Optimal bounded influence regression and scale M-estimators in the context of experimental design

Tadeusz Bednarski, Pedagogical Univ. in Zielona Góra and Polish Ac. of Sc.
Christine H. Müller, Carl von Ossietzky University Oldenburg [1]

September, 2000

Abstract

The problem of simultaneous robust estimation of regression and scale parameters in the linear regression model is studied in the context of experimental design. Optimal M estimates are given for a modified optimization problem of minimizing the asymptotic variances under bounded influence functions. This is done by reducing the multi-dimensional regression problem to the problem of estimating one-dimensional location and scale. For the location-scale case two subfamilies of optimal score functions are described in detail along with comparisons of the asymptotic variances and gross-error-sensitivities of the corresponding M estimators. It turns out that, even for small gross-error-sensitivities, one of the subfamilies provides variances which are close to those of the nonrobust maximum likelihood estimators.

Keywords: Regression-scale estimation, location-scale estimation, bounded influence function, asymptotic variances, optimality.

AMS subject classification: Primary 62F35; secondary 62J05.

1. Introduction

Assume a general linear model

$$Y_{nN} = x_{nN}^\top \beta + Z_{nN},\ n = 1, \dots, N,$$

where Y_{nN} are random observations, $x_{nN} \in \Re^r$ are known experimental conditions, $\beta \in \Re^r$ is an unknown parameter vector and Z_{nN} are random errors. Realizations of $Y_{1N}, \dots, Y_{NN}$ are denoted by $y_{1N}, \dots, y_{NN}$. We assume that for some unknown $\sigma \in \Re^+$ the errors Z_{nN}/σ are distributed according to a probability measure P for all $n = 1, \dots, N$. The aim is to estimate the unknown parameters β and σ.

[1]Research supported by the grant 436 POL 17/18/96 of the Deutsche Forschungsgemeinschaft and partly by grant 2 P03A 016 08 of KBN

If P has a density f then the maximum likelihood estimator for (β, σ) is given by

$$\begin{aligned}(\widehat{\beta}, \widehat{\sigma}) &\in \arg\max_{\beta,\sigma} \ln \prod_{n=1}^{N} \frac{1}{\sigma} f\left(\frac{y_n - x_n^\top \beta}{\sigma}\right) \\ &= \arg\max_{\beta,\sigma} \sum_{n=1}^{N} \left[-\ln\sigma + \ln f\left(\frac{y_n - x_n^\top \beta}{\sigma}\right)\right].\end{aligned}$$

This estimator is generalized by general M estimators which are given by

$$(\widehat{\beta}, \widehat{\sigma}) \in \arg\min_{\beta,\sigma} \sum_{n=1}^{N} \left[-\ln\sigma + \rho\left(\frac{y_n - x_n^\top \beta}{\sigma}, x_n\right)\right]. \tag{1.1}$$

Set

$$\psi(z,x) := -\frac{\partial}{\partial z}\,\rho(z,x), \quad \psi_\beta(z,x) := \frac{1}{\sigma}\,\psi(z,x)\,x,$$

and

$$\psi_\sigma(z,x) := \frac{1}{\sigma}\,(z\,\psi(z,x) - 1).$$

Necessary conditions for a solution $(\widehat{\beta}, \widehat{\sigma})$ of (1.1) are

$$\sum_{n=1}^{N} \psi_\beta\left(\frac{y_n - x_n^\top \widehat{\beta}}{\widehat{\sigma}}, x_n\right) = 0 \tag{1.2}$$

and

$$\sum_{n=1}^{N} \psi_\sigma\left(\frac{y_n - x_n^\top \widehat{\beta}}{\widehat{\sigma}}, x_n\right) = 0. \tag{1.3}$$

The simultaneous robust estimators of location and scale in the regression model were recently considered in Bednarski and Zontek (1994, 1996) and by Huggins (1993) in connection with estimation of fixed effects and random components in mixed models.

Here we will study the behavior of the simultaneous M estimators in the context of experimental design when the sample size N tends to infinity and P is the standard normal distribution. For that purpose we assume that the design $x_{1N}, \ldots, x_{NN}$ is converging to an asymptotic design measure δ with finite support $\mathrm{supp}(\delta)$ in the following sense

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e_{x_{nN}}(x) = \delta(\{x\})$$

for all $x \in \mathrm{supp}(\delta)$, where $e_{x_{nN}}$ denotes the Dirac measure on x_{nN}.

To obtain consistency of the M estimators we need the assumptions

$$\int \psi(z,x) P(dz) = 0 \qquad \text{(consistency for } \beta\text{)}, \tag{1.4}$$

and

$$\int \psi(z,x)\, z\, P(dz) = 1 \qquad \text{(consistency for } \sigma\text{)}, \tag{1.5}$$

for all $x \in \operatorname{supp}(\delta)$. Sufficient for the consistency condition (1.4) is that $\psi(\cdot, x)$ is an odd function for all $x \in \operatorname{supp}(\delta)$. If $\psi(\cdot, x)$ is odd then we have additionally

$$\int \psi(z,x)\, z^2\, P(dz) = 0 \tag{1.6}$$

and

$$\int \psi(z,x)^2\, z\, P(dz) = 0 \tag{1.7}$$

for all $x \in \operatorname{supp}(\delta)$.

Set

$$\begin{aligned}
\Lambda_\beta &:= \frac{1}{\sigma}\, z\, x \ \text{ and } \Lambda_\sigma := \frac{1}{\sigma}(\, z^2 - 1), \\
M(\psi,\delta) &:= \int (\psi_\beta(z,x)^\top, \psi_\sigma(z,x))^\top (\Lambda_\beta(z)^\top, \Lambda_\sigma(z))\, P(dz)\, \delta(dx), \\
Q(\psi,\delta) &:= \int (\psi_\beta(z,x)^\top, \psi_\sigma(z,x))^\top (\psi_\beta(z,x)^\top, \psi_\sigma(z,x))\, P(dz)\, \delta(dx), \\
M_\beta(\psi,\delta) &:= \mathcal{I}(\delta) := \int x\, x^\top\, \delta(dx), \\
M_\sigma(\psi,\delta) &:= \int (z\, \psi(z,x) - 1)\, (z^2 - 1)\, P(dz)\, \delta(dx), \\
Q_\beta(\psi,\delta) &:= \int \psi(z,x)^2\, x\, x^\top\, P(dz)\, \delta(dx), \\
Q_\sigma(\psi,\delta) &:= \int (z\, \psi(z,x) - 1)^2\, P(dz)\, \delta(dx).
\end{aligned}$$

Under the assumptions (1.4) - (1.7) we have

$$M(\psi,\delta) = \frac{1}{\sigma^2} \begin{pmatrix} M_\beta(\psi,\delta) & 0 \\ 0 & M_\sigma(\psi,\delta) \end{pmatrix}$$

and

$$Q(\psi,\delta) = \frac{1}{\sigma^2} \begin{pmatrix} Q_\beta(\psi,\delta) & 0 \\ 0 & Q_\sigma(\psi,\delta) \end{pmatrix}.$$

Then, under suitable regularity conditions, the asymptotic covariance matrix of a simultaneous M estimator $(\widehat{\beta}, \widehat{\sigma})$ has the form (see Hampel et al. 1986)

$$\sigma^2 \begin{pmatrix} V_\beta(\psi,\delta) & 0 \\ 0 & V_\sigma(\psi,\delta) \end{pmatrix},$$

where

$$V_\beta(\psi,\delta) := M_\beta(\psi,\delta)^{-1}\, Q_\beta(\psi,\delta)\, M_\beta(\psi,\delta)^{-1} = \mathcal{I}(\delta)^{-1}\, Q_\beta(\psi,\delta)\, \mathcal{I}(\delta)^{-1}$$

and

$$V_\sigma(\psi,\delta) := Q_\sigma(\psi,\delta)/M_\sigma(\psi,\delta)^2.$$

The influence function of the M estimator is (see Hampel et al. 1986)

$$\begin{pmatrix} IF_\beta(z,x,\psi,\delta) \\ IF_\sigma(z,x,\psi,\delta) \end{pmatrix},$$

where

$$IF_\beta(z,x,\psi,\delta) := M_\beta(\psi,\delta)^{-1}\psi(z,x)\,x = \mathcal{I}(\delta)^{-1}\,x\,\psi(z,x)$$

and

$$IF_\sigma(z,x,\psi,\delta) := (z\,\psi(z,x)-1)/M_\sigma(\psi,\delta).$$

The aim is to characterize score functions ψ which simultaneously minimizes

$$trV_\beta(\psi,\delta) \text{ and } V_\sigma(\psi,\delta)$$

under the side conditions that

$$\sup_{z,x}|IF_\beta(z,x,\psi,\delta)| \le b_\beta \text{ and } \sup_{z,x}|IF_\sigma(z,x,\psi,\delta)| \le b_\sigma, \tag{1.8}$$

i.e. that the gross-error-sensitivities are bounded.

If we are only interested in optimal estimation of the regression parameter β and do not require the consistency of σ (condition (1.5)), then we have the optimality problem of minimizing $\mathrm{tr}V_\beta(\psi,\delta)$ under the side conditions $\sup_{z,x}|IF_\beta(z,x,\psi,\delta)| \le b_\beta$ and (1.4). The optimality criterium as well as the side conditions are given by functions which are convex in the score functions ψ. Solutions of this optimality problem were characterized by Hampel (1978), Krasker (1980), Bickel (1981, 1984), Huber (1983), Rieder (1985, 1987, 1994). However, the so-called Hampel-Krasker solutions are given only implicitely. Explicite solutions were derived in Kurotschka and Müller (1992) and Müller (1994, 1997) for special experimental design situations.

For estimating simultaneously the regression and scale parameter, up to now no optimal score functions were derived. The reason is that $V_\sigma(\psi,\delta)$ is not anymore convex in ψ, so that the results of convex analysis cannot be applied as in the pure regression case. However we solve special variants of the above optimization problem and derive score functions ψ which are optimal in a modified sense. In Section 2 we modify the problem so that a reduction to the one-dimensional location-scale problem is possible. Section 3 contains the main results. We characterize there optimal score functions for simultaneous location and scale estimation. A detailed study, which includes comparisons of variances and gross-error-sensitivities is limited to two important special subfamilies of solutions for the modified optimization problem. Section 4 summarizes applications of the main results to the regression models.

2. Reduction to the location-scale case

In the case of estimating simultaneously the regression and scale parameter, we have the problem that $V_\sigma(\psi,\delta)$ is not convex in ψ, so that results of convex analysis cannot be applied directly. One way to solve the optimization problem nevertheless is to fix the quantity $M_\sigma(\psi,\delta)$ by setting $M_\sigma(\psi,\delta) = m$, where m is some given real number. Then we have only to minimize $\mathrm{tr}V_\beta(\psi,\delta)$ and $Q_\sigma(\psi,\delta)$ under the side conditions. The consistency requirement (1.5) for σ and the condition $M_\sigma(\psi,\delta) = m$ imply that $|IF_\beta(z,x,\psi,\delta)| \le b_\beta$ if and only if $|\psi(z,x)| \le$

$b_\beta/|\mathcal{I}(\delta)^{-1}x|$ as well as $|IF_\sigma(z,x,\psi,\delta)| \le b_\sigma$ if and only if $|z\psi(z,x)-1| \le b_\sigma m$. Since additionally $trV_\beta(\psi,\delta) = \int \psi(z,x)^2 \, |\mathcal{I}(\delta)^{-1}x|^2 \, P(dz) \, \delta(dx)$ we can regard this minimization problem for each design point x separately. This means that a score function ψ minimizes $trV_\beta(\psi,\delta)$ and $Q_\sigma(\psi,\delta)$ under the conditions (1.8) if it minimizes $\int \psi(z,x)^2 \, P(dz)$ and $\int (z\psi(z,x)-1)^2 \, P(dz)$ under the conditions $\sup_z |\psi(z,x)| \le b_\beta/|\mathcal{I}(\delta)^{-1}x|$ and $\sup_z |z\psi(z,x)-1| \le b_\sigma m$ for all $x \in \mathrm{supp}(\delta)$. We will see in Section 3 that the minimization at each point x can be carried out effectively so that we can reduce the multi-dimensional problem to the one-dimensional case.

A less formal reduction to the simple location scale estimation was carried out in Bednarski and Zontek (1996) where the primary goal was to construct robust estimators satisfying such utility conditions as reliability of variance assessment and computation feasibility. The guiding methodology stemmed there from the Fréchet differentiability requirement for M-functionals. It narrowed the class of usable score functions ψ considerably. The class turned out to contain optimal score functions for both location and scale in a sense.

3. Location-scale estimation

In the location-scale case the variances are

$$V_\beta(\psi) := Q_\beta(\psi) = \int \psi(z)^2 \, P(dz)$$

and

$$V_\sigma(\psi) := Q_\sigma(\psi)/M_\sigma(\psi)^2,$$

where

$$M_\sigma(\psi) := \int (z\psi(z)-1)\,(z^2-1)\,P(dz) = \int z^3\,\psi(z)\,P(dz) - 1$$

and

$$Q_\sigma(\psi) := \int (z\psi(z)-1)^2 \, P(dz) = \int z^2\psi(z)^2 \, P(dz) - 1.$$

The influence functions for location and scale are respectively

$$IF_\beta(z,\psi) := \psi(z)$$

and

$$IF_\sigma(z,\psi) := (z\,\psi(z)-1)/M_\sigma(\psi).$$

Since, as it was mentioned earlier, $V_\sigma(\psi)$ is not convex in ψ but $Q_\sigma(\psi)$ is convex we will regard the problem of minimizing $Q_\beta(\psi)$ and $Q_\sigma(\psi)$ under the side conditions

(3.9) ψ is an odd function (consistency for location),

(3.10) $\int z\,\psi(z)\,P(dz) = 1$ (consistency for scale),

(3.11) $|\psi(z)| \le l$ for all $z \in \Re$,

(3.12) $|z\,\psi(z) - 1| \le s$ for all $z \in \Re$.

Conditions (3.11) and (3.12) ensure that the influence functions for location and scale are bounded. While Condition (3.11) ensures that l itself is the bound for the location gross-error-sensitivity, Condition (3.12) provides that $s/M_\sigma(\psi)$ is the bound for the scale gross-error-sensitivity. Since later $M_\sigma(\psi)$ is fixed as a constant m, the quantity $M_\sigma(\psi)$ is not used in Condition (3.12).

Let Ψ be the family of functions satisfying (3.9), (3.11) and (3.12).

Proposition 1. Assume that $\psi_* \in \Psi$ satisfies the consistency condition (3.10) and $\int z\,\psi_*(z)\,P(dz)$ is in the interior of

$$\left\{\int z\,\psi(z)\,P(dz);\ \psi \in \Psi\right\}.$$

Then ψ_* minimizes $Q_\beta(\psi)$ within Ψ if and only if there exists $b \in I\!R$ so that ψ_* minimizes $\int (b\,z - \psi(z))^2\,P(dz)$ within Ψ.

Proof. Notice that Ψ is a convex set, Q_β is a convex function and the side condition (3.10) is given by a linear functional of ψ. The optimization problem is then convex and well-posed. Therefore by the Lagrange principle a score function ψ_* minimizes $Q_\sigma(\psi)$ under the side conditions if and only if there exists a multiplier $b \in I\!R$ so that

$$\begin{aligned}(3.13)\qquad & Q_\beta(\psi) + 2b(1 - \int z\,\psi(z)\,P(dz)) \\ \geq\ & Q_\beta(\psi_*) + 2b(1 - \int z\,\psi_*(z)\,P(dz)) \\ =\ & Q_\beta(\psi_*)\end{aligned}$$

for all $\psi \in \Psi$ (see e.g. Neustadt 1976, p. 81-84). Because of

$$\begin{aligned}& \int (b\,z - \psi(z))^2\,P(dz) \\ =\ & b^2 - 2b + 2b(1 - \int z\,\psi(z)\,P(dz)) + Q_\beta(\psi)\end{aligned}$$

the minimization of (3.13) within Ψ is equivalent to the minimization of $\int (b\,z - \psi(z))^2\,P(dz)$ within Ψ. $\square$

If ψ_* minimizes $\int (b\,z - \psi(z))^2\,P(dz)$ within Ψ, then ψ_* is of the form $\psi_* = bz$ as long as it is possible under the side conditions. Since we have $Q_\beta(\psi) = V_\beta(\psi)$, the Proposition 1 concerns as well the minimization of the variance $V_\beta(\psi)$ of location estimation. For scale estimation $Q_\sigma(\psi)$ and $V_\sigma(\psi)$ do not coincide. To minimize $V_\sigma(\psi)$ we fix the denominator's value. Hence, we use the additional side condition $M_\sigma(\psi) = m$, or equivalently

$$\int z^3 \psi(z)\,P(dz) = m + 1. \qquad (3.14)$$

Then score functions minimizing $Q_\sigma(\psi)$ ($V_\sigma(\psi)$) under these side conditions depend on m and are denoted by ψ_m^*. If ψ_m^* is known for all possible m then the optimal solution can be found by minimizing $Q_\sigma(\psi_m^*)/m^2$ with respect to m.

Proposition 2. Assume that $\psi_m^* \in \Psi$ satisfies the consistency condition (3.10) and the condition (3.14) for a given m and $(\int z\psi_m^*(z)\,P(dz), \int z^3\psi_m^*(z)\,P(dz))$ lies in the interior of

$$\left\{ \left(\int z\,\psi(z)\,P(dz), \int z^3\psi(z)\,P(dz) \right);\ \psi \in \Psi \right\}.$$

Then ψ_m^* minimizes $Q_\sigma(\psi)$ within Ψ under the side conditions (3.10) and (3.14) if and only if there exists $a, b \in I\!R$ so that ψ_m^* minimizes $\int(a + bz^2 - z\psi(z))^2\,P(dz)$ within the class Ψ .

Proof. Q_σ is a convex function and the side conditions (3.10) and (3.14) are given by a linear functional of ψ. As in the proof of Proposition 1, the Lagrange principle provides that a score function ψ_m^* minimizes $Q_\sigma(\psi)$ under the side conditions if and only if there exists Lagrange multipliers $a, b \in I\!R$ so that

$$\begin{aligned}(3.15)\quad & Q_\sigma(\psi) + 2(a-1)(1 - \int z\,\psi(z)\,P(dz)) + 2b(m + 1 - \int z^3\psi(z)\,P(dz)) \\ \geq\ & Q_\sigma(\psi_m^*) + 2(a-1)(1 - \int z\,\psi_m^*(z)\,P(dz)) + 2b(m + 1 - \int z^3\psi_m^*(z)\,P(dz)) \\ =\ & Q_\sigma(\psi_m^*)\end{aligned}$$

for all $\psi \in \Psi$. Again the minimization of (3.15) within Ψ is equivalent to minimization of $\int(a + b\,z^2 - z\,\psi(z))^2\,P(dz)$ within Ψ. □

According to Proposition 2 the optimal score function ψ_m^* should be of the form $\psi_m^*(z) = \frac{a}{z} + bz$ as long as the inequalities in (3.11) and (3.12) are satisfied. The unknown constants a and b should be determined by the consistency condition (3.10) and the condition for the denominator (3.14). This however means that a and b are determined by two nonlinear equations of complicated form.

Here we shall consider only the two limiting cases, where either a or b is 0. This means that we shall study the behavior of score functions ψ_0 and ψ_1 of the form $\psi_0 = \frac{a}{z}$ and $\psi_1(z) = b\,z$, respectively, when the inequalities (3.11) and (3.12) are satisfied. For these score functions the constants a and b are given as roots of one-dimensional functions which satisfy some convexity properties so that the constants can be calculated by Newton's method. Moreover, these score functions are solutions of modified optimization problems: ψ_0 minimizes $Q_\sigma(\psi)$ under the conditions (3.9) - (3.12), while ψ_1 minimizes $Q_\sigma(\psi)$ under the conditions (3.9) - (3.12), (3.14), where in condition (3.14) a special value m is used. Let us note that ψ_1 minimizes $Q_\beta(\psi) = V_\beta(\psi)$ under the conditions (3.9) - (3.12).

We start considerations with score functions ψ_0. The Proposition 2 implies that ψ_0 has to be of the form

$$\psi_0(z) := \begin{cases} l\, sgn(z) & \text{for } |z| \le \frac{a}{l}, \\ \frac{a}{z} & \text{for } |z| > \frac{a}{l}, \end{cases}$$

where a satisfies the consistency condition (3.9) and so it solves the equation

$$2\Phi'(0) - 2\Phi'(a/l) + 2a/l(1 - \Phi(a/l)) - \frac{1}{l} = 0.$$

Putting $t_0 = a/l$ we get that t_0 is a positive root of

$$g_0(t) := 2\Phi'(0) - 2\Phi'(t) + 2t(1 - \Phi(t)) - \frac{1}{l},$$

where Φ is the standard normal distribution function. The function g_0 is strictly increasing and convex with $g_0(0) < 0$. Since the boundary condition (3.12) implies $a \le s+1$, that is $t_0 \le \frac{s+1}{l}$, a root t_0 with this property exists if and only if the inequality $g_0(\frac{s+1}{l}) \ge 0$ is satisfied. We state this fact in the following theorem.

Theorem 1. If $s \ge 1$ and $g_0(\frac{s+1}{l}) \ge 0$, then ψ_0 with $a = t_0 l$ and $g_0(t_0) = 0$ minimizes $Q_\sigma(\psi)$ under the side conditions (3.9) - (3.12). The only root $t_0 > 0$ of $g_0(t)$ can be determined by Newton's method.

Proof. Under the condition (3.10) we have

$$\int (a - z\,\psi(z))^2\, P(dz) = a^2 - 2a + 1 + Q_\sigma(\psi)$$

so that the minimization of $Q_\sigma(\psi)$ under the conditions (3.9) - (3.12) is equivalent to the minimization of $\int (a - z\,\psi(z))^2\, P(dz)$ under the same conditions. □

Note that for ψ_0 and the root t_0 we have the following expressions providing the variances of the location and scale estimator:

$$\begin{aligned} Q_\beta(\psi_0) &= V_\beta(\psi_0) = l^2\,[2\Phi(t_0) - 1] + 2a^2\left[\Phi(t_0) + t_0^{-1}\Phi'(t_0) - 1\right], \\ Q_\sigma(\psi_0) &= l^2\left[2\Phi(t_0) - 1 - 2t_0\Phi'(t_0)\right] + 2a^2\,[1 - \Phi(t_0)] - 1, \\ M_\sigma(\psi_0) &= 2a\left[1 - \Phi(t_0) + t_0\Phi'(t_0)\right] - 1 + 2l\left[2\Phi'(0) - 2\Phi'(t_0) - t_0^2\Phi'(t)\right]. \end{aligned}$$

Consider now score functions of the form $\psi_1(z) = b\,z$. The boundary conditions (3.11) and (3.12) yield two possibilities:

$$\psi_1^1(z) := \begin{cases} b_1\, z & \text{for } |z| < \frac{l}{b_1}, \\ l\,\text{sgn}(z) & \text{for } \frac{l}{b_1} \le |z| \le \frac{s+1}{l}, \\ \frac{s+1}{z} & \text{for } |z| > \frac{s+1}{l}, \end{cases}$$

or

$$\psi_1^2(z) := \begin{cases} b_2\, z & \text{for } |z| < \sqrt{\frac{s+1}{b_2}}, \\ \frac{s+1}{z} & \text{for } |z| \ge \sqrt{\frac{s+1}{b_2}}, \end{cases},$$

where the coefficients b_1 and b_2 are to be determined by the consistency condition (3.10).

We shall study first ψ_1^1. By (3.10) we need to find the root t_1 of

$$g_1(t) := 2\Phi(t) - 1 - 2\,t\,\Phi'\left(\frac{s+1}{l}\right) + 2\,t\,\frac{s+1}{l}\left(1 - \Phi\left(\frac{s+1}{l}\right)\right) - \frac{t}{l},$$

where $t_1 = l/b_1$. Since by formula for $\psi_1^1(t)$, necessarily $\frac{l}{b_1} < \frac{s+1}{l}$, the root has to be in the interval $(0, (s+1)/l]$.

Notice that by $g_1(0) = 0$ and the concavity of g_1 ($g_1''(t) = -2t\,\Phi'(t) < 0$) the existence of the root is equivalent to $g_1'(0) > 0$ and $g_1(\frac{s+1}{l}) < 0$. The two inequalities can be written as

$$-\frac{2\Phi(d)-1}{d} > -2\Phi'(d) + 2\,d\,(1-\Phi(d)) - \frac{1}{l} > -2\,\Phi'(0).$$

for $d = \frac{s+1}{l}$.

Discussion concerning ψ_1^2 is similar. The consistency requirement (3.10) is reduced to the equation $g_2(t_2) = 0$, where

$$g_2(t) := 2\Phi(t) - 1 - 2\,t\,\Phi'(t) + 2\,t^2\,(1-\Phi(t)) - \frac{t^2}{s+1}$$

and $t_2 = \sqrt{\frac{s+1}{b_2}}$. By (3.11) we obtain $t_2 \leq \frac{l}{b_2}$. Again, by elementary calculus $g_2(0) = 0$ and $\lim_{t\to\infty} g_2(t) = -\infty$. Moreover, for $s \geq 1$, the derivative

$$g_2'(t) = 2t\left(2 - 2\Phi(t) - \frac{1}{s+1}\right)$$

is positive in the vicinity of 0 and it has a unique root $t^* = \Phi^{-1}\left(1 - \frac{1}{2(s+1)}\right)$. Therefore g_2 has a unique root in the interval $[t^*, \infty)$ and moreover the formula

$$g_2''(t) = \frac{g_2'(t)}{t} - 4t\,\Phi'(t)\ ,$$

implies that the function g_2 is there concave. Since the root t_2 of $g_2(t)$ must satisfy $t_2 \leq \frac{l}{b_2}$, or equivalently $t_2 \geq \frac{s+1}{l}$, $g_2(d)$ has to be nonnegative for $d = \frac{s+1}{l}$, which can be written as

$$-\frac{2\Phi(d)-1}{d} \leq -2\Phi'(d) + 2\,d\,(1-\Phi(d)) - \frac{1}{l},$$

where $d = (s+1)/l$.

The above considerations provide that we can define the following score function:

$$\psi_1(z) := \begin{cases} \psi_1^1(z) & \text{if } -\frac{2\Phi(d)-1}{d} > -2\Phi'(d) + 2\,d\,(1-\Phi(d)) - \frac{1}{l}, \\ \psi_1^2(z) & \text{if } -\frac{2\Phi(d)-1}{d} \leq -2\Phi'(d) + 2\,d\,(1-\Phi(d)) - \frac{1}{l}, \end{cases}$$

where b_1 and b_2 in ψ_1^1 and ψ_1^2 are defined as above. The following theorem summarizes the above argumentation.

Theorem 2. If $s \geq 1$, $g_0(\frac{s+1}{l}) > 0$, and $m = \int z^3 \psi_1(z)\,P(dz) - 1$, then ψ_1 minimizes $Q_\beta(\psi)$ under the side conditions (3.9) - (3.12) and $Q_\sigma(\psi)$ under the side conditions (3.9) - (3.12) and $\int z^3 \psi(z)\,P(dz) = m + 1$. The unique roots t_1 and t_2 appearing in ψ_1 can be calculated by Newton's method.

Proof. Under the side condition $\int z^3\psi(z)\,P(dz) = \int z^3\psi_1(z)\,P(dz)$ we have

$$\begin{aligned}&\int (b\,z^2 - z\,\psi(z))^2\,P(dz)\\ &= \; b^2 \int z^4\,P(dz) - 2\,b \int z^3\,\psi_1(z)\,P(dz)) + Q_\sigma(\psi) + 1.\end{aligned}$$

Setting $b = b_1$ and $b = b_2$ the minimization of $Q_\sigma(\psi)$ under the side conditions (3.9) - (3.12) and $\int z^3\psi(z)\,P(dz) = \int z^3\psi_1(z)\,P(dz)$ is equivalent to the minimization of $\int(b\,z^2 - z\,\psi(z))^2\,P(dz)$ under the side conditions. The assertion for $Q_\beta(\psi)$ follows from Proposition 1. □

The following expressions provide the variances of the location and scale estimator for ψ_1^1 and for ψ_1^2

$$\begin{aligned}Q_\beta(\psi_1^1) &= V_\beta(\psi_1^1) = b_1^2\left[2\Phi(t_1) - 1 - 2t_1\Phi'(t_1)\right]\\ &\quad +2l^2\left[\Phi(d) - \Phi(t_1)\right] + 2(s+1)^2\left[d^{-1}\Phi'(d) + \Phi(d) - 1\right],\\ Q_\sigma(\psi_1^1) &= b_1^2\left[6\Phi(t_1) - 3 - 6t_1\Phi'(t_1) - 2t_1^3\Phi'(t_1)\right]\\ &\quad + 2l^2\left[\Phi(d) - d\Phi'(d) - \Phi(t_1) + t_1\Phi'(t_1)\right] + 2(s+1)^2\left[1 - \Phi(d)\right] - 1,\\ M_\sigma(\psi_1^1) &= b_1\left[6\Phi(t_1) - 3 - 6t_1\Phi'(t_1) - 2t_1^3\Phi'(t_1)\right]\\ &\quad + 2l\left[2\Phi'(t_1) + t_1^2\Phi'(t_1) - 2\Phi'(d) - d^2\Phi'(d)\right]\\ &\quad + 2(s+1)\left[1 - \Phi(d) + d\Phi'(d)\right] - 1,\end{aligned}$$

$$\begin{aligned}Q_\beta(\psi_1^2) &= V_\beta(\psi_1^2) = b_2^2\left[2\Phi(t_2) - 1 - 2t_2\Phi'(t_2)\right] + 2(s+1)^2\left[t_2^{-1}\Phi'(t_2) + \Phi(t_2) - 1\right],\\ Q_\sigma(\psi_1^2) &= b_2^2\left[6\Phi(t_2) - 3 - 6t_2\Phi'(t_2) - 2t_2^3\Phi'(t_2)\right]\\ &\quad + 2(s+1)^2\left[1 - \Phi(t_2)\right] - 1,\\ M_\sigma(\psi_1^2) &= b_2\left[6\Phi(t_2) - 3 - 6t_2\Phi'(t_2) - 2t_2^3\Phi'(t_2)\right]\\ &\quad + 2(s+1)\left[1 - \Phi(t_2) + t_2\Phi'(t_2)\right] - 1,\end{aligned}$$

where t_1 and t_2 are roots of functions g_1 and g_2, respectively, and $b_1 = l/t_1$, $b_2 = (s+1)/t_2^2$.

Theorems 1 and 2 require that the bounds l and s for the gross-error-sensitivity have to satisfy $s \geq 1$ and

$$g_0\left(\frac{s+1}{l}\right) \geq 0, \tag{3.16}$$

where in Theorem 2 even strict inequality was required. In Figure 1 the set of all (l, s) satisfying (3.16) is presented. Since g_0 has a positive root $t_0 < \infty$ if and only if its limit in infinity is positive, that is $l > \sqrt{\pi/2}$, and condition (3.16) implies $t_0 \leq (s+1)/l$ (see the above considerations concerning ψ_0) a lower bound for l is $\sqrt{\pi/2} = 1.253314$. This is the same lower bound for the location gross-error-sensitivity as in the pure location case (see e.g. Huber 1981, Hampel et al. 1986, Rieder 1985, 1987, 1994). Depending on l the lower bound for s is $s = \max\{t_0 l - 1, 1\}$,

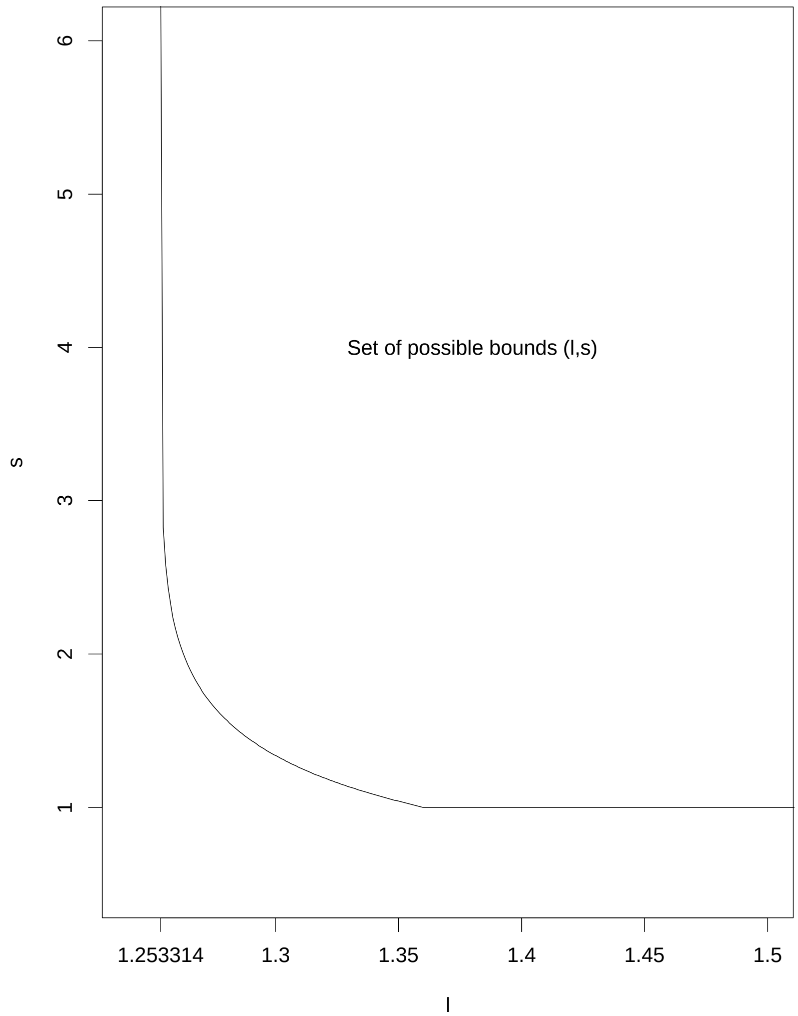

FIGURE 1. Set of possible bounds (l,s).

which is the boundary of the set in Figure 1. As long as we have $s + 1 = t_0 l$ there exists only one score function ψ satisfying the conditions (3.9) - (3.12). This score function is ψ_0 with $a = s + 1$. Hence, for location-scale estimation, ψ_0 plays a similar role as the median for pure location estimation.

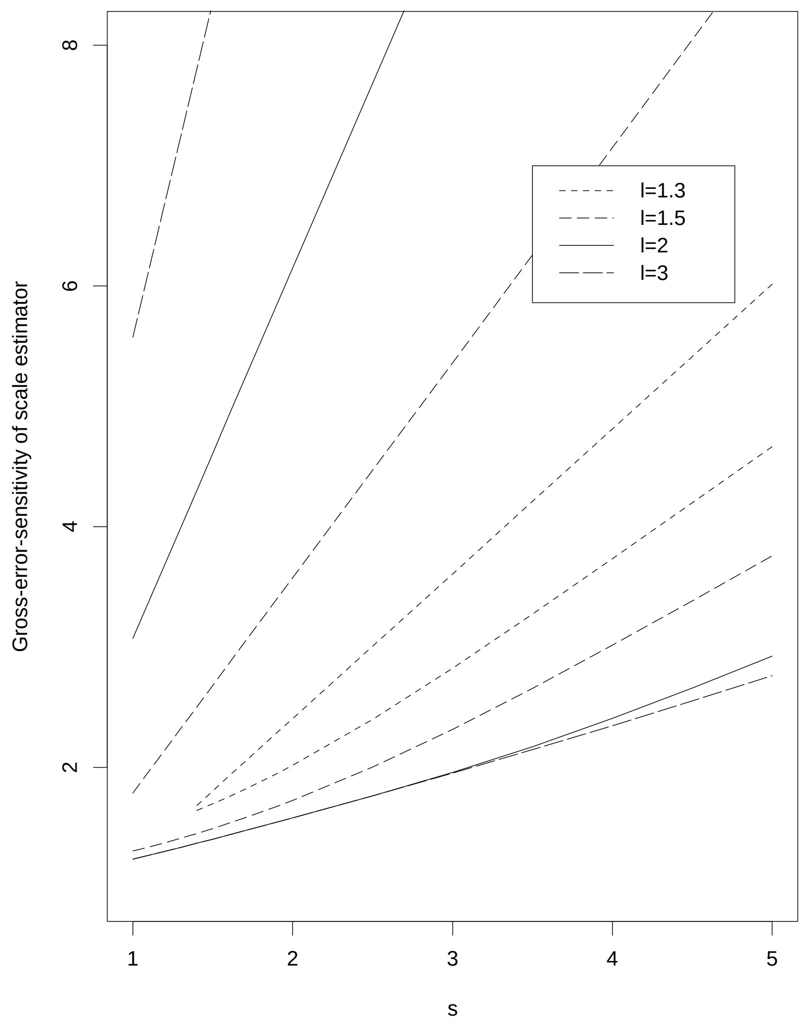

FIGURE 2. Scale gross-error-sensitivity against s: For every l, the upper curves correspond to ψ_0, while the lower ones correspond to ψ_1.

Since s is only the bound for $|z\,\psi(z) - 1|$ we have to divide s by $M_\sigma(\psi)$ to get the scale gross-error-sensitivity. Figure 2 shows how the scale gross-error-sensitivities of ψ_0 and ψ_1 depend on s. In general the scale gross-error-sensitivities of ψ_0 are much larger than that of ψ_1. Only for the case $s + 1 = t_0 l$, which is satisfied by l close to the lower bound $\sqrt{\pi/2}$, the scale gross-error-sensitivities coincide. This is shown as well by Figure 3 where the scale gross-error-sensitivities are plotted against l, the location gross-error-sensitivity.

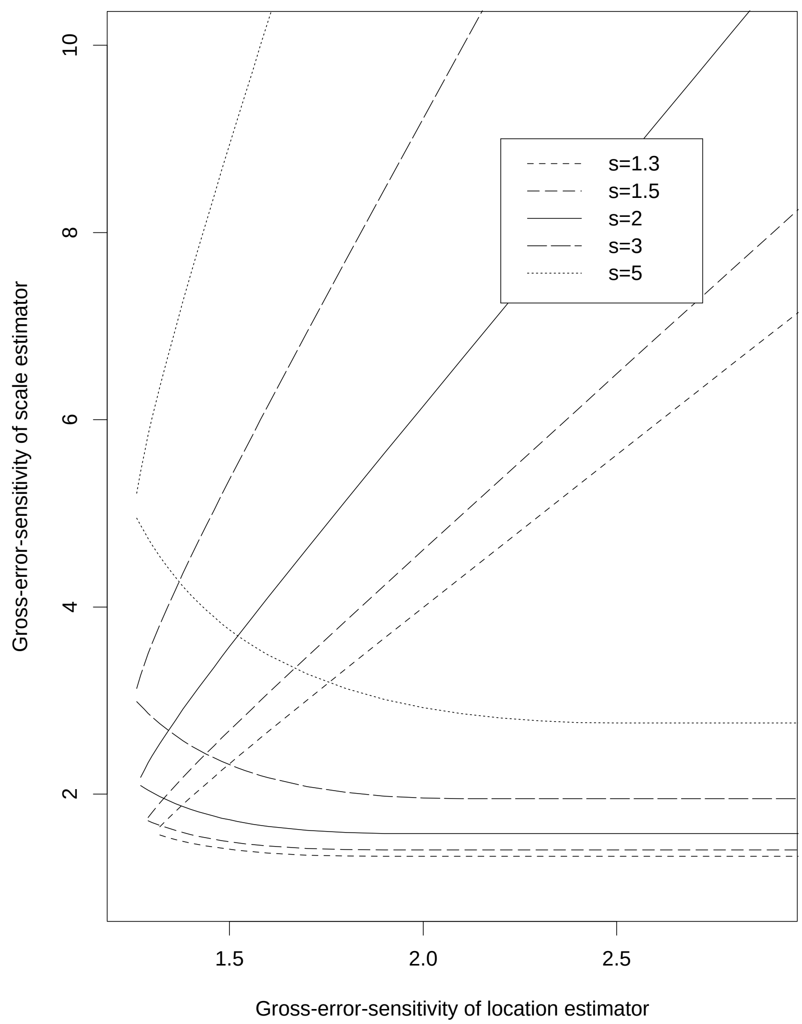

FIGURE 3. Scale gross-error-sensitivity against l: For every s, the upper curves correspond to ψ_0, while the lower ones correspond to ψ_1.

The effect that the gross error sensitivities of ψ_0 are larger than those of ψ_1 is caused by the fact that $M_\sigma(\psi_0)$ goes to zero while $M_\sigma(\psi_1)$ goes to 2, for $l \to \infty$. Since $Q_\sigma(\psi_0)$ and $M_\sigma(\psi_0)$ go to zero with the same rate the scale variance of ψ_0 goes to ∞ as $l \to \infty$. Hence the scale variances of ψ_0 are also much larger than those of ψ_1. See Figure 4. Again, only if l is approaching $\sqrt{\pi/2}$ then the differences disappear. Figure 4 shows as well that the scale variances of ψ_1 are approaching fastly the lower bound of 0.5, which is the variance of the maximum likelihood estimator for scale given by $\psi(z) = z$. This means that although ψ_1 is only the solution of minimizing $Q_\sigma(\psi)$ under the additional condition $\int z^3 \, \psi(z) \, P(dz) = \int z^3 \, \psi_1(z) \, P(dz)$ it cannot be very far away from the optimal solution $\psi^*_{m^*}$ with $m^* = \arg\min_m Q_\sigma(\psi^*_m)/m^2$.

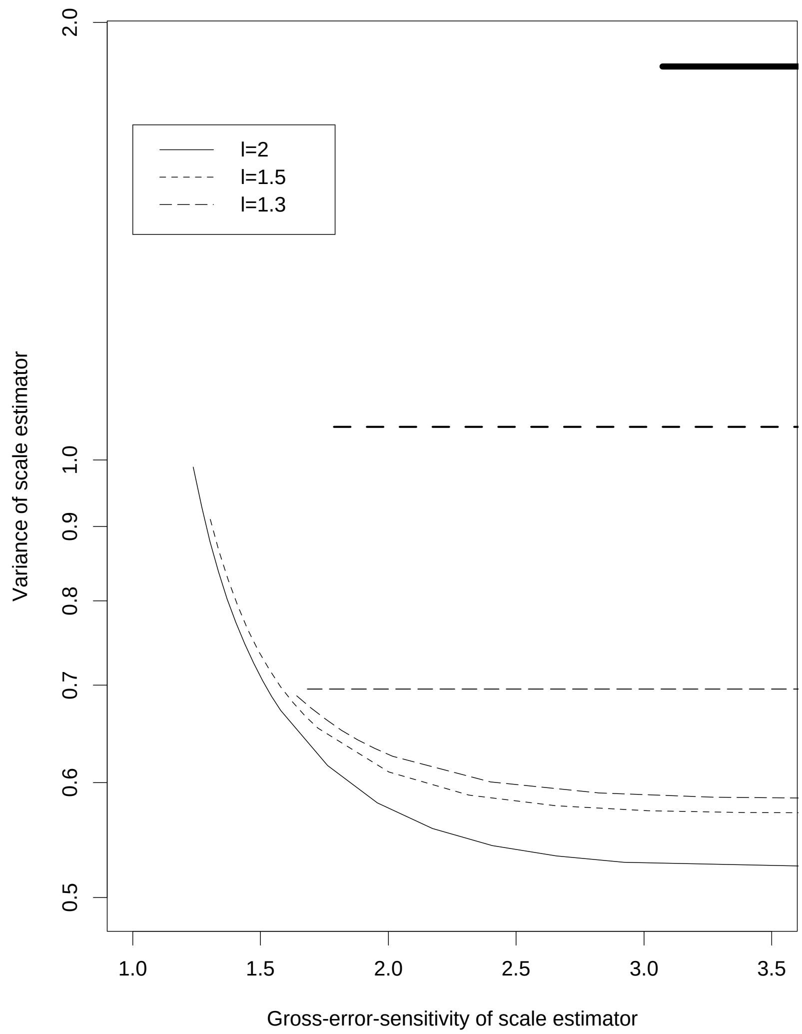

FIGURE 4. Scale variance against scale gross-error-sensitivity: For every l, the upper curves correspond to ψ_0, while the lower ones correspond to ψ_1.

Since ψ_1 minimizes $Q_\beta(\psi)$ under the necessary side conditions it is not surprising that the location variances of ψ_1 are much smaller than those of ψ_0 as shown in Figure 5. The curves of the location variances of ψ_0 appear as one straight line because ψ_0 is independent of s. The bound s determines only from which l the line starts. Again, the location variances of ψ_1 are approaching fastly the optimal value of 1 which is the variance of the maximum likelihood estimator for location given by $\psi(z) = z$.

Conclusion. The above results show that ψ_1 provides M estimators of location and scale which have a smaller asymptotic variance than the M estimators given by ψ_0 and this holds for all bounds on the gross-error-sensitivities. Even for small bounds on the gross-error-sensitivities, the variances given by ψ_1 are close to the variances of the nonrobust maximum likelihood

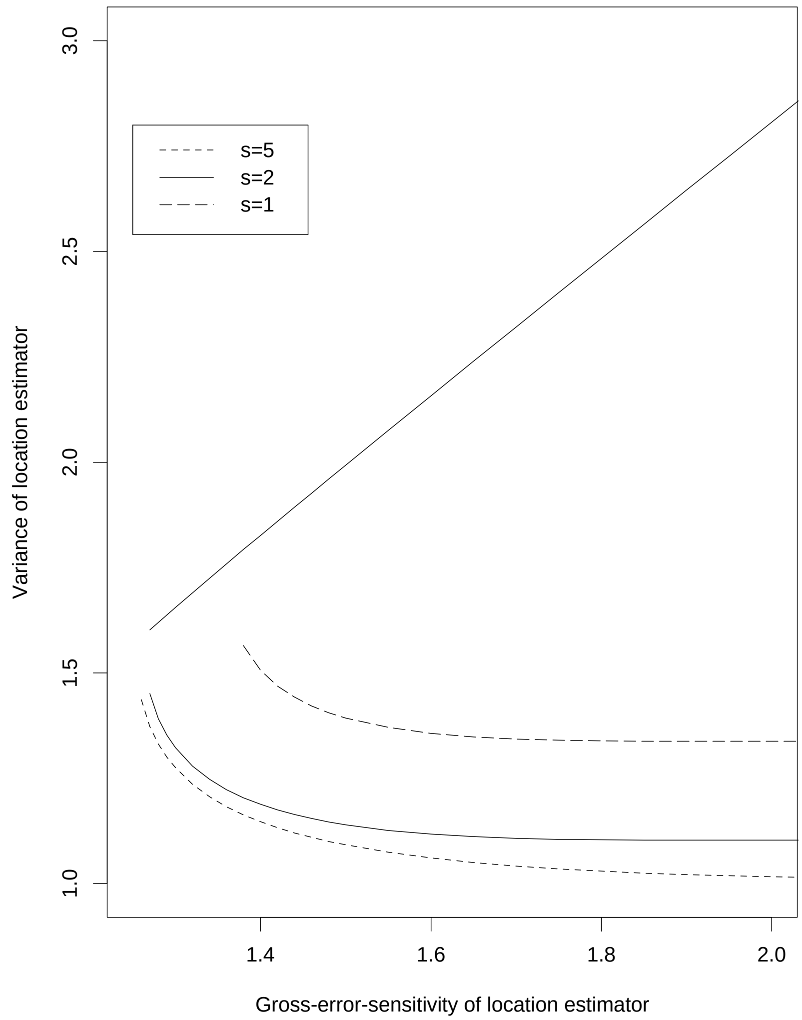

FIGURE 5. Location variance against location gross-error-sensitivity: For every s, the upper curves correspond to ψ_0, while the lower ones correspond to ψ_1.

estimators. The estimators given by ψ_1 are not optimal in the general setting of minimizing the variances under bounded influence functions for location and scale simultaneously. Nevertheless, we can expect that their behavior is not very different from the true optimal estimators which are difficult to calculate.

4. Applications to linear models

As it was pointed out in Section 2 the minimization of the variances under the conditions of consistency and bounded influence functions can be done for each design point x separately. This means that for each x we can use the results of Section 3 for determining $\psi(\cdot, x)$. In particular, we propose to use score functions of the form given by ψ_1.

Thereby, the bound l for the location gross-error-sensitivity becomes the bound $l(x) = b_\beta/|\mathcal{I}(\delta)^{-1}x|$, where b_β is the bound for the regression gross-error-sensitivity. This implies that the tuning constants b_1 and b_2 of the optimal score functions becomes constants $b_1(x)$ and $b_2(x)$ which depend on the design points x only via $|\mathcal{I}(\delta)^{-1}x|$. The same dependence of the tuning constant on x appears in the pure regression case although the optimal score functions in that case are of different type, namely of Huber type, i.e. $\psi(z,x) = z \min\left\{1, \frac{c(x)}{|z|}\right\}$ (see Kurotschka and Müller 1992, Müller 1994, 1997).

As was shown in Section 3 the gross-error-sensitivity bounds $(l(x), s)$ have to satisfy condition (3.16). This in particular means that $l(\cdot)$ have to be bounded from below by $\sqrt{\pi/2}$ which is only possible if $|\mathcal{I}(\delta)^{-1}x|$ is bounded from above. This implies that the regression gross-error-sensitivity b_β has to satisfy

$$b_\beta \geq \max\left\{|\mathcal{I}(\delta)^{-1}x|\sqrt{\frac{\pi}{2}};\ x \in \operatorname{supp}(\delta)\right\}.$$

The same condition is needed if the problem of estimating only the regression parameter is considered (see Kurotschka and Müller 1992, Müller 1997, p.114).

For linear models with qualitative factors as the ANOVA models $|\mathcal{I}(\delta)^{-1}x|$ is always bounded. However, for models with quantitative factors as polynomial regression models, $|\mathcal{I}(\delta)^{-1}x|$ is only bounded if the experimental conditions x are lying in a set $\mathcal{X}$ which is bounded. If x is assumed to be random, for example with normal distribution, then of course $\mathcal{X}$ is not bounded. However, in designed experiments the experimental conditions are given by the experimenter and then the experimental region $\mathcal{X}$ is bounded. Moreover, often experimental designs δ can be used so that $|\mathcal{I}(\delta)^{-1}x|$ has the same value for all $x \in \operatorname{supp}(\delta)$. Due to a equivalence theorem of Kiefer and Wolfowitz (1959, 1960) this is the case if δ is a A-optimal design, i.e. if δ minimizes the trace of $\mathcal{I}(\overline{\delta})$ for all $\overline{\delta}$ in a convex class of designs (see also Müller 1994, 1997, p.15).

References

(1) Bednarski, T. and Zontek, S. (1994). A note on robust estimation of parameters in mixed unbalanced models. In *Proceedings of the International Conference on Linear Statistical Inference LINSTAT'93*, eds. T. Caliński, R. Kala, Kluwer Academic Publishers, Dordrecht, 87-95.

(2) Bednarski, T. and Zontek, S. (1996). Robust estimation of parameters in a mixed unbalanced model. *Ann. Statist.* **24**, 1493-1510.

(3) Bickel, P.J. (1981). Quelque aspects de la statistique robuste. In École d'Été de Probabilités de St. Flour. *Springer Lecture Notes in Math.* **876**, 1-72.

(4) Bickel, P.J. (1984). Robust regression based on infinitesimal neighbourhoods. *Ann. Statist.* **12**, 1349-1368.

(5) Hampel, F.R. (1978). Optimally bounding the gross-error-sensitivity and the influence of position in factor space. *Proceedings of the ASA Statistical Computing Section*, ASA, Washington, D.C., 59-64.

(6) Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics - The Approach Based on Influence Functions.* John Wiley, New York.

(7) Huber, P.J. (1981). *Robust Statistics.* John Wiley, New York.

(8) Huber, P.J. (1983). Minimax aspects of bounded-influence regression (with discussion). *J. Amer. Statist. Assoc.* **78**, 66-80.

(9) Huggins, R.M. (1993). On the robust analysis of variance components models for pedigree data. *Austral. J. Statist.* **35**, 43-57.

(10) Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.* **30**, 271-294.

(11) Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Can. J. Math.* **12**, 363-366.

(12) Krasker, W.S. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48**, 1333-1346.

(13) Kurotschka, V. and Müller, Ch.H. (1992). Optimum robust estimation of linear aspects in conditionally contaminated linear models. *Ann. Statist.* **20**, 331-350.

(14) Müller, Ch.H. (1994). Optimal designs for robust estimation in conditionally contaminated linear models. *J. Statist. Plann. Inference.* **38**, 125-140.

(15) Müller, C.H. (1997). *Robust Planning and Analysis of Experiments.* Lecture Notes in Statistics **124**, Springer, New York.

(16) Neustadt, L.W. (1976). *Optimization - A Theory of Necessary Conditions.* Princeton University Press.

(17) Rieder, H. (1985). Robust estimation of functionals. *Technical Report.* Universität Bayreuth.

(18) Rieder, H. (1987). Robust regression estimators and their least favorable contamination curves. *Stat. Decis.* **5**, 307-336.

(19) Rieder, H. (1994). *Robust Asymptotic Statistics*, Springer, New York.

Carl von Ossietzky University Oldenburg
Department 6 of Mathematics
Postfach 2503
D-26111 Oldenburg
Germany
e-mail: mueller@math.uni-oldenburg.de

Polish Academy of Sciences
Institute of Mathematics
ul.Kopernika 18, 51-617 Wroclaw
Poland
e-mail: tabe@mat.impan.gov.pl