The Annals of Statistics
1992, Vol. 20, No. 1, 331–350

OPTIMUM ROBUST ESTIMATION OF LINEAR ASPECTS IN CONDITIONALLY CONTAMINATED LINEAR MODELS

BY VIKTOR KUROTSCHKA AND CHRISTINE MÜLLER

Freie Universität, Berlin

P. J. Bickel's approach to and results on estimating the parameter vector β of a conditionally contaminated linear regression model by asymptotically linear (AL) estimators $\hat{\beta}^*$ which have minimum trace of the asymptotic covariance matrix among all AL estimators with a given bound b on their asymptotic bias (MT-AL estimators with bias bound b) is here extended to conditionally contaminated general linear models and in particular for estimating arbitrary linear aspects $\varphi(\beta) = C\beta$ of β which are of actual interest in applications. Admitting that β itself is not identifiable in the model (also a practically important situation), a complete characterization of MT-AL estimators with bias bound b including the case where b is smallest possible is presented here, which extends and sharpens H. Rieder's characterization of all AL estimators with minimum asymptotic bias. These characterizations (Theorem 1) represent generalizations (in different directions) of those which define Hampel–Krasker estimators for β in linear regression models and admit (Theorem 2) explicit constructions of MT-AL estimators under generally applicable model assumption. Obviously, even in linear regression models, $\hat{\varphi}^* = C\hat{\beta}^*$ is not an MT-AL estimator for φ if $\hat{\beta}^*$ is one for β (there does not even exist an AL estimator nor an M estimator for β, if β is not identifiable in the model). Examples such as quadratic regression illustrate the not at all obvious relation between $\hat{\beta}^*$ and $\hat{\varphi}^*$, demonstrate the applicability of the general results and show explicitly the influence of the parametrization and the underlying design of the linear model.

1. Introduction. There exist different approaches to robust estimation of the parameter vector $\beta \in \mathbb{R}^p$ of a contaminated linear regression model $Y_n = x_n^T\beta + Z_n$, $x_n \in \mathbb{R}^p$, $n = 1, \dots, N$. A selection of literature is given in the list of references. The aim of this paper is to show how, in particular, the generally applicable and mathematically attractive approach (because it is least restrictive) of Bickel (1981), (1984) and Rieder (1985), (1987) can be modified and extended to successfully attack the problem of estimating an arbitrary linear aspect $\varphi(\beta) = C\beta$ of the parameter vector β of a contaminated general linear model with an arbitrary linear parametrization and an arbitrary underlying design admitting β itself not being identifiable.

One can obviously interpret a linear regression model as an implicit representation of a general linear model by identifying the regressors $x_1^T, \dots, x_N^T$ with rows of a design matrix of a linear model. An attempt to translate by such an identification results on robust estimation of β in a linear regression model

Received August 1988; revised May 1991.

AMS 1980 *subject classifications.* Primary 62F35, 62G05; secondary 62J05.

Key words and phrases. Linear model, conditional contamination, robust estimation, linear aspect, Hampel–Krasker estimator, quadratic regression.

to those for β in a general linear model [as can be found in Hampel, Ronchetti, Rousseeuw and Stahel (1986)] translates also the usual distributional assumptions on $x_1, \dots, x_N$. The distributional assumption on $x_1, \dots, x_N$ defining the linear regression model to be stochastic [see Bickel (1984)] or, equivalently, to have a random carrier [see Maronna and Yohai (1981)] translated to corresponding assumptions on the elements of the design matrix, seems to create discrepancies in interpretations [see Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 308] and restrict application of linear models which are predominantly composed by a given parametrization and an underlying design.

In this paper the random carriers will be identified with the underlying design of the linear model where the notion of a design will be used in its most general sense concerning its application as well as its mathematical form.

To include besides the usual (linear, polynomial, trigonometric, spline) regression experiments also more important generalizations thereof such as experiments simultaneously influenced by qualitative and quantitative factors and to provide a sufficiently adequate basis for design problems intrinsically connected with (robust) estimation in (contaminated) linear models the following more explicit form of a general linear model will be considered:

$$(1.1) \qquad Y_{Nn}(t_n) = a^T(t_n)\beta + Z_{Nn}, \qquad n = 1, \dots, N, \; N \in \mathbb{N},$$

where $t_1, \dots, t_N$ are experimental conditions ranging in a set T, $a\colon T \to \mathbb{R}^p$ is the vector of (known) regression functions, $\beta \in \mathbb{R}^p$ is the vector of (unknown) parameters, Z_{Nn} are the error variables, $Y_{Nn}(t_n)$ are the corresponding observations.

The following relatively simple but practically relevant example of a linear model with two qualitative and K quantitative factors may help as perceptual background for the general statements (and the relevance) of the problem considered in this paper.

The regression functions and the parameter vector of a two-way cross-classified covariance model with linear covariates are defined by

$$a^T(t)\beta = \beta_{i_1}^{(1)} + \beta_{i_2}^{(2)} + x^T\beta^{(3)},$$

$$\text{with } t = (i_1, i_2, x)^T \in T \subset \{1, \dots, I_1\} \times \{1, \dots, I_2\} \times \mathbb{R}^K \subset \mathbb{R}^{K+2},$$

that is,

$$(1.2) \qquad \beta = \left(\beta_1^{(1)}, \dots, \beta_{I_1}^{(1)}, \beta_1^{(2)}, \dots, \beta_{I_2}^{(2)}, \beta_1^{(3)}, \dots, \beta_K^{(3)}\right)^T \in \mathbb{R}^{I_1+I_2+K}$$

and

$$a = \left(1_{\{i_1=1\}}, \dots, 1_{\{i_1=I_1\}}, 1_{\{i_2=1\}}, \dots, 1_{\{i_2=I_2\}}, \pi_3, \dots, \pi_{K+2}\right)^T,$$

where $1_{\{i_k=i\}}$ is the indicator function over the set $\{i_k = i\} \subset T$ and π_j is the jth coordinate mapping (projection on the jth coordinate) of T. The case $I_1 = I_2 = 1$, that is, $\beta_1^{(1)} = \cdots = \beta_{I_2}^{(2)} = \beta^0$ defines the linear regression model and for $\beta^0 = 0$, one obtains $a^T(t)\beta = x^T\beta$, the linear regression through the origin.

In the theory of general linear models $a^T(t)\beta$, $t \in T$, is called the linear parametrization of the model and a specific selection $t_1, \ldots, t_N \in T$, is the allocation of experimental conditions according to the underlying design of the experiment. The mathematical description of a design δ as a probability measure on $(T, \mathscr{T})$, where $\mathscr{T}$ is some suitable σ-algebra on T, serves essentially two different practical purposes, namely describing:

1.1. *Deterministic designs*. The probability measure δ on $(T, \mathscr{T})$ determines by its support $\operatorname{supp}(\delta)$ at which of the experimental conditions the experiment should be (or has been) performed. The weights of δ determine the relative frequencies of replications of the experimental conditions in the sequence of $n = 1, \ldots, N$ of different runs constituting the experiment of size $N \in \mathbb{N}$. [See literature on approximate designs and their interpretation for cases when $\operatorname{supp}(\delta)$ is too large and, in particular, infinite; see also Kurotschka (1987).]

1.2. *Randomized designs*. The probability measure δ on $(T, \mathscr{T})$ determines randomly through a sequence $T_1, \ldots, T_N$ of independent random elements each distributed according to δ, the experimental conditions $t_1, \ldots, t_N$ as realizations of $T_1, \ldots, T_N$ at which the experiment should be (or has been) performed (and how frequently) in the sequence $n = 1, \ldots, N$ of different runs constituting the experiment of size $N \in \mathbb{N}$.

The formal theory of robust estimation of an arbitrary identifiable linear aspect $\varphi(\beta) = C\beta$, where $C \times \mathbb{R}^{s \times p}$, as presented here, is independent of either of the above interpretations of δ (because of its asymptotic character) and independent of the fact whether the design is subject to choice or given by the particular experimental setup. The allocations $d_N := (t_1, \ldots, t_N)^T$, $N \in \mathbb{N}$, of the experimental conditions may be viewed (completely in a formal way) as possible values of the (measurable) vector $D_N = (T_1, \ldots, T_N)^T := (\pi_1, \ldots, \pi_N)^T$ of projections π_n on the (natural) product space $(T^\infty, \mathscr{T}^\infty, \delta^\infty)$ and a specific observation $(y_{N1}, \ldots, y_{NN})^T$ as a particular value of $Y_N(D_N) := (Y_{N1}(T_1), \ldots, Y_{NN}(T_N))^T$, $N \in \mathbb{N}$.

The notion of identifiability of $\varphi(\beta) = C\beta$ in the model with an allocation $\{t_1, \ldots, t_N\}$ is classically (in the Gauss–Markov theory) defined by the linear estimableness of $\varphi(\beta)$. But there it means actually (and is mathematically equivalent to the fact) that the knowledge of $a^T(t)\beta$ for all $t \in \{t_1, \ldots, t_N\}$ implies the knowledge of $\varphi(\beta) = C\beta$. This notion may be extended within the more general framework considered here (including obvious interpretations in general asymptotic approaches) to:

Definition 1. A linear aspect $\varphi(\beta) = C\beta$ of β will be called *identifiable in the linear model*, defined by a linear parametrization $a^T(t)\beta$, $t \in T$, and an underlying design δ, if and only if $a^T(t)\beta = 0$ for all $t \in \mathrm{supp}(\delta)$ implies $\varphi(\beta) = 0$.

From this definition, one sees in particular that the desirable economic design δ [namely, those with small supp(δ) saving the number of changes of experimental conditions but providing efficiency of statistical procedures by replications) will in general leave β unidentified even if β itself is an *identifiable parametrization* (see Example 3.3 in Section 3), that is, if $a^T(t)\beta = 0$ for all $t \in T$ implies $\beta = 0$ [which in practical examples is not always the case, see (1.2)].

In the classical approach to estimation problems in general linear models, the deviations $Z_{Nn}(t_n) := Y_{Nn}(t_n) - a^T(t_n)\beta$ of the observations from the true response of the experiment to the experimental conditions described by $t_n \in T$ are modeled as stochastically independent, normal and, in particular, symmetrical around zero and independent of $t_n \in \mathrm{supp}(\delta)$ distributed random variables, that is, $P^{Z_{Nn}(t_n)} = n_{(0,\sigma^2)}$, where we set $\sigma^2 = 1$ for simplicity.

But it obviously appears more realistic (not only to include the possibility of appearance of gross errors, but also some deviation due to the imperfect description of the response of the experiment to some values $t \in T$) to admit contaminations of this classical assumption and to assume

$$P^{Z_{Nn}(t_n)}(dz) = (1 - \varepsilon_N(t_n))n_{(0,1)}(dz) + \varepsilon_N(t_n)Q(dz, t_n),$$

where the distribution $Q(dz, t_n)$ may depend on a particular value $t_n \in \mathrm{supp}(\delta)$ and is not necessarily symmetric around zero.

Reasonable (i.e., technically tractable and for applications somehow sufficiently realistic) assumptions on such conditional contaminations are the following [see Bickel's ac0 stochastic linear regression models in (1984) and Rieder's (1985, 1987) (c, 1) models and the origin of them in Huber (1983)]:

1. The amount of conditional contamination given $t \in \mathrm{supp}(\delta)$ described by $\varepsilon_N(t) \geq 0$ should decrease with the sample size N so that $\varepsilon_N(t)$ is bounded of order $N^{-1/2}$ for all $t \in \mathrm{supp}(\delta)$ or at least if supp(δ) is infinite in the mean over all $t \in \mathrm{supp}(\delta)$ with respect to δ, that is,

$$\int \varepsilon_N(t)\delta(dt) \leq O(N^{-1/2})$$

(see the Markov inequality for possible practical interpretation).

2. The form of conditional contamination given $t \in \mathrm{supp}(\delta)$ described by some (not necessarily symmetric around zero) conditional distribution $Q(dz, t)$ given $t \in \mathrm{supp}(\delta)$ should have a (Lebesgue) density and therefore a conditional density $f(z, t)$ given $t \in \mathrm{supp}(\delta)$ with respect to the standard normal measure $n_{(0,1)}$ such that the sequence of the joint distributions of $(Z_{N1}(T_1), T_1), \ldots, (Z_{NN}(T_N), T_N)$, $N \geq 1$, is contiguous to $(P^N)_{N \geq 1}$, where $P := (n_{(0,1)} \otimes \delta)$.

Setting $\varepsilon_N(t) = N^{-1/2}R\varepsilon(t)$, the formal description of these assumptions may be given by:

$(Z_{N1}(T_1), T_1), \dots, (Z_{NN}(T_N), T_N)$ are independent and identically distributed according to $P_N = Q_N \otimes \delta$, $N \geq 1$, where

$$\begin{aligned} Q_N(dz, t) &:= P^{Z_{Nn}(T_n)|T_n = t} = P^{Z_{Nn}(t)} \\ &= (1 - N^{-1/2} R\varepsilon(t)) n_{(0,1)}(dz) + N^{-1/2} R\varepsilon(t) Q(dz, t) \\ &= (1 + N^{-1/2} R\varepsilon(t)(f(z,t) - 1)) n_{(0,1)}(dz) \end{aligned} \tag{1.3}$$

with $\varepsilon\colon T \to \mathbb{R}^{(+)}$ such that $\int \varepsilon \, d\delta \leq 1$ and $\|\varepsilon(f-1)\|_\infty < \infty$ ($\|q\|_\infty$ denoting the $n_{(0,1)} \otimes \delta$-ess sup of $|q|$).

Of course, these assumptions define a whole class $\mathscr{P}_R$ of sequences $(P_N^N)_{N \geq 1}$ of distributions parametrized by ε and f which can be viewed as a contamination neighbourhood of the central normal model $(P^N)_{N \geq 1}$ with radius R motivating the name *conditionally contaminated linear models*.

Bickel's (1981, 1984) and Rieder's (1985, 1987) essential generalization (motivated by Le Cam's general approach to asymptotics) of Huber's M-estimators to asymptotically linear estimators with some influence function for estimating the parameter vector β of a contaminated linear regression model extends here for the considered general problem obviously as follows.

DEFINITION 2. Let $\varphi(\beta) = C\beta$ be an identifiable linear aspect of the parameter β of a conditionally contaminated linear model (1.1) and (1.3). Then $\hat{\varphi} = (\hat{\varphi}_N)_{N \geq 1}$ will be called an *asymptotically linear* or briefly an *AL estimator* for φ with *influence function* ψ if and only if there exists a function

$$\psi \in \Psi(C) := \left\{ \psi\colon \mathbb{R} \times T \to \mathbb{R}^s,\ \int |\psi|^2 \, dP < \infty,\ \int \psi \, dn_{(0,1)} = 0,\ \int \psi a^T \zeta \, dP = C \right\}$$

such that

$$N^{1/2} \left[\hat{\varphi}_N(Y_N(D_N), D_N) - \varphi(\beta) - N^{-1} \sum_{n=1}^{N} \psi(Y_{Nn}(T_n) - a^T(T_n)\beta, T_n) \right] \to 0$$

in probability $(P^N)_{N \geq 1}$. Here ζ is defined by $\zeta(z,t) = z$ to avoid variables (z,t) of ψ.

Under assumptions (1.3) on $\mathscr{P}_R$, no modification of Bickel's [(1981), page 18, (1984), page 1351–1355] or Rieder's [(1987), page 317–321] arguments are necessary to establish the asymptotic normality of AL estimators $\hat{\varphi}$ for φ, that is, the validity of

$$\mathscr{L}\left(N^{1/2} \left\{ \hat{\varphi}_N - \varphi - \int \psi \, dP_N \right\} \middle| P_N^N \right) \to N\left(0, \int \psi \psi^T \, dP\right) \quad \text{for all } (P_N^N)_{N \geq 1} \in \mathscr{P}_R$$

as can be seen, for instance, from the results of Behnen and Neuhaus [(1975), page 1350]. Therefore, one obtains immediately the two relevant components of the asymptotic mean square error of an AL estimator $\hat{\varphi}$ for φ in terms of its influence function ψ, namely the trace of its asymptotic covariance matrix

$V(\hat{\varphi}) = \int \psi\psi^T \, dP$, that is, $\operatorname{tr} V(\hat{\varphi}) = \int \psi^T\psi \, dP$, and its asymptotic bias $B_R(\hat{\varphi}) = \sup\{\limsup_{N\to\infty} |N^{1/2}\int \psi \, dP_N|;\ (P_N^N)_{N\geq 1} \in \mathscr{P}_R\} = R\|\psi\|_\infty$, where $\|\psi\|_\infty$ stands for the P-ess sup of $|\psi|$.

Hampel's (1978) and Krasker's (1980) natural multivariate extension of Hampel's (1968) widely discussed and appreciated approach to optimum robust estimation, namely, bounding the asymptotic bias (or the infinitesimal sensitivity) of the estimators (robustness) and minimizing the trace of their asymptotic covariance matrix (optimality), motivates the following definition and the resulting estimation problems.

DEFINITION 3. An AL estimator $\hat{\varphi}^*$ for $\varphi(\beta) = C\beta$ with influence curve $\psi^* \in \Psi(C)$ will be called a *minimum trace-AL estimator* (*MT-AL estimator*) for φ with bias bound b if and only if $\hat{\varphi}^* \in \arg\min\{\operatorname{tr} V(\hat{\varphi});\ \hat{\varphi}$ AL estimator for φ with $B_R(\hat{\varphi}) \leq b\}$ or in terms of influence functions if and only if $\psi^* \in \arg\min\{\operatorname{tr}\int \psi\psi^T \, dP;\ \psi \in \Psi(C)$ with $\|\psi\|_\infty \leq b/R\}$.

Here $\arg\min\{G(x);\ x \in X\}$ stands for the set of all arguments $x^* \in X$ of G which minimize G on X.

With these definitions, the problem considered in this paper is to characterize (as explicitly as possible) the MT-AL estimators with given bias bound for arbitrary (identifiable) linear aspects $\varphi(\beta) = C\beta$ in conditionally contaminated linear models with any given linear parametrization $a^T(t)\beta$, $t \in T$, and for an arbitrary underlying design δ and in particular for those having finite support.

Theorem 1 gives the general solution of this problem by establishing the existence and uniqueness, and by describing the general form of the influence functions of MT-AL estimators with arbitrary bias bound b including the boundary case where b is minimum possible (MT-AL estimators with minimum infinitesimal sensitivity). Under some assumptions on the parametrization and the underlying design of the linear model which do not seriously restrict the relevant models for applications (see Remark 4), Theorem 2 provides a more explicit and constructive characterization of the MT-AL estimators. The examples on quadratic regression in Section 3 demonstrate the applicability of Theorem 2 and serve (besides being of interest in their own right) to illustrate the particular features of the general results presented in this paper and in particular their practical relevance. Among others the examples show the not-at-all obvious relation between MT-AL estimators for β and those for $\varphi = C\beta$. But they also exhibit explicitly the (different) influence of the (different) underlying designs on the (different) estimation problems in a conditionally contaminated linear model and indicate the importance of design problems.

The general setup and the general results of this paper therefore serve also as a basis for investigating optimal designs of conditionally contaminated general linear models. First results on these design problems have been obtained by Müller (1987) and further work including constructions of different classes of examples is in progress.

2. The main results. Lemma 1 clarifies the identifiability of a linear aspect $\varphi(\beta) = C\beta$ in terms of its estimability by AL estimators.

LEMMA 1. *$\varphi(\beta) = C\beta$ is identifiable in the model if and only if $\Psi(C) \neq \emptyset$.*

REMARK 1. Lemma 1 shows among other things, that for linear models in which β itself is not identifiable, an AL estimator of an identifiable aspect $\varphi(\beta) = C\beta$ cannot be defined by $\hat{\varphi} = C\hat{\beta}$ with $\hat{\beta}$ being an AL (or M) estimator, because the set of influence functions for β is empty.

The generalization of Rieder's [(1985), Theorem 3.7(a)] characterization of all AL estimators for the parametervector β of a linear regression model with minimum asymptotic bias is stated here as a lemma (Lemma 2) because it appears here as an auxiliary result. It reduces the problem of characterizing MT-AL estimators (or other optimal robust estimators) with bias bound equal to the minimum possible by characterizing the class of influence functions of all AL estimators with minimum asymptotic bias.

For the sake of brevity and without loss of generality, we omit the constant R (i.e., $R := 1$) and call $\|\psi\|_\infty$ the asymptotic bias and $b_0(C) := \min\{\|\psi\|_\infty; \psi \in \Psi(C)\}$ the minimum asymptotic bias.

LEMMA 2. *Let $b_0(C)$ be the minimum asymptotic bias of AL estimators for a linear identifiable $\varphi(\beta) = C\beta$, then:*

(i) *There exists a matrix $Q_1 \in \mathbb{R}^{s\times p}$, so that*

$$
\begin{aligned}
b_0(C) &= \operatorname{tr} Q_1 C^T\left(\int |Q_1 a|\, d\delta\right)^{-1}(\pi/2)^{1/2} \\
&= \max\left\{\operatorname{tr} QC^T\left(\int |Qa|\, d\delta\right)^{-1}(\pi/2)^{1/2};\ Q \in \mathbb{R}^{s\times p} \text{ with } QC^T \neq 0\right\}.
\end{aligned}
\tag{2.1}
$$

(ii) *If Q_1 satisfies* (i), *then the influence function of every AL estimator for $\varphi(\beta) = C\beta$ with asymptotic bias equal to $b_0(C)$ coincides with*

$$
\psi_1 = Q_1 a |Q_1 a|^{-1} b_0(C) \operatorname{sgn}(\zeta) 1_{T_1}, \qquad P\text{-a.e. on } \mathbb{R} \times T_1, \tag{2.2}
$$

where $T_1 := \{t \in \operatorname{supp}(\delta);\ Q_1 a(t) \neq 0\}$ and 1_{T_1} is the indicator function over T_1.

REMARK 2. Because in general (even when estimating the parameter vector β itself) the set T_1 defined in Lemma 2 may have probability $\delta(T_1) < 1$, the influence curve ψ^* of an AL estimator with minimum asymptotic bias may not be P-a.e. determined by the function ψ_1 of Lemma 2 [see examples (3.3b), (3.6a) and (3.6b)]. There are particular cases of appropriately chosen designs δ with respect to a given linear parametrization of the model for which ψ^* coincides P-a.e. with ψ_1 [see examples (3.3a), (3.9) and (3.12)]. Another example of such a particular case where ψ_1 and ψ^* coincide P-a.e. is formulated as the assumption of Theorem 2 in Ronchetti and Rousseeuw (1985) for estimating β in a linear regression model through the origin [i.e., $a(t) = a(x_1, \dots, x_p) = (x_1, \dots, x_p)^T = t \in \mathbb{R}^p = T$] assuming that the δ on $(T, \mathscr{T}) = (R^p, \mathscr{B}^p)$ is such that $\int (xx^T/\|x\|)\delta(dx) = kE$, where E is the identity matrix and $k \in \mathbb{R}$. Such particular designs are generally known as isotropic designs with respect to some given parametrization of the linear model.

THEOREM 1. *Let $b_0(C)$ be the minimum asymptotic bias of AL estimators for $\varphi(\beta) = C\beta$, then*:

(i) *For every $b \geq b_0(C)$, an MT-AL estimator $\hat{\varphi}^*$ for $\varphi(\beta) = C\beta$ with bias bound b exists and is P-unique in the following sense*: *If ψ^* and ψ^{**} are influence functions of two MT-AL estimators, then $\psi^* = \psi^{**}$ P-a.e.*

(ii) *$\hat{\varphi}^*$ is an MT-AL estimator for $\varphi(\beta) = C\beta$ with bias bound $b > b_0(C)$ if and only if the influence function ψ^* of $\hat{\varphi}^*$ is of the form*

$$\psi^* = Q^* a \operatorname{sgn}(\zeta) \min\{|\zeta|, b|Q^* a|^{-1}\}, \qquad P\text{-a.e.}, \tag{2.3}$$

where $Q^* \in \mathbb{R}^{s \times p}$ is a solution of $Q^* \int a a^T [2\Phi(b|Q^* a|^{-1}) - 1]\, d\delta = C$.

(iii) *$\hat{\varphi}^*$ is an MT-AL estimator for $\varphi(\beta) = C\beta$ with bias bound $b = b_0(C)$ if and only if the influence function ψ^* of $\hat{\varphi}^*$ is P-a.e. of the form*

$$\begin{aligned} \psi^* = & \sum_{m=1}^{M-1} Q_m a |Q_m a|^{-1} b \operatorname{sgn}(\zeta) 1_{\mathbb{R} \times T_m} \\ & + Q_M a \operatorname{sgn}(\zeta) \min\{|\zeta|, b|Q_M a|^{-1}\} 1_{\mathbb{R} \times T_M}, \end{aligned} \tag{2.4}$$

where for $m = 1, \dots, M-1$, $Q_m \in \mathbb{R}^{s \times p}$ are solutions of

$$Q_m (2/\pi)^{1/2} b \int_{T_m} a a^T |Q_m a|^{-1}\, d\delta = C_m \neq 0$$

and $Q_M \in \mathbb{R}^{s \times p}$ is a solution of $Q_M \int_{T_M} a a^T [2\Phi(b|Q_M a|^{-1}) - 1]\, d\delta = C_M$ with $T_m := \{t \in \operatorname{supp}(\delta) \setminus \bigcup_{k=1}^{m-1} T_k;\ Q_m a(t) \neq 0\}$, $m = 1, \dots, M \geq 2$, and $C = \sum_{m=1}^{M} C_m$.

REMARK 3. As a referee pointed out to us, the if part concerning the optimality of an influence function of form (2.3) of Theorem 1(ii) for linear regression models can be deduced from Samarov's (1985) results (by setting his quantity M equal to $C^T C$). Also the estimation of a part β_1 of the parameter vector $\beta = (\beta_1^T, \beta_2^T)^T$ given in Hampel, Ronchetti, Rousseeuw and Stahel [(1986), Section 4.4] can be regarded as a particular case of the if direction of Theorem 1(ii).

THEOREM 2. *If the design δ has finite support $\{t_1, \dots, t_r\}$ and the design matrix $A_\delta := ((a_j(t_i))_{i=1,\dots,r}^{j=1,\dots,p})$ has rank r such that $\varphi(\beta) = C\beta$ is identifiable, then*:

(i) *The minimum asymptotic bias $b_0(C)$ of an AL estimator is equal to $b_0(C) = \max\{(\pi/2)^{1/2} |C I^-(\delta) a(t_i)|;\ i = 1, \dots, r\}$, where $I^-(\delta)$ is any general-*

ized inverse of the information matrix $I(\delta) := \int a(t)a^T(t)\,\delta(dt)$ *of* δ.

(ii) *The influence function of a MT-AL estimator with bias bound* $b \geq b_0(C)$ *is P-a.e. equal to* ψ^* *defined pointwise by*

$$\psi^*(z,t) = \begin{cases} CI^-(\delta)a(t)\operatorname{sgn}(z)\left(\dfrac{\pi}{2}\right)^{1/2}, \\ \qquad \text{for all } t \in \{t_1,\dots,t_r\} \text{ with } \ |CI^-(\delta)a(t)|\left(\dfrac{\pi}{2}\right)^{1/2} = b, \\ CI^-(\delta)a(t)\dfrac{\operatorname{sgn}(z)\min\{|z|, by(t)\}}{|CI^-(\delta)a(t)|y(t)}, \\ \qquad \text{for all } t \in \{t_1,\dots,t_r\} \text{ with } \ 0 < |CI^-(\delta)a(t)|\left(\dfrac{\pi}{2}\right)^{1/2} < b, \\ 0, \quad \text{for all other } t \in T, \end{cases}$$

where $y(t)$ *is a positive fixed point of* $f_t(y) := (2\Phi(by) - 1)|CI^-(\delta)a(t)|^{-1}$, *that is, the largest coordinate* $y = y(t)$ *at which the straight line*

$$g_t(y) = \frac{|CI^-(\delta)a(t)|}{2}y + \frac{1}{2}$$

intersects the (scaled) standard normal distribution function $h(y) = \Phi(by)$.

REMARK 4. Theorem 2 shows that for designs δ with minimum support, that is, designs with support points $t_1,\dots,t_r$ which in particular create linear independent vectors $a(t_1),\dots,a(t_r)$ of regression coefficients, the influence functions of MT-AL estimators for every identifiable aspect φ will have at most two different branches. That two different branches may actually occur when estimating with minimum asymptotic bias is demonstrated by examples (3.6a, b) even if estimating the whole parameter vector β in case of its identifiability [see (3.3b)]. Fortunately, designs with minimum support are of particular practical interest in applications (as already pointed out) so that the explicit characterization of the minimum bias and of influence functions given in Theorem 2 is of particular practical relevance.

3. Example: Quadratic regression. To be able to compare estimators for β with those for $\varphi(\beta) = C\beta$, the parameter β has to be identifiable in the model and to illuminate this comparison, C has to be as simple as possible (but not trivial), therefore, consider a conditionally contaminated quadratic regression model $X(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + Z(t)$, $t \in T \subset \mathbb{R}$, that is, $a(t) = (1, t, t^2)^T$ and $\beta = (\beta_0, \beta_1, \beta_2)^T$, and let $\varphi(\beta) = (\beta_1, \beta_2)^T$, that is, $C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. A reasonable classical optimum design on $T = [-1, 1]$ for β is

$$\delta = \tfrac{1}{4}e_{-1} + \tfrac{1}{2}e_0 + \tfrac{1}{4}e_1,$$

where e_t is the Dirac measure (unit point mass) over t.

To realize the influence of a design on the form of the influence function ψ^* of a MT-AL estimator, the following slight change of δ (actually only of its weights) is very informative:

$$\bar{\delta} = \tfrac{2}{7}e_{-1} + \tfrac{3}{7}e_0 + \tfrac{2}{7}e_1.$$

Straightforward application of Theorem 2 by computing $I^-(\delta)a(t)$ solves the following estimation problems.

3.1. *Estimation of* $\beta = (\beta_1, \beta_2, \beta_3)^T$. The minimum asymptotic bias $b_0(E)$ (here E = identity matrix) of an AL-estimator for β is equal to

$$b_0(E) = \begin{cases} 2\pi^{1/2}, & \text{for the design } \delta, \\ \tfrac{7}{3}\pi^{1/2}, & \text{for the design } \bar{\delta}. \end{cases} \tag{3.1}$$

3.1.A. The (P-unique) influence function ψ^* of a MT-AL estimator for β with bias bound $b > b_0(E)$ is equal to

$$\psi^*(z,t) = \frac{1}{2^{-1/2}}\frac{1}{u}\operatorname{sgn}(z)\min\{|z|, bu\}\cdot\begin{cases} (0,-1,1)^T, & \text{for } t = -1, \\ (0,1,1)^T, & \text{for } t = 1, \\ (1,0,-1)^T, & \text{for } t = 0 \end{cases} \tag{3.2a}$$

for the design δ, where u is the positive solution of $2^{1/2}u + 1/2 = \Phi(bu)$, that is, $u \approx 0.349$ for $b = 2b_0(E)$ and

$$\psi^*(z,t) = \begin{cases} 2^{-1/2}u^{-1}\operatorname{sgn}(z)\min\{|z|, bu\}(0,-1,1)^T, & \text{for } t = -1, \\ 2^{-1/2}u^{-1}\operatorname{sgn}(z)\min\{|z|, bu\}(0,1,1)^T, & \text{for } t = 1, \\ 2^{-1/2}v^{-1}\operatorname{sgn}(z)\min\{|z|, bv\}(1,0,-1)^T, & \text{for } t = 0 \end{cases} \tag{3.2b}$$

for the design $\bar{\delta}$, where u and v are the positive solutions of $7(2^{-1/2})u/4 + 1/2 = \Phi(bu)$ and $7(2^{-1/2})v/3 + 1/2 = \Phi(bv)$, that is, $u \approx 0.404$ and $v \approx 0.299$ for $b = 2b_0(E)$.

3.1.B. The (P-unique) influence function ψ_0^* of a MT-AL estimator for β with bias bound $b = b_0(E)$, that is, with minimum asymptotic bias, is equal to

$$\psi_0^*(z,t) = (2\pi)^{1/2}\operatorname{sgn}(z)\cdot\begin{cases} (0,-1,1)^T, & \text{for } t = -1, \\ (0,1,1)^T, & \text{for } t = 1, \\ (1,0,-1)^T, & \text{for } t = 0 \end{cases} \tag{3.3a}$$

for the design δ and

$$(3.3b)\quad \psi_0^*(z,t) = \begin{cases} 2^{-1/2}u^{-1}\operatorname{sgn}(z)\min\{|z|, \frac{7}{3}\pi^{1/2}u\}(0,-1,1)^T, & \text{for } t=-1,\\ 2^{-1/2}u^{-1}\operatorname{sgn}(z)\min\{|z|, \frac{7}{3}\pi^{1/2}u\}(0,1,1)^T, & \text{for } t=1,\\ \frac{7}{3}(\pi/2)^{1/2}\operatorname{sgn}(z)(1,0,-1)^T, & \text{for } t=0 \end{cases}$$

for the design $\bar{\delta}$, where u is the positive solution of $7(2^{-1/2})u/4 + 1/2 = \Phi(7\pi^{1/2}u/3)$, that is, $u \approx 0.339$. Note that for δ,

$$Q_1 = \begin{pmatrix} 1 & 0 & -1\\ 0 & 1 & 0\\ -1 & 0 & 2 \end{pmatrix}, \qquad Q_2 = 0,\ C_1 = E,\ C_2 = 0,$$

and for $\bar{\delta}$,

$$Q_1 = \begin{pmatrix} 1 & 0 & -1\\ 0 & 0 & 0\\ -1 & 0 & 1 \end{pmatrix}, \qquad Q_2 = 2^{-1/2}u^{-1}\begin{pmatrix} 0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{pmatrix},$$

$$C_1 = \begin{pmatrix} 1 & 0 & 0\\ 0 & 0 & 0\\ -1 & 0 & 0 \end{pmatrix}, \qquad C_2 = \begin{pmatrix} 0 & 0 & 0\\ 0 & 1 & 0\\ 1 & 0 & 1 \end{pmatrix},$$

fulfill the characterizing properties of Theorem 1(iii).

3.2. *Estimating* $\varphi(\beta) = C\beta = (\beta_1, \beta_2)^T$. The minimum asymptotic bias $b_0(C)$ of an AL-estimator for φ is equal to

$$(3.4)\qquad b_0(C) = \begin{cases} 2\pi^{1/2}, & \text{for the design } \delta,\\ \frac{7}{4}\pi^{1/2}, & \text{for the design } \bar{\delta}. \end{cases}$$

3.2.A. The (P-unique) influence function ψ_C^* of a MT-AL estimator for φ with bias bound $b > b_0(C)$ is equal to

$$(3.5a)\quad \psi_C^*(z,t) = \begin{cases} 2^{-1/2}u^{-1}\operatorname{sgn}(z)\min\{|z|, bu\}(-1,1)^T, & \text{for } t=-1,\\ 2^{-1/2}u^{-1}\operatorname{sgn}(z)\min\{|z|, bu\}(1,1)^T, & \text{for } t=1,\\ v^{-1}\operatorname{sgn}(z)\min\{|z|, bv\}(0,-1)^T, & \text{for } t=0 \end{cases}$$

for the design δ, where u and v are positive solutions of $2^{1/2}u + 1/2 = \Phi(bu)$ and $v + 1/2 = \Phi(bv)$, that is $u \approx 0.349$ and $v = 0.500$ for $b = 2b_0(C)$ and

$$(3.5b)\quad \psi_C^*(z,t) = \begin{cases} 2^{-1/2}u^{-1}\operatorname{sgn}(z)\min\{|z|, bu\}(-1,1)^T, & \text{for } t=-1,\\ 2^{-1/2}u^{-1}\operatorname{sgn}(z)\min\{|z|, bu\}(1,1)^T, & \text{for } t=1,\\ v^{-1}\operatorname{sgn}(z)\min\{|z|, bv\}(0,-1)^T, & \text{for } t=0, \end{cases}$$

for the design $\bar{\delta}$, where u and v are positive solutions of $7(2^{-1/2})u/4 + 1/2 = \Phi(bu)$ and $7v/6 + 1/2 = \Phi(bv)$, that is, $u \approx 0.399$ and $v \approx 0.425$ for $b = 2b_0(C)$.

3.2.B. The (P-unique) influence function $\psi^*_{C,0}$ of a MT-AL estimator for φ with bias bound $b = b_0(C)$, that is, with minimum asymptotic bias, is equal to

$$(3.6a)\quad \psi^*_{C,0}(z,t) = \begin{cases} (2\pi)^{1/2}\operatorname{sgn}(z)(-1,1)^T, & \text{for } t = -1, \\ (2\pi)^{1/2}\operatorname{sgn}(z)(1,1)^T, & \text{for } t = 1, \\ v^{-1}\operatorname{sgn}(z)\min\{|z|, 2\pi^{1/2}v\}(0,-1)^T, & \text{for } t = 0, \end{cases}$$

for the design δ, where v is the positive solution of $v + 1/2 = \Phi(2\pi^{1/2}v)$, that is, $v \approx 0.441$ and

$$(3.6b)\quad \psi^*_{C,0}(z,t) = \begin{cases} \frac{7}{4}(\pi/2)^{1/2}\operatorname{sgn}(z)(-1,1)^T, & \text{for } t = -1, \\ \frac{7}{4}(\pi/2)^{1/2}\operatorname{sgn}(z)(1,1)^T, & \text{for } t = 1, \\ v^{-1}\operatorname{sgn}(z)\min\{|z|, \frac{7}{4}\pi^{1/2}v\}(0,-1)^T, & \text{for } t = 0, \end{cases}$$

for the design $\bar{\delta}$, where v is the positive solution of $7v/6 + 1/2 = \Phi(7\pi^{1/2}v/4)$, that is, $v \approx 0.194$. Note that for δ and $\bar{\delta}$,

$$Q_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad Q_2 = v^{-1}\begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix},$$
$$C_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \qquad C_2 = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

fulfill the characterizing properties of Theorem 1(iii).

3.3. *Estimating* $\varphi(\beta) = C_1\beta = \beta_1$, *that is*, $C_1 = (0,1,0)$. Under the design $\delta = e_{-1}/4 + e_0/2 + e_1/4$, the minimum asymptotic bias for $\varphi(\beta) = C_1\beta = \beta_1$ is equal to

$$(3.7)\qquad b_0(C_1) = (2\pi)^{1/2}.$$

For this design the (P-unique) influence function of a MT-AL estimator with bias bound b for $\varphi(\beta) = \beta_1$ is equal to

$$(3.8)\quad \psi^*_{C_1}(z,t) = u^{-1}\operatorname{sgn}(z)\min\{|z|, bu\}t1_{\{-1,1\}}(t) \quad \text{for } b > b_0(C_1),$$

where u is the positive solution of $u + 1/2 = \Phi(bu)$, that is, $u = 0.493$ for $b = 2b_0(C_1)$ and is equal to

$$(3.9)\qquad \psi^*_{C_1,0}(z,t) = 2(\pi/2)^{1/2}\operatorname{sgn}(z)t1_{\{-1,1\}}(t) \quad \text{for } b = b_0(C_1).$$

Note that under the classical optimum design $\delta^* = e_{-1}/2 + e_1/2$ for estimating the linear aspect $\varphi(\beta) = \beta_1$, the parameter vector β itself is not identifiable. Under this design the minimum asymptotic bias of an AL estimator for $\varphi(\beta) = \beta_1$ is equal to

$$b_0(C_1) = (\pi/2)^{1/2} \tag{3.10}$$

The (P-unique) influence function of a MT-AL estimator with bias bound b is equal to

$$\psi^*_{C_1}(z,t) = u^{-1}\operatorname{sgn}(z)\min\{|z|, bu\}t1_{[-1,1]}(t) \quad \text{for } b > b_0(C_1), \tag{3.11}$$

where u is the positive solution of $u/2 + 1/2 = \Phi(bu)$, that is, $u \approx 0.987$ for $b = 2b_0(C_1)$, and is equal to

$$\psi^*_{C_1,0}(z,t) = (\pi/2)^{1/2}\operatorname{sgn}(z)t1_{[-1,1]}(t) \quad \text{for } b = b_0(C_1). \tag{3.12}$$

A comparison of these worked out examples, which have their own interest in applications, may help one to appreciate the general set up and the general results presented in this paper, in particular, with regard to design problems in conditionally contaminated linear models. The examples show in particular how different designs of the linear model may influence the size of the asymptotic bias as well as the shape of the influence functions of MT-AL estimators (and therefore of M-estimators if these exist) and how differently if different linear aspects φ of β are of actual interest. Note in particular that there is neither an evident relation between ψ^* and ψ^*_C nor between ψ^*_0 and $\psi^*_{C,0}$ reflecting the simple relation between β and $\varphi = C\beta$.

4. Proofs.

PROOF OF LEMMA 1. If $\varphi(\beta) = C\beta$ is identifiable, then the Gauss–Markov estimator for $\varphi(\beta)$ exists and is an AL estimator for $\varphi(\beta)$ with influence function $\psi_\infty = C(\int aa^T\,d\delta)^- a\zeta \in \Psi(C)$, that is, $\Psi(C) \neq \varnothing$, (here A^- denotes an arbitrary g-inverse of A, that is, a matrix A^- with $AA^-A = A$).

Assume $\psi(C) \neq \varnothing$, that is, there exists a $\psi_0 \in \Psi(C)$. Then $a^T(t)\beta = 0$ for δ-a.e. $t \in T$ implies $C\beta = \int \psi_0 a^T\zeta\,dP\,\beta = \int \psi_0 a^T\beta\zeta\,dP = 0$ which shows identifiability of $\varphi(\beta) = C\beta$. □

PROOF OF LEMMA 2. The proof can be based on the same standard arguments as used by Rieder (1985). But here they need some modification [because the matrix C defining $\varphi(\beta) = C(\beta)$ is not equal to the identity matrix and is not even square] and some extension (because it is not assumed that β is identifiable under δ nor that β is at all an identifiable parametrization). Partitioning $C = (c_1|\cdots|c_p)$ and $Q = (q_1|\cdots|q_p) \in \mathbb{R}^{s\times p}$ and considering $g_i(\psi) := c_i - \int \psi a_i\zeta\,dP$ as functions on $\Psi := \{\psi\colon \mathbb{R}\times T \to \mathbb{R}^s;\ \int|\psi|^2\,dP < \infty,\ \int \psi\,dn_{(0,1)} = 0\}$, one obtains $\operatorname{tr} QC^T - \int \psi^T Qa\zeta\,dP = \sum_{i=1}^p q_i^T g_i(\psi)$. This implies for all $\psi \in \Psi(C)$,

$$\operatorname{tr} QC^T = \int \psi^T Qa\zeta\,dP \le \|\psi\|_\infty \int |Qa|\,d\delta(\pi/2)^{-1/2}, \tag{4.1}$$

that is, $b_0(C) \geq \operatorname{tr} QC^T(\int |Qa|\, d\delta)^{-1}(\pi/2)^{1/2}$ for all Q with $QC^T \neq 0$ (note that $QC^T \neq 0$ implies $\int |Qa|\, d\delta > 0$).

The existence of Q_1 satisfying (i) can be proved by first showing the existence of $\tilde{\psi} \in \Psi(C)$ such that $\|\tilde{\psi}\|_\infty = \inf\{\|\psi\|_\infty;\ \psi \in \Psi(C)\}$ and then deducing the existence of Q_1 via the Lagrange principle which provides the existence of Lagrange multipliers $q_{11}, \dots, q_{1p} \in \mathbb{R}^s$ satisfying $\|\psi\|_\infty + \sum_{i=1}^p q_{1i}^T g_i(\psi) \geq \|\tilde{\psi}\|_\infty + \sum_{i=1}^p q_{1i}^T g_i(\tilde{\psi}) = \|\tilde{\psi}\|_\infty$ for all $\psi \in \Psi$.

Whereas the proof of the first part is standard (weak compactness arguments, as in the proof of Theorem 1), the application of the Lagrange principle requires the linear independence of $g_1, \dots, g_p$ on Ψ, that is, in the here considered case, the linear independence of the regression functions $a_1, \dots, a_p$ on $\operatorname{supp}(\delta)$. But this excludes the practically more realistic models and problems. To overcome this difficulty, the following transformation argument appears helpful:

Let $\bar{a} = (\bar{a}_1, \dots, \bar{a}_r)$ be a basis of $\operatorname{lin}\{a_1, \dots, a_p\}$ on $\operatorname{supp}(\delta)$, then there exist $A \in \mathbb{R}^{p \times r}$ and a generalized inverse A^- of A such that $a = A\bar{a}$, $\bar{a} = A^- a$ on $\operatorname{supp}(\delta)$ and $C(A^-)^T A^T = C$.

Because for all $\psi \in \Psi(C)$, one has $\int \psi \bar{a}^T \zeta\, dP = \int \psi a^T \zeta\, dP (A^-)^T = C(A^-)^T =: \bar{C}$ and for all

$$\bar{\psi} \in \bar{\Psi}(\bar{C}) := \left\{\psi: \mathbb{R} \times T \to \mathbb{R}^s;\ \int |\psi|^2\, dP < \infty,\ \int \psi\, dn_{(0,1)} = 0,\ \int \psi \bar{a}^T \zeta\, dP = \bar{C}\right\},$$

one has $\int \bar{\psi} a^T \zeta\, dP = \int \bar{\psi} \bar{a}^T \zeta\, dP A^T = \bar{C} A^T = C(A^-)^T A^T = C$, it follows that $\bar{\Psi}(\bar{C}) = \Psi(C)$ and, in particular, $b_0(C) = \min\{\|\bar{\psi}\|;\ \bar{\psi} \in \bar{\Psi}(\bar{C})\}$.

This shows that the minimum bias of AL-estimators in the model with arbitrary regression functions a coincides with that in the transformed model with linear independent regression functions $\bar{a}$ on $\operatorname{supp}(\delta)$.

Therefore, the Lagrange principle applies to the transformed model: There exists $\bar{Q}_1 \in \mathbb{R}^{s \times r}$ with $\bar{Q}_1 \bar{C}^T \neq 0$ which satisfies (i) with a replaced by $\bar{a}$. But because $\bar{C} = C(A^-)^T$ and $\bar{a} = A^- a$ on $\operatorname{supp}(\delta)$, there exists $Q_1 := \bar{Q}_1 A^- \in \mathbb{R}^{s \times p}$ with $Q_1 C^T = \bar{Q}_1 A^- C^T = \bar{Q}_1 \bar{C}_1^T \neq 0$ satisfying (i) with the originally given regression functions a (note that $Q_1 a = \bar{Q}_1 \bar{a}$ by construction).

The statement (ii) follows from the Cauchy–Schwarz and Hölder inequalities: Equality in (4.1) holds if and only if ψ coincides with ψ_1 for all $t \in \operatorname{supp}(\delta)$ for which $Q_1 a(t) \neq 0$. □

Proof of Theorem 1. (i) The weak compactness of $\Psi_b(C) := \{\psi \in \psi(C);\ \|\psi\|_\infty \leq b\}$ [see also Bickel (1984), page 1355, Rieder (1985), page 35] implies for every sequence $(\psi_n \in \Psi_b(C))_{n \geq 1}$ such that

$$\lim_{n \to \infty} \operatorname{tr} \int \psi_n \psi_n^T\, dP = i := \inf\left\{\operatorname{tr} \int \psi \psi^T\, dP;\ \psi \in \Psi_b(C)\right\}$$

the existence of a subsequence $(\psi_{n(k)})_{k \geq 1}$ of $(\psi_n)_{n \geq 1}$ and of an influence function $\psi^* \in \Psi_b(C)$ with

$$\begin{aligned}\operatorname{tr}\int \psi^*\psi^{*T}\,dP &= \lim_{k\to\infty}\int \psi_{n(k)}^T\psi^*\,dP \le \lim_{k\to\infty}\int |\psi_{n(k)}|\,|\psi^*|\,dP\\ &\le \lim_{k\to\infty}\left(\int|\psi_{n(k)}|^2\,dP\right)^{1/2}\left(\int|\psi^*|^2\,dP\right)^{1/2}\\ &= i^{1/2}\left(\operatorname{tr}\int\psi^*\psi^{*T}\,dP\right)^{1/2}.\end{aligned}$$

This proves the existence of an influence function ψ^* of an MT-AL estimator for φ with bias bound $b \ge b_0(C)$. The P-uniqueness of ψ^* follows from (ii) and (iii).

An AL estimator $\hat{\varphi}^* = (\hat{\varphi}_N^*)_{N\ge 1}$ with influence function ψ^* can be obtained by extending Rieder's [(1985), page 71] construction of an AL estimator for the parameter vector of a linear regression model to the general problem as follows: For a realization $d_N = (t_1,\dots,t_N)^T$ of the design δ and the realization $y_N := (y_{N1},\dots,y_{NN})^T$ of the corresponding observation $Y_N(d_N) = (Y_{N1}(t_1),\dots,Y_{NN}(t_N))^T$, define $\hat{\varphi}_N^*(y_N, d_N) := C(A^-)^T\beta_N^*(y_N, d_N) + N^{-1}\sum_{n=1}^N \psi^*(y_{Nn} - a(t_n)^T(A^-)^T\beta_N^*(y_N, d_N), t_n)$, where $\beta_N^*(y_N, d_N)$ is a discretized version in the sense of Rieder [(1985), page 71] of $(\overline{A}(d_N)^T\overline{A}(d_N))^-\overline{A}(d_N)^Ty_N$ with $\overline{A}(d_N) := (a(t_1)|\cdots|a(t_N))^T(A^-)^T$ and A as in Lemma 2. The asymptotic linearity of $(\hat{\varphi}_N^*)_{N\ge1}$ can then be proved by introducing the following transformation:

For $i = 1,\dots,s$, let $C^i := e_i^TC(A^-)^T \in \mathbb{R}^{1\times r}$, $\psi^i := e_i^T\psi^*$, $\hat{\varphi}_N^i := e_i^T\hat{\varphi}_N^*$, where e_i is the ith unit vector of $\mathbb{R}^s$.

Then there exist $B \in \mathbb{R}^{r-1\times r}$, $\psi_N^B\colon \mathbb{R}\times T \to \mathbb{R}^{r-1}$, $\psi^B := \mathbb{R}\times T\to\mathbb{R}^{r-1}$ and $\hat{\varphi}_N^B\colon \mathbb{R}^N\times T^N \to \mathbb{R}^{r-1}$ so that

$$\overline{C} := \begin{pmatrix} C^i\\ B\end{pmatrix} \in \mathbb{R}^{r\times r}\ \text{ is regular,}\qquad \overline{\psi} := \overline{C}^{-1}\begin{pmatrix}\psi^i\\ \psi^B\end{pmatrix},\qquad \overline{\psi}_N := \overline{C}^{-1}\begin{pmatrix}\psi^i\\ \psi_N^B\end{pmatrix}$$

fulfill conditions (4.23), (4.24) and (4.25) in Rieder [(1985), page 65] and

$$\begin{aligned}\overline{\varphi}_N(y_N, d_N) &:= \overline{C}^{-1}\begin{pmatrix}\hat{\varphi}_N^i(y_N, d_N)\\ \hat{\varphi}_N^B(y_N, d_N)\end{pmatrix}\\ &= \beta_N^*(y_N, d_N)\\ &\quad + N^{-1}\sum_{n=1}^N \overline{\psi}_N\big(y_{Nn} - a(t_n)^T(A^-)^T\beta_N^*(y_N, d_N), t_n\big).\end{aligned}$$

Therefore, Theorem 4.10 in Rieder (1985) provides

$$N^{1/2}\left[\overline{\varphi}_N(Y_N(D_N), D_N) - \overline{\beta} - N^{-1}\sum_{n=1}^N \overline{\psi}\big(Y_{Nn}(T_n) - \overline{a}(T_n)^T\overline{\beta}, T_n\big)\right] \to 0 \quad \text{in probability}$$

with $\overline{\beta} := A^T\beta \in \mathbb{R}^r$ and $\overline{a} := A^-a\colon T\to\mathbb{R}^r$.

Premultiplication with $\bar{e}_1\overline{C}$ where $\bar{e}_i$ is the ith unit vector in $\mathbb{R}^r$ yields the asymptotic linearity of $(\hat{\varphi}_N^i)_{N\geq 1}$ for $e_i^T C\beta$ with influence functions ψ^i and thus completes the proof of (i).

(ii) The problem of characterizing $\psi^* \in \arg\min\{\operatorname{tr}\int\psi\psi^T\,dP;\ \psi \in \Psi(C), \|\psi\|_\infty \leq b\}$ for $b > b_0(C)$ can be solved by modifying and extending arguments of Bickel [(1981), page 35] and Rieder [(1987), page 334], see also the original arguments in Hampel [(1968), pages 51–52] and Krasker [(1980), page 1342]. For a solution by the Lagrange principle, let $g_i\colon \Psi\colon \to \mathbb{R}^s$, $i = 1, \dots, p$, be as in Lemma 2. Then the Lagrange principle provides the existence of Lagrange multipliers $q_1, \dots, q_p \in \mathbb{R}^s$ and $q \in \mathbb{R}^+$ satisfying

$$\begin{aligned}
&\operatorname{tr}\int\psi\psi^T\,dP + 2\sum_{i=1}^{p} q_i^T g_i(\psi) + q(\|\psi\|_\infty - b)\\
&\quad\geq \operatorname{tr}\int\psi^*(\psi^*)^T\,dP + 2\sum_{i=1}^{p} q_i^T g_i(\psi^*) + q(\|\psi^*\|_\infty - b)\\
&\quad= \operatorname{tr}\int\psi^*(\psi^*)^T\,dP \quad \text{for all } \psi \in \Psi.
\end{aligned}$$

In case of linearly dependent $g_1, \dots, g_p$, reparametrize the model as in Lemma 2 by $a \to \bar{a} := A^- a$ and $\overline{C} := C(A^-)^T$. From the above inequality with accordingly redefined functions $\bar{g}_1, \dots, \bar{g}_r$ and the corresponding matrix $\overline{Q} := (\bar{q}_1|\cdots|\bar{q}_r) \in \mathbb{R}^{s\times r}$ of Lagrange multipliers, one obtains for every $\psi \in \Psi$ with $\|\psi\|_\infty \leq b$:

$$\begin{aligned}
&\int\left|\psi - \overline{Q}\bar{a}\zeta\right|^2 dP + 2\operatorname{tr}\overline{QC}^T\\
&\quad\geq \operatorname{tr}\int\psi\psi^T\,dP - 2\int\psi^T\overline{Q}\bar{a}\zeta\,dP + 2\operatorname{tr}\overline{QC}^T\\
&\qquad+ \int\bar{a}^T\overline{Q}^T\overline{Q}\bar{a}\zeta^2\,dP + q(\|\psi\|_\infty - b)\\
&\quad\geq \int\left|\psi^* - \overline{Q}\bar{a}\zeta\right|^2 dP + 2\operatorname{tr}\overline{QC}^T
\end{aligned}$$

because $\sum_{i=1}^r \bar{q}_i\bar{g}_i(\psi) = \operatorname{tr}\overline{QC}^T - \int\psi^T\overline{Q}\bar{a}\zeta\,dP$ by definition of $\bar{g}_1, \dots, \bar{g}_r$. This reduces the original problem to the characterization of

$$\psi^* \in \arg\min\left\{\int\left|\psi - \overline{Q}\bar{a}\zeta\right|^2 dP; \psi \in \Psi, \|\psi\|_\infty \leq b\right\} \quad \text{for } b > b_0(C).$$

Because $\int|\psi - \overline{Q}\bar{a}\zeta|^2\,dP = \int|\psi - Q^* a\zeta|^2\,dP$ for $Q^* := \overline{Q}A^- \in \mathbb{R}^{s\times p}$ and because $\|\psi\|_\infty \leq b$ implies $|\psi| \leq b$ P-a.e., every solution of the reduced minimization problem is obviously P-a.e. equal to

$$\begin{aligned}
\psi^*(z,t) &= Q^* a(t) z 1_{\{|Q^* a(t) z| \leq b\}}(z,t) + b\frac{Q^* a(t) z}{|Q^* a(t) z|} 1_{\{|Q^* a(t) z| > b\}}(z,t)\\
&= Q^* a(t)\operatorname{sgn}(z)\min\left\{|z|, b|Q^* a(t)|^{-1}\right\}.
\end{aligned}$$

The characterization of Q^* in (ii) follows from $\psi^* \in \Psi(C)$ which implies that

$$C = \int \psi^* a^T \zeta \, dP = Q^* \int a a^T |\zeta| \min\{|\zeta|, b|Q^* a|^{-1}\} d(n_{(0,1)} \otimes \delta)$$
$$= Q^* \int a a^T \left[2\Phi\left(b|Q^* a|^{-1}\right) - 1\right] d\delta.$$

[Note that the last equality is a consequence of the following well-known identity: $\int |x| \min\{|x|, k\} n_{(0,1)}(dx) = 2\Phi(k) - 1$.]

The equality of $\arg\min\{\int |\psi - Q^* a \zeta|^2 \, dP;\ \psi \in \Psi(C),\ \|\psi\|_\infty \le b\}$ and $\arg\min\{\operatorname{tr} \int \psi \psi^T \, dP;\ \psi \in \Psi(C),\ \|\psi\|_\infty \le b\}$ completes the proof [note that $b > b_0(C)$ follows from Lemma 2 because ψ^* is not of form (2.2)].

(iii) According to Lemma 2, the influence function of every AL estimator of $\varphi(\beta) = C(\beta)$ with minimum asymptotic bias $b_0(C)$ is P-a.e. equal to $\psi = \psi_1 + \psi_2$ with $\psi_2 := \psi 1_{\operatorname{supp}(\delta) \setminus T_1}$ and ψ_1 and T_1 are as in Lemma 2. For the two trivial cases $T_1 = \operatorname{supp}(\delta)$ and $T_1 \neq \operatorname{supp}(\delta)$ but $\int \psi_1 a^T \zeta \, dP = C$, that is, $\psi_1 \in \Psi(C)$, the P-unique influence function of an MT-AL estimator with bias bound $b_0(C)$ is obviously P-a.e. equal to ψ_1 [i.e., (2.4) holds with $M = 2$, $Q_2 = 0$]. For the second of these two trivial cases, $\operatorname{tr} \int \psi \psi^T \, dP = \int \psi^T \psi \, dP = \int \psi_1^T \psi_1 \, dP + \int \psi_2^T \psi_2 \, dP \ge \int \psi_1^T \psi_1 \, dP$ provides the proof.

The remaining more general case [see examples (3.3b), (3.6a), (3.6b) and Theorem 2 together with its proof] occurs when $T_1 \neq \operatorname{supp}(\delta)$ and $\int \psi_1 a^T \zeta \, dP = C_1 \neq C$, that is, $\psi_1 \notin \Psi(C)$ for T_1 and ψ_1 defined in Lemma 2.

In this case, the original problem to determine $\psi^* = \psi_1 + \psi_2^*$ so that $\psi^* \in \arg\min\{\operatorname{tr} \int \psi \psi^T \, dP;\ \psi \in \Psi(C),\ \|\psi\|_\infty = b_0(C)\}$ is obviously equivalent to the construction of ψ_2^* so that $\psi_2^* \in \arg\min\{\operatorname{tr} \int \psi_2 \psi_2^T \, dP;\ \psi_2 \in \Psi(C_2),\ \psi_2 1_{T_1} = 0,\ \|\psi_2\|_\infty \le b_0(C)\}$, where $\Psi(C_2)$ is as in Definition 2 with C replaced by $C_2 := C - C_1$.

This extremum problem for ψ_2^* can be identified with the problem of constructing the influence function of a MT-AL estimator for $\varphi_2(\beta) := C_2 \beta$ with bias bound $b_0(C)$ under the restriction δ_2 of the design δ_1 on $T_2 := \operatorname{supp}(\delta) \setminus T_1$, defined by $\delta_2(S) := \delta(S \cap T_2)$. The replacement of δ by δ_2 in $\Psi(C_2)$ discards the side condition $\psi_2 1_{T_1} = 0$ in the above minimization problem and redefines $\Psi(C_2)$ to be the set $\Psi_2(C_2) := \{\psi: \mathbb{R} \times T_2 \to \mathbb{R}^s;\ \int |\psi|^2 \, d(n_{(0,1)} \otimes \delta_2) < \infty,\ \int \psi \, dn_{(0,1)} = 0,\ \int \psi a^T \zeta \, d(n_{(0,1)} \otimes \delta_2) = C_2\}$ of all influence functions for estimating φ_2 under the design δ_2. Obviously for their minimum asymptotic bias $b_{20}(C_2)$, one obtains

$$b_{20}(C_2) \le (n_{(0,1)} \otimes \delta_2) - \operatorname{ess\,sup}(|\psi^*|)$$
$$\le (n_{(0,1)} \otimes \delta) - \operatorname{ess\,sup}(|\psi^*|) = b_0(C).$$

If $b_{20}(C_2) < b_0(C)$, then Theorem 1(ii) applied to the problem of estimating φ_2 with the set $\Psi_2(C_2)$ of influence functions establishes the validity of (iii) with $M = 2$.

If $b_{20}(C_2) = b_0(C)$, then the arguments at the beginning of the proof of (iii) apply to the modified problem partitioning T_2 in T_{21} and $T_3 := T_2 \setminus T_{21}$ and characterizing $\psi_2 = \psi_{21} 1_{T_{21}} + \psi_3 1_{T_3}$ on $T_2 := T \setminus T_1 = T \setminus \{t \in \operatorname{supp}(\delta);\ Q_1 a(t) \neq 0\}$ according to Lemma 2.

Note that the characterization of Q_1 in Lemma 2(i) translates to that given in Theorem 1(iii) and analogously to that for Q_2. The resulting two new trivial cases lead to the validity and uniqueness of (2.4) with $M = 3$ but $Q_3 = 0$.

The more general case restarts the chain of arguments which ends after at most $r - 1$ steps, where r is the dimension of the linear hull of $a_1, \dots, a_p$ on $\operatorname{supp}(\delta)$. This follows immediately from the construction of the partition $T_1, T_2 := T_{21}, \dots, T_M$ of T based on Lemma 2 and proves (iii) together with the corresponding uniqueness statement in (i). □

Proof of Theorem 2. The main idea behind Theorem 2 is, that under the condition posed on A_δ, a coordinatewise solution of the matrix equations for Q^* and for $Q_1, \dots, Q_M$ of Theorem 1 can be constructed by (real-valued) fixed points $y(t)$, $t \in \{t_1, \dots, t_r\}$. To spare the reader the technical construction of the solution, the proof is given here by showing that the following definitions fulfill the characterizing properties of Theorem 1.

Setting for the quantities in Theorem 1(iii) with $M = 2$,

$$T_1 := \{t \in \operatorname{supp}(\delta); |CI^-(\delta)a(t)|(\pi/2)^{1/2} = b\},$$

$$T_2 := \{t \in \operatorname{supp}(\delta); |CI^-(\delta)a(t)|(\pi/2)^{1/2} < b\} = \operatorname{supp}(\delta) \setminus T_1,$$

and

$$C_1 := CI^-(\delta)\int_{T_1} a(t)a^T(t)\delta(dt),$$

$$C_2 := CI^-(\delta)\int_{T_2} a(t)a^T(t)\delta(dt) = C - C_1,$$

the matrices

$$Q_1 := C\left(\int_{T_1} a(t)a^T(t)\delta(dt)\right)^-,$$

$$Q_2 := C\left(\int_{S_2} a(t)a^T(t)[2\Phi(by(t)) - 1]\delta(dt)\right)^-$$

$$\text{with } S_2 := T_2 \setminus \{t \in \operatorname{supp}(\delta); CI^-(\delta)a(t) = 0\},$$

will solve the equations given in Theorem 1(iii):

$$C_1 = Q_1\int_{T_1} a(t)a^T(t)|Q_1 a(t)|^{-1} b(2/\pi)^{1/2}\delta(dt),$$

$$C_2 = Q_2\int_{T_2} a(t)a^T(t)\left[2\Phi\left(b|Q_2 a(t)|^{-1}\right) - 1\right]\delta(dt),$$

identifying $b^* := \max\{|CI^-(\delta)a(t)|(\pi/2)^{1/2};\ t \in \operatorname{supp}(\delta)\}$ to be the minimum bias $b_0(C)$.

Note that for all $b > b^*$, the set T_1 as defined above is empty so that Theorem 1(ii) applies.

The following general argument supports the above statement.

The integrals defining $I(\delta)$, Q_1 and Q_2 are of the form

$$\int a(t)a^T(t)k(t)\delta(dt) = A_\delta^T D A_\delta$$

with $D = \operatorname{diag}(d_1, \ldots, d_r)$, where $d_i = k(t_i)\delta\{t_i\}$, $t_i \in \operatorname{supp}(\delta)$, $k(t_i) \in \mathbb{R}$. The assumption that rank $A_\delta = r$ implies the existence of (specific) generalized inverses of A_δ and of $A_\delta^T D A_\delta$ so that $A_\delta A_\delta^- = E$ and $(A_\delta^T D A_\delta)^- = A_\delta^- D^-(A_\delta^-)^T$. The identifiability of φ implies the existence of $K \in \mathbb{R}^{s\times r}$ so that $C = KA_\delta$. Therefore it holds true that $C(A_\delta^T D A_\delta)^- A_\delta^T = KA_\delta A_\delta^- D^-(A_\delta^-)^T A_\delta^T = KD^- = K\operatorname{diag}(d_1^-, \ldots, d_r^-)$, where $d_i^- = d_i^{-1}$ if $d_i \neq 0$ and 0 otherwise. This representation applied to $CI^-(\delta)$, Q_1 and Q_2 shows that $Q_2 a(t) = CI^-(\delta)a(t)(2\Phi(by(t)) - 1)^{-1}$ for $t \in S_2$ and 0 otherwise, that is, $|Q_2 a(t)| = y^{-1}(t)$ by definition of $y(t)$, $t \in S_2$, and $Q_1 a(t) = CI^-(\delta)a(t)$ for $t \in T_1$ and 0 otherwise.

This proves that indeed Q_1 and Q_2 are solutions of the matrix equation of Theorem 1 and ψ^* from Theorem 2 is of the form presented in Theorem 1, namely characterized in (iii) if $b = b^* = b_0(C)$ and in (ii) if $b > b^* = b_0(C)$. □

Acknowledgments. We thank H. Rieder for providing us his latest preprints and his encouraging interest. We also owe thanks to the referees whose questions, comments and recommendations helped to improve the readability of the paper.

REFERENCES

BICKEL, P. J. (1981). Quelques aspects de la statistique robuste. *Ecole d'Eté de Probabilités de St. Flour IX. Lecture Notes in Math.* **876** 1–72. Springer, New York.

BICKEL, P. J. (1984). Robust regression based on infinitesimal neighbourhoods. *Ann. Statist.* **12** 1349–1368.

BEHNEN, K. and NEUHAUS, G. (1975). A central limit theorem under contiguous alternatives. *Ann. Statist.* **3** 1349–1353.

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. dissertation, Univ. California, Berkeley.

HAMPEL, F. R. (1978). Optimally bounding the gross-error-sensitivity and the influence of position in factor space. In *Proceedings of the Statistical Computing Section* 59–64. Amer. Statist. Assoc., Washington, D.C.

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

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

HUBER, P. J. (1983). Minimax aspects of bounded-influence regression (with discussion). *J. Amer. Statist. Assoc.* **78** 66–80.

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

KUROTSCHKA, V. (1987). Characterizations and examples of optimal experiments with qualitative and quantitative factors. *Model-Oriented Data Analysis. Lecture Notes in Econom. and Math. Systems* **297** 53–71. Springer, New York.

MARONNA, R. A. and YOHAI, V. J. (1981). Asymptotic behaviour of general *M*-estimates for regression and scale with random carriers. *Z. Wahrsch. Verw. Gebiete* **58** 7–20.

MÜLLER, C. (1987). Optimale Versuchspläne für robuste Schätzfunktionen in linearen Modellen. Ph.D. dissertation, Free Univ., Berlin.

RIEDER, H. (1985). Robust estimation of functionals. Technical report, Univ. Bayreuth.

RIEDER, H. (1987). Robust regression estimators and their least favorable contamination curves. *Statist. Decisions* **5** 307–336.

RONCHETTI, E. and ROUSSEEUW, P. J. (1985). Change-of-variance sensitivities in regression analysis. *Z. Wahrsch. Verw. Gebiete* **68** 503–519.

SAMAROV, A. M. (1985). Bounded-influence regression via local minimax mean squared error. *J. Amer. Statist. Assoc.* **80** 1032–1040.

MATHEMATISCHES INSTITUT
FREIE UNIVERSITÄT BERLIN
ARNIMALLEE 2-6
1000 BERLIN 33
GERMANY