TRABAJOS DE ESTADISTICA
Vol. 6, núm. 2, 1991, pp. 67 a 85

EFFICIENCY RATE AND LOCAL DEFICIENCY OF HUBER'S LOCATION ESTIMATORS AND OF THE α-ESTIMATORS

A. RUBIO
Departamento de Matemáticas
Universidad de Extremadura

J. Á. VÍŠEK
Institute of Information Theory and Automation
Czechoslovak Academy of Sciences

ABSTRACT

The paper studies the problem of selecting an estimator with (approximately) minimal asymptotic variance. For every fixed contamination level there is usually just one such estimator in the considered family. Using the first and the second derivative of the asymptotic variance with respect to the parameter which parametrizes the family of estimators the paper gives two examples how to select the estimator and gives an approximation to a loss which we suffer when we use the estimator with approximately minimal asymptotic variance instead of the estimator with the precisely minimal.

Key words: Efficiency rate, local deficiency, level of contamination, Huber's location estimators, α-estimators.
A.M.S. Classification: 62F35, 62J99.

RESUMEN

En este artículo se estudia el problema de elección de un estimador con aproximadamente mínima varianza asintótica. Para cada nivel de contaminación fijado habrá, por lo general un único estimador de éstos en la familia considerada. En este artículo se dan dos ejemplos sobre cómo elegir el estimador mediante el uso de la primera y segunda derivada de la varianza asintótica con respecto al parámetro que parametriza la familia de estimadores; también

Recibido: enero 1991.
Revisado: octubre 1991.

se da una aproximación de la pérdida sufrida cuando se utiliza el estimador con aproximadamente mínima varianza asintótica, en lugar del estimador con exactamente la mínima varianza.

Título: Indice de eficiencia y deficiencia local de los estimadores de localización de Huber y de los α-estimadores.
Palabras clave: Indice de eficiencia, deficiencia local, nivel de contaminación. Estimadores de localización de Huber, α-estimadores.
Clasificación A.M.S.: 62F35, 62J99.

1. Introduction

In 1964 P. J. Huber has opened by his pioneer paper "Robust estimation of location parameter" a new region of the mathematical statistics. He has found a minimax solution of the location problem in the mixture model of data.

Let us recall this result to be able to enlighten goals of this paper. At first we need some basic notations.

Let N denote the set of all positive integers and R the real line. $\mathscr{B}$ is assumed to be the Borel σ-álgebra of the subsets of R and $\mathscr{M}$ the set of all probability measures defined on $(R, \mathscr{B})$. Finally, denote by $\mathscr{F}$ the set of all one-dimensional distribution functions. The Huber result was derived under the following assumption:

Assumption A. *Let $G \in \mathscr{F}$ have a density $g(t)$ with a convex support. Moreover, let g be twice continuously differentiable with $-\log g(t)$ strictly convex of the support of g.*

Without loss of generality let $\sup\{t \in R : g'(t) > 0\} = 0$. Finally, for any $\varepsilon \in [0, 1]$, define the model of contamination of data

$$\mathscr{P}(\varepsilon) = \{F \in \mathscr{F} : F = (1 - \varepsilon)G + \varepsilon H; \quad H \in \mathscr{F}\} \tag{1}$$

and put $\mathscr{P} = \{\mathscr{P}(\varepsilon)\}_{\varepsilon \in [0, 1]}$. The value ε may be interpreted as a contamination level. This value may be estimated (see Víšek (1985)) but generally it is unknown but fixed, being given by the physical circumstances. The experience of applied statisticians says that it lies ussually between 1 % and 10 % —see Hampel et al 1986. Now, let random variables $X_1, ..., X_n$ be independent and identically distributed according to $F(x - \Delta)$ where $F \in \mathscr{P}(\varepsilon)$ for some fixed $\varepsilon \in [0, 1]$ and some $\Delta \in R$. Let $\hat{\Delta}_{\varepsilon, n}$

$(X_1, ..., X_n)$ be an estimator of Δ and assume that its asymptotic distribution and the variance of a random variable distributed according to it exist for all $F \in \mathscr{P}(\varepsilon)$. Then denote this variance by as. var $_F(\hat{\Delta}_{\varepsilon, n})$ and put for $\tau \in [0, 1]$

$$V(\hat{\Delta}_{\varepsilon, n}, \tau) = \sup_{F \in \mathscr{P}(\tau)} \text{as. var}_F(\hat{\Delta}_{\varepsilon, n})$$

The task is to find an estimator such that $V(\hat{\Delta}_{\varepsilon,n}, \varepsilon)$ is minimal among all unbiased estimator of Δ. Huber has shown that such estimator $\hat{\Delta}_{\varepsilon, n}$ can be found as a solution of

$$\sum_{i=1}^{n} \psi_\varepsilon(X_i - t) = 0$$

where

$$\psi_\varepsilon(t) = \begin{matrix} -k(\varepsilon) & \{t \in R : g'(t)/g(t) \geqslant k(\varepsilon)\} \\ -g'(t)/g(t) & \{t \in R : |g'(t)/g(t)| < k(\varepsilon)\} \\ k(\varepsilon) & \{t \in R : g'(t)/g(t) \leqslant -k(\varepsilon)\} \end{matrix}$$

and $k(\varepsilon)$ is related to ε and to $g(t)$ by the equation

$$(1 - \varepsilon)^{-1} = \int_{t_0(\varepsilon)}^{t_1(\varepsilon)} g(t)\, dt + \frac{g(t_0(\varepsilon)) + g(t_1(\varepsilon))}{k(\varepsilon)} \tag{2}$$

with $t_0(\varepsilon) < t_1(\varepsilon)$ being the endpoints of interval $\{t \in R : |g'(t)/g(t)| \leqslant k(\varepsilon)\}$, i.e. for finite interval we have

$$g'(t(\varepsilon))/g(t_i(\varepsilon)) = (-1)^i k(\varepsilon)$$

for $i = 0,1$ (see Huber (1964)). Notice that for $\varepsilon > 0$ either $t_0(\varepsilon) > -\infty$ or $t_1(\varepsilon) < \infty$ and e.g. for the symmetric density g both inequalities hold. In other words, $\hat{\Delta}_{\varepsilon, n}$ is the maximum likelihood estimator for the density

$$f_\varepsilon(t) = \begin{matrix} (1 - \varepsilon) g(t_0(\varepsilon)) \exp\{k(\varepsilon)(t - t_0(\varepsilon))\} & \text{for } t \leqslant t_0(\varepsilon), \\ (1 - \varepsilon) g(t) & \text{for } t_0(\varepsilon) < t < t_1(\varepsilon), \\ (1 - \varepsilon) g(t_1(\varepsilon)) \exp\{-k(\varepsilon)(t - t_1(\varepsilon))\} & \text{for } t \geqslant t_1(\varepsilon). \end{matrix}$$

So, we have at our disposal the whole family of estimators $S_{0, n} = \{\hat{\Delta}_{\varepsilon, n}\}_{\varepsilon \in [0, 1]}$.

Let us assume that the "true" (but unknown) value of the contamination level is ε_0 and that we have selected (for processing our data) from the family $S_{0,n}$ the estimator $\hat{\Delta}_{\varepsilon_1,n}$. Then our loss is given by the difference

$$V(\hat{\Delta}_{\varepsilon_1,n}, \varepsilon_0) - V(\hat{\Delta}_{\varepsilon_0,n}, \varepsilon_0)$$

(because if we know the true value of the contamination level ε_0 we would select $\hat{\Delta}_{\varepsilon_0,n}$ instead of $\hat{\Delta}_{\varepsilon_1,n}$). The paper offers an approximation to this loss. Since we shall consider only the asymptotic variances let us write $V(\hat{\Delta}_\varepsilon, \varepsilon)$ instead of $V(\hat{\Delta}_{\varepsilon,n}, \varepsilon)$.

Another source of whole families of estimators is e.g. the minimal distance estimation. The authors usually consider a whole family of distances, frequently parametrized by a real (sometimes multidimensional) parameter, say α, $\alpha \in I$, $I \subset R$. It implies that they then obtain a family of estimators, say $\mathscr{L} = \{\hat{T}_\alpha\}_{\alpha \in I}$.

Describing the contamination of data again by some model, e. g. by $\mathscr{P}(\varepsilon_0)$, there is a question which estimator $\hat{T}_\alpha$ should be selected. One possibility how to establish a relation between the parameter α of the family $\mathscr{L}$ and the parameter representing the contamination level is to evaluate the derivative $\frac{\partial}{\partial \alpha} V(\hat{T}_\alpha, \varepsilon_0)$ and to find an $\alpha(\varepsilon_0)$, for which it turns to be zero. Naturally, having learnt that the estimator $\hat{T}_\alpha(\varepsilon_0)$ is the optimal one for the contamination level ε_0 we have to solve again the question how large is the difference

$$V(\hat{T}_{\alpha_1}, \varepsilon_0) - V(\hat{T}_{\alpha(\varepsilon_0)}, \varepsilon_0)$$

for some α_1 from the neighbourhood of $\alpha(\varepsilon_0)$ because not knowing the contamination level ε_0 one selects only an approximately optimal estimator $\hat{T}_{\alpha_1}$.

2. Efficiency rate and local deficiency

Definition 1. *For a real interval Γ let $S = \{\hat{\theta}_{\gamma,n}\}_{\gamma \in \Gamma}$ be a family of estimators such that for any $(\gamma, \varepsilon) \in \Gamma \times (0, 1)$ there is a neighbourhood $O(\gamma)$ such that $V(\hat{\theta}_\nu, \varepsilon)$ exists for all $\nu \in O(\gamma)$. Then if for some $(\gamma, \varepsilon) \in \Gamma \times (0, 1)$ the limit*

$$\lim_{\nu \to \gamma} \frac{V(\hat{\theta}_\nu, \varepsilon) - V(\hat{\theta}_\gamma, \varepsilon)}{\nu - \gamma}$$

exists, we shall call it the efficiency rate of family S with respect to the model $\mathscr{P}$ at the point (γ, ε) and denote it $\mathscr{E}\mathscr{R}(\gamma, \varepsilon)$. Moreover, if for some $\varepsilon \in (0, 1)$ and some $k \in N$ there are points $\gamma_i \in \Gamma$, $i = 1, 2, \ldots, k$ such that $\mathscr{E}\mathscr{R}(\gamma_i, \varepsilon) = 0$, find j, $1 \leqslant j \leqslant k$ such that $V(\hat{\theta}_{\gamma_j}, \varepsilon) = \min_{1 \leqslant i \leqslant k} V(\hat{\theta}_{\gamma_i}, \varepsilon)$. If then there is a neighbornhood $O(\gamma_j)$ such that $\mathscr{E}\mathscr{R}(v, \varepsilon)$ exists for all $v \in O(\gamma_j)$ and also the limit

$$\lim_{v \to \gamma_j} \frac{\mathscr{E}\mathscr{R}(v, \varepsilon)}{v - \gamma_j}$$

exists, we shall call it the local deficiency $\mathscr{L}\mathscr{D}(\varepsilon)$ of $\mathscr{S}$ with respect to the model $\mathscr{P}$ at the point ε. Moreover we shall say that the estimator $\hat{\theta}_{\gamma_j}$ is the optimal one in $\mathscr{S}$ for the mixture model $\mathscr{P}(\varepsilon)$.

Remark 1. The efficiency rate $\mathscr{E}\mathscr{R}(v, \varepsilon)$ (which is nothing else than $\frac{\partial V(\hat{\theta}_v, \varepsilon)}{\partial v}$) represents the slope of the curve $V(\hat{\theta}_v, \varepsilon)$ as a function of v for a fixed $\varepsilon \in (0, 1)$. Therefore in the neighborhood $O(\gamma_j)$ it indicates how quickly we may achieve the efficiency when changing v. Similarly, using the local deficiency we may write an approximation to (4) in the form

$$\frac{1}{2} \mathscr{L}\mathscr{D}(\varepsilon)(v - \gamma_j)^2.$$

On the other hand, (4) represents the deficiency of the estimator which we have selected, with respect to the best possible estimator. It inspired the name of the second characteristic.

Remark 2. As follows from the cases studied below we may expect to meet (frequently) with situation when $k = 1$, i.e. that there is just one point at which the efficiency rate is equal to zero.

Remark 3. It the past twenty years further models of contamination appeared. They were based e.g. on the Prokhorov or the Kolmogorov distance, on the 2-alternating (and 2-monotone) Choquet capacities or on a combination of the mixture model and the model with neighborhoods implied by the total variation.

Moreover, the idea of describing the asymptotic behaviour of statistical procedure by the efficiency rate and the local dificiency can be applied on the statistical test as well (see Víšek (1987)). Then is it necessary to give the definition of the efficiency rate and of the local deficiency in a little more general form.

3. Efficiency rate and local deficiency of Huber's location estimator

In what follows we shall derive the efficiency rate and the local deficiency for the Huber estimators of location and of the α-estimators. We shall assume further that the density g is symmetric to avoid a case when $t_0(\varepsilon)$ or $t_1(\varepsilon)$ can have infinite absolute value. The cases when either $t_0(\varepsilon) = -\infty$ or $t_1(\varepsilon) = \infty$ need to be treated separately; however it is only a technical matter. Let us denote for any random variable $\mathscr{I}$ and any distribution function $F \in \mathscr{F}$ by $E_F \mathscr{I}$ the mean value of $\mathscr{I}$ with respect to F if this meean value exists. In the above mentioned paper Huber gave for any $\varepsilon \in (0, 1)$ a formula for the supremum of the asymptotic variances of the location estimator $\hat{\Delta}_{\varepsilon, n}$ namely

$$V(\hat{\Delta}_\varepsilon, \varepsilon) = \frac{(1-\varepsilon)E_G\psi_\varepsilon^2 + \varepsilon \cdot k^2(\varepsilon)}{((1-\varepsilon)E_G\psi'_\varepsilon)^2}$$

where $\psi'_\varepsilon = \dfrac{d\psi_\varepsilon(t)}{dt}$ (this derivative exists under the Assumption A and for any $\varepsilon \in (0, 1)$ it is —as well as ψ— bounded). Following nearly slavishly Huber's derivation one finds that

$$V(\hat{\Delta}_\gamma, \varepsilon) = \frac{(1-\varepsilon)E_G\psi_\gamma^2 + \varepsilon k^2(\gamma)}{((1-\varepsilon)E_G\psi'_\gamma)^2}. \tag{5}$$

To be able to evaluate the efficiency rate and local deficiency of Huber's estimator of location (if any) we shall need the first derivative of $k(\gamma)$ with respect to γ. It will be found in the next assertion.

Assertion 1. *Let the Assumption A be fulfilled. Define for any fixed $z < 0$ and $t > z$ a function $r_z(t)$ by*

$$r_z(t) = \int_z^t g(y)\, dy + \frac{g^2(z)}{g'(z)} - \frac{g^2(t)}{g'(t)}.$$

Then $r_z(t)$ is differentiable and strictly decreasing.

PROOF: A straightforward computation gives

$$\frac{dr_z(t)}{dt} = \frac{g^2(t) \cdot g''(t) - g(t)[g'(t)]^2}{[g'(t)]^2}. \tag{6}$$

Due to requirement of the strict convexity of $-\log g(t)$ we have $g(t) \cdot g''(t) - [g'(t))]^2 < 0$ for any $t \in R$ and the proof follows. □

In what follows let us write $r_z'(t)$ instead of $\frac{dr_z(t)}{dt}$.

Lemma 1. *Under the Assumption A we have*

$$\left[\frac{dk(\gamma)}{d\gamma}\right]_{\gamma=\gamma_1} = -\frac{k^2(\gamma_1)}{(1-\gamma_1)^2}\{g(t_0(\gamma_1)) + g(t_1((\gamma_1))\}^{-1}$$

PROOF: Let $\gamma_1, \gamma_2 \in (0, 1)$ and $\gamma_1 < \gamma_2$. Then we have from (2)

$$0<(1-\gamma_2)^{-1}-(1-\gamma_1)^{-1}=\int_{t_0(\gamma_2)}^{t_2(\gamma_2)} g(t)dt+\frac{g(t_0(\gamma_2))+g(t_1(\gamma_2))}{k(\gamma_2)} -$$

$$-\int_{t_0(\gamma_1)}^{t_1(\gamma_1)} g(t)\,dt - \frac{g(t_0(\gamma_1)) + (t_1(\gamma_1))}{k(\gamma_1)}.$$

From the ASSERTION 1 it follows that $t_0(\gamma_1) < t_0(\gamma_2)$ and $t_1(\gamma_2) < t_1(\gamma_1)$. Moreover, from (3) we have $t_0(\gamma_2) < t_1(\gamma_2)$. So we have arrived at

$$-\infty < t_0(\gamma_1) < t_0(\gamma_2) < t_1(\gamma_2) < t_1(\gamma_1) < \infty. \tag{8}$$

Observe that $r_z'(t)$ does not depend on z (see (6)) and taking into account (8) one may rewrite (7) into the form

$$\frac{\gamma_2 - \gamma_1}{(1-\gamma_1)(1-\gamma_2)} = -\left\{\int_{t_0(\gamma_1)}^{t_0(\gamma_2)} r_z'(t)\,dt + \int_{t_1(\gamma_1)}^{t_1(\gamma_1)} r_z'(t)\,dt\right\}.$$

and since $r_z'(t)$ is strictly negative it implies that for $\gamma_2 \searrow \gamma_1$ we have $t_i(\gamma_2) \to t_i(\gamma_1)$ for $i = 1, 2$, and hence due to (3) also $k(\gamma_2) \to k(\gamma_1)$. Now put

$$M = \sup\{r_z'(t) : t \in T = [t_0(\gamma_1), t_0(\gamma_2)] \cup [t_1(\gamma_2), t_1(\gamma_1)]\}.$$

Since $r_z'(t)$ is negative and continuous on the closed and bounded (and hence compact) set T we have $M < 0$. But it implies

$$\gamma_2 - \gamma_1 \geqslant -(1 - \gamma_1)(1 - \gamma_2)M\{t_0(\gamma_2) - t_0(\gamma_1) + t_1(\gamma_1) - t_1(\gamma_2)\} > 0.$$

Let R^* be the image of the support S of the density $g(t)$ obtained by $-g'(t)/g(t)$. Since the ASSUMPTION A implies that $g'(t)/g(t)$ is striclty monotone we may define $\xi(y): R^* \to S$ as an inverse mapping to $-g'(t)/g(t)$. Due to positivity of $g(t)$ on S and existence and continuity of $g''(t)$ (see the ASSUMPTION A) there is a continuous derivative of ξ, say ξ'. Now taking into account (3) we may rewrite (2) into the form

$$(1 - \gamma)^{-1} = \int_{\xi(-k(\gamma))}^{\xi(k(\gamma))} g(t)\, dt + \frac{g(\xi(-k(\gamma))) + g(\xi(k(\gamma)))}{k(\gamma)}.$$

Moreover we have

$$\int_{\xi(-k(\gamma_1))}^{\xi(-k(\gamma_2))} g(t)\, dt = \int_{\xi(-k(\gamma_1))}^{\xi(-k(\gamma_2))} \{g(\xi(-k(\gamma_1))) + + g'(\lambda_{\gamma_1, \gamma_2, v})[v - \xi(-k(\gamma_1))]\}\, dv \tag{10}$$

for a suitable point $\lambda_{\gamma_1, \gamma_2, v}$ which is an element of $[\xi(-k(\gamma_1)), v]$. Denote

$$R(\gamma_2) = \int_{\xi(-k(\gamma_1))}^{\xi(-k(\gamma_2))} g'(\lambda_{\gamma_1, \gamma_2, v})[v - \xi(-k(\gamma_1))]\, du + + \int_{\xi(k(\gamma_2))}^{\xi(k(\gamma_1))} g'(\eta_{\gamma_1, \gamma_2, u})[\xi(k(\gamma_1)) - u\, du$$

where again $\eta_{\gamma_1, \gamma_2, u}$ is an appropriately selected point from $[u, \xi(k(\gamma_1))]$. Since $g'(y)$ is continuous, $v \in [\xi(-k(\gamma_1)), \xi(-k(\gamma_2))]$ and $u \in [\xi(k(\gamma_2)), \xi(k(\gamma_1))]$ we may find for a fixed (and sufficiently small) $\Delta > 0$ finite constants K_1 and K_2 (depending only on Δ) such that for any $\gamma_2 \in (\gamma_1, \gamma_1 + \Delta)$ we have

$$\mathrm{K}_1\{[t_0(\gamma_2) - t_0(\gamma_1)]^2 + [t_1(\gamma_1) - t_1(\gamma_2)]^2\} \leqslant R(\gamma_2) \leqslant \leqslant K_2\{[t_0(\gamma_2) - t_0(\gamma_1)]^2 + [t_1(\gamma_1) - t_1(\gamma_2)]^2\}$$

But it together with (8) and (9) implies that

$$\lim_{\gamma_2 \searrow \gamma_1} \frac{1}{\gamma_2 - \gamma_1} R(\gamma_2) = 0 \tag{11}$$

because $t_i(\gamma_2) \to t_i(\gamma_1)$ for $\gamma_2 \to \gamma_1$ and $i = 0, 1$. Using (10) we may rewvite (7) as follows.

$$\frac{\gamma_2 - \gamma_1}{(1-\gamma_1)(1-\gamma_2)} = -g(\xi(-k(\gamma_1)))[\xi(-k(\gamma_2)) - \xi(-k(\gamma_1))] - $$
$$- g(\xi(k(\gamma_1)))[\xi(k(\gamma_1)) - \xi(k(\gamma_2))] - R(\gamma_2) +$$
$$+ \frac{g(\xi(-k(\gamma_2))) + g(\xi(k(\gamma_2)))}{k(\gamma_2)} - \frac{g(\xi(-k(\gamma_1))) + g(\xi(k(\gamma_1)))}{k(\gamma_1)}.$$

Now, using the Mean Value Theorem, and selecting appropriately points $v_{\gamma 2}$, $w_{\gamma 2} \in (-k(\gamma_1), -k(\gamma_2))$, $z_{\gamma 2}$, $y_{\gamma 2} \in (k(\gamma_2), k(\gamma_1))$, $v_{\gamma 2} \in (\xi(-k(\gamma_1)), \xi(-k(\gamma_2)))$ and $\tau_{\gamma 2} \in (\xi(k(\gamma_2)), \xi(k(\gamma_1)))$ we obtain (subscript γ_2 should emphasize that γ_1 is assumed fixed and we are looking for a limit when $\gamma_2 \searrow \gamma_1$)

$$\frac{\gamma_2 - \gamma_1}{(1-\gamma_1)(1-\gamma_2)} + R(\gamma_2) = \{-g(\xi(-k(\gamma_1))) \cdot \xi'(v_{\gamma_2}) -$$
$$- g(\xi(k(\gamma_1))) \cdot \xi'(z_{\gamma_2}) + T(\gamma_1, \gamma_2)\} \times (k(\gamma_1) - k(\gamma_2))$$

where

$$T(\gamma_2, \gamma_1) = \frac{g(\xi(-k(\gamma_1))) + g(\xi(k(\gamma_1))) + k(\gamma_1)[g'(v_{\gamma_2}) \cdot \xi'(w_{\gamma_2}) - g'(\tau_{\gamma_2}) \cdot \xi'(y_{\gamma_2})]}{k(\gamma_1) \cdot k(\gamma_2)}.$$

Taking into account (3) and continuity of g and ξ'

$$\lim_{\gamma_2 \searrow \gamma_1} \left\{ -g(\xi(-k(\gamma_1))) \cdot \xi'(v_{\gamma_2}) + \frac{g'(v_{\gamma_2}) \cdot \xi'(w_{\gamma_2})}{k(\gamma_2)} \right\} = 0$$

and

$$\lim_{\gamma_2 \searrow \gamma_1} \left\{ -g(\xi(k(\gamma_1))) \cdot \xi'(z_{\gamma_2}) - \frac{g'(\tau_{\gamma_2}) \cdot \xi'(y_{\gamma_2})}{k(\gamma_2)} \right\} = 0.$$

But then we conclude from (11) and from the continuity of g and ξ' that

$$\frac{1}{(1-\gamma_1)^2} = \frac{g(t_0(\gamma_1)) + g(t_1(\gamma_1))}{k_2(\gamma_1)} \cdot \lim_{\gamma_2 \searrow \gamma_1} \frac{k(\gamma_1) - k(\gamma_2)}{\gamma_2 - \gamma_1}$$

Along similar lines we may obtain the same equation for the left-hand side limit and the proof follows. □

Remark 4. From the LEMMA 1 it is clear that under the Assumption A the derivative $\frac{d^2k(\gamma)}{d\gamma^2}$ also exists but as we shall see later we shall not need the second derivative of $k(\gamma)$ although it may seem that for evaluation of the second derivative of the supremum of asymptotic variances $V(\hat{\Delta}_\gamma, \varepsilon)$ (with respect to γ) we need it.

It what follows we shall use the LEMMA 1 only in the sense that it guarantees that the first derivative of $k(\gamma)$ with respect to γ exists (and is strictly negative for all $\gamma \in (0, 1)$). The explicit formula for this derivative is useful only when we need to evaluate its numerical value, e.g. when evaluating the efficiency rate or the local deficiency.

According to the Definition 1 we have for Huber's estimator of location

$$\mathscr{E}\mathscr{R}(\gamma, \varepsilon) = \lim_{v \to \gamma} (v-\gamma)^{-1}(1-\varepsilon)^{-2}\{(1-\varepsilon)[(E_G\psi'_\gamma)^2 E_G\psi_v^2 - (E_G\psi'_v)^2 E_G\psi_\gamma^2] + \varepsilon[k^2(v)(E_G\psi'_\gamma)^2 - k^2(\gamma)(E_G\psi'_v)^2]\}(E_G\psi'_\gamma)^{-2}(E_G\psi'_\gamma)^2 \tag{12}$$

if this limit exists. Hence the following lemma will be helpfull.

Lemma 2. *Under the assumption A we have*

$$\frac{dE\psi_\gamma^2}{d\gamma} = 2 \cdot k(\gamma)[G(t_0(\gamma)) + 1 - G(t_1(\gamma))] \frac{dk(\gamma)}{d\gamma} \tag{13}$$

and

$$\frac{dE\psi'_\gamma}{d\gamma} = [g(t_0(\gamma)) + g(t_1(\gamma))] \frac{dk(\gamma)}{d\gamma}\,. \tag{14}$$

PROOF: Earlier than we start to prove assertion of the lemma let us recall that $\xi(z)$ is the inverse function to $\psi_0(t) = -\frac{g'(t)}{g(t)}$ and hence $\xi'(z)$ exists under the Assumption A. From the definition of ψ_γ we have for $\gamma < v$

$$E_G\psi_v^2 - E_G\psi_\gamma^2 = G(\xi(-k(v)))k^2(v) - G(\xi(-k(\gamma)))k^2(\gamma) - \int_{\xi(-k(\gamma))}^{\xi(-k(v))} \psi_\gamma^2(t)g(t)\,dt - \int_{\xi(k(v))}^{\xi(k(\gamma))} \psi_\gamma^2(t)g(t)\,dt + [1 - G(\xi(k(v)))]k^2(v) - [1 - G(\xi(k(\gamma)))]k^2(\gamma)$$

because

$$\xi(-k(\gamma)) < \xi(-k(\nu)) < \xi(k(\nu)) < \xi(k(\gamma))$$

(see (8)). Then we easy verify that

$$\lim_{\nu \searrow \gamma} (\nu - \gamma)^{-1}[G(\xi(-k(\nu)))k^2(\nu) - G(\xi(-k(\gamma)))k^2(\gamma)] =$$

$$= \lim_{\nu \searrow \gamma} (\nu-\gamma)^{-1}$$

$$\{[G(\xi(-k(\nu))) - G(\xi(-k(\gamma)))]k^2(\nu) + G(\xi(-k(\gamma)))[k^2(\nu) - k^2(\gamma)]\} =$$

$$= g(\xi(-k(\gamma)))\xi'(-k(\gamma))k^2(\gamma)\frac{\mathrm{d}k(\gamma)}{d\gamma} + 2G(\xi(-k(\gamma)))k(\gamma)\frac{dk(\gamma)}{d\gamma}. \tag{16}$$

Making use of the Theorem about differentiating an indefinite integral with respect to its boundaries (see Hewitt, Stromberg (1965), theorem 18.3) we arrive at

$$\lim_{\nu \searrow \gamma} (\nu - \gamma)^{-1} \int_{\xi(-k(\gamma))}^{\xi(-k(\nu))} \psi_\gamma^2(t)g(t)\,dt =$$

$$= \psi_\gamma^2(\xi(-k(\gamma)))g(\xi(-k(\gamma))) \cdot \xi'(-k(\gamma))\frac{dk(\gamma)}{d\gamma}. \tag{17}$$

Realize that the integral from (17) stays in (15) with the negative sign and that $\psi_\gamma^2(\xi(-k(\gamma)) = k^2(\gamma)$. Hence when we shall look for the difference of the right-hand sides of (16) and (17). We obtain

$$2k(\gamma)G(\xi(-k(\gamma)))\frac{dk(\gamma)}{d\gamma} = 2k(\gamma) \cdot G(t_0(\gamma)) \cdot \frac{dk(\gamma)}{d\gamma}.$$

Deriving a similar expression for the third term and for the difference of the forth and the fifth terms in (15) one concludes the proof of (13) for $\nu \searrow \gamma$. The proof of (13) for $\nu \nearrow \gamma$ can be carried out nearly along the same lines.

Now the proof of (14). Let $\nu < \gamma$. Then

$$E_G\psi'_\nu - E_G\psi'_\gamma = \int_{\xi(-k(\nu))}^{\xi(-k(\gamma))} \psi'_\nu(t)g(t)\,dt + \int_{\xi(k(\gamma))}^{\xi(k(\nu))} \psi'_\nu(t)g(t)\,dt. \tag{18}$$

Using once again the Theorem about differentiating an indefinite integral we obtain

$$\lim_{v \to \gamma} (v - \gamma)^{-1}[E_G\psi'_v - E_G\psi'_\gamma] = \{\psi'_\gamma(\xi(-k(\gamma))) \cdot g(\xi(-k(\gamma))) \cdot \xi'(-k(\gamma)) + \\ + \psi'_\gamma(\xi(k(\gamma)))g(\xi(k(\gamma))) \cdot \xi'(k(\gamma))\} \frac{dk(\gamma)}{d\gamma}$$

(notice that the boundaries of the first integral in (18) are in the opposite order —with respect to v and γ— then of the second one). Since $\psi'_\gamma(\xi(k(\gamma))) = [\xi'(k(\gamma))]^{-1}$, $\xi(-k(\gamma)) = t_0(\gamma)$ and $\xi(k(\gamma)) = t_1(\gamma)$ the proof of (14) for $v \nearrow \gamma$ follows. Deriving the same expression for $v \searrow \gamma$ one concludes the proof of (14).

Theorem 1. *Under the Assumption A we have for the family $\mathscr{S}_{0,n}$ and for the model of contamination $\mathscr{P}$*

$$\mathscr{E}\mathscr{R}(\gamma, \varepsilon) = 2 \cdot \frac{dk(\gamma)}{d\gamma} \Big\{ - V(\hat{\Delta}_\gamma, \varepsilon)[g(t_0(\gamma)) + g(t_1(\gamma))](E_G\psi'_\gamma)^{-1} + \\ + k(\gamma)[[G(t_0(\gamma)) - G(t_1(\gamma))](1 - \varepsilon) + 1](1 - \varepsilon)^{-2}(E_G\psi'_\gamma)^{-2} \Big\}$$

for any $(\gamma, \varepsilon) \in (0, 1)^2$. Moreover, the local deficiency is given by the formula

$$\mathscr{L}\mathscr{D}(\varepsilon) = 2\left[\frac{dk(\gamma)}{d\gamma}\right]^2_{\gamma=\varepsilon} (E\psi'_\varepsilon)^{-1}\{-V(\hat{\Delta}_\varepsilon, \varepsilon)[g'(t_1(\varepsilon))\xi'(k(\varepsilon)) - \\ - g'(t_0(\varepsilon))\xi'(-k(\varepsilon))] + [[G(t_0(\varepsilon)) - G(t_1(\varepsilon))]](1 - \varepsilon) + 1 - \\ - k(\varepsilon)[g(t_0(\varepsilon))\xi'(-k(\varepsilon)) + g(t_1(\varepsilon))\xi'(k(\varepsilon))](1 - \varepsilon)^{-2}(E\psi'_\varepsilon)^{-1} - \\ - k(\varepsilon)[[G(t_0(\varepsilon)) - G(t_1(\varepsilon))](1 - \varepsilon) + 1][g(t_0(\varepsilon)) + \\ + g(t_1(\varepsilon))](1 - \varepsilon)^{-2}(E\psi'_\varepsilon)^{-2}\}. \qquad (20)$$

PROOF: Let us rewrite the right-hand side of (12) into the form

$$(v-\gamma)^{-1}(1-\varepsilon)^{-2}\{(1-\varepsilon)[(E_G\psi'_\gamma)^2[E_G\psi_v^2 - E_G\psi_\gamma^2] + E_G\psi_\gamma^2[(E_G\psi'_\gamma)^2 - (E_G\psi'_v)^2]] + \\ + \varepsilon[k^2(v)[(E_G\psi'_\gamma)^2 - (E_G\psi'_v)^2] + (E_G\psi'_v)^2[k^2(v) - k^2(\gamma)]]\}(E_G\psi'_v)^2(E_G\psi'_\gamma)^{-2}.$$

(Since the all mean values will be taken with respect to G let us drop the subscript G.) Now taking the limit for $v \to \gamma$ and using the LEMMA 1 one achieves

$$(1-\varepsilon)^{-2}\Big\{(1-\varepsilon)\Big[(E\psi'_\gamma)^2 \cdot 2k(\gamma)[G(t_0(\gamma)) + 1 - G(t_1(\gamma))]\frac{dk(\gamma)}{d\gamma} - E\psi_\gamma^2 \cdot 2 \cdot E\psi'_\gamma[g(t_0(\gamma)) + g(t_1(\gamma))]\frac{dk(\gamma)}{d\gamma}\Big] - \varepsilon\Big[k^2(\gamma)\cdot 2 \cdot E\psi'_\gamma[g(t_0(\gamma)) + g(t_1(\gamma))]\frac{dk(\gamma)}{d\gamma} - (E\psi'_\gamma)^2 \cdot 2 \cdot k(\gamma)\cdot\frac{dk(\gamma)}{d\gamma}\Big]\Big\}\cdot(E\psi'_\gamma)^{-4}.$$

Taking into account (5) one verifies (19). Earlier than we will continue with the second part of the proof of Theorem let us recall that Huber's estimator $\hat{\Delta}_{\varepsilon,n}$ is the minimax estimator which implies that $V(\hat{\Delta}_\varepsilon, \varepsilon) = \min_{0<\gamma<1} V(\hat{\Delta}_\gamma, \varepsilon)$ and hence (due to existence of the partial derivative of $V(\hat{\Delta}_\gamma, \varepsilon)$ with respect to γ)

$$\mathscr{E}\mathscr{R}(\varepsilon, \varepsilon) = \left[\frac{\partial V(\hat{\Delta}_\gamma, \varepsilon)}{\partial \gamma}\right]_{\gamma=\varepsilon} = 0.$$

Now, to find the formula for the local deficiency let us write down the derivative of the righ-hand side of (19). We obtain

$$\frac{d^2k(\gamma)}{d\gamma^2}\cdot\mathscr{E}\mathscr{R}(\gamma,\varepsilon)\cdot\left[\frac{dk(\gamma)}{d\gamma}\right]^{-1} + 2\frac{dk(\gamma)}{d\gamma}\Big\{\Big[-\frac{\partial V(\hat{\Delta}_\gamma,\varepsilon)}{\partial\gamma}[g(t_0(\gamma)) + g(t_1(\gamma))] - V(\hat{\Delta}_\gamma,\varepsilon)[-g'(t_0(\gamma))\cdot\xi'(-k(\gamma)) + g'(t_1(\gamma))\xi'(k(\gamma))]\frac{dk(\gamma)}{d\gamma}\Big](E\psi'_\gamma)^{-1} + V(\hat{\Delta}_\gamma,\varepsilon)[g(t_0(\gamma)) + g(t_1(\gamma))](E\psi'_\gamma)^{-2}\cdot\frac{dE\psi'_\gamma}{d\gamma} + \Big\{\frac{dk(\gamma)}{d\gamma}[[G(t_0(\gamma)) - G(t_1(\gamma))](1-\varepsilon) + 1] + k(\gamma)[-g(t_0(\gamma))\xi'(-k(\gamma)) - g(t_1(\gamma))\xi'(k(\gamma))](1-\varepsilon)\frac{dk(\gamma)}{d\gamma}\Big\}(1-\varepsilon)^{-2}(E\psi'_\gamma)^{-2} -$$

$$- 2k(\gamma)[[G(t_0(\gamma)) - G(t_1(\gamma))](1 - \varepsilon) + 1](1 - \varepsilon)^{-2}(E\psi'_\gamma)^{-3}\frac{dE\,\psi'_\gamma}{d\gamma}\Big\}.$$

For $\gamma = \varepsilon$ the first two terms of (21) are equal to zero. Moreover notice that the sum of the fourth term and of the half of the last term (i.e. without the factor "2") gives

$$\left\{-\mathscr{E}\mathscr{R}(\gamma,\ \varepsilon)\cdot(E\psi'_\varepsilon)^{-1}\cdot\frac{dE\psi'_\gamma}{d\gamma}\right\}_{\gamma=\varepsilon}$$

and hence it disappears for $\gamma = \varepsilon$, too. The rest of (21) may be written in the form

$$2\left[\frac{dk(\gamma)}{d\gamma}\right]^2(E\psi'_\gamma)^{-1}\{-V(\hat{\Delta}_\gamma,\varepsilon)[g'(t_1(\gamma))\xi'(k(\gamma)) - g'(t_0(\gamma))\xi'(-k(\gamma))] +$$
$$+ [[G(t_0(\gamma)) - G(t_1(\gamma))](1 - \varepsilon) + 1 -$$
$$- k(\gamma)[g(t_0(\gamma))\xi'(-k(\gamma)) + g(t_1(\gamma))\xi'(k(\gamma))](1 - \varepsilon)](1 - \varepsilon)^{-2}(E\psi'_\gamma)^{-1} -$$
$$- k(\gamma)[[G(t_0(\gamma)) - G(t_1(\gamma))](1 - \varepsilon) + 1][g(t_0(\gamma)) + g(t_1(\gamma))](1 - \varepsilon)^{-2}(E\psi'_\gamma)^{-2}\}.$$

Substituting ε instead of γ gives (20).

Remark 5. The numerical evaluation of (19) and (20) may be sometimes considerably simplified by the utilization of formula (2), written in the form

$$[G(t_0(\varepsilon)) - G(t_1(\varepsilon))](1 - \varepsilon) + 1 = k^{-1}(\varepsilon)(1 - \varepsilon)[g(t_0(\varepsilon)) + g(t_1(\varepsilon))].$$

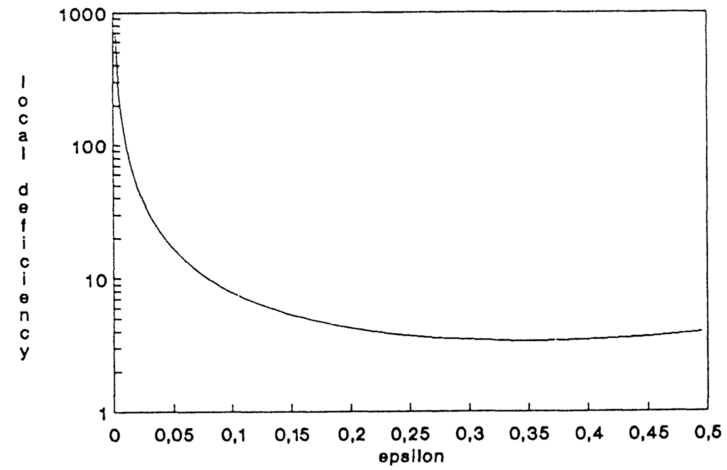

Figure 1. The local deficiency of the family of Huber's estimators for $\varepsilon \in (0,\ 0.5)$

4. Efficiency rate and local deficiency of the α-estimators

To illustrate the usefulness of the introduced notions we shall apply them on a family of estimators which was constructed without any apparent connection with a model of contamination. The α-estimators have been chosen for this purpose. For the definition and more information about their properties see Vajda (1984) and references given there. For the convenience of reader let us recall some basic facts. The α-estimators are estimators of the minimal distance type where a modified version of α-divergence was used as a measure of distance between the empirical distribution function and a distribution function from an assumed family $\{F_\theta\}_{\theta \in \Theta}$. The α-divergence of two probability measures, say P and Q (defined on a measurable space (Ω, β)) is the mean value of their transformed likelihood ratio. In other words, the α-divergence is a special case of f-divergence introduced by Csiszár in 1967. The transformation of the likelihood ratio has the form

$$t_\alpha(u) = [\text{sign}(1 - \alpha)](1 - u^\alpha)/\alpha$$

for $\alpha \in (0, 1) \cup (1, \infty)$ with $t_0(u) = -\ln u$. It is clear that for the case when P is an empirical distribution and Q an absolutely continuous one we obtain the α-divergence is equal to ∞. Hence a modification of the α-divergence is inevitable. One possibility is to restrict ourselves, when defining α-divergence, on a sequence of algebras $\{\beta_n\}_{n=1}^{\infty} \to \beta$ —see Vajda (1984)— to obtain a reasonable result. The family $\{F_\theta\}_{\theta \in \Theta}$ (usually denoted as the projection family) is selected to explain the data and need not necessarily include the distribution function which generated data. To give the reader an idea about behaviour of the α-estimators let us recall that their influence functions are frequently of redescending type and that they are close to the influence function of the well-known M-estimators with the redescending ψ function. If we compare the figure 4.1 of Vajda (1984) which gives the influence function of the α-estimator for $\{F_\theta\}_{\mu \in R}$ and $\alpha = 0.2$, with the figure of influence function of A 25 of Andrews et al. (1972), we find that the influence functions are very similar. We shall restrict ourselves on the family of the normal distribution with $\mu \in R$ and $\sigma^2 = 1$, i.e. we will consider the family $\{N(\mu, 1)\}_{\mu \in R}$. Then the α-estimators are defined as follows

$$= \frac{1}{n} \sum_{i=1}^{n} X_i \qquad \text{for } \alpha = 0,$$

$$\hat{\mu}_\alpha = \operatorname{argmin}_{t \in R} \sum_{i=1}^{n} \exp\left\{-\frac{\alpha(X_i - t)^2}{2}\right\} \quad \text{for } \alpha \in (0, 1).$$

So as the family S we shall consider $\{\hat{\mu}_\alpha\}_{\alpha \in [0, 1)}$. (Naturally $\hat{\mu}_\alpha$ depends also on the sample size n but because we shall consider as the criterial function again the asymptotic variance of $\hat{\mu}_\alpha$, this dependence need not be emphasized in the notation.) By the model of contamination we shall understand again $P = \{P(\varepsilon)\}_{\varepsilon \in [0, 1]}$. Finally, by $V(\hat{\mu}_\alpha, \varepsilon)$ we shall denote, similarly as above, the asymptotic variance of $\hat{\mu}_\alpha$. Since in Vajda (1984) the influence function of the α-estimators was derived a simple computation gives for our case

$$V(\hat{\mu}_\alpha, \varepsilon) = \frac{(1+\alpha)^3}{(1-\varepsilon)^2} \{(2\alpha + 1)^{-3/2}(1 - \varepsilon) + \varepsilon\alpha^{-1} \exp\{-1\}\}.$$

Taking the derivative we obtain

$$\mathscr{E}\mathscr{R}(\hat{\mu}_\alpha, \varepsilon) = \left[\frac{1+\alpha}{1-\varepsilon}\right]^2 \left\{\frac{3\alpha(1-\varepsilon)}{(2\alpha+1)^{5/2}} + \frac{2\alpha - 1}{\alpha^2} \cdot \varepsilon \cdot \exp\{-1\}\right\}.$$

It allows to establish a relation between ε and α, namely to find such $\alpha(\varepsilon)$ for which $\mathscr{E}\mathscr{R}(\hat{\mu}_{\alpha(\varepsilon)}, \varepsilon) = 0$. We find immediately that it is simpler to give $\varepsilon(\alpha)$ (i.e. for a fixed α find such $\varepsilon(\alpha)$ for which $\mathscr{E}\mathscr{R}(\hat{\mu}_\alpha, \varepsilon(\alpha)) = 0$). We obtain

$$\varepsilon(\alpha) = \left[1 + \frac{\exp\{-1\}(2\alpha + 1)^{5/2}(1 - 2\alpha)}{3\alpha^3}\right]^{-1}.$$

The Figure 2 presents this relation and shows that for ε increasing from

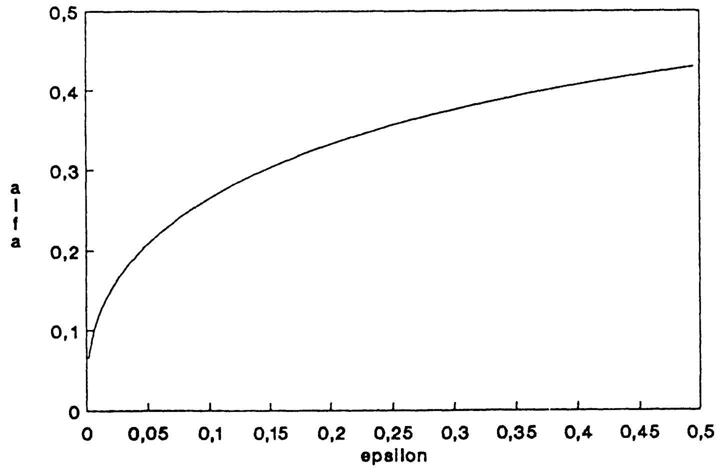

Figure 2. Dependence of the optimal α on ε for the α-estimators.

0 to 0.5 $\alpha(\varepsilon)$ increases from 0 to 0.43. In other words, it means that for any contamination level (represented by ε) we can find $\alpha \in (0, 0.5)$ for $\mathscr{E}\mathscr{R}(\hat{\mu}_{\alpha(\varepsilon)}, \varepsilon) = 0$. It implies that if we show that for the corresponding values ε and $\alpha(\varepsilon)$ the local deficiency is positive, we can find for any contamination level ε such value $\alpha(\varepsilon)$ that $\hat{\mu}_{\alpha(\varepsilon)}$ has minimal possible variance among $\{\hat{\mu}_\alpha\}_{\alpha \in (0,\,0.5)}$. Let us recall that it is reasonable to take into account only $\varepsilon \in (0, 0.5)$ because for $\varepsilon \in (0.5, 1)$ we have a larger part of contamination among data then the part of proper data and it seems not to be generally senseful.

Now let us derive the formula for the local deficiency. We arrive at

$$\mathscr{L}\mathscr{D}(\varepsilon) = \left[\frac{1 + \alpha(\varepsilon)}{1 - \varepsilon}\right]^2 \left\{3(1 - \varepsilon)\frac{1 - 3\alpha}{(2\alpha(\varepsilon) + 1)^{7/2}} + \right.$$

$$\left. + \varepsilon \frac{2(1 - \alpha(\varepsilon))}{\alpha^3(\varepsilon)} \exp\{-1\}\right\}.$$

The Figure 3 exhibits the dependence of the local deficiency on the contamination level ε. We see that $\mathscr{L}\mathscr{D}(\varepsilon)$ is positive for $\varepsilon \in (0, 1)$ and it implies that $V(\hat{\mu}_\alpha, \varepsilon(\alpha))$ is indeed minimal among all $V(\hat{\mu}_\nu, \varepsilon(\alpha))$, $\nu \in (0, 0.5)$.,

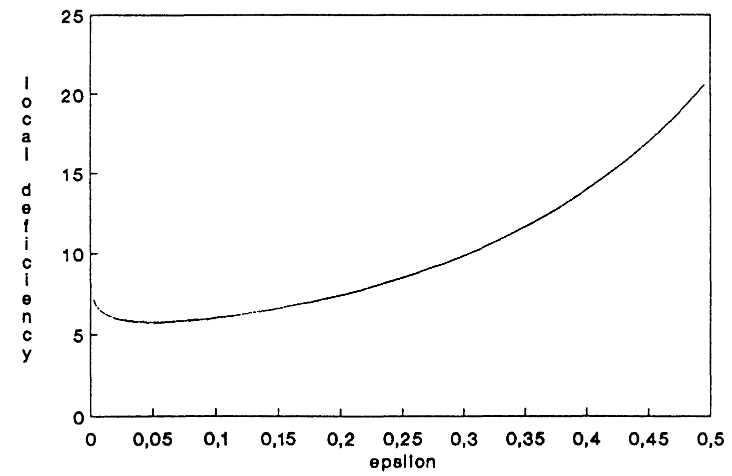

Figure 3. The local deficiency of the α-estimators.

5. Concluding remarks

The second example considering the family of the α-estimators demonstrated that using the efficiency rate we may find the estimator with the minimal (asymptotic) variance in given family of estimators — variances being taken with respect to a model of contamination. Naturally, it would be possible to give another examples of the applications of the introduced notions, for instance the optimal choice of regression quantiles for given contamination level. The authors hope to do so in a next paper. For the extensive example of the application of the notions in the statistical testing we refer to Víšek (1987).

Using the local deficiency we can write the approximation to the difference $V(\hat{\mu}_\alpha, \varepsilon) - V(\hat{\mu}_{\alpha(\varepsilon)}, \varepsilon)$ in the form $\frac{1}{2} \mathscr{L}\mathscr{D}(\varepsilon)[\alpha - \alpha(\varepsilon)]^2$, i.e. the approximation to our loss.

It may be also interesting to compare the local deficiency of different families of estimators. The Figure 4 offers such comparison of the local deficiency of Huber's estimators of location and of the α-estimators. We may observe that for ε less than 0.125 the local deficiency of the α-estimators is less than that one of Huber's estimators of location. On the other hand for the values of $\varepsilon \in (0.125, 0.5)$ the local deficiency of the α-estimators is somewhat larger. As it was already mentioned the present point of view of the applied statisticians is that the usual contamination

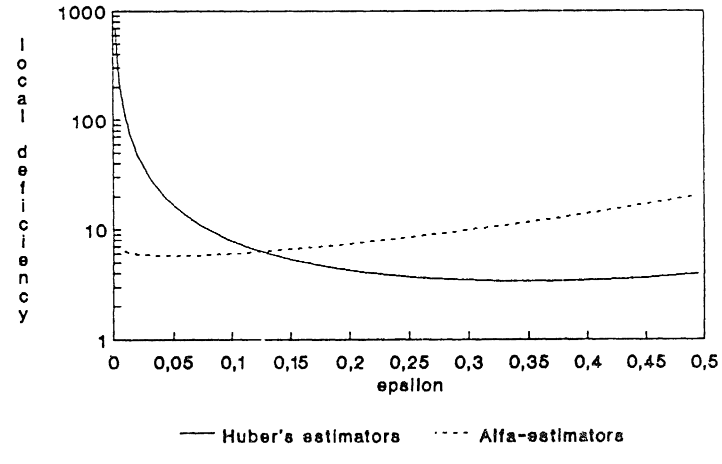

Figure 4. Comparison of the local deficiency of Huber's estimators and of α-estimators.

level is somewhere between 1 % and 10 % (even for very carerfully measured data — see Hampel et al. (1986)). Taking into account all these facts it may be preferable to use the α-estimators instead of Huber's ones, especially for the small values of ε.

References

ANDREWS, D. F.; BICKEL, P. J.; HAMPEL, F. R.; HUBER, P. J.; ROGER, W. H., and TUKEY, J. W. (1972): *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, N.J.

CSISZÁR, I. (1967): "Information type measures of difference of probability distributions and indirect observations". *Studia Sci. Math. Hungar.*, 2, pp. 209-318.

HAMPEL, F. R. (1968): "Contributions to the theory of robust estimation". *Unpublished dissertation*, Univ. of California, Berkely.

HAMPEL, F. R.; RONCHETTI, E. M.; ROUSSEUW, P. J., and STAHEL, W. A. (1986): *Robust Statistics — The Approach Based on Influence Functions*. J. Wiley & Sons, New York.

HAMPEL, F. R.; ROUSSEUW, P. J., and RONCHETTI, E. M. (1981): "The Change-of-variance curve and optimal redescending *M*-estimators". *JASA*, 1981, Volume 76, Number 375, pp. 643-648.

HEWITT E., and STROMBERG, K. (1965): *Real and Abstract Analysis*, Springer-Verlag, Berlin-Heidelberg-New York.

HUBER, P. J. (1964): "Robust eswtimation of location parameter". *Ann. Math. Statist.*, 35, pp. 73-101.

VAJDA, I. (1984): "Asymptotic efficiency and robustness of *D*-estimators", *Kybernetika*, Volume 20 (1984), Number 5, pp. 358-375.

VÍŠEK, J. Á. (1985): "Estimating contamination level". *Proc. of the 5th Pannonian Symp. on Math. Stat.*, Visegrád, Hungary, pp. 401-414.

VÍŠEK, J. Á. (1987): "Efficiency rate and local deficiency of the most powerful test in the model of contaminacy with general neighbourhoods", *Kybernetika*, Volume 23 (1987), Number 5, pp. 370-381.