma^2(1 + \frac{1}{2}q_{0,p}^2),$$

where $s_p^2(s)$ is given by (22).

Let us recall that the χ function (24) was introduced in Huber (1964). As we have already mentioned, Table 1 in Section 2 implies that, in the case of the Huber M -functions, a -quantiles (10) provide good approximations to normal quantiles only when the ratio of variances of F and G is kept smaller than 1.5. This ratio is beyond our control when σ^2 is unknown. Therefore, the estimator (20), considered in Proposition 4, is recommended in estimation of normal quantiles only in the case of probit M -functions given by (1).

By Lemma 3, the proper balance, independent from the unknown scale, can be automatically retained by using M -functions with scaling parameters. Hence, in the case of unknown scale factor σ , one should minimize the empirical version of functionals (15) under condition (16) and consider the scaled a -quantile functional (14). Then both functionals (14), based on the probit M -functions with

any $s > 0$ or on Huber M -functions with $s \leq 1.5$, can be safely used. The resulting estimators, empirical scaled a -quantiles, are given by

$$(25) \quad \begin{aligned} \hat{Q}_{p,G_\sigma}^{\text{sc}} &= \mathcal{Q}_{p,G_{\hat{\sigma}_n}}^{\text{sc}}(\hat{F}_n) \\ &= \mathcal{Q}_{1/2,G_{\hat{\sigma}_n}}(\hat{F}_n) + S_1 \cdot (\mathcal{Q}_{p,G_{\hat{\sigma}_n}}(\hat{F}_n) - \mathcal{Q}_{1/2,G_{\hat{\sigma}_n}}(\hat{F}_n)). \end{aligned}$$

Let

$$\sigma_{p,\text{sc}}^2 = \mathbf{c} \Sigma_{p,\text{sc}} \mathbf{c}',$$

where

$$\mathbf{c} = [1 - S_1, S_1, 0],$$

S_1 is given by (14) and $\Sigma_{p,\text{sc}}$ is the covariance matrix of $\{\mathcal{Q}_{1/2,G_{\hat{\sigma}_n}}(\hat{F}_n), \mathcal{Q}_{p,G_{\hat{\sigma}_n}}(\hat{F}_n), \hat{\sigma}_n\}$ given by the sandwich theorem. With this notation we have the following proposition, similar to Proposition 4.

PROPOSITION 5. *Let $M_{p,G}$ be an M -function given by (9) with $\text{Var}(G) < \infty$ and let $\chi(y)$ be a continuous and bounded function for estimating the scale parameter $S(F)$. Let Y_1, \dots, Y_n be i.i.d. r.v.'s with c.d.f. F . Then, the estimator $\hat{Q}_{p,G_\sigma}^{\text{sc}}$ is asymptotically normal and*

$$\lim_{n \rightarrow \infty} \Pr \left(\sqrt{n} \frac{\hat{Q}_{p,G_\sigma}^{\text{sc}} - \mathcal{Q}_{p,G_\sigma}^{\text{sc}}(F)}{\sigma_{p,\text{sc}}} < t \right) = \Phi(t).$$

By Theorem 2 below, if $F(\cdot) = \Phi(\frac{\cdot - \mu}{\sigma})$ for some $\mu \in \mathbb{R}$, $\sigma > 0$ and if the probit M -functions with parameter s are used then, for every $p \in (0, 1)$, the asymptotic variance of the estimator $\hat{Q}_{p,G_\sigma}^{\text{sc}}$ of p -quantiles approaches the asymptotic variance of the maximum likelihood estimator when $s \rightarrow \infty$. Hence, in the case of translation and scale parameters, both unknown, the empirical scaled a -quantiles $\hat{Q}_{p,G_\sigma}^{\text{sc}}$ are robust and highly efficient.

THEOREM 2. *Let $F(\cdot) = \Phi(\frac{\cdot - \mu}{\sigma})$ for some $\mu \in \mathbb{R}$ and $\sigma > 0$. If M -functions are probit M -functions M_{p,ϕ_s} with parameter s and the χ function in (16) is given by (24) then, for every $p \in (0, 1)$, we have*

$$\lim_{s \rightarrow \infty} \sigma_{p,\text{sc}}^2(s) = \sigma^2 \left(1 + \frac{1}{2} q_{0,p}^2 \right).$$

4. Simulations. To show the behavior of empirical a -quantiles and scaled empirical a -quantiles in the case of unknown σ , we report results of 1000 simulations for $p = 0.8$ and $p = 0.95$ for sample size $N = 100$. Table 2 contains results for the standard normal and Table 3 for the standardized χ_{20}^2 probability distribution, respectively.

TABLE 2

Means and standard deviations of estimators of quantiles for $p = 0.8$ ($q_{0.8} = 0.8416$) and $p = 0.95$ ($q_{0.95} = 1.6449$) for the standard normal population distribution

p	N	M	$\hat{E}(a-q)$	$\hat{E}(sc-a-q)$	$SD(a-q)$	$SD(sc-a-q)$	$SD(ML)$
0.8	100	P	0.8451	0.8353	0.1199	0.1199	0.1163
		H	0.8423	0.8457	0.1199	0.1214	
		A	0.8424		0.1364		
0.95	100	P	1.6319	1.6332	0.1587	0.1561	0.1533
		H	1.6400	1.6458	0.1622	0.1639	
		A	1.6533		0.2087		

For each generated sample of size N reported in Table 2, we calculated five estimators of p -quantiles: two empirical a -quantiles, two empirical scaled a -quantiles and the empirical quantiles. Each type of estimator was calculated for three M -functions: probit, Huber and the absolute value function. In Table 3 we additionally calculated maximum likelihood estimators (ML). Let us note that the estimators corresponding to the absolute value function coincide with the empirical quantiles, both in the case of a -quantiles and in the case of the scaled a -quantiles. Therefore, we report only one case.

Empirical a -quantiles (20) are referred to in the tables in columns marked $a-q$ and empirical scaled a -quantiles (25) in columns marked $sc-a-q$. References to M -functions M_Φ , M_{U_1} and M_A are made in rows of the tables. Column M refers to labels of the M -functions: P for the probit M -function with parameter $s = 1$, H for the Huber M -function with parameter $s = 1$ and A for the absolute value function. In columns marked $\hat{E}(ML)$ and $SD(ML)$ we report in Table 2 the corresponding theoretical standard deviations of the maximum likelihood estimator \hat{q}_p given by (2) and in Table 3 empirical means and standard deviations of this estimator.

The obtained empirical standard deviations of empirical a -quantiles and empirical scaled a -quantiles are very close to the theoretical standard deviations of the

TABLE 3

Means and standard deviations of estimators of quantiles for $p = 0.8$ ($q_{0.8} = 0.7965$) and $p = 0.95$ ($q_{0.95} = 1.8041$) for the standardized χ^2_{20} population distribution

p	N	M	$\hat{E}(a-q)$	$\hat{E}(sc-a-q)$	$\hat{E}(ML)$	$SD(a-q)$	$SD(sc-a-q)$	$SD(ML)$
0.8	100	P	0.8093	0.8041	0.8432	0.1371	0.1437	0.1408
		H	0.8030	0.8091		0.1355	0.1414	
		A	0.7956			0.1595		
0.95	100	P	1.6790	1.6916	1.6425	0.2157	0.2007	0.1975
		H	1.7214	1.7345		0.2319	0.2269	
		A	1.8173			0.2835		

maximum likelihood estimators. Our simulations also confirm that the considered estimators have variances much lower than the variances of the empirical quantiles. Since our estimators are robust, they can be safely used to estimate quantiles of distribution functions F being approximately normal, as in the case reported in Table 3. As noted earlier, this is an important feature as, by the CLT, approximately normal distributions frequently occur in practice. Let us note that, in contrast to the maximum likelihood estimator, our α -quantile estimators show how the scaled χ_{20}^2 distribution differs from the normal one.

APPENDIX: PROOFS

PROOF OF PROPOSITION 2. The first part of the proposition is a simple consequence of Lemma 1. The second part of the proposition follows from Cramér’s characterization of the normal distribution [Feller (1966), Section 15.8, Theorem 1]. \square

PROOF OF LEMMA 3. The proof follows from Lemma 1 after changing variables $Y' = Y/\sigma$ in (12). \square

PROOF OF PROPOSITION 3. The proof is similar to the proof of Proposition 2. \square

PROOF OF PROPOSITIONS 4 AND 5. The asymptotic normality of the considered estimators follows by direct application of the sandwich and delta theorems. \square

PROOF OF THEOREM 1. By applying L’Hôpital’s rule we get

$$\begin{aligned} \lim_{s \rightarrow \infty} s \left(2\Phi \left(\frac{y - \mu - \sqrt{s^2 + \sigma^2} q_{0,p}}{s} \right) - 1 + (2p - 1) \right) \\ = 2(y - \mu)\phi(q_{0,p}), \end{aligned}$$

where $\phi = \Phi'$. Moreover, by continuity and symmetry of ϕ , we have

$$\lim_{s \rightarrow \infty} \phi \left(\frac{y - \mu - \sqrt{s^2 + \sigma^2} q_{0,p}}{s} \right) = \phi(q_{0,p}).$$

Let us note that $S_0(F)'$ in (23) is given by

$$S_0(F)' = \frac{d}{d\sigma} \sqrt{\frac{\sigma^2}{\sigma^2 + s^2}} = \frac{s^2}{(\sqrt{\sigma^2 + s^2})^3}$$

and that

$$\lim_{s \rightarrow \infty} \chi_s(y) = y^2 - 1.$$

Moreover, we have

$$\lim_{s \rightarrow \infty} \mathbf{b}(s) = (1, 0, q_{0,p}).$$

Hence, and by the Lebesgue dominated convergence theorem, we obtain

$$\left\{ \lim_{\substack{s \rightarrow \infty \\ j=1,2,3}} S_{p,ij}(s) \mathbf{b}_i(s) \mathbf{b}_j(s) \right\}_{i=1,2,3} = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{2} q_{0,p}^2 \end{pmatrix}.$$

The sum of all elements of the right-hand side array equals the limit of $s_p^2(s)$ and this completes the proof of the theorem. \square

PROOF OF THEOREM 2. The proof is similar to the proof of Theorem 1. By applying L'Hôpital's rule we get

$$\begin{aligned} & \lim_{s \rightarrow \infty} s \left(2\Phi \left(\frac{y - \mu - \sigma \sqrt{1 + s^2} q_{0,p}}{s\sigma} \right) + (2p - 2) \right) \\ &= 2 \left(\frac{y - \mu}{\sigma} \right) \phi(-q_{0,p}). \end{aligned}$$

Moreover, by the continuity of ϕ we have

$$\lim_{s \rightarrow \infty} \phi \left(\frac{y - \mu - \sigma \sqrt{1 + s^2} q_{0,p}}{s\sigma} \right) = \phi(-q_{0,p}).$$

Let us note that in the present case we have

$$\lim_{s \rightarrow \infty} s \cdot S_1(s) = \lim_{s \rightarrow \infty} s / \sqrt{1 + s^2} = 1.$$

Hence, and by the Lebesgue dominated convergence theorem, we obtain

$$\left\{ \lim_{\substack{s \rightarrow \infty \\ j=1,2,3}} \Sigma_{p,sc,ij}(s) \mathbf{c}_i(s) \mathbf{c}_j(s) \right\}_{i=1,2,3} = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \frac{\sigma^2}{2} q_{0,p}^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The sum of all elements of the array on the right-hand side equals the required limit and this concludes the proof of the theorem. \square

Acknowledgments. We are grateful to Malcolm Hudson for naming the probit M -function, to Wolfgang Wefelmeyer for valuable comments on an early version of the paper and to an Associate Editor and two referees for suggestions which led to considerable improvement of the presentation of this article.

REFERENCES

- ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176.
- ANTOCH, J. and JANSSEN, P. (1989). Nonparametric regression M -quantiles. *Statist. Probab. Lett.* **8** 355–362.
- AZZALINI, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* **68** 326–328.
- BASSETT, G. and KOENKER, R. (1978). Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc.* **73** 618–622.
- BRECKLING, J. and CHAMBERS, R. (1988). M -quantiles. *Biometrika* **75** 761–771.
- CHAUDHURI, P. (1992). Generalized regression quantiles: Forming a useful toolkit for robust linear regression. In *L_1 -Statistical Analysis and Related Methods* (Y. Dodge, ed.) 169–185. North-Holland, Amsterdam.
- CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872.
- CHENG, C. and PARZEN, E. (1997). Unified estimators of smooth quantile and quantile density functions. *J. Statist. Plann. Inference* **59** 291–307.
- EFRON, B. (1991). Regression percentiles using asymmetric squared error loss. *Statist. Sinica* **1** 93–125.
- FALK, M. (1984). Relative deficiency of kernel type estimators of quantiles. *Ann. Statist.* **12** 261–268.
- FALK, M. and REISS, R.-D. (1989). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *Ann. Probab.* **17** 362–371.
- FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.
- HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.
- HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.
- HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.
- JONES, M. C. (1994). Expectiles and M -quantiles are quantiles. *Statist. Probab. Lett.* **20** 149–153.
- KAIGH, W. D. and CHENG, C. (1991). Subsampling quantile estimators and uniformity criteria. *Comm. Statist. Theory Methods* **20** 539–560.
- KEMPERMAN, J. H. B. (1987). The median for a finite measure on a Banach space. In *Statistical Data Analysis Based on the L_1 -Norm and Related Methods* (Y. Dodge, ed.) 217–230. North-Holland, Amsterdam.
- KOLTCHINSKII, V. I. (1997). M -estimation, convexity and quantiles. *Ann. Statist.* **25** 435–477.
- KOZEK, A. S. (2001). On M -estimators and normal quantiles. Available at <http://laurel.ocs.mq.edu.au/~akozek/rqp.pdf>.
- KOZEK, A. S. and PAWLAK, M. (2002). Distribution-free consistency of kernel nonparametric M -estimators. *Statist. Probab. Lett.* **58** 343–353.
- MACK, Y. P. (1987). Bahadur's representation of sample quantiles based on smoothed estimates of a distribution function. *Probab. Math. Statist.* **8** 183–189.
- MILASEVIC, P. and DUCHARME, G. R. (1987). Uniqueness of the spatial median. *Ann. Statist.* **15** 1332–1333.
- NADARAYA, È. (1964). Some new estimates for distribution functions. *Theory Probab. Appl.* **9** 497–500.
- NEWHEY, W. K. and POWELL, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55** 819–847.

- PARZEN, E. (1979). Nonparametric statistical data modeling (with discussion). *J. Amer. Statist. Assoc.* **74** 105–131.
- PFANZAGL, J. (1976). Investigating the quantile of an unknown distribution. In *Contributions to Applied Statistics* (W. J. Ziegler, ed.) 111–126. Birkhäuser, Basel.
- RALESCU, S. S. (1992). Asymptotic deviations between perturbed empirical and quantile processes. *J. Statist. Plann. Inference* **32** 243–258.
- REISS, R.-D. (1989). *Approximate Distributions of Order Statistics*. Springer, New York.
- VAN DER VAART, A. (1994). Weak convergence of smoothed empirical processes. *Scand. J. Statist.* **21** 501–504.
- VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.
- YUKICH, J. E. (1992). Weak convergence of smoothed empirical processes. *Scand. J. Statist.* **19** 271–279.

DEPARTMENT OF STATISTICS
MACQUARIE UNIVERSITY
SYDNEY, NSW 2109
AUSTRALIA
E-MAIL: akozek@efs.mq.edu.au