

OPTIMAL BIAS BOUNDS FOR ROBUST ESTIMATION IN LINEAR MODELS

CHRISTINE H. MÜLLER

Freie Universität Berlin

1. Mathematisches Institut

Arnimallee 2-6

D-14195 Berlin

Germany

Abstract. A conditionally contaminated linear model $Y(t) = x(t)'\beta + Z(t)$ is considered where the errors $Z(t)$ may have different contaminated normal distributions for different experimental conditions t . Estimating the unknown parameter β or a linear aspect $\varphi(\beta) = C\beta$ in such a model, an asymptotic bias will appear. Bounding the maximum asymptotic bias by some bias bound b , optimal robust estimators and optimal designs can be derived by minimizing the trace of the asymptotic covariance matrix (see Bickel, 1984; Rieder, 1987; Kurotschka and Müller, 1992; Müller, 1992a). While the optimal designs, which are the classical A -optimal designs, do not depend on the bias bound b , the optimal robust estimators depend strongly on b and the trace of their asymptotic covariance matrix increases if the bias bound decreases. In this paper optimal bias bounds are derived by minimizing the asymptotic mean squared error or its generalization. In particular at A -optimal designs the optimal bias bounds are easy to compute. For two examples the optimal bias bounds are given.

Key words: Conditionally contaminated linear models, Robust estimation, A -optimal designs, Optimal bias bounds, Mean squared error.

1. Introduction

A general linear model

$$Y_{nN} = x(t_{nN})'\beta + Z_{nN}, \quad n = 1, \dots, N,$$

is considered, where Y_{nN} are observations, $t_{nN} \in T$ are experimental conditions, $x : T \rightarrow \mathbb{R}^p$ is a known 'regression' function, $\beta \in \mathbb{R}^p$ is an unknown parameter vector, and Z_{nN} are error variables. In classical linear models it is assumed that the error variables Z_{1N}, \dots, Z_{NN} are independent and identically distributed. Usually it is assumed that they are normally distributed with mean 0 and known or unknown variance σ^2 , i.e.

$$\frac{1}{\sigma} Z_{nN} \sim P = \mathcal{N}(0, 1).$$

But if some outlying observations (gross errors) may appear, the normal distribution is not correct. Then, even for designed experiments, a conditionally contaminated linear model is adequate (see Bickel, 1984; Rieder, 1987; Kurotschka and Müller, 1992; Müller, 1992b). In such a model it is assumed that the error variables Z_{1N}, \dots, Z_{NN} are independent and are distributed according to a contaminated normal distribution, where the contamination may be different for different experimental

conditions, i.e.

$$\frac{1}{\sigma} Z_{nN} \sim Q_{nN}(dz) = [1 - N^{-1/2} \epsilon(t_{nN})] P(dz) + N^{-1/2} \epsilon(t_{nN}) g(z, t_{nN}) P(dz),$$

with $\sum_{n=1}^N \epsilon(t_{nN}) \leq N$ for almost all $N \in \mathbb{N}$, $\int g(z, t) P(dz) = 1$, $g(z, t) \geq 0$ for all $z \in \mathbb{R}$, $t \in T$. Thereby the markov kernel $g(\cdot, t)P$ models the form and $\epsilon(t) \geq 0$ the proportion of contamination. The set \mathcal{P} of all sequences $(Q^N = \otimes_{n=1}^N Q_{nN})_{N \in \mathbb{N}}$ defines a conditional contamination neighbourhood around the classical model $(P^N)_{N \in \mathbb{N}}$.

To estimate in this model a linear aspect $\varphi(\beta) = C\beta$, $C \in \mathbb{R}^{s \times p}$, one can use a one-step- M -estimator. An estimator $\hat{\varphi}_N : \mathbb{R}^N \times T^N \rightarrow \mathbb{R}^s$ is called a one-step- M -estimator for $\varphi(\beta) = C\beta$ with a score function $\psi : \mathbb{R} \times T \rightarrow \mathbb{R}^s$, an initial estimator $\hat{\beta}_N^0 : \mathbb{R}^N \times T^N \rightarrow \mathbb{R}^p$ for β and a variance estimator $\hat{\sigma}_N^2 : \mathbb{R}^N \times T^N \rightarrow \mathbb{R}^+$ for σ^2 , if

$$\hat{\varphi}_N(y_N, d_N) = C \hat{\beta}_N^0(y_N, d_N) + \frac{1}{N} \sum_{n=1}^N \psi \left(\frac{y_{nN} - x(t_{nN})' \hat{\beta}_N^0(y_N, d_N)}{\hat{\sigma}_N}, t_{nN} \right) \hat{\sigma}_N$$

(see Bickel, 1975; Rieder, 1985; and Müller, 1992b). Thereby the initial and the variance estimators can be robust or non-robust estimators. For example the initial estimator can be the least squares estimator or some M -estimator and the variance estimator can be the mean squared residuals or Huber's Proposal 2 (Huber, 1964). Because the asymptotic behaviour of the one-step- M -estimators does neither depend on the initial estimator nor on the variance estimator and here the robustness property is derived from the asymptotic distribution we also can use non-robust initial and variance estimators. But if the estimator should also satisfy a finite sample robustness property then the initial and variance estimator should also satisfy the finite sample robustness property.

If we assume that the design $d_N = (t_{1N}, \dots, t_{nN})$ converges to an asymptotic design measure δ in the following sense

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N e_{t_{nN}}(t) = \delta(t) \text{ for all } t \in \text{supp}(\delta), \quad (1)$$

then the one-step- M -estimator with the score function $\psi_\infty(z, t) = C I(\delta)^- x(t)z$ behaves asymptotically like the Gauss-Markov estimator for $\varphi(\beta) = C\beta$. Here e_t denotes the Dirac measure on t and $I(\delta) = \int x(t)x(t)' d\delta(t)$ is the information matrix of the design δ while $I(\delta)^-$ is a g -inverse of $I(\delta)$. In the conditionally contaminated linear model the one-step- M -estimator with the score function ψ_∞ , and therefore also the Gauss-Markov estimator, has an unbounded asymptotic bias. But robust estimators should have a bounded asymptotic bias and this is the case for all one-step- M -estimators with bounded score function ψ . This was shown by Bickel (1984) and Rieder (1985, 1987) for estimating the whole parameter vector β and by Kurotschka and Müller (1992) for estimating a linear aspect $\varphi(\beta) = C\beta$. Moreover, they derived optimal robust estimators by minimizing the trace of the asymptotic covariance matrix under the side condition that the asymptotic bias is bounded by

some bias bound b . Hence the optimal robust estimators depend on the bias bound b .

Similar optimal robust estimators can be also obtained by deriving the influence functions of the estimators (see Hampel et al., 1986). Basing on this approach, Samarov (1985) proposed optimal bounds b for the robust estimators by minimizing an approximated mean squared error at finite samples. But Samarov calculated the optimal bounds only for few estimators which, in particular for designed experiments, are not the optimal estimators.

In Section 3 of this paper we propose optimal bounds for planned experiments and show that for experiments at optimal designs the bounds are very easy to calculate. Because this approach is based on results of Kurotschka and Müller (1992) and Müller (1992a) concerning optimal robust estimators and optimal designs for robust estimation, at first these results will be briefly repeated in Section 2. In Section 4 we will give the optimal bias bounds for the linear regression model and a one-way lay-out with 3 treatments and a control treatment.

2. Optimal Robust Estimators with Given Bias Bound

Under some regularity conditions (see Müller, 1992b), in particular under condition (1) and the condition $\psi \in \Psi(\delta, C)$, where

$$\Psi(\delta, C) = \{\psi: \mathbf{R} \times T \rightarrow \mathbf{R}^s; \int \psi(z, \cdot) P(dz) = 0, \int \psi(z, t) x(t)' z P(dz) \delta(dt) = C\},$$

a one-step- M -estimator with a score function ψ is asymptotically normally distributed for all contaminated error distributions, i.e.

$$\mathcal{L}(\sqrt{N}(\hat{\varphi}_N - \varphi(\beta)) | Q_\beta^N) \xrightarrow{N \rightarrow \infty} \mathcal{N}(b(\psi, (Q^N)_{N \in \mathbf{N}}), \sigma^2 V(\psi, \delta)) \quad (2)$$

for all $(Q^N)_{N \in \mathbf{N}} \in \mathcal{P}$.

Thereby the asymptotic covariance matrix is given by

$$V(\psi, \delta) = \int \psi \psi' d(P \otimes \delta)$$

and the maximum asymptotic bias satisfies

$$\sup\{|b(\psi, (Q^N)_{N \in \mathbf{N}})|; (Q^N)_{N \in \mathbf{N}} \in \mathcal{P}\} = \sigma \|\psi\|_\infty \quad (3)$$

(see Bickel, 1984; Rieder, 1985, 1987; Kurotschka and Müller, 1992; and Müller, 1992b). This shows that the asymptotic behaviour of a one-step- M -estimator is completely determined by its score function.

Optimal robust estimators for a given bias bound b for the maximum asymptotic bias at a given design δ are those which have a score function $\psi_{b, \delta}$ solving

$$\min\{\text{tr } V(\psi, \delta); \psi \in \Psi(\delta, C), \|\psi\|_\infty \leq b\}.$$

The solution $\psi_{b, \delta}$ is $P \otimes \delta$ -unique and exists if and only if b is greater than or equal to $b_0(\delta) = \min\{\|\psi\|_\infty; \psi \in \Psi(\delta, C)\}$. A general, but very implicit characterization of

these optimal score functions $\psi_{b,\delta}$ was given by Hampel (1978) and Krasker (1980). More explicit characterizations for special designs were given by Kurotschka and Müller (1992) and Müller (1992a). In Müller (1992a) it was also shown that the classical A -optimal designs are optimal for these optimal robust estimators. It means that a design δ^* solving

$$\min\{\text{tr } C I(\delta)^{-1} C'; \delta \in \Delta\}$$

is also a solution of

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

and this holds for all bias bounds $b \geq b_0(\delta^*)$. Thereby we have $b_0(\delta^*) \leq b_0(\delta)$ for all $\delta \in \Delta$.

3. Optimal Bias Bounds

According to (2), the asymptotic mean squared error of a one-step- M -estimator with a score function ψ at a design δ is equal to

$$\sigma^2 (b(\psi, (Q^N)_{N \in \mathbb{N}})^2 + \text{tr } V(\psi, \delta)),$$

and, according to (3), the maximum asymptotic mean squared error is equal to

$$MSE(\psi, \delta) = \sigma^2 [\|\psi\|_\infty^2 + \text{tr } V(\psi, \delta)].$$

To give more or less weight on the maximum asymptotic bias, the mean squared error criterion can be generalized in the following sense

$$GMSE(\psi, \delta) = \alpha w(\|\psi\|_\infty) + \gamma \text{tr } V(\psi, \delta),$$

where $\alpha, \gamma \geq 0$ and $w: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is a convex function.

Now the optimal bias bound can be defined as the solution b^* of

$$\min\{GMSE(\psi_{b,\delta^*}, \delta^*); b \geq b_0(\delta^*)\}. \quad (4)$$

In particular, for $\alpha = \gamma$ and $w(b) = b^2$, the solution of (4) will be denoted by b_2 . It is the optimal bias bound for the mean squared error criterion, i.e. is a solution of

$$\min\{MSE(\psi_{b,\delta^*}, \delta^*); b \geq b_0(\delta^*)\}.$$

The solution of (4) for $\alpha = \gamma$ and $w(b) = b$ will be denoted by b_1 .

In Müller (1992a) it was shown that the score functions ψ_{b,δ^*} of the optimal robust estimators with bias bound b at the A -optimal design δ^* have a very simple form and that the trace of the corresponding asymptotic covariance matrix, $\text{tr } V(\psi_{b,\delta^*}, \delta^*)$, is a decreasing, convex function of b with a known first and second derivative. Hence $GMSE(\psi_{b,\delta^*}, \delta^*)$ is also a convex function of b and the minimum point can be calculated by Newton's method.

If we want to minimize $GMSE(\psi_{b,\delta}, \delta)$ with respect to b for an arbitrary, non- A -optimal design, then the problem becomes very complicated, since no explicit

formulas for the first and the second derivative are known. The advantage of the above definition of an optimal bias bound follows also from the fact that it provides the estimator and the design which minimizes the generalized mean squared error within all estimators and designs, i.e. $(\psi_{b^*, \delta^*}, \delta^*)$ is the solution of

$$\min\{GMSE(\psi, \delta); \psi \in \Psi(\delta, C) \text{ and } \delta \in \Delta\}.$$

4. Examples

4.1. LINEAR REGRESSION

In a linear regression model

$$Y(t) = \beta_0 + \beta_1 t + Z(t) \text{ with } t \in [-1, 1],$$

the A -optimal design for estimating the whole parameter vector β is $\delta^* = \frac{1}{2}(e_{-1} + e_1)$. According to Müller (1992a), the score function of the optimal robust estimator for β with the bias bound $b > b_0(\delta^*) = \sqrt{\pi}$ at δ^* has the form

$$\psi_{b, \delta^*}(z, t) = \frac{\text{sgn}(z)}{y_b} \frac{\min\{|z|, b y_b\}}{\sqrt{2}} (1, t)',$$

where y_b satisfies $y_b = [2\Phi(b y_b) - 1]/\sqrt{2} > 0$, while Φ denotes the standard normal distribution function.

As the optimal bias bound with respect to the mean squared error criterion we get $b_2 = 1.8289$. For this bias bound the fixed point y_b is equal to $y_{b_2} = 0.2386$. The optimal bias bound for the generalized mean squared error criterion with $\alpha = \gamma$ and $w(b) = b$ is $b_1 = 1.9934$, where $y_{b_1} = 0.4317$.

4.2. ONE-WAY LAY-OUT

In a one-way lay-out model

$$Y(i) = \beta_i + Z(i), \text{ for } i = 1, \dots, 4,$$

with a control treatment, say 1, and three other treatments, say 2, 3, 4, let consider a linear aspect of the form $\varphi(\beta) = (\beta_1, \beta_2 - \beta_1, \beta_3 - \beta_1, \beta_4 - \beta_1)'$. The A -optimal design for estimating this aspect is $\delta^* = \frac{1}{5}(2e_1 + e_2 + e_3 + e_4)$. According to Müller (1992a) the score function of the optimal robust estimator for $\varphi(\beta)$ with the bias bound $b > b_0(\delta^*) = 5\sqrt{\pi/2}$ at δ^* has the form

$$\psi_{b, \delta^*}(z, t) = \text{sgn}(z) \frac{\min\{|z|, b y_b\}}{y_b} \cdot \begin{cases} \frac{1}{2}(1, -1, -1, -1)' & \text{for } t = 1 \\ (e_1(t), \dots, e_4(t))' & \text{for } t \neq 1 \end{cases},$$

where y_b satisfies $y_b = [2\Phi(b y_b) - 1]/5 > 0$.

As the optimal bias bound with respect to the mean squared error criterion we get $b_2 = 6.4660$ with $y_{b_2} = 0.0675$. The optimal bias bound for the generalized mean squared error criterion with $\alpha = \gamma$ and $w(b) = b$ is $b_1 = 8.1950$ with $y_{b_1} = 0.1644$.

Acknowledgements

The author thanks the referees for their comments and suggestions which improved the presentation of the paper.

References

- Bickel, P.J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association* **70**, 428–434.
- Bickel, P.J. (1984). Robust regression based on infinitesimal neighbourhoods. *The Annals of Statistics* **12**, 1349–1368.
- Hampel, F.R. (1978). Optimally bounding the gross-error-sensitivity and the influence of position in factor space. In: *Proceedings of the ASA Statistical Computing Section*. ASA, Washington, D.C., 59–64.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics - The Approach Based on Influence Functions*. Wiley, New York.
- Huber, P.J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics* **35**, 73–101.
- Krasker, W.S. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48**, 1333–1346.
- Kurotschka, V. and Müller, Ch.H. (1992). Optimum robust estimation of linear aspects in conditionally contaminated linear models. *The Annals of Statistics* **20**, 331–350.
- Müller, Ch.H. (1992a). Optimal designs for robust estimation in conditionally contaminated linear models. *Journal of Statistical Planning and Inference*. To appear.
- Müller, Ch.H. (1992b). One-step- M -estimators in conditionally contaminated linear models. Preprint No. A-92-11, Freie Universität Berlin, Fachbereich Mathematik. Submitted for publication.
- Samarov, A.M. (1985). Bounded-influence regression via local minimax mean squared error. *Journal of the American Statistical Association* **80**, 1032–1040.
- Rieder, H. (1985). Robust estimation of functionals. Technical Report, Universität Bayreuth.
- Rieder, H. (1987). Robust regression estimators and their least favorable contamination curves. *Statistics and Decisions* **5**, 307–336.