

# Optimal designs for robust estimation in conditionally contaminated linear models

Christine Müller

*Freie Universität Berlin, Berlin, Germany*

Received 9 October 1991; revised manuscript received 30 October 1992

Recommended by F. Pukelsheim

## Abstract

A generalization of the classical A-optimality criterion for designs is derived by defining optimal designs for those asymptotically linear (AL-) estimators which are optimally robust in the sense of minimizing the trace of the covariance matrix under bounded bias in an infinitesimal conditionally contaminated normal linear model. It is proved that the A-optimal designs are also optimal in the generalized, robust sense. For the proof special characterizations of the influence functions of the optimal robust AL-estimators for A-optimal designs and designs with finite support based on characterizations in Hampel (*Proc. ASA Stat. Comp. Section*, 1978), Krasker (*Econometrica* **48**, 1980) and Kurotschka and Müller (*Ann. Statist.* **20**, 1992) are investigated. In particular a very simple form of optimal influence functions at A-optimal designs is derived. This provides the side result that for estimating the whole parameter vector at A-optimal designs the Huber and the Hampel–Krasker estimator coincide.

*AMS Subject Classification:* Primary 62F35, 62K05; secondary 62J05.

*Key words and phrases:* Linear model; infinitesimal contamination neighbourhood; robust estimation; Hampel–Krasker estimator; Huber estimator; linear aspect; optimal design; A-optimality.

## 1. Introduction and main result

We consider a general linear model  $X_{nN}(t_n) = a(t_n)^T \beta + Z_{nN}$ ,  $n = 1, \dots, N$ ,  $N \in \mathbb{N}$ , where

$t_1, \dots, t_N \in T$  are different experimental conditions,  
 $a: T \rightarrow \mathbb{R}^r$  the vector of known ‘regression’ functions,  
 $\beta \in \mathbb{R}^r$  the vector of unknown parameters,  
 $Z_{nN}$  the error variables,  
 $X_{nN}(t_n)$  the observations at  $t_n$ .

*Correspondence to:* C. Müller, Freie Universität Berlin, Fachbereich Mathematik WE01, Institut für Mathematik, Arnimallee 2–6, 1000 Berlin (West) 33, Germany.

Here  $t_1, \dots, t_N$  are realizations of independent random elements  $T_1, \dots, T_N$  which are identically distributed according to a given probability measure  $\delta$  on  $(T, \mathcal{F})$  where  $\mathcal{F}$  is a suitable  $\sigma$ -algebra on the experimental region  $T$ . In classical models,  $(Z_{1N}, T_1), \dots, (Z_{NN}, T_N)$  are independent and identically distributed according to  $P_\delta := n_{(0,1)} \otimes \delta$  where  $n_{(0,1)}$  is the standard normal probability measure on  $\mathbb{R}$ . To describe contamination of the error variables  $Z_{nN}$ , we take  $(Z_{1N}, T_1), \dots, (Z_{NN}, T_N)$  to be independent and identically distributed according to a contaminated distribution  $P_{N,\delta} := k_N \otimes \delta$  where the Markov kernel  $k_N$  has the form

$$k_N(dx, t) = (1 - N^{-1/2} \varepsilon(t)) n_{(0,1)}(dx) + N^{-1/2} \varepsilon(t) k(dx, t)$$

for almost all  $N \in \mathbb{N}$ . The function  $\varepsilon: T \rightarrow [0, \infty)$  determines conditionally on  $t \in T$  the probabilities of contamination, and the Markov kernel  $k$  determines conditionally on  $t \in T$  the form of the contamination distribution. On the average the proportion of contamination shall be less than a positive constant  $R$ , i.e.,  $\int \varepsilon d\delta \leq R$ , providing a conditionally contaminated linear model, also studied by Bickel (1984), Rieder (1985, 1987) and Kurotschka and Müller (1992). For contiguity properties the conditional contamination distribution  $k$  is taken to have the form  $k(dx, t) = f(x, t) n_{(0,1)}(dx)$ , with  $\text{ess sup} |\varepsilon(t)(f(x, t) - 1)| < \infty$ . The distributions  $((P_{N,\delta})_{N \in \mathbb{N}}^N)$  of  $((Z_{1N}, T_1), \dots, (Z_{NN}, T_N))_{N \in \mathbb{N}}$  define a class  $\mathbb{U}_{R,\delta}$  which states an infinitesimal contamination neighbourhood around the normal model  $((P_\delta)_{N \in \mathbb{N}}^N)$ , with radius  $R$ .

If the experimenter has the option between several distributions  $\delta$  then  $\delta$  can be subjected to optimization criteria and will be interpreted as a design, because  $\delta$  provides random designs  $(T_1, \dots, T_N)$  which have asymptotic properties like a deterministic design  $\delta$ , see Kurotschka and Müller (1992). Here such ‘designs’  $\delta$  are characterized which are optimal for robust estimation of a linear aspect  $\varphi(\beta) = C\beta$  ( $C \in \mathbb{R}^{s \times r}$ ) of the unknown parameter vector  $\beta$  in the conditionally contaminated linear model.

To estimate  $\beta$  with observations  $X_{1N}(T_1), \dots, X_{NN}(T_N)$  in the contaminated linear model, Bickel (1981, 1984) and Rieder (1985, 1987) investigated the asymptotic behaviour of asymptotically linear (AL-) estimator sequences, define optimal robust AL-estimators as those which minimize the trace of the asymptotic covariance matrix among all AL-estimators, with asymptotic bias bounded by some given constant, and characterize those optimal robust AL-estimators. In particular if the bias bound is greater than the minimum possible bias, the optimal robust AL-estimators have influence functions which Hampel (1978) and Krasker (1980) have proposed. These results can be extended to estimation of a linear aspect  $\varphi(\beta) = C\beta \in \mathbb{R}^s$  as Kurotschka and Müller (1992) have shown:

AL-estimators  $(\hat{\varphi}_N)_{N \in \mathbb{N}}$  for  $\varphi(\beta) = C\beta$  with influence function  $\psi$  are estimators which fulfill

$$N^{1/2} \left[ \hat{\varphi}_N - \varphi(\beta) - N^{-1} \sum_{n=1}^N \psi(X_{nN}(T_n) - a(T_n)^T \beta, T_n) \right] \xrightarrow{N \rightarrow \infty} 0,$$

in probability  $((P_\delta)^N)_{N \in \mathbb{N}}$  where  $\psi$  is an element of the set

$$\begin{aligned} \Psi(\delta) := \{ \psi : \mathbb{R} \times T \rightarrow \mathbb{R}^s; \int |\psi(x, t)|^2 P_\delta(dx, dt) < \infty, \\ \int \psi(x, \cdot) n_{(0, 1)}(dx) = 0, \int \psi(x, t) a(t)^T x P_\delta(dx, dt) = C \}. \end{aligned}$$

Asymptotically AL-estimators for  $\varphi(\beta)$  are normally distributed, i.e.,

$$L[N^{1/2}(\hat{\varphi}_N - \varphi(\beta) - \int \psi dP_{N, \delta}) | (P_{N, \delta})^N] \xrightarrow{N \rightarrow \infty} N(0, \int \psi \psi^T dP_\delta),$$

for all  $((P_{N, \delta})^N)_{N \in \mathbb{N}} \in \mathcal{U}_{R, \delta}$ , with maximum asymptotic bias

$$\sup \left\{ \limsup_{N \rightarrow \infty} |N^{1/2} \int \psi dP_{N, \delta}|; ((P_{N, \delta})^N)_{N \in \mathbb{N}} \in \mathcal{U}_{R, \delta} \right\} = R \|\psi\|_\delta,$$

where  $\|\psi\|_\delta$  denotes the  $P_\delta$ -essential supremum of  $|\psi(x, t)|$ .

The Gauss–Markov estimator is an AL-estimator with unbounded asymptotic bias because its influence function is  $\psi(x, t) = CI(\delta)^- a(t)x$ , where  $I(\delta) := \int a(t)a(t)^T \delta(dt)$  is the information matrix and  $M^- \in \mathbb{R}^n \times m$  denotes a  $g$ -inverse of  $M \in \mathbb{R}^{m \times n}$ , i.e.,  $MM^-M = M$ . For robust AL-estimators the asymptotic bias should be bounded by some given bound  $b$  (without loss of generality we set  $R = 1$ ). Optimal robust AL-estimators with bias bound  $b$  are those which have an influence function  $\psi_{b, \delta}$  solving

$$\psi_{b, \delta} = \arg \min \{ \text{tr } V(\psi, \delta); \psi \in \Psi_b(\delta) \}, \tag{1}$$

where  $\text{tr } V(\psi, \delta)$  is the trace of the asymptotic covariance matrix  $V(\psi, \delta) := \int \psi \psi^T dP_\delta$  and  $\Psi_b(\delta) := \{ \psi \in \Psi(\delta); \|\psi\|_\delta \leq b \}$ .

Hence, if  $\Delta$  is a subset of  $\Delta^*$ , the set of all distribution  $\delta$  on  $T$  at which  $\varphi(\beta)$  is identifiable i.e., there exists  $K \in \mathbb{R}^s \times r$  with  $C = KI(\delta)$ , we call a design  $\delta^*$  optimal in  $\Delta$  for robust estimation with bias bound  $b$  if and only if at  $\delta^*$  the trace of the asymptotic covariance matrix of the optimal robust AL-estimator with bias bound  $b$  is minimized among all  $\delta \in \Delta$ .

**Definition.** A design  $\delta^*$  is called optimal in  $\Delta$  for robust estimation with bias bound  $b$ , briefly called  $A(b)$ -optimal in  $\Delta$ , if and only if

$$\delta^* = \arg \min \{ \inf \{ \text{tr } V(\psi, \delta); \psi \in \Psi_b(\delta) \}; \delta \in \Delta \} \quad (\inf \phi := \infty). \tag{2}$$

Because for  $b \geq b_0(\delta) := \min \{ \|\psi\|_\delta; \psi \in \Psi(\delta) \}$  a solution of (1) always exists (see, for example, Bickel, 1984, p. 1355; or Kurotschka and Müller, 1992, Theorem 1) every solution of (2) is a solution of

$$\delta^* = \arg \min \{ \text{tr } V(\psi_{b, \delta}, \delta); \delta \in \Delta \text{ with } b_0(\delta) \leq b \}$$

and vice versa. Because for  $b = \infty$  and  $\delta \in \Delta^*$  the influence functions of the Gauss–Markov estimators  $\psi_{\infty, \delta}(x, t) := CI(\delta)^{-1}a(t)x$  are solutions of (1), the well-known classical *A-optimal designs*  $\delta_A$  which are defined by

$$\delta_A = \arg \min \{ \text{tr } CI(\delta)^{-1}C^T; \delta \in \Delta \}$$

are optimal for estimation without bias bound, i.e.,  $A(\infty)$ -optimal, so that  $A(b)$ -optimality is a generalization of  $A$ -optimality. The following theorem shows that  $A$ -optimal designs are also optimal for robust estimation with bounded bias.

**Theorem.** Let  $b_0(\Delta) := \min\{b_0(\delta); \delta \in \Delta\}$  and either

- (A)  $\Delta = \{\delta \in \Delta^*; \delta \text{ has a finite support}\}$  and  $b \geq b_0(\Delta)$ , or
- (B)  $\Delta = \Delta^*$ ,  $b = b_0(\Delta)$  and  $a: T \rightarrow \mathbb{R}^r$  is continuous, or
- (C)  $\Delta = \Delta^*$ ,  $b \geq b_0(\Delta)$ ,  $a: T \rightarrow \mathbb{R}^r$  is continuous and  $T$  compact.

If  $\Delta$  includes an  $A$ -optimal design, then  $\delta^*$  is  $A$ -optimal in  $\Delta$  if and only if  $\delta^*$  is optimal in  $\Delta$  for robust estimation with bias bound  $b$ , i.e.,  $A(b)$ -optimal in  $\Delta$ .

Essential to the proof of the theorem are calculations of the trace of the covariance matrix  $V(\psi_{b, \delta}, \delta)$ . They are based on special characterizations of the optimal robust influence functions  $\psi_{b, \delta}$  for  $A$ -optimal designs and for designs with finite support. These characterizations use the characterizations for arbitrary linear aspects  $\varphi(\beta) = C\beta$  given in Kuratschka and Müller (1992). In particular it is shown that for  $A$ -optimal designs the optimal robust influence functions have the simplest form, only depending on a single one-dimensional fixed point. This entails that at  $A$ -optimal designs, estimating the whole parameter vector  $\beta$  with bias bound greater than the minimum bias bound, i.e.,  $b > b_0(\delta) := \min\{\|\psi\|_\delta; \psi \in \mathcal{P}(\delta)\}$ , the Hampel–Krusker influence functions of Hampel (1978) and Krasker (1980) coincide with the influence functions of Huber (1973) (see also Huber, 1981; and Hampel et al., 1986). Hence Huber’s estimators appear to be optimal in this context for special designs. For  $b = b_0(\delta)$  and  $A$ -optimal designs there is also a relation to the result of Ronchetti and Rousseeuw (1985) concerning most  $B_u$ -robust estimators.

Because the characterizations of optimal robust influence functions may be of interest in their own right especially for  $A$ -optimal designs they are given in Section 2. In Section 3 the theorem is proved under the various conditions (A), (B) and (C). In Sections 4 and 5 two examples, the quadratic regression and a two-way classification model, are given.

## 2. Characterization of optimal robust estimators

Let  $\Phi$  denote the distribution function of the standard normal distribution and let  $y(t)$  be a positive fixed point of

$$f_t(y) := (2\Phi(by) - 1) |CI(\delta)^{-1}a(t)|^{-1}.$$

For designs  $\delta$  which have a finite support with linearly independent regressors Kurotschka and Müller (1992, Theorem 2) show that  $\psi_{b,\delta}^*$  with

$$\psi_{b,\delta}^*(x, t) := \begin{cases} CI(\delta)^- a(t) \operatorname{sgn}(x) (\frac{1}{2}\pi)^{1/2} & \text{for } b = |CI(\delta)^- a(t)| (\frac{1}{2}\pi)^{1/2}, \\ CI(\delta)^- a(t) \frac{\operatorname{sgn}(x) \min\{|x|, by(t)\}}{|CI(\delta)^- a(t)|y(t)} & \text{for } b > |CI(\delta)^- a(t)| (\frac{1}{2}\pi)^{1/2} > 0, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

is a solution of (1). Thus it is the influence function of an optimal robust AL-estimator with bias bound  $b$  at  $\delta$ . Lemma 1 below shows that influence functions of the form (3) are also optimal for arbitrary A-optimal designs. (Note that there exist A-optimal designs with a support on which the regressors are not linearly independent, see for instance the example in Section 4.) The support  $\operatorname{supp}(\delta)$  of  $\delta$  is the smallest closed set  $D \subset T$  with  $\delta(D) = 1$ . Moreover we define  $\Delta_D := \{\delta \in \Delta^*; \operatorname{supp}(\delta) = D\}$  for  $D \subset T$ .

**Lemma 1.** *Let  $a: D \rightarrow \mathbb{R}^r$  be continuous and  $\delta \in \Delta_D$ .*

(a) *Then we have  $b_0(\delta) \geq (\operatorname{tr} CI(\delta)^- C^T \frac{1}{2}\pi)^{1/2}$ , with equality if and only if  $\delta$  is A-optimal in  $\Delta_D$ .*

(b) *If  $\delta$  is A-optimal in  $\Delta_D$  and  $b \geq b_0(\delta)$ , then  $\psi_{b,\delta}^*$  is a solution of (1).*

**Proof.** Because of

$$b_0(\delta) = \max \{ \operatorname{tr} QC^T (\int |Qa(t)| \delta(dt))^{-1} (\frac{1}{2}\pi)^{1/2}; Q \in \mathbb{R}^{s \times r} \text{ with } QC^T \neq 0 \}$$

(see Kurotschka and Müller, 1992, Lemma 2; and for  $C = E_{r \times r}$  the identity matrix, Rieder, 1985, Theorem 3.7), one gets for  $Q = CI(\delta)^-$  and  $\psi_{0,\delta} := Qa(t) \operatorname{sgn}(x) (\frac{1}{2}\pi)^{1/2}$

$$\begin{aligned} \|\psi_{0,\delta}\|_\delta &\geq b_0(\delta) \geq \operatorname{tr} QC^T (\int |Qa(t)| \delta(dt))^{-1} (\frac{1}{2}\pi)^{1/2} \\ &\geq \operatorname{tr} QC^T (\int |Qa(t)|^2 \delta(dt))^{-1/2} (\frac{1}{2}\pi)^{1/2} = (\operatorname{tr} CI(\delta)^- C^T \frac{1}{2}\pi)^{1/2}, \end{aligned}$$

with equality if and only if  $|CI(\delta)^- a(t)|^2 = \operatorname{tr} CI(\delta)^- C^T$   $\delta$ -a.e., i.e.,  $\sup_{t \in D} |CI(\delta)^- a(t)|^2 = \operatorname{tr} CI(\delta)^- C^T$ . Since the *equivalence theorem for linear optimality* is also true for estimating  $\varphi(\beta) = C\beta$  in  $\Delta_D$  (see, for example, Federov, 1972, p. 125), assertion (a) follows. Moreover if  $\delta$  is A-optimal in  $\Delta_D$ , then for  $b = b_0(\delta)$  with  $Q_1 = CI(\delta)^-$  we have

$$\psi_{b,\delta}^*(x, t) = \psi_{0,\delta}(x, t) = Q_1 a(t) |Q_1 a(t)|^{-1} b \operatorname{sgn}(x) \quad P_\delta\text{-a.e.},$$

where

$$C = \int Q_1 a(t) a(t)^T |Q_1 a(t)|^{-1} b(2/\pi)^{1/2} \delta(dt).$$

Furthermore for  $b > b_0(\delta)$  with  $Q_1 = CI(\delta)^{-1} (|CI(\delta)^{-1} a(t)| y(t))^{-1}$  for some  $t$  fulfilling  $|CI(\delta)^{-1} a(t)|^2 = \text{tr } CI(\delta)^{-1} C^T$  we have

$$\psi_{b,\delta}^*(x, t) = Q_1 a(t) \text{sgn}(x) \min\{|x|, b|Q_1 a(t)|^{-1}\} \quad P_\delta\text{-a.e.},$$

where

$$C = \int Q_1 a(t) a(t)^T (2\Phi(b|Q_1 a(t)|^{-1}) - 1) \delta(dt).$$

Hence Theorem 1 in Kurotschka and Müller (1992) proves assertion (b) (see also for  $C = E_{r \times r}$ , Bickel, 1981, 1984; and Rieder, 1985, 1987).  $\square$

**Remark 1.** Because A-optimal designs are characterized by  $|CI(\delta)^{-1} a(t)|^2 = \text{tr } CI(\delta)^{-1} C^T$   $\delta$ -a.e., form (3) of the optimal influence functions reduces for A-optimal designs to

$$\psi_{b,\delta}^*(x, t) := \begin{cases} CI(\delta)^{-1} a(t) \text{sgn}(x) (\frac{1}{2}\pi)^{1/2} & \text{for } b = (\text{tr } CI(\delta)^{-1} C^T \frac{1}{2}\pi)^{1/2}, \\ CI(\delta)^{-1} a(t) \frac{\text{sgn}(x) \min\{|x|, b y_b\}}{\sqrt{\text{tr } CI(\delta)^{-1} C^T y_b}} & \text{for } b > (\text{tr } CI(\delta)^{-1} C^T \frac{1}{2}\pi)^{1/2}, \end{cases}$$

where  $y_b$  is a positive fixed point of

$$f_b(y) := (2\Phi(by) - 1) (\text{tr } CI(\delta)^{-1} C^T)^{-1/2}.$$

Therefore, for A-optimal designs  $\delta$  and  $b > b_0(\delta)$ , only one one-dimensional fixed point  $y_b$  for  $\psi_{b,\delta}^*$  must be determined, and  $\psi_{b,\delta}^*(x, t)$  is of the form  $Qa(t)\eta(x)$  with  $\eta(x) = \text{sgn}(x) \cdot \min\{|x|, c\}$  which Huber (1973) has proposed for estimating the whole parameter vector  $\beta$ . For estimating  $\beta$  with bias bound  $b = b_0(\delta)$ , Theorem 2 of Ronchetti and Rousseeuw (1985) can be deduced from Lemma 1, replacing  $\delta$  and  $a(t)$  by the modified design  $\delta^m(dt) := (\int |a(t)| \delta(dt))^{-1} |a(t)| \delta(dt)$  and the modified regressors  $a^m(t) := |a(t)|^{-1} a(t)$ , so that  $\delta^m$  is A-optimal for the modified regressors  $a^m(t)$  (see Müller, 1992, p. 113).

**Remark 2.** If  $F: [0, \infty)^3 \rightarrow \mathbb{R}$  is defined by

$$F(a, b, y) := a(2\Phi(by) - 1) - y,$$

and  $f$  implicitly given by  $F(a, b, f(a, b)) = 0$  and  $f(a, b) > 0$  then we have  $y(t) = f(|CI(\delta)^{-1} a(t)|^{-1}, b)$ . Further, with

$$g(y) := \int \min\{|x|, y\}^2 n_{(0,1)}(dx)$$

and

$$v_b(a) := \begin{cases} \frac{1}{2}\pi & \text{for } a = \pi/(2b^2), \\ g(bf(a^{1/2}, b)) a f(a^{1/2}, b)^{-2} & \text{for } a > \pi/(2b^2), \end{cases}$$

we get, for influence functions  $\psi_{b,\delta}^*$  of form (3) at designs with finite support  $D$ ,

$$\text{tr } V(\psi_{b,\delta}^*, \delta) = \sum_{t \in D} |CI(\delta)^{-1} a(t)|^2 \delta(\{t\}) v_b(|CI(\delta)^{-1} a(t)|^{-2}). \quad (4)$$

For designs  $\delta$  which are A-optimal in  $\Delta_D$  this reduces to

$$\operatorname{tr} V(\psi_{b,\delta}^*, \delta) = \operatorname{tr} CI(\delta)^{-1} C^T v_b((\operatorname{tr} CI(\delta)^{-1} C^T)^{-1}), \quad (5)$$

where

$$\lim_{b \rightarrow \infty} v_b((\operatorname{tr} CI(\delta)^{-1} C^T)^{-1}) = 1, \quad \lim_{b \downarrow b_0(\delta)} v_b((\operatorname{tr} CI(\delta)^{-1} C^T)^{-1}) = \frac{1}{2}\pi.$$

The latter is seen as follows. Setting

$$r(y) := g(y)(2\Phi(y) - 1)^{-2},$$

and

$$b(\varepsilon) := \frac{1}{\varepsilon} \Phi^{-1}\left(\frac{\varepsilon + a}{2a}\right),$$

we have  $v_b(a^2) = r(bf(a, b))$ ,  $F(a, b(\varepsilon), \varepsilon) = 0$  and, with the rule of L'Hôpital,

$$\lim_{y \downarrow 0} r(y) = \frac{1}{2}\pi \quad \text{and} \quad \lim_{\varepsilon \downarrow 0} b(\varepsilon) = \frac{1}{a} \sqrt{\frac{\pi}{2}}.$$

In particular this provides

$$\lim_{b \downarrow a^{-1}\sqrt{\pi/2}} f(a, b) = \lim_{\varepsilon \downarrow 0} f(a, b(\varepsilon)) = \lim_{\varepsilon \downarrow 0} \varepsilon = 0.$$

Therefore, and because  $f(a, b)$  is strictly increasing in  $b$  ( $\partial/\partial b f(a, b) > 0$ ), we get

$$\lim_{b \downarrow \sqrt{\pi/(2a)}} v_b(a) = \frac{1}{2}\pi \quad \text{and} \quad \lim_{b \rightarrow \infty} v_b(a) = 1.$$

Setting  $a(\varepsilon) := \varepsilon(2\Phi(b\varepsilon) - 1)^{-1}$ , we can see similarly that  $v_b(\cdot)$  is continuous on  $[\pi/(2b^2), \infty)$  with

$$\lim_{a \rightarrow \infty} v_b(a) = 1.$$

**Remark 3.** If for finite support  $D = \{\tau_1, \dots, \tau_I\}$  the regressors  $a(\tau_1), \dots, a(\tau_I)$  are not linearly independent, i.e.,  $p = \operatorname{rg}(A_D) < I = \operatorname{card}(D)$  where  $A_D := (a(\tau_1) | \dots | a(\tau_I))^T$ , then there exists  $B_D = (b_1 | \dots | b_I)^T \in \mathbb{R}^{I \times I-p}$  with  $\operatorname{rg}(A_D | B_D) = I$  and  $A_D^T B_D = 0_{r \times I-p}$ . Defining

$$b(t) := 0 \in \mathbb{R}^{I-p} \quad \text{for } t \in T/D,$$

$$b(t) := b_i \in \mathbb{R}^{I-p} \quad \text{for } t = \tau_i, \quad i = 1, \dots, I,$$

we can regard the extended model  $X^*(t) = (a(t)^T | b(t)^T) \beta^* + Z^*$  with  $\beta^* \in \mathbb{R}^{r+I-p}$  instead of the model  $X(t) = a(t)^T \beta + Z$ , and the aspect  $\varphi_G(\beta^*) = (C | G) \beta^* \in \mathbb{R}^s$  with  $G \in \mathbb{R}^{s \times I-p}$  instead of  $\varphi(\beta) = C\beta \in \mathbb{R}^s$ . In the extended model the regressors  $(a(\tau_1)^T | b(\tau_1)^T)^T, \dots, (a(\tau_I)^T | b(\tau_I)^T)^T$  are linearly independent, whence Theorem 2

of Kurotschka and Müller (1992) can be applied. For application of that theorem define

$$J(\delta) := \int \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} (a(t)^T | b(t)^T) \delta(dt),$$

$$\Psi_G(\delta) := \{\psi \in \Psi(\delta); \int \psi(x, t) b(t)^T x P_\delta(dx, dt) = G\},$$

$$b_0(\delta, G) := \min \{\|\psi\|_\delta; \psi \in \Psi_G(\delta)\},$$

$$\gamma(t, G) := (C|G)J(\delta)^{-1} \begin{pmatrix} a(t) \\ b(t) \end{pmatrix},$$

and

$$\psi_{G, b, \delta}^*(x, t) := \begin{cases} \gamma(t, G) \operatorname{sgn}(x) (\frac{1}{2}\pi)^{1/2} & \text{for } b = |\gamma(t, G)| (\frac{1}{2}\pi)^{1/2}, \\ \gamma(t, G) \frac{\operatorname{sgn}(x) \min\{|x|, b y_G(t)\}}{|\gamma(t, G)| y_G(t)} & \text{for } b > |\gamma(t, G)| (\frac{1}{2}\pi)^{1/2} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where  $y_G(t)$  is a positive fixed point of

$$f_{t, G}(y) := (2\Phi(by) - 1) |\gamma(t, G)|^{-1}.$$

Then the influence function  $\psi_{G, b, \delta}^*$  solves

$$\psi_{G, b, \delta}^* = \arg \min \{\operatorname{tr} V(\psi, \delta); \psi \in \Psi_G(\delta) \text{ with } \|\psi\|_\delta \leq b\}.$$

Since

$$\operatorname{tr} V(\psi_{G, b, \delta}^*, \delta) = \sum_{t \in D} |\gamma(t, G)|^2 \delta(\{t\}) v_b(|\gamma(t, G)|^{-2})$$

is continuous, as a function of  $G$  on the compact set

$$\{G \in \mathbb{R}^{s \times I - p}; b_0(\delta, G) \leq b\}$$

(see Remark 2 for continuity of  $v_b$ ), and because of

$$\Psi(\delta) = \bigcup \{\Psi_G(\delta); G \in \mathbb{R}^{s \times I - p}\},$$

and

$$\begin{aligned} & \min \{\operatorname{tr} V(\psi, \delta); \psi \in \Psi_b(\delta)\} \\ &= \min \{\operatorname{tr} V(\psi_{G, b, \delta}^*, \delta); G \in \mathbb{R}^{s \times I - p} \text{ with } b_0(\delta, G) \leq b\}, \end{aligned}$$

we get the following lemma:

**Lemma 2.** *If  $\delta \in \Delta_D$  and  $p = \operatorname{rg}(A_D) < I = \operatorname{card}(D) < \infty$ , then we have*

$$b_0(\delta) = \min \{\max \{|\gamma(t, G)| (\frac{1}{2}\pi)^{1/2}; t \in D\}; G \in \mathbb{R}^{s \times I - p}\},$$

and for  $b \geq b_0(\delta)$  there exists  $G(b) \in \mathbb{R}^{s \times I - p}$  such that  $\psi_{G(b), b, \delta}^*$  is a solution of (1).



### 3. Proof of the theorem

**Proof under condition (A).** If the functions  $f$ ,  $v_b$  and  $r$  are defined as in Remark 2,  $h(y) := y\Phi(-y) - \Phi'(y)$  and  $s_b(a) := v_b(a^2)$  we have the following derivatives

$$\begin{aligned} r'(y) &= [2\Phi(y) - 1]^{-3} 4h(y) [2\Phi(y) - 1 - 2y\Phi'(y)] < 0, \\ s'_b(a) &= 4h(bf(a, b))b [2\Phi(bf(a, b)) - 1]^{-1} < 0, \\ s''_b(a) &> 0 \quad \text{for } a > b^{-1}(\frac{1}{2}\pi)^{1/2}. \end{aligned}$$

This shows that  $v_b$  is strictly decreasing and strictly convex. Let  $\delta_A$  be an A-optimal design in  $\Delta$  and  $\delta \in \Delta_D$  with  $\text{rg}(A_D) = \text{card}(D)$ . Then, applying Theorem 2 of Kurotschka and Müller (1992) and Lemma 1, we get from (4) and (5) in Remark 2:

$$\begin{aligned} \min\{\text{tr } V(\psi, \delta); \psi \in \Psi_b(\delta)\} &= \text{tr } V(\psi_{b, \delta}^*, \delta) \\ &= \text{tr } CI(\delta)^{-} C^T \sum_{t \in D} |CI(\delta)^{-} a(t)|^2 \delta(\{t\}) (\text{tr } CI(\delta)^{-} C^T)^{-1} v_b(|CI(\delta)^{-} a(t)|^{-2}) \\ &\geq \text{tr } CI(\delta)^{-} C^T v_b((\text{tr } CI(\delta)^{-} C^T)^{-1} \sum_{t \in D} \delta(\{t\})) \\ &\geq \text{tr } CI(\delta_A)^{-} C^T v_b((\text{tr } CI(\delta_A)^{-} C^T)^{-1}) \\ &= \min\{\text{tr } V(\psi, \delta_A); \psi \in \Psi_b(\delta_A)\}, \end{aligned}$$

with equality if and only if  $\delta$  is also A-optimal in  $\Delta$ . Because of Remark 3 and

$$\text{tr}(C|G)J(\delta)^{-}(C|G)^T \geq \text{tr } CI(\delta)^{-} C^T,$$

where  $G$  and  $J$  are defined as in Remark 3, a similar inequality is true for  $\delta \in \Delta_D$  with  $\text{rg}(A_D) < \text{card}(D)$ :

$$\begin{aligned} \min\{\text{tr } V(\psi, \delta); \psi \in \Psi_b(\delta)\} &= \text{tr } V(\psi_{G, b, \delta}^*, \delta) \\ &\geq \text{tr}(C|G)J(\delta)^{-}(C|G)^T v_b((\text{tr}(C|G)J(\delta)^{-}(C|G)^T)^{-1}) \\ &\geq \text{tr } CI(\delta)^{-} C^T v_b((\text{tr } CI(\delta)^{-} C^T)^{-1}) \\ &\geq \min\{\text{tr } V(\psi, \delta_A); \psi \in \Psi_b(\delta_A)\}. \quad \square \end{aligned}$$

**Proof under condition (B).** If  $\delta_A$  is A-optimal in  $\Delta$ , then Lemma 1 yields  $b_0(\delta) \geq b_0(\delta_A)$  for all  $\delta \in \Delta$ , with equality if and only if  $\delta$  is also A-optimal. Hence we obtain  $b = b_0(\Delta) = b_0(\delta_A)$ , and  $\delta$  is A-optimal in  $\Delta$  if and only if  $\Psi_b(\delta) \neq \emptyset$ . Because of (5) in Remark 2 the assertion is proved.  $\square$

**Proof under condition (C).** If it can be shown that for every  $\delta \in \Delta$  there exists a  $\delta_F \in \Delta$  with finite support fulfilling  $I(\delta) = I(\delta_F)$  and

$$\min\{\text{tr } V(\psi, \delta_F); \psi \in \Psi_b(\delta_F)\} \leq \min\{\text{tr } V(\psi, \delta); \psi \in \Psi_b(\delta)\},$$

then the theorem under condition (C) follows from the theorem under condition (A). To show this let  $\delta \in \mathcal{A}$  be arbitrary. Then by Theorem 1 in Kurotschka and Müller (1992) there exists a solution  $\psi_{b,\delta}$  of (1), a disjoint partition  $T_1, \dots, T_M$  of  $T$ , and matrices  $Q_1, \dots, Q_M \in \mathbb{R}^{s \times r}$  such that, for  $m=1, \dots, M-1$ ,

$$\psi_{b,\delta}(x, t) = Q_m a(t) |Q_m a(t)|^{-1} b \operatorname{sgn}(x) \quad P_\delta\text{-a.e. on } \mathbb{R} \times T_m,$$

while for  $m=M$

$$\psi_{b,\delta}(x, t) = Q_M a(t) \operatorname{sgn}(x) \min\{|x|, b|Q_M a(t)|^{-1}\} \quad P_\delta\text{-a.e. on } \mathbb{R} \times T_M$$

$$\text{with } T_M = \bigcap_{m=1}^{M-1} \{t \in T; Q_m a(t) = 0\}.$$

Setting

$$d\delta_m = \delta(T_m)^{-1} 1_{T_m} d\delta,$$

there exists  $\delta_{n,m} \in \mathcal{A}$ ,  $n \in \mathbb{N}$ ,  $m=1, \dots, M$  with finite support such that

$$\int f d\delta_{n,m} \rightarrow \int f d\delta_m, \quad n \rightarrow \infty,$$

for all continuous and bounded functions  $f$ . This holds especially for

$$f_m(t) := Q_m a(t) a(t)^T |Q_m a(t)|^{-1},$$

$$f_*(t) := Q_M a(t) a(t)^T (2\Phi(b|Q_M a(t)|^{-1}) - 1),$$

$$f_0(t) := |Q_M a(t)|^2 g(b|Q_M a(t)|^{-1}),$$

$$f_\infty(t) := a(t) a(t)^T,$$

where  $g$  is defined as in Remark 2. Because the sets

$$S := \{(\varrho a^T, a a^T) \in \mathbb{R}^{s \times r} \times \mathbb{R}^{r \times r}; a \in a(T), \varrho \in \mathbb{R}^s \text{ with } |\varrho| \leq 1\}$$

and

$$S_* := \{(f_*(t), f_0(t), f_\infty(t)); t \in T_M\} \subset \mathbb{R}^{s \times r} \times \mathbb{R} \times \mathbb{R}^{r \times r}$$

are compact, the theorem of Carathéodory (see, for example, Silvey, 1980, p. 72) provides for  $m=1, \dots, M-1$  with  $p \leq s \cdot r + r \cdot r + 1$  the existence of

$$Q_{m1}, \dots, Q_{mp} \in \{\varrho \in \mathbb{R}^s; |\varrho| \leq 1\},$$

$$a_{m1}, \dots, a_{mp} \in a(T),$$

$$\alpha_{m1}, \dots, \alpha_{mp} \in [0, 1], \quad \sum_{i=1}^p \alpha_{mi} = 1,$$

with

$$\left( \int f_m d\delta_m, \int f_\infty d\delta_m \right) = \sum_{i=1}^p \alpha_{mi} (Q_{mi} a_{mi}^T, a_{mi} a_{mi}^T).$$

Moreover, if  $T_M \neq \emptyset$  it secures for  $m = M$  with  $q \leq s \cdot r + r \cdot r + 2$  the existence of

$$\tau_1, \dots, \tau_q \in T_M, \quad a(\tau_i) \neq a(\tau_j) \text{ for } i \neq j,$$

$$\alpha(\tau_1), \dots, \alpha(\tau_q), \quad \sum_{j=1}^q \alpha(\tau_j) = 1,$$

with

$$\left( \int f_* d\delta_M, \int f_0 d\delta_M, \int f_\infty d\delta_M \right) = \sum_{j=1}^q \alpha(\tau_j) (f_*(\tau_j), f_0(\tau_j), f_\infty(\tau_j)).$$

Setting

$$R_0 := \{a_{mi}; i = 1, \dots, p, m = 1, \dots, M-1\}$$

and

$$R_* := \{a_{mi}; i = 1, \dots, p, m = 1, \dots, M-1\} \cup a(T_\tau) \quad \text{with } T_\tau = \{\tau_1, \dots, \tau_q\},$$

there exist  $T_0 \subset T$  and  $T_* \subset T$  with  $\text{card } T_0 = \text{card } R_0$ ,  $a(T_0) = R_0$ ,  $T_\tau \subset T_*$ ,  $\text{card } T_* = \text{card } R_*$  and  $a(T_*) = R_*$ . Define for  $T_M = \emptyset$

$$\delta_F(t) := \sum_{m=1}^{M-1} \delta(T_m) \sum_{i=1}^p \alpha_{mi} 1_{\{a_{mi}=a(t)\}}(t) 1_{T_0}(t),$$

$$\psi_F(x, t) := \delta_F(t)^{-1} \sum_{m=1}^{M-1} \delta(T_m) \sum_{i=1}^p \alpha_{mi} Q_{mi} 1_{\{a_{mi}=a(t)\}}(t) b \operatorname{sgn}(x) 1_{T_0}(t),$$

and for  $T_M \neq \emptyset$

$$\delta_F(t) := \delta(T_M) \alpha(t) 1_{T_\tau}(t)$$

$$+ \sum_{m=1}^{M-1} \delta(T_m) \sum_{i=1}^p \alpha_{mi} 1_{\{a_{mi}=a(t)\}}(t) 1_{T_*}(t),$$

$$\psi_F(x, t) := \delta_F(t)^{-1} \delta(T_M) \alpha(t) Q_M a(t) \operatorname{sgn}(x) \min\{|x|, b |Q_M a(t)|^{-1}\} 1_{T_\tau}(t)$$

$$+ \delta_F(t)^{-1} \sum_{m=1}^{M-1} \delta(T_m) \sum_{i=1}^p \alpha_{mi} Q_{mi} 1_{\{a_{mi}=a(t)\}}(t) b \operatorname{sgn}(x) 1_{T_*}(t).$$

Then we have  $\delta_F(T_0) = 1$  for  $T_M = \emptyset$ ,  $\delta_F(T_*) = 1$  for  $T_M \neq \emptyset$ ,  $\psi_F \in \Psi_b(\delta_F)$  and  $I(\delta_F) = I(\delta)$ . Because of

$$(\alpha + (1-\alpha)\zeta)^2 \leq \alpha + (1-\alpha)\zeta^2 \quad \text{for all } \alpha \in [0, 1] \text{ and } \zeta \in \mathbb{R},$$

we also get

$$\min\{\operatorname{tr} V(\psi, \delta_F); \psi \in \Psi_b(\delta_F)\} \leq \operatorname{tr} V(\psi_F, \delta_F) = \int |\psi_F|^2 dP_\delta,$$

$$\leq \int |\psi_{b,\delta}|^2 dP_\delta = \min\{\operatorname{tr} V(\psi, \delta); \psi \in \Psi_b(\delta)\}. \quad \square$$

**4. Example: Quadratic regression**

For the quadratic regression model  $X(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + Z(t)$  where  $\beta = (\beta_0, \beta_1, \beta_2)^T$  and  $a(t) = (1, t, t^2)^T$  with  $t \in T = [-1, 1]$ , Kurotschka and Müller (1992) derived the influence functions of optimal robust AL-estimators with bias bound  $b$  for several designs and aspects. According to Lemma 1, the influence functions with the simplest form appear at the A-optimal designs, namely for estimating the whole parameter vector  $\beta$  at the A-optimal design  $\delta_A = \frac{1}{4}e_{-1} + \frac{1}{2}e_0 + \frac{1}{4}e_1$ , and for estimating the aspect  $\varphi(\beta) = \beta_1$  at the A-optimal design  $\delta_A^1 = \frac{1}{2}e_{-1} + \frac{1}{2}e_1$ . For the A-optimal design  $\delta_A$  for estimating  $\beta$  the trace of the asymptotic covariance matrix of the optimal AL-estimator with bias bound  $b$  is plotted for several values of the bias bound  $b$  in Figure 1. For comparison also the trace of the asymptotic covariance matrix of the optimal AL-estimator with bias bound  $b$  at the uniform design with equispaced three-point support for estimating  $\beta$ , namely at  $\delta = \frac{1}{3}e_{-1} + \frac{1}{3}e_0 + \frac{1}{3}e_1$ , is plotted. The traces of the covariance matrices are

$$\text{tr } V(\psi_{b, \delta_A}^*, \delta_A) = \begin{cases} 8 \cdot \frac{1}{2} \pi \approx 12.57 & \text{for } b = b_0(\delta_A) = 2\sqrt{\pi} \approx 3.54, \\ 8v_b(\frac{1}{8}) & \text{for } b \geq b_0(\delta_A), \\ \text{tr } I(\delta_A)^{-1} = 8 & \text{for } b = \infty, \end{cases}$$

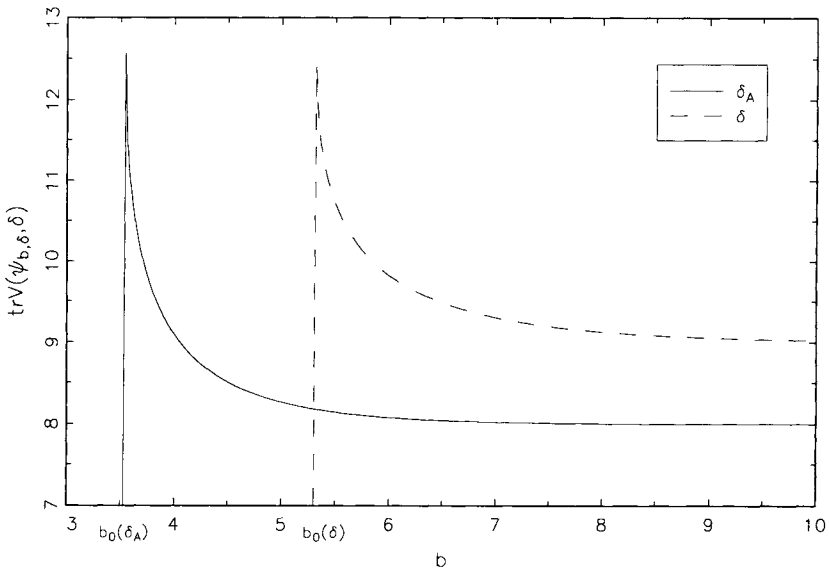


Fig. 1. The traces of the asymptotic covariance matrices of the optimal robust AL-estimators for the quadratic regression model. These are plotted versus the bias bounds  $b$  for the A-optimal design  $\delta_A$  and the uniform design  $\delta$  with equispaced three-point support. The minimum bias bounds  $b_0(\delta_A)$  and  $b_0(\delta)$  are marked with vertical lines.

and

$$\text{tr } V(\psi_{b, \delta}^*, \delta) = \begin{cases} 3v_b(\frac{2}{9}) + 6 \cdot \frac{1}{2}\pi \approx 12.43 & \text{for } b = b_0(\delta) = 3\sqrt{\pi} \approx 5.32, \\ 3v_b(\frac{2}{9}) + 6v_b(\frac{1}{18}) & \text{for } b \geq b_0(\delta), \\ \text{tr } I(\delta)^{-1} = 9 & \text{for } b = \infty. \end{cases}$$

### 5. Example: Two-way classification

Consider a two-way classification model without interactions  $X(i, j) = \alpha_i + \beta_j + Z(i, j)$ ,  $i = 1, 2, 3, j = 1, 2$ , with three levels for the first factor and two levels for the second factor. Thus we have

$$\beta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)^T, \\ a(t) = a(i, j) = (1_{\{1\}}(i), 1_{\{2\}}(i), 1_{\{3\}}(i), 1_{\{1\}}(j), 1_{\{2\}}(j))^T,$$

with

$$t = (i, j) \in T = \{1, 2, 3\} \times \{1, 2\}.$$

For estimating the contrasts  $\varphi(\beta) = (\alpha_1 - \alpha_2, \alpha_1 - \alpha_3)^T$  the A-optimal design in  $\mathcal{A}^*$  is

$$\delta_A = \frac{1}{2(2^{1/2} + 2)} \cdot (2^{1/2} e_{(1,1)} + 2^{1/2} e_{(1,2)} + e_{(2,1)} + e_{(2,2)} + e_{(3,1)} + e_{(3,2)}).$$

According to Lemma 1, at this design the influence function of the optimal robust AL-estimator with bias bound  $b$  is, for  $b = b_0(\delta_A) = (2^{1/2} + 1)\pi^{1/2} \approx 4.28$ ,

$$\psi_{b, \delta_A}^*(x, (i, j)) = \text{sgn}(x) (\frac{1}{2}\pi)^{1/2} (2^{1/2} + 1) \cdot \begin{cases} (1, 1)^T & \text{for } i = 1, \\ (-2^{1/2}, 0)^T & \text{for } i = 2, \\ (0, -2^{1/2})^T & \text{for } i = 3, \end{cases}$$

and for  $b > b_0(\delta_A)$ ,

$$\psi_{b, \delta_A}^*(x, (i, j)) = \frac{\text{sgn}(x) \min\{|x|, bu\}}{2^{1/2}u} \cdot \begin{cases} (1, 1)^T & \text{for } i = 1, \\ (-2^{1/2}, 0)^T & \text{for } i = 2, \\ (0, -2^{1/2})^T & \text{for } i = 3, \end{cases}$$

with  $0 < u = (2\Phi(bu) - 1)(2 + 2^{1/2})^{-1}$ .

The trace of the asymptotic covariance matrix of the AL-estimators with these influence functions results in

$$\text{tr } V(\psi_{b, \delta_A}^*, \delta_A) = \begin{cases} (2^{1/2} + 1)^2 \pi \approx 18.31 & \text{for } b = b_0(\delta_A), \\ (2 + 2^{1/2})^2 v_b((2 + 2^{1/2})^{-2}) & \text{for } b \geq b_0(\delta_A), \\ \text{tr } CI(\delta_A)^{-1} C^T = (2 + 2^{1/2})^2 \approx 11.66 & \text{for } b = \infty. \end{cases}$$

For several values of the bias bound  $b$  the trace of the covariance matrix is plotted in Figure 2.

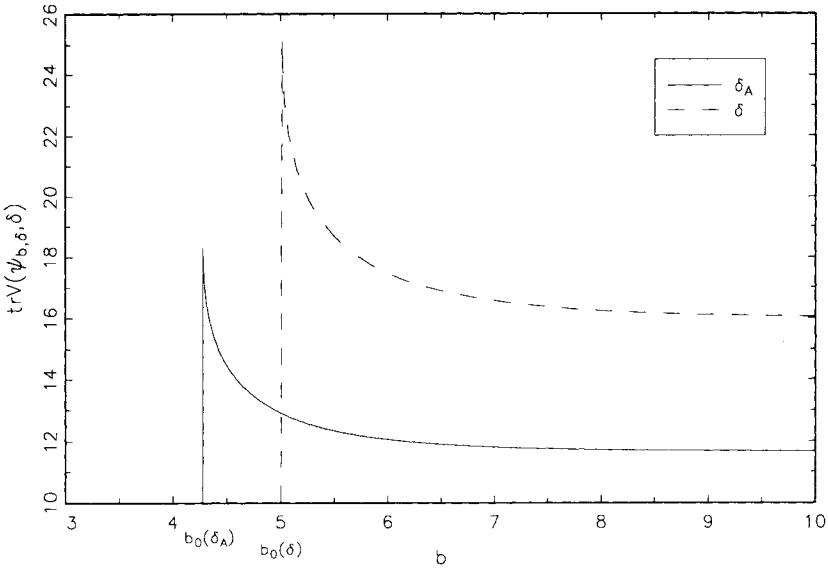


Fig. 2. The traces of the asymptotic covariance matrices of the optimal robust AL-estimators for the two-way classification model. These are plotted versus the bias bounds  $b$  for the A-optimal design  $\delta_A$  and the uniform design  $\delta$  with minimum support. The minimum bias bounds  $b_0(\delta_A)$  and  $b_0(\delta)$  are marked with vertical lines.

At the design  $\delta_A$  the regressors are not linearly independent. An uniform design with minimum support for estimating  $\varphi(\beta) = (\alpha_1 - \alpha_2, \alpha_1 - \alpha_3)^T$  so that the regressors are linearly independent is  $\delta = \frac{1}{4}e_{(1,1)} + \frac{1}{4}e_{(1,2)} + \frac{1}{4}e_{(2,1)} + \frac{1}{4}e_{(3,2)}$ . This design  $\delta$  is A-optimal in the restricted class  $\mathcal{A}_D$  with  $D = \{(1, 1), (1, 2), (2, 1), (3, 2)\}$ . Therefore we use Lemma 1 to calculate the influence function of the optimal robust AL-estimator with bias bound  $b$ . It provides for  $b = b_0(\delta) = 4(\frac{1}{2}\pi)^{1/2} \approx 5.01$ ,

$$\psi_{b,\delta}^*(x, (i, j)) = \text{sgn}(x) 4(\frac{1}{2}\pi)^{1/2} \cdot \begin{cases} (1, 0)^T & \text{for } (i, j) = (1, 1), \\ (0, 1)^T & \text{for } (i, j) = (1, 2), \\ (-1, 0)^T & \text{for } (i, j) = (2, 1), \\ (0, -1)^T & \text{for } (i, j) = (3, 2), \end{cases}$$

and for  $b > b_0(\delta)$ ,

$$\psi_{b,\delta}^*(x, (i, j)) = \frac{\text{sgn}(x) \min\{|x|, bv\}}{v} \cdot \begin{cases} (1, 0)^T & \text{for } (i, j) = (1, 1) \\ (0, 1)^T & \text{for } (i, j) = (1, 2), \\ (-1, 0)^T & \text{for } (i, j) = (2, 1), \\ (0, -1)^T & \text{for } (i, j) = (3, 2), \end{cases}$$

with  $0 < v = (2\Phi(bv) - 1)\frac{1}{4}$ . The corresponding trace of the asymptotic covariance matrix is (see Figure 2)

$$\text{tr } V(\psi_{b,\delta}^*, \delta) = \begin{cases} 8\pi \approx 25.13 & \text{for } b = b_0(\delta), \\ 16 v_b(\frac{1}{16}) & \text{for } b \geq b_0(\delta), \\ CI(\delta)^{-1} C^T = 16 & \text{for } b = \infty. \end{cases}$$

## Acknowledgement

The author would like to thank V. Kurotschka for valuable discussions and H. Rieder for providing her his latest preprints and his encouraging interest. She also owes thanks to the coordinating editor who helped to improve the presentation of the paper.

## References

- Bickel, P.J. (1981). Quelques aspects de la statistique robuste. In: *École d'Été de Probabilités de St. Flour*. Springer Lecture Notes in Math. **876**, 1–72.
- Bickel, P.J. (1984). Robust regression based on infinitesimal neighbourhoods. *Ann. Statist.* **12**, 1349–1368.
- Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.
- 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.
- Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986). *Robust Statistics — The Approach Based on Influence Functions*. John Wiley, New York.
- Huber, P.J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1**, 799–821.
- Huber, P.J. (1981). *Robust Statistics*. John Wiley, New York.
- Krasker, W.S. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48**, 1333–1346.
- Kurotschka, V. and C. Müller (1992). Optimum robust estimation of linear aspects in conditionally contaminated linear models. *Ann. Statist.* **20**, 331–350.
- Müller, C. (1992). Robust estimation with minimum bias and A-optimal designs. In: Pázman, A. and J. Volaufová, eds., *PROBASTAT'91, Proceedings of the International Conference on Probability and Mathematical Statistics, Bratislava 1991*, 109–115.
- Rieder, H. (1985). Robust estimation of functionals. Technical Report. Universität Bayreuth.
- Rieder, H. (1987). Robust regression estimators and their least favorable contamination curves. *Stat. Decis.* **5**, 307–336.
- Ronchetti, E. and P.J. Rousseeuw (1985). Change-of-variance sensitivities in regression analysis. *Z. Wahrsch. verw. Geb.* **68**, 503–519.
- Silvey, S.D. (1980). *Optimal Design — An Introduction to the Theory for Parameter Estimation*. Chapman and Hall, New York.