Rieder, Helmut

Working Paper

Neighborhoods as nuisance parameters? Robustness vs. semiparametrics

SFB 373 Discussion Paper, No. 1999,75

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Rieder, Helmut (1999) : Neighborhoods as nuisance parameters? Robustness vs. semiparametrics, SFB 373 Discussion Paper, No. 1999,75, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10046636

This Version is available at:
https://hdl.handle.net/10419/61698

Neighborhoods as Nuisance Parameters? Robustness vs. Semiparametrics

Helmut Rieder *
University of Bayreuth, Germany

24 September 1999

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, or fails, to solve the robust MSE problem for gross error models.

Key Words and Phrases: Hellinger, total variation, and contamination neighborhoods; 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 minimax and convolution theorems.
AMS 1991 classification: 62F35, 62G35.

* This paper was written during a visit to Sonderforschungsbereich 373 ("Quantifikation und Simulation ökonomischer Prozesse") at Humboldt–Universität zu Berlin. Financial support from Deutsche Forschungsgemeinschaft is gratefully acknowledged. I thank Wolfgang Härdle for his hospitality.

1 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 set $\Theta \subset \mathbb{R}^k$, whereas ν acts as nuisance parameter, which, for each θ, ranges over some set H_θ; typically, subsets of some infinite dimensional function spaces. The observations $x_1, \ldots, x_n \sim Q_{\theta,\nu}$ are assumed to be independent identically distributed. Estimators of θ may be any functions $S_n : \Omega^n \to \mathbb{R}^k$ which are product-$\mathcal{B}^n$, Borel-$\mathbb{B}^k$ measurable. 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 $Q_{\theta_0,\nu_0} = Q$, and denote expectation and covariance under Q by E and $\mathcal{C}$, respectively. Also the spaces L_2, L_2^k, and L_∞ of square integrable real-, $\mathbb{R}^k$-valued, and essentially bounded functions, respectively, refer to $Q = Q_{\theta_0,\nu_0}$.

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 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 is some path $t \mapsto \nu_t \in H_{\theta_0 + ta}$ so that, as $t \to 0$ in $\mathbb{R}$,

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

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 have expectation zero. The covariance $\mathcal{I} = \mathcal{C}(\Lambda)$ is assumed of full rank k.

As for complete technical details, maybe in slightly different notations, the reader please consult standard textbooks on asymptotic statistics such as Bickel et al. (1993; chapters 2–3), Rieder[1] (1994; chapters 2–4), van der Vaart (1998; chapter 25). This recommendation also holds for the following notions, basic properties and results, to be summarized in this section.

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

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

[1] HR, subsequently

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

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

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 $Q_n(a,g) = Q_{\theta_0+t_n a,\nu_{t_n}}$ for $t_n = 1/\sqrt{n}$, their distributions under $Q_n^n(a,g)$ converge weakly as $n \to \infty$, for every $a \in \mathbb{R}^k$ and $g \in \partial_2\mathcal{Q}$,

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

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

In view of (1.3), it is easy to see that the projection $\Pi(\psi)$ on $(\operatorname{clin}\partial\mathcal{Q})^k$ must be the same for all $\psi \in \Psi$—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 $\mathcal{J} = \mathcal{C}\big(\Lambda - \Pi_2(\Lambda)\big)$ is called the Fisher information of model $\mathcal{Q}$ for the parameter θ at (θ_0, ν_0). Nonsingularity of this covariance $\mathcal{J}$ is equivalent to the existence of influence curves (that is, $\Psi \neq \emptyset$, which we want to assume).

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,

$$\mathcal{Q}_{\nu_0} = \{\, Q_{\theta,\nu_0} \mid \theta \in \Theta \,\} \tag{1.8}$$

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 and Fisher information of $\mathcal{Q}_{\nu_0}$ at θ_0 are given by, respectively,

$$\hat{\varrho} = \mathcal{I}^{-1}\Lambda, \qquad \mathcal{I} = \mathcal{C}(\Lambda) \tag{1.9}$$

The following bound of $\mathcal{J}$ by $\mathcal{I}$ holds in the positive definite sense,

$$\mathcal{C}(\hat{\varrho}) = \mathcal{I}^{-1} \leq \mathcal{J}^{-1} = \mathcal{C}(\varrho) \tag{1.10}$$

with equality iff $\varrho = \hat{\varrho}$, which is the case iff $\Pi_2(\Lambda) = 0$. This is the case of *adaptivity*. However, adaptation must still be achieved by some suitable estimator construction. ////

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

$$a'\Lambda \notin \underline{c}\mathrm{lin}(\partial_2 \mathcal{Q} + \text{constants}) \qquad \forall\, a \in \mathbb{R}^k,\, a \neq 0 \tag{1.11}$$

where $\underline{c}$lin denotes the closed linear span in L_1; see Shen (1995; Theorem 1). In the finite dimensional case, where $\partial_2 \mathcal{Q} = \{\, b'\Delta \mid b \in \mathbb{R}^\ell \,\}$ with the scores function $\Delta \in L_2^\ell$ for ν at (θ_0, ν_0), condition (1.11) is equivalent to full rank $k + \ell$ of the total Fisher information matrix $\mathcal{I} = \mathcal{C}\,(\Lambda', \Delta')'$. In this case, explicit constructions have been provided by HR (1994), Remark 4.2.11 and 5.5(8), 5.5(9), setting $D = (\mathbb{I}_k, 0_{k\times\ell})$ there. ////

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,

$$\mathcal{C}(\psi) \geq \mathcal{J}^{-1} = \mathcal{C}(\varrho) \qquad \forall\, \psi \in \Psi \tag{1.12}$$

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 1.3 [nonlinear projection] These theorems require some structure of the tangent set $\partial\mathcal{Q}$, to be a linear space or at least a convex cone. Anyway, the projection is generally that on the closed linear space $\mathrm{clin}\,\partial\mathcal{Q}$.

One exception is Theorem 9.2.2 of Pfanzagl and Wefelmeyer (1982) about asymptotically median unbiased estimators, in terms of the projection on a closed convex cone $\partial\mathcal{Q}$. No projection of influence curves on any tangent set appears in Theorem 4.1(A) of HR (1981b), which provides another nonstandard asymptotic minimax bound. Both results use the Neyman–Pearson lemma (classical, respectively for capacities—infinitesimal gross error neighborhoods), some pseudo loss functions (confidence probabilities), and are restricted to the estimation of a real valued functional, respectively one real parameter. ////

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 without 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, a neighborhood model $\mathcal{Q}$ is obtained,

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

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 so already for members of the ideal model $Q = P_\zeta \in U(\theta;r)$ with ζ close to θ (if, as usual, 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,

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

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

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

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

Let us fix $\theta_0 \in \Theta$ and $\nu_0 = 0$, and write P for the previous $Q = Q_{\theta_0,\nu_0} = P_{\theta_0}$. 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 $\partial_2 \mathcal{Q}_*$,

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

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

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

which first appeared explicitly in Bickel (1981). Note that $\mathcal{G}_c \subset \mathcal{G}_v \subset \mathcal{G}_c - \mathcal{G}_c$. In the following, the scores function Λ is that of the ideal model $\mathcal{P}$, for θ at θ_0.

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

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

PROOF Invoke bounded approximations $\Lambda^{(t)}$ of Λ such that $\mathrm{E}\,\Lambda^{(t)} = 0$ and, as $t \to 0$, $\sup|\Lambda^{(t)}| = \mathrm{o}(1/t)$ 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 $\nu_t = tg$ in defining measures Q_t 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 are members of the neighborhoods $U_*(\theta_0 + ta; tr)$, at least 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_{θ_0+ta}, such that $d_v(\widetilde{P}_{\theta_0+ta}, P_{\theta_0+ta}) = \mathrm{o}(t)$ and

$$Q_t \in \widetilde{U}_c(\theta_0 + ta; tr) \tag{2.11}$$

for the tr-contamination ball $\widetilde{U}_c(\theta_0 + ta; tr)$ around $\widetilde{P}_{\theta_0+ta}$; namely, $\widetilde{P}_{\theta_0+ta}$ with P-density $1 + t_r a'\Lambda^{(t)}$, $t_r = t/(1 - tr)$.

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). ////

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, $\Lambda - \Pi_2(\Lambda) = 0$ in (1.7); in particular, adaptivity fails drastically.

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,

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

with scaling matrix

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

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). Now $\mathcal{K} \neq \mathcal{C}\big(\Lambda - \widetilde{\Pi}_2(\Lambda)\big)$, 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 only 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, $\mathcal{I} = \mathcal{C}(\Lambda) = (\mathcal{I}_{i,j})$ 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*

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

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 $g \in \mathcal{G}_h$,

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

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

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

$$2\,r < \min_{j=1,\ldots,k} \mathrm{E}\,|\Lambda_j| \tag{3.8}$$

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

$$\Lambda_j^{(r)} = \gamma_j' \vee \Lambda_j \wedge \gamma_j'' \tag{3.9}$$

where the clipping constants $\gamma_j' < 0 < \gamma_j''$ *are uniquely determined by*

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

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

In case $k = 1$, in order to minimize $\mathrm{E}(\Lambda - g)^2$ for $g \in \mathcal{G}_v$, we set up a Lagrangian $\mathrm{E}\big((\Lambda - g)^2 + 2\alpha g + 2\beta|g|\big)$ 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 $\tilde{q} = \gamma' \vee \Lambda \wedge \gamma''$ by means of the unique solutions $\gamma' < 0 < \gamma''$ of $\mathrm{E}(\gamma' - \Lambda)_+ = r = \mathrm{E}(\Lambda - \gamma'')_+$, which is a matter of continuity (dominated convergence theorem), monotony (strict), and the intermediate value theorem. We shall verify that this $\tilde{q}$ minimzes $\mathrm{E}\,q^2$ subject to $\mathrm{E}\,q = 0$, $\mathrm{E}\,|\Lambda - q| \le 2r$.

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

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

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

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

$$\Lambda_j^{(r)} = (\Lambda_j + r) \wedge \alpha_j \tag{3.12}$$

with clipping constant $\alpha_j > 0$ *uniquely determined by*

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

In (3.11), $\inf_P$ denotes the P-essential infimum.

PROOF Obviously, $\Lambda_j - \widetilde{\pi}_2(\Lambda_j) = 0$ iff $\Lambda_j \ge -r$ 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 $\mathrm{E}\,q = 0$, $q \le \Lambda + r$, for which we minimize a Lagrangian $\mathrm{E}\,q^2 - 2\alpha\,\mathrm{E}\,q = \mathrm{E}(q - \alpha)^2 + \text{constant}$, subject to $q \le \Lambda + r$. Doing this pointwise, the necessary form seems $\tilde{q} = (\Lambda + r) \wedge \alpha$.

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 $\alpha > 0$, which we use to define $\tilde{q} = (\Lambda + r) \wedge \alpha$. (Only in case $r = 0$, may α 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 $\tilde{q} < \alpha$. Thus $(\alpha - \tilde{q})(\alpha - q)$ is always greater or equal to $(\alpha - \tilde{q})^2$. Consequentially, $\mathrm{E}(-\tilde{q})(q - \tilde{q}) = \mathrm{E}(\alpha - \tilde{q})(q - \alpha + \alpha - \tilde{q}) \leq 0$; which is (3.4).////

Remark 3.6 In Theorems 3.4 and 3.5, condition (3.8), respectively (3.11), ensures that $\mathrm{E}\, \Lambda_j^{(r)} \Lambda_j > 0$. This may be seen by rewriting $\mathrm{E}\, \Lambda_j^{(r)} \Lambda_j$ as $\mathrm{E}\, |\Lambda_j^{(r)}|^2$ plus $r(\gamma_j'' - \gamma_j')$, where $r(\gamma_j'' - \gamma_j') > 0$ unless $r = 0$, respectively as $\mathrm{E}\, |\Lambda_j^{(r)}|^2$ plus $\mathrm{E}\, \Lambda_j^{(r)} (\Lambda_j + r - \Lambda_j^{(r)})$, where $\Lambda_j + r \leq \alpha_j$ a.e. P only if $r = 0$.

Whether condition (3.8), respectively (3.11), for dimension $k > 1$ already implies that $\det \mathcal{K} \neq 0$, hence existence of $\tilde{\varrho}_v$, respectively of $\tilde{\varrho}_c$, is unclear.////

4 The Robust Influence Curve in Comparison

We shall prove, respectively disprove, the semiparametric recipe (3.1)–(3.2) by comparison of results. How does the semiparametric influence curve $\tilde{\varrho}_*$ compare with the *robust* influence curve η_* that, by definition, minimizes asymptotic maximum 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—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

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

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 (1981b) 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, $\omega_h^2(\psi) = 8\,\mathrm{maxev}\,\mathcal{C}(\psi)$. In view of the Cramér–Rao bound (1.12), therefore, Hellinger risk $\mathrm{MSE}_h(.\,;\beta,r)$ is minimized by the canonical IC (1.9): $\hat{\varrho} = \mathcal{I}^{-1}\Lambda$, for every $\beta, r \in [\,0,\infty)$. Theorem 3.3 thus yields the following coincidence.

Theorem 4.2 *Under condition* (3.5): $8\,r^2 < \min_{j=1,\ldots,k} \mathcal{I}_{j,j}$ *of its existence, the semiparametric IC* $\tilde{\varrho}_h$ *is 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* $\beta \in [\,0,\infty)$.

The coincidence, in principle, 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) are more suitable for robustness.

Remark 4.3 Identity (4.3) implies equality in (1.10), with the robust IC η_h 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 $\mathrm{MSE}_h(\eta_h;\beta,r) = \mathrm{tr}\,\mathcal{I}^{-1} + 8\beta r^2\,\mathrm{maxev}\,\mathcal{I}^{-1}$ is clearly larger than $\mathrm{tr}\,\mathcal{I}^{-1} = \mathrm{MSE}_h(\hat{\varrho};\beta,0)$ (if only $\beta r > 0$). Despite of $\eta_h = \hat{\varrho}$ attaining both sides, in models $\mathcal{Q}_h$ and $\mathcal{P}$, respectively, strict inequality holds; that is, Hellinger neighborhoods do not go for free. ////

4.2 Relations for Total Variation

Case $k = 1$ 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,

$$\eta_v = c' \vee A\Lambda \wedge c'' \tag{4.4}$$

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

$$\beta r^2(c'' - c') = \mathrm{E}(c' - A\Lambda)_+ \tag{4.5}$$

Theorem 4.4 *Assume that*

$$0 < r < \mathrm{E}\,\Lambda_+ \tag{4.6}$$

Then the semiparametric IC $\tilde{\varrho}_v$ is the robust IC η_v minimzing $\mathrm{MSE}_v(.\,;\beta,r)$, *iff the bias weight $\beta = \beta(r)$ is chosen such that*

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

where $\gamma' = \gamma'(r) < 0 < \gamma''(r) = \gamma''$ are the solutions to (3.10), *that is,*

$$\mathrm{E}(\gamma' - \Lambda)_+ = r = \mathrm{E}(\Lambda - \gamma'')_+$$

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

Thus $\tilde{\varrho}_v$ attains form (4.4) with $c' = \gamma' A$ and $c'' = \gamma'' A$; in particular, $\beta r^2(c'' - c') = \beta r^2(\gamma'' - \gamma')A$. Since $Ar = A\,\mathrm{E}(\gamma' - \Lambda)_+ = \mathrm{E}(c' - A\Lambda)_+$ by (3.10), condition (4.5) is the same as (4.7). ////

This result justifies the semiparametric recipe if one accepts bias weight (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

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

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

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

since $\tilde{\varrho}_v$ is of form (4.4) and (4.5), 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 $\beta_{\min} = 4.8662$ at $r_{\min} = 0.1668$. In particular, no radius r_1 for which $\beta(r_1) = 1$ exists.

The radius $R(r)$ rises from 0 towards ∞ at $1/\sqrt{2\pi}$. Since $R(r)/r = \sqrt{\beta(r)}$ is always larger than $\sqrt{\beta_{\min}}$, 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 $\beta(r) > \beta_{\min}$, is typically even more pessimistic. ////

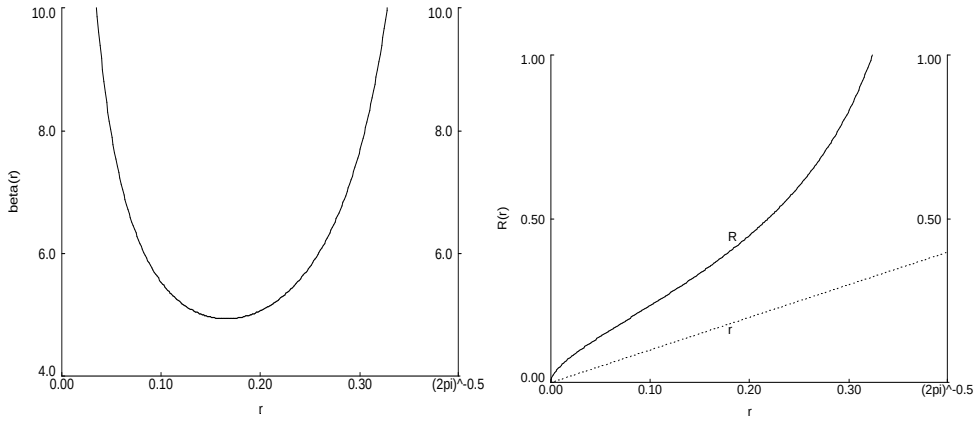

Figure 1: Bias weight $\beta(r)$ and radius $R(r)$ versus radius $0 < r < 1/\sqrt{2\pi}$, for total variation neighborhoods $U_v(\theta, r/\sqrt{n})$ about the ideal location model $P_\theta = \mathcal{N}(\theta, 1)$.

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* (4.6): $0 < r < \mathrm{E}\,\Lambda_+$. *Then the semiparametric IC* $\tilde{\varrho}_v$ *is the robust IC* η_v *iff half-width*

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

is employed in the confidence risk (4.10).

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)—but 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 $r > 0$ is the same as (4.6). ////

Case $k > 1$ 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, in case $s = 2$,

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

whereas in case $s = \infty$, all differences $c_j'' - c_j'$ are 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 $\gamma_j' < 0 < \gamma_j''$ defined by (3.10), and $(A^{j,i})^{-1} = \mathcal{K}$ given by (3.2), the semiparametric IC $\tilde{\varrho}_v$ has the coordinates

$$\tilde{\varrho}_j = A^{j,1}\,\gamma_1' \vee \Lambda_1 \wedge \gamma_1'' + \cdots + A^{j,k}\,\gamma_k' \vee \Lambda_k \wedge \gamma_k'' \tag{4.16}$$

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,

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

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

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

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

According to Theorem 3.5 on the other hand, the semiparametric IC $\tilde{\varrho}_c$ has the coordinates

$$\tilde{\varrho}_j = A^{j,1}\,(\Lambda_1 + r) \wedge \alpha_1 + \cdots + A^{j,k}\,(\Lambda_k + r) \wedge \alpha_k \tag{4.19}$$

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

Since in robustness respects, contamination and total variation have always turned out very similar, it is surprising that the semiparametric recipe (3.1) and (3.2) may give reasonable results for one model but not the other. The intuitive convex combinations, however, have been used in robust statistics from the start, prior to any other type of neighborhoods.

5 Unresolved: Robust Adaptation

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 $Q_{\theta,\nu}^n$, 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 formulated in HR (1981 a; section 3). 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 estimators $\sqrt{n}\,(S_n - \theta)$ 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 $Q_{\theta,\nu}$.

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 $|\psi| \leq \sup|\eta_c|$. 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 with respect to a sub-σ-algebra of $\mathcal{B}$).

The construction problem has not been solved. Towards such a robust adaptation, consistency of kernel density estimators has to be investigated over shrinking neighborhoods $U_*(\theta, \nu; r/\sqrt{n})$, as estimators of the density ν and scores $\Lambda_\nu = -\dot{\nu}/\nu$ belonging to the ideal center measure $Q_{\theta,\nu}$. For these purposes, the quantities ν and Λ_ν need not be estimated over the full range but only where $|\Lambda_\nu| \leq$ some bound.

Acknowledgement I thank Peter Ruckdeschel for plotting 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., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.

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

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

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

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

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

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

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

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

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

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