The Annals of Statistics
1998, Vol. 26, No. 3, 1126–1146

OPTIMUM ROBUST TESTING IN LINEAR MODELS

BY CHRISTINE MÜLLER

Georg-August-University of Göttingen

Robust tests for linear models are derived via Wald-type tests that are based on asymptotically linear estimators. For a robustness criterion, the maximum asymptotic bias of the level of the test for distributions in a shrinking contamination neighborhood is used. By also regarding the asymptotic power of the test, admissible robust tests and most-efficient robust tests are derived. For the greatest efficiency, the determinant of the covariance matrix of the underlying estimator is minimized. Also, most-robust tests are derived. It is shown that at the classical D-optimal designs, the most-robust tests and the most-efficient robust tests have a very simple form. Moreover, the D-optimal designs provide the highest robustness and the highest efficiency under robustness constraints across all designs. So, D-optimal designs are also the optimal designs for robust testing. Two examples are considered for which the most-robust tests and the most-efficient robust tests are given.

1. Introduction. A general linear model

$$Y_N = A(d_N)\beta + Z_N \tag{1.1}$$

is considered, where $Y_N = (Y_{1N}, \dots, Y_{NN})'$ is the vector of real-valued observations, $Z_N = (Z_{1N}, \dots, Z_{NN})'$ the vector of error variables, $\beta \in \mathbb{R}^r$ an unknown parameter vector, $d_N = (t_{1N}, \dots, t_{NN})' \in \mathcal{T}^N$ the vector of known experimental conditions, that is, the design, $A(d_N) = (a(t_{1N}), \dots, a(t_{NN}))' \in \mathbb{R}^{N\times r}$ the design matrix and a: $\mathcal{T} \to \mathbb{R}^r$ the vector of known "regression" functions. Realizations of the N-dimensional random vectors Y_N and Z_N are denoted by $y_N = (y_{1N}, \dots, y_{NN})'$ and $z_N = (z_{1N}, \dots, z_{NN})'$, respectively. For this linear model, a hypothesis of the general form H_0: $L\beta = l$ should be tested against the alternative H_1: $L\beta \neq l$, where $l \in \mathbb{R}^s$ and $L \in \mathbb{R}^{s\times r}$ is of rank s. For these hypotheses, it is known that the classical F-tests are very sensitive to outliers and to other deviations from normality. Therefore other tests, which are robust against these violations of the model assumptions, should be used.

For testing a hypothesis of the form H_0: $\beta = 0$ or H_0: $\beta_1 = 0$, where β_1 is a subvector of β, there exist several attempts to derive tests that are robust against outliers. Many such tests were derived by transferring the robustness criteria for estimators to the tests. See, for example, Hampel, Ronchetti, Rousseeuw and Stahel (1986), Staudte and Sheather (1990), Markatou and Hettmansperger (1990), Akritas (1991), Markatou, Stahel and Ronchetti (1991), Silvapulle (1992a, b), Coakley and Hettmansperger (1992, 1994), Rieder (1994), Markatou and He (1994), Heritier and Ronchetti (1994),

Received October 1995; revised November 1997.

AMS 1991 *subject classifications*. Primary 62F35, 62K05; secondary 62J05, 62J10.

Key words and phrases. Linear model, shrinking contamination, robust tests, asymptotically linear estimators, bias of the level, most robustness, efficiency, D-optimality, optimal design.

Markatou and Manos (1996). Up to now, only Ronchetti (1982) and Rieder (1994), Section 5.4, considered the question of how to derive robust tests to be as efficient as possible. Ronchetti (1982) derived most-efficient robust tests within the class of τ-tests. These τ-tests have a complicated asymptotic distribution, and so only an implicit characterization of the tests is possible. See also Chapter 7 of Hampel, Ronchetti, Rousseeuw and Stahel (1986). Rieder (1994) defined most-efficient robust tests within the Wald-type tests that are based on asymptotically linear estimators of β. However, a representation of the tests that can be used in practice was not given. Moreover, an open problem is how to derive most-robust tests, that is, tests with maximum robustness.

The purpose of this paper is to characterize explicitly most-efficient robust tests and most-robust tests in order that they can easily be used in practice. To do this, we use Wald-type tests that are based on asymptotically linear estimators for $L\beta$. For a robustness concept, we make use of the concept already proposed in Müller (1992a, c), which is similar to that of Rieder (1994). This concept expresses robustness in terms of the asymptotic bias of the level of the test for distributions in a shrinking contamination neighborhood. This robustness concept is closely related to that of Heritier and Ronchetti (1994), which is based on the influence function of the test statistic. However, Heritier and Ronchetti need the strong assumption that the test statistic is Fréchet differentiable. This excludes important tests such as the most-robust tests covered here.

Using either our robustness concept or that of Heritier and Ronchetti to derive most-efficient robust tests and most-robust tests leads to optimization problems for influence functions. In these optimization problems, the efficiency criterion and the robustness criterion are not convex functions of the influence function. This differs from robust estimation, where, for every design, methods of convex analysis can be used to characterize most-efficient robust estimators and most-robust estimators [see Hampel (1978), Krasker (1980), Bickel (1981, 1984), Rieder (1985, 1987, 1994), Kurotschka and Müller (1992)]. Although convex analysis can not be applied to the Wald-type tests considered here, optimal tests can still be derived. To do this, we combine the efficiency of the test with the efficiency of the design.

The paper is organized as follows. In Section 2, the Wald-type tests based on asymptotically linear estimators, called ALE-tests, are defined and their asymptotic behavior in contaminated linear models is investigated. From this asymptotic behavior, the robustness measure and the efficiency measure for ALE-tests are derived in Section 3. In Section 4, the most-robust tests and their corresponding designs are given. In particular, it is shown that the classical D-optimal designs provide the highest robustness. In Section 5, most-efficient tests and designs are derived under robustness constraints. At first, admissible tests are characterized by generalizing a result of Krasker and Welsch (1982). Then most-efficient robust tests are derived by reducing the power (the efficiency) of the test to the determinant of the asymptotic covariance matrix of the underlying estimator. It turns out that for most-efficient robust tests, the D-optimal designs are again optimal. In Section 6, two examples are given that show the applicability of the results for models with qualitative factors and for models with quantitative factors. In the first example, optimal robust tests for testing the equality of the effects in a one-way lay-out model are presented. The second example deals with a testing problem in the quadratic regression model. All proofs are given in Section 7.

2. Definition and asymptotic behavior of ALE-tests. Recall that the test statistic of the classical F-test for testing H_0: $L\beta = l$ in the model (1.1) is

$$\tau_N^{LS}(y_N, d_N) = N\big(\hat{\varphi}_N^{LS}(y_N, d_N) - l\big)' C_N^{LS}(y_N, d_N)^{-1}\big(\hat{\varphi}_N^{LS}(y_N, d_N) - l\big),$$

where $\hat{\varphi}_N^{LS}$ is the Gauss–Markov estimator for $\varphi(\beta) = L\beta$ and $C_N^{LS}(y_N, d_N)$ converges in probability to the asymptotic covariance matrix of the Gauss–Markov estimator [see, e.g., Christensen (1987), page 40 ff.]. This test statistic can be generalized by replacing the Gauss–Markov estimator $\hat{\varphi}_N^{LS}$ by an asymptotically linear estimator. Many well-known estimators such as R-estimators, M-estimators and one-step M-estimators are asymptotically linear. Moreover, they are usually asymptotically linear for contiguous alternatives of H_0: $L\beta = l$, that is, for alternatives of the form $\beta_N = \beta + N^{-1/2}\overline{\beta}$ with $L\beta = l$. To define the asymptotic linearity of an estimator for contiguous alternatives, we assume that the sequence of designs $(d_N)_{N\in\mathbb{N}}$ converges to an asymptotic design measure δ with finite support in the following sense:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e_{t_{nN}}\big(\{t\}\big) = \delta\big(\{t\}\big)$$

for all $t \in \operatorname{supp}(\delta)$, where $\operatorname{supp}(\delta)$ is the support of δ and e_t is the Dirac measure on $t \in \mathscr{T}$. The measure δ is also called a design. Moreover, we assume that the error variables $Z_{1N}, \dots, Z_{NN}$ have unknown variance σ^2 so that $Z_{1N}/\sigma, \dots, Z_{NN}/\sigma$ are independent and identically distributed according to a symmetric distribution $P = P^{Z_{nN}/\sigma}$ with mean 0, variance 1, and bounded Lebesgue density $f(z)$. Here P_β^N denotes the distribution of Y_N.

DEFINITION 2.1. An estimator $\hat{\varphi}_N$ for $\varphi(\beta) = L\beta$ is asymptotically linear for contiguous alternatives of H_0: $L\beta = l$ with influence function ψ if ψ: $\mathbb{R} \times \mathscr{T} \to \mathbb{R}^s$ satisfies the conditions

$$\int \psi(z, t) P(dz) = 0 \quad \text{for all } t \in \mathscr{T}, \tag{2.1}$$

$$\int \psi(z, t) a(t)' z P(dz)\, \delta(dt) = L, \tag{2.2}$$

$$\int |\psi(z, t)|^2 P(dz)\, \delta(dt) < \infty, \tag{2.3}$$

and

$$\lim_{N\to\infty} P^N_{\beta_N}\left(\left\{y_N \in \mathbb{R}^N; \sqrt{N}\left|\hat{\varphi}(y_N, d_N) - \varphi(\beta_N) - \frac{\sigma}{N}\sum_{n=1}^{N}\psi\left(\frac{y_{nN} - a(t_{nN})'\beta_N}{\sigma}, t_{nN}\right)\right| > \varepsilon\right\}\right) = 0 \tag{2.4}$$

for all $\varepsilon > 0$, $\sigma \in \mathbb{R}^+$ and $\beta_N = \beta + N^{-1/2}\overline{\beta}$ with $\beta, \overline{\beta} \in \mathbb{R}^r$ and $L\beta = l$.

For example, one-step M-estimators for $L\beta$ are asymptotically linear. These estimators generalize the M-estimators and have the form

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

where $\hat{\beta}^0_N$ is an initial estimator for β. In particular, (2.4) holds for one-step M-estimators if the score functions ψ satisfy condition (2.2), (2.3) and

(2.5) $\psi(z, t) = \psi_0(z, t) + h(t)\operatorname{sgn}(z)$, where, for all $t \in T_0$, $\psi_0(\cdot, t)$ is antisymmetric, continuous and there exists $\lambda_1(t), \dots, \lambda_\Lambda(t)$ so that $\psi_0(\cdot, t)$ has bounded and continuous derivatives on $\mathbb{R} \setminus \{\lambda_1(t), \dots, \lambda_\Lambda(t)\}$.

This can be seen by considering the results of Müller (1994b), who shows (2.4) for $\beta_N = \beta$, that is, $\overline{\beta} = 0$. Müller's proof, however, provides that (2.4) also holds for all $\beta_N = \beta + N^{-1/2}\overline{\beta}$ if $P^N_{\beta_N}$ is contiguous to P^N_β. This, for example, is the case for the normal distribution $P = n_{(0,1)}$. Hence, for the rest of the paper we will assume that $P^N_{\beta_N}$ is contiguous to P^N_β.

We frequently have ψ of the form $\psi(z, t) = Ma(t)\rho(z, t)$, where $M \in \mathbb{R}^{s\times r}$ and ρ: $\mathbb{R} \times \mathcal{T} \to \mathbb{R}$. In this case, (2.5) concerns only ρ. In particular, this is the case for M-estimators for β, where $L = E_{r\times r}$ ($E_{r\times r}$ denotes the identity matrix) and $M^{-1} = \int a(t)a(t)'\rho(z, t)zP(dz)\,\delta(dt)$. Note that score functions ρ satisfying (2.5) can be discontinuous and thus the corresponding M-functionals are not Fréchet differentiable.

The asymptotic covariance matrix of an asymptotically linear estimator is $\sigma^2 C(\psi, \delta)$, where

$$C(\psi, \delta) := \int \psi(z, t)\psi(z, t)'P(dz)\,\delta(dt).$$

Hence, in addition to the Gauss–Markov estimator, the matrix C^{LS}_N in the F-test statistic should be replaced by $\sigma^2 C(\psi, \delta)$ or by any C_N: $\mathbb{R}^N \times \mathcal{T}^N \to \mathbb{R}^{s\times s}$ such that $C_N(Y_N, d_N)$ converges in probability to $\sigma^2 C(\psi, \delta)$. The resulting generalized F-test statistic for testing H_0: $L\beta = l$ is then of Wald-type [cf. Wald (1943), Markatou, Stahel and Ronchetti (1991), Silvapulle (1992a), Heritier and Ronchetti (1994)]. Such tests are called *ALE-tests* and their corresponding test statistics are called *ALE-test statistics* because they are based on asymptotically (A) linear (L) estimators (E) [cf. Rieder (1994), page 153]. For testing subparameters, that is, hypotheses of the form H_0: $(0_{s\times r-s}\,|\,E_{s\times s})\beta = 0$, Wald-type tests and thus ALE-tests are asymptotically equivalent to the score-type tests [see Heritier and Ronchetti (1994), Markatou and Manos (1996)].

DEFINITION 2.2. A test statistic τ_N: $\mathbb{R}^N \times \mathscr{T}^N \to \mathbb{R}$ is called an ALE-test statistic with influence function ψ for testing H_0: $L\beta = l$ if

$$\tau_N(y_N, d_N) = N\big(\hat{\varphi}_N(y_N, d_N) - l\big)' C_N(y_N, d_N)^{-1}\big(\hat{\varphi}_N(y_N, d_N) - l\big),$$

where $\hat{\varphi}_N$ is asymptotically linear for contiguous alternatives of H_0: $L\beta = l$ with influence function ψ and $\lim_{N\to\infty} C_N = \sigma^2 C(\psi, \delta)$ in probability for $(P_\beta^N)_{N\in\mathbb{N}}$ with $L\beta = l$. A test based on an ALE-test statistic is called an ALE-test.

To derive the robustness and efficiency properties of ALE-tests, we study the asymptotic behavior of ALE-test statistics in contaminated linear models, where the contamination models the occurence of outliers. In contaminated linear models, the errors $Z_{1N}/\sigma, \ldots, Z_{NN}/\sigma$ are independent but distributed according to a contaminated distribution of the form

$$\begin{aligned} Q_{nN}(dz) &:= P^{Z_{nN}/\sigma}(dz) \\ &= \big(1 - N^{-1/2}\varepsilon c(t_{nN})\big)P(dz) + N^{-1/2}\varepsilon c(t_{nN}) g(z, t_{nN}) P(dz), \end{aligned}$$

where $\varepsilon > 0$, $c(t) \geq 0$, $g(z, t) \geq 0$, $\int c(t)\,\delta(dt) \leq 1$ and $\int g(z, t) P(dz) = 1$ for all $t \in \mathscr{T}$ and $z \in \mathbb{R}$ [compare with Bickel (1981, 1984) and Rieder (1985, 1987, 1994)]. The set of all sequences $(Q^N_{\varepsilon, c, g})_{N\in\mathbb{N}} = (Q^N)_{N\in\mathbb{N}} := (\bigotimes_{n=1}^N Q_{nN})_{N\in\mathbb{N}}$ defines the shrinking contamination neighborhood $\mathscr{U}_\varepsilon$ around the ideal (central) model $(P^N)_{N\in\mathbb{N}}$. If Q^N is the distribution of Z_N, then Q^N_β denotes the distribution of Y_N. To derive power properties, we regard not only fixed β with $L\beta = l$ but also sequences $(\beta_N)_{N\in\mathbb{N}}$ of contiguous alternatives.

THEOREM 2.1. *If τ_N is an ALE-test statistic for testing H_0: $L\beta = l$ with influence function ψ, then τ_N has an asymptotic chi-squared distribution, that is,*

$$\begin{aligned} \mathscr{L}(\tau_N | Q^N_{\beta_N}) \longrightarrow \chi^2\Big(s, &\big[\gamma + \sigma b\big(\psi, (Q^N)_{N\in\mathbb{N}}\big)\big]' \big[\sigma^2 C(\psi, \delta)\big]^{-1} \\ &\times \big[\gamma + \sigma b\big(\psi, (Q^N)_{N\in\mathbb{N}}\big)\big]\Big) \qquad \text{as } N \to \infty \end{aligned}$$

for all $(Q^N)_{N\in\mathbb{N}} = (Q^N_{\varepsilon, c, g})_{N\in\mathbb{N}} \in \mathscr{U}_\varepsilon$ and all $\beta_N = \beta + N^{-1/2}\overline{\beta} \in \mathbb{R}^r$ with $L\beta_N = l + N^{-1/2}\gamma$, where

$$b\big(\psi, (Q^N_{\varepsilon, c, g})_{N\in\mathbb{N}}\big) := \varepsilon \int \psi(z, t) c(t) g(z, t) P(dz)\,\delta(dt)$$

and $\chi^2(s, \lambda)$ *is the chi-squared distribution with s degrees of freedom and noncentrality parameter* λ.

3. Efficiency and robustness of ALE-tests. Because $b(\psi, (Q^N)_{N\in\mathbb{N}}) = 0$ if $(Q^N)_{N\in\mathbb{N}} = (P^N)_{N\in\mathbb{N}}$, Theorem 2.1 provides the asymptotic level and the asymptotic power for ideal distributions P^N. In particular, the rejection set should be chosen to be $(\chi^2_{1-\alpha,\,s,\,0}, \infty)$ in order to obtain an asymptotic α-level test for the ideal distribution. Here $\chi^2_{1-\alpha,\,s,\,0}$ refers to the $(1-\alpha)$-quantile of the chi-squared distribution $\chi^2(s, 0)$. The asymptotic power for contiguous ideal alternatives $(P^N_{\beta_N})_{N\in\mathbb{N}}$ with $L\beta_N = l + N^{-1/2}\gamma \neq l$ is thus given by

$$\lim_{N\to\infty} P^N_{\beta_N}\big(\tau_N > \chi^2_{1-\alpha,\,s,\,0}\big) = 1 - \mathscr{X}^2_{s,\,\gamma'[\sigma^2 C(\psi,\delta)]^{-1}\gamma}\big(\chi^2_{1-\alpha,\,s,\,0}\big), \tag{3.1}$$

where $\mathscr{X}_{s,\,\lambda}$ is the distribution function of the chi-squared distribution $\chi^2(s, \lambda)$. We will regard this asymptotic power (3.1) as a measure for the efficiency of an ALE-test. This power is maximized when $\gamma'[\sigma^2 C(\psi, \delta)]^{-1}\gamma$ is maximized for all $\gamma \in \mathbb{R}^s$.

As a measure for the robustness of an ALE-test, we regard its maximum bias of the asymptotic level for contaminated distributions Q^N_β with $L\beta = l$, that is,

$$b_\varepsilon(\psi) := \max\Big\{\lim_{N\to\infty}\big(Q^N_\beta\big(\tau_N > \chi^2_{1-\alpha,\,s,\,0}\big) - \alpha\big);\ (Q^N)_{N\in\mathbb{N}} \in \mathscr{U}_\varepsilon \text{ and } \beta \in \mathbb{R}^r \text{ with } L\beta = l\Big\}.$$

The bias $b_\varepsilon(\psi)$ is bounded if and only if $\psi' C(\psi, \delta)^{-1}\psi$ is bounded. This is shown in the following theorem.

THEOREM 3.1. *For an ALE-test for testing* H_0: $L\beta = l$ *with influence function* ψ, *the following inequalities are equivalent*:

(i) $$b_\varepsilon(\psi) \le 1 - \mathscr{X}^2_{s,b}(\chi^2_{1-\alpha,\,s,\,0}) - \alpha,$$

(ii) $$\varepsilon^2 \|\psi' C(\psi, \delta)^{-1}\psi\|_\delta \le b,$$

where $\|\psi' C(\psi, \delta)^{-1}\psi\|_\delta := \max_{(z,\,t)\in\mathbb{R}\times\operatorname{supp}(\delta)} \psi(z, t)' C(\psi, \delta)^{-1}\psi(z, t)$.

Theorem 3.1 shows that the asymptotic bias of a test based on a M-estimator for β is bounded if the self-standardized gross-error-sensitivity of the M-estimator is bounded [for the definition of the self-standardized gross-error-sensitivity see Krasker and Welsch (1982), Ronchetti and Rousseeuw (1985), Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. This result corresponds to a result of Heritier and Ronchetti (1994), who derived the robustness of tests via influence functions by assuming Fréchet differentiability of the test statistic.

In Sections 4 and 5, we consider the problem of deriving tests and designs which maximize the asymptotic power and minimize the maximum asymptotic bias. This problem results in optimization problems within the set of influence functions because the maximum asymptotic bias of the level and the asymptotic power depend on the ALE-test only via its influence function. The influence functions are typically of the form $\psi(z,t) = Ma(t)\rho(z,t)$ (see Section 2). Therefore, we will look at optimization problems for

$$\begin{aligned}\Psi^*(\delta, L) := \big\{\psi\colon \mathbb{R}\times\mathscr{T} \to \mathbb{R}^s;\ &\psi \text{ satisfies conditions (2.1)–(2.3) and}\\ &\psi(z,t) = Ma(t)\rho(z,t) \text{ for all } (z,t)\in\mathbb{R}\times\mathscr{T} \text{ for some}\\ &M\in\mathbb{R}^{s\times r} \text{ and } \rho\colon \mathbb{R}\times\mathscr{T}\to\mathbb{R}\big\}.\end{aligned}$$

It is not clear if every ψ lying in $\Psi^*(\delta, L)$ is the influence function of an ALE-test. But if we find an optimal solution ψ_* in $\Psi^*(\delta, L)$ which satisfies the conditions (2.2),(2.3) and (2.5), then we know from the remarks above that a one-step M-estimator with score function ψ_* is asymptotically linear with influence function ψ_*; that is, ψ_* is also an influence function of an ALE-test. We will see that this is always the case.

In order to find the optimal design, we will regard the set of all designs, such that the linear aspect $\varphi(\beta) = L\beta$ is identifiable; that is,

$$\Delta(L) := \{\delta;\ \text{there exists } K\in\mathbb{R}^{s\times r} \text{ with } L = K\ I(\delta)\}.$$

We will occasionally consider the set of designs with fixed support $\mathscr{D}\subset\mathscr{T}$ as well; that is,

$$\Delta_{\mathscr{D}} := \big\{\delta;\ \operatorname{supp}(\delta) = \mathscr{D}\big\}.$$

4. Most-robust tests and designs. In Theorem 3.1, it was shown that the asymptotic bias of the level of an α level ALE-test for testing $H_0\colon L\beta = l$ is a strictly increasing function of $\|\psi' C(\psi,\delta)^{-1}\psi\|_\delta$, so that most-robust tests and designs should minimize $\|\psi' C(\psi,\delta)^{-1}\psi\|_\delta$. Thereby, let

$$b_0^T(\delta, L) := \min\big\{\big\|\psi' C(\psi,\delta)^{-1}\psi\big\|_\delta;\ \psi\in\Psi^*(\delta, L)\big\}$$

be the minimum asymptotic bias for testing $L\beta$ at δ.

Definition 4.1. (a) An ALE-test with influence function ψ_0 is most-robust for testing $L\beta$ at $\delta\in\Delta(L)$ if

$$\psi_0 \in \arg\min\big\{\big\|\psi' C(\psi,\delta)^{-1}\psi\big\|_\delta;\ \psi\in\Psi^*(\delta, L)\big\}.$$

(b) A design δ_0 is most-robust for testing $L\beta$ in $\Delta \subset \Delta(L)$ if

$$\delta_0 \in \arg\min\{b_0^T(\delta, L);\ \delta \in \Delta\}.$$

Note that the optimization criterion $\|\psi' C(\psi, \delta)^{-1}\psi\|_\delta$ is not convex in ψ, but it is invariant with respect to nonsingular transformations of $L\beta$.

The following theorem shows that the highest robustness is achieved at the classical D-optimal designs. Recall that the classical D-optimal design is a design for which the determinant of the covariance matrix of the nonrobust Gauss–Markov estimator and the volume of the ellipsoid given by a fixed power value of the nonrobust F-test are minimized. When we generalize the designs d_N to design measures δ, this means that a design δ_D is D-optimal for $L\beta$ in a set Δ of designs if and only if

$$\delta_D \in \arg\min\{\det(LI(\delta)^- L');\ \delta \in \Delta\},$$

where $\det(M)$ is the determinant of a matrix M, M^- is the g-inverse of a matrix M; that is, $MM^-M = M$, and $I(\delta) := \int a(t)a(t)'\delta(dt)$ is the information matrix of the design δ [see, e.g., Pázman (1986), page 100 or Pukelsheim (1993), page 135].

THEOREM 4.1. (i) *We have* $b_0^T(\delta, L) \geq s$ *for all* $\delta \in \Delta(L)$.

(ii) *If* δ_D *is* D*-optimal for* $L\beta$ *in* $\Delta_{\mathscr{D}} \subset \Delta(L)$, *then* $b_0^T(\delta_D, L) = s$; *that is,* δ_D *is most-robust for testing* $L\beta$ *in* $\Delta(L)$, *and* ψ_{01} *given by* $\psi_{01}(z, t) := LI(\delta_D)^- a(t)\operatorname{sgn}(z)\sqrt{\pi/2}$ *is the influence function of a most-robust ALE-test for testing* $L\beta$ *at* δ_D.

Note that ψ_{01} is the influence function of the L_1 estimator and that the M-functional of this estimator is not Fréchet differentiable. Hence ψ_{01} cannot be treated using the approach of Heritier and Ronchetti (1994).

Theorem 4.1 holds also for designs which are D-optimal in a modified model. In this case, we define for a function $h\colon \mathscr{T} \to \mathbb{R}^+ \setminus \{0\}$

$$a_h(t) := h(t)^{-1}a(t), \qquad L_h := LJ_h^- I_h(\delta),$$

where

$$J_h := \int a_h(t)a(t)'\delta(dt), \qquad I_h(\delta) := \int a_h(t)a_h(t)'\,\delta(dt).$$

COROLLARY 4.2. *If there exists some function* $h\colon \mathscr{T} \to \mathbb{R}^+ \setminus \{0\}$ *so that* δ *is* D*-optimal in* $\Delta_{\operatorname{supp}(\delta)} \subset \Delta(L)$ *for* $\varphi_h(\beta) = L_h\beta$ *in the modified model* $Y_h(t) = a_h(t)'\beta + Z$, *then* $b_0^T(\delta, L) = s$; *that is,* δ *is most-robust for testing* $L\beta$ *in* $\Delta(L)$, *and* ψ_0 *given by* $\psi_0(z, t) = L_h I_h(\delta)^- a_h(t)\operatorname{sgn}(z)\sqrt{\pi/2}$ *is the influence function of a most-robust ALE-test for testing* $L\beta$ *at* δ.

Note that Theorem 4.1 and Corollary 4.2 are analogous to Lemma 1(a) in Müller (1994a) and Theorem 2 in Müller (1992b), respectively, which show that for estimation, the minimum bias is attained at A-optimal designs. Using

$h(t) = |a(t)|$, Theorem 4.1(i) and Corollary 4.2 yield Theorem 2′ of Ronchetti and Rousseeuw (1985) [or Proposition 1(ii) in Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 318] which concerns estimators with minimum self-standardized gross-error-sensitivity.

5. Most-efficient robust tests and designs. A good test should not only have a small bias of the level but also a high efficiency, that is, a high power at least for the ideal model. Because small bias and high power are contrary properties, we use the usual approach of maximizing the power at the ideal model under the side condition that the bias is bounded by some bias bound. Because the asymptotic power at the ideal model and the maximum asymptotic bias of an ALE-test are given by $\gamma' C(\psi, \delta)^{-1}\gamma$ with $\gamma \in \mathbb{R}^s$ and $\|\psi' C(\psi, \delta)^{-1}\psi\|_\delta$, respectively, we will maximize $\gamma' C(\psi, \delta)^{-1}\gamma$ for all $\gamma \in \mathbb{R}^s$ under the side condition that $\|\psi' C(\psi, \delta)^{-1}\psi\|_\delta \le b$. If we have no bias bound, that is, $b = \infty$, then the influence function of the Gauss–Markov estimator, that is, ψ_{GM} given by $\psi_{GM}(z, t) := LI(\delta)^- a(t)z$, maximizes $\gamma' C(\psi, \delta)^{-1}\gamma$ within $\Psi^*(\delta, L)$ for all $\gamma \in \mathbb{R}^s$. But for $b < \infty$, it is not possible in general to find an influence function which maximizes $\gamma' C(\psi, \delta)^{-1}\gamma$ uniformly for all $\gamma \in \mathbb{R}^s$, that is, which minimizes $C(\psi, \delta)$ in the positive-semidefinite sense. We can characterize admissible solutions, but in order to find optimal solutions, we have to consider a real-valued function of $C(\psi, \delta)$. For this real-valued function, we will use the determinant, since the minimization of $\det(C(\psi, \delta))$ is invariant with respect to nonsingular transformations of the linear aspect $\varphi(\beta) = L\beta$. Then, $\det(C(\psi, \delta))$ is not convex in ψ. However, the side condition is not convex in ψ as well. Therefore, the optimization results from convex analysis are not applicable in any case. As in classical design theory, the solutions which minimize $\det(C(\psi, \delta))$ are called "D-optimal."

DEFINITION 5.1. (a) An ALE-test with influence function $\psi_{b,\delta}$ is admissible for testing $L\beta$ at δ with bias bound b if $\|\psi'_{b,\delta} C(\psi_{b,\delta}, \delta)^{-1}\psi_{b,\delta}\|_\delta \le b$ and $C(\psi, \delta) = C(\psi_{b,\delta}, \delta)$ for all $\psi \in \Psi^*(\delta, L)$ with $\|\psi' C(\psi, \delta)^{-1}\psi\|_\delta \le b$ and $C(\psi, \delta) \le C(\psi_{b,\delta}, \delta)$.

(b) An ALE-test with influence function $\psi_{b,\delta}$ is D-optimal for testing $L\beta$ at δ with bias bound b if

$$\psi_{b,\delta} \in \arg\min\left\{\det\bigl(C(\psi, \delta)\bigr) : \psi \in \Psi^*(\delta, L) \text{ with } \left\|\psi' C(\psi, \delta)^{-1}\psi\right\|_\delta \le b\right\}.$$

(c) A design δ_b is D-optimal in Δ for testing $L\beta$ with bias bound b if

$$\delta_b \in \arg\min\Bigl\{\min\bigl\{\det(C(\psi, \delta));\ \psi \in \Psi^*(\delta, L) \text{ with } \|\psi' C(\psi, \delta)^{-1}\psi\|_\delta \le b\bigr\};\ \delta \in \Delta\Bigr\}.$$

In deriving admissible and D-optimal tests, we will assume that the ideal distribution P is the standard normal distribution, that is, $P = n_{(0,1)}$. The distribution function of $n_{(0,1)}$ will be denoted by Φ.

We first characterize the influence function of an admissible ALE-test by generalizing a result of Krasker and Welsch (1982) concerning admissible estimators for estimating β with a self-standardized gross-error-sensitivity bounded by b. For this we use the following matrices:

$$M_b(B) := \int a(t)a(t)'\left[2\Phi\left(\frac{\sqrt{b}}{|Ba(t)|}\right) - 1\right]\delta(dt),$$

$$Q_b(B) := \int a(t)a(t)' g\left(\frac{\sqrt{b}}{|Ba(t)|}\right)\delta(dt),$$

$$M_{b,0}(B) := \int a(t)a(t)'\frac{\sqrt{b}}{|Ba(t)|}\sqrt{2/\pi}\,\delta(dt),$$

$$Q_{b,0}(B) := \int a(t)a(t)'\left(\frac{\sqrt{b}}{|Ba(t)|}\right)^2\delta(dt).$$

THEOREM 5.1. *Let $\delta \in \Delta(L)$, $B_b \in \mathbb{R}^{s\times r}$ and*

$$\psi_{b,\delta}(z,t) := LM_b(B_b)^- a(t)\min\left\{|z|, \frac{\sqrt{b}}{|B_b a(t)|}\right\}\operatorname{sgn}(z), \tag{5.1}$$

where

$$\begin{aligned} a(t)'B_b'B_b a(t) = a(t)'M_b(B_b)^- L'\left[LM_b(B_b)^- Q_b(B_b)M_b(B_b)^- L'\right]^{-1} \\ \times LM_b(B_b)^- a(t) \quad \textit{for all } t \in \operatorname{supp}(\delta) \end{aligned} \tag{5.2}$$

or

$$\psi_{b,\delta}(z,t) := LM_{b,0}(B_b)^- a(t)\frac{\sqrt{b}}{|B_b a(t)|}\operatorname{sgn}(z), \tag{5.3}$$

where

$$\begin{aligned} &a(t)'B_b'B_b a(t) \\ &\quad = a(t)'M_{b,0}(B_b)^- L'\left[LM_{b,0}(B_b)^- Q_{b,0}(B_b)M_{b,0}(B_b)^- L'\right]^{-1} \\ &\qquad \times LM_{b,0}(B_b)^- a(t) \quad \textit{for all } t \in \operatorname{supp}(\delta). \end{aligned} \tag{5.4}$$

Then the ALE-test with influence function $\psi_{b,\delta}$ is admissible for testing $L\beta$ at δ with bias bound b.

Theorem 5.1 can be used to show that at a D-optimal design δ_D, the influence functions of admissible ALE-tests for testing $L\beta$ with bias bound b have a very simple form. These influence functions are given by

$$\psi_{b,\delta_D}(z,t) = \begin{cases} LI(\delta_D)^- a(t)\operatorname{sgn}(z)\sqrt{\pi/2}, & \text{for } b = s, \\ LI(\delta_D)^- a(t)\operatorname{sgn}(z)\dfrac{\min\{|z|, \sqrt{b}y_b\}}{2\Phi(\sqrt{b}y_b) - 1}, & \text{for } b > s, \end{cases} \tag{5.5}$$

where

$$y_b^2 = \frac{1}{s}g\left(\sqrt{b}y_b\right) > 0 \tag{5.6}$$

with

$$g(y) := \int \min\{|z|, y\}^2 P(dz).$$

This is stated in the following lemma. This lemma is analogous to Lemma 1(b) in Müller (1994a), which shows that for estimation with bias bound at A-optimal designs, the influence functions of A-optimal AL-estimators have a simple form. Thereby, instead of a formula for the trace of the asymptotic covariance matrix, we have a simple formula for the asymptotic covariance matrix of the underlying AL-estimator. For this purpose, we define $\overline{W}$: $(0, \infty)\times(0, \infty) \to \mathbb{R}$ by $\overline{W}(c, y) = g(y)/c - y^2$ and $\overline{w}$: $(0, 1) \to (0, \infty)$ implicitely by $\overline{W}(c, \overline{w}(c)) = 0$. Note that y_b of (5.6) satisfies $y_b = (1/\sqrt{b})\overline{w}(s/b)$. Set also

$$\overline{v}(c) = \begin{cases} \dfrac{(2\Phi(\overline{w}(c)) - 1)^2}{g(\overline{w}(c))}, & \text{for } c < 1, \\ \dfrac{2}{\pi}, & \text{for } c = 1. \end{cases}$$

Note also that a solution of (5.6) always exists and can be easily calculated. This is because the function $\overline{\omega}$: $(0, \infty) \to (0, \infty)$ given by $\overline{\omega}(y) = g(y)/c - y^2$ for $c > 0$ has at most one positive root which exists if and only if $c < 1$. The root lies in $S := \{y \in [0, \infty);\ 2\Phi(-y) \le c\}$, where $\overline{\omega}$ is concave on S. Hence, the root can be calculated by Newton's method starting with an interior point of S.

LEMMA 5.2. *If δ_D is D-optimal for $L\beta$ in $\Delta_{\mathscr{D}} \subset \Delta(L)$ and $b \ge s$, then the ALE-test with influence function ψ_{b,δ_D} given by* (5.5) *is admissible for testing $L\beta$ with bias bound b at δ_D and its asymptotic covariance matrix satisfies*

$$C(\psi_{b,\delta_D}, \delta_D) = \frac{1}{\overline{v}(s/b)} LI(\delta_D)^- L'.$$

If $a(\tau_1), \dots, a(\tau_I)$ are linearly independent and δ is not D-optimal in $\Delta_{\mathrm{supp}(\delta)}$, then we have a modification of Lemma 5.2 which gives at least a lower bound for the asymptotic covariance matrix. For ψ with $\psi(z, t) = Ma(t)\rho(z, t)$, set

$$v_\rho(t) := \frac{\left(\int \rho(z, t) z P(dz)\right)^2}{\int \rho(z, t)^2 P(dz)}$$

and

$$\delta_\rho(\{t\}) := \frac{1}{\sum_{\tau\in\mathscr{D}} v_\rho(\tau)\delta(\{\tau\})}\, v_\rho(t)\delta(\{t\}).$$

LEMMA 5.3. *If $\delta \in \Delta_{\mathscr{D}} \subset \Delta(L)$, $\mathscr{D} = \{\tau_1, \dots, \tau_I\}$, $a(\tau_1), \dots, a(\tau_I)$ are linearly independent and $\psi \in \Psi^*(\delta, L)$ with $\|\psi' C(\psi, \delta)^{-1}\psi\|_\delta \le b$ and $\psi(z, t) =$*

$Ma(t)\rho(z,t)$, then

$$C(\psi,\delta) \geq \frac{1}{\overline{v}(s/b)} LI(\delta_\rho)^{-}L'.$$

Using Lemmas 5.2 and 5.3, we can show that at a D-optimal design δ_D we have D-optimality of an ALE-test with influence function ψ_{b,δ_D} given by (5.5). Moreover, the D-optimal design is also D-optimal for testing with bias bound b in $\Delta_{\mathscr{D}}$. Furthermore, this holds not only within $\Delta_{\mathscr{D}}$ with linearly independent regressors but also for any D-optimal design within designs of $\Delta(L)$ with finite support. Therefore, we have a theorem which is analogous to the theorem in Müller (1994a), which shows that the A-optimal designs are also optimal for robust estimation.

THEOREM 5.4. *If $\Delta_0 = \{\delta \in \Delta(L); \mathrm{supp}(\delta)$ is finite$\}$, $b \geq s$ and δ_D is D-optimal for $L\beta$ in Δ_0, then δ_D is D-optimal in Δ_0 for testing $L\beta$ with bias bound b and an ALE-test with influence function ψ_{b,δ_D} given by* (5.5) *is asymptotically D-optimal for testing $L\beta$ at δ_D with bias bound b.*

6. Examples.

EXAMPLE 6.1 (One-way layout). In a one-way layout model with four levels, the observation at $t_{nN} = i$ is given by

$$Y_{nN} = \beta_i + Z_{nN} = a(t_{nN})'\beta + Z_{nN}$$

for $n = 1, \ldots, N$, where

$$\beta = (\beta_1, \beta_2, \beta_3, \beta_4)' \in \mathbb{R}^4,$$

$$a(t) = \big(1_1(t), 1_2(t), 1_3(t), 1_4(t)\big)' \in \mathbb{R}^4$$

and

$$\mathscr{T} = \{1, 2, 3, 4\}.$$

The hypothesis H_0: $L\beta := (\beta_2 - \beta_1, \beta_3 - \beta_1, \beta_4 - \beta_1)' = 0$, or equivalently H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4$, versus H_1: $L\beta = (\beta_2 - \beta_1, \beta_3 - \beta_1, \beta_4 - \beta_1)' \neq 0$ is usually tested in this model. The D-optimal design for $L\beta$ in $\Delta(L)$ is, according to Lemma 7.1,

$$\delta_D = \tfrac{1}{4}(e_1 + e_2 + e_3 + e_4).$$

According to Theorem 4.1, this design δ_D is most-robust for testing H_0: $L\beta = 0$, that is,

$$b_0^T(\delta_D, L) = 3 = \min\{b_0^T(\delta, L); \delta \in \Delta(L)\},$$

and the influence function of a most-robust ALE-test at δ_D has the form

$$\psi_{01}(z,t) = \begin{cases} (-1,-1,-1)'\,\mathrm{sgn}(z)\,4\sqrt{\pi/2}, & \text{for } t = 1, \\ (1_2(t), 1_3(t), 1_4(t))'\,\mathrm{sgn}(z)\,4\sqrt{\pi/2}, & \text{for } t \neq 1. \end{cases} \tag{6.1}$$

According to Theorem 5.4, for every $b \geq b_0^T(\delta_D, L) = 3$, the D-optimal design δ_D is also D-optimal for testing with bias bound b, where the influence function of a D-optimal ALE-test for testing at δ_D with bias bound b is given by (5.5). For $b = b_0^T(\delta_D, L) = 3$, it coincides with ψ_{01} given by (6.1), and for $b > b_0^T(\delta_D, L) = 3$, it has the form

$$\psi_{b,\,\delta_D}(z, t) = \begin{cases} (-1, -1, -1)' \operatorname{sgn}(z)\, 4 \dfrac{\min\{|z|, \sqrt{b} v_b\}}{2\Phi(\sqrt{b} v_b) - 1}, & \text{for } t = 1, \\[2ex] (1_{\{2\}}(t), 1_{\{3\}}(t), 1_{\{4\}}(t))' \operatorname{sgn}(z)\, 4 \dfrac{\min\{|z|, \sqrt{b} v_b\}}{2\Phi(\sqrt{b} v_b) - 1}, & \text{for } t \neq 1, \end{cases}$$

with $v_b^2 = (1/3)\ g(\sqrt{b} v_b) > 0$; that is, $v_b = (1/\sqrt{b})\overline{w}(3/b)$. For example, for $b = 4$ we have $v_b = 0.240$.

EXAMPLE 6.2 (Quadratic regression). In a quadratic regression model, the observation at t_{nN} is given by

$$Y_{nN} = \beta_0 + \beta_1 t_{nN} + \beta_2 t_{nN}^2 + Z_{nN} = a(t_{nN})'\beta + Z_{nN}$$

for $n = 1, \dots, N$, where $\beta = (\beta_0, \beta_1, \beta_2)' \in \mathbb{R}^3$ and $a(t) = (1, t, t^2)' \in \mathbb{R}^3$. Assume that $\mathscr{T} = [0, 1]$ and that the interesting aspect is $L\beta := (\beta_0, \beta_1 + \beta_2)' \in \mathbb{R}^2$. In particular, testing the hypothesis H_0: $L\beta = (l_1, 0)'$ means that we test the hypothesis that the quadratic function is symmetric on $[0, 1]$ and is equal to l_1 at the end points $t = 0$ and $t = 1$. The D-optimal design for $L\beta$ in $\Delta_{\mathscr{D}}$ with $\mathscr{D} = \{0, 1\}$ is, according to Lemma 7.1,

$$\delta_D = \tfrac{1}{2}(e_0 + e_1).$$

Hence, according to Theorem 4.1, δ_D is most-robust for testing H_0: $L\beta = 0$ in $\Delta(L)$ with

$$b_0^T(\delta_D, L) = 2 = \min\{b_0^T(\delta, L); \delta \in \Delta(L)\},$$

and the influence function of a most-robust ALE-test at δ_D has the form

$$\psi_{01}(z, t) = \begin{cases} (1, -1)' \operatorname{sgn}(z)\, 2\sqrt{\pi/2}, & \text{for } t = 0, \\ (0, 1)' \operatorname{sgn}(z)\, 2\sqrt{\pi/2}, & \text{for } t = 1. \end{cases} \tag{6.2}$$

According to Theorem 5.4, for every $b \geq b_0^T(\delta_D, L) = 2$, the D-optimal design δ_D is also D-optimal for testing with bias bound b, where the influence function of a D-optimal ALE-test statistic for testing at δ_D with bias bound b is given by (5.5). For $b = b_0^T(\delta_D, L) = 2$, this influence function coincides with ψ_{01} given by (6.2), and for $b > b_0^T(\delta_D, L) = 2$, it has the form

$$\psi_{b,\,\delta_D}(z, t) = \begin{cases} (1, -1)' \operatorname{sgn}(z)\, 2 \dfrac{\min\{|z|, \sqrt{b}\ v_b\}}{2\Phi(\sqrt{b} v_b) - 1}, & \text{for } t = 0, \\[2ex] (0, 1)' \operatorname{sgn}(z)\, 2 \dfrac{\min\{|z|, \sqrt{b}\ v_b\}}{2\Phi(\sqrt{b} v_b) - 1}, & \text{for } t = 1, \end{cases}$$

with $v_b^2 = (1/2)g(\sqrt{b}v_b) > 0$, that is, $v_b = (1/\sqrt{b})\overline{w}(2/b)$. For example, for $b = 3$ we have $v_b = 0.377$.

7. Proofs.

PROOF OF THEOREM 2.1. Consider the following part of τ_N:

$$\sqrt{N}C_N^{-1/2}(\hat{\varphi}_N - l) = \sqrt{N}C_N^{-1/2}(\hat{\varphi}_N - \varphi(\beta_N)) + C_N^{-1/2}\gamma.$$

Because $(P_{\beta_N}^N)_{N\in\mathbb{N}}$ is contiguous to $(P_\beta^N)_{N\in\mathbb{N}}$, the covariance estimator C_N converges to $\sigma^2 C(\psi, \delta)$ in probability $(P_{\beta_N}^N)_{N\in\mathbb{N}}$. Thus, because $\hat{\varphi}_N$ is weakly asymptotically linear for contiguous alternatives, the distribution $\mathscr{L}(\sqrt{N}C_N^{-1/2}(\hat{\varphi}_N - \varphi(\beta_N))|P_{\beta_N}^N)$ behaves asymptotically like

$$\mathscr{L}\left(N^{-1/2}\left[\sigma^2 C(\psi, \delta)\right]^{-1/2} \sum_{n=1}^{N} \sigma\psi\left(\frac{y_{nN} - a(t_{nN})'\beta_N}{\sigma}, t_{nN}\right)\middle| P_{\beta_N}^N\right)$$
$$= \mathscr{L}\left(N^{-1/2} \sum_{n=1}^{N} C(\psi, \delta)^{-1/2}\psi(z_{nN}, t_{nN})\middle| P^N\right).$$

The third lemma of LeCam [see Hájek and Šidák (1967), page 208] states that $\mathscr{L}(N^{-1/2}\sum_{n=1}^N C(\psi, \delta)^{-1/2}\psi(z_{nN}, t_{nN})|Q^N)$ is asymptotically normally distributed with mean $C(\psi, \delta)^{-1/2}b(\psi, (Q^N)_{N\in\mathbb{N}})$ and covariance matrix $E_{s\times s}$, and in particular, that $(Q^N)_{N\in\mathbb{N}}$ is contiguous to $(P^N)_{N\in\mathbb{N}}$ for all $(Q^N)_{N\in\mathbb{N}} \in \mathscr{U}_\varepsilon$. Then $(Q_{\beta_N}^N)_{N\in\mathbb{N}}$ is also contiguous to $(P_{\beta_N}^N)_{N\in\mathbb{N}}$ such that $C_N^{-1/2}\gamma$ converges to $[\sigma^2 C(\psi, \delta)]^{-1/2}\gamma$ in probability for $(Q_{\beta_N}^N)_{N\in\mathbb{N}}$, and

$$\mathscr{L}\left(\sqrt{N}\left[\sigma^2 C(\psi, \delta)\right]^{-1/2}(\hat{\varphi}_N - l))\middle| Q_{\beta_N}^N\right)$$

is asymptotically normally distributed with mean $[\sigma^2 C(\psi, \delta)]^{-1/2}(\gamma + \sigma b(\psi, (Q^N)_{N\in\mathbb{N}})$ and covariance matrix $E_{s\times s}$. □

PROOF OF THEOREM 3.1. From Theorem 2.1 and the fact that $\mathscr{X}_{s,b}(k)$ is decreasing in b for all $k > 0$ we get at once that (i) is equivalent to

$$\max\left\{b(\psi, (Q^N)_{N\in\mathbb{N}})' C(\psi, \delta)^{-1} b(\psi, (Q^N)_{N\in\mathbb{N}}); \; (Q^N)_{N\in\mathbb{N}} \in \mathscr{U}_\varepsilon\right\} \le b. \tag{7.1}$$

The equivalence of (7.1) and (ii) follows from the results in Rieder (1985, 1994) and Kurotschka and Müller (1992) concerning the maximum asymptotic bias of asymptotically linear estimators in shrinking contamination neighborhoods by interpreting the function $C(\psi, \delta)^{-1/2}\psi$ as influence function. □

To derive the assertions for D-optimal designs, we need a modification of the equivalence theorem for D-optimality of Kiefer and Wolfowitz (1960), which was shown in Müller [(1997), page 15].

LEMMA 7.1. *Let $\Delta_{\mathscr{D}} \subset \Delta(L)$. Then δ_D is D-optimal for $L\beta$ in $\Delta_{\mathscr{D}}$ if and only if*

$$a(t)' I(\delta_D)^- L'[LI(\delta_D)^- L']^{-1} LI(\delta)^- a(t) = s \quad \text{for all } t \in \mathscr{D}.$$

PROOF OF THEOREM 4.1. Assertion (i) follows from a straightforward generalization of the proof of Theorem 2′ in Ronchetti and Rousseeuw (1985) [or Proposition 1(ii) in Hampel et al. (1986), page 318], and assertion (ii) follows from using Lemma 7.1. □

PROOF OF THEOREM 5.1. Because $L = KM_b(B_b)$ and $L = K_0 M_{b,0}(B_b)$ for some $K, K_0 \in \mathbb{R}^{s\times r}$, it is easy to see that $\psi_{b,\delta}$ is an element of $\Psi^*(\delta, L)$ with $\|\psi'_{b,\delta} C(\psi_{b,\delta}, \delta)^{-1} \psi_{b,\delta}\|_\delta \le b$. In order to show the admissibility of $\psi_{b,\delta}$, the proof of Krasker and Welsch (1982) is generalized.

At first assume

$$\psi_{b,\delta}(z,t) := LM_b(B_b)^- a(t) \min\left\{|z|, \frac{\sqrt{b}}{|B_b a(t)|}\right\} \operatorname{sgn}(z).$$

Set $V := LM_b(B_b)^- Q_b(B_b) M_b(B_b)^- L'$, $Q := V^{-1/2} LM_b(B_b)^-$, $D := V^{-1/2}L$ and $\psi_B(z,t) := V^{-1/2}\psi_{b,\delta}$. Because of (5.2), one obtains $|B_b a(t)| = |Qa(t)|$ for all $t \in \operatorname{supp}(\delta)$ and

$$\psi_B(z,t) = Qa(t) \min\left\{|z|, \frac{\sqrt{b}}{|Qa(t)|}\right\} \operatorname{sgn}(z).$$

Because of $\int \psi_B(z,t) a(t)' z P(dz)\, \delta(dt) = D$, Theorem 1 in Kurotschka and Müller (1992) provides

$$\psi_B \in \arg\min\bigl\{\operatorname{tr}\bigl(C(\psi,\delta)\bigr); \psi \in \Psi^*(\delta, D) \text{ with } \|\psi'\psi\|_\delta \le b\bigr\}. \tag{7.2}$$

Assume that there exists $\psi_0 \in \Psi^*(\delta, L)$ with

$$C(\psi_0, \delta) < C(\psi_{b,\delta}, \delta) \tag{7.3}$$

and

$$\bigl\|\psi'_0 C(\psi_0, \delta)^{-1} \psi_0\bigr\|_\delta \le b. \tag{7.4}$$

Then $\int V^{-1/2}\psi_0(z,t) a(t)' z P(dz)\, \delta(dt) = V^{-1/2} L = D$ and

$$V^{-1/2}\, C(\psi_0, \delta) V^{-1/2} < V^{-1/2} C(\psi_{b,\delta}, \delta) V^{-1/2} = E_{r\times r},$$

which, with (7.4), implies

$$\bigl\|\psi'_0 V^{-1/2} V^{-1/2} \psi_0\bigr\|_\delta \le \Bigl\|\psi'_0 V^{-1/2}\bigl(V^{-1/2} C(\psi_0,\delta) V^{-1/2}\bigr)^{-1} V^{-1/2}\psi_0\Bigr\|_\delta \le b.$$

Hence, property (7.2) provides for $V^{-1/2}\psi_0$,

$$\operatorname{tr}\bigl(V^{-1/2} C(\psi_0, \delta) V^{-1/2}\bigr) \ge \operatorname{tr}\bigl(V^{-1/2} C(\psi_{b,\delta}, \delta) V^{-1/2}\bigr),$$

which is a contradiction of (7.3).

For

$$\psi_{b,\delta}(z,t) := LM_{b,0}(B_b)^- a(t) \frac{\sqrt{b}}{|B_b a(t)|}\operatorname{sgn}(z),$$

the assertion follows as above. We have only to replace $M_b(B_b)$ and $Q_b(B_b)$ by $M_{b,0}(B_b)$ and $Q_{b,0}(B_b)$. □

PROOF OF LEMMA 5.2. Set

$$B_b = \begin{cases} \big(LI(\delta_D)^- L'\big)^{-1/2} LI(\delta_D)^- \dfrac{1}{\sqrt{s}y_b}, & \text{for } b > s,\\ \big(LI(\delta_D)^- L'\big)^{-1/2} LI(\delta_D)^-, & \text{for } b = s.\end{cases}$$

Then the equivalence theorem for D-optimality (Lemma 7.1) provides $|B_b a(t)| = 1/y_b$ for $b > s$ and $|B_b a(t)| = \sqrt{s}$ for $b = s$ for all $t \in \mathscr{D}$ so that the admissibility of ψ_{b,δ_D} follows from Theorem 5.1. Moreover, because of $\sqrt{b}y_b = \overline{w}(s/b)$, we obtain for $b > s$,

$$C(\psi_{b,\delta_D}, \delta_D) = \int \psi_{b,\delta_D}\psi_{b,\delta_D}\, d(P\otimes\delta_D) = \frac{1}{\overline{v}(s/b)} LI(\delta_D)^- L'$$

and for $b = s$,

$$C(\psi_{b,\delta_D}, \delta_D) = \int \psi_{b,\delta_D}\psi_{b,\delta_D}\, d(P\otimes\delta_D) = \frac{1}{\overline{v}(1)} LI(\delta_D)^- L'. \qquad \square$$

For $s = 1$, the admissible solution for a bias bound b is also universally optimal for the bias bound b. In particular, we get the following corollary of Theorem 5.1, which is useful in proving Lemma 5.3.

COROLLARY 7.2. *If*

$$\rho_b(z) = \begin{cases} \operatorname{sgn}(z), & \textit{for } b = 1,\\ \operatorname{sgn}(z)\min\{|z|, \sqrt{b}y\}, & \textit{for } b > 1,\end{cases}$$

where $y^2 = g(\sqrt{b}y) > 0$, *then we have*

$$\frac{\int \rho_b(z)^2 P(dz)}{\big(\int \rho_b(z) z P(dz)\big)^2} \le \frac{\int \rho(z)^2 P(dz)}{\big(\int \rho(z) z P(dz)\big)^2}$$

for all $\rho\colon \mathbb{R}\to\mathbb{R}$ *with* $\max_{z\in\mathbb{R}}(\rho(z)^2/\int \rho(y)^2 P(dy)) \le b$, $\int \rho(z)P(dz) = 0$ *and* $\int \rho(z)^2 P(dz) < \infty$.

PROOF. The assertion follows from Theorem 5.1 by setting $a(t) = 1$ for all $t \in \mathscr{T}$ and $L = 1$ and by using $B_b = 1/y$ for $b > 1$ and $B_b \neq 0$ arbitrarily for $b = 1$ [see also Corollary 9.8 in Müller (1995) or Corollary 8.2 in Müller (1997)]. □

LEMMA 7.3. $\overline{v}\colon (0,1] \to \mathbb{R}$ *is decreasing and concave.*

PROOF. Define

$$\tilde{w}(b) := \overline{w}\left(\frac{1}{b}\right) \quad \text{and} \quad u(b) := \overline{v}\left(\frac{1}{b}\right) \quad \text{for } b \geq 1,$$

$$l(y) := 2\Phi(y) - 1 - 2y\Phi'(y) \quad \text{and} \quad h(y) := y\Phi(-y) - \Phi'(y) \quad \text{for } y \in \mathbb{R},$$

and

$$w := \tilde{w}(b) \quad \text{for a fixed } b > 1.$$

Then $g(y) = l(y) + 2y^2\Phi(-y)$, $l(y) > 0$ and $h(y) < 0$ for $y > 0$. Using the implicit function theorem we get for fixed $b > 1$,

$$\tilde{w}'(b) = \frac{g(w)w}{2bl(w)} > 0.$$

This implies

$$u'(b) = \frac{-(2\Phi(w) - 1)2h(w)}{w} > 0 \tag{7.5}$$

and

$$u''(b) = \frac{\tilde{w}'(b)2}{w^2}\big[h(w)l(w) - \big(2\Phi(w) - 1\big)w\Phi(-w)\big] < 0.$$

In particular, we have

$$bu''(b) + 2u'(b) = \frac{1}{w\ l(w)}\big[h(w)l(w)g(w) - \big(2\Phi(w) - 1\big)f(w)\big] < 0 \tag{7.6}$$

because $f(y) := y\Phi(-y)g(y) + 4h(y)l(y) > 0$ for $y > 0$. See also Lemma 12.7 in Müller (1995) or Lemma A.11 in Müller (1997). Because of (7.5) and (7.6) we get

$$\overline{v}'(c) = u'\left(\frac{1}{c}\right)\frac{-1}{c^2} < 0$$

and

$$\overline{v}''(c) = \frac{1}{c^4}u''\left(\frac{1}{c}\right) + \frac{2}{c^3}u'\left(\frac{1}{c}\right) < 0. \qquad \square$$

PROOF OF LEMMA 5.3. Set

$$A_{\mathscr{D}} := (a(\tau_1), \dots, a(\tau_I))',$$

$$M(\rho) := \int a(t)a(t)'\rho(z,t)zP(dz)\,\delta(dt),$$

$$Q(\rho) := \int a(t)a(t)'\rho(z,t)^2P(dz)\,\delta(dt),$$

$$D := \operatorname{diag}\big(\delta(\{\tau_1\}), \dots, \delta(\{\tau_I\})\big),$$

$$D_1(\rho) := \operatorname{diag}\left(\int \rho(z,\tau_1) z P(dz), \dots, \int \rho(z,\tau_I) z P(dz)\right),$$

$$D_2(\rho) := \operatorname{diag}\left(\int \rho(z,\tau_1)^2 P(dz), \dots, \int \rho(z,\tau_I)^2 P(dz)\right),$$

and

$$D_\rho := \operatorname{diag}(\delta_\rho(\{\tau_1\}), \dots, \delta_\rho(\{\tau_I\})),$$

where $\operatorname{diag}(\lambda_1, \dots, \lambda_I)$ is the diagonal matrix with diagonal elements $\lambda_1, \dots, \lambda_I$. Because $a(\tau_1), \dots, a(\tau_I)$ are linearly independent, we have $A_{\mathscr{D}}(A'_{\mathscr{D}} B A_{\mathscr{D}})^- A'_{\mathscr{D}} = B^{-1}$ for all regular $B \in \mathbb{R}^{I \times I}$. Note that if $\psi \in \Psi^*(\delta, L)$ exists, then $L\beta$ is identifiable at δ so that $L = KA_{\mathscr{D}}$ for some $K \in \mathbb{R}^{s \times I}$ [see Lemma 1 in Kurotschka and Müller (1992)]. Then condition (2.2) in the definition of $\psi \in \Psi^*(\delta, L)$ provides $\psi(z,t) = LM(\rho)^- a(t)\rho(z,t)$ for all $(z,t) \in \mathbb{R} \times \mathscr{D}$. This implies

$$\begin{aligned} C(\psi,\delta) &= KD_1(\rho)^{-1} D_2(\rho) D_1(\rho)^{-1} D^{-1} K' \\ &= \frac{1}{\sum_{t \in \mathscr{D}} v_\rho(t)\delta(\{t\})} K D_\rho^{-1} K' = \frac{1}{\sum_{t \in \mathscr{D}} v_\rho(t)\delta(\{t\})} LI(\delta_\rho)^- L'. \end{aligned} \tag{7.7}$$

Moreover, for all $z \in \mathbb{R}$ and all $i = 1, \dots, I$, we have

$$\begin{aligned} b &\geq \psi(z,\tau_i)' C(\psi,\delta)^{-1} \psi(z,\tau_i) \\ &= \rho(z,\tau_i)^2 u_i' D_1(\rho)^{-1} D^{-1} K' \left(\sum_{t \in \mathscr{D}} v_\rho(t)\delta(\{t\})\right) \left[LI(\delta_\rho)^- L'\right]^{-1} KD_1(\rho)^{-1} D^{-1} u_i \\ &= \frac{\rho(z,\tau_i)^2}{\int \rho(y,\tau_i)^2 P(dy)} \frac{1}{\delta_\rho(\{\tau\})\delta(\{\tau_i\})} u_i' K' \left[LI(\delta_\rho)^- L'\right]^{-1} K u_i \\ &= \frac{\rho(z,\tau_i)^2}{\int \rho(y,\tau_i)^2 P(dy)} \frac{\delta_\rho(\{\tau\})}{\delta(\{\tau_i\})} a(\tau_i)' I(\delta_\rho)^- L' \left[LI(\delta_\rho)^- L'\right]^{-1} LI(\delta_\rho)^- a(\tau_i). \end{aligned}$$

Setting

$$b_i := \frac{b\delta(\{\tau_i\})}{\delta_\rho(\{\tau_i\}) a(\tau_i)' I(\delta_\rho)^- L' \left[LI(\delta_\rho)^- L'\right]^{-1} LI(\delta_\rho)^- a(\tau_i)}$$

we have for $i = 1, \dots, I$,

$$\max_{z \in \mathbb{R}} \frac{\rho(z,\tau_i)^2}{\int \rho(y,\tau_i)^2 P(dy)} \leq b_i$$

so that Theorem 4.1(i) applied to the one-dimensional case provides $b_i \geq 1$. Defining

$$\rho_0(z,\tau_i) := \begin{cases} \operatorname{sgn}(z), & \text{for } b_i = 1, \\ \operatorname{sgn}(z) \min\{|z|, \sqrt{b_i}\, y_i\}, & \text{for } b_i > 1, \end{cases}$$

where $y_i^2 = g(\sqrt{b_i}\, y_i) > 0$, we obtain according to Corollary 7.2,

$$v_\rho(\tau_i) \le v_{\rho_0}(\tau_i) = \overline{v}\left(\frac{1}{b_i}\right)$$

for $i = 1, \dots, I$. Note that $\sqrt{b_i}\, y_i = \overline{w}\,(1/b_i)$. Because according to Lemma 7.3 $\overline{v}$ is concave we therefore have

$$\begin{aligned}
&\sum_{i=1}^{I} v_\rho(\tau_i)\delta(\{\tau_i\}) \\
&\qquad \le \sum_{i=1}^{I} \overline{v}\left(\frac{1}{b_i}\right)\delta(\{\tau_i\}) \\
&\qquad \le \overline{v}\left(\sum_{i=1}^{I} \frac{1}{b_i}\delta(\{\tau_i\})\right) \\
&\qquad = \overline{v}\left(\frac{1}{b}\operatorname{tr}\left(\sum_{i=1}^{I} \left[LI(\delta_\rho)^- L'\right]^{-1} LI(\delta_\rho)^- a(\tau_i)a(\tau_i)' I(\delta_\rho)^- L'\delta_\rho(\{\tau_i\})\right)\right) \\
&\qquad = \overline{v}\left(\frac{1}{b}\operatorname{tr}(E_{s\times s})\right) = \overline{v}\left(\frac{s}{b}\right).
\end{aligned}$$

Then (7.7) implies the assertion. $\square$

PROOF OF THEOREM 5.4. Consider any $\delta \in \Delta$ with $\operatorname{supp}(\delta) = \{\tau_1, \dots, \tau_I\}$ and any $\psi \in \Psi^*(\delta, L)$ with $\|\psi' C(\psi, \delta)^{-1}\psi\|_\delta \le b$. If $a(\tau_1), \dots, a(\tau_I)$ are not linearly independent, then we can extend the regressors by $\tilde{a}(t)$ so that $\overline{a}(\tau_1), \dots, \overline{a}(\tau_I)$ are linearly independent, where $\overline{a}(t) = (a(t)', \tilde{a}(t)')'$. Then we have $\psi \in \overline{\Psi}^*(\delta, \overline{L})$ for some $\tilde{L}$, where $\overline{L} = (L, \tilde{L})$ and $\overline{\Psi}^*(\delta, \overline{L})$ is defined for the extended model given by $\overline{Y} = \overline{a}(t)'\overline{\beta} + Z$. Denoting $\overline{I}(\delta) = \int \overline{a}(t)\overline{a}(t)'\delta(dt)$, Lemma 5.3 provides

$$\begin{aligned}
\det\left(C(\psi, \delta)\right) &\ge \left(\frac{1}{\overline{v}(s/b)}\right)^s \det\left(\overline{L}\ \overline{I}(\delta_\rho)^- \overline{L}'\right) \\
&\ge \left(\frac{1}{\overline{v}(s/b)}\right)^s \det\left(LI(\delta_\rho)^- L'\right) \qquad (7.8)\\
&\ge \left(\frac{1}{\overline{v}(s/b)}\right)^s \det\left(LI(\delta_D)^- L'\right)
\end{aligned}$$

[see also Lemma A.15 in Müller (1987) for the property $\overline{L}\ \overline{I}(\delta_\rho)^- \overline{L}' \ge LI(\delta_\rho)^- L'$]. According to Lemma 5.2, the lower bound in (7.8) is attained by an ALE-test with influence function $\psi_{b,\,\delta_D}$ at δ_D. $\square$

Acknowledgments. The author thanks an Associate Editor and a referee for their valuable comments that improved the presentation of this paper.

REFERENCES

AKRITAS, M. G. (1991). Robust M estimation in the two-sample problem. *J. Amer. Statist. Assoc.* **86** 201–204.

BICKEL, P. J. (1981). Quelque aspects de la statistique robuste. *École d'Été de Probabilités de St. Flour. Lecture Notes in Math.* **876** 1–72. Springer, Berlin.

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

CHRISTENSEN, R. (1987). *Plane Answers to Complex Questions—The Theory of Linear Models.* Springer, New York.

COAKLEY, C. W. and HETTMANSPERGER, T. P. (1992). Breakdown bounds and expected resistance. *J. Nonparametr. Statist.* **1** 267–276.

COAKLEY, C. W. and HETTMANSPERGER, T. P. (1994). The maximum resistance of tests. *Austral. J. Statist.* **36** 225–233.

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

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

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.

HERITIER, S. and RONCHETTI, E. (1994). Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc.* **89** 897–904.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879.

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **12** 363–366.

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

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

KUROTSCHKA, V. and MÜLLER, C. H. (1992). Optimum robust estimation of linear aspects in conditionally contaminated linear models. *Ann. Statist.* **20** 331–350.

MARKATOU, M. and HE, X. (1994). Bounded influence and high breakdown point testing procedures in linear models. *J. Amer. Statist. Assoc.* **89** 543–549.

MARKATOU, M. and HETTMANSPERGER, T. P. (1990). Robust bounded-influence tests in linear models. *J. Amer. Statist. Assoc.* **85** 187–190.

MARKATOU, M. and MANOS, G. (1996). Robust tests in nonlinear regression models. *J. Statist. Plann. Inference* **55** 205–217.

MARKATOU, M., STAHEL, W. A. and RONCHETTI, E. (1991). Robust M-type testing procedures for linear models. In *Directions in Robust Statistics and Diagnostics I* (W. Stahel and S. Weisberg, eds.) 201–220. Springer, New York.

MÜLLER, C. H. (1987). Optimale Versuchspläne für robuste Schätzfunktionen in linearen Modellen. Ph.D. thesis. Freie Univ. Berlin.

MÜLLER, C. H. (1992a). L_1-estimation and testing in conditionally contaminated linear models. In *L_1-Statistical Analysis and Related Methods* (Y. Dodge, ed.) 69–76. North-Holland, Amsterdam.

MÜLLER, C. H. (1992b). Robust estimation with minimum bias and A-optimal designs. In *PROBASTAT'91, Proceedings of the International Conference on Probability and Mathematical Statistics, Bratislava 1991* (A. Pázman and J. Volaufová, eds.) 109–115.

MÜLLER, C. H. (1992c). Robust testing in conditionally contaminated linear models. Preprint No. A-92-12, Fachbereich Mathematik und Informatik, Freie Univ. Berlin.

MÜLLER, C. H. (1994a). Optimal designs for robust estimation in conditionally contaminated linear models. *J. Statist. Plann. Inference* **38** 125–140.

MÜLLER, C. H. (1994b). One-step-M-estimators in conditionally contaminated linear models. *Statist. Decisions* **12** 331–342.

MÜLLER, C. H. (1995). Outlier robust inference for planned experiments. Habilitation thesis. Freie Univ. Berlin.

MÜLLER, C. H. (1997). *Robust Planning and Analysis of Experiments. Lecture Notes in Statist.* **124**. Springer, New York.

PÁZMAN, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

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.

RIEDER, H. (1994). *Robust Asymptotic Statistics*. Springer, New York.

RONCHETTI, E. (1982). Robust testing in linear models: the infinitesimal approach. Ph.D. thesis, Swiss Federal Institute of Technology, Zürich.

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

SILVAPULLE, M. J. (1992a). Robust Wald-type tests of one-sided hypothesis in the linear model. *J. Amer. Statist. Assoc.* **87** 156–161.

SILVAPULLE, M. J. (1992b). Robust tests of inequality constraints and one-sided hypotheses in the linear model. *Biometrika* **79** 621–630.

STAUDTE, R. G. and SHEATHER, S. J. (1990). *Robust Estimation and Testing*. Wiley, New York.

WALD, A. (1943). Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **3** 426–482.

INSTITUTE FÜR MATHEMATISCHE STOCHASTIK
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
LOTZESTR. 13
37083 GÖTTINGEN
GERMANY
E-MAIL: chmuelle@math.uni-goettingen.de