Robust Linear Calibration

CHRISTOS P. KITSOS and CHRISTINE H. MÜLLER
Athens University of Business and Economics, Freie Universität Berlin

March 1995

Summary. We regard the simple linear calibration problem where only the response y of the regression line $y = \beta_0 + \beta_1 t$ is observed with errors. The experimental conditions t are observed without error. For the errors of the observations y we assume that there may be some gross errors providing outlying observations. This situation can be modeled by a conditionally contaminated regression model. In this model the classical calibration estimator based on the least squares estimator has an unbounded asymptotic bias. Therefore we introduce calibration estimators based on robust one-step-M-estimators which have a bounded asymptotic bias. For this class of estimators we discuss two problems: The optimal estimators and their corresponding optimal designs. We derive the locally optimal solutions and show that the maximin efficient designs for non-robust estimation and robust estimation coincide.

AMS 1991 subject classifications: 62J05, 62F35, 62G20, 62K05

Key words: Linear calibration, classical estimator, conditional contamination, robust estimation, one-step-M-estimator, local optimality, optimal design, asymptotic efficiency, maximin efficiency.

1 Introduction

In the linear calibration set-up the simple linear regression model

$$Y(t) = \beta_0 + \beta_1 t + Z = x(t)'\beta + Z \qquad \text{with } t \in \mathbb{R}$$

is assumed where $\beta = (\beta_0, \beta_1)'$ is unknown and $x(t) = (1, t)'$ is the known regression function. Given a favorite value y of $Y(t)$, the corresponding experimental condition t_y should be estimated so as to reproduce y if there is no error, $Z = 0$. In particular, if the expected value of the error Z is equal to zero, the experimental condition t_y should be that condition which provides $E(Y(t_y)) = y$. Hence, the interest is focused on the non-linear aspect

$$\varphi(\beta) := \varphi_y(\beta) := t_y = \frac{y - \beta_0}{\beta_1}$$

of the linear regression model. This non-linear aspect can be estimated by repeating the linear regression experiment at given experimental conditions $t_{1N},...,t_{NN} \in T := [-1,1]$ and estimating $\varphi(\beta)$ from the observations Y_{nN} given by

$$Y_{nN} = \beta_0 + \beta_1 t_{nN} + Z_{nN}.$$

In absolute calibration problems (see for example Buonaccorsi (1986), Osborne (1991) and Kitsos (1992)) it is assumed that the errors Z_{1N},...,Z_{NN} are independent and identically distributed with expectation $E(Z_{nN}) = 0$ and that the experimental conditions t_{nN} are without errors. For these problems the so-called "classical estimator" $\varphi(\widehat{\beta}_N^*)$ based on the least squares estimator $\widehat{\beta}_N^*$ for β should be used (see Eisenhart (1939), Osborne (1991)). But, if with some probability, at a given experimental condition t_{nN}, an outlying observation caused by a gross error may occur, then the expected value of Z_{nN} is not any more zero. Then a conditionally contaminated linear regression model is more adequate. In such a model it is assumed that the error variables $Z_{1N},...,Z_{NN}$ are independent and distributed according to a contaminated normal distribution where the contamination may be different for different experimental conditions. I.e.

$$(1.1) \quad \frac{1}{\sigma} Z_{nN} \sim Q_{nN}(dz) := (1 - N^{-1/2} R\, \epsilon(t_{nN}))\, n_{(0,1)}(dz) \\ + N^{-1/2} R\, \epsilon(t_{nN})\, g(z, t_{nN})\, n_{(0,1)}(dz)$$

where $n_{(0,1)}$ is the standard normal distribution, $R > 0$, $\frac{1}{N}\sum_{n=1}^{N} \epsilon(t_{nN}) \leq 1$ for almost all $N \in I\!N$ and $\int g(z,t)\, n_{(0,1)}(dz) = 1$, $g(z,t) \geq 0$ for all $z \in I\!R$, $t \in T$. Thereby the Markov kernel $g(\cdot,t)n_{(0,1)}$ models the form of contamination, and $\epsilon(t) \geq 0$ the amount of contamination. The set $\mathcal{P}_R$ of all sequences $(Q^N := \bigotimes_{n=1}^{N} Q_{nN})_{N\in I\!N}$ defines a contamination neighbourhood around the classical model $(P^N)_{N\in I\!N} := ((n_{(0,1)})^N)_{N\in I\!N}$. If $(Z_{1N},...,Z_{NN})$ is distributed according to Q^N and β is the true parameter vector, then we denote by Q_β^N the distribution of $(Y_{1N},...,Y_{NN})$, in particular by P_β^N the distribution of $(Y_{1N},...,Y_{NN})$ for the classical model.

Estimating in this model the non-linear aspect $\varphi(\beta)$ by the estimator $\varphi(\widehat{\beta}_N^*)$, where $\widehat{\beta}_N^*$ is the least squares estimator, an unbounded asymptotic bias appears. I.e. $\sqrt{N}(\varphi(\widehat{\beta}_N^*) - \varphi(\beta))$ is under $(Q_\beta^N)_{N\in I\!N}$ asymptotically normally distributed with mean (bias) $b_\beta((Q^N)_{N\in I\!N})$, where $\sup\{b_\beta((Q^N)_{N\in I\!N});\ (Q^N)_{N\in I\!N} \in \mathcal{P}_R\} = \infty$. But robust estimators have a bounded asymptotic bias, whence the estimator $\varphi(\widehat{\beta}_N^*)$ is not robust. Therefore in Section 2 robust estimators for $\varphi(\beta)$ are proposed for which the asymptotic bias is bounded. In Section 3 several strategies are discussed for obtaining optimal robust estimators and optimal designs for robust estimation. In particular a maximin efficiency criterion for robust estimation is proposed. For this criterion locally optimal robust estimators and locally optimal designs for robust estimation are defined as those estimators and designs which minimize the asymptotic variance under the side condition that the asymptotic bias is bounded by some local bound $b(\beta) < \infty$. Then the efficiency of a given robust estimator at a given design is measured by comparing its asymptotic variance with the asymptotic variance of the locally optimal robust estimator at the locally optimal design. This efficiency criterion is similar to the classical efficiency criterion for non-robust estimation of a non-linear aspect (see Silvey (1980)), where a given design is compared with the locally optimal design. In Section 4 the locally optimal robust estimators at the locally optimal designs are derived. In Section 5 the locally optimal solutions are used to show that the maximin efficient design for robust estimation coincides with the maximin efficient design for non-robust estimation which was derived by Kitsos (1992). All proofs are given in Section 6.

2 Estimators based on robust estimators for β

In a conditionally contaminated linear calibration model the non-linear aspect $\varphi(\beta)$ should be estimated with a robust estimator. Such an estimator can be of the form $\varphi(\widehat{\beta}_N)$ where $\widehat{\beta}_N$ is an optimal robust one-step-M-estimator for β. A one-step-M-estimators for β with score function $\psi : I\!R \times T \to I\!R^2$ has the form

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

where $d_N := (t_{1N}, ..., t_{NN})$ and $y_N := (y_{1N}, ..., y_{NN})$. The inital estimator $\widehat{\beta}_N^0$ may be the least squares estimator or some robust estimator for β, for example a M-estimator. The variance estimator $\widehat{\sigma}_N$ may be the mean squared residuals or some robust estimator for the variance, for example Huber's Proposal 2 (Huber (1964)).

If $d_N = (t_{1N}, ..., t_{NN})$ is the design for the calibration experiment then we can define a corresponding design measure by

$$\delta_N := \frac{1}{N} \sum_{n=1}^N e_{t_{nN}}$$

where e_t denotes the Dirac measure on t. Now, assume that a design measure δ and a finite set $T_0 \subset T$ exist, so that the supports of δ and all δ_N are included in T_0 and that for all $t \in T_0$ we have

$$\lim_{N \to \infty} \delta_N(t) = \delta(t).$$

Moreover assume that the score function ψ satisfies

(2.1) $\quad \psi(z, t) = \psi_0(z, t) + c(t)\mathrm{sgn}(z),$ where for all $t \in T_0$

$\psi_0(\cdot,t)$ is antisymmetric, continuous and there exists $l_1(t),...,l_L(t)$ so that $\psi_0(\cdot,t)$ has bounded and continuous derivatives on $I\!R \setminus \{l_1(t),...,l_L(t)\}$,

and

$$(2.2) \quad \int \psi(z,t)x(t)'zP(dz)\delta(dt) = E_{p\times p}$$

where $E_{p\times p}$ denotes the identity matrix.

Then under some tightness conditions on the initial and variance estimator the one-step-M-estimator $\widehat{\beta}_N$ is asymptotically normally distributed for all contaminated error distributions (see Müller (1994b)), i.e.

$$(2.3) \quad \mathcal{L}(\sqrt{N}(\widehat{\beta}_N - \beta)|Q_\beta^N) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(b(\psi,(Q^N)_{N\in I\!N}), \sigma^2 \int \psi\,\psi'\,d(P\otimes\delta)\,) \quad \text{for all } (Q^N)_{N\in I\!N} \in \mathcal{P}_R \text{ and all } \beta \in I\!R^2$$

with maximum asymptotic bias

$$\sup\{|b(\psi,(Q^N)_{N\in I\!N})|;\ (Q^N)_{N\in I\!N} \in \mathcal{P}_R\} = R\,\sigma\,\operatorname*{ess\,sup}_{P\otimes\delta} |\psi|.$$

The following theorem shows that an estimator $\varphi(\widehat{\beta}_N)$ based on an asymptotically normally distributed one-step-M-estimator $\widehat{\beta}_N$ is also asymptotically normally distributed. For that we have

$$B := \{\beta = (\beta_0,\beta_1)' \in I\!R^2; \beta_1 \neq 0\},$$

and

$$\dot{\varphi}_\beta := \frac{\partial}{\partial\beta}\,\varphi(\beta) = \frac{-1}{\beta_1}\,(1,\,\varphi(\beta)).$$

Lemma 2.1 *If $\widehat{\beta}_N$ is an asymptotically normally distributed one-step-M-estimator for β, then*

$$\sqrt{N}(\varphi(\widehat{\beta}_N) - \varphi(\beta)) - \sqrt{N}(\dot{\varphi}_\beta\widehat{\beta}_N - \dot{\varphi}_\beta\beta) \overset{N\to\infty}{\longrightarrow} 0$$

in probability for $(P_\beta^N)_{N\in I\!N}$ for all $\beta \in B$.

Using this lemma we get the following theorem.

Theorem 2.1 *If $\widehat{\beta}_N$ is an asymptotically normally distributed one-step-M-estimator for β with score function ψ, then also $\varphi(\widehat{\beta}_N)$ is asymptotically normally distributed for all contaminated error distributions and all $\beta \in B$, i.e.*

$$
\begin{aligned}
(2.4)\quad & \mathcal{L}(\sqrt{N}(\varphi(\widehat{\beta}_N) - \varphi(\beta))|Q_\beta^N) \\
& \xrightarrow{N\to\infty} \mathcal{N}(\dot{\varphi}_\beta\, b(\psi, (Q^N)_{N\in I\!N}),\, \sigma^2\, \dot{\varphi}_\beta \int \psi\psi'\, d(P\otimes\delta)\, \dot{\varphi}_\beta') \\
& \quad \textit{for all } (Q^N)_{N\in I\!N} \in \mathcal{P}_R \textit{ and all } \beta \in B,
\end{aligned}
$$

with maximum asymptotic bias

$$\sup\{|\dot{\varphi}_\beta\, b(\psi, (Q^N)_{N\in I\!N})|;\ (Q^N)_{N\in I\!N} \in \mathcal{P}_R\} = R\,\sigma\, \operatorname*{ess\,sup}_{P\otimes\delta} |\dot{\varphi}_\beta\, \psi| \quad \textit{for all } \beta \in B.$$

Without loss of generality we now can assume that the variance σ^2 as well as the radius R of the contamination neighbourhood $\mathcal{P}_R$ are equal to 1.

Recall that the least squares estimator behaves asymptotically like a one-step-M-estimator with score function

$$\psi_\infty(z,t) := I(\delta)^{-1} x(t)\, z$$

where $I(\delta) := \int x(t)\, x(t)'\, \delta(dt)$. Hence the maximum asymptotic bias of an estimator for $\varphi(\beta)$ which is based on the least squares estimator is equal to $\operatorname{ess\,sup}_{P\otimes\delta} |\dot{\varphi}_\beta \psi_\infty| = \infty$. For robust estimators the maximum asymptotic bias should be less than infinity. But this is only satisfied if the score function of the underlying one-step-M-estimator for β is bounded, i.e. if the underlying one-step-M-estimator is a robust estimator for β.

To get "good" robust estimators for $\varphi(\beta)$ we can base the estimator $\varphi(\widehat{\beta}_N)$ on an optimal robust one-step-M-estimator $\widehat{\beta}_N$ for β. This will not provide an optimal robust estimator for $\varphi(\beta)$ but it will provide a reasonable choice of a robust estimator for $\varphi(\beta)$ because a globally optimal robust estimator for $\varphi(\beta)$ does not exist (see Section 3). Optimal robust estimators $\widehat{\beta}_N^b$, say, for β are those estimators which minimize the trace of the asymptotic covariance matrix under the side condition that the maximum asymptotic bias is bounded by some bias bound b. Because only the score functions of the one-step-M-estimators determine the asymptotic behaviour, the optimal robust estimators are given by their score function. A general characteriztion of these optimal score functions is given in Krasker (1980). For designed experiments Kurotschka and Müller (1992) gave a more explicit characterization of these score functions. In particular if we use the design

$$\delta_p := p\, e_1 + (1-p)\, e_{-1}, \qquad \text{with } p \in (0,1),$$

then the optimal score function $\psi_{b,p}$ for robust estimation of β with bias bound b has the form

$$\psi_{b,p}(z,t) = \begin{cases} \operatorname{sgn}(z)\, \frac{b}{\sqrt{2}}\, (1,t)' & \text{for } \frac{\sqrt{\pi}}{2\,\delta_p(\{t\})} = b, \\ \operatorname{sgn}(z)\, \frac{\min\{|z|, b\, y_{b,p}(t)\}}{\sqrt{2}\, y_{b,p}(t)}\, (1,t)' & \text{for } \frac{\sqrt{\pi}}{2\,\delta_p(\{t\})} < b, \end{cases}$$

for $t \in \{1,-1\}$ where $y_{b,p}(t)$ is the positive solution of $y = \delta_p(\{t\})\sqrt{2}(2\Phi(b\,y)-1)$. This optimal score function exists if and only if $b \geq \max\{\frac{1}{\sqrt{2p}}, \frac{1}{\sqrt{2(1-p)}}\}\sqrt{\frac{\pi}{2}}$. Note that for deriving the optimal score function according to Theorem 2 of Kurotschka and Müller (1992) we have to calculate

$$(2.5) \quad I(\delta_p)^{-1}\, x(t) = \begin{cases} \frac{1}{2\,p}\, (1,1)' & \text{for } t = 1 \\ \\ \frac{1}{2\,(1-p)}\, (1,-1)' & \text{for } t = -1 \end{cases} .$$

The following lemma gives the maximum asymptotic bias and the asymptotic variance of the estimators based on optimal robust estimators $\widehat{\beta}_N^b$ at δ_p. For that set

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

where $f(a,b)$ is an implicit function given by $f(a,b) = (2\Phi(b\, f(a,b)) - 1)\, a$ and $g(v) := \int \min\{|z|, v\}^2 P(dz)$.

Lemma 2.2 *The maximum asymptotic bias of the estimator $\varphi(\widehat{\beta}_N^b)$ for $\varphi(\beta)$ is equal to*

$$(2.6) \quad b(\beta,p) := ess \sup_{P\otimes\delta_p} |\dot{\varphi}_\beta\, \psi_{b,p}| = \max\left\{ \frac{|1+\varphi(\beta)|}{|\beta_1|\sqrt{2}}, \frac{|1-\varphi(\beta)|}{|\beta_1|\sqrt{2}} \right\} b$$

which is independent of p. Its asymptotic variance is equal to

$$\begin{aligned}(2.7) \quad & \dot{\varphi}_\beta \int \psi_{b,p}\, \psi'_{b,p}\, d(P \otimes \delta_p)\, \dot{\varphi}'_\beta \\ = \quad & \frac{(1+\varphi(\beta))^2}{\beta_1^2 4p}\, v_b(2p^2) + \frac{(1-\varphi(\beta))^2}{\beta_1^2 4(1-p)}\, v_b(2(1-p)^2).\end{aligned}$$

3 Optimal robust estimators and designs

The asymptotic normality (2.4) of estimators of the form $\varphi(\widehat{\beta}_N)$ shows that the maximum asymptotic bias $\operatorname{ess\,sup}_{P\otimes\delta} |\dot{\varphi}_\beta\, \psi|$ and the asymptotic variance

$$V(\psi, \delta, \beta) := \dot{\varphi}_\beta \int \psi\psi'\, d(P \otimes \delta)\, \dot{\varphi}'_\beta$$

depend on β. If there is no bias bound for the maximum asymptotic bias then the matrix $\int \psi\, \psi'\, d(P\otimes\delta)$ is minimized uniformly by ψ_∞ within all score functions ψ satisfying condition (2.1) and (2.2). I.e. the variance $V(\psi_\infty, \delta, \beta)$ is the minimum for all β so that the estimator based on the least squares estimator is asymptotically optimal. (Recall from Section 2 the comment that the least squares estimator behaves asymptotically like an one-step-M-estimator with score function ψ_∞.) But if the maximum asymptotic bias is bounded by some bias bound b, then in principle it is impossible to minimize $\int \psi\, \psi'\, d(P\otimes\delta)$ uniformly, so that an optimal robust estimator cannot be found independently of β. Also optimal designs δ cannot be found independently of β, and this problem already occurs for classical non-robust estimation. All non-linear optimal experimental design theory suffer from this dependence on β so that different strategies for non-robust estimation were investigated to overcome this problem, see Ford et al. (1989) and for calibration Osborne (1991).

One strategy of avoiding dependence on β is the so-called maximin efficiency approach which was used for non-robust non-linear estimation problems by Silvey (1980) and Kitsos (1986). We extend this approach and define it as follows: If ψ_β is the score function of the locally optimal estimator for $\varphi(\beta)$ and δ_β is the locally optimal design, then the efficiency of a given estimator for $\varphi(\beta)$ with score function ψ at a given design δ is

$$E_\beta(\psi, \delta) := \frac{V(\psi_\beta, \delta_\beta, \beta)}{V(\psi, \delta, \beta)} \in [0, 1].$$

Then, a design δ and an estimator with score function ψ and bias bound b is called maximin efficient for robust estimation of $\varphi(\beta)$ with bias bound b if their minimum efficiency is greater than the minimum efficiency of any other design and estimator with bias bound b, i.e. which solve

$$(3.1)\quad \max\{\min_{\beta\in B} E_\beta(\psi,\delta);\ \delta\in\Delta,\ \psi\in\Psi_\delta \text{ with } \operatorname*{ess\,sup}_{P\otimes\delta}|\psi|\le b\}.$$

Thereby Δ is some given set of design measures, for example $\Delta=\{\delta;\ \text{supp}(\delta)\subset[-1,1]\}$, the set of all designs with support included in the intervall $[-1,1]$, and

$$\Psi_\delta := \{\psi : I\!R\times T\to I\!R^2;\ \psi \text{ satisfies condition (2.1) and (2.2)}\}$$

is the set of score functions providing asymptotically normally distributed one-step-M-estimators for β.

For deriving maximin efficient solutions for robust estimation a first step is to determine the locally optimal robust estimators and designs.

4 Locally optimal robust estimators and designs

Lemma 2.1 shows that locally an estimator $\varphi(\widehat{\beta}_N)$ based on an asymptotically normally distributed one-step-M-estimator $\widehat{\beta}_N$ behaves asymptotically like an one-step-M-estimators for the linear aspect $\gamma(\tilde{\beta})=\dot{\varphi}_\beta\tilde{\beta}$. Thereby a one-step-M-estimator $\widehat{\gamma}_N$, say, for a linear aspect $\gamma(\beta)=C\beta,\ C\in I\!R^{s\times 2}$, is defined as (see Müller (1994b))

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

where ψ now lies in

$$\begin{aligned}\Psi_\delta(C) \;:=\; &\{\psi: I\!R\times T\to I\!R^s;\ \psi \text{ satisfies condition 2.1 and}\\ &\int\psi(z,t)x(t)'zP(dz)\delta(dt)=C\}.\end{aligned}$$

Hence locally optimal robust estimators are the optimal robust estimators for the particular linear aspect $\gamma(\tilde{\beta})=\dot{\varphi}_\beta\tilde{\beta}$. Therefore locally optimal designs δ_β and locally optimal estimators ψ_β for robust estimation with local bias bound $b(\beta)$ in Δ are defined as solutions of the following minimization problem

$$\begin{aligned}(4.1)\quad &\min\{V(\psi,\delta,\beta);\ \delta\in\Delta,\ \psi: I\!R\times T\to I\!R^2 \text{ satisfies}\\ &\qquad\dot{\varphi}_\beta\,\psi\in\Psi_\delta(\dot{\varphi}_\beta) \text{ and } \operatorname*{ess\,sup}_{P\otimes\delta}|\dot{\varphi}_\beta\psi|\le b(\beta)\}.\end{aligned}$$

Note that if ψ lies in Ψ_δ then $\dot{\varphi}_\beta\,\psi$ lies in $\Psi_\delta(\dot{\varphi}_\beta)$ but that $\dot{\varphi}_\beta\psi \in \Psi_\delta(\dot{\varphi}_\beta)$ does not imply $\psi \in \Psi_\delta$. For example $\dot{\varphi}_\beta\,\beta$ may be identifiable at δ while the whole parameter vector β is not identifiable at δ.

Now, if there is no local bias bound $b(\beta)$ then the least squares estimator is the locally optimal estimator for all $\beta \in B$ and the design

$$\delta_\beta := p_\beta e_1 + (1-p_\beta)e_{-1},$$

with

$$p_\beta := \begin{cases} \frac{1}{2}\,(\varphi(\beta)+1) & \text{for } |\varphi(\beta)| \le 1 \\ \frac{1}{2\,\varphi(\beta)}\,(\varphi(\beta)+1) & \text{for } |\varphi(\beta)| > 1 \end{cases},$$

is the locally optimal one in $\Delta = \{\delta;\ \text{supp}(\delta) \in [-1,1]\}$. This shows the following theorem which is due to Kitsos (1992).

Theorem 4.1 *If $b(\beta) = \infty$ and $\Delta = \{\delta;\ supp(\delta) \in [-1,1]\}$ then $(\delta_\beta, \psi_\infty)$ is locally optimal, i.e. a solution of (4.1).*

Note if the whole parameter vector β is not identifiable at δ_β which is the case for $|\varphi(\beta)| = 1$ then ψ_∞ will be defined by $\psi_\infty(z,t) = I(\delta_\beta)^- x(t)\,z$ where M^- denotes a generalized inverse of a matrix M.

If $b(\beta) < \infty$ then the score functions of the locally optimal robust estimators have the following form

$$(4.2)\qquad \psi_{\beta,b(\beta)}(z,t) := \begin{cases} I(\delta_\beta)^- x(t)\,\frac{\text{sgn}(z)}{|\dot{\varphi}_\beta\, I(\delta_\beta)^- x(t)|}\,b(\beta) \\ \qquad\qquad \text{for } b(\beta) = \sqrt{\dot{\varphi}_\beta\, I(\delta_\beta)^-\,\dot{\varphi}_\beta'}\,\sqrt{\frac{\pi}{2}} \\ I(\delta_\beta)^- x(t)\frac{\text{sgn}(z)\,\min\{|z|, b(\beta)\,y_{b,\beta}\}}{|\dot{\varphi}_\beta\, I(\delta_\beta)^- x(t)|\,y_{b,\beta}} \\ \qquad\qquad \text{for } b(\beta) > \sqrt{\dot{\varphi}_\beta\, I(\delta_\beta)^-\,\dot{\varphi}_\beta'}\,\sqrt{\frac{\pi}{2}} \end{cases}$$

where $y_{b,\beta}$ is the positive solution of

$$y = (2\Phi(b(\beta)\,y) - 1)\cdot\left(\sqrt{\dot{\varphi}_\beta\, I(\delta_\beta)^-\,\dot{\varphi}_\beta'}\,\right)^{-1}.$$

Note that $|\dot{\varphi}_\beta\, I(\delta_\beta)^- x(t)|^2 = \dot{\varphi}_\beta\, I(\delta_\beta)^-\,\dot{\varphi}_\beta'$ for $t \in \{-1,1\}$, $\dot{\varphi}_\beta\, I(\delta_\beta)^-\,\dot{\varphi}_\beta' = \frac{1}{\beta_1^2}$ for $|\varphi(\beta)| \le 1$ and $\dot{\varphi}_\beta\, I(\delta_\beta)^-\,\dot{\varphi}_\beta' = \frac{\varphi(\beta)^2}{\beta_1^2}$ for $|\varphi(\beta)| > 1$ (see Kitsos (1992)).

Theorem 4.2 *If* $\sqrt{\dot\varphi_\beta\, I(\delta_\beta)^- \,\dot\varphi_\beta'}\sqrt{\frac{\pi}{2}} \leq b(\beta) < \infty$ *and* $\Delta = \{\delta;\ supp(\delta) \in [-1,1]\}$ *then* $(\delta_\beta, \psi_{\beta,b(\beta)})$ *is locally optimal, i.e. a solution of (4.1).*

Theorem 4.2 and Theorem 4.1 show that the locally optimal designs for robust estimation coincide with the locally optimal designs for non-robust estimation. Moreover they show that all locally optimal designs have a support included in the two-point set $\{-1, 1\}$.

5 Maximin efficient designs

For the calibration problem Kitsos (1992) showed that $\delta_{\frac{1}{2}}$ (i.e. the D- and A-optimal design for non-robust estimation of β) is the maximin efficient design in $\Delta := \{\delta;\ \text{supp}(\delta) \subset [0,1]\}$ for classical non-robust estimation of $\varphi(\beta)$. We are now going to show that $\delta_{\frac{1}{2}}$ is also maximin efficient for robust estimation.

If $\varphi(\widehat{\beta}_N)$ is a robust estimator at a given design δ which is for all $\beta \in B$ asymptotically normally distributed with asymptotic variance $V(\psi, \delta, \beta)$ and maximum asymptotic bias $b(\beta, \psi, \delta) := \text{ess}\sup_{P\otimes\delta} |\dot\varphi_\beta\, \psi|$ we get the efficiency of this estimator at the design δ by comparing it with the locally optimal robust one-step-M-estimators at the locally optimal design δ_β. Thereby for the locally optimal estimators we should use the same local bias bound $b(\beta)$ which is attained by the estimator $\varphi(\widehat{\beta}_N)$ at δ to make fair comparisons. I.e. we set for the local bias bound $b(\beta) = b(\beta, \psi, \delta)$. Then the asymptotic efficiency measure is

$$E_\beta(\psi, \delta) := \frac{V(\psi_{\beta,b(\beta)}, \delta_\beta, \beta)}{V(\psi, \delta, \beta)}$$

where $\psi_{\beta,b(\beta)}$ is the score function of the locally optimal one-step-M-estimator at δ_β with local bias bound $b(\beta) = b(\beta, \psi, \delta)$.

In Section 4 it was shown that all locally optimal designs in $\Delta = \{\delta;\ \text{supp}(\delta) \subset [-1,1]\}$ are concentrated on the two-point set $\{-1, 1\}$. Hence we can restrict our attention to the set

$$\Delta^0 := \{\delta;\ \text{supp}(\delta) \subset \{-1, 1\}\} = \{\delta_p;\ p \in (0, 1)\}$$

so that the maximin efficiency criterion of Section 3 becomes

$$\max\{\min_{\beta\in B} E_\beta(\psi, \delta);\ \delta \in \Delta^0,\ \psi \in \Psi_\delta \text{ with } \text{ess}\sup_{P\otimes\delta} |\psi| \leq b\}.$$

Moreover if we restrict ourselves only to estimators of the form $\varphi(\widehat{\beta}_N^b)$, which are based on optimal robust one-step-M-estimators for β with bias bound b, which have the score functions $\psi_{b,p}$ (see Section 2), then the maximin efficiency criterion can be reduced to a pure design problem of the following form

$$(5.1) \quad \max\{\min_{\beta\in B} E_\beta(\psi_{b,p}, \delta_p);\ p \in (0,1)\}.$$

Designs which solve this problem will be called maximin efficient designs for robust estimation with bias bound b. For $b = \infty$ this maximin efficiency criterion coincides with the classical maximin efficiency criterion for designs for non-robust estimation, as was regarded in Silvey (1980).

Recall that at the design δ_p the optimal robust estimator $\widehat{\beta}_N^b$ for β with bias bound b only exists if $b \geq \max\{\frac{1}{\sqrt{2p}}, \frac{1}{\sqrt{2(1-p)}}\}\sqrt{\frac{\pi}{2}}$ (see Section 2). Therefore, for deriving maximin efficient designs for robust estimation with bias bound b, we have to restrict Δ^0 to $\{\delta_p;\ p \in [0,1]$ with $\max\{\frac{1}{\sqrt{2p}}, \frac{1}{\sqrt{2(1-p)}}\}\sqrt{\frac{\pi}{2}} \leq b\}$ or to define $E_\beta(\psi_{b,p}, \delta_p) = 0$ for $\max\{\frac{1}{\sqrt{2p}}, \frac{1}{\sqrt{2(1-p)}}\}\sqrt{\frac{\pi}{2}} > b$, which means that there is no efficiency if there is no robust estimator for β with bias bound b. We think that defining $E_\beta(\psi_{b,p}, \delta_p) = 0$ for $\max\{\frac{1}{\sqrt{2p}}, \frac{1}{\sqrt{2(1-p)}}\}\sqrt{\frac{\pi}{2}} > b$ is the best way for deriving the maximin efficient design for robust estimation.

The following lemma shows that for $b \geq \max\{\frac{1}{\sqrt{2p}}, \frac{1}{\sqrt{2(1-p)}}\}\sqrt{\frac{\pi}{2}}$ always $b(\beta,p) := b(\beta, \psi_{b,p}, \delta_p) \geq \sqrt{\dot{\varphi}_\beta\, I(\delta_\beta)^{-1}\, \dot{\varphi}_\beta'}\sqrt{\frac{\pi}{2}}$ is satisfied. Hence in this case always the locally optimal robust estimator at δ_β exists and the efficiency $E_\beta(\psi_{b,p}, \delta_p)$ is well-defined.

Lemma 5.1 *If $b \geq \max\{\frac{1}{\sqrt{2p}}, \frac{1}{\sqrt{2(1-p)}}\}\sqrt{\frac{\pi}{2}}$ then for all $\beta \in B$ at δ_β the locally optimal robust estimator for $\varphi(\beta)$ with local bias bound $b(\beta,p)$ exists and its asymptotic variance satisfies*

$$V(\psi_{\beta,b(\beta)}, \delta_\beta, \beta) \geq \begin{cases} \frac{1}{\beta_1^2}\, v_b(2) & \text{for } |\varphi(\beta)| \leq 1 \\ \frac{\varphi(\beta)^2}{\beta_1^2}\, v_b(2) & \text{for } |\varphi(\beta)| > 1 \end{cases},$$

with equality if and only if $|\varphi(\beta)| = 1$.

Theorem 5.1 *If $b \geq \sqrt{\pi}$ then the design $\delta_{\frac{1}{2}}$ is maximin efficient for robust estimation with bias bound b, i.e. a solution of (5.1), and the maximin efficiency is equal to*

$$E(b) := \frac{v_b(2)}{2\, v_b(\frac{1}{2})}.$$

Figure 1: *Maximin efficiencies $E(b)$ for robust estimation with bias bound b versus the bias bound b*

Theorem 5.1 shows that the maximin efficient designs for robust estimation are always independent of the bias bound b. Moreover they coincide with the maximin efficient design for non-robust estimation which was derived in Kitsos (1992). Thereby, the derivation of the maximin design in Kitsos (1992) is a special case of the derivation of maximin design for robust estimation (see the proof of Theorem 5.1) because $b = \infty$ provides the non-robust case. In particular for $b = \infty$ we have $v_b(a) = 1$ for all $a > 0$, so that the maximin efficiency for non-robust estimation is equal to 0.5. For the minimum bias bound $b = \sqrt{\pi}$ the maximin efficiency is equal to 0.319 and increases for $b \to \infty$ quickly to the maximin efficiency for non-robust estimation as Figure 1 shows.

Note also that the maximin efficient design for robust estimation of $\varphi(\beta)$ is also the optimal design for robust estimation of the whole parameter vector β. Thereby the bias bound b for robust estimation of β should also satisfy $b \geq \sqrt{\pi}$. See Müller (1994a).

6 Proofs

Proof of Lemma 2.1

The non-linear function φ is continously differentiable on the set B. Hence the asymptotic normality of $\widehat{\beta}_N$ implies for all $\beta \in B$

$$\begin{aligned} &\sqrt{N}(\varphi(\widehat{\beta}_N) - \varphi(\beta)) - \sqrt{N}(\dot{\varphi}_\beta \widehat{\beta}_N - \dot{\varphi}_\beta \beta) \\ = &\sqrt{N}(|\widehat{\beta}_N - \beta|) \frac{\varphi(\widehat{\beta}_N) - \varphi(\beta) - \dot{\varphi}_\beta(\widehat{\beta}_N - \beta)}{|\widehat{\beta}_N - \beta|} \overset{N\to\infty}{\longrightarrow} 0. \square \end{aligned}$$

Proof of Theorem 2.1

The assertion follows from the asymptotic normality of $\widehat{\beta}_N$ given in (2.3) and Lemma 2.1. $\square$

Proof of Lemma 2.2

The maximum asymptotic bias and the asymptotic variance follow from the special form of $\psi_{b,p}$. Recall $R = 1 = \sigma$. $\square$

Proof of Theorem 4.1

The assertion follows from a result in Kitsos (1992) who showed that the design δ_β is locally optimal for classical estimators based on the least squares estimator which asymptotically behaves like an one-step-M-estimator with score function ψ_∞. Note that the local optimality of δ_β can be also seen by Elfving's (1952) theorem about c-optimal designs ($c = \dot{\varphi}_\beta$). The optimal weights of δ_β can be also deduced from a result of Pukelsheim and Torsney (1991). For the optimality of the score function ψ_∞, in particular in the case where the whole parameter vector β is not identifiable at δ_β, see also the proof of Theorem 4.2. $\square$

Proof of Theorem 4.2

The assertions follow from the results in Müller (1994a) about optimal designs for robust estimation of linear aspects. In particular there was shown that for robust estimation of a linear aspect $\gamma(\beta) = C\beta$ with bias bound b always (for all bias bounds b) at the A-optimal design the trace of the asymptotic covariance matrix is minimized. Recall that the A-optimal design for estimating $\gamma(\beta) = C\beta$ is that design which minimize $\operatorname{tr} C\, I(\delta)^- C'$. For estimating the special linear aspect $\gamma(\tilde{\beta}) = \dot{\varphi}_\beta \tilde{\beta}$ the A-optimal design coincides with the c-optimal design where $c = \dot{\varphi}_\beta$. Hence the design δ_β is optimal (see the proof of Theorem 4.1). Moreover from Lemma 1 in Müller (1994a) it follows that $\dot{\varphi}_\beta \psi_{\beta,b(\beta)}$ is the score function of the optimal robust estimator for $\gamma(\tilde{\beta}) = \dot{\varphi}_\beta \tilde{\beta}$ with bias bound $b(\beta)$. Because of $\psi_{\beta,\infty} = \psi_\infty$ we get besides the optimality of $\psi_{\beta,b(\beta)}$ also the optimality of ψ_∞ as a special case ($b(\beta) = \infty$). $\square$

Proof of Lemma 5.1

Because of

$$b(\beta) = b(\beta, p) := \operatorname*{ess\,sup}_{P \otimes \delta_p} |\dot\varphi_\beta \, \psi_{b,p}| = \max\left\{ \frac{|1+\varphi(\beta)|}{|\beta_1|\sqrt{2}}, \frac{|1-\varphi(\beta)|}{|\beta_1|\sqrt{2}} \right\} b$$

(see Lemma 2.2) and because of

$$\dot\varphi_\beta \, I(\delta_\beta)^- \, \dot\varphi_\beta' = \begin{cases} \frac{1}{\beta_1^2} & \text{for } |\varphi(\beta)| \leq 1, \\ \frac{\varphi(\beta)^2}{\beta_1^2} & \text{for } |\varphi(\beta)| > 1, \end{cases}$$

(see Kitsos (1992)), $b \geq \max\{\frac{1}{\sqrt{2}p}, \frac{1}{\sqrt{2}(1-p)}\}\sqrt{\frac{\pi}{2}} \geq \sqrt{\pi}$ implies

$$\begin{aligned} b(\beta) \;&\geq\; \max\left\{ \frac{|1+\varphi(\beta)|}{|\beta_1|\sqrt{2}}, \frac{|1-\varphi(\beta)|}{|\beta_1|\sqrt{2}} \right\} \sqrt{\pi} \\ &\geq\; \max\left\{ \frac{1}{|\beta_1|}, \frac{|\varphi(\beta)|}{|\beta_1|} \right\} \sqrt{\frac{\pi}{2}} \geq \sqrt{\dot\varphi_\beta \, I(\delta_\beta)^{-1} \, \dot\varphi_\beta'} \sqrt{\frac{\pi}{2}}. \end{aligned}$$

Hence, locally optimal robust estimators with local bias bound $b(\beta)$ at δ_β always exist (see Lemma 1 in Müller (1994a)). Their score functions satisfy (see (4.2) and (2.5))

$$\dot\varphi_\beta \; \psi_{\beta,b(\beta)}(z,t) := \begin{cases} -\operatorname{sgn}(\frac{1+t\,\varphi(\beta)}{\beta_1}) \operatorname{sgn}(z) \, b(\beta) \\ \qquad\qquad \text{for } b(\beta) = \sqrt{\dot\varphi_\beta \, I(\delta_\beta)^{-1} \, \dot\varphi_\beta'} \sqrt{\frac{\pi}{2}} \\ -\operatorname{sgn}(\frac{1+t\,\varphi(\beta)}{\beta_1}) \operatorname{sgn}(z) \, \frac{\min\{|z|, b(\beta)\, y_{b,\beta}\}}{y_{b,\beta}} \\ \qquad\qquad \text{for } b(\beta) > \sqrt{\dot\varphi_\beta \, I(\delta_\beta)^{-1} \, \dot\varphi_\beta'} \sqrt{\frac{\pi}{2}} \end{cases}.$$

Here $y_{b,\beta}$ are the positive solutions of

$$y = (2\Phi(b(\beta)\, y) - 1) \cdot \begin{cases} |\beta_1| & \text{for } |\varphi(\beta)| \leq 1 \\ |\frac{\beta_1}{\varphi(\beta)}| & \text{for } |\varphi(\beta)| > 1 \end{cases},$$

i.e. $y_{b,\beta} = f(a, b(\beta))$, with $a = |\beta_1|$ for $\varphi(\beta) \leq 1$ and $a = |\frac{\beta_1}{\varphi(\beta)}|$ for $\varphi(\beta) > 1$.

The definition of $f(a,b)$ shows that $f(a, \frac{b}{a})/a$ is independent of a so that we have $v_{\frac{b}{a}}(a^2) = v_b(1)$ for all $b \geq \sqrt{\frac{\pi}{2}}$ and $a > 0$. In particular, we have $v_{\frac{b\sqrt{2}}{a}}(a^2) = v_b(2)$ for all

$b \geq \sqrt{\frac{\pi}{2}}$ and $a > 0$. Moreover, $v_b(a)$ as function of b is strictly decreasing in b (see Müller (1994a)), and

$$b(\beta) \leq \begin{cases} \frac{1}{|\beta_1|}\, b\sqrt{2} & \text{for } |\varphi(\beta)| \leq 1 \\ \frac{|\varphi(\beta)|}{|\beta_1|}\, b\sqrt{2} & \text{for } |\varphi(\beta)| > 1 \end{cases},$$

with equality if and only if $|\varphi(\beta)| = 1$. Thus, we get for the asymptotic variances of the locally optimal robust estimators

$$\begin{aligned} &\dot\varphi_\beta \int \psi_{\beta,b(\beta)}\psi'_{\beta,b(\beta)}\, d(P \otimes \delta_\beta)\ \dot\varphi'_\beta \\ &= \begin{cases} \frac{1}{\beta_1^2}\, v_{b(\beta)}(\beta_1^2) & \text{for } |\varphi(\beta)| \leq 1 \\ \frac{\varphi(\beta)^2}{\beta_1^2}\, v_{b(\beta)}(\frac{\beta_1^2}{\varphi(\beta)^2}) & \text{for } |\varphi(\beta)| > 1 \end{cases} \\ &\geq \begin{cases} \frac{1}{\beta_1^2}\, v_{\frac{b\sqrt{2}}{|\beta_1|}}(\beta_1^2) & \text{for } |\varphi(\beta)| \leq 1 \\ \frac{\varphi(\beta)^2}{\beta_1^2}\, v_{\frac{|\varphi(\beta)| b\sqrt{2}}{|\beta_1|}}(\frac{\beta_1^2}{\varphi(\beta)^2}) & \text{for } |\varphi(\beta)| > 1 \end{cases} \\ &= \begin{cases} \frac{1}{\beta_1^2}\, v_b(2) & \text{for } |\varphi(\beta)| \leq 1 \\ \frac{\varphi(\beta)^2}{\beta_1^2}\, v_b(2) & \text{for } |\varphi(\beta)| > 1 \end{cases}, \end{aligned}$$

with equality if and only if $|\varphi(\beta)| = 1$. □

Proof of Theorem 5.1

Set

$$v_1 := \frac{v_b(2p^2)}{v_b(2)} \text{ and } v_{-1} := \frac{v_b(2(1-p)^2)}{v_b(2)}.$$

Then we obtain with (2.7) in Lemma 2.2 and (5.2) in Lemma 5.1 for the inverse of the relative efficiency of the robust estimator $\varphi(\widehat{\beta}_N)$ at the design δ_p

$$\begin{aligned} (6.1)\quad E_\beta(\psi_{b,p}, \delta_p)^{-1} &= \frac{\dot\varphi_\beta \int \psi_{b,p}\psi'_{b,p}\, d(P \otimes \delta_p)\ \dot\varphi'_\beta}{\dot\varphi_\beta \int \psi_{\beta,b(\beta)}\psi'_{\beta,b(\beta)}\, d(P \otimes \delta_\beta)\ \dot\varphi'_\beta} \\ &= \begin{cases} \frac{(1+\varphi(\beta))^2}{4p}\, v_1 + \frac{(1-\varphi(\beta))^2}{4(1-p)}\, v_{-1} & \text{for } |\varphi(\beta)| \leq 1 \\ \frac{1}{\varphi(\beta)^2}[\frac{(1+\varphi(\beta))^2}{4p}\, v_1 + \frac{(1-\varphi(\beta))^2}{4(1-p)}\, v_{-1}] & \text{for } |\varphi(\beta)| > 1 \end{cases} \\ &\leq \max\left\{\frac{v_1}{p}, \frac{v_{-1}}{1-p}\right\} \cdot \frac{1}{4} \cdot \begin{cases} (1+\varphi(\beta))^2 + (1-\varphi(\beta))^2 & \text{for } |\varphi(\beta)| \leq 1 \\ \frac{1}{\varphi(\beta)^2}[(1+\varphi(\beta))^2 + (1-\varphi(\beta))^2] & \text{for } |\varphi(\beta)| > 1 \end{cases} \\ &\leq \max\left\{\frac{v_1}{p}, \frac{v_{-1}}{1-p}\right\}, \end{aligned}$$

with equality if and only if $\varphi(\beta) = 1$ and $\frac{v_1}{p} \geq \frac{v_{-1}}{(1-p)}$, or $\varphi(\beta) = -1$ and $\frac{v_1}{p} \leq \frac{v_{-1}}{(1-p)}$. Because $v_b(a)$ is strictly decreasing in a (see Müller (1994a)) the following inequalities hold:

$$\frac{v_1}{p} = \frac{v_b(2p^2)}{p\, v_b(2)} \geq \frac{2\, v_b(\frac{1}{2})}{v_b(2)} \qquad \text{if and only if} \qquad p \leq \frac{1}{2},$$

and

$$\frac{v_{-1}}{1-p} = \frac{v_b(2(1-p)^2)}{(1-p)\, v_b(2)} \leq \frac{2\, v_b(\frac{1}{2})}{v_b(2)} \qquad \text{if and only if} \qquad p \geq \frac{1}{2}.$$

Hence, $\max\{\frac{v_1}{p}, \frac{v_{-1}}{1-p}\}$ attains its minimum at $p = \frac{1}{2}$. This yields

$$\delta_{\frac{1}{2}} = \arg\max\left\{\min_{\beta \in B}\ E_\beta(\psi_{b,p}, \delta_p);\ p \in (0,1)\right\},$$

and

$$\max\left\{\min_{\beta \in B}\ E_\beta(\psi_{b,p}, \delta_p);\ p \in (0,1)\right\} = \frac{v_b(2)}{2\ v_b(\frac{1}{2})}.$$

Note that $v_1 = 1$ and $v_{-1} = 1$ for $b = \infty$, so that the valuation in (6.1) generalizes the valuation in Kitsos (1992). □

References

[1] BUONACCORSI, J.P. (1986). Design considerations for calibration. *Technometrics* **28**, 149-155.

[2] EISENHART, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. *Ann. Math. Statist.* **10**, 162-186.

[3] ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23**, 255-262.

[4] FORD, I., KITSOS, C.P. and TITTERINGTON, D.M. (1989). Recent advances in nonlinear experimental design. *Technometrics* **13**, 49-60.

[5] HUBER, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.

[6] KITSOS, C.P. (1986). Design and inference for nonlinear problems. *Ph.D. thesis*, University of Glasgow, U.K.

[7] KITSOS, C.P. (1992). Quasi-sequential procedures for the calibration problem. In *COMPSTAT 1992, Vol. 2*, Y. Dodge and J. Whittaker (eds.), 227-231, Physica-Verlag.

[8] KRASKER, W.S. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48**, 1333-1346.

[9] KUROTSCHKA, V. and MÜLLER, Ch.H. (1992). Optimum robust estimation of linear aspects in conditionally contaminated linear models. *Ann. Statist.* **20**, 331-350.

[10] MÜLLER, Ch.H. (1994a). Optimal designs for robust estimation in conditionally contaminated linear models. *J. Statist. Plann. Inference.* **38**, 125-140.

[11] MÜLLER, Ch.H. (1994b). One-step-M-estimators in conditionally contaminated linear models. *Stat. Decis.* **12**, 331-342.

[12] OSBORNE, C. (1991). Statistical calibration: A review. *Int. Statist. Rev.* **59**, 309-336.

[13] PUKELSHEIM, F. and TORSNEY, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Ann. Statist.* **19**, 1614-1625.

[14] SILVEY, S.D. (1980). *Optimal Design.* Chapman and Hall, London.

CHRISTOS P. KITSOS
Athens University of Business and Economics
Department of Statistics
Patission St. 76
104 34 Athens
Greece

CHRISTINE H. MÜLLER
Freie Universität Berlin
1. Mathematisches Institut
Arnimallee 2-6
D-14195 Berlin-Dahlem
Germany