Robust estimation of non-linear aspects

Christos P. Kitsos
Athens University of Business and Economics, Athens, Greece
and
Christine H. Müller
Free University of Berlin, Berlin, Germany

Summary

As a first step for dealing with efficient robust estimation in non-linear models, we regard the problem of efficient robust estimation of non-linear aspects (functions) $\varphi(\beta)$ of the unknown parameter β of a linear model. For robust estimation of a general non-linear aspect we propose estimators which are based on one-step-M-estimators and derive their asymptotic behaviour at the contaminated linear model, where the errors have contaminated normal distributions. The asymptotic behaviour provides criteria for robustness and optimality of the estimators and the corresponding designs. Because it is impossible to find globally optimal robust estimators and designs locally optimal solutions are used for efficiency comparisons. Simple formulas for the efficiency rates are given for the general case. Using these results the efficiency rates for estimating robustly the relative variation of a circadian rhythm are calculated. These efficiency rates are very similar to those for non-robust estimation although on principle there is an important difference.

Key words: Contaminated linear model, non-linear aspect, locally optimal robust estimators, locally optimal designs, efficiency rate, circadian rhythm.

1 Introduction

A general linear model

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

is considered where Y_{nN} are observations, $t_{nN} \in T$ are experimental conditions, $x : T \to \mathbb{R}^p$ is a known "regression" function, $\beta \in B \subset \mathbb{R}^p$ is an unknown parameter vector, Z_{nN} are error variables. In this model a general non-linear aspect $\varphi(\beta)$ of β shall be estimated, where $\varphi : B \to \mathbb{R}^r$ is a given non-linear function.

In classical linear models it is assumed that the error variables $Z_{1N}, ..., Z_{NN}$ are independent and identically distributed and often it is assumed that they are normally distributed with mean 0 and known or unknown variance σ^2, i.e.

$$\frac{1}{\sigma} Z_{nN} \sim P := n_{(0,1)},$$

where $n_{(0,1)}$ denotes the standard normal distribution. But if some outlying observations (gross errors) may appear the normal distribution is not correct and the classical estimators based on least squares estimators may be baised very much.

For estimating the whole unknown parameter vector several approaches were developed to derive outlier robust estimators (see Hampel et al. (1986) and Rousseeuw and Leroy (1987)). One often used approach is the approach of Hampel (1978) and Krasker (1980) who derived the robustness criterion by calculating the influence function of the estimator, where mainly M-estimators were regarded. A similar approach which leads to the same robustness criterion is the more model oriented and more general approach of Bickel (1981, 1984) and Rieder (1985, 1987). They studied the asymptotic behaviour of general asymptotically linear estimators at contaminated linear models and in particular at conditionally contaminated linear models where the errors Z_{nN} can have different contaminated normal distributions for different experimental conditions. The class of asymptotically linear estimators in particular includes one-step-M-estimators which generalize M-estimators. This approach is especially appropiate for designed experiments and for estimation general linear aspects $C\beta$, $C \in I\!R^{r \times p}$, of β as Kurotschka and Müller (1992) showed. Therefore we here also use this model oriented approach to study the robustness properties and efficiency properties of estimators for general non-linear aspects $\varphi(\beta)$ of β. Because some problems in non-linear models can be transformed in a linear model with an interesting non-linear aspect (see also the example in Section 4 of this paper) the approach of this paper may also a first step for regarding the efficiency of robust estimators in non-linear models.

In Section 2 we describe the conditionally contaminated linear model and introduce the class of regarded estimators for the non-linear aspect which are based on one-step-M-estimators. In this section we also derive the asymptotic behaviour of the estimators at the contaminated linear model which provides the robustness and optimality criterion. As for classical non-robust estimation (see for example Silvey (1980), Buonaccorsi et al. (1986), Kitsos (1986, 1992), Kitsos et al. (1988), McDonald and Studden (1990)) it is impossible to find globally optimal designs for robust estimation. Moreover, in opposite to non-robust estimation, it is also impossible to find globally optimal robust estimators. Only locally optimal designs and robust estimators can be derived and used for efficiency comparisons as is shown in Section 3. In this section it is also shown that the locally optimal solutions have a simple form so that general formulas for the efficiency rates can be derived. In Section 4 the efficiency rates are investigated for the problem of estimating the relative variation of a circadian rhythm. The efficiency comparisons show that on principle there is an important difference to those for classical non-robust estimation which were given by Kitsos et al. (1988). But nevertheless the calculated efficiency rates are very similar to those for classical non-robust estimation. This may imply, as we discuss in Section 5, that strategies which were developed for deriving optimal designs for non-robust non-linear estimation problems can be transferred to robust non-linear estimation problems.

2 Robust estimators for non-linear aspects and their asymptotic behaviour

For robust estimation of a non-linear aspect $\varphi(\beta)$ we consider estimators of the form $\widehat{\varphi}_N = \varphi(\widehat{\beta}_N)$ where $\widehat{\beta}_N$ is an estimator for β. An appropriate class of estimators for β which includes the least squares estimator as well as many robust estimators is the class of one-step-M-estimators as were regarded in Bickel (1975) and Müller (1992). These one-step-M-estimators are given by

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

where $d_N := (t_{1N}, ..., t_{NN})'$, $y_N := (y_{1N}, ..., y_{NN})'$ and $\psi : I\!R \times T \to I\!R^p$ is called the score function of the estimator. The initial estimator $\widehat{\beta}_N^0$ and the variance estimator $\widehat{\sigma}_N$ 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)).

For these estimators we now derive their asymptotic behaviour at a conditionally contaminated linear model which was also regarded by Bickel (1981, 1984), Rieder (1987) and Müller (1992, 1994). In such a model it is assumed that the error variables $Z_{1N}, ..., Z_{NN}$ of the linear model are also independent but distributed according to a contaminated normal distribution where the contamination may be different for different experimental conditions. I.e.

$$(2.1) \quad \frac{1}{\sigma} Z_{nN} \sim Q_{nN}(dz) := (1 - N^{-1/2} R\, \epsilon(t_{nN}))\, n_{(0,1)}(dz)$$

$$+ \, N^{-1/2} R \, \epsilon(t_{nN}) \, g(z, t_{nN}) \, n_{(0,1)}(dz)$$

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

To derive the asymptotic behaviour of the estimators we assume that a finite set T_0 and an asymptotic design measure δ exist with $\delta(T_0) = 1$, $\{t_{1N}, ..., t_{NN}\} \subset T_0$ for all $N \in I\!N$, $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e_{t_{nN}}(\{t\}) = \delta(\{t\})$ for all $t \in T_0$ and

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

where e_t denotes the Dirac measure on t and $E_{p\times p}$ the identity matrix. Then under some regularity conditions on the initial and the variance estimator an one-step-M-estimator $\widehat{\beta}_N$ for β is asymptotically normally distributed in the conditionally contaminated linear model (see Müller (1992)). If $\varphi(\beta)$ is continuously differentiable on an open set $B \subset I\!R^p$, where

$$\dot{\varphi}_\beta := \frac{\partial}{\partial\beta}\, \varphi(\beta) \in I\!R^{r\times p}$$

denotes the matrix of partial derivatives of φ with respect to the elements of β, then also $\varphi(\widehat{\beta}_N)$ is asymptotically normally distributed in the conditionally contaminated linear model. I.e. for all $\beta \in B$ and all $(Q^N)_{N\in I\!N} \in \mathcal{P}_R$ we have

$$\text{(2.3)} \qquad \mathcal{L}(\sqrt{N}(\varphi(\widehat{\beta}_N) - \varphi(\beta)) | Q_\beta^N) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(\dot{\varphi}_\beta \, b(\psi, (Q^N)_{N\in I\!N}), \sigma^2 \, V(\psi, \delta, \beta))$$

with

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

and maximum asymptotic bias

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

If the score function of $\widehat{\beta}_N$ satisfies ess $\sup_{P\otimes\delta} |\psi| < \infty$, i.e. $\widehat{\beta}_N$ is robust, then also ess $\sup_{P\otimes\delta} |\dot{\varphi}_\beta\, \psi| < \infty$ so that $\varphi(\widehat{\beta}_N)$ is robust. Moreover we can set without loss of generality $R = 1 = \sigma$.

3 Efficiency of robust estimators and designs

In (2.3) it is shown that the maximum asymptotic bias $\operatorname{ess\,sup}_{P\otimes\delta} |\dot{\varphi}_\beta\, \psi|$ and the asymptotic covariance matrix $V(\psi, \delta, \beta)$ of an estimator $\varphi(\widehat{\beta}_N)$ based on an one-step-M-estimator $\widehat{\beta}_N$ for β with score function ψ 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 $\psi_\infty(z, t) := I(\delta)^{-1} x(t) z$ within all score functions ψ fulfilling condition (2.2), where we have $I(\delta) := \int x(t) x(t)'\, \delta(dt)$. I.e. the covariance matrix $V(\psi_\infty, \delta, \beta)$ is the minimum for all β so that the estimator based on the least squares estimator is asymptotically optimal. (Note that the least squares estimator behaves asymptotically like an one-step-M-estimator with score function ψ_∞.) But if the maximum asymptotic bias should be bounded by some bias bound b, 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 classical non-linear optimal experimental design theory suffer on this β-dependence so that strategies were invented to overcome this problem. See for an overview Ford et al. (1989).

One basic strategy is to compare a given estimator and design with the locally optimal estimators and designs. This provides the efficiency rate of the estimator and the design and this efficiency rate shall be investigated in this section.

For deriving the locally optimal estimators and designs for robust estimation we regard the local behaviour of an estimator $\varphi(\widehat{\beta}_N)$ based on an robust one-step-M-estimator $\widehat{\beta}_N$ for β with score function ψ. Locally, i.e. for given β, this estimator behaves asymptotically like

$$\dot{\varphi}_\beta \widehat{\beta}_N \;\;=\;\; \dot{\varphi}_\beta \widehat{\beta}^0_N + \frac{1}{N}\sum\nolimits_{n=1}^N \dot{\varphi}_\beta \psi \left(\frac{y_{nN}-x(t_{nN})'\widehat{\beta}^0_N(y_N,d_N)}{\widehat{\sigma}_N}, t_{nN} \right) \widehat{\sigma}_N,$$

i.e. like an one-step-M-estimator for the linear aspect $\tilde{\varphi}(\tilde{\beta}) = \dot{\varphi}_\beta \tilde{\beta}$ with score function $\dot{\varphi}_\beta \psi$ (see Müller (1992)). Now we can try to find locally optimal robust estimators within all these robust one-step-M-estimators for the linear aspect $\tilde{\varphi}(\tilde{\beta}) = \dot{\varphi}_\beta \tilde{\beta}$. I.e., given a bias bound b for estimating this linear aspect, we have to find a score function $\psi_{\beta,b,\delta}$ which solves

$$\psi_{\beta,b,\delta} = \arg\min\{\mathrm{tr}\ V(\psi,\delta,\beta);\ \dot{\varphi}_\beta\psi \in \Psi_\delta(\dot{\varphi}_\beta) \text{ with } \operatorname*{ess\,sup}_{P\otimes\delta} |\dot{\varphi}_\beta\psi| \leq b\},$$

where $\Psi_\delta(\dot{\varphi}_\beta)$ is the class of all score functions for estimating the linear aspect $\tilde{\varphi}(\tilde{\beta}) = \dot{\varphi}_\beta \tilde{\beta}$ and in particular satisfying $\int \psi(z,t)x(t)'zP(dz)\delta(dt) = \dot{\varphi}_\beta$.

Similarly locally optimal designs for robust estimation with bias bound b can be defined as solutions $\delta_{\beta,b}$ of

$$\delta_{\beta,b} = \arg\min\{\mathrm{tr}\, V(\psi_{\beta,b,\delta},\delta,\beta);\ \delta \in \Delta\}$$

where Δ is the set of the regarded design measures.

For estimating a general linear aspect $C\beta$ the solutions concerning optimal robust estimators are characterized in Kurotschka and Müller (1992) and the solutions concerning optimals designs for robust estimation are characterized in Müller (1994). In particular in Müller (1994) it is shown that the classical A-optimal designs, which minimize $tr\, C\, I(\delta)^- C'$ (where $I(\delta)^-$ denotes the g-inverse of $I(\delta)$), are also optimal for robust estimation and this holds independently of the bias bound b. Therefore the locally optimal designs for estimation without bias bound are also locally optimal for robust estimation with bias bound b, i.e.

$$\delta_{\beta,b} = \delta_\beta := \ \arg\ \min\{\mathrm{tr}\ \dot{\varphi}_\beta\, I(\delta)^-\dot{\varphi}_\beta';\ \delta \in \Delta\}.$$

Moreover, the score functions

$$\psi_{\beta,b} := \psi_{\beta,b,\delta_\beta}$$

of the locally optimal robust estimators with bias bound b at the locally optimal designs δ_β have a very simple form, namely (see Müller (1994))

$$(3.1)\quad \psi_{\beta,b}(z,t)=\begin{cases} I(\delta_\beta)^- x(t)\,\mathrm{sgn}(z)\sqrt{\frac{\pi}{2}} & \text{for } b=\sqrt{\mathrm{tr}\,\dot\varphi_\beta\, I(\delta_\beta)^-\dot\varphi_\beta'\,\frac{\pi}{2}}\\ I(\delta_\beta)^- x(t)\,\mathrm{sgn}(z)\frac{\min\{|z|,b\,y_{b,\beta}\}}{y_{b,\beta}\sqrt{\mathrm{tr}\,\dot\varphi_\beta\, I(\delta_\beta)^-\dot\varphi_\beta'}} & \text{for } b>\sqrt{\mathrm{tr}\,\dot\varphi_\beta\, I(\delta_\beta)^-\dot\varphi_\beta'\,\frac{\pi}{2}} \end{cases}.$$

Thereby, $y_{b,\beta}$ is the positive solution of $y\sqrt{\mathrm{tr}\ \dot\varphi_\beta\ I(\delta_\beta)^-\ \dot\varphi_\beta'}=2\,\Phi(b\,y)-1$ while Φ denotes the distribution function of the normal distribution $n_{(0,1)}$. Hence, locally optimal robust estimators at locally optimal designs are easy to compute and can be applied for efficiency comparisons.

Now, a given design δ and a given estimator $\varphi(\widehat\beta_N)$ for $\varphi(\beta)$, where $\widehat\beta_N$ is a robust one-step-M-estimator for β with bias bound b and score function ψ, can be compared with the locally optimal robust estimator at the locally optimal design. Because, according to (2.3), the maximum asymptotic bias of $\varphi(\widehat\beta_N)$ is equal to

$$b(\beta):=b(\beta,\psi,\delta):=\operatorname*{ess\,sup}_{P\otimes\delta}|\dot\varphi_\beta\,\psi|,$$

i.e. depends on β, the estimator $\varphi(\widehat\beta_N)$ at the design δ should be only compared with locally optimal robust estimators with the same local bias bound $b(\beta)$ at the locally optimal design δ_β, to make fair comparisons. Hence the asymptotic efficiency measure is defined as

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

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

If δ is an A-optimal design for estimating β then the score function ψ_b of the optimal robust one-step-M-estimator for β at δ with bias bound b has the form (see Müller (1994))

$$(3.2)\quad \psi_b(z,t)=\begin{cases} I(\delta)^{-1}x(t)\,\mathrm{sgn}(z)\sqrt{\frac{\pi}{2}} & \text{for } b=\sqrt{\mathrm{tr}\,I(\delta)^{-1}\,\frac{\pi}{2}}\\ I(\delta)^{-1}x(t)\,\mathrm{sgn}(z)\frac{\min\{|z|,b\,y_b\}}{y_b\sqrt{\mathrm{tr}\,I(\delta)^{-1}}} & \text{for } b>\sqrt{\mathrm{tr}\,I(\delta)^{-1}\,\frac{\pi}{2}} \end{cases},$$

where y_b is the positive solution of $y\sqrt{\text{tr } I(\delta)^{-1}} = 2\,\Phi(b\,y) - 1$. Hence, in this case with the equivalence theorem for linear optimality (see Federov (1972), p. 125), we get for the local bias bounds

$$(3.3) \quad b(\beta) \;=\; \frac{b}{\sqrt{\text{tr} I(\delta)^{-1}}} \operatorname*{ess\,sup}_{\delta} |\dot{\varphi}_\beta \; I(\delta)^{-1}\, x(t)| = \operatorname*{ess\,sup}_{\delta} \frac{|\dot{\varphi}_\beta \; I(\delta)^{-1}\, x(t)|}{|I(\delta)^{-1}\, x(t)|}\; b,$$

where $b(\beta) \geq \sqrt{\text{tr } \dot{\varphi}_\beta \; I(\delta_\beta)^{-} \; \dot{\varphi}_\beta' \, \frac{\pi}{2}}$ for all $b \geq \sqrt{\text{tr } I(\delta)^{-1} \, \frac{\pi}{2}}$. Then the efficiency measure is

$$E_\beta(\psi_b, \delta) = \begin{cases} \frac{g(b(\beta)\, y_{b,\beta})}{y_{b,\beta}^2 \text{ tr } \dot{\varphi}_\beta \; I(\delta)^{-1} \; \dot{\varphi}_\beta'} \; \frac{2}{\pi} & \text{for } b = \sqrt{\text{tr } I(\delta)^{-1} \, \frac{\pi}{2}} \\ \frac{g(b(\beta)\, y_{b,\beta})}{g(b\, y_b)} \cdot \frac{y_b^2}{y_{b,\beta}^2} \cdot \frac{\text{tr } I(\delta)^{-1}}{\text{tr } \dot{\varphi}_\beta \; I(\delta)^{-1} \; \dot{\varphi}_\beta'} & \text{for } b > \sqrt{\text{tr } I(\delta)^{-1} \, \frac{\pi}{2}} \end{cases}.$$

Because of $\lim_{v\to\infty} g(v) = 1$, $\lim_{b\to\infty} y_b = \frac{1}{\sqrt{\text{tr } I(\delta)^{-1}}}$ and $\lim_{b\to\infty} y_{b,\beta} = \frac{1}{\sqrt{\text{tr } \dot{\varphi}_\beta \, I(\delta_\beta)^{-} \, \dot{\varphi}_\beta'}}$ it is easy to see that for $b \to \infty$ the efficiency $E_\beta(\psi_b, \delta)$ tends to

$$(3.4) \quad \frac{\text{tr } \dot{\varphi}_\beta \; I(\delta_\beta)^{-} \; \dot{\varphi}_\beta'}{\text{tr } \dot{\varphi}_\beta \; I(\delta)^{-1} \; \dot{\varphi}_\beta'},$$

i.e. to the design efficiency of δ relative to δ_β as appeared in classical non-robust estimation (see Kitsos et al. (1988)).

4 Example: The rhythmometry problem

Consider the cosinor model of a circadian rhythms which was regarded in Kitsos et al. (1988). In such a model the observations are given by the non-linear model

$$Y_{nN} = \theta_0 + \theta_1 \cos(2\pi t_{nN} + \theta_2) + Z_{nN}$$

with $t_{nN} \in T := [0, 1)$ and θ_0 being the mean, θ_1 the amplitude and θ_2 the phase. In Kitsos et al. (1988) the interest was focussed on estimating the ratio

$$g = g(\theta_0, \theta_1) = \frac{\theta_1}{\theta_0}$$

as it gives a potentially useful measure of relative variation, falling usually in practice within the range $0 < g < 1$.

By reparameterization one gets a linear model

$$\begin{aligned} Y_{nN} &= \beta_0 + \beta_1 \cos(2\pi t_{nN}) + \beta_2 \sin(2\pi t_{nN}) + Z_{nN} \\ &= x(t_{nN})'\beta + Z_{nN} \end{aligned}$$

where $\beta_0 = \theta_0,\ \beta_1 = \theta_1 \cos(\theta_2),\ \beta_2 = -\theta_1 \sin(\theta_2),\ x(t) = (1, \cos(2\pi t), \sin(2\pi t))'$ and $\beta = (\beta_0, \beta_1, \beta_2)'$. Therefore the aspect $g(\theta_0, \theta_1)$ of the non-linear cosinor model is a non-linear aspect

$$\varphi(\beta) := \frac{\sqrt{\beta_1^2 + \beta_2^2}}{\beta_0} = g(\theta_0, \theta_1)$$

of a linear model. Note that $\dot{\varphi}_\beta = \frac{1}{\theta_0}(-\frac{\theta_1}{\theta_0}, \frac{\beta_1}{\theta_1}, \frac{\beta_2}{\theta_1})$.

In Kitsos et al. (1988) is was assumed that the errors Z_{nN} are normally distributed with mean 0 and constant variance σ^2. But there may exist some outlying observations so that the errors have a conditionally contaminated normal distribution of form (2.1). Without loss of generality we can here assume that the variance σ^2 as well as the radius R of the contamination neighbourhood $\mathcal{P}_R$ are equal to 1.

In a conditionally contaminated cosinor model the non-linear aspect $\varphi(\beta)$ should be estimated with a robust estimator, for example with an estimator of the form $\varphi(\widehat{\beta}_N)$, where $\widehat{\beta}_N$ is an optimal robust one-step-M-estimator for β. If the design

$$\delta = \frac{1}{3}(e_0 + e_{\frac{1}{3}} + e_{\frac{2}{3}})$$

is used then this design is A- and D-optimal for estimating β (see Kitsos et al. (1988)) and the score function of the optimal robust one-step-M-estimator for β with bias bound $b > \sqrt{5\frac{\pi}{2}}$ results according to (3.2) in

$$\psi_b(z,t) = (1, 2\cos(2\pi t), 2\sin(2\pi t))' \, \frac{\operatorname{sgn}(z) \min\{|z|, b\, y_b\}}{\sqrt{5} y_b}.$$

Thereby y_b is the positive solution of $y\sqrt{5} = 2\Phi(b\,y) - 1$, i.e. $y_b = 0.216$ for $b = 3$ and $y_b = 0.444$ for $b = 6$, and $\sqrt{5\frac{\pi}{2}} = \sqrt{\operatorname{tr} I(\delta)^{-1}\frac{\pi}{2}}$ is the minimum possible bias bound. Then, according to (3.3) the maximum asymptotic bias of the estimator $\varphi(\widehat{\beta}_N)$ is

$$b(\beta) = \frac{b}{\sqrt{5}\theta_0} \max\left\{\left|\frac{-\theta_1}{\theta_0} + 2\cos(2\pi t + \theta_2)\right|;\ t \in \{0, \tfrac{1}{3}, \tfrac{2}{3}\}\right\}.$$

These estimator should be compared with the locally optimal robust one-step-M-estimators for $\varphi(\beta)$ with local bias bounds $b(\beta)$ at the locally optimal designs.

Locally optimal designs, i.e. A-optimal designs for estimating $\dot{\varphi}_\beta\beta$, are (see Kitsos et al. (1988))

$$\text{(4.1)} \quad \delta_\beta = \frac{1}{2}\left(\left(1 - \frac{\theta_1}{\theta_0}\right) e_{\frac{-\theta_2}{2\pi}} + \left(1 + \frac{\theta_1}{\theta_0}\right) e_{\frac{\pi-\theta_2}{2\pi}}\right)$$

if $0 < \frac{\theta_1}{\theta_0} < 1$. According to (3.1) the score function of the locally optimal robust one-step-M-estimator for $\varphi(\beta)$ with bias bound $b(\beta)$ at δ_β has the form

$$\text{(4.2)} \quad \psi_{\beta,b(\beta)}(z,t) = I(\delta_\beta)^- \begin{pmatrix} 1 \\ \cos(2\pi t) \\ \sin(2\pi t) \end{pmatrix} \frac{\operatorname{sgn}(z)\ \min\{|z|, b(\beta)\,y_{b,\beta}\}\ \theta_0}{y_{b,\beta}},$$

where $y_{b,\beta}$ is the positive solution of $\frac{y}{\theta_0} = 2\Phi(b(\beta)\,y) - 1$. Always this solution will exist because $b(\beta) > \frac{1}{\theta_0}\sqrt{\frac{\pi}{2}}$ if $b > \sqrt{5\frac{\pi}{2}}$, where $\frac{1}{\theta_0}\sqrt{\frac{\pi}{2}} = \sqrt{\operatorname{tr} \dot{\varphi}_\beta\, I(\delta_\beta)^- \dot{\varphi}_\beta' \frac{\pi}{2}}$ is the locally minimum possible bias bound.

Table 1 shows the efficiencies $E_\beta(\psi_b, \delta)$ of the estimator $\varphi(\widehat{\beta}_N)$ for nine different values for θ_1, five different values for θ_2 and three different values for b including the case $b = \infty$, i.e. the classical non-robust case. For θ_0 only the value $\theta_0 = 1$ was used because the efficiencies are independent of θ_0. This is due to the fact that

$$\begin{aligned} b(\beta)\,y_{b,\beta} &= \frac{b}{\sqrt{5}} \max\left\{\left|\frac{-\theta_1}{\theta_0} + 2\cos(2\pi t + \theta_2)\right|;\ t \in \{0, \tfrac{1}{3}, \tfrac{2}{3}\}\right\} \frac{y_{b,\beta}}{\theta_0} \\ &= \frac{b}{\sqrt{5}} \max\left\{\left|\frac{-\theta_1}{\theta_0} + 2\cos(2\pi t + \theta_2)\right|;\ t \in \{0, \tfrac{1}{3}, \tfrac{2}{3}\}\right\} (2\Phi(b(\beta)\,y_{b,\beta}) - 1) \end{aligned}$$

Table 1: Relative efficiencies $E_\beta(\psi_b, \delta)$ and, in brackets, local bias bounds $b(\beta)$ for the rhythmometry problem with $\theta_0 = 1$

	$b=3$					$b=6$		$b=\infty$
θ_1	$\theta_2 = 0$	$\theta_2 = \frac{\pi}{12}$	$\theta_2 = \frac{\pi}{6}$	$\theta_2 = \frac{\pi}{4}$	$\theta_2 = \frac{\pi}{3}$	$\theta_2 = 0$	$\theta_2 = \frac{\pi}{3}$	
0.1	0.414	0.414	0.414	0.414	0.413	0.497		0.498
	(2.549)	(2.458)	(2.458)	(2.726)	(2.817)	(5.098)	(5.635)	
0.2	0.408	0.409	0.408	0.407	0.407	0.490		0.490
	(2.415)	(2.324)	(2.592)	(2.860)	(2.952)	(4.830)	(5.903)	
0.3	0.399	0.399	0.398	0.397	0.397	0.478		0.478
	(2.281)	(2.300)	(2.726)	(2.994)	(3.086)	(4.562)	(6.172)	
0.4	0.388	0.386	0.385	0.385	0.384	0.462		0.463
	(2.147)	(2.434)	(2.860)	(3.129)	(3.220)	(4.293)	(6.440)	
0.5	0.374	0.370	0.369	0.369	0.369	0.444		0.444
	(2.012)	(2.568)	(2.995)	(3.263)	(3.354)	(4.025)	(6.708)	
0.6	0.355	0.352	0.352	0.352	0.352	0.423		0.424
	(2.147)	(2.702)	(3.129)	(3.397)	(3.488)	(4.293)	(6.977)	
0.7	0.335	0.334	0.334	0.333	0.333	0.401		0.402
	(2.281)	(2.837)	(3.263)	(3.531)	(3.622)	(4.562)	(7.245)	
0.8	0.316	0.315	0.315	0.315	0.315	0.378		0.379
	(2.415)	(2.971)	(3.397)	(3.665)	(3.757)	(4.830)	(7.513)	
0.9	0.296	0.296	0.296	0.296	0.296	0.355		0.356
	(2.549)	(3.105)	(3.531)	(3.799)	(3.891)	(5.098)	(7.782)	

and

$$
\begin{aligned}
y_{b,\beta}^2 \operatorname{tr} \dot{\varphi}_\beta I(\delta)^{-1} \dot{\varphi}_\beta' &= \operatorname{tr} \dot{\varphi}_\beta I(\delta_\beta)^{-} \dot{\varphi}_\beta' \quad y_{b,\beta}^2 \quad \frac{\operatorname{tr} \dot{\varphi}_\beta I(\delta)^{-1} \dot{\varphi}_\beta'}{\operatorname{tr} \dot{\varphi}_\beta I(\delta_\beta)^{-} \dot{\varphi}_\beta'} \\
&= \frac{y_{b,\beta}^2}{\theta_0^2} \left(\frac{\theta_1^2}{\theta_0^2} + 2 \right) \\
&= (2\Phi(b(\beta)\, y_{b,\beta}) - 1)^2 \left(\frac{\theta_1^2}{\theta_0^2} + 2 \right)
\end{aligned}
$$

depend only on the ratio $\frac{\theta_1}{\theta_0}$ and the phase θ_2. But for non-robust estimation, i.e. for $b = \infty$, the efficiencies (given by (3.4)) depend only on the ratio $\frac{\theta_1}{\theta_0}$ and not on the mean θ_0 and the phase θ_2 as already Kitsos et al. (1988) have shown. Hence the main difference between robust and non-robust estimation is the additional dependence on the phase θ_2 for robust estimation. For $b < \infty$ Table 1 provides also the values for the local bias bounds $b(\beta)$.

5 Discussion

As Table 1 for the rhythmometry problem shows, the efficiencies for robust estimation ($b < \infty$) are very similar to the efficiencies for classical non-robust estimation ($b = \infty$), in particular for the bias bound $b = 6$ but also for the bias bound $b = 3$ which is near to the minimum possible bias bound of $\sqrt{5\frac{\pi}{2}} = 2.802$. The main difference to non-robust estimation is that the efficiency rates depend additionally on the phase θ_2. But the influence of the phase θ_2 is very low, although also those values θ_2 were used which produce the largest ($\theta_2 = \frac{\pi}{3}$) and the smallest ($\theta_2 = 0$, at least for $\frac{\theta_1}{\theta_0} \geq 0.5$) values for the local bias bounds $b(\beta)$. Even at $b = 6$ the efficiencies were identical at $\theta_2 = 0$ and $\theta_2 = \frac{\pi}{3}$, and this coincides with the behaviour of the efficiencies for non-robust estimation.

From the results for the rhythmometry problem one can expect that for general non-linear problems the efficiencies for robust estimation are very similar to the efficiencies for non-robust estimation. Hence the strategies which were investigated for deriving optimal designs for non-robust non-linear estimation problems as maximin efficient designs, fully sequential designs, batch sequential designs and two-stage designs (see Silvey (1980), Ford et al. (1989), Kitsos (1986, 1989, 1992)) may be also used for robust non-linear estimation problems. These strategies may also provide strategies for deriving particular robust estimators. For example if we have a two-stage design with N_1 observations at the first stage and N_2 observations at the second stage, then we can use in the second stage the locally optimal design $\delta_{\widehat{\beta}^1}$ and the locally optimal robust estimator with score function $\psi_{\widehat{\beta}^1,b}$, where the estimator $\widehat{\beta}^1$ for β is obtained in the first stage.

For the rhythmometry problem a very practicable two-stage design is the following design: In the first stage a four point design of the form $d_4 := (\tau, \tau + \frac{1}{4}, \tau + \frac{1}{2}, \tau + \frac{3}{4})'$ with $\tau \in [0, \frac{1}{4})$ is used. Because of $\cos(2\pi t)^2 + \sin(2\pi t)^2 = 1$ this design is D-optimal for non-robust estimation of β for all $\tau \in [0, \frac{1}{4})$ (see Federov (1972), p.75). Moreover this design is approximately locally optimal for all relevant β and for $\tau = 0$ the least squares estimator $\widehat{\beta}_4^*$, say, for estimating β has a very simple form, namely

$$\widehat{\beta}_4^*((y_1, y_2, y_3, y_4)', d_4) = \begin{pmatrix} \frac{1}{4}(y_1 + y_2 + y_3 + y_4) \\ \frac{1}{2}(y_1 - y_3) \\ \frac{1}{2}(y_2 - y_4) \end{pmatrix}.$$

Basing on this estimator in the second stage the locally optimal design $\delta_{\widehat{\beta}_4^*}$ given by (4.1) and the locally optimal robust one-step-M-estimator with score function $\psi_{\widehat{\beta}_4^*,b}$ given by (4.2) can be used. Note that at the locally optimal design $\delta_{\widehat{\beta}_4^*}$ the whole parameter vector β is not anymore estimable. But this design has besides its greater efficiency the advantage that only at two different experimental conditions the observations are made, which is more practicable than a three or four point design, in particular for medical investigations of a circadian rhythm.

REFERENCES

[1] BICKEL, P.J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association* **70**, 428-434.

[2] BICKEL, P.J. (1981). Quelque aspects de la statistique robuste. In École d'Été de Probabilités de St. Flour. *Springer Lecture Notes in Mathematics* **876**, 1-72.

[3] BICKEL, P.J. (1984). Robust regression based on infinitesimal neighbourhoods. *The Annals of Statistics* **12**, 1349-1368.

[4] BUONACCORSI, J.P. and IYER, H.K. (1986). Optimal designs for ratios of linear combinations in the general linear model. *Journal of Statistical Planning and Inference* **13**, 345-356.

[5] FEDEROV, V.V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

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

[7] 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.

[8] HAMPEL, F.R., RONCHETTI, E.M., ROUSSEEUW, P.J. and STAHEL, W.A. (1986). *Robust Statistics - The Approach Based on Influence Functions.* John Wiley, New York.

[9] HUBER, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics* **35**, 73-101.

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

[11] KITSOS, C.P. (1989). Fully-sequential procedures in nonlinear design problems. *Comp. Statist. & Data Analysis* **8**, 13-19.

[12] 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.

[13] KITSOS, C.P., TITTERINGTON, D.M. and TORSNEY, B. (1988). An optimal design problem in rhythmometry. *Biometrics* **44**, 657-671.

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

[15] 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.

[16] McDONALD, G.C. and STUDDEN, W.J. (1990). Design aspects of regression-based ratio estimation. *Technometrics* **32**, 417-424.

[17] MÜLLER, Ch.H. (1992). One-step-M-estimators in conditionally contaminated linear models. *Preprint No. A-92-11*, Freie Universität Berlin, Fachbereich Mathematik. *To appear in Statistics & Decisions.*

[18] MÜLLER, Ch.H. (1994). Optimal designs for robust estimation in conditionally contaminated linear models. *Journal of Statistical Planning and Inference* **38**, 125-140.

[19] RIEDER, H. (1987). Robust regression estimators and their least favorable contamination curves. *Statistics & Decisions* **5**, 307-336.

[20] ROUSSEEUW, P.J. and LEROY, A.M. (1987). *Robust Regression and Outlier Detection.* John Wiley, New York.

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