Connections between Semiparametrics and Robustness

Helmut Rieder

University of Bayreuth, Germany

Tsinghua University, Beijing, 23 and 28 May 2013

Topics

1) Robust influence curves for models with ∞-dim. nuisance parameter; e.g., semiparametric regression (Cox), mixture models (Neyman–Scott).

2) Adaptiveness (Stein's necessary condition) of robust estimators w.r.t. a finite-dim. nuisance parameter; e.g., location, linear regression, and ARMA.

3) Semiparametric treatment of gross error deviations from an ideal model as an ∞-dim. nuisance parameter, by projection on balls; for testing, an asymptotic version of the Huber–Strassen maximin result is thus obtained.

4) Uniform and nonuniform asymptotic normality of robust and adaptive estimators, respectively, in regression and time series models.

5) Fragility of optimal one-sided tests and confidence limits obtained for convex tangent cones, by projection on cones, as opposed to stability of corresponding procedures, even two-sided, for linear tangent spaces.

6) Control of the unknown neighborhood radius, a nuisance parameter in robustness.

Model $\mathcal{Q} = \{\, Q_{\theta,\nu} \,\}$ with parameter of interest $\theta \in \Theta$ open $\subset$ some $\mathbb{R}^k$, and nuisance parameter $\nu \in H_\theta$. Differentiability at any fixed (θ_0, ν_0)

$$dQ_{\theta_0+ta,\nu_t^g} \approx \bigl(1 + t(a'\Lambda + g)\bigr)\, dQ_{\theta_0,\nu_0} \quad \text{as } t \to 0 \tag{1}$$

in direction $a \in \mathbb{R}^k$, along paths $t \mapsto \nu_t^g$. Tangents $g \in L_2(Q_{\theta_0,\nu_0})$, $g \perp 1$,

$$\partial_1 \mathcal{Q} = \{\, a'\Lambda \mid a \in \mathbb{R}^k \,\}, \quad \partial_2 \mathcal{Q} \text{ a cone}, \quad \partial \mathcal{Q} = \partial_1 \mathcal{Q} + \partial_2 \mathcal{Q}$$

Fisher information of $\mathcal{Q}_{\nu_0}$ (ν fixed to ν_0) about θ at θ_0: $\quad \mathcal{I} = \operatorname{Cov} \Lambda$

Bickel (1982), Bickel et al. (1993), v.d.Vaart (1998)
Rieder (1994), Shen (1995)

Influence curves at Q_{θ_0,ν_0}:

$$\psi \in L_2^k, \quad \mathrm{E}\,\psi = 0, \quad \mathrm{E}\,\psi\Lambda' = \mathbb{I}_k, \quad \mathrm{E}\,\psi g = 0 \;\; \forall g \in \partial_2 \mathcal{Q} \tag{2}$$

F-consistent diff. functionals: $\quad T(Q_{\theta_0+ta,\nu_t^g}) - \theta_0 \approx \mathrm{E}\,\psi(a'\Lambda + g)\, t = ta$

AL estimators: $\quad n^{1/2}(S_n - \theta_0) \approx n^{-1/2} \sum_{i=1}^n \psi(x_i)$ in $Q^{(n)}_{\theta_0,\nu_0}$-probability, such that

$$\sqrt{n}\,(S_n - \theta_0) \longrightarrow \mathcal{N}(a, \operatorname{Cov}\psi) \quad \text{in law under } Q_n^{(n)}(a,g)$$

where $Q_n(a,g) = Q_{\theta_0+s_n a,\, \nu_{s_n}^g}$, at scale $s_n = 1/\sqrt{n}$, $n =$ sample size.

Let Π, Π_2 denote the (coordinatewise) orthogonal projections from $L_2^k(Q_{\theta_0,\nu_0})$ on the closed linear spans $c\ell \operatorname{lin} \partial\mathcal{Q} = \partial_1\mathcal{Q} + c\ell \operatorname{lin} \partial_2\mathcal{Q}$ and $c\ell \operatorname{lin} \partial_2\mathcal{Q}$, respectively. Unique projection on $c\ell \operatorname{lin} \partial\mathcal{Q}$ of all ICs:

$$\Pi(\psi) = \psi_{\text{eff}} := \mathcal{J}^{-1}\bar{\Lambda} \qquad \forall\, \psi \text{ IC} \tag{3}$$

where $\bar{\Lambda} := \Lambda - \Pi_2(\Lambda)$ (model $\mathcal{Q}$). $\psi_{\text{class}} := \mathcal{I}^{-1}\Lambda$ (model $\mathcal{Q}_{\nu_0}$).
Fisher informations of $\mathcal{Q}$ and $\mathcal{Q}_{\nu_0}$ for θ at (θ_0, ν_0) and θ_0, respectively:

$$\mathcal{J} = \operatorname{Cov}\bar{\Lambda} = \mathcal{I} - \operatorname{Cov}\Pi_2(\Lambda) \leqslant \mathcal{I} = \operatorname{Cov}\Lambda \tag{4}$$

Asymptotic covariance bound for AL estimators with IC ψ:

$$\operatorname{Cov}\psi \geqslant \mathcal{J}^{-1} = \operatorname{Cov}\psi_{\text{eff}}\,, \qquad \text{attained iff } \; \psi = \psi_{\text{eff}} \tag{5}$$

Information bounds: $\mathcal{J}^{-1}$ (model $\mathcal{Q}$) $\geqslant$ $\mathcal{I}^{-1}$ (model $\mathcal{Q}_{\nu_0}$).
Classical adaptivity (necessary condition): Stein (1956)

$$\mathcal{J}^{-1} = \mathcal{I}^{-1} \iff \Lambda \perp \partial_2\mathcal{Q} \iff \psi_{\text{eff}} = \psi_{\text{class}} \tag{6}$$

ICs exist iff $\mathcal{J} > 0$ iff $a'\Lambda \notin c\ell \operatorname{lin} \partial_2\mathcal{Q}$ $\forall a \in \mathbb{R}^k, a \neq 0$. Shen (1995), Rieder (2000)
Bounded ICs exist iff $a'\Lambda \notin$ L_1-closure $c\ell_1 \operatorname{lin} \partial_2\mathcal{Q}$ $\forall a \in \mathbb{R}^k, a \neq 0$.

1.3 Semiparametric Regression: $P_{\theta,\nu} = Q^{(W_{\theta,\nu},Z)}$ = law of observations $(W_{\theta,\nu}, Z)$, where Z is some k-dim. covariate and $W_{\theta,\nu}$ are the responses. Optimally bounded ICs at (θ,ν) are of (sufficient) form

$$\varrho = (A\Lambda - \xi - a)\,\min\Big\{1, \frac{b}{|A\Lambda - \xi - a|}\Big\} \tag{7}$$

for some $b \in (0,\infty)$, $A \in \mathbb{R}^{k\times k}$, $\xi \in c\ell\,\mathrm{lin}\,\partial_2\mathcal{Q}$, and some a.
If the joint law of $(W_{\theta,\nu}, Z)$ is distorted (errors-in-variables): $a \in \mathbb{R}^k$.
If only the conditional laws $Q^{W_{\theta,\nu}|Z=z}(dw)$ may be distorted and the marginal $Q^Z(dz)$ is kept ideal (error-free-variables), then $a : \mathbb{R}^k \to \mathbb{R}^k$, such that $\mathrm{E}(\varrho|Z) = 0$.

Remark Actually, condition $\varrho \perp \partial_2\mathcal{Q}$ (infinite-dim.) allows only approximations of the optimal ρ: Assuming a CONS $g_1, g_2, \ldots$ of $c\ell\,\mathrm{lin}\,\partial_2\mathcal{Q}$, one can prove the existence of $A_m \in \mathbb{R}^{k\times k}$, $\xi_m \in \mathrm{lin}\{g_1,\ldots,g_m\}$, and a_m, such that the IC ϱ_m of form (7)—now $\perp \partial_2\mathcal{Q}$ weakened to $\varrho_m \perp g_1,\ldots,g_m$—tend in $L_2(P)$ to the optimal ϱ (not necessarily of form (7)). Shen (1995), Ruckdeschel, Hable, Rieder (2010)

Cox regression: Response variables $W_{\theta,\nu} = \left(T_{\theta,\nu} \wedge C, \mathbf{1}_{\{T_{\theta,\nu} \leqslant C\}}\right)$ from survival times $T_{\theta,\nu}$ and a bounded censoring time C; $T_{\theta,\nu}$ and C stoch. independent given Z.
The cumulative hazard function of $T_{\theta,\nu}\,|Z$ assumed of form $e^{\theta' Z}\nu$ for some $\theta \in \mathbb{R}^k$ and unknown, abs. continuous baseline hazard function ν.
Then the parametric scores function Λ at (θ, ν) is

$$\Lambda\big((y,\delta), z\big) = \delta z - z e^{\theta' z}\, \nu(y) \tag{8}$$

and $\partial_2 \mathcal{Q} = B\, L_2(\nu)$ for the operator B defined by

$$B\zeta\colon \big((y,\delta), z\big) \longmapsto \delta\, \zeta(y) - e^{\theta' z} \int_{[0,y]} \zeta\, d\nu\,, \qquad \zeta \in L_2(\nu) \tag{9}$$

The projection on $c\ell\, \partial_2 \mathcal{Q}$ equals $\Pi_2 = B(B^* B)^{-1} B^*$, B^* the adjoint, where $(B^* B)^{-1} B^*(y) = \mathrm{E}(Z|Y = y, \delta = 1)$. Estimation of θ, since $\Pi_2(\Lambda) \neq 0$, is not adaptive w.r.t. ν. Bickel, Klaassen et al. (1993), van der Vart (1998)

Remark To the IC ϱ of form (7), a robust version of the Cox PLE is constructed, using the order statistics to $T_{\theta,\nu} \wedge C$, as an M-estimator with the random weights $\min\left\{1, \frac{b}{|A\Lambda - \xi - a|}\right\}$ evaluated at a starting estimate $(\tilde{\theta}, \tilde{\nu})$, and a weighted Breslow estimate of ν employing the same weights. Ruckdeschel, Hable, Rieder (2010)

1.4 Exponential mixture models: $Q_{\theta,\nu}(dx) = \int M_{\theta,u}(dx)\,\nu(du)$, each $M_{\theta,u}(dx)$ a pm with μ-density $f(x,\theta,u) = \exp\{u' T_\theta(x) + S_\theta(x) - b(\theta,u)\}$ and distribution $\nu(du)$ of the incidental parameter. Setting dot$=\partial/\partial\theta$,

$$\Lambda(X,\theta,\nu) = \dot{T}(X,\theta)'\,\mathrm{E}(U|T) + \dot{S}(X,\theta) - \mathrm{E}(\dot{b}(\theta,U)|T) \qquad (10)$$

$$\partial_2\mathcal{Q} = \left\{\, w(X) \in L_2 \;\middle|\; \mathrm{E}\,w(X) = 0\,,\;\; w(X) \text{ is } \sigma(T)\text{-measurable} \,\right\} \qquad (11)$$

$$\Pi_2\colon\, h(X) \longmapsto \mathrm{E}(h(X)|\,T(X,\theta)) - \mathrm{E}\,h(X) \qquad (12)$$

where $\partial_2\mathcal{Q} = c\ell\,\mathrm{lin}\,\partial_2\mathcal{Q} = c\ell_1\,\mathrm{lin}\,\partial_2\mathcal{Q}$. Bickel, Klaassen et al. (1993))

Optimally robust IC of necessary and sufficient form (7): Shen (1995), Fischer (2006)

$$\varrho = \Gamma\min\{1, \frac{b}{|\Gamma|}\}\,, \qquad \Gamma = A\Lambda - \xi - a\,, \quad \Lambda = \Lambda(X,\theta,)$$

with $\xi \in L_2(T(X,\theta))$ and $a \in \mathbb{R}$ determined such that $\mathrm{E}(\varrho|\,T) = 0$.

Special case: $T(X,\theta) = T(X)$ and $S(X,\theta) = \theta' S(x)$. Then

$$\Lambda = S - \mathrm{E}(\dot{b}|\,T), \quad \bar{\Lambda} = S - \mathrm{E}(S|\,T) \qquad (13)$$

The conditional density of X on $T=t$ w.r.t. $\mu(dx|\,T=t)$ not depending on ν(?), Λ(?) and $\bar{\Lambda}$(!) do not depend on ν: classical adaptivity. More generally, $\xi \in L_2(T(X))$ and $a \in \mathbb{R}$ such that $\mathrm{E}(\varrho|\,T) = 0$ do not depend on ν: robust adaptivity (§2). Shen (1995), Fischer (2006)

1.4 Finite Dimensional Case: In case $\nu \in H_\theta \subset$ some $\mathbb{R}^m$, differentiability (1) is assumed with

$$\partial_2 \mathcal{Q} = \{\, b'\Delta \mid b \in \mathbb{R}^m \,\} \tag{14}$$

for some nuisance scores $\Delta \in L_2^m(Q_{\theta_0,\nu_0})$, $\mathrm{E}\,\Delta = 0$, $\mathcal{D} := \mathrm{Cov}\,\Delta > 0$. Fisher information at (θ_0, ν_0) for the full parameter (θ, ν) is

$$\mathcal{H} = \mathrm{Cov}\begin{pmatrix}\Lambda\\ \Delta\end{pmatrix} = \begin{pmatrix}\mathcal{I} & \mathcal{C}\\ \mathcal{C}' & \mathcal{D}\end{pmatrix}, \qquad \mathcal{C} = \mathrm{E}\,\Lambda\Delta' \tag{15}$$

where $\det\mathcal{H} = \det\mathcal{D}\det\mathcal{J}$, $\mathcal{J} = \mathcal{I} - \mathcal{C}\,\mathcal{D}^{-1}\mathcal{C}'$, and $\Pi_2\Lambda = \mathcal{C}\,\mathcal{D}^{-1}\Delta$. Then

Neyman (1951): $C(\alpha)$-tests

$$\psi_{\mathrm{eff}} = \mathcal{J}^{-1}(\Lambda - \mathcal{C}\,\mathcal{D}^{-1}\Delta) \tag{16}$$

and $\psi_{\mathrm{eff}} =$ first k coordinates of $\mathcal{H}^{-1}\binom{\Lambda}{\Delta} = \psi_{\mathrm{class}}^{\mathrm{full}}$ for the full parameter.
Adaptivity $\iff \mathcal{C} = \mathrm{E}\,\Lambda\Delta' = 0$ (symmetric in main/nuisance parameter).

Minmax MSE problems for AL estimators in robust neighborhood models:

$$\mathrm{MSE}_*(\psi, r) = \mathrm{E}\,|\psi|^2 + r^2\,\omega_*^2(\psi) = \min!$$

that, in addition to the asymptotic variance $\mathrm{Cov}\,\psi$, involve the maximum asymptotic biasses ω_* generated by shrinking $r/\sqrt{n}$-neighborhoods about Q_{θ_0,ν_0}, $\omega_* =$ sup-norm and variants, e.g., integral of sectionwise supnorms, Rieder (1994) and refer to the following two sets of ICs, respectively:

1. model $\mathcal{Q}_{\nu_0}$ (no nuisance ν): $\psi \in L_2^k(Q_{\theta_0,\nu_0})$, $\mathrm{E}\,\psi = 0$, $\mathrm{E}\,\psi\Lambda' = \mathbb{I}_k$
2. model $\mathcal{Q}$ (with nuisance ν): in addition, $\mathrm{E}\,\psi g = 0\ \ \forall\, g \in \partial_2\mathcal{Q}$

Due to strict convexity, the minimizers ϱ_1 and ϱ_2, respectively, are unique, and minMSE1 $\leqslant$ minMSE2. Robust adaptivity (extending classical):

$$\text{minMSE1} = \text{minMSE2} \iff \varrho_1 = \varrho_2 \iff \varrho_1 \perp \partial_2\mathcal{Q} \qquad (17)$$

Nonadaptivity (quantitative): $\dfrac{\text{minMSE2}}{\text{minMSE1}} - 1.$

2.2 Symmetric Location

Beran (1974), Stone (1975)

$$Q_{\theta,f}(dx) = f(x-\theta)\,\lambda(dx), \quad \theta \in \mathbb{R} \tag{18}$$

f symmetric, $\mathcal{I}_f^{\mathrm{loc}} = \int (\Lambda_f^{\mathrm{loc}})^2 f\,d\lambda < \infty$, $\Lambda_f^{\mathrm{loc}} = -\dot{f}/f$, $dF = f\,d\lambda$.
For $\theta_0 = 0$, $f = f_0$ fixed, $\partial_2\mathcal{Q} = \{\, g \in L_2(F) \mid \mathrm{E}\, g = 0,\ g \text{ symmetric} \,\}$.
By symmetry, $\Lambda_f^{\mathrm{loc}} = -\dot{f}/f$ (odd) $\perp g$ (symmetric) in $L_2(F)$: $\Longrightarrow$
$\Lambda_f^{\mathrm{loc}} \notin c\ell_1 \operatorname{lin} \partial_2\mathcal{Q}$ and classical adaptivity holds.

Robust ICs, for known $\nu_0 = f$, Huber (1981), Hampel et al. (1985), Rieder (1994)

$$\varrho(x) = A\,\Lambda_f^{\mathrm{loc}}(x)\min\{1, c\,|\Lambda_f^{\mathrm{loc}}(x)|^{-1}\}$$

are all odd (like Λ_f^{loc}), hence $\varrho \perp \partial_2\mathcal{Q}$: $\Longrightarrow$ robust adaptivity

Remark Adaptive constructions that not only achieve asymptotic linearity with the robust IC in the ideal model but uniform asymptotic normality over shrinking neighborhoods not yet solved completely. Shen (1994), Stabla (2005)

2.3 Regression and Scale

Kohl (2005)

$$Q_{\theta,\sigma}(dx, dy) = \frac{1}{\sigma} f\left(\frac{y - x'\theta}{\sigma}\right) \lambda(dy) K(dx) \tag{19}$$

Assumptions: F symmetric, finite Fisher information of location $\mathcal{I}_f^{\rm loc}$ and scale $\mathcal{I}_f^{\rm sc} = \int (\Lambda_f^{\rm sc})^2 \, dF$, where $\Lambda_f^{\rm sc}(u) = u\Lambda^{\rm loc}(u) - 1$; $\int xx' \, K(dx) > 0$.
Classical adaptivity holds (i.e., w.r.t. σ and w.r.t. θ)—due to symmetry of F—and extends to robust adaptivity w.r.t. σ and w.r.t. θ, in connection with the bias terms

$$\omega_{c,0}(\psi) = \omega_{c,1}(\psi) = \sup_{x,u} |\psi(x, u)| \tag{20}$$

$$\omega_{c,2}^2(\psi) = \int \sup_u |\psi(x, u)|^2 \, K(dx) \tag{21}$$

These biasses are generated by contamination neighborhoods (Tukey, $* = c$), which are unconditional ($t = 0$), or errors-in-variables, or are average conditional, error-free-variables, ($t = \alpha = 1$), respectively by average square conditional neighborhoods ($t = \alpha = 2$, $* = c$).

Bickel (1984), Rieder (1987)

Robust ICs for regression and scale

F symmetric, $t=0$ and $t=\alpha=1$

θ main, σ nuisance:

$$\varrho_{\rm rg}(x,u) = A_{\rm rg} x \Lambda_f^{loc}(u)\, w_{\rm rg}(x,u) \tag{22}$$

$$w_{\rm rg}(x,u) = \min\{\,1, b_{\rm rg}\, |A_{\rm rg} x \Lambda_f^{\rm loc}(u)|^{-1}\} \tag{23}$$

$$A_{\rm rg}^{-1} = \mathrm{E}\, xx' \Lambda_f^{\rm loc}(u)^2 w_{\rm rg}(x,u) \tag{24}$$

$$r^2 b_{\rm rg} = \mathrm{E}\big(|A_{\rm rg} x \Lambda_f^{\rm loc}(u)| - b_{\rm rg}\big)_+ \tag{25}$$

σ main, θ nuisance:

$$\varrho_{\rm sc}(u) = A_{\rm sc}(\Lambda_f^{\rm sc}(u) - z_{\rm sc}) w_{\rm sc}(u) \tag{26}$$

$$w_{\rm sc}(u) = \min\{\,1, c_{\rm sc}\, |\Lambda_f^{\rm sc}(u) - z_{\rm sc}|^{-1}\} \tag{27}$$

$$z_{\rm sc} = \mathrm{E}\, \Lambda_f^{\rm sc} w_{\rm sc} / \,\mathrm{E}\, w_{\rm sc} \tag{28}$$

$$A_{\rm sc}^{-1} = \mathrm{E}(\Lambda_f^{\rm sc} - z_{\rm sc})^2 w_{\rm sc} \tag{29}$$

$$r^2 c_{\rm sc} = \mathrm{E}\big(|\Lambda_f^{\rm sc} - z_{\rm sc}| - c_{\rm sc}\big)_+ \tag{30}$$

Full parameter (θ,σ): $\varrho = \binom{\varrho_{\rm rg}}{\varrho_{\rm sc}}$, but weights $w_{\rm rg}$, $w_{\rm sc}$ both replaced by

$$w(x,u) = \min\Big\{1, b\, \Big||A_{\rm rg} x|^2 (\Lambda_f^{loc}(u))^2 + A_{\rm sc}^2 (\Lambda_f^{\rm sc}(u) - z_{\rm sc})^2\Big|^{-1/2}\Big\} \tag{31}$$

where

$$r^2 b = \mathrm{E}\Big(\,\Big||A_{\rm rg} x|^2 (\Lambda_f^{loc}(u))^2 + A_{\rm sc}^2 (\Lambda_f^{\rm sc}(u) - z_{\rm sc})^2\Big|^{1/2} - b\Big)_+ \tag{32}$$

Especially, if $F = \mathcal{N}(0,1)$, then $|\varrho_{\rm rg}(x,u)| \sim 1/u$ (x fixed, $|u| \to \infty$).

Regression with intercept as nuisance parameter

$$Q_{\theta,\,\mu}(dx, dy) = f(y - \mu - x'\theta)\,\lambda(dy)\,K(dx) \tag{33}$$

Assumptions: F symmetric, $\mathcal{I}_f^{\mathrm{loc}} < \infty$, $\int xx'\,K(dx) > 0$, $\int x\,K(dx) = 0$.

Classical adaptivity, even if K is asymmetric.

Robust adaptivity for average square conditional neighborhoods
$t = \alpha = 2$, $* = c$, even if K is asymmetric.

Robust adaptivity for unconditional neighborhoods $* = c$, $t = 0$ and average conditional neighborhoods $* = c$, $t = \alpha = 1$, if K symmetric.

For asymmetric K, $* = c$, $t = 0$ or $t = \alpha = 1$, no robust adaptivity, since

$$\mathrm{E}\,\varrho_{\mathrm{rg}}\,\Lambda_f^{\mathrm{loc}} = A_{\mathrm{rg}}\,\mathrm{E}\,x\Lambda_f^{\mathrm{loc}}(u)^2 \min\{\,1, b_{\mathrm{rg}}\,|A_{\mathrm{rg}}\Lambda_f^{\mathrm{loc}}(u)|^{-1}\} \neq 0 \tag{34}$$

For a 2-point asymmetric K, nonadaptivity may be up to 300% Kohl (2005)

Robust ICs in model $\mathcal{Q}$ are of form (15), (16), (18) with $A_{\mathrm{rg}}x$ replaced by $A_{\mathrm{rg}}x + A_\mu$, and (17) by

$$A_{\mathrm{rg}}\,\mathrm{E}\,xx'\Lambda_f^{\mathrm{loc}}(u)^2 w = \mathbb{I}_k - A_\mu\,\mathrm{E}\,x'\Lambda_f^{\mathrm{loc}}(u)^2 w \tag{35}$$

$$A_\mu\,\mathrm{E}\,\Lambda_f^{\mathrm{loc}}(u)^2 w = -A_{\mathrm{rg}}\,\mathrm{E}\,x\Lambda_f^{\mathrm{loc}}(u)^2 w \tag{36}$$

2.3 ARMA(p, q): $\boldsymbol{\phi(B)(X_t - \mu) = \xi(B)V_t}$ $t \in \mathbb{Z}$, B backshift

Innovations V_t i.i.d.$\sim F$, $\mathcal{I}_F^{\rm loc} < \infty$, $\int u\, F(du) = 0$, $\int u^2\, F(du) < \infty$.
Stationarity and invertibility assumption: $\phi(z)\xi(z) \neq \ \forall\, |z| \leqslant 1$,
ϕ, ξ relatively prime ($\Rightarrow$ positive Fisher information), $\phi_p\, \xi_q \neq 0$.
Influence $\psi(x_{\leqslant t})$ of observation x_t given the past $x_{<t} := (x_{t-1}, x_{t-2}, \dots)$.
Influence curves $\psi(x_{\leqslant t})$ of AL estimators as in (2), but $\mathrm{E}\big(\psi(x_{\leqslant t})\big|x_{<t}\big) = 0$
(stationary, ergodic martingale differences). Jeganathan (1982), Staab (1984), Rieder (2003)
Differentiability (1) now refers to transition densities of the ideal model $\mathcal{P}$.
Joint law of $x_{\leqslant n}$: $Q^{(n)}(dx_{\leqslant n}) = \prod_{j=1}^n Q^{(n,j|<j)}(dx_j|x_{<j})Q^{(n,0)}(dx_{\leqslant 0})$
Neighborhoods ($* = c,\ t = \varepsilon$) of radius $r_n = r\, s_n = r\, n^{-1/2}$ about the ideal transition distributions $P^{(n,j|<j)}(dx_j|x_{<j})$ with contamination curve ε:

$$Q^{(n,j|<j)}(dx_j|x_{<j}) = \\ (1 - r_n\, \varepsilon(x_{<j}))P^{(n,j|<j)}(dx_j|x_{<j}) + r_n\, \varepsilon(x_{<j})M^{(n,j|<j)}(dx_j|x_{<j}) \qquad (37)$$

where $M^{(n,j|<j)}(dx_j|x_{<j})$ any kernel, initial distribution (of $x_{\leqslant 0}$) left ideal.
$\alpha = 1$: $\mathrm{E}\,\varepsilon \leqslant 1$, $\alpha = 2$: $\mathrm{E}\,\varepsilon^2 \leqslant 1$ Bickel (1984), Rieder (1987) for regression

Bias terms for $* = c$ and $t = \varepsilon$, respectively $t = \alpha = 1, 2$:

$$\omega_{c,\varepsilon}(\psi) = \mathrm{E}\,\varepsilon(x_{\leqslant 0}) \sup_{x_1} |\psi(x_1, x_{\leqslant 0})| \tag{38}$$

$$\omega_{c,1}(\psi) = \|\psi\|_\infty\,, \quad \omega^2_{c,2}(\psi) = \mathrm{E} \sup_{x_1} |\psi(x_1, x_{\leqslant 0})|^2 \tag{39}$$

Transition scores: $\Lambda_1 = \Lambda_f^{\mathrm{loc}}(V_1)\big(H_1', \tau\big)'$ where $\tau = \phi(1)/\xi(1)$ and

$$H_1' = \big(-B\phi^{-1}(B), \dots, -B^p\phi^{-1}(B); B\xi^{-1}(B), \dots, B^q\xi^{-1}(B)\big)\,V_1 \tag{40}$$

Denoting $\mathcal{K} = \mathrm{Cov}\,H_1$, Fisher information is: $\mathcal{I} = \mathcal{I}_F^{\mathrm{loc}}\,\mathrm{diag}(\mathcal{K}, \tau^2)$

$\Longrightarrow$ classical adaptivity (w.r.t. μ and w.r.t. (ϕ, ξ))

Analogy to regression with intercept on identifying H_1 as regressor.
Robust ICs are of regression type form (15), (16), (18), (28), (29).
In the model with parameter $(\phi, \xi, \mu = 0)$:

$$\varrho_{c,\alpha} = AH_1(\Lambda_f^{\mathrm{loc}}(V_1) - \vartheta_\alpha)\,w_\alpha\,, \quad w_\alpha = \min\Big\{\,1, \frac{\beta_\alpha}{|\Lambda_f^{\mathrm{loc}}(V_1) - \vartheta_\alpha|}\,\Big\} \tag{41}$$

$\alpha = 1$: $\beta_1 = b/|AH_1|$, $\vartheta_1 = \vartheta_1(H_1)$ Hampel-type

$\alpha = 2$: $\beta_2 =$ constant, $\vartheta_2 =$ constant Huber-type

Robust Adaptivity for ARMA

1) Estimation of (ϕ, ξ), nuisance parameter μ:

Robust adaptivity in case $\alpha = 2$, in case $\alpha = 1$ if F is symmetric.

In fact, $\mathrm{E}\, H_1 = 0$, and H_1, $\Lambda_f^{\mathrm{loc}}(V_1)$ are stochastically independent, so

$$\mathrm{E}\, AH_1(\Lambda_f^{\mathrm{loc}} - \vartheta_2)^2 \min\{1, \tfrac{\beta_2}{|\Lambda_f^{\mathrm{loc}} - \vartheta_2|}\} = 0 \qquad (\alpha = 2)$$

But

$$\mathrm{E}\, AH_1(\Lambda_f^{\mathrm{loc}} - \vartheta_1)^2 \min\{1, \tfrac{b/|AH|}{|\Lambda_f^{\mathrm{loc}} - \vartheta_1(H)|}\} = 0 \qquad (\alpha = 1)$$

where $\vartheta_1(H_1) = \vartheta_1(-H_1)$, requires $\mathcal{L}_F(H_1)$, resp. F, to be symmetric.

Nonadaptivity for AR(1), MA(1) with asymmetric $F = \mathrm{Gumbel}(\gamma, 1)$,
$\gamma = -\mathrm{di}\Gamma(1)$ ($\Rightarrow \int v\,dF(v) = 0$), at most 3%! Kohl (2005)

2) Estimation of μ with nuisance parameter (ϕ, ξ):

Robust adaptivity for $\alpha = 1, 2$

robust IC: $\varrho_{c,12} = A\Lambda_f^{\mathrm{loc}}(V_1) \min\{1, \beta_{12}\, |\Lambda_f^{\mathrm{loc}}(V_1)|^{-1}\}$ $\alpha = 1, 2$

2.5 ARCH(p): $X_t = \sigma(1 + a_1 X_{t-1}^2 + \ldots + a_p X_{t-p}^2)^{1/2} V_t \qquad t \in \mathbb{Z}$

Innovations V_t i.i.d.$\sim F$, $\mathcal{I}_F^{\rm sc} < \infty$, $\int v dF(v) = 0$, $\int v^2 dF(v) = 1$
Stationarity, ergodicity: $\mathrm{E} \log V_t^2 + \log \sigma^2 + \log \max_j a_j < 0$

Estimation of a, nuisance parameter σ:
No adaptivity—neither classical nor robust ($* = c$, $\alpha = 1$).
ARCH(1) with $F = \mathrm{logNormal}(\delta, \gamma)$ with $\delta = -e^{\gamma^2/2}$ ($\Rightarrow \int v dF(v) = 0$):
Nonadaptivity increases with $r \in [0, \infty)$. Kohl (2005), MonteCarlo

$a_1 = 1$, $\gamma = .5$: $.3 \uparrow .4$, $\qquad a_1 = 10$, $\gamma = .5$: $25 \uparrow 160$.

Conclusion *Classical adaptivity extends to robust adaptivity for neighborhoods* $* = c$, $\alpha = 2$, *for neighborhoods* $* = c$, $t = 0$, $\alpha = 1$ *some additional symmetry of the ideal model may be needed.*

Neighborhood model $\mathcal{Q} = \{Q \mid Q \in U_*(P_\theta, r),\ \theta \in \Theta\}$ of neighborhoods about the elements of an ideal model $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$. Writing

$$Q_{\theta,\nu} = P_\theta + \nu\,, \quad \nu := Q - P_\theta \quad \text{for} \quad Q \in U_*(P_\theta, r) \tag{42}$$

puts $\mathcal{Q}$ into semiparametric model form: main parameter θ, nuisance parameter $\nu = Q - P_\theta \in H_\theta := U_*(P_\theta, r) - P_\theta$; in particular, $\mathcal{P} = \mathcal{Q}_{\nu_0=0}$.

Remark We assume a true θ (idealistic approach), so the law Q may be referred to this θ. Conversely, given Q, the inclusion $Q \in U_*(P_\theta, r)$ may not define θ uniquely.

Neigborhoods $U_c(\theta, r) = \{(1-r)P + r\,M \mid M \text{ any probability}\}$ (convex contamination) and balls $U_*(\theta, r) = \{\,Q \mid d_*(Q, P_\theta) \leqslant r\,\}$ in the Hellinger and total variation metrics, which are defined by

$$\sqrt{2}\, d_h^2(Q, P) = \left\|\sqrt{dQ} - \sqrt{dP}\,\right\|_2, \quad 2\, d_v(Q, P) = \|dQ - dP\|_1 \tag{43}$$

3.1 Proposition 1 *Fix θ_0, $\nu_0 = 0$. Then $\partial_2 \mathcal{Q}_* = rG_*$ for $* = h, v, c$, where $G_* =$ all functions $g \in L_2(P_{\theta_0})$, $\mathrm{E}\, g = 0$, such that, respectively,*

$$(h)\ \ \mathrm{E}\, g^2 \leqslant 8 \qquad (v)\ \ \mathrm{E}\, |g| \leqslant 2 \qquad (c)\ \ g \geqslant -1 \tag{44}$$

In particular, if $r > 0$: $c\ell \operatorname{lin} \partial_2 \mathcal{Q}_* = L_2(P_{\theta_0})$, so $\Pi_2(\Lambda) = \Lambda$ and $\bar{\Lambda} = 0$.

We therefore dispense with the linear span and define the sp-robust IC

$$\tilde\varrho_* := \mathcal{C}^{-1}\tilde\Lambda\,, \qquad \tilde\Lambda = \Lambda - \widetilde\Pi_2(\Lambda)\,, \qquad \mathcal{C} = \mathrm{E}\,\tilde\Lambda\Lambda' \tag{45}$$

in analogy to ψ_{eff}, but employing nonlinear projection $\widetilde\Pi_2 \colon L_2^k \to (rG_*)^k$ onto closed convex sets; the radius r is assumed so small that $\det\mathcal{C} \neq 0$.

3.2 Lemma 1 *Let $\tilde{\mathcal{G}}$, $\hat{\mathcal{G}}$, $\bar{\mathcal{G}}$ be some nonempty closed convex: subset, cone, and linear subspace, respectively, of some Hilbert space $\mathcal{H}$. Then, for any $\kappa \in \mathcal{H}$, the unique best approximations $\tilde\kappa \in \tilde{\mathcal{G}}$, $\hat\kappa \in \hat{\mathcal{G}}$, $\bar\kappa \in \bar{\mathcal{G}}$ of κ are characterized by*

$$\langle\kappa - \tilde\kappa|g\rangle \leqslant \langle\kappa - \tilde\kappa|\tilde\kappa\rangle, \quad \langle\kappa - \hat\kappa|g\rangle \leqslant \langle\kappa - \hat\kappa|\hat\kappa\rangle = 0\,, \quad \langle\kappa - \bar\kappa|g\rangle = 0 \tag{46}$$

for all g in $\tilde{\mathcal{G}}$, $\hat{\mathcal{G}}$, and $\bar{\mathcal{G}}$, respectively.

3.2 Theorem 2 (h) *If* $8\,r^2 < \min_{j=1,\dots,k}\mathcal{I}_{j,j}$ *then* $\tilde\varrho_h = \mathcal{I}^{-1}\Lambda$.

(v) *If* $2\,r < \min_{j=1,\dots,k}\mathrm{E}\,|\Lambda_j|$ *then*

$$\tilde\Lambda_j^{(v)} = v_j' \vee \Lambda_j \wedge v_j'' \quad \text{where} \quad \mathrm{E}(v_j' - \Lambda_j)_+ = r = \mathrm{E}(\Lambda_j - v_j'')_+ \tag{47}$$

(c) *If* $r < -\max_{j=1,\dots,k}\inf_{P_{\theta_0}}\Lambda_j$ *then*

$$\tilde\Lambda_j^{(c)} = (\Lambda_j + r) \wedge u_j \qquad \textit{where} \quad \mathrm{E}\big((\Lambda_j + r)\wedge u_j\big) = 0 \tag{48}$$

Hellinger balls Since $\mathrm{MSE}_h(\psi, r) = \operatorname{tr} \mathrm{Cov}\, \psi + 8r^2 \operatorname{maxev} \mathrm{Cov}\, \psi$ and $\mathrm{Cov}\, \psi \geqslant \mathcal{I}^{-1}$ $\forall$ ICs in model $\mathcal{P}$, the robust IC is $\varrho_h = \mathcal{I}^{-1}\Lambda$ for all $r \geqslant 0$. Thus $\tilde{\varrho}_h = \varrho_h$ whenever $\tilde{\varrho}_h$ is defined.

Remark Despite $\varrho_h = \psi_{\text{class}}$, model $\mathcal{P}$ is not adaptive w.r.t. Hellinger neighborhoods since $\mathrm{MSE}_h(\varrho_h, r) = \operatorname{tr} \mathcal{I}^{-1} + 8r^2 \operatorname{maxev} \mathcal{I}^{-1} > \mathrm{MSE}_h(\varrho_h, 0)$ for $r > 0$.

Contamination neighborhoods Risk $\mathrm{MSE}_c(\psi, r) = \|\psi\|_2^2 + r^2 \|\psi\|_\infty^2$ is uniquely minimized by the robust IC ϱ_c,

$$\varrho_c = (A\Lambda - a) \min\big\{1, b\, |A\Lambda - a|^{-1}\big\} \tag{49}$$

where Rieder (1994)

$$r^2 b = \mathrm{E}\big(|A\Lambda - a| - b\big)_+ \tag{50}$$

The sp-robust IC $\tilde{\varrho}_c$ (exchanging linear combination and clipping) has the coordinates

$$\tilde{\varrho}_{c,j} = C^-_{j,1} (\Lambda_1 + r) \wedge u_1 + \cdots + C^-_{j,k} (\Lambda_k + r) \wedge u_k \tag{51}$$

with upper clipping constants u_j from (48) and $(C^-_{j,i}) = \mathcal{C}^{-1}$ from (45).
In general, due to only one-sided (upper) bounds: $\mathrm{MSE}_c(\tilde{\varrho}_c, r) = \infty$!

Total variation balls—dimension $k = 1$

Robust IC minimizing $\mathrm{MSE}_v(\psi, r) = \|\psi\|_2^2 + r^2(\sup\psi - \inf\psi)^2$ is given by

$$\varrho_v = c' \vee A\Lambda \wedge c'' \tag{52}$$

where

$$r^2(c'' - c') = \mathrm{E}(c' - A\Lambda)_+ = \mathrm{E}(A\Lambda - c'')_+ \tag{53}$$

3.3 Theorem 1 *The sp-robust IC $\tilde{\varrho}_v$ for $r < \mathrm{E}\,\Lambda_+$ coincides with the robust IC ϱ_v for*

$$\tilde{r} = \sqrt{\frac{r}{v_r'' - v_r'}} \tag{54}$$

where

$$\mathrm{E}(v_r' - \Lambda)_+ = r = \mathrm{E}(\Lambda - v_r'')_+ \tag{55}$$

Rieder (2000)

3.3 Example 2 In case $P_\theta = \mathcal{N}(\theta, 1)$, $\tilde{\varrho}_v$ turns out pessimistic since

$$\tilde{r}/r \geqslant 2.2 \ \ \forall r < 1/\sqrt{2\pi}\,, \quad \text{and } \tilde{r}/r \uparrow \infty \text{ as } r \downarrow 0 \text{ or } \uparrow 1/\sqrt{2\pi}$$

MSE-evaluation desirable.

Total variation balls—dimension $k > 1$

Robust IC minimizing $\mathrm{MSE}_v(\psi, r) = \|\psi\|_2^2 + r^2\omega_{v;s}^2$ for $s = 2, \infty$ with

$$\omega_{v;2}^2(\psi) = \textstyle\sum_{j=1}^k(\sup\psi_j - \inf\psi_j)^2, \quad \text{respectively}$$
$$\omega_{v;\infty}^2(\psi) = \max_{j=1,\dots,k}(\sup\psi_j - \inf\psi_j)^2,$$

has coordinates of form $\varrho_{v,j} = c_j' \vee A_j\Lambda \wedge c_j''$ where, for variant $s = 2$,

$$r^2(c_j'' - c_j') = \mathrm{E}(c_j' - A_j\Lambda)_+ = \mathrm{E}(A_j\Lambda - c_j'')_+ \quad \forall\, j = 1, \dots k$$

respectively, for variant $s = \infty$, $\forall\, j = 1, \dots k$,

$$r^2(c_j'' - c_j') = \textstyle\sum_{i=1}^k \mathrm{E}(c_i' - A_i\Lambda)_+ = \sum_{i=1}^k \mathrm{E}(A_i\Lambda - c_i'')_+$$

Sp-robust IC $\tilde\varrho_v$ has the coordinates:

$$\tilde\varrho_{v,j} = C_{j,1}^-\, v_1' \vee \Lambda_1 \wedge v_1'' + \cdots + C_{j,k}^-\, v_k' \vee \Lambda_k \wedge v_k'' \tag{56}$$

where $\mathrm{E}(v_j' - \Lambda_j)_+ = r = \mathrm{E}(\Lambda_j - v_j'')_+$ and $(C_{j,i}^-) = \mathcal{C}^{-1}$ from (45), (47).
Thus the order of clipping and linear combination is interchanged again.
$\tilde\varrho_v$ is suboptimal but still sensibly robust. A MSE-comparison desirable.

Dimension $k=1$. Given any probability P, we consider local asymptotic alternatives along tangents $g \in L_2(P)$, $\int g\,dP = \mathrm{E}\,g = \langle g|1\rangle = 0$,

$$dP_{n,g} \approx (1+s_n g)\,dP\,, \qquad s_n = 1/\sqrt{n} \tag{57}$$

E.g., by P-densities: $\big(\frac{1}{2}sg + (1-\frac{1}{4}s^2\|g\|^2)^{1/2}\big)^2$, or simply: $1+sg$, if $\|g\|_\infty < \infty$.

Observations $x_1,\ldots,x_n$ i.i.d. $\sim P_{n,g}$.

Let $G_0, G_1 \subset L_2 \cap \{\mathrm{E}=0\}$, $G_0 \cap G_1 = \emptyset$. Fix any $g := (g_0, g_1) \in G_0 \times G_1$.

The simple asymptotic testing problem H_{g_0} vs. K_{g_1} at level $\alpha \in (0,1)$ is:

$$\liminf_{n\to\infty} \int \delta_n\,dP^n_{n,g_1} = \max! \qquad \text{s.t.} \quad \limsup_{n\to\infty} \int \delta_n\,dP^n_{n,g_0} \leqslant \alpha \tag{58}$$

Denoting $g_{10} := g_1 - g_0$, the optimal test is $\delta_g = (\delta_{n,g})$,

$$\delta_{n,g} = \mathbf{1}\Big(s_n \sum_{i=1}^n g_{10}(x_i) > \|g_{10}\|\,u_\alpha + \langle g_{10}|g_0\rangle\Big) \tag{59}$$

where $\|.\| = \|.\|_2 = \langle .|.\rangle^{1/2}$, and u_α = standard normal upper α point: $\Phi(-u_\alpha) = \alpha$.

δ_g achieves asymptotic size α and power $\Phi(-u_\alpha + \|g_{10}\|)$ under H_{g_0}, K_{g_1}.

The tests $\delta_{n,g}$ are unique up to terms $\to 0$ in P^n-probability.

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

$$\inf_{g_1 \in G_1} \liminf_{n\to\infty} \int \delta_n \, dP^n_{n,g_1} = \max ! \quad \text{s.t.} \quad \sup_{g_0 \in G_0} \limsup_{n\to\infty} \int \delta_n \, dP^n_{n,g_0} \leqslant \alpha \qquad (60)$$

Assume now G_0, G_1 closed, convex. Pass to $G_{10} := G_1 - G_0$, which set is convex, but need not be closed if $\dim L_2(P) > 1$. We assume G_{10} closed and pick $q_{10} := q_1 - q_0$ the unique minimum norm element of G_{10}.

3.4 Theorem 1[saddle point for testing]

Then the maxmin asy. testing problem H_{G_0} vs. K_{G_1} at level α has saddle point (q, δ_q), and the maxmin asy. power $= \Phi(-u_\alpha + \|q_{10}\|)$.
Any other pair $g = (g_0, g_1)$ in $G_0 \times G_1$ achieving $g_{10} = q_{10}$ also provides a saddle point (g, δ_g), and necessarily $\delta_g = \delta_q$.

Proof Based on LAN, this is the statistical equivalent of the first characterization in (46) with $\kappa = 0$, for the minimum norm element of closed convex sets.

Given some scores $\Lambda \in L_2(P_{\theta_0})$, $\int \Lambda \, dP_{\theta_0} = 0$, and $\tau \in \mathbb{R}, \neq 0$, enlarge the parametric alternatives $dP_{\theta_0 + s_n\tau} \approx (1 + s_n \tau \Lambda) dP_{\theta_0}$ to $P_{n,g}$, by the nuisance parameter $g \in G_0$, respectively $g \in \tau\Lambda + G_1$. Then

$$q_{10} = \tau\Lambda - \widetilde{\Pi}_2(\tau\Lambda | G_0 - G_1) \qquad (61)$$

To test neighborhoods $U_*(\theta_0, s_n r_0)$ and $U_*(\theta_0 + s_n\tau, s_n r_1)$ about $P = P_{\theta_0}$ and $P_{\theta_0+s_n\tau}$ of type $* = h, v, c$ with possibly different radii $s_n r_0$ and $s_n r_1$, respectively, employ the tangent balls G_* defined in (37) and put

$$G_{*,0} = r_0 G_* \,, \qquad G_{*,1} = \tau\Lambda + r_1 G_* \tag{62}$$

Abbreviate $H_* := H_{G_{*,0}}$ and $K_* := K_{G_{*,1}}$.

3.5 Theorem 1 [Hellinger balls, $* = h$] *Let* $8\, r^2 < \tau^2 \, \|\Lambda\|^2$. *Then the least favorable tangent pair* $q_h = (q_{h,0}, q_{h,1})$ *in* $G_{h,0} \times G_{h,1}$ *is unique,*

$$q_{h,0} = r_0 \gamma\Lambda \,, \quad q_{h,1} = \tau\Lambda - r_1\gamma\Lambda \quad \textit{where } \gamma = \sqrt{8}\,\|\Lambda\|^{-1} \tag{63}$$

The maxmin test $\delta_{q_h} = (\delta_{n,q_h})$ *for* H_h *vs.* K_h *is*

$$\delta_{n,q_h} = \mathbf{1}\Big(s_n \|\Lambda\|^{-1} \sum_{i=1}^{n} \Lambda(x_i) > u_\alpha + \sqrt{8}\, r_0 \Big) \tag{64}$$

Maxmin asymptotic power $= \Phi\big(-u_\alpha + \tau\,\|\Lambda\| - \sqrt{8}\, r\,\big)$.

Remarks a) Despite of classical test statistics, no adaptivity w.r.t. Hellinger balls.
b) No Huber–Strassen least favorable pairs $dQ_1 = \pi\, dQ_0$ to compare with: Birgè (1980)

$$Q_0(\pi > t) \geqslant Q'(\pi > t), \; Q_1(\pi > t) \leqslant Q''(\pi > t) \quad \forall Q' \in \mathcal{Q}_0, Q'' \in \mathcal{Q}_1, \forall\, t > 0 \tag{65}$$

3.5 Theorem 2 [Total variation balls, $* = v$] *Let* $2\,r < \tau\,\mathrm{E}\,|\Lambda|$.

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* $q_{v,0} = r_0\,\tilde g_v$, $q_{v,1} = \tau\Lambda - r_1\,\tilde g_v$ *for the tangent* $\tilde g_v$ *defined by*

$$r\,\tilde g_v = \tau\,(\Lambda - v'')_+ - \tau\,(v' - \Lambda)_+ \tag{66}$$

and $v' < 0 < v''$ *determined by* $\tau\,\mathrm{E}(v' - \Lambda)_+ = r = \tau\,\mathrm{E}(\Lambda - v'')_+$.
A tangent pair $g_{v,0} = r_0 g_0$, $g_{v,1} = \tau\Lambda - g_1$ *is least favorable iff*

$$r_0\,g_0^+ + r_1\,g_1^+ = \tau\,(\Lambda - v'')_+\,, \qquad r_0\,g_0^- + r_1\,g_1^- = \tau\,(v' - \Lambda)_+ \tag{67}$$

With $\Lambda^{(v)} := v' \vee \Lambda \wedge v''$, *the maxmin test* $\delta_{q_v} = (\delta_{n,q_v})$ *for* H_v *vs.* K_v *is*

$$\delta_{n,q_v} = \mathbf{1}\Big(s_n \sum_{i=1}^n \Lambda^{(v)}(x_i) > \|\Lambda^{(v)}\|\,u_\alpha + r_0\,(v'' - v')\Big) \tag{68}$$

Maxmin asy. power $= \Phi\big(-u_\alpha + \tau\,\|\Lambda^{(v)}\|\,\big)$.

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

3.5 Theorem 3 [Contamination, $* = c$] *Let* $r_0 < \mathrm{E}\big(\tau\Lambda - (r_1 - r_0)\big)_+$.
a) *The least favorable tangent pair* $q_c = (q_{c,0}, q_{c,1})$ *in* $G_{c,0} \times G_{c,1}$ *is unique,*

$$q_{c,0} = \tau(\Lambda - c'')_+ - r_0\,, \quad q_{c,1} = \tau\Lambda + \tau(c' - \Lambda)_+ - r_1 \qquad (69)$$

where $c' < z := (r_1 - r_0)/\tau < c''$ *are determined by* $\mathrm{E}\, q_{c,0} = \mathrm{E}\, q_{c,1} = 0$.
Based on $\Lambda^{(c)} := c' \vee \Lambda \wedge c'' - z$, *the maxmin test* $\delta_{q_c} = (\delta_{n,q_c})$ *for* H_c *vs.* K_c *is*

$$\delta_{n,q_c} = \mathbf{1}\Big(s_n \sum_{i=1}^{n} \Lambda^{(c)}(x_i) > \|\Lambda^{(c)}\|\, u_\alpha + r_0\,(c'' - z)\Big) \qquad (70)$$

Maxmin asy. power $= \Phi\big(-u_\alpha + \tau\,\|\Lambda^{(c)}\|\,\big)$.

b) δ_{q_c} *coincides with the robust asy. 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 asy. minimum power over* $U_c\big(P_{\theta_0+\tau/\sqrt{n}}\,; r_1/\sqrt{n}\,\big)$ *subject to asy. maximum size* $\leqslant \alpha$ *over* $U_c\big(P_{\theta_0}; r_0/\sqrt{n}\,\big)$.

Huber (1964), (1968), Huber–Carol (1970), Huber–Strassen (1973), Rieder (1978), (2000)

Summary

ESTIMATION

Hellinger ($* = h$): SpM (semiparametric method) yields the optimally robust IC.

Total variation ($* = v$), parameter dim $k = 1$: SpM yields a suboptimal IC of optimally robust form (for a different radius).

Total variation ($* = v$), parameter dim $k > 1$: SpM eases the problem by exchanging the order of clipping and linear combination of coordinates. The sp-robust IC thus obtained is reasonably robust under MSE.

Contamination ($* = c$): SpM fails, yielding unbounded ICs, $\mathrm{MSE} = \infty$.

TESTING a one-dimensional parameter

Total variation, contamination ($* = c, v$): SpM yields the optimally robust—maxmin—asymptotic tests of *Huber–Strassen* form.

Hellinger ($* = h$): SpM yields a maxmin asymptotic test—although, at finite sample size, no Huber–Strassen pairs exist.

Adaptive constructions by Beran (1976), Kreiß (1987) for ARMA, and by Drost, Klaassen, Wercker (1997, 1998) for ARCH, GARCH, TAR, such that for all F, θ

$$\mathcal{L}_{F,\theta}\big\{(n\mathcal{I}_{F,\theta})^{1/2}(S_n-\theta)\big\}\longrightarrow\mathcal{N}(0,\mathbb{I}_k) \tag{71}$$

Adaptation w.r.t. symmetric innovation distribution.

Nonuniformity Klaassen (1980)

1-dim location, F symmetric, $\mathcal{I}_F^{\mathrm{loc}}<\infty$, $S_n\colon\mathbb{R}^n\to\mathbb{R}$ translation equivariant, sample size n fixed. Then $\forall\,\varepsilon>0\;\forall\,x>0$

$$\inf_{G\in B_c^{\mathrm{s,i}}(F,\varepsilon)} G^n\big\{\big|(n\mathcal{I}_G)^{1/2}S_n\big|\leqslant x\big\}=0<2\,\Phi(x)-1 \tag{72}$$

where $\quad B_c^{\mathrm{s,i}}(F,\varepsilon)=\big\{(1-\varepsilon)F+\varepsilon H\;\big|\;H\text{ symmetric, }\mathcal{I}_H^{\mathrm{loc}}<\infty\big\}$

Extensions to other models? Practical use of adaptive estimators? Robustness? Bickel (1981), (1982), Huber (1996)

Models Location, scale (nonidentifiable), linear regression, ARMA having a finite Fisher information of the form

$$\mathcal{I}_{F,\theta} = \mathcal{I}_F^{\mathrm{loc/sc}} \sigma_F^2 \, \mathcal{K}_\theta \tag{73}$$

Factor $\sigma_F^2 = \int v^2 \, F(dv)$, where $\mu_F = \int v \, F(dv) = 0$, appears only in MA, AR, ARMA.

Huber (1981)

$$\mathcal{I}_F^{\mathrm{loc}} := \sup_{\varphi \in \mathcal{C}_c^1} \Big(\int \dot\varphi \, dF \Big)^2 \Big/ \int \varphi^2 dF \tag{74}$$

$\mathcal{C}_c^1 :=$ all continuosly differentiable functions of compact support. Then: $\mathcal{I}_F^{\mathrm{loc}} < \infty$ iff $dF = f \, d\lambda$, f abs. continuous and $\int (\Lambda_F^{\mathrm{loc}})^2 \, dF < \infty$, in which case $\mathcal{I}_F^{\mathrm{loc}} = \int (\Lambda_F^{\mathrm{loc}})^2 \, dF$.

Ruckdeschel, Rieder (2010)

$$\mathcal{I}_F^{\mathrm{sc}} := \sup_{\varphi \in \mathcal{C}_{1,c}} \Big(\int v \dot\varphi(v) \, dF \Big)^2 \Big/ \int \varphi^2 dF \tag{75}$$

$\mathcal{C}_{1,c} :=$ all functions with continuous derivative of compact support. Then: $\mathcal{I}_F^{\mathrm{sc}} < \infty$ iff $dF = f \, d\lambda$ on $\mathbb{R} \setminus \{0\}$, $v \mapsto v \, f(v)$ is abs. continuous and $\int_{\neq 0} (\Lambda_F^{\mathrm{sc}})^2 \, dF < \infty$ where $\Lambda_F^{\mathrm{sc}} = v \, \Lambda_F^{\mathrm{loc}} - 1$, in which case $\mathcal{I}_F^{\mathrm{sc}} = \int_{\neq 0} (\Lambda_F^{\mathrm{sc}})^2 \, dF$.

$\Longrightarrow$ $\mathcal{I}_F^{\mathrm{loc/sc}}$ is convex and weakly l.s.c. but not u.s.c. !

Kolmogorov metric = sup-norm distance between c.d.f.'s on $\mathbb{R}^k$

4.3 Theorem 1 (location, scale, linear regression, MA)

Assume $\mathcal{I}_F^{\mathrm{loc/sc}} < \infty$, $S_n \colon \mathbb{R}^n \to \mathbb{R}^k$ *any estimator, n fixed. Then* $\forall\, \varepsilon > 0$

$$\sup_{G \in B_c^{\mathrm{s,i}}(F,\varepsilon)} d_\kappa\Big(\mathcal{L}_{G,\theta}\big\{(n\mathcal{I}_{G,\theta}^{1/2}(S_n - \theta)\big\}, \mathcal{N}(0,\mathbb{I}_k)\Big) \geqslant 1 - \frac{1}{2^k} - \kappa_n \qquad (76)$$

where

$$\kappa_n := d_\kappa\Big(\mathcal{L}_{F,\theta}\big\{(n\mathcal{I}_{F,\theta}^{1/2}(S_n - \theta)\big\}, \mathcal{N}(0,\mathbb{I}_k)\Big) \qquad (77)$$

and $B_c^{\mathrm{s,i}}(F,\varepsilon)$ = *all* $(1-\varepsilon)F + \varepsilon H$ *with H symmetric,* $\mathcal{I}_H^{\mathrm{loc/sc}} < \infty$, *and, in case MA, in addition* $\mu_H = 0$, $\sigma_H^2 \in (0,\infty)$.

Remarks a) No equivariance, no symmetry assumptions.

b) Use $G_m = (1-\varepsilon_m)F + \varepsilon_m/2\,\big(\mathcal{N}(-a,\sigma_m^2) + \mathcal{N}(a,\sigma_m^2)\big)$ with $\varepsilon_m, \sigma_m^2 \to 0$ such that $\mathcal{I}_{G_m}^{\mathrm{loc/sc}} \to \infty$ and, in case of MA, $\sigma_{G_m}^2 \to \sigma_F^2$. In these models, the joint law of observations is d_v-continuous in the innovation distribution. Pass to d_κ, which is scale invariant and metrizes weak convergence to $\mathcal{N}(0,\mathbb{I}_k)$.

c) $$1 - 2^{-k} = d_\kappa\big(1_0, \mathcal{N}(0,\mathbb{I}_k)\big)$$

4.3 Theorem 2 (location, scale, linear regression, MA, AR, ARMA)

If $\mathcal{L}_{F,\theta}\{(n\mathcal{I}_{F,\theta}^{1/2}(S_n-\theta)\} \longrightarrow \mathcal{N}(0,\mathbb{I}_k)$ *then, for any* $\varepsilon_n \to 0$,

$$\liminf_{n\to\infty} \sup_{G\in B_c^{\mathrm{s,i}}(F,\varepsilon_n)} d_\kappa\Big(\mathcal{L}_{G,\theta}\{(n\mathcal{I}_{G,\theta}^{1/2}(S_n-\theta)\}, \mathcal{N}(0,\mathbb{I}_k)\Big) \geqslant 1-\frac{1}{2^k} \quad (78)$$

where $B_c^{\mathrm{s,i}}(F,\varepsilon) =$ *all* $(1-\varepsilon)F+\varepsilon H$ *with* H *symmetric,* $\mathcal{I}_H^{\mathrm{loc/sc}} < \infty$, *and, in the cases MA, AR, ARMA, in addition* $\mu_H = 0$, $\sigma_H^2 \in (0,\infty)$. *In the case of AR, ARMA, the functions* S_n *are required to be continuous.*

Remarks a) Adaptive constructions S_n are smooth in the observations.
b) In AR, ARMA, i.e. MA(∞), the joint law of the observations is not d_v-continuous in the innovation distribution. Instead, we derive bounds in L_2 which translate to Prokhorov distance d_π via Strassen (1965)

$$d_\pi\big(\mathcal{L}(Y),\mathcal{L}(X)\big) \leqslant \sqrt{\|Y-X\|_2} \quad (79)$$

Invoke continuity of S_n and, again, switch to d_κ.
c) ARCH? GARCH?

Let $(\mathcal{M}, d)$ be any metric space, balls $B(F,r)$ (open/closed).
For any given function $\alpha\colon \mathcal{M} \to \mathbb{R}$ consider

$$\beta(F,r) := \sup\{\, \alpha(G) \mid G \in B(F,r)\,\} \tag{80}$$

which, for fixed F, increases in r.

4.4 Lemma 1 *The function β satisfies*

$$\beta(F, r-0) \leqslant \liminf_{G\to F} \beta(G,r) \leqslant \limsup_{G\to F} \beta(G,r) \leqslant \beta(F, r+0) \tag{81}$$

with "$=$" except for countably many values of r, depending on F.

Follows from $\quad B(F, r-\delta) \subset B(G,r) \subset B(F, r+\delta) \quad$ if $\quad \delta = d(G,F)$.

Remarks a) Robust risk (max Var, max MSE, min FisherInfo) continuous.
b) Weak dependence of robust estimators and minmaxrisk on the unknown radius r of neighborhoods as a nuisance parameter. Rieder, Ruckdeschel, Kohl (2008)
c) Based on uniform tightness of the empirical process, uniformly asymptotically normal constructions of robust estimators in the independent case,

$$\mathcal{L}_{Q_n^n}\{n^{1/2}(S_n - T(Q_n))\} \longrightarrow \mathcal{N}(0, \mathrm{Cov}\,\varrho_\theta) \tag{82}$$

for all sequences Q_n out of neighborhoods $U_*(\theta, r_n)$ about P_θ, $r_n = r\,n^{-1/2}$, $0 < r < \infty$.
d) Difficulties under dependence; need neighborhoods smaller than (2.30).

Functional $T\colon \mathcal{P} \longrightarrow \mathbb{R}$, defined on a family $\mathcal{P}$ of pm's on some sample space $(\Omega, \mathcal{A})$. Observations $x_1, \dots, x_n$ i.i.d. $\sim$ any $P \in \mathcal{P}$.
Want most accurate tests and confidence statements about $T(P)$.
Fix any $P = P_0 \in \mathcal{P}$. Local alternatives at P within $\mathcal{P}$ are defined by

$$\sqrt{dP_{g,s}} = \bigl(1 + \tfrac{s}{2} g\bigr)\sqrt{dP} + \mathrm{o}(s) \quad \text{as } s \downarrow 0 \tag{83}$$

Tangent set $\mathcal{G}$ of all $g \in L_2(P)$, $g \perp 1$, $P_{g,s} \in \mathcal{P}$ for small $s > 0$.
$\mathcal{G}$ is a cone, vertex at 0 (i.e., $\gamma g \in \mathcal{G}$ whenever $g \in \mathcal{G}$, $\gamma \geqslant 0$), and will be assumed also convex (i.e., $\gamma_1 g_1 + \gamma_2 g_2 \in \mathcal{G}$ for all $g_0, g_1 \in \mathcal{G}$, $\gamma_0, \gamma_1 \geqslant 0$).
Differentiability of T: There is some $\kappa \in L_2(P)$ such that for all $g \in \mathcal{G}$,

$$T(P_{g,s}) = T(P) + s\langle \kappa | g \rangle + \mathrm{o}(s) \quad \text{as } s \downarrow 0 \tag{84}$$

κ is nonunique, but $\bar\kappa =$ the orthoprojection of κ onto $c\ell \operatorname{lin} \mathcal{G}$ is unique.
In addition, let $\hat\kappa =$ the (nonorthogonal) projection of κ onto $c\ell\, \mathcal{G}$.

Literature The $*$-Theorem 25.20, for $\mathcal{G}$ a cone, and LAM-Theorem 25.21, for $\mathcal{G}$ a convex cone, by v.d.Vaart (1998) are both in terms of $\bar\kappa$, not $\hat\kappa$.
For $\mathcal{G}$ a (closed) convex cone, Pfanzagl+Wefelmeyer (1982) state optimal 2-sided confidence bounds, and Janssen (1999) optimal 1-sided tests, in terms of $\hat\kappa$, but, in the proofs, assume $-\mathcal{G} \subset \mathcal{G}$, whence $\mathcal{G}$ linear, resp. $\kappa - \hat\kappa \perp \mathcal{G}$, whence $\hat\kappa = \bar\kappa$.

Characterizations of $\bar{\kappa} \in c\ell \operatorname{lin} \mathcal{G}$ and $\hat{\kappa} \in c\ell\, \mathcal{G}$ as in (46) by, respectively,

$$\kappa - \bar{\kappa} \perp \mathcal{G}\,; \text{ that is, } \langle \kappa | g \rangle = \langle \bar{\kappa} | g \rangle \ \ \forall g \in \mathcal{G} \tag{85}$$

$$\langle \kappa | \hat{\kappa} \rangle = \langle \hat{\kappa} | \hat{\kappa} \rangle \quad \text{and} \quad \langle \kappa | g \rangle \leqslant \langle \hat{\kappa} | g \rangle \ \ \forall g \in \mathcal{G} \tag{86}$$

Bounds based on $\hat{\kappa}$ are sharper since $\|\hat{\kappa}\| < \|\bar{\kappa}\|$ unless $\bar{\kappa} = \hat{\kappa}$.
We shall assume either

a) $\mathcal{G} = \hat{\mathcal{G}}$ a closed convex cone, vertex at 0, OR
b) $\mathcal{G} = \bar{\mathcal{G}}$ a closed linear space.

For comparison, let $P = P_0 \in \hat{\mathcal{P}} \subset \bar{\mathcal{P}}$, where the smaller model $\hat{\mathcal{P}}$ has tangent set a closed convex cone $\hat{\mathcal{G}}$, and the tangent set of the larger model $\bar{\mathcal{P}}$ is the closed linear span $\bar{\mathcal{G}} = c\ell \operatorname{lin} \hat{\mathcal{G}}$ of $\hat{\mathcal{G}}$. We assume that

$$\bar{\kappa} \in \bar{\mathcal{G}} \setminus \hat{\mathcal{G}} \ \ (\text{i.e. } \bar{\kappa} \neq \hat{\kappa}) \ \text{ and } \ \hat{\kappa} \neq 0\,. \tag{87}$$

5.1 Example 1 Let $P = \mathcal{N}(0,1)$ and $\kappa(x) = x$ the identity on the real line; κ may be interpreted the influence curve at P of the expectation functional as well as of the one-sample normal scores rank functional.
As tangent sets at P, consider $\hat{\mathcal{G}}$ and $\bar{\mathcal{G}}$, the convex hull and linear span, respectively, of the two tangents $g_1(x) = \operatorname{sign}(x)$ and $g_2(x) = \mu \operatorname{sign}(x)\, 1_{(|x| \leqslant a)}$, with a and $\mu = \mu_a$ in $(0,\infty)$ such that $\|g_2\| = \|g_1\| = \|\kappa\| = 1$.
By a minimization w.r.t. $a \in (0,\infty)$, it may be achieved that $\|\hat{\kappa}\| = .85\, \|\bar{\kappa}\|$.

Given $P \in \mathcal{P}$, the n i.i.d. observations $x_i \sim Q_n = P_{g,t/\sqrt{n}}$, $n \geqslant 1$, for any $t \in (0,\infty)$, any tangent $g \in \mathcal{G}$ at P, one-sided hypotheses about Q_n are

$J^0: \quad Q_n = P \iff g = 0$ and, employing the functional T,

$$J : \lim_{n\to\infty} \sqrt{n}\,\big(T(Q_n) - T(P)\big) = 0 \iff \langle \kappa | g \rangle = 0$$

$$H : \lim_{n\to\infty} \sqrt{n}\,\big(T(Q_n) - T(P)\big) \leqslant 0 \iff \langle \kappa | g \rangle \leqslant 0$$

$$K : \lim_{n\to\infty} \sqrt{n}\,\big(T(Q_n) - T(P)\big) \geqslant c \iff \langle \kappa | g \rangle \geqslant c \in (0,\infty) \text{ fixed.}$$

In case $P \in \hat{\mathcal{P}} \subset \bar{\mathcal{P}}$ and corresponding tangent sets $\bar{\mathcal{G}} = c\ell \operatorname{lin} \hat{\mathcal{G}}$, the corresponding hypotheses obviously satisfy $J^0 \subset \hat{J} \subset \bar{J}$, $\hat{H} \subset \bar{H}$, $\hat{K} \subset \bar{K}$. Depending on the choice $\mathcal{G} = \hat{\mathcal{G}}$ or $\bar{\mathcal{G}}$, put $\tilde{\kappa} = \hat{\kappa}$, respectively $\tilde{\kappa} = \bar{\kappa}$. We consider sequences $\tau = (\tau_n)$ of tests τ_n at sample size n.

5.2 Theorem 1 [J^0 vs. K] *If* $\limsup_{n\to\infty} \int \tau_n \, dP^n \leqslant \alpha$ *then*

$$\inf_K \limsup_{n\to\infty} \int \tau_n \, dQ_n^n \leqslant \Phi\Big(-u_\alpha + \tfrac{c}{\|\tilde{\kappa}\|}\Big) \tag{88}$$

The power bound is achieved uniquely—up to $\mathrm{o}_{P^n}(n^0)$*—by the tests*

$$\tilde{\tau}_n = \mathbf{1}\Big\{\tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \tilde{\kappa}(x_i) > \|\tilde{\kappa}\|\, u_\alpha\Big\} \tag{89}$$

[$\bar{H}$ vs. $\bar{K}$] *In case* $\mathcal{G} = \bar{\mathcal{G}}$ *moreover* $\sup_{\bar{H}} \limsup_{n\to\infty} \int \bar{\tau}_n \, dQ_n^n \leqslant \alpha$

Proof The closed convex set $G_1 =$ all $g_1 \in \mathcal{G}$ such that $\langle \kappa | g_1 \rangle \geqslant c$ has minimum norm element $q_1 = \tilde{t}\tilde{\kappa}$ with $\tilde{t} = c/\|\tilde{\kappa}\|^2$. Thus 3.4 Theorem 1 provides the unique asymptotic maxmin test $\tilde{\tau}$ for J^0 vs. K. To enlarge the null J^0 to J or H, set $G_0 =$ all $g_0 \in \mathcal{G}$ such that $\langle \kappa | g_0 \rangle = 0$, respectively $\leqslant 0$.
In case $\mathcal{G} = \bar{\mathcal{G}}$, $q_1 = q_{10} = q_1 - q_0$, with $q_0 = 0$, turns out of minimum norm also in $\bar{G}_{10} = \bar{G}_1 - \bar{G}_0$. This is true since $c \leqslant \langle \bar{\kappa} | g_1 \rangle - \langle \bar{\kappa} | g_0 \rangle \implies \|q_{10}\|^2 \leqslant \langle q_{10} | g_{10} \rangle$ for all $g_{10} \in \bar{G}_{10}$, and thus (46). 3.4 Theorem 1 now applies again for $\bar{H}$ vs. $\bar{K}$.
In case $\mathcal{G} = \hat{\mathcal{G}}$, minimization of the norm in $\hat{G}_{10} = \hat{G}_1 - \hat{G}_0$ is yet unsolved.

For $\bar{\mathcal{G}} = c\ell \operatorname{lin} \hat{\mathcal{G}}$ note that $\hat{\kappa} \neq \bar{\kappa}$ iff $\langle \kappa | g \rangle < \langle \hat{\kappa} | g \rangle$ for some $g \in \hat{\mathcal{G}}$.

5.2 Theorem 2 *In case $\mathcal{G} = \hat{\mathcal{G}}$ assume some $g_0 \in \hat{\mathcal{G}}$ such that*

$$\langle \kappa | g_0 \rangle \leqslant 0 < \langle \hat{\kappa} | g_0 \rangle \tag{90}$$

Then $\sup_{\hat{J}} \liminf_{n\to\infty} \int \hat{\tau}_n \, dQ_n = 1$ [level breakdown of $\hat{\tau}$ on $\hat{J}$]

5.2 Example 3 In 5.1 Example 1, although $\hat{\kappa} \neq \bar{\kappa}$, condition (90) is not fulfilled. But replace tangent g_2 there by $g_3(x) = \delta\, 1_{(0,a]}(x) - \eta\, 1_{(a,\infty)}(x) = -g_3(-x)$, $x \geqslant 0$, where the constants may be determined such that $\|g_3\| = 1$. Then g_3 achieves (90).

5.2 Remark 4 [$\bar{\tau}$ for $\hat{H}$ vs. $\hat{K}$] *In case $\bar{\mathcal{G}} = c\ell \operatorname{lin} \hat{\mathcal{G}}$, since $P \in \hat{H} \subset \bar{H}$, the test $\bar{\tau}$ achieves $\sup_{\hat{H}}$ asy.level of $\bar{\tau}_n = \alpha$ and, with $\inf_{\hat{K}}$ attained at $\hat{q}_1 = \hat{t}\hat{\kappa} \in \hat{K} \subset \bar{K}$,*

$$\inf_{\hat{K}} \text{asy.power of } \bar{\tau}_n = \Phi\Big(-u_\alpha + \frac{c}{\|\bar{\kappa}\|}\Big) \;(\bar{\tau} \text{ best?}) < \Phi\Big(-u_\alpha + \frac{c}{\|\hat{\kappa}\|}\Big) \tag{91}$$

Given P, $\mathcal{G}$, T differentiable under $P_{n,g,t} := P_{g,t/\sqrt{n}}$, as in (83), (84). Consider estimator sequences $S = (S_n)$ which, for certain tangents $g \in \mathcal{G}$, asymptotically have median $\geqslant T$ or $\leqslant T$ such that, respectively, $\forall\, t > 0$,

$$\limsup_{n\to\infty} P^n_{n,g,t}\big\{S_n < T(P_{n,g,t})\big\} \leqslant \tfrac{1}{2} \tag{92}$$

$$\limsup_{n\to\infty} P^n_{n,g,t}\big\{S_n > T(P_{n,g,t})\big\} \leqslant \tfrac{1}{2} \tag{93}$$

We assume $\mathcal{G}$ closed, a) a convex cone $\hat{\mathcal{G}}$, or b) a linear space $\bar{\mathcal{G}}$.

5.3 Theorem 1 a) $\mathcal{G} = \hat{\mathcal{G}}$: *If* (92) *holds for* $g = \hat\kappa$, *then* $\forall\, c > 0$

$$\limsup_{n\to\infty} P^n\big\{\,T(P) > S_n - \tfrac{c}{\sqrt{n}}\,\big\} \leqslant \Phi\big(\tfrac{c}{\|\hat\kappa\|}\big) \tag{94}$$

The upper bound is attained by $\hat S$ $\forall\, c > 0$, *iff*

$$\sqrt{n}\,\big(\hat S_n - T(P)\big)_+ = \big(n^{-1/2}\textstyle\sum_{i=1}^n \hat\kappa(x_i)\big)_+ + \mathrm{o}_{P^n}(n^0) \tag{95}$$

b) $\mathcal{G} = \bar{\mathcal{G}}$: *Under* (92) *for* $g = \bar\kappa$ *and* (93) *for* $g = -\bar\kappa$, *then* $\forall\, c', c'' > 0$,

$$\limsup_{n\to\infty} P^n\big\{S_n - \tfrac{c''}{\sqrt{n}} < T(P) < S_n + \tfrac{c'}{\sqrt{n}}\big\} \leqslant \Phi\big(\tfrac{c''}{\|\bar\kappa\|}\big) - \Phi\big(\tfrac{-c'}{\|\bar\kappa\|}\big) \tag{96}$$

The upper bound is attained by $\bar S$ $\forall\, c', c'' > 0$ *iff*

$$\sqrt{n}\,\big(\hat S_n - T(P)\big) = n^{-1/2}\textstyle\sum_{i=1}^n \bar\kappa(x_i) + \mathrm{o}_{P^n}(n^0) \tag{97}$$

Estimators such that, with any $\eta \in L_2(P)$, $\eta \perp 1$,

$$\sqrt{n}\,\big(S_n - T(P)\big) = n^{-1/2} \textstyle\sum_{i=1}^n \eta(x_i) + \mathrm{o}_{P^n}(n^0) \tag{98}$$

for all tangents $g \in \mathcal{G}$, all $t > 0$, are asymptotically normal

$$\mathcal{L}_{P^n_{n,g,t}}\big\{\sqrt{n}\,\big(S_n - T(P_{n,g,t})\big)\big\} \longrightarrow \mathcal{N}\big(t\langle \eta - \kappa | g\rangle, \|\eta\|^2\big) \tag{99}$$

5.4 Corollary 2 [$\mathcal{G} = \bar{\mathcal{G}}$, stability of $\bar{S}$]: *The estimator* $\bar{S}$ *achieves*

$$P^n_{n,g,t}\big\{\bar{S}_n - \tfrac{c''}{\sqrt{n}} < T(P_{n,g,t}) < \bar{S}_n + \tfrac{c'}{\sqrt{n}}\big\} \longrightarrow \Phi\big(\tfrac{c''}{\|\bar{\kappa}\|}\big) - \Phi\big(\tfrac{-c'}{\|\bar{\kappa}\|}\big) \tag{100}$$

for all $g \in \bar{\mathcal{G}}$, $t > 0$, *and all* $c', c'' \geqslant 0$; *in particular, is asy. median unbiased achieving* $\lim_n = \frac{1}{2}$ *in* (92), (93) $\forall\, g \in \bar{\mathcal{G}}$.

In case $\mathcal{G} = \hat{\mathcal{G}} \subset \bar{\mathcal{G}}$ and $\hat{\kappa} \neq \bar{\kappa}$, $\exists\, g_1 \in \hat{\mathcal{G}}$ such that $0 < \langle \kappa | g_1\rangle < \langle \hat{\kappa} | g_1\rangle$.
Consequently, no opt. estimator $\hat{S}$ of form (95) may fulfill condition (93).
Moreover, as lower confidence limit, $\hat{S}$ breaks down under $P_{n,g_1,t}$.

5.4 Proposition 3 [$\mathcal{G} = \hat{\mathcal{G}}$, positive asy. bias of $\hat{S}$]: $\exists\, g_1 \in \hat{\mathcal{G}}$ *such that any optimal estimator* $\hat{S}$ *of form* (95) *satisfies*, $\forall\, c \geqslant 0$,

$$\lim_{t\to\infty} \lim_{n\to\infty} P^n_{n,g_1,t}\big\{T(P_{n,g_1,t}) \geqslant S_n - \tfrac{c}{\sqrt{n}}\big\} = 0 < \Phi\big(\tfrac{c}{\|\hat{\kappa}\|}\big) \tag{101}$$

in particular, violates (93) *as* $\lim_t \lim_n P^n_{n,g_1,t}\big\{\hat{S}_n > T(P_{n,g_1,t})\big\} = 1 > \frac{1}{2}$.

Models:

Rieder, Kohl, Ruckdeschel (2008)

- Location: $y = \theta + u$, $u \sim \mathcal{N}_k(0, \mathbb{I}_k)$, $P_0 = \mathcal{N}_k(0, \mathbb{I}_k) = P$
- Scale ($k = 1$): $y = \sigma u$, $u \sim \mathcal{N}(0, 1) = P_1 = P$
- Regression ($k \geqslant 1$): $\quad y = x\,\theta + u; \quad x,\ u$ sto. indep.

$$u \sim \mathcal{N}(0,1)\ ,\ x \sim K(dx)$$
$$P = P_0(dx, du) = K(dx)\,\mathcal{N}(0,1)(du)$$

  For $\alpha = 2$, coincidence of results with 1-dim. location.
  For $\alpha = 1$, assume K spherically (elliptically) symmetric.
- ARMA(p, q)-models (with shift) are covered, setting $K = \mathcal{L}(H)$.
  Ideal innovations i.i.d. $\sim \mathcal{N}(0, 1)$, then K multivariate normal.
  For $\alpha = 2$, coincidence of results with 1-dim. location.
- ARCH(1): $y_t = \sqrt{1 + \theta y_{t-1}^2}\, u_t$, u_t i.i.d. $\sim \mathcal{N}(0, 1)$
  For $\alpha = 2$, coincidence of results with 1-dim. scale.

Neigborhoods About P :

- (1-dim. location) symmetric contamination nbd of size $s \in [0,1)$:
$$F = (1-s)\,\mathcal{N}(0,1) + s\,H, \qquad H \text{ symmetric}$$
- $r/\sqrt{n}$ - nbds at sample size n:
$$Q_n = (1 - \tfrac{r}{\sqrt{n}})P + \tfrac{r}{\sqrt{n}}\,H$$
(location, unconditional regression; scale: H symmetric)
- conditional regression $r/\sqrt{n}$ - nbds, with radius curve $\varepsilon(x)$:
$$Q_n(du\,|\,x) = (1 - \tfrac{r}{\sqrt{n}}\,\varepsilon(x))\Phi(du) + \tfrac{r}{\sqrt{n}}\,\varepsilon(x)\,H(du\,|\,x),$$
- in time series: contaminated transition probabilities
$Q_n(dy_t|\bar{y}_{t-1})$ where $\bar{y}_{t-1} := y_{t-1}, \dots, y_1$
$$= (1 - \tfrac{r}{\sqrt{n}}\,\varepsilon(\overline{y}_{t-1}))P(dy_t|\bar{y}_{t-1}) + \tfrac{r}{\sqrt{n}}\,\varepsilon(\overline{y}_{t-1})\,H_n(dy_t\mid\overline{y}_{t-1})$$
- $\|\varepsilon\|_\alpha \leqslant 1$: $\mathrm{E}\,\varepsilon \leqslant 1\ (\alpha = 1)$, $\mathrm{E}\,\varepsilon^2 \leqslant 1\ (\alpha = 2)$, $\varepsilon \leqslant 1\ (\alpha = \infty)$
E is taken under the ideal measure P, resp. ideal regressor distr.

Relative Maximum Risk Over Neighborhoods:

We use the estimate which is optimally robust for the neighborhood model of an assumed radius while this radius may not be true.

- relative Var (in Huber[64] model):

  (minmax) M-estimates of location, $\sum_{i=1}^{n} \psi(y_i - S_n) \approx 0$

$$\mathrm{relVar}\,(\psi_{s_0}, s) = \frac{\mathrm{maxVar}\,(\psi_{s_0}, s)}{\mathrm{maxVar}\,(\psi_s, s)}, \qquad 0 \leqslant s < 1$$

- relative MSE ($r/\sqrt{n}$ - neighborhoods, Ri[94]):
  (minmax) asy. linear estimates with influence curves

$$\sqrt{n}\,(S_n - \theta) - n^{-1/2} \sum_{i=1}^{n} \eta(y_i) \longrightarrow 0 \quad \text{in } P\text{-prob.}$$

$$\mathrm{relMSE}\,(\eta_{r_0}, r) = \frac{\mathrm{maxMSE}\,(\eta_{r_0}, r)}{\mathrm{maxMSE}\,(\eta_r, r)}, \qquad 0 \leqslant r < \infty$$

Location (1-dim)

Minimax asymptotic variance

- Minimax M-estimate for $s \in [0,1)$:

$$\psi_s(u) = (-m_s) \vee u \wedge m_s\,, \qquad \tfrac{s}{1-s}\, m_s = \mathrm{E}(|u| - m_s)_+$$

- Maximal asymptotic variance of ψ_{s_0} under s:

$$\mathrm{maxVar}\,(\psi_{s_0}, s) = \frac{(1-s)\,\mathrm{E}\,\psi_{s_0}^2 + s\,m_0^2}{\big[(1-s)\,\mathrm{E}\,\psi'_{m_0}\big]^2}$$

- Median $(s=1)$: $\psi_1(u) = \mathrm{sign}\,(u) = \lim\limits_{s\to 1} \frac{1}{m_s}\psi_s(u)$,

$$\mathrm{maxVar}\,(\psi_1, s) = \tfrac{\pi}{2(1-s)^2}\,, \qquad \mathrm{relVar}\,(\psi_1, s) \longrightarrow 1 \quad (s \to 1)$$

Minimax asymptotic MSE

- Minimax IC for $r \in [0, \infty)$: $\quad \eta_r(u) = A_r\, u \min\left\{1, \frac{c_r}{|u|}\right\}$,

$$1 = A_r \operatorname{E} u^2 \min\left\{1, \frac{c_r}{|u|}\right\}, \qquad r^2\, c_r = \operatorname{E}\left(|u| - c_r\right)_+$$

- Median $(r = \infty)$: $\eta_\infty(u) = b_{\min} \operatorname{sign}(u)$,
- Minimal bias (of ALE): $b_{\min} = \sqrt{\frac{\pi}{2}}$
- Maximal MSE of η_{r_0} under r:

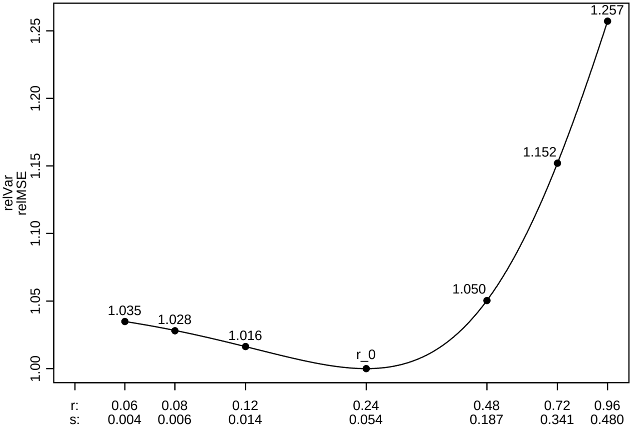
$$\operatorname{maxMSE}(\eta_{r_0}, r) = A_{r_0}^2\ \operatorname{E} \min\left\{u^2,\ c_{r_0}^2\right\}\ +\ r^2\ A_{r_0}^2 c_{r_0}^2$$

Coincidence

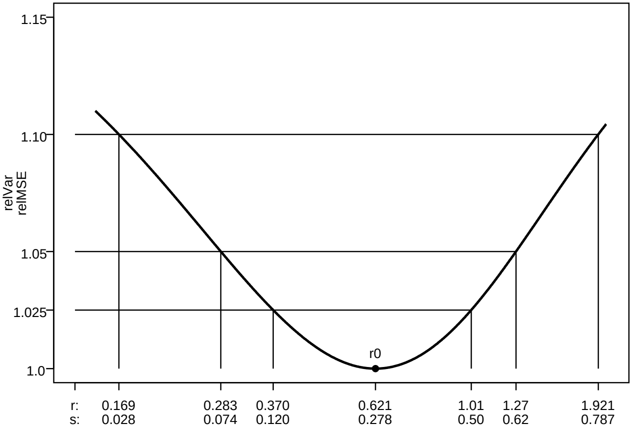
$$\begin{aligned}(1 - s) \operatorname{maxMSE}(\eta_{r_0}, r) &= \operatorname{maxVar}(\psi_{s_0}, s)\\ \Longrightarrow \operatorname{relVar}(\psi_{s_0}, s) &= \operatorname{relMSE}(\eta_{r_0}, r)\end{aligned}$$

where r and s correspond via $\quad s = r^2/(1 + r^2)$.

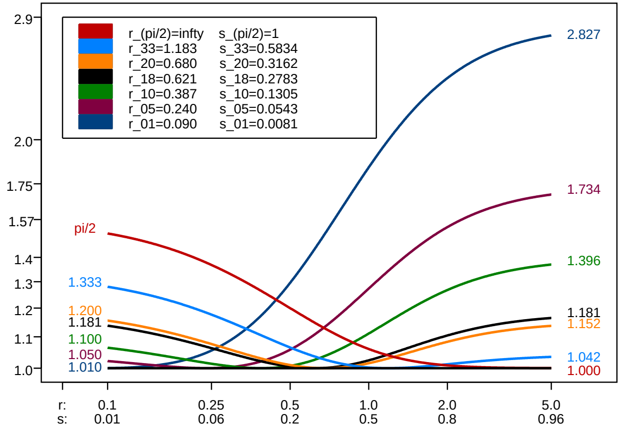
relVar, relMSE: 1-Dimensional Location (Var=1.050 at r, s = 0)



relVar, relMSE: 1-Dimensional Location (Var = 1.181 at r, s = 0)



1-dim. Location: relMSE, relVar vs. r, rsp. s=r^2/(1+r^2)



Location (k-dim)

Minimax asymptotic MSE

- Minimax IC for $r \in [0, \infty)$: $\eta_r(u) = \alpha_r u \min\left\{1, \frac{c_r}{|u|}\right\}$,

$$k = \alpha_r \operatorname{E}|u|^2 \min\left\{1, \frac{c_r}{|u|}\right\}, \qquad r^2 c_r = \operatorname{E}(|u| - c_r)_+$$

- min-L_1 ($r = \infty$): $\sum_{i=1}^n |u_i - \hat\theta|_2 = \min_\theta !$ $\quad \eta_\infty(u) = b_{\min} \frac{u}{|u|}$
- Minimal bias (of ALE):

$$b_{\min} = \frac{k}{\operatorname{E}|\Lambda|} = \frac{k\,\Gamma(\frac{k}{2})}{\sqrt{2}\,\Gamma(\frac{k+1}{2})}, \qquad \frac{b_{\min}}{\sqrt{k}} \to 1, \ \frac{\operatorname{E}|\eta_\infty|^2}{k} \to 1$$

- Maximal MSE of η_{r_0} under r:

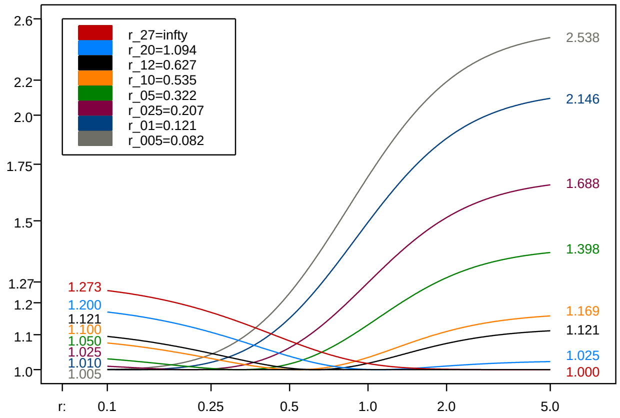
$$\operatorname{maxMSE}(\eta_{r_0}, r) = \alpha_{r_0}^2 \operatorname{E}\min\{|u|^2, c_{r_0}^2\} + r^2 \alpha_{r_0}^2 c_{r_0}^2$$

- Relative MSE: η_∞ becomes radius–minimax

$$\lim_{k\to\infty} \frac{\operatorname{maxMSE}(\eta_{r_0}, r)}{\operatorname{maxMSE}(\eta_\infty, r)} = 1$$

uniformly in $0 \leqslant r_0, r \leqslant$ any $r_1 < \infty$.

relMSE: 2-Dimensional Location



Regression (k-dim)

Minimax asymptotic MSE ($* = c, \alpha = 1$)

- Minimax IC for $r \in [0, \infty)$: $\eta_r(x,u) = \alpha_r x\, u \min\left\{1, \frac{c_r}{|xu|}\right\}$,

$$k = \alpha_r \operatorname{E} |x|^2 u^2 \min\left\{1, \frac{c_r}{|xu|}\right\}, \qquad r^2\, c_r = \operatorname{E}\left(|xu| - c_r\right)_+$$

- weighted min-L_1 ($r = \infty$): $\eta_\infty(x,u) = b_{\min} \frac{x}{|x|} \operatorname{sign}(u)$
- Minimal bias (of ALE): $b_{\min} = \frac{k}{\operatorname{E}|\Lambda|} = \sqrt{\frac{\pi}{2}}\, \frac{k}{\operatorname{E}|x|}$
- Maximal MSE of η_{r_0} under r:

$$\operatorname{maxMSE}(\eta_{r_0}, r) = \alpha_{r_0}^2 \operatorname{E} \min\{|x|^2 u^2,\ c_{r_0}^2\} \ +\ r^2\, \alpha_{r_0}^2 c_{r_0}^2$$

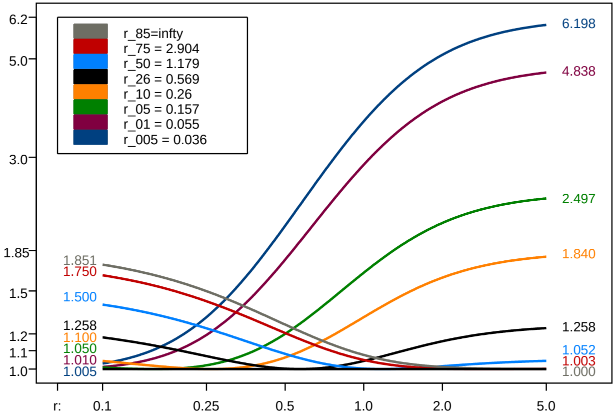
- RelMSE same for all θ, but depends on $K(dx)$.

Convergence to 1-dim. location

$$\lim_{k\to\infty} \operatorname{relMSE}(\eta_{r_0}, r) = \operatorname{relMSE}(\eta_{r_0}^{\mathrm{1loc}}, r)$$

uniformly for $0 \leqslant r_0, r \leqslant$ any $r_1 < \infty$, as $k \to \infty$.
In case ($* = c, \alpha = 2$) limit attained $\forall\, k \geqslant 1$.

relMSE: Regression (*=c, alpha=1, K normal, dim=3)



Scale (1-dim)

Minimax asymptotic MSE for $(* = c)$ contamination balls

- Minimax IC for $r \in [0, \infty)$: $\quad \eta_r(u) = A_r(u^2 - \alpha_r^2) \min\left\{1, \dfrac{c_r}{|u^2 - \alpha_r^2|}\right\}$

$$\begin{aligned} 0 &= \mathrm{E}(u^2 - \alpha_r^2) \min\left\{1, \tfrac{c_r}{|u^2 - \alpha_r^2|}\right\}, \\ A_r^{-1} &= \mathrm{E}(u^2 - \alpha_r^2)^2 \min\left\{1, \tfrac{c_r}{|u^2 - \alpha_r^2|}\right\}, \\ r^2 c_r &= \mathrm{E}(|u^2 - \alpha_r^2| - c_r)_+ \end{aligned}$$

- MAD $(r = \infty)$: $\quad \eta_\infty(u) = b_{\min} \operatorname{sign}(|u| - \alpha_\infty), \quad \hat{\theta} = \alpha_\infty^{-1} \operatorname{med}(|u_i|)$
- Minimal bias (of ALE): $\quad b_{\min} = (4\alpha_\infty \varphi(\alpha_\infty))^{-1} = 1.166$
- $0 < \alpha_r$ decreasing from $\alpha_0 = 1$ to $\alpha_\infty := \Phi^{-1}(3/4) = 0.674$
- clipping of $|u|$ only from above for $r \leqslant 0.92$;
  clipping of $|u|$ from below and above iff $r \geqslant 0.92$
- For $r_0, r \in [0, \infty)$, the maximal MSE is

$$\operatorname{maxMSE}(\eta_{r_0}, r) = A_{r_0}^2 \, \mathrm{E} \min\{|u^2 - \alpha_{r_0}^2|^2, \, c_{r_0}^2\} \; + \; r^2 A_{r_0}^2 c_{r_0}^2$$

Minimax asymptotic MSE for $(* = v)$ contamination balls

- Minimax IC for $r \in [0, \infty)$:

$$\eta_r(u) = A_r\{[g_r \vee u^2 \wedge (g_r + c_r)] - 1\}$$
$$0 = \mathrm{E}(g_r - u^2)_+ - \mathrm{E}(u^2 - g_r - c_r)_+$$
$$1 = A_r\,\mathrm{E}\,u^2\big\{[g_r \vee u^2 \wedge (g_r + c_r)] - 1\big\}$$
$$r^2\,c_r = \mathrm{E}(g_r - u^2)_+$$

- MADv $(r = \infty)$:

$$\eta_\infty(u) = \omega_v^{\min}\big\{P(|u| < 1)\,1(|u| > 1) - P(|u| > 1)\,1(|u| < 1)\big\}$$

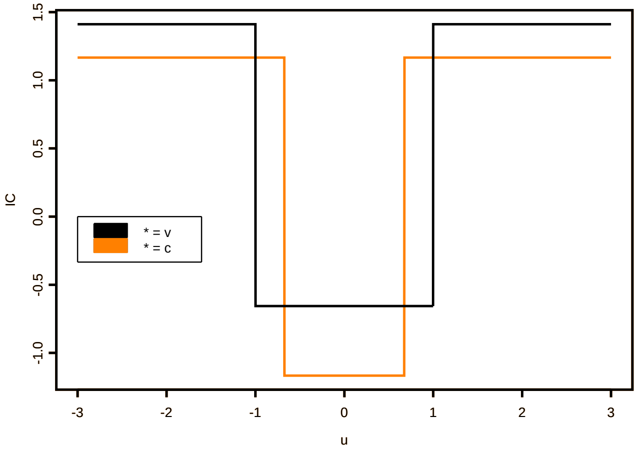
- Minimal bias (of ALE):

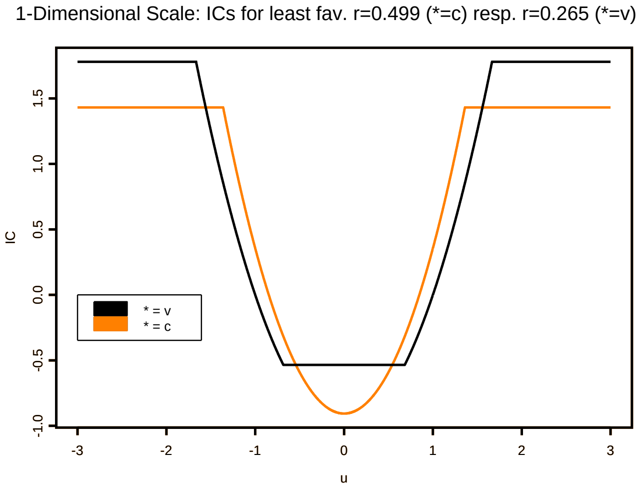
$$\omega_v^{\min} = (\mathrm{E}\,\Lambda_+)^{-1} = \sqrt{\tfrac{\pi}{2}\,e} \approx 2.066$$

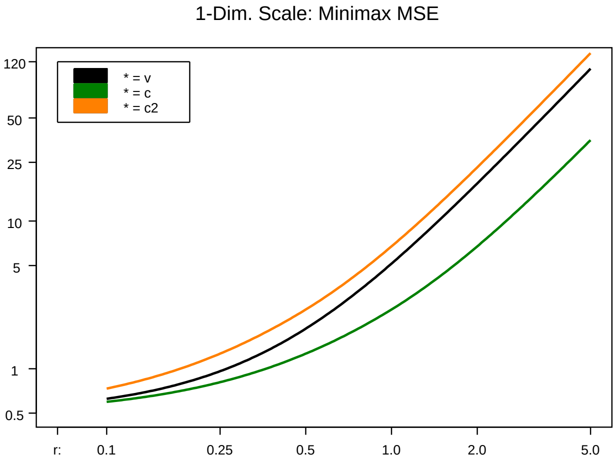
- clipping of $|u|$ always from above and below
- For $r_0, r \in [0, \infty)$, the maximal MSE is

$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = A_{r_0}^2\,\mathrm{E}\,\big\{[g_{r_0} \vee u^2 \wedge (g_{r_0} + c_{r_0})] - 1\big\}^2 + r^2 A_{r_0}^2 c_{r_0}^2$$

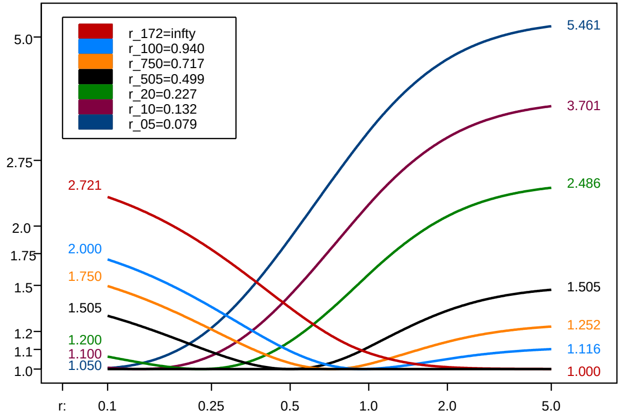
1-Dimensional Scale: IC-comparison for lower cases

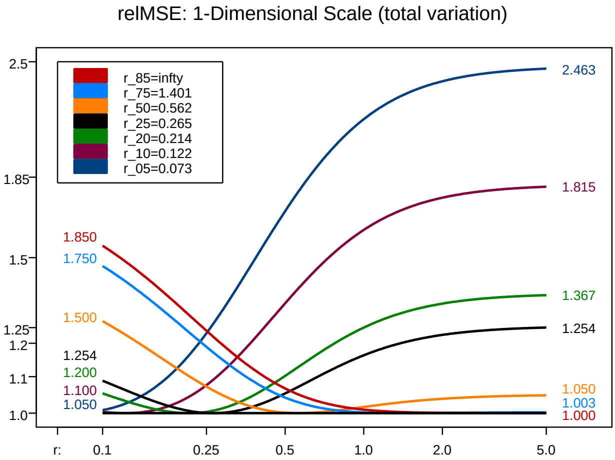



1-Dimensional Scale: ICs for least fav. r=0.499 (*=c) resp. r=0.265 (*=v)
IC
u
-1.0
-0.5
0.0
0.5
1.0
1.5
-3
-2
-1
0
1
2
3
* = v
* = c


1-Dim. Scale: Minimax MSE
* = v
* = c
* = c2
120
50
25
10
5
1
0.5
r:
0.1
0.25
0.5
1.0
2.0
5.0

relMSE: 1-Dimensional Scale (contamination)




relMSE: 1-Dimensional Scale (total variation)
r_85=infty
r_75=1.401
r_50=0.562
r_25=0.265
r_20=0.214
r_10=0.122
r_05=0.073
2.5
1.85
1.5
1.25
1.2
1.1
1.0
1.850
1.750
1.500
1.254
1.200
1.100
1.050
2.463
1.815
1.367
1.254
1.050
1.003
1.000
r:
0.1
0.25
0.5
1.0
2.0
5.0

Summary

1) Estimation of the unknown radius hardly pays, provided one employs the radius-minimax estimator. The increase of its risk with respect to the radius-optimal procedure is moderate to small.
In all our models, it is $\leqslant 12.5\%$, if the radius may be specified to belong to some interval $[\frac{1}{3}r, 3r]$ for any r.

2) The minimax radii are small: 5–6% contamination, at sample size 100.

3) The radius-minimax estimator for completely unknown radius stays the same for a variety of convex risks which are homogeneous in bias and (square root) variance; e.g., L_p-loss, confidence levels. Ruckdeschel, Rieder (2004)

Rieder, Kohl Ruckdeschel (2008)

Beran, R.J. (1974): Asymptotically efficient and adaptive rank estimates in location models. *Ann. Statist.* **2** 63–74.
Beran, R.J. (1976): Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.* **11** 432–452.
Bickel, P.J. (1982): On adaptive estimation. *Ann. Statist.* **10** 647–671.
Bickel, P.J. et al. (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.
Birgé, L. (1980): Approximation dans les espaces métrique et théorie de l'estimation. Ph.D. Thesis, University of Paris.
Drost, F.C., Klaassen, C.A.J., Werker, B.J.M. (1997): Adaptive estimation in time-series models. *Ann. Statist.* **25** 786–817.
Fischer, J. (2006): *Robuste Schätzung im semiparametrischen Mixture-Modell*. Diploma thesis, U Bayreuth.
Hampel, F.R. et al. (1986): *Robust Statistics—The Approach Based on Influence Functions*. Wiley, New York.
Huber, P.J. (1981): *Robust Statistics*. Wiley, New York.
Huber, P.J. and Strassen, V.(1973): Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263.
Huber–Carol, C. (1970): *Étude asymptotique de tests robustes*. Thèse de Doctorat, ETH Zürich.
Kohl, M. (2005): *Numerical Contributions to the Asymptotic Theory of Robustness*. PhD thesis. University of Bayreuth.
Kohl, M. and Ruckdeschel, P. (2008): ROptEst: Opt. robust estimation. *R ver. 0.6.3*. URL http://robast.r-forge.r-project.org
Kreiss, J.-P. (1987): On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.
Pfanzagl, J., Wefelmeyer, W. (1982): *Contributions to a General Asymptotic Statistical Theory*. Springer LN in Statistics #13.
Rieder, H. (1977): Least favorable pairs for special capacities. *Ann. Statist.* **5** 909–921.
Rieder, H. (1978): A robust asymptotic testing model. *Ann. Statist.* **6** 1080–1094.
Rieder, H. (1994): *Robust Asymptotic Statistics*. Springer, New York.
Rieder, H. (2000): Neighborhoods as nuisance parameters? Robustness vs. semiparametrics. Discussion Paper Nr. 25, SFB 373.
Rieder, H. (2000): One-sided confidence about functionals over tangent cones. Discussion Paper Nr. 26, SFB 373.
Rieder, H., Kohl, M. and Ruckdeschel, P. (2008): The cost of not knowing the radius. *Stat. Meth.& Appl.* **17** 13–40.
Ruckdeschel, P. and Rieder, H. (2004): Optimal influence curves for general loss functions. *Statistics&Decisions* **22** 201223.
Ruckdeschel, P. (2006): A motivation for $1/\sqrt{n}$-shrinking neighborhoods. *Metrika* **63** 295-307.
Ruckdeschel, P., Hable, R., Rieder, H. (2010): Optimal robust ICs in semiparametric regression. *JSPI* **140** 226–245.
Ruckdeschel, Kohl, Rieder (2010): Infinitesimally Robust estimation in general smooth models. *Stat. Meth.& Appl.* **19** 333–354.
Ruckdeschel, P. and Rieder, H. (2010): Fisher information of scale. *Statistics and Probability Letters* **80** 1881-1885.
Shen, L.Z. (1994): Optimal robust estimates for semiparametric symmetric location models. *Statistics&Decisions* **12** 113–124.
Shen, L.Z. (1995): On optimal B-robust influence functions in semiparametric models. *Ann. Statist.* **23** 968–989.
Stabla, Th. (2005): *Robuste adaptive Schätzung*. Diploma thesis, U Bayreuth.
Stein, C. (1956): *Efficient nonparametric estimation and testing*. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I , 187–195. UC Press, Berkeley and Los Angeles.
Stone, C. (1975): Adaptive maximum likelihood estimation for a location parameter. *Ann. Statist.* **3** 267–284.
van der Vaart, A.W. (1998): *Asymptotic Statistics*. CUP, Cambridge.