ONE-STEP-M-ESTIMATORS IN CONDITIONALLY CONTAMINATED LINEAR MODELS

Christine H. Müller

Abstract

The notion of one-step-M-estimators is generalized for estimating general linear aspects of homoscedastic or heteroscedastic conditionally contaminated linear models with general parametrization and general underlying design. For such contaminated linear models it is shown that the generalized one-step-M-estimators at random and deterministic designs with finite support are asymptotically normally distributed with maximum asymptotic bias and asymptotic covariance matrix only depending on the score function of the one-step-M-estimator and the true variances of the model. This is obtained by deriving their asymptotic linearity under relative simple assumptions which are easy to verify and which admit a wide scope of applications. An example demonstrates the applicability of the results.

1 Introduction

A general linear model

$$X_N = A(d_N)\beta + Z_N$$

is considered where $X_N = (X_{1N}, ..., X_{NN})^T \in \mathbb{R}^N$ is the vector of observations, $Z_N = (Z_{1N}, ..., Z_{NN})^T \in \mathbb{R}^N$ the vector of error variables, $\beta \in \mathbb{R}^p$ an unknown parameter vector, $d_N = (t_{1N}, ..., t_{NN})^T \in T^N$ the vector of experimental conditions, i.e. the design, $a : T \to \mathbb{R}^p$ the vector of known "regression" functions, $A(d_N) = (a(t_{1N}), ..., a(t_{NN}))^T \in \mathbb{R}^{N \times p}$ the design matrix. The set of used experimental conditions T should be finite and remain the same with growing sample size. For deterministic designs d_N we assume that the corresponding design measures $\delta_N = \frac{1}{N} \sum_{n=1}^{N} e_{t_{nN}}$ (e_t denoting the Dirac measure on $t \in T$) converge weakly to

a design measure δ for $N \to \infty$. But besides those deterministic designs we also regard random designs $d_N = (t_{1N}, ..., t_{NN})^T$ where $t_{1N}, ..., t_{NN}$ are realizations of random variables $T_{1N}, ..., T_{NN}$ which are independent and identically distributed according to a design measure δ.

The ideal linear model: For deterministic designs we assume that in the ideal (central) model the error variables $Z_{1N}, ..., Z_{NN}$ have known or unknown variances $\sigma^2(t_{1N}), ..., \sigma^2(t_{NN})$ so that $\frac{1}{\sigma(t_{1N})} Z_{1N}, ..., \frac{1}{\sigma(t_{NN})} Z_{NN}$ are independent and identically distributed according to a symmetric distribution $P = P^{Z_{nN}/\sigma(t_{nN})}$ with mean 0, variance 1 and bounded Lebesgue density $f(z)$. For random designs it is assumed that $(Z_{1N}, T_{1N}), ..., (Z_{NN}, T_{NN})$ are independent and identically distributed where the distribution $P^{Z_{nN}/\sigma(t_{nN})|T_{nN}=t_{nN}}$ of $\frac{1}{\sigma(t_{nN})} Z_{nN}$ given $T_{nN} = t_{nN}$ is a symmetric distribution P with mean 0, variance 1 and bounded Lebesgue density $f(z)$.

But if some outlying observations and gross errors may occur then a conditionally contaminated linear model is more adequate than the ideal model, in particular for designed experiments. The homoscedastic conditionally contaminated model with random designs, which was regarded in Bickel (1981, 1984), Rieder (1985, 1987), Müller (1987) and Kurotschka and Müller (1992), can be generalized to the heteroscedastic case with random designs as well as with deterministic designs as follows.

The conditionally contaminated linear model: For random designs the distribution of $\frac{1}{\sigma(t)} Z_{nN}$ given $T_{nN} = t$ is a contaminated distribution of the form

$$Q_N(dz, t) := P^{Z_{nN}/\sigma(t)|T_{nN}=t}(dz) = (1 + N^{-1/2} Rq(z, t)) P(dz)$$

and for deterministic designs the distribution of $\frac{1}{\sigma(t_{nN})} Z_{nN}$ is a contaminated distribution of the form

$$Q_{nN}(dz) := P^{Z_{nN}/\sigma(t_{nN})}(dz) = (1 + N^{-1/2} Rq(z, t_{nN})) P(dz)$$

where $R > 0$ and

$$q \in \mathcal{Q} := \{q : \mathbb{R} \times T \to \mathbb{R}; \text{ ess sup}_{P \otimes \delta} |q(z, t)| < \infty, \int q(z, \cdot) P(dz) = 0,$$

$$q(\cdot, t) \geq -\epsilon(t) \text{ for all } t \in T \text{ for some } \epsilon : T \to [0, \infty] \text{ with } \int \epsilon d\delta \leq 1\}.$$

All sequences $(Q^N)_{N \in \mathbb{N}} = ((Q_N \otimes \delta)^N)_{N \in \mathbb{N}}$ and $(Q^N)_{N \in \mathbb{N}} = ((Q_{1N} \otimes ... \otimes Q_{NN}))_{N \in \mathbb{N}}$, respectively, define the conditional contamination neighbourhood $\mathcal{P}_R$ with radius R around the ideal model $((P \otimes \delta)^N)_{N \in \mathbb{N}}$ and $(P^N)_{N \in \mathbb{N}}$, respectively.

For such general contaminated linear models we propose in Section 2 generalized one-step-M-estimators for estimating a linear aspect $\varphi(\beta) = C\beta$, $C \in \mathbb{R}^{r \times p}$, of the unknown paramenter vector β. In Section 3 we show the asymptotic normality of these estimators

by extending Bickel's (1975) proof of asymptotic linearity of one-step Huber estimators at the ideal model. Here the asymptotic linearity of the estimators is only used for deriving the asymptotic normality of the estimators at the conditionally contaminated linear model because it provides the relevant robustness criteria for designed experiments (see Kurotschka and Müller (1992) and Müller (1992)). But the asymptotic linearity of the estimators may also have other applications in particular for other neighbourhoods of the ideal model as were regarded by Rieder (1980, 1985, 1987, 1993). The asymptotic linearity of the estimators is shown under simple assumptions which are easy to verify and which include all relevant cases in particular for optimal robust estimation at the conditionally contaminated model. Thereby, in opposite to Rieder (1980, 1985, 1993) and Schick (1987) we do not assume that the initial estimator of the one-step-M-estimator is discretized, and in opposite to Simpson et al. (1992) we also allow score functions of the one-step-M-estimators which are unbounded or not continuous. In particular we allow score functions basing on sign functions which play an important role for optimal robust estimation. An example in Section 4 demonstrates the applicability of the results in particular for optimal robust estimation. All proofs are given in Section 5.

2   One-step-M-estimators

We regard one-step-M-estimators because they are very general, their calculation is easy and they easily can be extended for estimation of arbitrary linear aspects $\varphi(\beta) = C\beta$ of β (see Müller (1987) and Kurotschka and Müller (1992)). To include also the case that the variances $\sigma^2(t)$ may be unknown and different here one-step-M-estimators are defined as follows.

DEFINITION 2.1 *An estimator $\widehat{\varphi}_N : I\!\!R^N \times T^N \to I\!\!R^r$ is called an one-step-M-estimator for $\varphi(\beta) = C\beta$ with score function $\psi : I\!\!R \times T \to I\!\!R^r$, initial estimator $\widehat{\beta}_N : I\!\!R^N \times T^N \to I\!\!R^p$ for β and variance estimators $\widehat{\sigma}_N^2(t) : I\!\!R^N \times T^N \to I\!\!R^+$ for $\sigma^2(t)$, $t \in T$, iff*

$$\widehat{\varphi}(x_N, d_N) = C\widehat{\beta}_N(x_N, d_N)$$
$$+ \frac{1}{N} \sum_{n=1}^{N} \psi\left(\frac{x_{nN} - a(t_{nN})^T \widehat{\beta}_N(x_N, d_N)}{\widehat{\sigma}_N(t_{nN})(x_N, d_N)}, t_{nN} \right) \widehat{\sigma}_N(t_{nN})(x_N, d_N).$$

Note that these one-step-M-estimators are scale equivariant if the variance estimators and the initial estimator are scale equivariant. Possible scale equivariant initial and variance estimators are the least squares estimators and the corresponding empirical variance estimators. But to ensure good robustness properties, also at small sample sizes, better initial estimators and variance estimators are scale equivariant robust estimators, for example high breakdown point estimators. But for the asymptotic considerations of this paper only the following conditions on the initial estimator and the variance estimators are relevant:

(2.1) $\sqrt{N}(a(\tau_1), ..., a(\tau_I))^T(\widehat{\beta}_N - \beta)$ is tight at the ideal model where $T = \{\tau_1, ..., \tau_I\}$.

(2.2) $\sqrt{N}(\widehat{\sigma}_N(t) - \sigma(t))$ is tight at the ideal model for all $t \in T$.

Condition (2.1) is due to the possibility that β itself may be not identifiable.

For the score functions the following simple conditions are assumed (compare the conditions given in Maronna and Yohai (1981)):

(2.3) $\psi(z,t) = \psi_0(z,t) + c(t)\mathrm{sgn}(z)$ where for all $t \in T$
          $\psi_0(\cdot,t)$ is antisymmetric, continuous and there exists $l_1(t), ..., l_L(t)$ so that
          $\psi_0(\cdot,t)$ has bounded and continuous derivatives on $I\!R \setminus \{l_1(t), ..., l_L(t)\}$,

(2.4) $\int \psi(z,t)a(t)^T z P(dz)\delta(dt) = C.$

Condition (2.3) can be generalized as Lemma 2.1 below shows but in this form it is easy to check. The condition (2.4) is due to the aspect which shall be estimated and can be often fulfilled for a given score function by multiplying with a suitable matrix. For example the score functions of the Hampel-Krasker estimator (see Hampel (1978), Krasker (1980)), of the Krasker-Welsch estimator (see Krasker and Welsch (1982)) and of all optimal robust estimators for linear aspects characterized in Müller (1987), Kurotschka and Müller (1992) and Müller (1992) fulfill condition (2.3) and (2.4). In particular the score functions of all robust estimators with minimum asymptotic bias fulfill these conditions.

LEMMA 2.1 *If $\psi = (\psi_1, ..., \psi_r)^T : I\!R \times T \to I\!R^r$ satisfies condition (2.3) then*

(2.5) $\lambda(b,s,t,\sigma) := \int \psi\left(\dfrac{z\sigma - a(t)^T b}{\sigma + s}, t\right)(\sigma + s)P(dz)$ *is for all $t \in T, \sigma \in I\!R^+$*
          continuously differentiable in a neighbourhood of $(b,s) = (0,0)$ where

$$\frac{\partial}{\partial(b,s)}\lambda(b,s,t,\sigma)/_{(b,s)=(0,0)} = (-\int \psi(z,t)a(t)^T z P(dz) \mid 0_{r\times 1}),$$

and there exists $K \in I\!R^+$ and antisymmetric and monotone increasing functions $\psi_j^+(\cdot,t) : I\!R \to I\!R$ and $\psi_j^-(\cdot,t) : I\!R \to I\!R$ with

(2.6) $\psi_j(\cdot,t) = \psi_j^+(\cdot,t) - \psi_j^-(\cdot,t),$

(2.7) $\psi_j^\pm\left(\dfrac{z}{\sigma+s}, t\right)(\sigma+s)$ *is as function of s continuous on $\left[\dfrac{-\sigma}{4}, \dfrac{\sigma}{4}\right]$ for all $z \in I\!R$,*

(2.8) $\int \psi_j^\pm(z,t)^2 P(dz) < \infty,$

(2.9) $\int \left|\psi_j^\pm\left(z + \dfrac{k+h}{\sigma}, t\right) - \psi_j^\pm\left(z + \dfrac{k}{\sigma}, t\right)\right|^2 P(dz) \le K\dfrac{|h|}{\sigma},$

(2.10) $\left|\int \left(\psi_j^\pm\left(\dfrac{z\sigma + k + h}{\sigma + s}, t\right) - \psi_j^\pm\left(\dfrac{z\sigma + k}{\sigma + s}, t\right)\right) P(dz)\right| \le K\dfrac{|h|}{\sigma},$

(2.11) $\int \left|\psi_j^\pm\left(\dfrac{z\sigma + h}{\sigma + s + \rho}, t\right)(\sigma + s + \rho) - \psi_j^\pm\left(\dfrac{z\sigma + h}{\sigma + s}, t\right)(\sigma + s)\right|^2 P(dz) \le K|\rho|^2,$

for all $j \in \{1, ..., r\}$, $t \in T$, $k \in I\!R$, $\sigma \in I\!R^+$ and $h, s, \rho \in [\frac{-\sigma}{4}, \frac{\sigma}{4}]$.

3 Asymptotic linearity and normality

The following theorem shows the asymptotic linearity of the generalized one-step-M-estimators. To formulate this property for random designs as well as for deterministic designs let

$$\psi_{nN}(b,s) := \psi\left(\frac{X_{nN} - a(t_{nN})^T(\beta + b)}{\sigma(t_{nN}) + s}, t_{nN}\right)(\sigma(t_{nN}) + s) \text{ for deterministic designs}$$

and

$$\psi_{nN}(b,s) := \psi\left(\frac{X_{nN} - a(T_{nN})^T(\beta + b)}{\sigma(T_{nN}) + s}, T_{nN}\right)(\sigma(T_{nN}) + s) \text{ for random designs.}$$

THEOREM 3.1 *If $\widehat{\varphi}_N$ is an one-step-M-estimator for $\varphi(\beta) = C\beta$ with score function ψ satisfying conditions (2.3) and (2.4), or (2.4), (2.5) - (2.11), initial estimator $\widehat{\beta}_N$ satisfying condition (2.1) and variance estimators $\widehat{\sigma}_N(t)$ satisfying condition (2.2) then*

$$P(|\sqrt{N}(\widehat{\varphi}_N - \varphi(\beta)) - \frac{1}{N}\sum_{n=1}^{N}\psi_{nN}(0,0)| > \epsilon) \overset{N\to\infty}{\longrightarrow} 0 \text{ for all } \epsilon > 0 \text{ and all } \beta \in I\!\!R^p$$

at the ideal model.

The asymptotic linearity of the one-step-M-estimators implies their asymptotic normality at the conditionally contaminated linear model. As in Bickel (1981), Rieder (1985, 1987), Müller (1987) and Kurotschka and Müller (1992) this follows immediately by the Third Lemma of LeCam (see Hájek and Šidák (1967) p.208). Only for deterministic designs additionally Lindeberg's condition and weak convergence of the design measures δ_N have to be used.

THEOREM 3.2 *Let $\widehat{\varphi}_N$ be an one-step-M-estimator for $\varphi(\beta) = C\beta$ with score function ψ satisfying conditions (2.3) and (2.4), initial estimator $\widehat{\beta}_N$ satisfying condition (2.1) and variance estimators $\widehat{\sigma}_N(t)$ satisfying condition (2.2). Then*

$$\mathcal{L}\left(\sqrt{N}(\widehat{\varphi}_N - \varphi(\beta))|Q^N\right)$$
$$\overset{N\to\infty}{\longrightarrow} \mathcal{N}\left(R\int \psi(z,t)\sigma(t)q(z,t)P(dz)\delta(dt), \int \psi(z,t)\psi(z,t)^T\sigma(t)^2 P(dz)\delta(dt)\right)$$
for all $(Q^N)_{N\in I\!\!N} \in \mathcal{P}_R$,

with maximum asymptotic bias

$$\sup\left\{R\int \psi(z,t)\sigma(t)q(z,t)P(dz)\delta(dt); q \in \mathcal{Q}\right\} = R \operatorname*{ess\,sup}_{P\otimes\delta} |\psi(z,t)\sigma(t)|.$$

Robust one-step-M-estimators should have a bounded maximum asymptotic bias. From Theorem 3.2 it is clear that this is only possible if the score function itself is bounded. Optimality criteria for robust one-step-M-estimators can be derived from the asymptotic covariance matrix. See for the homoscedastic case with known variance and random design Bickel (1981, 1984), Rieder (1985, 1987), Müller (1987) and Kurotschka and Müller (1992). In particular Theorem 3.2 shows that the results in Müller (1987, 1992) concerning optimal random designs for robust estimation in homoscedastic normal linear models with known variances are also true for homoscedastic normal models with unknown variances and for deterministic designs.

4   Example

Consider an one-way classification model with three treatments 1,2,3 where the observations are given by

$$X(t) = 1_{\{1\}}(t)\beta_1 + 1_{\{2\}}(t)\beta_2 + 1_{\{3\}}(t)\beta_3 + Z(t) = a(t)^T\beta + Z(t)$$

with $t \in \{1,2,3\} = T$, $a(t) = (1_{\{1\}}(t), 1_{\{2\}}(t), 1_{\{3\}}(t))^T$ and $\beta = (\beta_1, \beta_2, \beta_3)^T$. Assume that at the ideal model the error variables $Z(t)$ are normal distributed with mean 0 and unknown variance $\sigma(t)^2$. For estimating the linear aspect $\varphi(\beta) = (\beta_1, \beta_2 - \beta_1, \beta_3 - \beta_1)^T$ a simple design measure is $\delta = \frac{1}{3}(e_1 + e_2 + e_3)$. A deterministic design sequence $(d_N)_{N \in \mathbb{N}}$ for which the corresponding design measures $(\delta_N)_{N \in \mathbb{N}}$ converges weakly to this design measure δ is for example given by $d_3 = (1,2,3)$, $d_4 = (1,2,3,1)$, $d_5 = (1,2,3,1,2)$, $d_6 = (1,2,3,1,2,3)$, $d_7 = (1,2,3,1,2,3,1)$ and so forth.

An optimal most robust score function for estimating $\varphi(\beta)$ at the design measure δ in the homoscedastic model is (see Müller (1987), Kurotschka and Müller (1992))

$$\psi(z,t) = \begin{cases} (1,-1,-1)^T \operatorname{sgn}(z)\, 3\,\sqrt{\frac{\pi}{2}} & \text{for } t = 1, \\[2mm] (0,1,0)^T \operatorname{sgn}(z)\, \frac{\min\{|z|,b\,y\}}{y} & \text{for } t = 2, \\[2mm] (0,0,1)^T \operatorname{sgn}(z)\, \frac{\min\{|z|,b\,y\}}{y} & \text{for } t = 3, \end{cases}$$

with $b = 3\sqrt{\frac{3\pi}{2}}$ and $y = (2\Phi(by) - 1)^{\frac{1}{3}} \approx 0.32$ where Φ denotes the standard normal distribution function. This score function satisfies condition (2.3) and (2.4) so that an one-step-M-estimator with this score function, with initial estimator satisfying (2.1) and variance estimators satisfying (2.2) is according to Theorem 3.1 asymptotically linear. Moreover, according to Theorem 3.2, at the conditionally contaminated model with radius R this estimator is also asymptotically normally distributed with a maximum asymptotic bias equal to $R \cdot 3\sqrt{\frac{3\pi}{2}} \cdot \max\{\sigma(1), \sigma(2), \sigma(3)\}$. Another score function which also satisfies condition (2.3) and (2.4) and is optimal for robust estimation with bias bound $2b$ (see Müller (1987),

Kurotschka and Müller (1992)) is given by

$$
\psi(z,t) = \begin{cases}
(1,-1,-1)^T \, \mathrm{sgn}(z) \, \frac{\min\{|z|,2\,b\,v\}}{\sqrt{3}\,v} & \text{for } t = 1, \\[2mm]
(0,1,0)^T \, \mathrm{sgn}(z) \, \frac{\min\{|z|,2\,b\,w\}}{w} & \text{for } t = 2, \\[2mm]
(0,0,1)^T \, \mathrm{sgn}(z) \, \frac{\min\{|z|,2\,b\,w\}}{w} & \text{for } t = 3,
\end{cases}
$$

where again $b = 3\sqrt{\frac{3\pi}{2}}$ and $v = (2\Phi(2\,b\,v) - 1)\frac{1}{3\sqrt{3}} \approx 0.19$ and $w = (2\Phi(2\,b\,w) - 1)\frac{1}{3} \approx 0.33$. At the conditionally contaminated model an one-step-M-estimator with this score function is asymptotically normally distributed with maximum asymptotic bias equal to $R \cdot 6\sqrt{\frac{3\pi}{2}} \cdot \max\{\sigma(1), \sigma(2), \sigma(3)\}$.

5 Proofs

Proof of Lemma 2.1

Set $\psi_*(z,t) = c(t)\mathrm{sgn}(z)$ then $\psi = \psi_0 + \psi_*$ and condition (2.5) follows for ψ_* by integrating and for ψ_0 by exchanging integration and differentiation. Condition (2.3) also implies that the increasings parts $\psi_{0j}^{\pm}$ of the components ψ_{0j} of ψ_0 are antisymmetric, continuous and differentiable with bounded and continuous derivative on $I\!\!R \setminus \{l_1(t), ..., l_L(t)\}$, i.e. there exists $K_1 > 1$ with

$$
|\psi_{0j}^{\pm}(z + h, t) - \psi_{0j}^{\pm}(z, t)| \leq K_1 |h|
$$

and

$$
|\psi_{0j}^{\pm}(z, t)| \leq K_1 |z|
$$

for all $j \in \{1, ..., r\}$, $z, h \in I\!\!R$, $t \in T$. Setting $K_0 = 8K_1^2$ then for $\psi = \psi_0$ conditions (2.7) - (2.10) are obviously fulfilled and condition (2.11) follows from

$$
\int \left| \psi_{0j}^{\pm}\left(\frac{z\sigma + h}{\sigma + s + \rho}, t\right)(\sigma + s + \rho) - \psi_{0j}^{\pm}\left(\frac{z\sigma + h}{\sigma + s}, t\right)(\sigma + s) \right|^2 P(dz)
$$

$$
\leq |\rho|^2 \int \left(\left|\frac{z\sigma + h}{\sigma + s}\right| K_1 + \left|\psi_{0j}^{\pm}\left(\frac{z\sigma + h}{\sigma + s}, t\right)\right| \right)^2 P(dz)
$$

$$
\leq |\rho|^2 \, 4\, K_1^2 \, \frac{\sigma^2 + h^2}{(\sigma + s)^2} \leq |\rho|^2 K_0.
$$

For $\psi = \psi_*$ the conditions (2.9) - (2.11) are fulfilled for $K_* = 2\max\{1, \sup\{f(z); z \in I\!\!R\}\} \cdot \max\{c, c^2\}$ where f is the Lebesgue density of P and $c = \max\{|c(t)|; t \in T\}$. Note that for $h \geq 0$

$$
\int \left| \mathrm{sgn}\left(\frac{z\sigma + k + h}{\sigma + s}\right) - \mathrm{sgn}\left(\frac{z\sigma + k}{\sigma + s}\right) \right| P(dz)
$$

$$= \int \left| \operatorname{sgn}\left(\frac{z\sigma + k + h}{\sigma}\right) - \operatorname{sgn}\left(\frac{z\sigma + k}{\sigma}\right) \right|^2 P(dz)$$

$$= 2\left(P\left(\left(-\infty, \frac{-k}{\sigma}\right]\right) - P\left(\left(-\infty, \frac{-(k+h)}{\sigma}\right]\right) \right) \leq 2\sup\{f(z); z \in I\!R\}\frac{h}{\sigma}.$$

For general $\psi = \psi_0 + \psi_*$ conditions (2.8) - (2.11) are fulfilled for $K = 4\max\{K_0, K_*\}$ using Hölder inequality for condition (2.8), (2.9) and (2.11).

Proof of Theorem 3.1

If $T = \{\tau_1, ..., \tau_I\}$ and $(a(\tau_1), ..., a(\tau_I))$ is not of full rank then one can regard a transformed model $X_{nN}(t_{nN}) = \overline{a}(t_{nN})^T\overline{\beta} + Z_{nN}$ where $\overline{\beta}$ is identifiable and $(\overline{a}(\tau_1), ..., \overline{a}(\tau_I))$ is of full rank (see Müller (1987) and Kurotschka and Müller (1992)). Therefore without loss of generality one can assume that $(a(\tau_1), ..., a(\tau_I))$ is of full rank and that condition (2.1) implies tightness of $\sqrt{N}(\widehat{\beta}_N - \beta)$ itself.

Setting

$$\xi(t) := \int \psi(z, t)a(t)^T z P(dz),$$

$$\begin{aligned} \psi_{nNt}(b, s) &:= \psi_{nN}(b, s)1_{\{t\}}(t_{nN}) \qquad \text{for deterministic designs,} \\ \psi_{nNt}(b, s) &:= \psi_{nN}(b, s)1_{\{t\}}(T_{nN}) \qquad \text{for random designs,} \end{aligned}$$

and

$$\begin{aligned} \lambda_{nNt}(b, s) &:= E(\psi_{nNt}(b, s)) \\ &= \begin{cases} \lambda(b, s, t, \sigma(t))1_{\{t\}}(t_{nN}) & \text{for deterministic designs,} \\ \lambda(b, s, t, \sigma(t))\delta(\{t\}) & \text{for random designs,} \end{cases} \end{aligned}$$

one gets with condition (2.4)

$$\left| \sqrt{N}\left(\widehat{\varphi}_N - \varphi(\beta) - \frac{1}{N}\sum_{n=1}^N \psi_{nN}(0, 0) \right) \right|$$

$$(5.12) \quad \leq \quad \sum_{t \in T}\left| \sqrt{N}\,\xi(t)\,\delta(\{t\})\,(\widehat{\beta}_N - \beta) \right.$$

$$\left. + \frac{1}{\sqrt{N}}\sum_{n=1}^N (\lambda_{nNt}(\widehat{\beta}_N - \beta, \widehat{\sigma}_N(t) - \sigma(t)) - \lambda_{nNt}(0, 0)) \right|$$

$$(5.13) \quad + \sum_{t \in T}\left| \frac{1}{\sqrt{N}}\sum_{n=1}^N (\psi_{nNt}(\widehat{\beta}_N - \beta, \widehat{\sigma}_N(t) - \sigma(t)) - \psi_{nNt}(0, 0) \right.$$

$$\left. - \lambda_{nNt}(\widehat{\beta}_N - \beta, \widehat{\sigma}_N(t) - \sigma(t)) + \lambda_{nNt}(0, 0)) \right|.$$

The first term (5.12) is for deterministic designs less than

$$\sum_{t \in T}\left| \xi(t)\left(\delta(\{t\}) - \frac{1}{N}\sum_{n=1}^N 1_{\{t\}}(t_{nN}) \right)\sqrt{N}(\widehat{\beta}_N - \beta) \right|$$

$$+ \sum_{t \in T}\left(\frac{1}{N}\sum_{n=1}^N 1_{\{t\}}(t_{nN}) \right) \cdot \sqrt{N}\left| \begin{pmatrix} \widehat{\beta}_N - \beta \\ \widehat{\sigma}_N(t) - \sigma(t) \end{pmatrix} \right| \cdot \frac{|\Lambda(t)|}{\left| \begin{pmatrix} \widehat{\beta}_N - \beta \\ \widehat{\sigma}_N(t) - \sigma(t) \end{pmatrix} \right|}$$

and for random designs less than

$$\sum_{t \in T} \delta(\{t\}) \cdot \sqrt{N} \left| \left(\begin{matrix} \widehat{\beta}_N - \beta \\ \widehat{\sigma}_N(t) - \sigma(t) \end{matrix} \right) \right| \cdot \frac{|\Lambda(t)|}{\left| \left(\begin{matrix} \widehat{\beta}_N - \beta \\ \widehat{\sigma}_N(t) - \sigma(t) \end{matrix} \right) \right|}$$

where

$$\Lambda(t) := \xi(t) \, (\widehat{\beta}_N - \beta) + \lambda(\widehat{\beta}_N - \beta, \widehat{\sigma}_N(t) - \sigma(t), t, \sigma(t)) - \lambda(0, 0, t, \sigma(t)).$$

Thus because of condition (2.5), (2.1) and (2.2) and because of the weak convergence of $\delta_N = \frac{1}{N} \sum_{n=1}^{N} e_{t_{nN}}$ the first term (5.12) converges for deterministic designs as well as for random designs to zero.

To show that the second term (5.13) also converges to zero one can show this for the increasing components ψ_j^+ and ψ_j^- separately, i.e. one can assume without loss of generality that for every $t \in T$ the function $\psi(\cdot, t)$ is monotone increasing with values in $\mathbb{R}$. Although ψ may be not of the form $\psi(z, t) = c(t)\eta(z)$ with $\eta : \mathbb{R} \to \mathbb{R}$ the proof now follows along the lines of the proof in Bickel (1975):

Because $\sqrt{N}(\widehat{\beta}_N - \beta)$ is tight there exists for every $\gamma > 0$ a $k(\gamma) > 0$ with $P(|\widehat{\beta}_N - \beta|_\infty > \frac{k(\gamma)}{\sqrt{N}}) < \frac{\gamma}{7}$ where $|\beta|_\infty$ denotes the maximum of the absolute values of the coordinates of β. The cube $[-\frac{k(\gamma)}{\sqrt{N}}, \frac{k(\gamma)}{\sqrt{N}}]^p$ can be discretized by $b_N^1, ..., b_N^{L(\gamma)}$ so that for every b from the cube there exists b_N^l with $|b_N^l - b| \le \frac{\gamma}{\sqrt{N}}$. Setting for deterministic designs

$$\nu_{nNt}(b, h, s) := \psi \left(\frac{X_{nN} - a(t)^T(\beta + b) + h}{\sigma(t) + s}, t \right) (\sigma(t) + s) 1_{\{t\}}(t_{nN})$$

and for random designs

$$\nu_{nNt}(b, h, s) := \psi \left(\frac{X_{nN} - a(t)^T(\beta + b) + h}{\sigma(t) + s}, t \right) (\sigma(t) + s) 1_{\{t\}}(T_{nN})$$

and

$$S_{Nt}(b, h, s) := \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \left(\nu_{nNt}(b, h, s) - E(\nu_{nNt}(b, h, s)) \right)$$

one gets for every b with $|b_N^l - b| \le \frac{\gamma}{\sqrt{N}}$, because ψ is monotone increasing,

$$|\nu_{nNt}(b_N^l, 0, s) - \nu_{nNt}(b, 0, s)| \le \nu_{nNt}(b_N^l, h_{Nt}, s) - \nu_{nNt}(b_N^l, -h_{Nt}, s)$$

for all $s \in \mathbb{R}$ where $h_{Nt} = |a(t)| \frac{\gamma}{\sqrt{N}}$. This implies

$$|S_{Nt}(0, 0, 0) - S_{Nt}(b, 0, s)|$$
$$\le |S_{Nt}(0, 0, 0) - S_{Nt}(b_N^l, 0, 0)| + |S_{Nt}(b_N^l, 0, 0) - S_{Nt}(b_N^l, 0, s)|$$
$$+ |S_{Nt}(b_N^l, h_{Nt}, s) - S_{Nt}(b_N^l, h_{Nt}, 0)| + |S_{Nt}(b_N^l, h_{Nt}, 0) - S_{Nt}(b_N^l, -h_{Nt}, 0)|$$
$$+ |S_{Nt}(b_N^l, -h_{Nt}, 0) - S_{Nt}(b_N^l, -h_{Nt}, s)|$$
$$+ \frac{2}{\sqrt{N}} \sum_{n=1}^{N} |E(\nu_{nNt}(b_N^l, h_{Nt}, s) - \nu_{nNt}(b_N^l, -h_{Nt}, s))|.$$

Condition (2.10) implies for every $s \in [-\frac{\sigma(t)}{4}, \frac{\sigma(t)}{4}]$

(5.14) $\dfrac{2}{\sqrt{N}} \sum_{n=1}^{N} |E(\nu_{nNt}(b_N^l, h_{Nt}, s) - \nu_{nNt}(b_N^l, -h_{Nt}, s))|$

$\qquad \leq 4\sqrt{N} K |h_{Nt}| \dfrac{\sigma(t) + s}{\sigma(t)} \leq 5K |a(t)| \gamma.$

If I is the number of experimental conditions t in T one gets with Chebyshev inequality and condition (2.9) for every $\epsilon > 0$

(5.15) $P(|S_{Nt}(0,0,0) - S_{Nt}(b_N^l, 0, 0)| > \dfrac{\epsilon}{6I})$

$\qquad \leq \dfrac{6^2 I^2}{\epsilon^2} K \sigma(t) |a(t)^T b_N^l| \leq \dfrac{6^2 I^2}{\epsilon^2} K \sigma(t) |a(t)| \sqrt{p}\, k(\gamma) \dfrac{1}{\sqrt{N}},$

(5.16) $P(|S_{Nt}(b_N^l, h_{Nt}, 0) - S_{Nt}(b_N^l, -h_{Nt}, 0)| > \dfrac{\epsilon}{6I})$

$\qquad \leq \dfrac{6^2 I^2}{\epsilon^2} K \sigma(t) 2 |h_{Nt}| \leq \dfrac{6^2 I^2}{\epsilon^2} K \sigma(t) 2 |a(t)| \gamma \dfrac{1}{\sqrt{N}}$

and for every $h_0 \in \mathbb{R}$

(5.17) $P(|S_{Nt}(b_N^l, \frac{h_0}{\sqrt{N}}, 0) - S_{Nt}(0,0,0)| > \epsilon)$

$\qquad \leq \dfrac{1}{\epsilon^2} K \sigma(t) (|a(t)^T b_N^l| + |\frac{h_0}{\sqrt{N}}|)$

$\qquad \leq \dfrac{1}{\epsilon^2} K \sigma(t) (|a(t)| \sqrt{p}\, k(\gamma) + |h_0|) \dfrac{1}{\sqrt{N}}.$

Furthermore condition(2.11) provide for every $\rho \in [-\frac{\sigma(t)}{4}, \frac{\sigma(t)}{4}]$ and every $h_0 \in \mathbb{R}$

(5.18) $E(|S_{Nt}(b_N^l, \frac{h_0}{\sqrt{N}}, s) - S_{Nt}(b_N^l, \frac{h_0}{\sqrt{N}}, s + \rho)|^2) \leq K |\rho|^2.$

Because of condition (2.8) the sequence $S_{Nt}(0,0,0)$, $N \in \mathbb{N}$, is tight so that with (5.17) and (5.18) for all t, l and h_0 the sequence of stochastic processes $S_{Nt}(b_N^l, \frac{h_0}{\sqrt{N}}, \cdot) : [-\frac{\sigma(t)}{4}, \frac{\sigma(t)}{4}] \to \mathbb{R}$, $N \in \mathbb{N}$, is tight (see Billingsley (1968) p.95). Therefore for every $\epsilon > 0$ and every $\gamma > 0$ there exist $N_{\epsilon,\gamma} \in \mathbb{N}$ and $\eta_{\epsilon,\gamma}$ with $\min\{\frac{\sigma(t)}{4} ; t \in T\} \geq \eta_{\epsilon,\gamma} > 0$ so that

$\qquad P(\sup\{|S_{Nt}(b_N^l, 0, 0) - S_{Nt}(b_N^l, 0, s)| ; |s| \leq \eta_{\epsilon,\gamma}\} > \frac{\epsilon}{6I}) < \frac{\gamma}{7 \cdot I \cdot L(\gamma)},$

$\qquad P(\sup\{|S_{Nt}(b_N^l, h_{Nt}, 0) - S_{Nt}(b_N^l, h_{Nt}, s)| ; |s| \leq \eta_{\epsilon,\gamma}\} > \frac{\epsilon}{6I}) < \frac{\gamma}{7 \cdot I \cdot L(\gamma)},$

$\qquad P(\sup\{|S_{Nt}(b_N^l, -h_{Nt}, 0) - S_{Nt}(b_N^l, -h_{Nt}, s)| ; |s| \leq \eta_{\epsilon,\gamma}\} > \frac{\epsilon}{6I}) < \frac{\gamma}{7 \cdot I \cdot L(\gamma)}$

and

$\qquad P(|\hat{\sigma}_N(t) - \sigma(t)| > \eta_{\epsilon,\gamma}) < \dfrac{\gamma}{7I}$

for all $t \in T$ and $N \geq N_{\epsilon,\gamma}$. Moreover $N_{\epsilon,\gamma}$ can be chosen so that the probabilities in (5.15) and (5.16) are also less than $\frac{\gamma}{7 \cdot I \cdot L(\gamma)}$ for $N \geq N_{\epsilon,\gamma}$. Then one gets for the term (5.13) for all

$\epsilon > 0$ and all $0 < \gamma < \frac{\epsilon}{30 \cdot I \cdot K \cdot \max\{|a(t)|; t \in T\}}$ (see inequality (5.14))

$$P\left(\sum_{t \in T} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\psi_{nNt}(\widehat{\beta}_N - \beta, \widehat{\sigma}_N(t) - \sigma(t)) - \psi_{nNt}(0,0) \right.\right.$$
$$\left.\left. - \lambda_{nNt}(\widehat{\beta}_N - \beta, \widehat{\sigma}_N(t) - \sigma(t)) + \lambda_{nNt}(0,0)) \right| > \epsilon \right)$$

$$\leq \sum_{t \in T} P(\sup\{|S_{Nt}(0,0,0) - S_{Nt}(b,0,s)|; |b|_\infty \leq \tfrac{k(\gamma)}{\sqrt{N}}, |s| \leq \eta_{\epsilon,\gamma}\} > \tfrac{\epsilon}{I}) + 2\tfrac{\gamma}{7}$$

$$\leq \sum_{t \in T} \sum_{l=1}^{L(\gamma)} P(\sup\{|S_{Nt}(0,0,0) - S_{Nt}(b,0,s)|; |b - b_N^l| \leq \tfrac{\gamma}{\sqrt{N}}, |s| \leq \eta_{\epsilon,\gamma}\} > \tfrac{\epsilon}{I})$$
$$+ 2\tfrac{\gamma}{7}$$

$$\leq \gamma \qquad \text{for all } N \geq N_{\epsilon,\gamma}.$$

References

[1] BICKEL, P.J.: One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70**, 428-434 (1975).

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

[3] BICKEL, P.J.: Robust regression based on infinitesimal neighbourhoods. *Ann. Statist.* **12**, 1349-1368 (1984).

[4] BILLINGSLEY, P.: *Convergence of Probability Measures.* John Wiley, New York (1968).

[5] HÁJEK, J. and ŠIDÁK, Z.: *Theory of Rank Tests.* Academic Press, New York (1967).

[6] HAMPEL, F.R.: 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 (1978).

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

[8] KRASKER, W.S. and WELSCH, R.E.: Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.* **77**, 595-604 (1982).

[9] KUROTSCHKA, V. and MÜLLER, Ch.: Optimum robust estimation of linear aspects in conditionally contaminated linear models. *To appear in Ann. Statist.* (1992).

[10] MARONNA, R.A. and YOHAI, V.J.: Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Z. Wahrsch. verw. Gebiete* **58**, 7-20 (1981).

[11] MÜLLER, Ch.: Optimale Versuchspläne für robuste Schätzfunktionen in linearen Modellen. *Ph. D. thesis.* Freie Universität Berlin (1987).

[12] MÜLLER, Ch.: Optimal designs for robust estimation in conditionally contaminated linear models. To appear in *J. Statist. Plann. Inference* (1992).

[13] RIEDER, H.: Estimates derived from robust tests. *Ann. Statist.* **8**, 106-115 (1980).

[14] RIEDER, H.: Robust estimation of functionals. *Technical Report.* Universität Bayreuth (1985).

[15] RIEDER, H.: Robust regression estimators and their least favorable contamination curves. *Stat. Decis.* **5**, 307-336 (1987).

[16] RIEDER, H.: *Robust Asymptotic Statistics*, Springer, New York, to appear (1993).

[17] SCHICK, A.: A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference* **16**, 89-105 (1987).

[18] SIMPSON, D.G., RUPPERT, D. and CARROLL, R.J.: On one-step GM estimates and stability of inferences in linear regression. *J. Amer. Statist. Assoc.* **87**, 439-450 (1992).

Christine H. Müller
Freie Universität Berlin
Fachbereich Mathematik und Informatik
1. Mathematisches Institut
Arnimallee 2-6
D-14195 Berlin
Germany