Neighborhoods as Nuisance Parameters? Robustness vs. Semiparametrics.

Helmut Rieder
University of Bayreuth, Germany

25 March 2000

Abstract

Deviations from the center within a robust neighborhood may naturally be considered an infinite dimensional nuisance parameter. Thus, in principle, the semiparametric method may be tried, which is to compute the scores function for the main parameter minus its orthogonal projection on the closed linear tangent space for the nuisance parameter, and then rescale for Fisher consistency. We derive such a semiparametric influence curve by nonlinear projection on the tangent balls arising in robust statistics.

This semiparametric IC is compared with the robust IC that minimizes maximum weighted mean square error of asymptotically linear estimators over infinitesimal neighborhoods. For Hellinger balls, the two coincide (with the classical one). In the total variation model, the semiparametric IC solves the robust MSE problem for a particular bias weight. In the case of contamination neighborhoods, the semiparametric IC is bounded only from above. Due to an interchange of truncation and linear combination, the discrepancy increases with the dimension. Thus, despite of striking similarities, the semiparametric method falls short, respectively fails, to solve the minimax MSE estimation problem for the gross error models.

Moreover, for testing hypotheses which are defined by two closed and convex sets of tangents, we furnish a saddle point via projection on these sets. In the cases of total variation and contamination neighborhoods, the robust asymptotic tests based on least favorable pairs are recovered. Therefore, the two approaches agree in the testing context.

Key Words and Phrases: Infinitesimal neighborhoods; Hellinger, total variation, contamination; semiparametric models; tangent spaces, cones, and balls; projection; influence curves; Fisher consistency; canonical influence curve; Hampel–Krasker influence curve; differentiable functionals; asymptotically linear estimators; Cramér–Rao bound; maximum mean square error; asymptotic minmax and convolution theorems; $C(\alpha)$- and Wald tests; least favorable pairs; robust asymptotic tests; saddle points.
AMS/MSC-2000 classification: 62F35, 62G35.

1 Introduction: The Semiparametric Setup

We need to set up the standard semiparametric framework, which employs some family $\mathcal{Q}$ in the set $\mathcal{M}$ of all probabilities on some sample space $(\Omega, \mathcal{B})$,

$$\mathcal{Q} = \{\, Q_{\theta,\nu} \mid \theta \in \Theta\,,\ \nu \in H_\theta \,\} \subset \mathcal{M} \tag{1.1}$$

The parameter θ of interest is finite (k-)dimensional, out of some open parameter set $\Theta \subset \mathbb{R}^k$, whereas ν acts as nuisance parameter. For each θ, ν ranges over some set H_θ; typically, subsets of some infinite dimensional function spaces. The observations $x_1, \dots, x_n \sim Q_{\theta,\nu}$ are assumed to be independent identically distributed. Estimators of θ may be any functions $S_n\colon \Omega^n \to \mathbb{R}^k$ which are product measurable $\mathcal{B}^n$, Borel $\mathbb{B}^k$. Let us fix (θ_0, ν_0), the true but unknown values of main and nuisance parameter.

In this generality, optimality results for the estimation of θ_0 can only be derived in an approximate way, that is, asymptotically as the sample size n tends to infinity. Moreover, to obtain meaningful results at all, estimators, which now are estimator sequences $S = (S_n)$, have to be judged locally about (θ_0, ν_0). Subsequently, this fixed parameter will be omitted whenever feasible. Thus, we put $Q_{\theta_0,\nu_0} = Q$. Expectation and covariance under Q are denoted by E and C, respectively. Also the spaces L_2 and L_∞ of square integrable and essentially bounded real functions, respectively, refer to the fixed $Q = Q_{\theta_0,\nu_0}$. The corresponding spaces of $\mathbb{R}^k$ valued functions are denoted by L_2^k and L_∞^k, respectively.

For the local asymptotics a certain smoothness of the parametric model is required, in the sense of mean square differentiability at (θ_0, ν_0) of square root densities: There exists some function $\Lambda \in L_2^k$—the scores function for the main parameter θ at (θ_0, ν_0)—such that for each $a \in \mathbb{R}^k$, and for each function $g \in \partial_2 \mathcal{Q}$ there exists some path $t \mapsto \nu_t^g \in H_{\theta_0 + ta}$, such that, as $t \to 0$ in $\mathbb{R}$,

$$\sqrt{dQ_{\theta_0+ta,\nu_t^g}} = \bigl(1 + \tfrac{1}{2}\, t\, (a'\Lambda + g)\bigr) \sqrt{dQ_{\theta_0,\nu_0}} + \mathrm{o}(t) \tag{1.2}$$

In this context, the tangent set $\partial\mathcal{Q} = \partial_1\mathcal{Q} + \partial_2\mathcal{Q}$ of the model $\mathcal{Q}$ at (θ_0, ν_0) enters, where $\partial_1\mathcal{Q} = \{\, a'\Lambda \mid a \in \mathbb{R}^k \,\}$ is the tangent space (linear, closed) for the first parameter component, and $\partial_2\mathcal{Q} \subset L_2$ denotes the tangent set for the nuisance component; all tangents in either class $\partial_*\mathcal{Q}$ necessarily of expectation zero. The covariance $\mathcal{I} = \mathrm{C}\,\Lambda$, which is the Fisher information of the ν_0-section $\mathcal{Q}_{\nu_0}$ of model $\mathcal{Q}$ for the parameter θ at θ_0, confer (1.10) and (1.11) below, is assumed of full rank k.

As for complete technical details, maybe in slightly different notations, the reader may consult standard textbooks on asymptotic statistics such as Bickel et al. (1993; chapters 2–3), Rieder[1] (1994; chapters 2–4), and van der Vaart (1998;

[1] HR, subsequently

chapter 25). This recommendation also holds for the following notions, to be summarized in this section.

Influence functions ψ or, in robust terminology, influence curves ψ for model $\mathcal{Q}$ at (θ_0, ν_0) are defined by the conditions

$$\psi \in L_2^k, \quad \mathrm{E}\,\psi = 0\,, \quad \mathrm{E}\,\psi\Lambda' = \mathbb{I}_k\,, \quad \mathrm{E}\,\psi g = 0 \;\; \forall g \in \partial_2\mathcal{Q} \tag{1.3}$$

where $\mathbb{I}_k$ denotes the $k \times k$ identity matrix. The set of all influence curves for model $\mathcal{Q}$ at (θ_0, ν_0) is denoted by $\Psi = \Psi_{\theta_0,\nu_0}$.

On the one hand, influence curves go with functionals $T\colon \mathcal{Q} \to \mathbb{R}^k$ which are differentiable, with respect to model $\mathcal{Q}$ at (θ_0, ν_0) in accordance with (1.2), and Fisher consistent for the main parameter, such that

$$T(Q_{\theta_0+ta,\nu_t}) = T(Q_{\theta_0,\nu_0}) + \mathrm{E}\,\psi(a'\Lambda + g)\,t + \mathrm{o}(t) = \theta_0 + ta + \mathrm{o}(t) \tag{1.4}$$

On the other hand, influence curves go with asymptotically linear estimators. These are estimators $S = (S_n)$ that have an expansion

$$\sqrt{n}\,(S_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(x_i) \; + \mathrm{o}_{Q^n}(n^0) \tag{1.5}$$

where the remainder tends to zero in probability, under the sequence of product measures Q^n. Such estimators are asymptotically normal in accordance with (1.2): Setting $Q_n(a,g) = Q_{\theta_0+s_n a,\,\nu^g_{s_n}}$ for $s_n = 1/\sqrt{n}$, their distributions under $Q_n^n(a,g)$ converge weakly as $n \to \infty$, for every $a \in \mathbb{R}^k$ and $g \in \partial_2\mathcal{Q}$,

$$\sqrt{n}\,(S_n - \theta_0)(Q_n^n(a,g)) \xrightarrow{\;\mathrm{w}\;} \mathcal{N}(a, \mathrm{C}\,\psi) \tag{1.6}$$

Given any $\psi \in \Psi$, at least locally valid constructions to achieve (1.4) and (1.5) are $T(M) = \theta_0 + 2\int \psi\,\sqrt{dQ}\,\sqrt{dM}$ and $S_n = \theta_0 + 1/n \sum \psi(x_i)$.

For either tangent set $\partial_*\mathcal{Q}$ let $\mathrm{lin}\,\partial_*\mathcal{Q}$ and $c\ell\,\mathrm{lin}\,\partial_*\mathcal{Q}$ denote the linear span, respectively the closed linear span, of $\partial_*\mathcal{Q}$ in L_2. Thus, $c\ell\,\mathrm{lin}\,\partial_1\mathcal{Q} = \partial_1\mathcal{Q}$, and $c\ell\,\mathrm{lin}\,\partial\mathcal{Q} = \partial_1\mathcal{Q} + c\ell\,\mathrm{lin}\,\partial_2\mathcal{Q}$ as $\dim\partial_1\mathcal{Q}$ is finite. Introduce the orthogonal projection $\pi_*\colon L_2 \to c\ell\,\mathrm{lin}\,\partial_*\mathcal{Q}$ on $c\ell\,\mathrm{lin}\,\partial_*\mathcal{Q}$, and $\Pi_*\colon L_2^k \to (c\ell\,\mathrm{lin}\,\partial_*\mathcal{Q})^k$ the orthogonal projection in the product space; then $\Pi_* = (\pi_*, \dots, \pi_*)'$, acting coordinatewise.

In view of (1.3), the projection $\Pi(\psi)$ on $(c\ell\,\mathrm{lin}\,\partial\mathcal{Q})^k$ must be the same for every $\psi \in \Psi$—the shortest, or canonical, influence curve ϱ. In fact,

$$\Pi(\psi) = \varrho = \mathcal{J}^{-1}\big(\Lambda - \Pi_2(\Lambda)\big) \qquad \forall\,\psi \in \Psi \tag{1.7}$$

where

$$\mathcal{J} = \mathrm{C}\big(\Lambda - \Pi_2(\Lambda)\big) \tag{1.8}$$

denotes the Fisher information of model $\mathcal{Q}$ for the parameter θ at (θ_0, ν_0).

By a little argument, the existence of influence curves may be seen to be equivalent to regularity, that is, positive definiteness, of $\mathcal{J}$,

$$\begin{aligned} \Psi \neq \emptyset &\iff \mathcal{J} > 0 \\ &\iff a'\Lambda \notin c\ell \operatorname{lin} \partial_2 \mathcal{Q} \quad \forall\, a \in \mathbb{R}^k ,\ a \neq 0 \end{aligned} \tag{1.9}$$

which condition we want to assume fulfilled subsequently.

Remark 1.1 [adaptivity] With the nuisance parameter ν fixed to ν_0, the ν_0-section $\mathcal{Q}_{\nu_0}$ of model $\mathcal{Q}$ is a model without nuisance parameter,

$$\mathcal{Q}_{\nu_0} = \{\, Q_{\theta,\nu_0} \mid \theta \in \Theta \,\} \tag{1.10}$$

satisfying (1.2) with $\partial_2 \mathcal{Q}_{\nu_0} = \{0\}$ and $\partial \mathcal{Q}_{\nu_0} = \partial_1 \mathcal{Q}$. Consequentially, the canonical influence curve $\hat{\varrho}$ and the Fisher information of model $\mathcal{Q}_{\nu_0}$ for the parameter θ at θ_0 are given by, respectively,

$$\hat{\varrho} = \mathcal{I}^{-1}\Lambda , \qquad \mathcal{I} = \mathrm{C}\,\Lambda \tag{1.11}$$

The following bound of $\mathcal{J}$ by $\mathcal{I}$, in the positive definite sense, is an immediate consequence of (1.7) and (1.11),

$$\mathrm{C}\,\hat{\varrho} = \mathcal{I}^{-1} \leq \mathcal{J}^{-1} = \mathrm{C}\,\varrho \tag{1.12}$$

where the lower bound is attained iff $\varrho = \hat{\varrho}$, which in turn holds iff $\Pi_2(\Lambda) = 0$. This is the case of adaptivity. The construction of adaptive estimators is a major subject of semiparametric theory; confer Bickel (1982; Sections 3 and 4), Klaassen (1987), Schick (1986), and the references mentioned therein. ////

Remark 1.2 [bounded influence curves] Existence of bounded ICs $\psi \in \Psi$, which may become relevant for robustness in semiparametric models, proves equivalent to the following condition

$$a'\Lambda \notin c\ell' \operatorname{lin} \partial_2 \mathcal{Q} \qquad \forall\, a \in \mathbb{R}^k,\ a \neq 0 \tag{1.13}$$

where $c\ell' \operatorname{lin}$ denotes the closed linear span in L_1. This follows from Theorem 1 of Shen (1995) on observing that his condition (S'), with $c\ell' \operatorname{lin}(\partial_2 \mathcal{Q} + \text{constants})$ in the place of $c\ell' \operatorname{lin} \partial_2 \mathcal{Q}$, because $\mathrm{E}\,\Lambda = 0$ and $\mathrm{E}\,g = 0\ \forall g \in \partial_2 \mathcal{Q}$, in fact simplifies to (1.13).

Naturally, as any bounded influence curve is an influence curve, respectively as the L_1-closure in (1.13) is larger than the L_2-closure in (1.9), condition (1.13) is generally stronger than (1.9). When $\operatorname{lin} \partial_2 \mathcal{Q}$ has finite dimension, however, it is closed in both L_1 and L_2, and consequentially, the existence of influence curves automatically implies the existence of bounded ones. ////

Remark 1.3 [finite dimensions] In case $H_\theta \subset \mathbb{R}^m$ for some finite dimension m, suppose the square root densities of model $\mathcal{Q}$ are L_2-differentiable at (θ_0, ν_0) with respect to the full parameter (θ, ν), such that (1.2) is satisfied with paths $\nu_t = \nu_0 + tb$ and with $g = b'\Delta$, where $\Delta \in L_2^m$, $\mathrm{E}\,\Delta = 0$, denotes the scores function for the nuisance parameter ν at (θ_0, ν_0).

Then $\partial_2 \mathcal{Q} = \{\, b'\Delta \mid b \in \mathbb{R}^m \,\}$, and the Fisher information $\mathcal{H}$ of model $\mathcal{Q}$ for the full parameter (θ, ν) at (θ_0, ν_0) is

$$\mathcal{H} = \mathrm{C}\begin{pmatrix} \Lambda \\ \Delta \end{pmatrix} = \begin{pmatrix} \mathcal{I} & \mathcal{C} \\ \mathcal{C}' & \mathcal{D} \end{pmatrix} \qquad \text{where } \mathcal{C} = \mathrm{E}\,\Lambda\,\Delta' \tag{1.14}$$

Like $\mathcal{I} = \mathrm{C}\,\Lambda > 0$, the covariance $\mathcal{D} = \mathrm{C}\,\Delta$, which is the Fisher information of the θ_0-section $\mathcal{Q}_{\theta_0}$ of model $\mathcal{Q}$ for the parameter ν at ν_0, may be assumed of full rank m.

Then $\Pi_2(\Lambda) = \mathcal{C}\,\mathcal{D}^{-1}\Delta$ and $\mathcal{J} = \mathrm{C}\big(\Lambda - \Pi_2(\Lambda)\big) = \mathcal{I} - \mathcal{C}\,\mathcal{D}^{-1}\mathcal{C}'$ (familiar from $C(\alpha)$-tests). Moreover, $\mathcal{J} > 0$ iff $\mathcal{H} > 0$, since $\det \mathcal{H} = \det \mathcal{D} \det \mathcal{J}$. Condition (1.13), too, because $\mathcal{D} > 0$, is obviously equivalent to $\operatorname{rk} \mathcal{H} = k + m$.

In this case,

$$\varrho = \mathcal{J}^{-1}\big(\Lambda - \mathcal{C}\,\mathcal{D}^{-1}\Delta\big) = D\,\mathcal{H}^{-1}\begin{pmatrix} \Lambda \\ \Delta \end{pmatrix} \tag{1.15}$$

where

$$\mathcal{J} = \mathcal{I} - \mathcal{C}\,\mathcal{D}^{-1}\mathcal{C}' \quad \text{ and } \quad D = (\mathbb{I}_k, 0_{k\times m}) \tag{1.16}$$

defines the shortest influence curve, in fact, the first component of the shortest influence curve (usually ascribed to the MLE) for the full parameter.

Starting from this function $\mathcal{H}^{-1}\binom{\Lambda}{\Delta}$, bounded influence curves have been constructed explicitly by HR (1994), Remark 4.2.11 and 5.5(8), 5.5(9), if the matrix D there is specialized to the projection matrix $(\mathbb{I}_k, 0_{k\times m})$. ////

Closely related to the orthogonal projection (1.7) of influence curves leading to the canonical IC ϱ is the Cramér–Rao bound for the covariance,

$$\mathrm{C}\,\psi \geq \mathcal{J}^{-1} = \mathrm{C}\,\varrho \qquad \forall\,\psi \in \Psi \tag{1.17}$$

in the positive definite sense, with equality iff $\psi = \varrho$. In view of (1.6), this bound concerns the asymptotic covariance of asymptotically linear estimators. Thus, the asymptotically linear estimator with canonical influence curve ϱ at (θ_0, ν_0) is the asymptotically most accurate to estimate θ_0, in model $\mathcal{Q}$.

That this optimality is not restricted to estimators which are asymptotically linear, but need to fulfill only a regularity condition weaker than asymptotic linearity, or may even be arbitrary measurable, is the subject of the convolution and asymptotic minimax theorems, respectively; confer, for example, Bickel et al. (1993; Theorem 3.3.2), HR (1994; Theorems 4.3.2, 4.3.4), van der Vaart (1998; Theorems 25.20, 25.21, Lemma 25.25).

Remark 1.4 [nonlinear projection] These optimality theorems require some structure of the tangent set $\partial\mathcal{Q}$, to be a linear space or at least a convex cone. In spite of the special structure, the projection in terms of which the bounds are stated, is generally that on the closed linear span $c\ell \operatorname{lin} \partial\mathcal{Q}$.

One exception is the concentration bound for asymptotically median unbiased estimators by Pfanzagl and Wefelmeyer (1982; Theorem 9.2.2), in terms of the projection on a closed convex cone $\partial\mathcal{Q}$. In HR (2000) we however show that the bound may not possibly be attained, and derive a suitable one-sided bound that is still based on the projection on $c\ell\, \partial\mathcal{Q}$ (as opposed to $c\ell \operatorname{lin} \partial\mathcal{Q}$). ////

2 The Infinitesimal Robust Setup

In robust statistics, we start with an ideal model $\mathcal{P} = \{ P_\theta \mid \theta \in \Theta \}$ which is smoothly parametrized by some finite (k-)dimensional parameter θ out of an open subset $\Theta \subset \mathbb{R}^k$; that is, $\mathcal{P}$ is some model as assumed in Section 1 but deprived of its nuisance parameter. Since we do not believe in such a model $\mathcal{P}$ strictly, we enlarge its elements P_θ to certain neighborhoods $U(\theta; r) \subset \mathcal{M}$ of radius r. Then the i.i.d. observations, under the hypothesis θ, may be allowed to follow any law $Q \in U(\theta; r)$, while still θ has to be estimated. Thus, the neighborhood model $\mathcal{Q}$ is obtained,

$$\mathcal{Q} = \big\{ Q \;\big|\; \theta \in \Theta ,\, Q \in U(\theta; r) \big\} \tag{2.1}$$

Model $\mathcal{Q}$ is clearly semiparametric: The deviation $Q - P_\theta$ of $Q \in U(\theta; r)$ from the ideal P_θ has entered as nuisance parameter ν, ranging over the sets of differences $H_\theta = \{ Q - P_\theta \mid Q \in U(\theta; r) \}$, where $Q = Q_{\theta,\nu}$ with $\nu \in H_\theta$. In particular, the ideal model $\mathcal{P}$ is the ν_0 section of model $\mathcal{Q}$ for $\nu_0 = 0$.

Remark 2.1 [nonidentifiability] If one does not start with a true θ, but the real law Q, and seeks θ depending on Q, one runs into the identifiability problem: The equation $Q = Q_{\theta, Q - P_\theta} = P_\theta + Q - P_\theta$ has multiple solutions θ. This is the case already for members $Q = P_\zeta$ of the ideal model with ζ close to θ such that $P_\zeta \in U(\theta; r)$ (usually, the parametrization is continuos relative to the neighborhoods).

This problem has been dealt with by means of functionals that are Fisher consistent at the ideal model and extend the parametrization to the neighborhoods. Actually, both approaches lead to the same optimally robust influence curves and procedures—once the choice of functional is subjected to robustness criteria; confer HR (1994; preface, subsection 4.3.3). ////

We specify the neighborhoods $U(\theta; r)$ to be balls around P_θ of radius r in Hellinger or total variation distance, or contamination neighborhoods,

$$U_*(\theta; r) = \big\{ Q \in \mathcal{M} \;\big|\; d_*(Q, P_\theta) \le r \big\} \tag{2.2}$$

$$U_c(\theta; r) = \big\{ Q = (1 - r)_+\, P_\theta + (1 \wedge r)\, M \;\big|\; M \in \mathcal{M} \big\} \tag{2.3}$$

where the Hellinger and total variation metrics d_h and d_v are given by

$$2\, d_h^2(Q,P) = \int \big|\sqrt{dQ} - \sqrt{dP}\,\big|^2 , \qquad 2\, d_v(Q,P) = \int |dQ - dP| \tag{2.4}$$

Let us fix $\theta_0 \in \Theta$ and $\nu_0 = 0$, and write P for the previous $Q = Q_{\theta_0,\nu_0} = P_{\theta_0}$. In the sequel, the scores function Λ is that of the ideal model $\mathcal{P}$, for θ at θ_0.

Towards the differentiability (1.2) of the neighborhood model $\mathcal{Q}_*$ at $(\theta_0, 0)$, depending on the type of neighborhoods $U_*(\theta_0; r)$, we introduce the following balls $\mathcal{G}_* = \mathcal{G}_*(\theta_0; r)$ as candidate tangent sets $\partial_2 \mathcal{Q}_*$,

$$\mathcal{G}_h = \big\{\, g \in L_2 \;\big|\; \mathrm{E}\, g = 0,\ \mathrm{E}\, g^2 \le 8 r^2 \,\big\} \tag{2.5}$$

$$\mathcal{G}_v = \big\{\, g \in L_2 \;\big|\; \mathrm{E}\, g = 0,\ \mathrm{E}\, |g| \le 2r \,\big\} \tag{2.6}$$

$$\mathcal{G}_c = \big\{\, g \in L_2 \;\big|\; \mathrm{E}\, g = 0,\ g \ge -r \,\big\} \tag{2.7}$$

where $\mathcal{G}_h \subset \sqrt{2}\, \mathcal{G}_v$ as $d_v \le \sqrt{2}\, d_h$, and $\mathcal{G}_c \subset \mathcal{G}_v = \mathcal{G}_c - \mathcal{G}_c$ by (7.36) below.

Proposition 2.2 *The tangent sets at* $(\theta_0, 0)$ *of the neighborhood model* $\mathcal{Q}_*$, *for* $* = h, v, c$, *are*

$$\partial_1 \mathcal{Q}_* = \{\, a'\Lambda \mid a \in \mathbb{R}^k \,\}, \quad \partial_2 \mathcal{Q}_* = \mathcal{G}_*, \quad \partial \mathcal{Q}_* = \partial_1 \mathcal{Q}_* + \partial_2 \mathcal{Q}_* \tag{2.8}$$

PROOF Invoke bounded approximations $\Lambda^{(t)}$ of Λ such that $\mathrm{E}\,\Lambda^{(t)} = 0$ and, as $t \to 0$, $\sup |\Lambda^{(t)}| = \mathrm{o}(t^{-1})$ and $\mathrm{E}\, |\Lambda^{(t)} - \Lambda|^2 \to 0$. Given $a \in \mathbb{R}^k$ and any bounded $g \in \mathcal{G}_*$, employ the path $\nu_t^g = tg$ in defining measures $Q_t = Q_{\theta_0 + ta, tg}$ by

$$dQ_t = \big(1 + t(a'\Lambda^{(t)} + g)\big)\, dP \tag{2.9}$$

Then mean square differentiability (1.2) is satisfied, and these probabilities belong to the neighborhoods $U_*(\theta_0 + ta; tr)$ in the following, entirely acceptable sense,

$$d_*(Q_t, P_{\theta_0 + ta}) \le tr + \mathrm{o}(t) \tag{2.10}$$

in the cases $* = h, v$. In the case $* = c$, there exist approximations $\widetilde{P}_{\theta_0 + ta}$ of $P_{\theta_0 + ta}$, namely, $\widetilde{P}_{\theta_0 + ta}$ with P density $1 + t_r a'\Lambda^{(t)}$, $t_r = t/(1 - tr)$, such that

$$d_v(\widetilde{P}_{\theta_0 + ta}, P_{\theta_0 + ta}) = \mathrm{o}(t) \qquad \text{and} \qquad Q_t \in \widetilde{U}_c(\theta_0 + ta; tr) \tag{2.11}$$

for the tr contamination balls $\widetilde{U}_c(\theta_0 + ta; tr)$ about $\widetilde{P}_{\theta_0 + ta}$.

In either case, we pass to the closure of $\mathcal{G}_* \cap L_\infty$ in L_2, which is $\mathcal{G}_*$. The technical details needed in this proof may be found in HR (1994): Remark 4.2.3, Lemma 4.2.4, Lemma 5.3.1, and proof to Theorem 5.4.1 (a). ////

Although not yet explicitly as tangent sets, the balls $\mathcal{G}_*$ appear in Bickel (1981) and in other work on infinitesimal neighborhoods by that time.

The tangent sets $\mathcal{G}_*$ are closed convex, and the smallest cone and linear space containing either $\mathcal{G}_*$ is already the full tangent space $L_2 \cap \{\mathrm{E} = 0\}$, provided only that $r > 0$. Consequentially, $\Lambda - \Pi_2(\Lambda) = 0$ and $\mathcal{J} = 0$ in (1.7); in particular, adaptivity fails drastically. The canonical IC ϱ is undefined.

3 The Semiparametric Influence Curve

In the robust setup, we therefore modify definition (1.7) of canonical influence curve, replacing π_2 by the nonlinear projection $\widetilde{\pi}_2 \colon L_2 \to \partial_2 \mathcal{Q}_*$ on $\partial_2 \mathcal{Q}_* = \mathcal{G}_*$ itself. Correspondingly, Π_2 is replaced by $\widetilde{\Pi}_2 = (\widetilde{\pi}_2, \ldots, \widetilde{\pi}_2)' \colon L_2^k \to (\partial_2 \mathcal{Q}_*)^k$, defined coordinatewise. Thus, we obtain the following function $\tilde{\varrho}_*$, which we call the semiparametric influence curve,

$$\tilde{\varrho} = \mathcal{K}^{-1}\bigl(\Lambda - \widetilde{\Pi}_2(\Lambda)\bigr) \tag{3.1}$$

with scaling matrix

$$\mathcal{K} = \mathrm{E}\bigl(\Lambda - \widetilde{\Pi}_2(\Lambda)\bigr)\Lambda' \tag{3.2}$$

The definition of $\tilde{\varrho}$ requires that $\det \mathcal{K} \neq 0$. Rescaling of $\Lambda - \widetilde{\Pi}_2(\Lambda)$ by $\mathcal{K}$ ensures that $\mathrm{E}\,\tilde{\varrho}\Lambda' = \mathbb{I}_k$ (Fisher consistency). Note that $\mathcal{K} \neq \mathrm{C}\bigl(\Lambda - \widetilde{\Pi}_2(\Lambda)\bigr)$, since residuals are no longer orthogonal to the approximating ball.

Remark 3.1 The modified projection recipe (3.1)–(3.2) seems intuitively plausible but is based only on analogy. The semiparametric influence curve has not been derived as—but may be checked against—a mathematical solution to some suitable extension of the Cramér–Rao bound, or convolution and asymptotic minimax theorems, in the semiparametric, respectively robust, setup with full tangent balls. ////

The following approximation lemma is well-known and will be applied to the balls $G = \mathcal{G}_*$, the space $X = L_2$, and the coordinates x of Λ; then $\tilde{g} = \widetilde{\pi}_2(\Lambda_j)$.

Lemma 3.2 *Let G be a nonempty closed and convex subset of some Hilbert space X, and $x \in X$. Then the minimum norm problem*

$$|x - g|^2 = \min ! \qquad g \in G \tag{3.3}$$

has a unique solution $\tilde{g} \in G$, which is characterized by

$$\langle x - \tilde{g} | g - \tilde{g} \rangle \leq 0 \qquad \forall\, g \in G \tag{3.4}$$

In the sequel, $\mathcal{I} = \mathrm{C}\,\Lambda = (\mathcal{I}_{i,j})$ and $\hat{\varrho} = \mathcal{I}^{-1}\Lambda$ denote Fisher information (of full rank k) and the canonical influence curve, of the ideal model $\mathcal{P}$ at θ_0.

We now determine the semiparametric influence curves $\tilde{\varrho}_h$, $\tilde{\varrho}_v$, $\tilde{\varrho}_c$ for the Hellinger, total variation, and contamination neighborhood models, respectively.

Theorem 3.3 [Hellinger model] *The semiparametric IC $\tilde{\varrho}_h$ exists iff*

$$8\,r^2 < \min\nolimits_{j=1,\ldots,k} \mathcal{I}_{j,j} \tag{3.5}$$

And then

$$\tilde{\varrho}_h = \hat{\varrho} = \mathcal{I}^{-1}\Lambda \tag{3.6}$$

PROOF In the case $k = 1$ we have $\widetilde{\pi}_2 = \gamma\Lambda$ with $\gamma =$ positive root of the minimum of 1 and $8\,r^2/\mathcal{I}$. Indeed, by Cauchy–Schwarz, for every $g \in \mathcal{G}_h$,

$$\langle \Lambda - \gamma\Lambda | g\rangle = (1-\gamma)\langle\Lambda|g\rangle \le (1-\gamma)\sqrt{8}\, r\,\mathcal{I}^{1/2} = (1-\gamma)\gamma\mathcal{I} = \langle\Lambda - \gamma\Lambda|\gamma\Lambda\rangle \quad (3.7)$$

For general $k \ge 1$, this implies that $\Lambda - \widetilde{\Pi}_2(\Lambda) = D\Lambda$ and $\mathcal{K} = D\mathcal{I}$ with matrix $D = \operatorname{diag}(1-\gamma_j)$, where $0 \le \gamma_j \le 1$, and $\gamma_j = 1$ iff $\mathcal{I}_{j,j} \le 8r^2$. ////

Theorem 3.4 [total variation] *The semiparametric IC $\tilde{\varrho}_v$ exists only if*

$$2\,r < \min_{j=1,\dots,k} \operatorname{E}|\Lambda_j| \quad (3.8)$$

And then $\Lambda^{(v)} = \Lambda - \widetilde{\Pi}_2(\Lambda)$ has coordinates

$$\Lambda_j^{(v)} = v_j' \vee \Lambda_j \wedge v_j'' \quad (3.9)$$

where the clipping constants $v_j' < 0 < v_j''$ are uniquely determined by

$$\operatorname{E}(v_j' - \Lambda_j)_+ = r = \operatorname{E}(\Lambda_j - v_j'')_+ \quad (3.10)$$

PROOF Obviously, $\Lambda_j - \widetilde{\pi}_2(\Lambda_j) = 0$ iff $\operatorname{E}|\Lambda_j| \le 2r$. Thus assume (3.8).

In case $k = 1$, in order to minimize $\operatorname{E}(\Lambda - g)^2$ for $g \in \mathcal{G}_v$, we set up a Lagrangian $\operatorname{E}\big((\Lambda - g)^2 + 2\alpha g + 2\beta|g|\big)$ with some unspecified real multipliers, and try to minimize the integrand $I(g) = (\Lambda - g)^2 + 2\alpha g + 2\beta|g|$ at each point.

A minimizing value $\tilde{g} = 0$ means that $\Lambda^2 \le (\Lambda - g)^2 + 2\alpha g + 2\beta g$ for all numbers $g > 0$; that is, $\Lambda - \alpha \le \beta$, and $\Lambda^2 \le (\Lambda - g)^2 + 2\alpha g - 2\beta g$ for all numbers $g < 0$; that is, $\Lambda - \alpha \ge -\beta$. This is the case when $\Lambda - \tilde{g} = \Lambda$.

If $\tilde{g} > 0$, then the derivative $dI(\tilde{g}) = 0$ gives $\Lambda - \tilde{g} = \alpha + \beta$. If $\tilde{g} < 0$, $dI(\tilde{g}) = 0$ gives $\Lambda - \tilde{g} = \alpha - \beta$. These are the cases when $\Lambda - \alpha > \beta$, respectively when $\Lambda - \alpha < -\beta$.

Altogether, $\Lambda - \tilde{g} = (-\beta) \vee (\Lambda - \alpha) \wedge \beta + \alpha = (\alpha - \beta) \vee \Lambda \wedge (\alpha + \beta)$ seems to be the necessary form of $\tilde{q} = \Lambda - \tilde{g}$.

Now define $\tilde{q} = v' \vee \Lambda \wedge v''$ by means of the unique solutions $v' < 0 < v''$ of $\operatorname{E}(v' - \Lambda)_+ = r = \operatorname{E}(\Lambda - v'')_+$, which is a matter of continuity (dominated convergence theorem), monotony (strict), and the intermediate value theorem. We shall verify that this $\tilde{q}$ minimzes $\operatorname{E}q^2$ subject to $\operatorname{E}q = 0$, $\operatorname{E}|\Lambda - q| \le 2r$.

By the definition of $\tilde{q}$, $\operatorname{E}(\Lambda - q)\tilde{q} \le v'' \operatorname{E}(\Lambda - q)_+ - v' \operatorname{E}(q - \Lambda)_+$, which is less or equal $r(v'' - v') = \operatorname{E}(\Lambda - \tilde{q})\tilde{q}$. Thus $\operatorname{E}(-\tilde{q})(q - \tilde{q}) \le 0$, which is (3.4). ////

Theorem 3.5 [contamination] *The semiparametric IC $\tilde{\varrho}_c$ exists only if*

$$r < -\max_{j=1,\dots,k} \inf_P \Lambda_j \quad (3.11)$$

where $\inf_P$ denotes the P essential infimum. And then $\Lambda^{(c)} = \Lambda - \widetilde{\Pi}_2(\Lambda)$ has coordinates

$$\Lambda_j^{(c)} = (\Lambda_j + r) \wedge u_j \quad (3.12)$$

with clipping constant $u_j > 0$ *uniquely determined by*

$$0 = \mathrm{E}\,(\Lambda_j + r) \wedge u_j \tag{3.13}$$

PROOF Obviously, $\Lambda_j - \widetilde{\pi}_2(\Lambda_j) = 0$ iff $\Lambda_j \geq -r$ a.e. P. Thus assume (3.11).

In case $k = 1$, in order to minimize $\mathrm{E}(\Lambda - g)^2$ for $g \in \mathcal{G}_c$, we pass to the equivalent problem of minimizing $\mathrm{E}\,q^2$ subject to $\mathrm{E}\,q = 0$, $q \leq \Lambda + r$, for which we minimize a Lagrangian $\mathrm{E}\,q^2 - 2u\,\mathrm{E}\,q = \mathrm{E}(q-u)^2 + \text{constant}$, subject to $q \leq \Lambda + r$. Doing this pointwise, the necessary form seems $\tilde{q} = (\Lambda + r) \wedge u$.

Now consider the function $f(s) = \mathrm{E}(\Lambda + r) \wedge s$ for $s \geq 0$. It is monotone, continuous [dominated convergence applies since $-(\Lambda + r)_- \leq f \leq (\Lambda + r)_+$], and has limits $-\mathrm{E}(\Lambda + r)_- < 0$ and $r \geq 0$ at 0 and ∞, respectively. Thus f has a zero $u > 0$, which we use to define $\tilde{q} = (\Lambda + r) \wedge u$. (Only in case $r = 0$, may u be nonunique, but then $\tilde{q} = \Lambda$.) By construction, $\tilde{q}$ satisfies the side conditions $\mathrm{E}\,q = 0$, $q \leq \Lambda + r$.

To prove $\tilde{q}$ optimal, let $q \in L_2$ be any such function. Then $q \leq \tilde{q} = \Lambda + r$ as soon as $\tilde{q} < u$. Thus $(u - \tilde{q})(u - q)$ is always greater or equal to $(u - \tilde{q})^2$. Consequentially, $\mathrm{E}(-\tilde{q})(q - \tilde{q}) = \mathrm{E}(u - \tilde{q})(q - u + u - \tilde{q}) \leq 0$; which is (3.4). ////

Remark 3.6 In Theorems 3.4 and 3.5, conditions (3.8) and (3.11), respectively, ensure that $\mathrm{E}\,\Lambda_j^{(*)}\Lambda_j > 0$ for $* = v, c$. This may be seen by writing

$$\mathrm{E}\,\Lambda_j^{(v)}\Lambda_j = \mathrm{E}\,\Lambda_j^{(v)}\Lambda_j^{(v)} + r\,(v_j'' - v_j') \tag{3.14}$$

where $r\,(v_j'' - v_j') > 0$ unless $r = 0$ (and then $\Lambda_j^{(v)} = \Lambda_j$, and $\mathcal{I}_{j,j} > 0$), respectively by writing

$$\mathrm{E}\,\Lambda_j^{(c)}\Lambda_j = \mathrm{E}\,\Lambda_j^{(c)}\Lambda_j^{(c)} + \mathrm{E}\,\Lambda_j^{(c)}(\Lambda_j + r - \Lambda_j^{(c)}) \tag{3.15}$$

where $\Lambda_j + r \leq u_j$ a.e. P only if $r = 0$ (and again $\Lambda_j^{(c)} = \Lambda_j$, $\mathcal{I}_{j,j} > 0$).

However, whether condition (3.8), respectively (3.11), for dimension $k > 1$ already imply the nonsingularity of $\mathcal{K}$, hence the existence of $\tilde{\varrho}_v$, respectively of $\tilde{\varrho}_c$, is unclear. ////

Remark 3.7 The optimization problems of this section resemble those that determine robust influence curves, however with three distinctions:

(1) The approximation $\mathrm{E}\,|\Lambda - g|^2 = \min!$, instead of $\mathrm{E}\,|\psi|^2 = \min!$.

(2) The L_2, L_1, and $\inf_P$ bounds on tangents translate into bounds on influence curves in the dual norm $\sup_{g \in \mathcal{G}_*} |\mathrm{E}\,\psi g|$, for $* = h, v, c$, respectively.

(3) There is no condition on tangents that would correspond to the Fisher consistency $\mathrm{E}\,\psi\Lambda' = \mathbb{I}_k$ of influence curves. ////

4 Comparison of Semiparametric and Robust Estimators

We shall prove, respectively disprove, the semiparametric recipe (3.1)–(3.2) by comparison of results. How does the semiparametric estimator—the asymptotically linear estimator with semiparametric influence curve $\tilde{\varrho}_*$—compare with the robust estimator—the asymptotically linear estimator with robust influence curve η_* that, by definition, minimizes maximum asymptotic mean square error of asymptotically linear estimators? The maximum is evaluated over shrinking neighborhoods $U_*(\theta_0; r/\sqrt{n}\,)$, as the sample size n tends to infinity, with starting radius $r \ge 0$—henceforth, radius r—fixed. For asymptotically linear estimators, this maximum asymptotic MSE naturally extends the covariance criterion employed in the Cramér–Rao bound to the infinitesimal robust setup.

Remark 4.1 An extension of asymptotic maximum MSE over neighborhoods, from asymptotically linear to arbitrary estimators $S = (S_n)$, employing a risk such as

$$\lim_{b\to\infty} \lim_{c\to\infty} \limsup_{n\to\infty} \sup_{|t|\le c} \sup_{Q\in U_n(t;r)} \int b \wedge |R_n|^2 \, dQ^n \tag{4.1}$$

where $U_n(t;r) = U_*(\theta_0 + t/\sqrt{n}\,, r/\sqrt{n}\,)$ of fixed radius r, and $R_n = \sqrt{n}\,(S_n - \theta_0)$, has not been achieved. Theorem 4.1(A) of HR (1981 b), which admits arbitrary estimators, is restricted to one sided confidence probabilities, dimension $k = 1$, and total variation, contamination neighborhoods (for which least favorable testing pairs exist). Therefore, except in this special robust setup, our comparison of semiparametric and robust ICs is tied with asymptotically linear estimators.////

For the estimation of θ_0, over shrinking neighborhoods $U_*(\theta_0; r/\sqrt{n}\,)$, radius r, we consider a weighted MSE with nonnegative bias weight β. In the case of estimators of θ_0 that are asymptotically linear with influence curves ψ at θ_0, the maximum asymptotic weighted mean square error is

$$\mathrm{MSE}_*(\psi; \beta, r) = \mathrm{E}\,|\psi|^2 + \beta r^2 \omega_*^2(\psi) \tag{4.2}$$

As for the derivation of this risk with weight $\beta = 1$, the bias terms $\omega_*(\psi)$, and the minimization of $\mathrm{MSE}_*(\psi; \beta, r)$ for $\psi \in \Psi$, which determines the robust IC η_* uniquely, please confer HR (1994; chapter 5, subsection 5.5.2).

The influence curves $\Psi = \Psi_{\theta_0}$, and asymptotic linearity of estimators, are defined with respect to the ideal model $\mathcal{P}$ at θ_0.

4.1 Coincidence in Hellinger Model

Hellinger bias, according to HR (1994; Proposition 5.5.3), is given in terms of the maximum eigenvalue of the covariance, $\omega_h^2(\psi) = 8 \max\mathrm{ev}\, \mathrm{C}\,\psi$. In view of the Cramér–Rao bound (1.17), therefore, Hellinger risk $\mathrm{MSE}_h(\,.\,; \beta, r)$ is minimized by the canonical IC (1.11): $\hat{\varrho} = \mathcal{I}^{-1}\Lambda$, for every $\beta, r \in [0,\infty)$. Theorem 3.3 thus shows the following coincidence.

Theorem 4.2 *Assume* (3.5): $8\,r^2 < \min_{j=1,\dots,k} \mathcal{I}_{j,j}$. *Then the semiparametric IC* $\tilde{\varrho}_h$ *agrees with the robust IC* η_h,

$$\tilde{\varrho}_h = \hat{\varrho} = \mathcal{I}^{-1}\Lambda = \eta_h \tag{4.3}$$

minimizing $\mathrm{MSE}_h(\,.\,;\beta,r)$, *for every* $\beta \in [0,\infty)$.

In principle, the coincidence justifies the semiparametric recipe. The value of this result, however, is somewhat diminished since Hellinger neighborhoods, in certain respects, are deemed too small; confer Bickel (1981; Théorème 8) and HR (1994; Example 6.1.1). The gross error neighborhoods (total variation, contamination) seem in practice more suitable for robustness.

Remark 4.3 Identity (4.3) implies equality $\mathrm{C}\,\hat{\varrho} = \mathrm{C}\,\eta_h$ in (1.12), with the semiparametric and robust IC $\eta_h = \hat{\varrho}$ replacing the canonical IC ϱ, which might suggest adaptivity. However, due to bias, covariance alone does not define the right risk in the Hellinger model $\mathcal{Q}_h$, which is why MSE_h is used. But clearly,

$$\mathrm{MSE}_h(\eta_h;\beta,r) = \mathrm{tr}\,\mathcal{I}^{-1} + 8\,\beta r^2 \,\mathrm{max\,ev}\,\mathcal{I}^{-1} > \mathrm{tr}\,\mathcal{I}^{-1} = \mathrm{MSE}_h(\hat{\varrho};\beta,0) \tag{4.4}$$

if only $\beta r > 0$. Despite $\eta_h = \hat{\varrho}$ achieves minimum MSE in model $\mathcal{Q}_h$ as well as in $\mathcal{P}$, strict inequality holds in (4.4), so adaptivity is violated: Hellinger neighborhoods do not go for free. ////

4.2 Relations for Total Variation

Dimension $k = 1$

Total variation bias in one dimension, according to HR (1994; Proposition 5.5.3), is $\omega_v(\psi) = \sup_P \psi - \inf_P \psi$. The robust IC η_v minimizing $\mathrm{MSE}_v(\,.\,;\beta,r)$ is given by HR (1994; Theorem 5.5.7), with βr^2 replacing β there. Thus,

$$\eta_v = c' \vee A\Lambda \wedge c'' \tag{4.5}$$

for any numers $c' < 0 < c''$ and A such that $\mathrm{E}\,\eta_v = 0$, $\mathrm{E}\,\eta_v\Lambda = 1$, and

$$\beta r^2 (c'' - c') = \mathrm{E}(c' - A\Lambda)_+ \tag{4.6}$$

The following result justifies the semiparametric recipe (3.1)–(3.2) if one accepts the particular bias weight implicitly defined by (4.7).

Theorem 4.4 *Assume* (3.8): $r < \mathrm{E}\,\Lambda_+$. *Then the semiparametric IC* $\tilde{\varrho}_v$ *agrees with the robust IC* η_v *minimizing* $\mathrm{MSE}_v(\,.\,;\beta,r)$, *iff bias weight* $\beta = \beta(r)$ *is chosen such that*

$$\beta^{-1} = r\,(v'' - v') \tag{4.7}$$

where $v' = v'(r) < 0 < v''(r) = v''$ *are determined by*

$$\mathrm{E}(v' - \Lambda)_+ = r = \mathrm{E}(\Lambda - v'')_+ \tag{3.10}$$

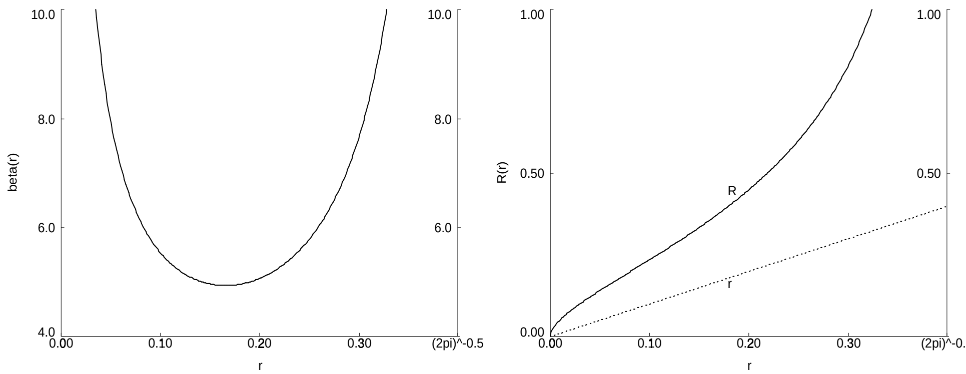

Figure 1: Bias weight $\beta(r)$ and radius $R(r)$ versus radius $0 < r < 1/\sqrt{2\pi}$, for total variation neighborhoods $U_v(\theta, r/\sqrt{n})$ about the ideal location model $P_\theta = \mathcal{N}(\theta, 1)$.

PROOF Theorem 3.4 supplies $\tilde{\varrho}_v = A\, v' \vee \Lambda \wedge v''$ with clipping constants v', v'' determined by (3.10) and rescaling constant $A^{-1} = \mathcal{K} > 0$ (Remark 3.6).

Thus $\tilde{\varrho}_v$ attains form (4.5) with $c' = v'A$ and $c'' = v''A$; in particular, $\beta r^2(c'' - c') = \beta r^2(v'' - v')A$. Since $Ar = A\,\mathrm{E}(v' - \Lambda)_+ = \mathrm{E}(c' - A\Lambda)_+$ by (3.10), equation (4.6) is the same as (4.7). ////

Bias weight $\beta = 1$, in view of (4.1), is the most natural choice. Then the semiparametric IC $\tilde{\varrho}_v$ minimizes $\mathrm{MSE}_v(\,.\,;1, r_1)$, since it equals the robust IC η_v for this radius r_1 , iff

$$r_1^{-1} = v''(r_1) - v'(r_1) \tag{4.8}$$

Let us keep bias weight $\beta = 1$. Then the semiparametric IC $\tilde{\varrho}_v$ defined for radius r minimizes the risk $\mathrm{MSE}_v(\,.\,;1, R(r))$ for another radius $R(r)$ given by

$$R^2(r) = r\big/\big(v''(r) - v'(r)\big) = r^2\beta(r) \tag{4.9}$$

since $\tilde{\varrho}_v$ is of form (4.5) and (4.6), hence is the robust η_v , for this radius $R(r)$. Also, by (4.7), it holds that $R(r) = r\sqrt{\beta(r)}$, and (4.8) means that $R(r_1) = r_1$.

Example 4.5 For the standard normal location model $P_\theta = \mathcal{N}(\theta, 1)$, Figure 1 shows the bias weight $\beta(r)$ and the radius $R(r)$ defined by (4.7) and (4.9), respectively. The function $\beta(.)$ has singularities at 0 and the right boundary, which is $1/\sqrt{2\pi} = 0.3989$, and attains its minimum value $\beta_{\mathrm{min}} = 4.8662$ at $r_{\mathrm{min}} = 0.1668$. In particular, no radius r_1 for which $\beta(r_1) = 1$ exists.

The radius $R(r)$ descends to 0, hence $\beta(r) = \mathrm{o}(r^{-2})$, as $r \to 0$, and rises to ∞ as $r \to 1/\sqrt{2\pi}$. Since $R(r)/r = \sqrt{\beta(r)}$ is always larger than $\sqrt{\beta_{\mathrm{min}}}$, the semiparametric IC $\tilde{\varrho}_v$ safeguards against more than double the amount of contamination assumed in its definition (3.1)–(3.2), and, as $\beta(r) > \beta_{\mathrm{min}}$, is typically even more pessimistic. ////

Confidence risk

The asymptotic maximum risk considered in HR (1981 b), instead of mean square error, and bounded from below for arbitrary estimators (S_n), is based on right and left confidence probabilities as follows,

$$\lim_{c\to\infty} \limsup_{n\to\infty} \sup_{|t|\le c} \sup_{Q\in U_n(t;r)} Q^n(R_n < -\tau) \vee Q^n(R_n > \tau) \tag{4.10}$$

where $U_n(t;r) = U_v(\theta_0 + t/\sqrt{n}\,, r/\sqrt{n}\,)$ of fixed radius r, and $\tau \in (0,\infty)$ is some interval half-width. As already in (4.1), the standardization $R_n = \sqrt{n}\,(S_n - \theta_0)$ is needed only for the description of the asymptotic minimax estimator as an asymptotically linear one.

Theorem 4.6 *Assume* (3.8): $r < \mathrm{E}\,\Lambda_+$. *Then the semiparametric IC* $\tilde{\varrho}_v$ *agrees with the robust IC* η_v *with respect to confidence risk* (4.10) *iff we choose half-width*

$$\tau = \tau(r) = 1 \tag{4.11}$$

PROOF According to HR (1981 b; Theorems 4.1(A)–4.3; 1980; Theorem 3.1), for radius

$$r < \tau\,\mathrm{E}\,\Lambda_+ \tag{4.12}$$

the estimator (S_n) minimizing risk (4.10) is asymptotically linear at θ_0 with IC η_v of form (3.9) and (3.10), however, with r in (3.10) replaced by r/τ.

Thus, the semiparametric IC $\tilde{\varrho}_v$ is the robust IC η_v iff $\tau = 1$ in risk (4.10). And then, condition (4.12) on r is the same as (3.8). ////

Dimension $k > 1$

Exact total variation bias for more than one dimension is rather unwieldy, $\omega_v(\psi) = \sup_{|e|=1} \sup_P e'\psi - \inf_P e'\psi$, where $\sup_{|e|=1}$ extends over all unit vectors in $\mathbb{R}^k$; confer HR (1994; Proposition 5.5.3). Approximate versions $\omega_{v;2}^2(\psi)$ and $\omega_{v;\infty}(\psi)$ are defined by the Euclidean and sup norms in $\mathbb{R}^k$ of the vector of coordinate biasses $\omega_v(\psi_j)$, respectively, which bound the exact bias from below and above: $\omega_{v;\infty} \le \omega_v \le \omega_{v;2} \le \sqrt{k}\,\omega_{v;\infty}$. According to HR (1994; Theorems 5.5.6–7) on one hand, the robust ICs η_v minimizing either risk $\mathrm{MSE}_{v;s}(\,.\,;\beta,r)$ have the coordinates

$$\eta_j = c_j' \vee A_j\Lambda \wedge c_j'' \tag{4.13}$$

with any numbers $c_j' < 0 < c_j''$ and row vectors $A_j \in \mathbb{R}^k$ such that the side conditions $\mathrm{E}\,\eta_v = 0$ and $\mathrm{E}\,\eta_v\Lambda' = \mathbb{I}_k$ are met. Moreover, the clipping constants satisfy

$$\beta r^2(c_j'' - c_j') = \mathrm{E}\,(c_j' - A_j\Lambda_j)_+ \tag{4.14}$$

in case $s = 2$, whereas, in case $s = \infty$, the differences $c''_j - c'_j$ are all the same

$$\beta r^2 (c''_j - c'_j) = \mathrm{E}\,(c'_1 - A_1 \Lambda_1)_+ + \cdots + \mathrm{E}\,(c'_k - A_k \Lambda_k)_+ \tag{4.15}$$

By Theorem 3.4 on the other hand, with clipping constants $v'_j < 0 < v''_j$ defined by (3.10), and $(A^{j,i})^{-1} = \mathcal{K}$ by (3.2), the semiparametric IC $\tilde{\varrho}_v$ has coordinates

$$\tilde{\varrho}_j = A^{j,1}\, v'_1 \vee \Lambda_1 \wedge v''_1 + \cdots + A^{j,k}\, v'_k \vee \Lambda_k \wedge v''_k \tag{4.16}$$

Thus, the order of clipping and linear combination is interchanged in $\tilde{\varrho}_v$ and η_v. So $\tilde{\varrho}_v$ resembles, but does not exactly match, the robust η_v, therefore does not minimize either risk $\mathrm{MSE}_{v;s}(\,.\,;\beta, r)$, $s = 2, \infty$, if only $\beta r > 0$.

However, the bias terms $\omega_{v;s}$ are only bounds for the exact bias ω_v, while $\tilde{\varrho}_v$ ought to be compared with the minimizer of the exact risk $\mathrm{MSE}_v(\,.\,;\beta, r)$. And, at least, $\tilde{\varrho}_v$ has finite biasses $\omega_{v;s}(\tilde{\varrho}_v)$ and $\omega_v(\tilde{\varrho}_v)$, hence finite risks $\mathrm{MSE}_{v;s}(\tilde{\varrho}_v; \beta, r)$, and $\mathrm{MSE}_v(\tilde{\varrho}_v; \beta, r)$.

The relative increase of risk of the semiparametric IC $\tilde{\varrho}_v$ over that of the robust IC η_v remains to be investigated numerically—even in one dimension when $\beta \neq \beta(r)$. A suboptimal $\tilde{\varrho}_v$ may still be useful.

4.3 Discrepancy for Contamination

Contamination bias is $\omega_c(\psi) = \sup_P |\psi|$, the L_∞ norm. The robust IC η_c which minimizes $\mathrm{MSE}_c(\,.\,;\beta, r)$, by HR (1994; Theorem 5.5.6), is the Hampel–Krasker influence curve,

$$\eta_c = (A\Lambda - a)\, w\,, \qquad w = \min\Bigl\{1,\ \frac{b}{|A\Lambda - a|}\Bigr\} \tag{4.17}$$

with a particular bound, namely, the solution b to the equation

$$\beta r^2 b = \mathrm{E}\,\bigl(|A\Lambda - a| - b\bigr)_+ \tag{4.18}$$

which may be nonunique only if $\beta r = 0$ (in which case $\eta_c = \hat{\varrho}$).

The semiparametric IC $\tilde{\varrho}_c$, by Theorem 3.5, has coordinates

$$\tilde{\varrho}_j = A^{j,1}\,(\Lambda_1 + r) \wedge u_1 + \cdots + A^{j,k}\,(\Lambda_k + r) \wedge u_k \tag{4.19}$$

with upper clipping constants u_j defined by (3.13), and $(A^{j,i})^{-1} = \mathcal{K}$ by (3.2).

Thus, in general, $\tilde{\varrho}_c$ is unbounded so that the risk $\mathrm{MSE}_c(\tilde{\varrho}_c; \beta, r)$ becomes infinite, if only $\beta r > 0$ (the only interesting case).

The intuitive convex combinations, which have been used in robust statistics prior to any other type of neighborhoods, have always turned out very similar to total variation in robustness respects. It is therefore surprising that the semiparametric recipe (3.1), (3.2) may give reasonable results for one model but not the other. However, in the simplest testing context (one parameter, one-sided), the discrepancy disappears again; confer Remark 7.3 and Theorem 7.7.

5 Unresolved: Robust Adaptive Estimation

In the general semiparametric model of Section 1, given the canonical influence curves (1.7), one $\varrho_{\theta,\nu}$ for each parameter $\theta \in \Theta$, $\nu \in H_\theta$, the construction problem is to obtain an estimator (S_n) that, for each $\theta \in \Theta$ and $\nu \in H_\theta$, is asymptotically linear at (θ, ν) with prescribed IC $\varrho_{\theta,\nu}$.

Such estimators are automatically nonrobust in the same setup—asymptotic, infinitesimal—in which their efficiency is obtained.

For example, consider the model $dQ_{\theta,\nu}(x) = \nu(x-\theta)\,dx$ with location parameter $\theta \in \mathbb{R}$ and nuisance parameter ν any symmetric Lebesgue density of finite Fisher information of location, $\mathcal{I}_\nu = \int \Lambda_\nu^2(x)\,\nu(x)\,dx$, $\Lambda_\nu = -\dot\nu/\nu$; then $\Lambda_{\theta,\nu}(x) = \Lambda_\nu(x-\theta)$. In this model, adaptivity $\Pi_{2;\theta,\nu}(\Lambda_{\theta,\nu}) = 0$ holds by reasons of symmetry. Adaptive estimators have been constructed by Beran (1974) and Stone (1975) and, at each (θ, ν), have expansion (1.5) with influence curve $\varrho_{\theta,\nu}(x) = \hat\varrho_{\theta,\nu}(x) = \mathcal{I}_\nu^{-1}\Lambda_\nu(x-\theta)$. Hence, under $Q^n_{\theta,\nu}$, they achieve normal limit law $\mathcal{N}(0, \mathcal{I}_\nu^{-1})$, as if ν was known.

The assumption of exact symmetry, however, is very strict. In practice, one would accept a distribution function as symmetric if it only is in a small neighborhood of an exactly symmetric one. Such nonparametric hypotheses of approximate symmetry have been investigated by HR (1981 a; section 3) and generalized by Kakiuchi and Kimura (2000). If $Q_{\theta,\nu}$ is thus enlarged to a shrinking neighborhood $U_*(\theta, \nu; r/\sqrt{n})$, while still θ has to be estimated, the adaptive estimates $\sqrt{n}\,(S_n - \theta)$ are driven off from their limit $\mathcal{N}(0, \mathcal{I}_\nu^{-1})$ by some bias up to $\pm r\,\omega_*(\hat\varrho_{\theta,\nu})$ which, for gross error neighborhoods ($* = v, c$), may become infinite if only $\Lambda_\nu = -\dot\nu/\nu$ is unbounded.

This observation obviously extends to the general semiparametric model if the canonical influence curve $\varrho_{\theta,\nu}$ is unbounded.

Other robustness aspects, not considered in this paper, are breakdown point and qualitative robustness. Possibly related is Klaassen's result on the nonuniform convergence of adaptive estimators in the symmetric location case; confer Bickel's (1981) presentation. Pfanzagl and Wefelmeyer (1982; Proposition 9.4.1, Corollary 9.4.5) have similar results, which connect the nonuniformity with discontinuity of the Fisher information. On the contrary, it is easy to see (since the Lindeberg condition may be verified uniformly) that Huber's (1964) minimax location M-estimate tends to its normal limits uniformly on the corresponding symmetric contamination neighborhood.

In view of all this, it seems desirable to construct estimators not with the canonical influence curves $\varrho_{\theta,\nu}$ but robust influence curves $\eta_{\theta,\nu}$ instead, sacrificing a few percent efficiency under $Q_{\theta,\nu}$ to gain robustness against deviations from $Q_{\theta,\nu}$. The problem has also been adressed, and declared a field of future research, by Bickel et al. (1993; Introduction, p 4).

A first step in this direction has been made by Shen (1995; Theorem 2) who derives a bounded influence curve $\eta_{\theta,\nu} = \eta_c$ minimizing $\mathrm{E}\,|\psi|^2$ among all influence curves $\psi \in \Psi$, as defined in (1.3) for a general semiparametric model, subject to $|\psi| \le \sup |\eta_c|$. In some sense, the result may be viewed an extension of HR (1994; Theorem 5.5.1), from finite to infinite dimensional nuisance tangent space $\partial_2 \mathcal{Q}$ (of a certain kind; namely, an L_2 space of functions, expectation zero, and measurable relative to some sub σ algebra of $\mathcal{B}$).

The corresponding robust, and adaptive, estimator construction, however, has not been done yet.

6 Semiparametric $C(\alpha)$-Tests

The semiparametric approach is carried further to the testing of hypotheses about the parameter of interest. The optimal tests are generalized $C(\alpha)$-tests, which are based on residual scores after an orthogonal projection on the closed linear tangent space for the nuisance parameter. In connection with the robust tangent balls, the nonlinear projection on these balls will be employed instead.

6.1 $C(\alpha)$-Tests For Tangent Spaces

Recall the the general setup of Section 1: The semiparametric probability model $\mathcal{Q} = \{\, Q_{\theta,\nu} \mid \theta \in \Theta\, ,\ \nu \in H_\theta \,\}$ with main parameter θ, nuisance parameter ν, the fixed parameter value (θ_0, ν_0) and corresponding element $Q = Q_{\theta_0,\nu_0}$, the scores function $\Lambda \in L_2^k = L_2^k(Q)$ of $\mathcal{Q}$ for θ and the differentiablity (1.2) of $\mathcal{Q}$ at (θ_0, ν_0), the orthogonal projection $\Pi_2 \colon L_2^k \to (c\ell \operatorname{lin} \partial_2 \mathcal{Q})^k$, and the Fisher information $\mathcal{J} = \mathrm{C}\,\bar\Lambda$ of $\mathcal{Q}$ for θ at (θ_0, ν_0), where $\bar\Lambda$ denotes the residual scores

$$\bar\Lambda = \Lambda - \Pi_2(\Lambda) \tag{6.1}$$

Given some numbers $-\infty < z_1 < z_2 < \infty$ and $0 \le z_3 < z_4 < \infty$, local asymptotic one- and multisided hypotheses about the difference between the true θ and its reference value θ_0 are defined by

$$H' : \ e'\mathcal{J}^{1/2} a \le z_1 \qquad \text{vs.} \qquad K' : \ e'\mathcal{J}^{1/2} a \ge z_2 \tag{6.2}$$

$$H'' : \quad a'\mathcal{J} a \le z_3^2 \qquad \text{vs.} \qquad K'' : \quad a'\mathcal{J} a \ge z_4^2 \tag{6.3}$$

where $e \in \mathbb{R}^k$, $|e| = 1$, is some fixed unit vector, and $\mathcal{J}^{1/2} = A$ any $k \times k$ root of $\mathcal{J}$ such that $AA' = \mathcal{J}$.

The hypotheses concern the sequence of laws $Q_n \in \mathcal{Q}$ of the n i.i.d. observations $x_1, \ldots, x_n \sim Q_n$. It is assumed that, for any $a \in \mathbb{R}^k$ and $g \in \partial_2 \mathcal{Q}$, eventually (that is, for all but finitely many n),

$$Q_n = Q_n(a, g) = Q_{\theta_0 + s_n a,\, \nu^g_{s_n}} \tag{6.4}$$

where $s_n = 1/\sqrt{n}$ and $t \mapsto \nu_t^g \in H_{\theta_0 + ta}$ is some path with tangent g in (1.2).

We employ asymptotic tests $\delta = (\delta_n)$, that is, sequences of tests δ_n at sample size n. Their error probabilities will be evaluated under the n fold product measures Q_n^n, asymptotically, as n tends to infinity.

For $\alpha \in (0,1)$, let u_α denote the upper α point of the standard normal distribution Φ, such that $\Phi(-u_\alpha) = \alpha$. By $\chi^2(k, z^2)$ denote the χ^2 distribution with k degress of freedom and noncentrality z^2, respectively a random variable having this distribution, and by $c_\alpha(k, z^2)$ its upper α point.

Theorem 6.1 *Let $\delta = (\delta_n)$ be any sequence of tests.*

(a) *Then, in the one-sided case,*

$$\sup_{H'} \limsup_{n\to\infty} \int \delta_n \, dQ_n^n(a,g) \le \alpha \tag{6.5}$$

implies

$$\inf_{K'} \limsup_{n\to\infty} \int \delta_n \, dQ_n^n(a,g) \le \Phi\bigl(-u_\alpha + (z_2 - z_1)\bigr) \tag{6.6}$$

(b) *In the multisided case,*

$$\sup_{H''} \limsup_{n\to\infty} \int \delta_n \, dQ_n^n(a,g) \le \alpha \tag{6.7}$$

implies

$$\inf_{K''} \limsup_{n\to\infty} \int \delta_n \, dQ_n^n(a,g) \le \Pr\bigl(\chi^2(k, z_4^2) > c_\alpha(k, z_3^2)\bigr) \tag{6.8}$$

(c) *Bounds* (6.6) *and* (6.8), *with* $\limsup$ *replaced by* $\liminf$, *are achieved by the asymptotic tests*

$$\delta' = (\delta'_n), \qquad \delta'_n = \mathbf{I}\bigl(e'\mathcal{J}^{-1/2} Z_n > u_\alpha + z_1\bigr) \tag{6.9}$$

$$\delta'' = (\delta''_n), \qquad \delta''_n = \mathbf{I}\bigl(Z'_n \mathcal{J}^{-1} Z_n > c_\alpha(k, z_3^2)\bigr) \tag{6.10}$$

respectively, where $Z_n = 1/\sqrt{n}\, \sum_{i=1}^n \bar\Lambda(x_i)$.

PROOF The differentiability (1.2), for every $a \in \mathbb{R}^k$, $g \in \partial_2 \mathcal{Q}$, entails the following loglikelihood expansion,

$$\log \frac{dQ_n^n(a,g)}{dQ^n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n (a'\Lambda + g)(x_i) - \frac{1}{2} \|a'\Lambda + g\|^2 + \mathrm{o}_{Q^n}(n^0) \tag{6.11}$$

Thus, given $a \in \mathbb{R}^k$, the Fisher information $\|a'\Lambda + g\|^2$ at $t = 0$ of the one parameter family $\mathcal{Q}(a,g) = \{Q_{\theta_0 + ta,\, \nu_t^g}\}$ is minimized with respect to $g \in \partial_2 \mathcal{Q}$ by $g_a = -\pi_2(a'\Lambda) = -a'\Pi_2(\Lambda)$. Therefore, associating with each $a \in \mathbb{R}^k$ any path $\nu_t^a = \nu_t^{g_a}$, the sequence of k parameter submodels $\mathcal{Q}_n = \{\, Q_{n,a} \mid a \in \mathbb{R}^k \,\}$ of $\{\, Q_n(a,g) \mid a \in \mathbb{R}^k,\ g \in \partial_2 \mathcal{Q} \,\}$, consisting of the elements $Q_{n,a} = Q_n(a, g_a)$, will turn out least favorable.

In fact, as $a'\Lambda + g_a = a'\bar\Lambda$ and $\mathrm{C}\,\bar\Lambda = \mathcal{J}$, expansion (6.11) specializes to

$$\log \frac{dQ_{n,a}^n}{dQ^n} = \frac{a'}{\sqrt{n}} \sum_{i=1}^n \bar\Lambda(x_i) - \frac{1}{2} a'\mathcal{J} a + \mathrm{o}_{Q^n}(n^0) \tag{6.12}$$

Because of this asymptotic normality, of the sequence of product models $\mathcal{Q}_n^n$, Theorems 3.4.6, 3.4.11 of HR (1994) are in force and, subject to (6.5) and (6.7), respectively, furnish the power bounds (6.6) and (6.8), as well as the asymptotically most powerful level α tests δ' and δ'', for the sequence of submodels.

But, for arbitrary tangents $g \in \partial_2 \mathcal{Q}$, (6.11) implies the following asymptotic normality of Z_n under $Q_n^n(a,g)$,

$$Z_n(Q_n^n(a,g)) \xrightarrow{\;\mathrm{w}\;} \mathcal{N}\big(\mathrm{E}\,\bar\Lambda(a'\Lambda + g), \mathcal{J}\big) \tag{6.13}$$

where

$$\mathrm{E}\,\bar\Lambda(a'\Lambda + g) = \mathrm{E}\,\bar\Lambda\Lambda' a = \mathcal{J} a \tag{6.14}$$

since $\bar\Lambda$ is orthogonal to $\Pi_2(\Lambda)$ and g. Hence, the asymptotic error probabilities of the tests δ' and δ'' do not depend on $g \in \partial_2 \mathcal{Q}$. ////

Remark 6.2 The orthogonality of $\bar\Lambda$ on $\partial_2 \mathcal{Q}$ may be used a second time to construct test statistics that do not require knowledge of ν_0.

In the finite dimensional case, confer Remark 1.3, upon a regularization of the likelihoods, (total) scores function, and Fisher information, estimates of ν which are $\sqrt{n}$ consistent and suitably discretized may be inserted for ν_0; confer HR (1994; Lemmas 6.4.1 and 6.4.4). Thus Neyman's $C(\alpha)$-tests are obtained.

The test statistics Z_n may also be replaced by an estimator $S = (S_n)$ of θ which is asymptotically linear, in the sense of (1.5), at each $Q_{\theta,\nu}$, with canonical influence curve $\varrho_{\theta,\nu} = \mathcal{J}_{\theta,\nu}^{-1}\bar\Lambda_{\theta,\nu}$; confer HR (1994; Theorem 6.4.8). This leads to Wald's estimator tests λ' and λ'',

$$\lambda' = (\lambda_n'), \qquad \lambda_n' = \mathbf{I}\big(e'\mathcal{J}^{1/2}\sqrt{n}\,(S_n - \theta_0) > u_\alpha + z_1\big) \tag{6.15}$$

$$\lambda'' = (\lambda_n''), \qquad \lambda_n'' = \mathbf{I}\big(n\,(S_n - \theta_0)'\mathcal{J}\,(S_n - \theta_0) > c_\alpha(k, z_3^2)\big) \tag{6.16}$$

Like δ' and δ'', also the test sequences λ' and λ'' achieve maxmin asymptotic power subject to level α for H' vs. K', respectively for H'' vs. K''.

In the infinite dimensional case, the estimation of $\bar\Lambda_{\theta_0,\nu_0}$ and $\mathcal{J}_{\theta_0,\nu_0}$ (with θ_0 known, ν_0 unknown), and the construction of an asymptotically linear estimator with canonical influence curve $\varrho_{\theta,\nu} = \mathcal{J}_{\theta,\nu}^{-1}\bar\Lambda_{\theta,\nu}$ at $Q_{\theta,\nu}$ (at least for $\theta = \theta_0$ and every $\nu \in H_{\theta_0}$) is more difficult. The methods of Klaassen (1987) and the references mentioned therein may prove useful. ////

6.2 $C(\alpha)$-Tests For Tangent Balls

Recall the setup of Section 2: Starting from an ideal, smooth k parametric model $\mathcal{P} = \{\, P_\theta \mid \theta \in \Theta \,\}$ without nuisance parameter, Theorem 6.1 first specializes with $\partial_2 \mathcal{P} = \{0\}$. Thus, the classical test sequences $\hat\delta'$ and $\hat\delta''$,

$$\hat\delta_n' = \mathbf{I}\big(e'\mathcal{I}^{-1/2}\hat Z_n > u_\alpha + z_1\big) \tag{6.17}$$

$$\hat\delta_n'' = \mathbf{I}\big(\hat Z_n'\mathcal{I}^{-1}\hat Z_n > c_\alpha(k, z_3^2)\big) \tag{6.18}$$

based on $\hat{Z}_n = 1/\sqrt{n}\,\sum \Lambda(x_i)$, as well as the test sequences $\hat{\lambda}'$ and $\hat{\lambda}''$ employing an asymptotically linear estimator $\hat{S} = (\hat{S}_n)$ with influence curve $\hat{\varrho} = \mathcal{I}^{-1}\Lambda$ at $P = P_{\theta_0}$,

$$\hat{\lambda}'_n = \mathbf{I}\big(e'\mathcal{I}^{1/2}\sqrt{n}\,(\hat{S}_n - \theta_0) > u_\alpha + z_1\big) \tag{6.19}$$

$$\hat{\lambda}''_n = \mathbf{I}\big(n\,(\hat{S}_n - \theta_0)'\mathcal{I}\,(\hat{S}_n - \theta_0) > c_\alpha(k, z_3^2)\big) \tag{6.20}$$

achieve maxmin asymptotic power subject to level α, for the following parametric local asymptotic one- and multisided hypotheses about $\theta - \theta_0$, respectively,

$$\widehat{H}' :\ e'\mathcal{I}^{1/2}a \le z_1 \qquad \text{vs.} \qquad \widehat{K}' :\ e'\mathcal{I}^{1/2}a \ge z_2 \tag{6.21}$$

$$\widehat{H}'' :\quad a'\mathcal{I}a \le z_3^2 \qquad \text{vs.} \qquad \widehat{K}'' :\quad a'\mathcal{I}a \ge z_4^2 \tag{6.22}$$

Now enlarge the parametric measures P_θ to neighborhoods $U(\theta; r_0)$ under the null hypothesis, respectively $U(\theta; r_1)$ under the alternative. Thus, robust local asymptotic one- and multisided hypotheses $\widetilde{H}'$ vs. $\widetilde{K}'$, and $\widetilde{H}''$ vs. $\widetilde{K}''$ about $\theta - \theta_0$ are obtained. These concern the laws $Q_n \in U(\theta_0 + s_n a; s_n r_{0/1})$ at sample size n, where $s_n = 1/\sqrt{n}$, and $a \in \mathbb{R}^k$ is subject to the conditions of the corresponding parametric hypotheses $\widehat{H}'$, $\widehat{K}'$, $\widehat{H}''$, $\widehat{K}''$, respectively.

By this enlargement, size and power of the tests $\hat{\delta}'$ and $\hat{\delta}''$ will be affected without control. Conceptually, a robustification is appealing that interpretes model deviations as nuisance parameter. Then, to the neighborhood model $\mathcal{Q}$ of semiparametric form (2.1), Theorem 6.1 may again be applied, and leads to the semiparametric recipe: From Λ subtract the component $\Pi_2(\Lambda)$ explained by the nuisance parameter, and exchange the test statistics $\mathcal{I}^{-1/2}\hat{Z}_n$ based on Λ for the test statistics $\mathcal{J}^{-1/2}Z_n$ based on $\bar{\Lambda} = \Lambda - \Pi_2(\Lambda)$.

Remark 6.3 In the context of testing, contrary to estimation, there is no Fisher consistency requirement, that is, $\mathrm{E}\,\psi\Lambda' = \mathbb{I}_k$ in (1.3) and the corresponding standardization by $\mathcal{J}^{-1}$ in (1.7). The present standardization of $\bar{\Lambda}$ by $\mathcal{J}^{-1/2}$ shall achieve unit covariance of the limit normals to obtain invariance under the orthogonal group, which is needed in the proof of the maxmin testing result.////

However, in the case of Hellinger, total variation, and contamination neighborhoods, the tangent sets $\partial_2\mathcal{Q}_*$ determined by Proposition 2.2, where $* = h, v, c$, satisfy $c\ell\,\mathrm{lin}\,\partial_2\mathcal{Q}_* = L_2 \cap \{\mathrm{E} = 0\}$. Therefore, as in Section 3, we replace π_2 and Π_2 by the nonlinear projection $\widetilde{\pi}_2 : L_2 \to \mathcal{G}_*$ on $\partial_2\mathcal{Q}_* = \mathcal{G}_*$, respectively by $\widetilde{\Pi}_2 = (\widetilde{\pi}_2, \dots, \widetilde{\pi}_2)' : L_2^k \to \mathcal{G}_*^k$ (acting coordinatewise).

Actually, the situation is more complex for testing than for estimation in Section 3, since now two neighborhoods (null hypothesis, alternative) are involved. This will be clarified in Remark 7.3 below.

We first put $r = r_0 + r_1$ and naively project on $\mathcal{G}_*$ (of this radius). Thus, let

$$\tilde{\Lambda} = \Lambda - \widetilde{\Pi}_2(\Lambda) \tag{6.23}$$

and suppose that

$$\tilde{\mathcal{J}} = \mathrm{C}\,\tilde{\Lambda} > 0 \tag{6.24}$$

Then, based on $\tilde{Z}_n = 1/\sqrt{n}\,\sum \tilde{\Lambda}(x_i)$, the semiparametric approach leads to the scores statistics,

$$e'\tilde{\mathcal{J}}^{-1/2}\tilde{Z}_n\,, \qquad \tilde{Z}_n'\tilde{\mathcal{J}}^{-1}\tilde{Z}_n \tag{6.25}$$

for testing the robust one- and multisided hypotheses $\widetilde{H}'$ vs. $\widetilde{K}'$ and $\widetilde{H}''$ vs. $\widetilde{K}''$, respectively. The corresponding semiparametric estimator tests would employ the statistics

$$e'\tilde{\mathcal{J}}^{-1/2}\mathcal{K}\sqrt{n}\,(\tilde{S}_n-\theta_0)\,, \qquad n\,(\tilde{S}_n-\theta_0)'\mathcal{K}'\tilde{\mathcal{J}}^{-1}\mathcal{K}\,(\tilde{S}_n-\theta_0) \tag{6.26}$$

based on an asymptoticaly linear estimator $\tilde{S}=(\tilde{S}_n)$ with semiparametric influence curve $\tilde{\varrho}=\mathcal{K}^{-1}\tilde{\Lambda}$, provided $\mathcal{K}=\mathrm{E}\,\tilde{\Lambda}\Lambda'$ is regular; confer (3.1), (3.2).

The semiparametric asymptotic tests thus obtained are denoted by $\tilde{\delta}'$, $\tilde{\delta}''$, and $\tilde{\lambda}'$, $\tilde{\lambda}''$, respectively. The suitable choice of the critical values for their test statistics must however be left open so far.

Hellinger Model By Theorem 3.3, under condition (3.5): $8\,r^2 < \min \mathcal{I}_{j,j}$, we have $\tilde{\Lambda}=D\Lambda$ with regular matrix $D=\mathrm{diag}(1-\gamma_j)$, where $\gamma_j^2\mathcal{I}_{j,j}=8\,r^2$.

It follows that $\tilde{\mathcal{J}}=D\mathcal{I}D$, $\tilde{\mathcal{J}}^{1/2}=D\mathcal{I}^{1/2}$, and so $\tilde{\mathcal{J}}^{-1/2}\tilde{\Lambda}=\mathcal{I}^{-1/2}\Lambda$. Moreover, $\mathcal{K}=D\mathcal{I}$, hence $\mathcal{K}'\tilde{\mathcal{J}}^{-1}\mathcal{K}=\mathcal{I}$, and $\tilde{\varrho}=\hat{\varrho}=\mathcal{I}^{-1}\Lambda$ by Theorem 4.2.

Therefore, the semiparametric test statistics (6.25), (6.26) agree with the parametric test statistics in (6.17)–(6.20). The result matches Theorem 4.2.

Total Variation Model Under condition (3.8): $2\,r < \min \mathrm{E}\,|\Lambda_j|$, Theorem 3.4 furnishes $\tilde{\Lambda}$ with coordinates $\tilde{\Lambda}_j = v_j' \vee \Lambda_j \wedge v_j''$ and clipping constants are determined by (3.10). Thus the coordinates of $\tilde{\mathcal{J}}^{-1/2}\tilde{\Lambda}$ are linear combinations of $v_j' \vee \Lambda_j \wedge v_j''$, hence are bounded.

Boundedness of the semiparametric test statistics and influence curve $\tilde{\varrho}_v$, confer (4.16), ensures a minimal robustness of the corresponding semiparametric tests $\tilde{\delta}_v'$, $\tilde{\lambda}_v'$, for $\widetilde{H}_v'$ vs. $\widetilde{K}_v'$, and $\tilde{\delta}_v''$, $\tilde{\lambda}_v''$ for $\widetilde{H}_v''$ vs. $\widetilde{K}_v''$.

Contamination Model Under condition (3.11): $r < -\max \inf_P \Lambda_j$, Theorem 3.5 supplies $\tilde{\Lambda}_j = (\Lambda_j + r)\wedge u_j$, whose upper clipping constant u_j is defined by (3.13). Thus the coordinates of $\tilde{\mathcal{J}}^{-1/2}\tilde{\Lambda}$, certain linear combinations of $(\Lambda_j + r)\wedge u_j$, may be unbounded.

Unboundedness of the semiparametric test statistics and influence curve $\tilde{\varrho}_c$, confer (4.19), entails maximum asymptotic error probabilities 1 of the corresponding tests for the robust hypotheses; as with estimation in Subsection 4.3.

However, Remark 7.3 says that, instead on $\mathcal{G}_c = r\,G_c$, we must actually project on the set $r_0\,G_c - r_1\,G_c$ (which makes no difference in the Hellinger and total variation models.) The correct $\tilde{\Lambda}$ and $\tilde{\varrho}_c$, therefore, are determined by Theorem 7.7, and turn out bounded towards both sides.

Boundedness of the semiparametric test statistics and influence curve $\tilde{\varrho}_c$, now essentially of form (4.16), ensures some minimal robustness of the corresponding tests $\tilde{\delta}_c'$, $\tilde{\lambda}_c'$ for $\widetilde{H}_c'$ vs. $\widetilde{K}_c'$, and $\tilde{\delta}_c''$, $\tilde{\lambda}_c''$ for $\widetilde{H}_c''$ vs. $\widetilde{K}_c''$.

Multiparameter, Multisided Case In this general case, an exact evaluation of the asymptotic maximum size over $\widetilde{H}''$ and minimum power over $\widetilde{K}''$ of the derived semiparametric tests, and other tests based on quadratic forms in sums or in asymptotically linear estimators, is rather complicated; confer HR (1994; § 5.4, pp 192–194), especially equation (54) there. Optimization problems arise for the maximum eigenvalue of the information standardized covariance subject to bounds on the self-standardized sensitivity; see equation (55), p 194.

As these problems have not been solved yet, no optimally robust test is distinguished, in comparison of which the semiparametric tests might be judged.

It certainly is an advantage of the semiparametric approach to robust testing that it works in higher dimensions as it works in one, and that it yields test statistics which seem reasonably, if not optimally, robust.

One Parameter, One-Sided Case In the simplest case, a strong justification of the semiparametric approach is possible. Section 7 will establish optimal robustness: For the one parameter, one-sided, robust hypotheses $\widetilde{H}'$ vs. $\widetilde{K}'$, the semiparametric test $\tilde{\delta}'$ (and $\tilde{\lambda}'$) is asymptotically maxmin.

7 Saddle Points For Testing Convex Sets

Consider hypotheses which consist of local alternatives generated by any two disjoint closed convex sets G_0 and G_1 of tangents at some probability P. Picking the unique minimum norm element of $G_1 - G_0$, and the corresponding sequence of Neyman–Pearson tests, seems to fit the semiparametric projection arguments—and furnishes a saddle point.

The result applies to infinitesimal Hellinger, total variation, and contamination neighborhoods around P and a local alternative of P with fixed tangent, respectively. In the total variation and contamination cases, the maxmin asymptotic tests thus obtained by projection agree with the robust asymptotic tests based on the least favorable pairs in the sense of Huber and Strassen (1973).

7.1 Convex Sets Defining Local Alternatives

Let $P \in \mathcal{M}$ be some probability. Every tangent $\rho \in L_2 \cap \{\mathrm{E} = 0\}$ at P gives rise to a sequence of local alternatives $P_{n,\rho}$ of P such that, in the Hilbert space of square root densities,

$$\sqrt{dP_{n,\rho}} = \bigl(1 + \tfrac{1}{2} s_n \rho\bigr) \sqrt{dP} + \mathrm{o}(s_n) \qquad \text{as } n \to \infty \tag{7.1}$$

where $s_n = 1/\sqrt{n}$. Constructions to achieve (7.1) are

$$dP_{n,\rho} = \Bigl(\tfrac{1}{2} s_n \rho + \sqrt{1 - \tfrac{1}{4} s_n^2 \,\|\rho\|^2}\;\Bigr)^2 dP \tag{7.2}$$

or

$$dP_{n,\rho} = (1 + s_n \rho)\, dP \qquad \text{if } \rho \in L_\infty \tag{7.3}$$

Let $G_0, G_1 \subset L_2 \cap \{\mathrm{E} = 0\}$ be any two disjoint sets of tangents. The observations $x_1, \ldots, x_n$ at sample size n are assumed independent identically distributed with distribution Q_n. For fixed $g = (g_0, g_1) \in G_0 \times G_1$, preliminary simple asymptotic hypotheses concerning Q_n are that, eventually,

$$H_{g_0} : Q_n = P_{n,g_0} \qquad K_{g_1} : Q_n = P_{n,g_1} \tag{7.4}$$

As in Section 6, asymptotic tests $\delta = (\delta_n)$, that is, sequences of tests δ_n at sample size n, are employed, and their error probabilities are evaluated under the n fold product measures Q_n^n.

Then the testing problem H_{g_0} vs. K_{g_1} at level $\alpha \in (0,1)$,

$$\liminf_{n\to\infty} \int \delta_n \, dP_{n,g_1}^n = \max! \tag{7.5}$$

subject to

$$\limsup_{n\to\infty} \int \delta_n \, dP_{n,g_0}^n \le \alpha \tag{7.6}$$

has the solution $\delta_g = (\delta_{n,g})$,

$$\delta_{n,g} = \mathbf{I}\Bigl(\frac{1}{\sqrt{n}} \sum_{i=1}^n g_{10}(x_i) > \|g_{10}\| u_\alpha + \langle g_{10} | g_0 \rangle\Bigr) \tag{7.7}$$

where $g = (g_0, g_1)$, $g_{10} = g_1 - g_0$, and u_α denotes the standard normal upper α point. Under H_{g_0}, δ_g achieves asymptotic size α and under K_{g_1}, asymptotic power $\Phi(-u_\alpha + \|g_{10}\|)$. The tests $\delta_{n,g}$ are unique up to terms tending to 0 in P^n probability. All these statements follow from the loglikelihood expansion (7.13) below and HR (1994; Corollary 3.4.2[2]).

Put $H_{G_0} = \cup\{H_{g_0} \mid g_0 \in G_0\}$ and $K_{G_1} = \cup\{K_{g_1} \mid g_1 \in G_1\}$.

7.2 The Maxmin Test Result

Then the maxmin testing problem H_{G_0} vs. K_{G_1} at level $\alpha \in (0,1)$ is

$$\inf_{g_1 \in G_1} \liminf_{n\to\infty} \int \delta_n \, dP_{n,g_1}^n = \max! \tag{7.8}$$

subject to

$$\sup_{g_0 \in G_0} \limsup_{n\to\infty} \int \delta_n \, dP_{n,g_0}^n \le \alpha \tag{7.9}$$

[2] Note that $\sigma > 0$ must be assumed in part (b).

Convex Closed Tangent Sets The tangent sets

$$G_0, G_1 \subset L_2 \cap \{\mathrm{E}=0\}\,, \qquad G_0 \cap G_1 = \emptyset \tag{7.10}$$

are each assumed convex and closed in L_2. The set of differences

$$G_{10} = G_1 - G_0 \tag{7.11}$$

which is again convex, may not be closed if $\dim L_2 > 2$, and therefore explicitly assumed to be also closed. Then denote by $q_{10} = q_1 - q_0$ the unique minimum norm element of G_{10}; as $G_0 \cap G_1 = \emptyset$, we have $q_{10} \neq 0$.

Theorem 7.1 *The asymptotic testing problem H_{G_0} vs. K_{G_1} at level α has a saddle point at $q = (q_0, q_1)$, and the maxmin asymptotic power achieved by δ_q equals $\Phi(-u_\alpha + \|q_{10}\|)$.*

PROOF For any tangent ρ, the following loglikelihood expansion holds,

$$\log \frac{dP^n_{n,\rho}}{dP^n} = s_n \sum\nolimits_i \rho(x_i) - \tfrac{1}{2}\|\rho\|^2 + \mathrm{o}_{P^n}(n^0) \tag{7.12}$$

Hence

$$\log \frac{dP^n_{n,g_1}}{dP^n_{n,g_0}} = s_n \sum\nolimits_i g_{10}(x_i) + \mathrm{const} + \mathrm{o}_{P^n}(n^0) \tag{7.13}$$

by mutual contiguity, for every $g = (g_0, g_1) \in G_0 \times G_1$ and $g_{10} = g_1 - g_0$. Therefore, the test sequence δ_g is indeed the optimum one at level α for H_{g_0} vs. K_{g_1}; confer HR (1994; Corollary 3.4.2).

Let us evaluate δ_q, for any $q = (q_0, q_1) \in G_0 \times G_1$ fixed, under other tangents $\rho \in G_0 \cup G_1$. In view of (7.12), by one of LeCam's lemmas, confer HR (1994; Corollary 2.2.6), the sequence of test statistics $s_n \sum_i q_{10}(x_i)$ are asymptotically normal under $P^n_{n,\rho}$,

$$s_n \sum\nolimits_i q_{10}(x_i) \xrightarrow{\mathrm{w}} \mathcal{N}\big(\langle q_{10}|\rho\rangle, \|q_{10}\|^2\big) \tag{7.14}$$

hence

$$\lim_{n\to\infty} \int \delta_{n,q}\, dP^n_{n,\rho} \;=\; \Phi\Big(-u_\alpha + \frac{\langle q_{10}|\rho - q_0\rangle}{\|q_{10}\|}\Big) \tag{7.15}$$

Therefore, the asymptotic size under $g_0 \in G_0$ becomes maximal at $g_0 = q_0$, and the asymptotic power under $g_1 \in G_1$ becomes minimal at $g_1 = q_1$, iff

$$\langle q_{10}|q_{10} - g_{10}\rangle \leq 0 \qquad \forall\, g_{10} \in G_{10} \tag{7.16}$$

By Lemma 3.2, this characterizes the minimum norm element q_{10} of G_{10}. ////

While $q_{10} = q_1 - q_0$ is unique, there may exist other least favorable pairs of tangents $g = (g_0, g_1)$ in $G_0 \times G_1$ achieving the same $g_{10} = g_1 - g_0 = q_{10}$ of minimum norm in $G_{10} = G_1 - G_0$. But then $\delta_g = \delta_q$, by the following corollary. So the maxmin asymptotic level α test for H_{G_0} vs. K_{G_1} is unique.

Corollary 7.2 *Let $g = (g_0, g_1)$ and $q = (q_0, q_1)$ be two least favorable tangent pairs in $G_0 \times G_1$. Then*

$$\langle q_{10}|g_0\rangle = \langle q_{10}|q_0\rangle\,, \qquad \langle q_{10}|g_1\rangle = \langle q_{10}|q_1\rangle \tag{7.17}$$

PROOF By the saddle point, δ_q achieves asymptotic size $\le \alpha$ under H_{g_0} and asymptotic power $\ge \Phi(-u_\alpha + \|q_{10}\|) = \Phi(-u_\alpha + \|g_{10}\|)$ under K_{g_1}. However, strict inequalities cannot hold since δ_g is optimal for H_{g_0} vs. K_{g_1}. Inserting $\rho = g_0, g_1$ in (7.15) and (7.16), (7.17) follows. Hence, in particular, $\delta_g = \delta_q$. ////

7.3 Robust Asymptotic Tests

In the setup of Section 2, with $P = P_{\theta_0}$, the normed robust tangent balls G_* are

$$G_h = \bigl\{\, g \in L_2 \bigm| \mathrm{E}\, g = 0\,,\ \mathrm{E}\, g^2 \le 8 \,\bigr\} \tag{7.18}$$

$$G_v = \bigl\{\, g \in L_2 \bigm| \mathrm{E}\, g = 0\,,\ \mathrm{E}\,|g| \le 2 \,\bigr\} \tag{7.19}$$

$$G_c = \bigl\{\, g \in L_2 \bigm| \mathrm{E}\, g = 0\,,\ g \ge -1 \,\bigr\} \tag{7.20}$$

Thus $\mathcal{G}_* = r G_*$ are the balls of radius r introduced in (2.5)–(2.7); $* = h, v, c$.

We assume parameter dimension $k = 1$. Invoke the scores function Λ for the parameter θ of the ideal model $\mathcal{P}$ at θ_0, and let numbers $r_0, r_1, \tau \in [0, \infty)$ be given. Then Theorem 7.1 is going to be applied to the tangent sets

$$G_{*,0} = r_0 G_*\,, \qquad G_{*,1} = \tau\Lambda + r_1 G_* \tag{7.21}$$

Remark 7.3 The minimum norm element $q_{*,10}$ of $G_{*,10} = G_{*,1} - G_{*,0}$, therefore, will be $\tau\Lambda$ minus its projection on the set of differences $r_0 G_* - r_1 G_*$. ////

Abbreviate the corresponding hypotheses by $H_* = H_{G_{*,0}}$ and $K_* = K_{G_{*,1}}$. As shown in the proof to Proposition 2.2, H_* and K_* represent the neighborhoods $U_*(\theta_0; s_n r_0)$ and $U_*(\theta_0 + s_n\tau; s_n r_1)$ about P_{θ_0} and $P_{\theta_0 + s_n\tau}$ of radii $s_n r_0$ and $s_n r_1$ respectively, up to some $\mathrm{o}(s_n)$ where $s_n = 1/\sqrt{n}$. Put $r = r_0 + r_1$.

Maxmin Tests for Hellinger Balls

Theorem 7.4 *Let*

$$8\, r^2 < \tau^2 \mathcal{I}\,, \quad \text{where } \mathcal{I} = \|\Lambda\|^2 \tag{7.22}$$

Then the least favorable tangent pair $q_h = (q_{h,0}, q_{h,1})$ in $G_{h,0} \times G_{h,1}$ is unique,

$$q_{h,0} = r_0 \gamma\Lambda\,, \quad q_{h,1} = \tau\Lambda - r_1\gamma\Lambda\,, \quad \text{where } \gamma = \sqrt{8}\,\|\Lambda\|^{-1} \tag{7.23}$$

The maxmin asymptotic level α test $\delta_{q_h} = (\delta_{n,q_h})$ for H_h vs. K_h is given by

$$\delta_{n,q_h} = \mathbf{I}\left(\frac{1}{\sqrt{n\mathcal{I}}} \sum_{i=1}^{n} \Lambda(x_i) > u_\alpha + \sqrt{8}\, r_0 \right) \tag{7.24}$$

and achieves maxmin asymptotic power $\Phi\bigl(-u_\alpha + \tau\|\Lambda\| - \sqrt{8}\, r\bigr)$.

PROOF Since G_h is symmetric convex, $G_{10} = \tau\Lambda + r_1 G_h - r_0 G_h = \tau\Lambda - r G_h$, and the minimum norm element q_{10} is supplied by $q_0 = r_0 \tilde{g}$, $q_1 = \tau\Lambda - r_1 \tilde{g}$, where $\tilde{g} \in G_h$ is the unique minimizer of $\|\tau\Lambda - r g\|$ among all $g \in G_h$.

The projection of $\tau\Lambda$ on $r G_h$ is determined by Theorem 3.3 and its proof, with Λ replaced by $\tau\Lambda$. Then condition (7.22) coincides with condition (3.5), and is equivalent to $r\tilde{g} \neq \tau\Lambda$, that is, $G_{h,0} \cap G_{h,1} = \emptyset$. Thus $\tilde{g} = \gamma\Lambda$.

With $q_{10} = (\tau - \gamma r)\Lambda$ and $\langle q_{10} | q_0 \rangle = \|q_{10}\| \sqrt{8}\, r_0$, Theorem 7.1 applies.

The pair (q_0, q_1) is unique: $q_0 = r_0 g_0$ and $q_1 = \tau\Lambda + r_1 g_1$ for arbitrary elements $g_0, g_1 \in G_h$ entails that $r\tilde{g} = r_0 g_0 - r_1 g_1$, and then $g_0 = -g_1 = \tilde{g}$ because $\|g_0\|, \|g_1\| \leq \sqrt{8} = \|\tilde{g}\|$ and the norm is strictly convex. ////

Thus the least favorable tangents—multiples of Λ—generate local alternatives within the parametric model $\mathcal{P}$, and the asymptotic maxmin test δ_{q_h} agrees with the asymptotic most powerful test for $(P^n_{\theta_0})$ vs. $(P^n_{\theta_0 + s_n \tau})$ at the smaller level $\Phi(-u_\alpha - \sqrt{8}\, r_0)$. The result compares with Theorem 4.2 and Remark 4.3.

Least Favorable Pairs of Probabilities For Hellinger balls, least favorable pairs of probabilities in the sense of Huber and Strassen (1973) do not exist; confer Birgé (1980).

For total variation and contamination neighborhoods, such Huber–Strassen pairs exist, While the least favorable pairs are not unique, their likelihood and its distribution under each of the two probabilities of least favorable pairs is unique; confer HR (1977). The Neyman–Pearson tests based on the likelihoods of the product measures of least favorable probability pairs furnish finite sample size, hence also asymptotic, maxmin tests.

The robust asymptotic tests derived from Huber–Strassen pairs have been evaluated by Huber–Carol (1970), HR (1978), Wang (1981), and Quang (1985).

Maxmin Tests for Total Variation Balls

Theorem 7.5 *Let*

$$2r < \tau \,\mathrm{E}\,|\Lambda| \tag{7.25}$$

(a) *Then a least favorable tangent pair* $q_v = (q_{v,0}, q_{v,1})$ *in* $G_{v,0} \times G_{v,1}$ *is given by*

$$q_{v,0} = r_0 \tilde{g}_v, \qquad q_{v,1} = \tau\Lambda - r_1 \tilde{g}_v \tag{7.26}$$

where

$$r\tilde{g}_v = \tau(\Lambda - v'')_+ - \tau(v' - \Lambda)_+ \tag{7.27}$$

with clipping constants $v' = v'(r/\tau) < 0 < v''(r/\tau) = v''$ *determined by*

$$\tau\,\mathrm{E}(v' - \Lambda)_+ = r = \tau\,\mathrm{E}(\Lambda - v'')_+ \tag{7.28}$$

Setting $\Lambda^{(v)} = v' \vee \Lambda \wedge v''$, *the maxmin asymptotic level* α *test* $\delta_{q_v} = (\delta_{n,q_v})$ *for* H_v *vs.* K_v *is given by*

$$\delta_{n,q_v} = \mathbf{I}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Lambda^{(v)}(x_i) > \|\Lambda^{(v)}\|\,u_\alpha + r_0\,(v'' - v')\right) \tag{7.29}$$

and achieves maxmin asymptotic power $\Phi\bigl(-u_\alpha + \tau\|\Lambda^{(v)}\|\bigr)$.

(b) *The test sequence* δ_{q_v} *coincides with the robust asymptotic test based on least favorable probability pairs for* $U_v\bigl(P_{\theta_0}; r_0/\sqrt{n}\bigr)$ *vs.* $U_v\bigl(P_{\theta_0+\tau/\sqrt{n}}; r_1/\sqrt{n}\bigr)$, *hence maximizes the asymptotic minimum power over* $U_v\bigl(P_{\theta_0+\tau/\sqrt{n}}; r_1/\sqrt{n}\bigr)$ *subject to asymptotic maximum size* $\le \alpha$ *over* $U_v\bigl(P_{\theta_0}; r_0/\sqrt{n}\bigr)$.

PROOF

(a) Also G_v is symmetric convex, so $G_{10} = \tau\Lambda + r_1 G_v - r_0 G_v = \tau\Lambda - r G_v$, and the minimum norm element q_{10} is supplied by $q_0 = r_0\,\tilde g$, $q_1 = \tau\Lambda - r_1\,\tilde g$, where $\tilde g \in G_v$ is the unique minimizer of $\|\tau\Lambda - r\,g\|$ among all $g \in G_v$.

The projection of $\tau\Lambda$ on rG_h is determined by Theorem 3.4, with $\tau\Lambda$ in the place of Λ. Then condition (7.25) coincides with condition (3.8), and is equivalent to $r\,\tilde g \ne \tau\Lambda$, that is, $G_{v,0} \cap G_{v,1} = \emptyset$. Thus $\tilde g$ is of form (7.27), (7.28).

With $q_{10} = \tau\Lambda^{(v)}$ and $\langle q_{10}|q_0\rangle = \tau r_0\,(v'' - v')$, Theorem 7.1 applies.

(b) We invoke the results of HR (1978), replacing $P_{-\tau_n}$ by P_0 in (2.8) there. This reduces 2τ to τ in that paper. Then the radius condition (2.6) of HR (1978): $r/\tau < \mathrm{E}\,\Lambda_+$, coincides with (7.25). Moreover, the clipping equations (3.9) of HR (1978) agree with (7.28), and then the function $\Lambda^{(v)}$ equals the function (3.10) of HR (1978).

Therefore, Theorems 3.4 and 4.1 of HR (1978) tell us that δ_{q_v} maximizes the asymptotic minimum power over $U_v(P_{\theta_0+s_n\tau}; s_n r_1)$ subject to asymptotic maximum size $\le \alpha$ over $U_v(P_{\theta_0}; s_n r_0)$. ////

Remark 7.6 Under condition (7.25), all least favorable pairs $g_v = (g_{v,0}, g_{v,1})$ of tangents in $G_{v,0} \times G_{v,1}$ are characterized by

$$g_{v,0} = r_0\,g_0\,, \qquad g_{v,1} = \tau\Lambda - r_1\,g_1 \tag{7.30}$$

where g_0 and g_1 may be any elements of G_v whose positive and negative parts make up those of $\tilde g_v$ given by (7.27) and (7.28) such that

$$r_0\,g_0^+ + r_1\,g_1^+ = \tau\,(\Lambda - v'')_+\,, \qquad r_0\,g_0^- + r_1\,g_1^- = \tau\,(v' - \Lambda)_+ \tag{7.31}$$

The least favorable tangent pair $q_v = (q_{v,0}, q_{v,1})$, which results from the special choice $g_0 = g_1 = \tilde g_v$, is not the only one in general. Other choices of g_0 and g_1 may be based on suitable partitions of the events $\{\Lambda > v''\}$ and $\{\Lambda < v'\}$.

For testing $U_v\bigl(P_{\theta_0}; r_0/\sqrt{n}\bigr)$ vs. $U_v\bigl(P_{\theta_0+\tau/\sqrt{n}}; r_1/\sqrt{n}\bigr)$, all least favorable pairs of probabilities have been characterized by HR (1977; Theorem 5.2). ////

Maxmin Tests for Contamination Neighborhoods

Theorem 7.7 *Let*

$$r_0 < \mathrm{E}\big(\tau\Lambda - (r_1 - r_0)\big)_+ \tag{7.32}$$

(a) *Then the least favorable tangent pair* $q_c = (q_{c,0}, q_{c,1})$ *in* $G_{c,0} \times G_{c,1}$ *is unique,*

$$q_{c,0} = \tau\,(\Lambda - c'')_+ - r_0\,, \qquad q_{c,1} = \begin{cases} \tau\Lambda + \tau\,(c' - \Lambda)_+ - r_1 \\ \tau\,(\Lambda \vee c') - r_1 \end{cases} \tag{7.33}$$

with clipping constants $c' = c'(r_1/\tau) < z < c''(r_0/\tau) = c''$ *determined by*

$$\tau\,\mathrm{E}(c' - \Lambda)_+ = r_1\,, \qquad \tau\,\mathrm{E}(\Lambda - c'')_+ = r_0 \tag{7.34}$$

where $z = (r_1 - r_0)/\tau$. *Setting* $\Lambda^{(c)} = c' \vee \Lambda \wedge c'' - z$, *the maxmin asymptotic level* α *test* $\delta_{q_c} = (\delta_{n,q_c})$ *for* H_c *vs.* K_c *is given by*

$$\delta_{n,q_c} = \mathbf{I}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Lambda^{(c)}(x_i) > \|\Lambda^{(c)}\|\,u_\alpha + r_0\,(c'' - z)\right) \tag{7.35}$$

and achieves maxmin asymptotic power $\Phi\big(-u_\alpha + \tau\|\Lambda^{(c)}\|\big)$.

(b) *The test sequence* δ_{q_c} *coincides with the robust asymptotic test based on least favorable probability pairs for* $U_c\big(P_{\theta_0}; r_0/\sqrt{n}\big)$ *vs.* $U_c\big(P_{\theta_0+\tau/\sqrt{n}}; r_1/\sqrt{n}\big)$, *hence maximizes the asymptotic minimum power over* $U_c\big(P_{\theta_0+\tau/\sqrt{n}}; r_1/\sqrt{n}\big)$ *subject to asymptotic maximum size* $\le \alpha$ *over* $U_c\big(P_{\theta_0}; r_0/\sqrt{n}\big)$.

PROOF

(a) We can show that $G_{10} = \tau\Lambda + r_1 G_c - r_0 G_c$ equals the closed set

$$\tau\Lambda - (r_1 - r_0) + \big\{\, g \in L_2 \;\big|\; \mathrm{E}\,g = r_1 - r_0\,,\ \mathrm{E}\,g^+ \le r_1\,,\ \mathrm{E}\,g^- \le r_0 \,\big\} \tag{7.36}$$

As $\mathrm{E}(\tau\Lambda - c)_+ = \mathrm{E}(c - \tau\Lambda)_+ - c$, radius condition (7.32) is equivalent to

$$r_1 < \mathrm{E}\big((r_1 - r_0) - \tau\Lambda\big)_+ \tag{7.37}$$

If (7.32) and (7.37) are violated, the zero function is in G_{10} as

$$0 = \tau\Lambda - (r_1 - r_0) + \big((r_1 - r_0) - \tau\Lambda\big)_+ - \big(\tau\Lambda - (r_1 - r_0)\big)_+$$

Under conditions (7.32) and (7.37), equivalently $c' < z = (r_1 - r_0)/\tau < c''$ for the solutions c' and c'' to (7.34), the function $q_{10} = q_{c,1} - q_{c,0}$ is nonzero,

$$\begin{aligned} q_{10} &= \tau\Lambda - (r_1 - r_0) + \tau\,(c' - \Lambda)_+ - \tau\,(\Lambda - c'')_+ \\ &= \tau\,(c' \vee \Lambda \wedge c'') - (r_1 - r_0) = \tau\Lambda^{(c)} \end{aligned} \tag{7.38}$$

and, by Lemma 3.2, the minimum norm element of G_{10}. In fact, for all $g_0 \in G_c$,

$$\begin{aligned} \langle \Lambda^{(c)} | r_0\, g_0 - q_{c,0} \rangle &= \bigl\langle c' \vee \Lambda \wedge c'' \bigm| r_0\,(1+g_0) - \tau\,(\Lambda - c'')_+ \bigr\rangle \\ &\le c'' r_0\, \mathrm{E}(1+g_0) - c''\tau\, \mathrm{E}(\Lambda - c'')_+ = 0 \end{aligned} \tag{7.39}$$

as $c' \vee \Lambda \wedge c'' \le c''$ and $1+g_0 \ge 0$, and by (7.34). Likewise, for all $g_1 \in G_c$,

$$\begin{aligned} \langle \Lambda^{(c)} | q_{c,1} - \tau\Lambda - r_1\, g_1 \rangle &= \bigl\langle c' \vee \Lambda \wedge c'' \bigm| \tau\,(c' - \Lambda)_+ - r_1\,(1+g_1) \bigr\rangle \\ &\le c'\tau\, \mathrm{E}(c' - \Lambda)_+ - c' r_1\, \mathrm{E}(1+g_1) = 0 \end{aligned} \tag{7.40}$$

With $\langle \Lambda^{(c)} | q_{c,0} \rangle = r_0\,(c'' - z)$, Theorem 7.1 applies.

Now let $(r_0\, g_0, \tau\Lambda + r_1\, g_1)$ be any least favorable tangent pair, that is, with elements $g_0, g_1 \in G_c$ such that $\tau\Lambda + r_1\, g_1 - r_0\, g_0 = q_{10}$. Then, in view of (7.38),

$$r_1\,(1+g_1) - r_0\,(1+g_0) = \tau\,(c' - \Lambda)_+ - \tau\,(\Lambda - c'')_+ \tag{7.41}$$

The RHS, since $c' < c''$, is a decomposition into positive and negative parts. As also $1+g_1 \ge 0$ and $1+g_0 \ge 0$ this implies that

$$\tau\,(c' - \Lambda)_+ \le r_1\,(1+g_1)\,, \qquad \tau\,(\Lambda - c'')_+ \le r_0\,(1+g_0) \tag{7.42}$$

But by (7.34), the functions compared have the same expectations. Hence strict inequalities cannot hold. It follows that

$$r_0\, g_0 = \tau\,(\Lambda - c'')_+ - r_0\,, \qquad r_1\, g_1 = \tau\,(c' - \Lambda)_+ - r_1 \tag{7.43}$$

which proves uniqueness of the least favorable tangent pair $q_c = (q_{c,0}, q_{c,1})$.

(b) The substitution of $P_{-\tau_n}$ by P_0 in HR (1978) reduces 2τ to τ there. Then the radius condition (2.6) of HR (1978) is (7.32). Moreover, the clipping equations (3.9) of HR (1978) agree with (7.34), and the present function $\Lambda^{(c)}$ equals the function defined by (3.10) in HR (1978).

Therefore, Theorems 3.4 and 4.1 of HR (1978) tell us that δ_{q_c} maximizes the asymptotic minimum power over $U_c(P_{\theta_0 + s_n \tau}; s_n r_1)$ subject to asymptotic maximum size $\le \alpha$ over $U_c(P_{\theta_0}; s_n r_0)$. ////

Remark 7.8 The radius condition (7.32), being equivalent to $\tau c'' > r_1 - r_0$ for c'' satisfying (7.34), is stronger than $\tau c'' > -r_0$. In turn, $\tau c'' > -r_0$ for c'' satisfying (7.34), can be shown to be equivalent to $r_0 < -\tau \inf_P \Lambda$.

Under this radius condition (3.11): $r_0 < -\tau \inf_P \Lambda$, Theorem 3.5 (with $\tau\Lambda$ in the place of Λ) yields the element $\tilde g_0$ of G_c minimizing $\|\tau\Lambda - r_0\, g\|$ among all $g \in G_c$:

$$r_0\, \tilde g_0 = \tau\Lambda - (\tau\Lambda + r_0) \wedge u = \tau\,(\Lambda - c'')_+ - r_0 \tag{7.44}$$

with u and $\tau c'' = u - r_0$ determined by $\mathrm{E}\, \tilde g_0 = 0$. Thus, $q_{c,0} = r_0\, \tilde g_0$.

Likewise, the radius condition (7.37), being equivalent to $\tau c' < r_1 - r_0$ for c' satisfying (7.34), implies that $\tau c' < r_1$, equivalently $r_1 < \tau \sup_P \Lambda$.

Under this radius condition (3.11): $r_1 < \tau \sup_P \Lambda$, Theorem 3.5 (with $-\tau\Lambda$ in the place of Λ) yields the element $\tilde g_1$ of G_c minimizing $\|\tau\Lambda + r_1 g\|$ among all $g \in G_c$. And then it may again be verified that $q_{c,1} = \tau\Lambda + r_1 \tilde g_1$.

Therefore, according to Lemma 3.2, it follows that, for all $g_0, g_1 \in G_c$,

$$\langle \tau\Lambda - r_0 \tilde g_0 | r_0 g_0 - r_0 \tilde g_0 \rangle \le 0, \qquad \langle \tau\Lambda + r_1 \tilde g_1 | r_1 \tilde g_1 - r_1 g_1 \rangle \le 0 \tag{7.45}$$

But the bounds (7.39) and (7.40) established in the preceeding proof tell us that this remains true for $\tau\Lambda^{(c)} = \tau\Lambda + r_1 \tilde g_1 - r_0 \tilde g_0$ in the place of $\tau\Lambda - r_0 \tilde g_0$, respectively of $\tau\Lambda + r_1 \tilde g_1$. This is remarkable since the two additional terms are always nonnegative,

$$\begin{aligned} \langle r_1 \tilde g_1 | r_0 g_0 - r_0 \tilde g_0 \rangle &= \big\langle \tau (c' - \Lambda)_+ - r_1 \big| r_0 g_0 + r_0 - \tau (\Lambda - c'')_+ \big\rangle \\ &= \tau r_0 \, \langle (c' - \Lambda)_+ | 1 + g_0 \rangle \ge 0 && (7.46) \\ \langle r_0 \tilde g_0 | r_1 g_1 - r_1 \tilde g_1 \rangle &= \big\langle \tau (\Lambda - c'')_+ - r_0 \big| r_1 g_1 + r_1 - \tau (c' - \Lambda)_+ \big\rangle \\ &= \tau r_1 \, \langle (\Lambda - c'')_+ | 1 + g_1 \rangle \ge 0 && (7.47) \end{aligned}$$

where use has been made of $c' < c''$, which is garanteed by the stronger radius condition (7.32), (7.37). ////

Remark 7.9 For testing $U_c\big(P_{\theta_0}; r_0/\sqrt{n}\big)$ vs. $U_c\big(P_{\theta_0+\tau/\sqrt{n}}; r_1/\sqrt{n}\big)$, all least favorable pairs of probabilities have been characterized in terms of their densities by HR (1977; Theorem 5.2). The uniqueness of the least favorable tangent pair $q_c = (q_{c,0}, q_{c,1})$ gives rise to the conjecture that, contrary to the total variation case,

$$\lim_{n\to\infty} \sqrt{n}\, d_h(Q''_{n,j}, Q'_{n,j}) = 0, \qquad j = 0, 1 \tag{7.48}$$

if $(Q'_{n,0}, Q'_{n,1})$ and $(Q''_{n,0}, Q''_{n,1})$ are any two, possibly different, least favorable probability pairs for $U_c\big(P_{\theta_0}; r_0/\sqrt{n}\big)$ vs. $U_c\big(P_{\theta_0+\tau/\sqrt{n}}; r_1/\sqrt{n}\big)$. ////

Remark 7.10 For shrinking contamination neighborhoods of a one-parameter family involving a finite dimensional nuisance parameter, the robust asymptotic tests based on least favorable pairs were investigated by Wang (1981). It would be interesting to derive the maxmin asymptotic tests by projection arguments, and also for total variation balls. ////

Acknowledgement A first version of the paper was written during a visit to Sonderforschungsbereich 373 ("Quantifikation und Simulation ökonomischer Prozesse") at Humboldt–Universität zu Berlin. Financial support by Deutsche Forschungsgemeinschaft is gratefully acknowledged. I thank Wolfgang Härdle for his hospitality. I am also indebted to Peter Ruckdeschel for his assistance and the plot of Figure 1.

References

[1] Beran, R.J. (1974): Asymptotically efficient and adaptive rank estimates in location models. Ann. Statist. **2** 63–74.

[2] Bickel, P.J. (1981): Quelques aspects de la statistique robuste. In *Ecole d'Eté de Probabilités de Saint Flour IX 1979* (P.L. Hennequin, ed.), 1–72. Lecture Notes in Mathematics #876. Springer, Berlin.

[3] Bickel, P.J. (1982): On adaptive estimation. Ann. Statist. **10** 647–671.

[4] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.

[5] Birgé, L. (1980): Approximation dans les espaces métrique et théorie de l'estimation. Ph.D. Thesis, University of Paris.

[6] Huber, P.J. (1964): Robust estimation of a location parameter. Ann. Math. Statist. **35** 73–101.

[7] Huber, P.J. and Strassen, V.(1973): Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263.

[8] Huber–Carol, C. (1970): *Étude asymptotique de tests robustes*. Thèse de Doctorat, ETH Zürich.

[9] Kakiuchi, I. and Kimura, M. (2000): Robust rank tests for k-sample approximate equality in the presence of gross errors. *J. Statist. Plann. Inf.* **82**, to appear.

[10] Klaassen, C.A.J. (1987): Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15** 1548–1562.

[11] Pfanzagl, J. and Wefelmeyer, W. (1982): *Contributions to a General Asymptotic Statistical Theory*. Lecture Notes in Statistics #13. Springer, Berlin.

[12] Quang, P.X. (1985): Robust sequential testing. *Ann. Statist.* **13** 638–649.

[13] Rieder, H. (1977): Least favorable pairs for special capacities. *Ann. Statist.* **5** 909–921.

[14] Rieder, H. (1978): A robust asymptotic testing model. *Ann. Statist.* **6** 1080–1094.

[15] Rieder, H. (1980): Estimates derived from robust tests. *Ann. Statist.* **8** 106–115.

[16] Rieder, H. (1981 a): Robustness of one- and two-sample rank tests against gross errors. Ann. Statist. **9** 245–265.

[17] Rieder, H. (1981 b): On local asymptotic minimaxity and admissibility in robust estimation. *Ann. Statist.* **9** 266–277.

[18] Rieder, H. (1994): *Robust Asymptotic Statistics*. Springer, New York.

[19] Rieder, H. (2000): One-sided confidence about functionals over tangent cones. Submitted for publication.

[20] Shen, L.Z. (1995): On optimal B-robust influence functions in semiparametric models. Ann. Statist. **23** 968–989.

[21] Schick, A. (1986): On asymptotically efficient estimation in semiparametric models. Ann. Statist. **14** 1139–1151.

[22] Stone, C. (1975): Adaptive maximum likelihood estimation for a location parameter. Ann. Statist. **3** 267–284.

[23] van der Vaart, A.W. (1998): *Asymptotic Statistics*. CUP, Cambridge.

[24] Wang, P.C.C. (1981): Robust asymptotic tests of statistical hypotheses involving nuisance parameters. *Ann. Statist.* **9** 1096–1106.

Department of Mathematics
University of Bayreuth, NW II
D-95440 Bayreuth, Germany
e-mail: helmut.rieder@uni-bayreuth.de