

The Cost Of Not Knowing The Radius

Helmut Rieder, Matthias Kohl, Peter Ruckdeschel

Department of Mathematics, University of Bayreuth,
e-mail: helmut.rieder@uni-bayreuth.de

The date of receipt and acceptance will be inserted by the editor

Abstract Robust Statistics considers the quality of statistical decisions in the presence of deviations from the ideal model, where deviations are modelled by neighborhoods of a certain size about the ideal model. We introduce a new concept of optimality (radius-minimaxity) if this size or radius is not precisely known:

For this notion, we determine the increase of the maximum risk over the minimax risk in the case that the optimally robust estimator for the false neighborhood radius is used. The maximum increase of the relative risk is minimized in the case that the radius is known only to belong to some interval $[r_l, r_u]$. We pursue this minmax approach for a number of ideal models and a variety of neighborhoods. Also, the effect of increasing parameter dimension is studied for these models.

The minimax increase of relative risk in case the radius is completely unknown, compared with that of the most robust procedure, is 18.1% vs. 57.1% and 50.5% vs. 172.1% for one-dimensional location and scale, respectively, and less than $1/3$ in other typical contamination models. In most models considered so far, the radius needs to be specified only up to a factor $\rho \leq \frac{1}{3}$, in order to keep the increase of relative risk below 12.5%, provided that the radius-minimax robust estimator is employed. The least favorable radii leading to the radius-minimax estimators turn out small: 5%–6% contamination, at sample size 100.

Key words Symmetric location and contamination; infinitesimal asymmetric neighborhoods; total variation, contamination; asymptotically linear estimators; influence curves; maximum asymptotic variance and mean

Correspondence to: Department of Mathematics, University of Bayreuth, NW II, D-95440 Bayreuth, Germany

square error; relative risk; inefficiency; least favorable radius; radius–minimax robust estimator; location, scale, regression models.

AMS/MSC-2000 classification: 62F35, 62G35, 62G05, 62J05.

1 Introduction and Summary

1.1 Statistical Folklore

It has been said that robust procedures depend but little on the tuning constants regulating the degree of robustness. However, the good-natured dependence has hardly ever been documented nor has it been investigated theoretically.

In robustness theory, the tuning constants are determined by the neighborhood radius via certain implicit equations, and the radius appears as a one-dim. nuisance parameter of robust neighborhood models. More abstractly, the model deviations may be treated as values of an infinite dim. nuisance parameter; cf. Rieder[HR henceforth] (2000). But the more elementary case of just the radius has not been considered by mathematical and semiparametric statistics.

Some textbooks create an impression contrary to data-analytic experience. For example, Witting and Müller–Funk (1995; Anmerkung 6.44) declare the choice of the clipping constant c to be of decisive importance and, in their Beispiel 7.4.5, state that the radius is unknown in practice. Since, however, the two quantities are connected, a light of arbitrariness is shed on robust procedures.

As for a theoretical indication of the weak dependence, the adaptive clipping by Beran (1981) and HR (1994; Remarks 6.4.6 and 6.4.9) may be recalled. The adaptive modification of clipping constants by means of a goodness-of-fit statistic would not show up in the asymptotic results. On closer inspection, this is caused by these clipping constants tending to infinity. Thus, the construction is essentially bound to infinitesimal Hellinger balls, which are no gross-error neighborhoods; cf. Bickel (1981; Théorème 8) and HR (1994; Example 6.1.1).

1.2 Our Approach

Instead, we propose to evaluate the maximum risk of the estimator which is optimally robust for a neighborhood of radius r_0 over a neighborhood of radius r , and relate the result to the minimax risk for that radius r . On division, the *inefficiency* is obtained—the limit of the ratio of sample sizes such as to achieve the same accuracy asymptotically. The inefficiency as a function of r is called the inefficiency curve of the estimator (1 at $r = r_0$). Inefficiency minus 1 is termed *subefficiency* (0 at $r = r_0$).

It can be proven (Theorem 2.1), that the inefficiency curves are bowl-shaped, smoothly increasing from the value 1 at $r = r_0$ towards both sides

to two relative maxima at the interval boundaries. Determination of r_0 so as to equate both boundary values will minimize the maximal subefficiency over r in the respective estimator class (M -estimates, asymptotically linear estimators).

The radius r_0 may be termed *least favorable* in the sense that the corresponding optimally robust estimator—besides being minimax for the particular neighborhood of radius r_0 —is *radius-minimax*, minimizing the maximal subefficiency over the radius range. It is the recommended robust estimator in case that the true radius r is unknown except to belong to the radius interval.

Remark 1.1 (a) There is no saddle point though. The subefficiency of the radius-minimax estimator is elsewhere worse (i.e., larger) than at r_0 , where it is 0, and equally worst (i.e., maximum) at the boundaries of the radius interval.

(b) Our approach is not restricted to the infinitesimal setup, as introduced in Subsection 2.1.2(b). The concept in principle applies to arbitrary models, risks, and neighborhood types. Only the numerical evaluations may be easier in some specifications than in others. \square

To appreciate these new notions, consider the simplest possible example, one-dim. location in the setup of infinitesimal neighborhoods: we obtain a least favorable (starting) radius of $r_0 = .62$, which is just 6.2% contamination at sample size $n = 100$. The minimax subefficiency is 18.1% and leads to an M -estimate as corresponding radius-minimax procedure with clipping height .719, which is very close to the H07-estimate of Andrews et al. (1972). Three general conclusions may be drawn from our results:

(i) The minimax subefficiency is small. Small in comparison with the most robust estimators, and small for practical purposes. Consistent estimation of the radius from the data hence seems neither necessary nor worthwhile—however under the provision that the radius-minimax robust estimator is employed.

(ii) The least favorable radii are small. This surprising fact seems to confirm Huber (1996; Sec. 28, p 61), who distinguishes robustness from diagnostics by its purpose to safeguard against—as opposed to find and identify—deviations from the assumptions; in particular, to safeguard against deviations below or near the limits of detectability. Like Huber (1996), the small least favorable radii we obtain might question the breakdown literature, which is concerned only with (stability under) large contamination and, at most, (efficiency under) zero contamination.

(iii) The radius-minimax procedure does not depend on the risk. With respect to contamination or total variation neighborhood systems about a given parametric model, in a reasonably large class of convex risks which are homogeneous in bias and (square root of) variance, compare the assumptions to Theorem 2.1 below, the radius-minimax procedure for completely unknown radius does not depend on the risk (Thm. 2.1(b)). Thus, in

these cases, a universal optimally robust procedure is obtained that neither depends on a radius nor on a loss function.

1.3 Comparison With Semiparametrics

The radius, albeit a one-dim. quantity, in connection with the robust neighborhoods has infinite dim. features, and so a comparison with a basic semiparametric model is recommended.

We assume the classical univariate location model with unknown symmetric error distribution F and density f of finite Fisher information $\mathcal{I}_F^{\text{loc}} = \int (\Lambda_F^{\text{loc}})^2 dF$, where $\Lambda_F^{\text{loc}} = -f'/f$, and consider the location M -estimate defined by some odd function $\psi_0: \mathbb{R} \rightarrow \mathbb{R}$; for example, $\psi_0 = \Lambda_{F_0}^{\text{loc}}$ for some other such law F_0 .

Then, provided certain weak regularity conditions are satisfied by ψ_0 and F , the M -estimate under observations i.i.d. $\sim F$ is asymptotically normal with asymptotic variance

$$\text{Var}_{\text{loc}}(\psi_0, F) = \frac{\int \psi_0^2 dF}{\left(\int \psi_0 \Lambda_F^{\text{loc}} dF\right)^2} \in (0, \infty) \quad (1.1)$$

However, if ψ_0 , on some nondegenerate interval, is absolutely continuous with a bounded derivative, we can show that

$$\sup_{F \in U_c^{s,i}(F_0, \varepsilon)} \text{Var}_{\text{loc}}(\psi_0, F) \cdot \mathcal{I}_F^{\text{loc}} = \infty \quad \forall \varepsilon \in (0, 1) \quad (1.2)$$

where $U_c^{s,i}(F_0, \varepsilon) = \{ (1 - \varepsilon)F_0 + \varepsilon H \mid H \text{ symmetric, } \mathcal{I}_H^{\text{loc}} < \infty \}$.

Thus, if only the nuisance parameter F changes arbitrarily little (in L_1), the inefficiency of the location M -estimate defined by ψ_0 may become infinite. For the proof, and the similar result for scale, cf. HR (2001)

In comparison with the radius as a nuisance parameter in robust statistics—the results of this study—the highly unstable situation is just the opposite. Further relations with semiparametrics are derived in HR (2000)

1.4 Uniform Convergence To The Normal Limit

Uniform convergence to the normal limit is an issue, in particular in connection with the large families of probabilities which make the models in semiparametrics and robustness, respectively.

But the desirable uniformity cannot be achieved by adaptive and fully efficient estimation. Using equivariance, Klaassen (1980) derives such a finite-sample result for the one-dim. location model. Consequences are noted by Bickel (1982; Remark 5.5) and Huber (1996; 1996; Sec. 28). Bickel (1981; Note, p 51) asks for extensions. Pfanzagl and Wefelmeyer (1982; Sec. 9.4)

derive an asymptotic version for real-valued smooth functionals. The following extension to k -dim. linear regression provides the asymptotic lower bound $1 - 2^{-k}$ in Kolmogorov distance.

Consider the regression model $P_\theta(dx, dy) = f(y - x'\theta) dy K(dx)$ with unknown parameter $\theta \in \mathbb{R}^k$, univariate error law $F(du) = f(u) du$ of finite Fisher information of location, and regressor law K such that the $k \times k$ matrix $\mathcal{K} = \int xx' K(dx)$ is regular. Then, for fixed F , the model is L_2 -differentiable at each θ with scores function $\Lambda_\theta(x, y) = \Lambda_F^{\text{loc}}(y - x'\theta) x$ and Fisher information $\mathcal{I}_\theta = \mathcal{I}_F^{\text{loc}} \mathcal{K}$.

By definition, the standardized laws of an adaptive estimator (S_n) are asymptotically standard normal such that, for each main/nuisance parameter pair (θ, F) ,

$$\mathcal{V}_{\theta, F}^n := \mathcal{L}_{\theta, F} \left\{ \sqrt{n} \mathcal{I}_{\theta, F}^{1/2} (S_n - \theta) \right\} \xrightarrow{\text{w}} \mathcal{N}(0, \mathbb{I}_k) \quad (1.3)$$

weakly, as $n \rightarrow \infty$, where $\mathbb{I}_k = k \times k$ identity matrix.

Fix θ and F_0 . Then, if (S_n) is an adaptive estimator, and $\varepsilon_n \in (0, 1)$ any sequence tending to 0, we can show that, in Kolmogorov distance d_κ ,

$$\liminf_{n \rightarrow \infty} \sup_{F \in U_c^{s,i}(F_0, \varepsilon_n)} d_\kappa(\mathcal{V}_{\theta, F}^n, \mathcal{N}(0, \mathbb{I}_k)) \geq 1 - \frac{1}{2^k} \quad (1.4)$$

where $U_c^{s,i}(F_0, \varepsilon_n) = \{ (1 - \varepsilon_n)F_0 + \varepsilon_n H \mid H \text{ symmetric, } \mathcal{I}_H^{\text{loc}} < \infty \}$.

Remark 1.2 The result is proved for more general i.i.d. models of location or scale structure in HR (2001); it is shown to hold also for MA(q)-models with innovation distribution F . A somewhat weaker extension to AR(p)- and ARMA(p, q)-models still renders convergence of the adaptive estimators of Beran (1976) and Kreiss (1987) nonuniform as above. \square

On second look not so much the estimators are to be blamed for (1.2) and (1.4) non-uniform asymptotic normality; actually, the law of any estimator S_n is uniformly continuous in total variation. Rather the standardization by Fisher information in (1.3) should be questioned because Fisher information of location/scale is vaguely lower semicontinuous, but not upper semicontinuous even in total variation.

In robust statistics, on the contrary, risk is evaluated uniformly by passing to maximum asymptotic variance/MSE and minimum Fisher information, respectively, over neighborhoods. By simple set inclusions, however, the passage to the supremum $g(x, r) = \sup\{f(y) \mid y \in B(x, r)\}$ of any function f over balls $B(x, r)$ already implies continuity relative to these balls of $g(\cdot, r)$ at x for almost all radii; namely, for those r such that $g(x, r - 0) = g(x, r + 0)$.

Uniform weak convergence of optimally robust estimators over neighborhoods as proved by Beran (1981), Millar (1981), Bickel (1981; Théorème 5), Huber (1981), and HR (1994; Chap. 6) also underlies the present investigation of asymptotic risk.

In the last analysis, both uniform convergence and the availability of a low-cost minimax strategy against misspecification of the radius seem to be consequences of the uniform risk evaluation over total variation type neighborhoods in robustness theory, and advantages of robust statistics in practice.

Outline In Section 2 of the paper, the theoretical setup is formulated and the concepts are formally defined. In Section 3, for a list of ideal models and neighborhood systems, the optimally robust estimators and their risk functions are determined. The numerical algorithms are described in Section 4. An appendix contains mathematical proofs in section A, tabulated numerical results in Section B, as well as a selection of four plots.

2 General concept

2.1 Setup

2.1.1 Ideal model For a measurable space (Ω, \mathcal{A}) and \mathcal{M}_1 , the set of probability measures on \mathcal{A} , let $\mathcal{P} = \{P_\theta, \theta \in \Theta\} \subset \mathcal{M}_1$ be a parametric model with open parameter set $\Theta \subset \mathbb{R}^k$. We assume \mathcal{P} to be L_2 -differentiable with derivative Λ_θ and Fisher information $\mathcal{I}_\theta > 0$ at every $\theta \in \Theta$. In the ideal model, the observations X_i are i.i.d. according to P_θ for some (unknown) $\theta \in \Theta$. We want to estimate $t(\theta)$, the value of a differentiable transformation $t: \Theta \rightarrow \mathbb{R}^p$ with $dt(\theta) = D_\theta$ for some matrix $D_\theta \in \mathbb{R}^{p \times k}$ with $\text{rk } D_\theta = p \leq k$. For estimators, we assume asymptotically linear estimators (ALE's); that is, sequences $S = (S_n)$ of estimators $S_n(X_1, \dots, X_n)$ such that

$$\sqrt{n} (S_n - t(\theta)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_\theta(X_i) + o_{P_\theta^n}(n^0) \quad (2.1)$$

as $n \rightarrow \infty$, for some (partial) influence curve (pIC) $\eta_\theta \in \Psi_\theta^{D_\theta}$, where

$$\Psi_\theta^{D_\theta} = \{ \eta_\theta \in L_2^p(P_\theta) \mid E_\theta \eta_\theta = 0, E_\theta \eta_\theta \Lambda_\theta' = D_\theta \} \quad (2.2)$$

As for this setup and definition, cf. HR (1994; section 4.2). In the somewhat more restricted one-dim. location setup of Huber (1964); i.e., symmetric P_θ and monotone Λ_θ , the procedures are further specialized to location M-estimators.

In the sequel, expectation will always be taken under the fixed ideal model distribution $P = P_\theta$; similarly, we put $\Lambda = \Lambda_\theta$ (scores), $\mathcal{I} = \mathcal{I}_\theta$ (Fisher information) and we omit θ whenever it is possible.

2.1.2 Neighborhoods [Nbd's in the sequel] As general in Robust Statistics, these ideal models are enlarged to nbd's; more specifically, we consider

symmetric contamination nbd's \mathcal{V} of fixed size $s \in [0, 1)$ about the ideal P , assumed symmetric about zero, consisting of all convex combinations

$$Q = (1 - s)P + sH \quad (2.3)$$

with arbitrary unknown probability H , symmetric about 0. These fixed nbd's, whose size does not depend on the sample size, are bound to one-dim. location and Huber's (1964) minimax asymptotic variance approach.

infinitesimal nbd's $\mathcal{U}_n = \mathcal{U}_*(\theta, r/\sqrt{n})$ of starting radius $r \in [0, \infty)$, given as the sequence of shrinking contamination ($* = c$) nbd's about P at sample size n , consisting of all

$$Q_n = (1 - r_n)P + r_n H_n \quad (2.4)$$

where H_n may be arbitrary unknown probabilities, and $r_n = r/\sqrt{n}$.

Likewise, infinitesimal total variation ($* = v$) nbd's are the sequences of shrinking balls about P , of radius $r_n = r/\sqrt{n}$ at sample size n , defined by

$$d_v(Q_n, P) \leq r_n, \quad \text{where } d_v(Q, P) = \sup_{A \in \mathcal{A}} |Q(A) - P(A)| \quad (2.5)$$

Infinitesimal nbd's are employed in the location, scale, and regression models 3.1.1, 3.1.2, and 3.1.4. In the scale model 3.1.2, they may as well be restricted by symmetry (that is, P , H_n , and Q_n all symmetric). In regression, these nbd's about $P(dx, du) = \Phi(\sigma_u^{-1} du) K(dx)$ are termed unconditional, or errors-in-variables, nbd's since also the regressor marginal is subject to distortion.

conditional regression nbd's $\mathcal{U}_{*,\alpha}$: Contrary to unconditional nbd's, they keep the ideal regressor distribution K , and only the conditional error law given x may change; to any Markov kernel $Q_n(du|x)$ which, for each x , is in the nbd about the ideal $\Phi(\sigma_u^{-1} du)$ of radius $r\varepsilon(x)/\sqrt{n}$. The function $\varepsilon: \mathbb{R}^k \rightarrow [0, \infty)$, which weights the radius depending on the regressor, is called radius curve.

We employ conditional, or error-free-variables, nbd's with varying radius curves ε subject to $L_\alpha(K)$ -norm $\|\varepsilon\|_\alpha \leq 1$ for $\alpha = 1, 2$, respectively. The cases $\alpha = 1, 2$ are named average, respectively average square, conditional.

For these nbd's, see HR (1994; Chap 4,7) and the references cited there.

2.1.3 Risk and Inefficiency The *asymptotic maximum MSE* of the ALE with pIC η_{r_0} that is optimal for an infinitesimal neighborhood of (starting) radius $r_0 \in [0, \infty)$, evaluated over an infinitesimal nbd of another (starting) radius $r \in [0, \infty)$ is

$$\max\text{MSE}(\eta_{r_0}, r) = \text{E}|\eta_{r_0}|^2 + r^2 \omega_{*,\alpha}^2(\eta_{r_0}) \quad (2.6)$$

where $* = c, v$ and $\alpha = 1, 2$. The bias terms $\omega_{*,\alpha}(\eta_{r_0})$ for the different models are defined and evaluated in HR (1994; Subsections 5.3.1 and 7.3.2).

The *MSE-inefficiency* is then obtained by division through the minimax asymptotic MSE for radius r ,

$$\text{relMSE}(\eta_{r_0}, r) = \frac{\max \text{MSE}(\eta_{r_0}, r)}{\max \text{MSE}(\eta_r, r)} \quad (2.7)$$

G-risk generalizes MSE as in Ruckdeschel and Rieder (2004): we in principle consider asymptotic maximal risk R on shrinking nbd's \mathcal{U}_n with respect to certain loss function $\ell: \mathbb{R}^p \rightarrow [0, \infty]$; that is,

$$R(S) = \lim_{M \rightarrow \infty} \lim_n \sup_{Q_n \in \mathcal{U}_n} \int M \wedge \ell(\sqrt{n}(S_n - \theta)) dQ_n^n \quad (2.8)$$

With the usual identification of S and η , let $\sigma^2(\eta) = \text{E}|\eta|^2$ denote the trace of the asymptotic covariance; then we generalize the MSE case, where $\ell(z) = |z|^2$ and $R(\eta) = \sigma^2(\eta) + r^2\omega(\eta)^2$, to losses ℓ leading to

$$R_G(\eta) = G(r\omega(\eta), \sigma(\eta)) \quad (2.9)$$

where we assume $G: (0, \infty] \times (0, \infty) \rightarrow (0, \infty]$ to be isotone in both arguments, totally differentiable, convex and

$$\inf \{ R_G(\eta) \mid \eta \in \Psi^D \} < \lim_{w \rightarrow \infty} G(w, s), \quad \text{for } s^2 \geq \text{tr } D\mathcal{I}^{-1}D'$$

Theorem 3.1 in Ruckdeschel and Rieder (2004) determines the pIC minimizing this G -risk; it is of the same form as the MSE solution but with a different equation (Ruckdeschel and Rieder 2004; (3.4)) determining the clipping height.

2.1.4 G-inefficiency generalizes relMSE and is defined as

$$\rho_G(r', r) := \frac{G(r\omega_{r'}, \sigma_{r'})}{G(r\omega_r, \sigma_r)} \quad (2.10)$$

where ω_r, σ_r^2 denote bias and variance of the G -optimal pIC for radius r . This inefficiency reflects the loss in efficiency w.r.t. G -risk when using the r' -optimal procedure instead of the r -optimal one in a situation with “true” radius r .

2.2 Unknown radius

The introduced inefficiencies serve to get rid of the dependence on r : For r varying in $[r_l, r_u]$, we define the *minimax G-inefficiency* as

$$\bar{\rho}_G := \inf_{r' \in (r_l, r_u)} \sup_{r \in (r_l, r_u)} \rho_G(r', r) \quad (2.11)$$

where if nothing else is stated, G defaults to MSE, and ρ_G is simply relMSE. A radius $r_0 = r_{G,0}$ which attains

$$\sup_{r \in (r_l, r_u)} \rho_G(r_0, r) = \bar{\rho}_G \quad (2.12)$$

is called *least favorable radius*, and the G -optimal pIC for the least favorable radius is called *radius-minimax*.

In addition to the true radius r being completely unknown (unrestricted radius interval, $r_l = 0$, $r_u = \infty$), we consider the cases that the user can specify the radius up to a factor of $1/3$ or $1/2$, that is any r_3 or r_2 such that the true radius r certainly would stay within $[\frac{1}{3}r_3, 3r_3]$ or $[\frac{1}{2}r_2, 2r_2]$, respectively. For any such interval, the least favorable r_0 (and thus, the corresponding radius-minimax estimator) may be found as in the unrestricted case. In a further step, least favorable values of r_3 and r_2 are determined; these are those radii that maximize the minimax subefficiencies over $[\frac{1}{3}r_3, 3r_3]$ and $[\frac{1}{2}r_2, 2r_2]$, respectively.

2.3 Theoretical results

Our algorithms employ that the least favorable radius $r_0 \in [r_l, r_u]$ is a zero of the mapping $r \mapsto \rho_G(r, r_l) - \rho_G(r, r_u)$. This uses the bowl-shape of the ρ_G curves as plotted in figures attached to this paper which follows from the fact that $s \mapsto \rho_G(r, s)$ is de/increasing in s for $s < r$ resp. $s > r$, and $r \mapsto \rho_G(r, s)$ is in/decreasing in r for $s < r$ resp. $s > r$. The following theorem (Ruckdeschel and Rieder 2004; Theorem 6.1) guarantees these monotonicities, as well as the risk-independence alluded to in Subsection 1.2. In particular, it covers the case of the MSE in all its assertions.

Theorem 2.1 (a) *Assume a risk of form (2.9) with G isotone in both arguments and totally differentiable, and that in addition G be homogeneous in the sense that there is some function $\mu_G: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that for all $\nu > 0$, $s > 0$, $w > 0$*

$$G(\nu w, \nu s) = \mu_G(\nu) G(w, s) \quad (2.13)$$

Then for each $r' > 0$,

$$\sup_{r \in (r_l, r_u)} \rho_G(r', r) \leq \mu_G\left(\frac{\omega_{r'}}{\omega_{r_u}}\right) \vee \mu_G\left(\frac{\sigma_{r'}}{\sigma_{r_l}}\right) \quad (2.14)$$

There is an $r_1 \in (r_l, r_u)$ depending on r_l, r_u , and G , such that

$$\mu_G\left(\frac{\omega_{r_1}}{\omega_{r_u}}\right) = \mu_G\left(\frac{\sigma_{r_1}}{\sigma_{r_l}}\right) \quad (2.15)$$

(b) For $r_l = 0$ $r_u = \infty$, and the corresponding least favorable radius $r_0 = r_{G,0}$, the following identity holds,

$$\sup_{r>0} \rho_G(r_0, r) = \mu_G\left(\frac{\omega_{r_0}}{\omega_{\min}}\right) = \mu_G\left(\frac{\sigma_{r_0}}{\sigma_{\min}}\right) = \bar{\rho}_G \quad (2.16)$$

where ω_{\min} denotes the bias of the most robust pIC and $\sigma_{\min}^2 = \text{tr } D\mathcal{I}^{-1}D'$. The radius-minimax pIC η_b does not depend on G and, for cases $* = c$ respectively $* = v$ and $k = 1$ attains form

$$\eta_b = (A\Lambda - a) \min\left\{1, \frac{b}{|A\Lambda - a|}\right\} \quad (2.17(c))$$

$$\eta_b = c \vee A\Lambda \wedge (c + b) \quad (2.17(v))$$

for Lagrange multipliers $A \in \mathbb{R}^{p \times k}$, $a \in \mathbb{R}^p$, $c \in (-b; 0)$ ensuring that η_b is a pIC, and with clipping height $b = b_G(r_{G,0})$ determined by

$$\frac{\omega(\eta_b)}{\omega_{\min}} = \frac{\sigma(\eta_b)}{\sigma_{\min}} \quad (2.18)$$

3 Optimally Robust Estimates and Their Inefficiency Curves

3.1 List of Ideal Models Considered

In this paper, we consider the following list of ideal models, which is of course not exhaustive, but somehow representative for our results. Further models, including the Binomial, Poisson, Gamma and Gaussian ARMA and ARCH have been considered in Kohl (2005).

3.1.1 k -dimensional normal location:

$$y_i = \theta + u_i \quad (3.1)$$

with parameter $\theta \in \mathbb{R}^k$, errors u_i i.i.d. $\sim \mathcal{N}(0, \sigma_u^2 \mathbb{I}_k)$, scale $\sigma_u \in (0, \infty)$ known. The scores are $\Lambda_\theta(y) = \sigma_u^{-2}(y - \theta)$ and $\mathcal{I}_\theta = \sigma_u^{-2} \mathbb{I}_k$ the Fisher information.

3.1.2 One-dimensional normal scale:

$$y_i = \theta u_i \quad (3.2)$$

with parameter $\theta \in (0, \infty)$, the errors u_i i.i.d. $\sim \mathcal{N}(0, 1)$. The scores and Fisher information are given by $\theta \Lambda_\theta(y) = \theta^{-2} y^2 - 1$ and $\mathcal{I}_\theta = 2\theta^{-2}$.

3.1.3 One-dimensional normal location and scale:

$$y_i = \mu + \sigma u_i \quad (3.3)$$

with parameter $\theta' = (\mu, \sigma) \in \mathbb{R} \times (0, \infty)$, the errors u_i i.i.d. $\sim \mathcal{N}(0, 1)$. The scores and Fisher information are given by

$$\Lambda_\theta(y) = \sigma^{-1} \begin{pmatrix} \sigma^{-1}(y - \mu) \\ \sigma^{-2}(y - \mu)^2 - 1 \end{pmatrix} \quad \mathcal{I}_\theta = \sigma^{-2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

3.1.4 k -dimensional normal linear regression:

$$y_i = x_i' \theta + u_i \quad (3.4)$$

with parameter $\theta \in \mathbb{R}^k$, the random regressors x_i i.i.d. $\sim K(dx)$ and errors u_i i.i.d. $\sim \mathcal{N}(0, \sigma_u^2)$ stochastically independent; scale $\sigma_u \in (0, \infty)$ known. Scores and Fisher information are $A_\theta(x, y) = \sigma_u^{-2}(y - x'\theta)x$ and $\mathcal{I}_\theta = \sigma_u^{-2} \mathbb{E}_K xx'$.

For K we employ $K = \mathcal{N}(0, \sigma_x^2 \mathbb{I}_k)$ and $K = \text{Ufo}_k(0, m_x)$, the uniform on a centered ball of radius m_x ; $\sigma_x, m_x \in (0, \infty)$.

Remark 3.1 (Simplifications as to the choice of the parameter) In models 3.1.1–3.1.4, the observations are i.i.d.. The inefficiencies turn out invariant under rescaling of the u_i and x_i , respectively. So we may fix

$$\sigma_u = 1, \quad \sigma_x = 1, \quad m_x = 1 \quad (3.5)$$

Moreover, $\theta = 0$ may be fixed in models 3.1.1 and 3.1.4, $\theta = 1$ in models 3.1.2 and $\theta = (0, 1)'$ in model 3.1.3, due to equivariance of these models.

For all considered models, we also spell out minimax robust estimators and minimum bias explicitly, based on the general form from HR (1994; Chapters 5 and 7).

3.2 Summary For Some Particular Models

3.2.1 One-Dimensional Robust Location

Nbd's of fixed size: In Huber's (1964, 1981) approach, the ideal standard normal location model is enlarged to symmetric contamination nbd's \mathcal{V} of fixed size $s \in [0, 1]$; in his model, we speak of 'size' instead of 'radius'. As estimators, location M -estimates are employed and judged by their maximum asymptotic variance.

In this setup, it is the optimally robust M -estimate for $s_0 = 27.8\%$ (least favorable) that minimizes the maximum subefficiency over $[0, 1]$. The minimax subefficiency of 18.1% improves on the 57.1% of the median (approximately optimal as $s \rightarrow 100\%$), and it even more improves on the 90.8% subefficiency (attained for $s \rightarrow 1$, vs. only 3.7% at $s = 0$) that goes with Huber's (1964) preferred clipping height $m_{s_1} = 1.5$ (belonging to the optimally robust M -estimate for symmetric contamination size s_1 only 3.76%). Rather, the H07-estimate with clipping height .70, which has survived in Sections 7.B.8 and 7.C.4 of the Princeton robustness study by Andrews et al. (1972), comes (in fact, very) close to the size–minimax M -estimate ($m_{s_0} = .719$) achieving maximum subefficiency 18.7% \approx 18.1%.

The subefficiency of the size–minimax M -estimate is the maximal 18.1% only at the unrealistic size boundaries 0 and 1. On more realistic size intervals (about $s_0 = .278$), it stays well below 18.1%: below 2.5% for

$.12 \leq s \leq .50$, below 5% for $.074 \leq s \leq .62$, and still below 10% for $.028 \leq s \leq .78$.

Thus, using the optimally robust M -estimate for $s_0 = 27.8\%$, as opposed to the mean, median, or Huber's proposal, one will not only stay within 18.1% of the minimax asymptotic variance over a symmetric contamination nbd of whatever size $s \in [0, 1)$ but, at the same time, within 2.5% of the minimax risk for any size $12\% \leq s \leq 50\%$, within 5% for any size $7.4\% \leq s \leq 62\%$, and still within 10% of the minimax risk for arbitrary size $2.8\% \leq s \leq 78\%$.

Remark 3.2 Via relation (3.6) below, $s_0 = .278$ corresponds to $r_0 = .62$, and the corresponding radius intervals about r_0 read:

$$.37 \leq r \leq 1.01 \text{ (2.5\%)}, .29 \leq r \leq 1.27 \text{ (5\%)}, .17 \leq r \leq 1.92 \text{ (10\%)}. \quad \square$$

Shrinking Nbd's: In the infinitesimal setup, despite of the conceptual differences to Huber's approach, the same well-known kind of optimally robust estimators are obtained. Not so well-known however is that also the maximum risks in both models agree (up to a factor $1 - s$), and hence the inefficiency curves coincide, via the following size/radius-relation:

$$s = r^2 / (1 + r^2) \quad (3.6)$$

Thus, the least favorable (starting) radius means $r_0 = .62$, which is just 6.2% contamination at sample size $n = 100$. The minimax subefficiency again is 18.1%. The subefficiency of the radius-minimax estimator stays below 2.5%, 5%, and 10%, in the contamination intervals: 3.7%–10.1%, 2.9%–12.7%, and 1.7%–19.2%, respectively, at sample size $n = 100$ (Remark 3.2). The 18.1% minimax subefficiency may be cut down to less than 8.9% and 4.5%, if the user can specify any r_3, r_2 such that the true radius r stays within $[\frac{1}{3}r_3, 3r_3]$ and $[\frac{1}{2}r_2, 2r_2]$, respectively. The least favorable radii are $r_3 = .55$ and $r_2 = .57$, defining the least favorable contamination ranges 1.8%–16.5% and 2.9%–11.4%, at $n = 100$, respectively.

Remark 3.3 Our numbers obtained in the asymptotic minimax MSE approach refer to contamination nbd's. With shrinking total variation balls instead, the same estimators are optimally robust for radii one-half those for contamination and with the same minimax risk. The inefficiency curves at r ($* = c$) and $r/2$ ($* = v$) thus agree and the radius-minimax procedure stays the same while the least favorable radius is halved. The coincidence extends to the k -dim. location and regression models of our study. \square

3.2.2 One-Dimensional Robust Scale centered at the standard normal already demonstrates the limitations of the minimax asymptotic variance approach; cf. Huber (1981; Sec. 5.7, p 124).

Apart from one-dim. location, therefore, including scale, the infinitesimal robust setup is used, employing ALE's, and maxMSE. In the scale model, the nbd's may further be restricted by symmetry.

If r is totally unknown, the minimax subefficiency is 50.5%, to be compared with the 172.1% of the median absolute deviation, and $r_0 = .50$ is the least favorable radius (5% contamination at $n = 100$). If the radius is known up to a factor of $\frac{1}{3}$ or $\frac{1}{2}$, the value 50.5% may be lowered to less than 20.8% and 9.9%, respectively. The corresponding least favorable radii $r_3 = .49$ and $r_2 = .56$ define the least favorable contamination ranges 1.6%–14.7% and 2.8%–11.2%, at sample size $n = 100$, respectively.

Remark 3.3 does not extend to the normal scale model, as its scales scores are even-symmetric. Also the optimally robust influence curve for total variation (spelled out here seemingly for the first time) differs from that for contamination of twice the radius; in particular, the new solution always involves clipping from below; cf. Subsection 3.4.

Nevertheless, the 1 : 2 relation seems to hold at least approximately for the least favorable radii: $r_0 = .27$, $r_3 = .24$, and $r_2 = .25$ (numerical evaluation). But the subefficiency numbers, too, are only about one half those for contamination: The minimax subefficiency 25.4% in case $\rho = 0$ compares with 85% maximum subefficiency of the most robust estimate, and drops to 11.5% and 5.6%, respectively, if $\rho = \frac{1}{3}, \frac{1}{2}$; cf. Subsection B.2.

Thus, robust scale estimation becomes even more stable under radius misspecification if it is based on, and employs the optimally robust procedures devised for, the larger total variation balls.

3.2.3 k -Dimensional Robust Location about the k -variate standard normal enlarged by r/\sqrt{n} -contamination nbd's has the minimax subefficiency of $r \in [0, \infty)$ unknown decrease from 18.1% for $k = 1$ to 12.1% for $k = 2$, and to 9.1% for $k = 3$. As k increases, the relative MSE-risks are squeezed towards 1 near the origin but, due to arbitrarily large supnorms of the optimally robust influence curves, spread out to the right. The minimal standardized bias of asymptotically linear estimators under contamination (the minimal supnorm of their influence curves) is

$$\omega_c^{\min} = k\Gamma\left(\frac{k}{2}\right) / \left(\sqrt{2}\Gamma\left(\frac{k+1}{2}\right)\right) \approx \sqrt{k} \quad \text{as } k \rightarrow \infty \quad (3.7)$$

and is achieved by the minimum L_1 -estimate. Also the trace of the covariance of this estimate equals approximately trace k of the inverse Fisher information; intuitively speaking, by its influence curve only one out of k spherical coordinates, the length, is sacrificed. Consequently, the minimum L_1 -estimate becomes the nearly optimal choice for larger dimension. For $k \geq 5$, its maximum subefficiency over the full radius range is less than 10.4%, for $k \geq 10$ less than 5.1%, and it stays within a factor of 2 of the minimax value, both subefficiencies decreasing to 0 as $k \rightarrow \infty$; cf. Subsection B.1 and the web-page to this article.

3.2.4 Infinitesimal Nbd Regression Unconditional, or errors-in-variables, nbd's are about the joint law of regressor and error; in particular, the regressor distribution may be distorted, too. Conditional, or error-free-variables,

regression nbd's, which go back to Huber (1983) and Bickel (1984), on the contrary keep the ideal K to have only the conditional error distribution given x distorted—by an amount $r\varepsilon(x)/\sqrt{n}$. As for more details on infinitesimal regression nbd's, unconditional and conditional, radius curves, cf. HR (1994; Chap. 7), who also provides the required MSE-optimality.

We employ conditional, or error-free-variables, nbd's with any radius curves ε subject to $L_\alpha(K)$ -norm $\|\varepsilon\|_\alpha \leq 1$ for $\alpha = 1, 2$. These cases obtain the attributes average, and average square, respectively.

For square average conditional contamination, Huber M -estimates are optimally robust. Independently of the regressor distribution, their relative risks turn out identical to those in the one-dim. location model with infinitesimal contamination nbd's; cf. Subsubsection 3.7.2. Minimax inefficiencies and least favorable radii, therefore, are the same as for one-dim. location.

Hampel–Krasker estimates are optimally robust in case of average conditional, as well as unconditional, contamination. The minimax subefficiency over the full radius range descends from the values 27.1% (K uniform) and 34.7% (K normal) for $k = 1$ to the value 18.1% (one-dim. location) as $k \rightarrow \infty$. Related numbers, e.g. the minimax subefficiency in case the radius can be specified up to factor $\frac{1}{3}$ or $\frac{1}{2}$, converge likewise. The least favorable radii r_0 approach the value .62 (one-dim. location) from below. For all dimensions, the minimax subefficiency cuts down the maximum subefficiency of the most robust estimate to less than $1/3$ its value; cf. the tables in Subsubsection B.4.

Remark 3.4 The limit for increasing dimension depends on our choice of regressor distributions, and may explained by the fact that $|x| \approx 1$ in the case $K(dx) = \text{Ufo}_k(0, m)$, and that $|x| \approx \sqrt{k}$ in the case $K(dx) = \mathcal{N}(0, \sigma^2 \mathbb{I}_k)$, if k is large.

If the spherically symmetric regressor distribution $K(dx)$ were chosen such that the distribution of $|x|$ for general dimension k is the same as for dimension $k = 1$, then the inefficiency numbers and least favorable radii would stay the same as for one dimension; that is, in the tables of Subsection B.4, the first lines would be valid for all other dimensions as well. \square

In the following subsections 3.3, 3.4, 3.5, and 3.6, the MSE-minimax IC η_r to radius $r \in [0, \infty)$ and $r = \infty, * = c/v$ as well as the minimum bias will be provided by HR (1994; Theorems 5.5.7, 5.5.1.b/5.5.5b), respectively. We will not mention this in each case.

3.3 One-Dimensional Location

3.3.1 Minimax Asymptotic Variance In the $k = 1$ dim. location model 3.1.1, we start with symmetric contamination nbd's 2.1.2 (a). The minimax M-estimate for size $s \in [0, 1)$ is defined by

$$\psi_s(u) = (-m_s) \vee u \wedge m_s, \quad \frac{s}{1-s} m_s = \mathbf{E}(|u| - m_s)_+ \quad (3.8)$$

as given by Huber (1964). For size $s = 1$, we take ψ_1 from the median,

$$\psi_1(u) = \text{sign}(u) \quad (3.9)$$

The maximal asymptotic variance of ψ_{s_0} (that is, of the M-estimate based on ψ_{s_0}) for fixed size $s_0 \in [0, 1)$ evaluated over a symmetric contamination nbd of fixed size $s \in [0, 1)$ is

$$\max\text{Var}(\psi_{s_0}, s) = \frac{(1-s)\mathbb{E}\psi_{s_0}^2 + sm_{s_0}^2}{[(1-s)\mathbb{E}\psi'_{s_0}]^2} \quad (3.10)$$

respectively, in the case of the median,

$$\max\text{Var}(\psi_1, s) = \frac{\pi}{2(1-s)^2} \quad (3.11)$$

The *Var-inefficiency*, for $s_0 \in [0, 1]$ and $s \in [0, 1)$, is

$$\text{relVar}(\psi_{s_0}, s) = \frac{\max\text{Var}(\psi_{s_0}, s)}{\max\text{Var}(\psi_s, s)} \quad (3.12)$$

Although $\max\text{Var}(\psi_1, 1) = \infty$, the median is approximately optimal for nbd size $s \rightarrow 1$, as not only $\psi_s/m_s \rightarrow \psi_1$ pointwise but, more conclusively, we show that

$$\lim_{s \rightarrow 1} \text{relVar}(\psi_1, s) = 1 \quad (3.13)$$

3.3.2 Minimax Asymptotic MSE for r/\sqrt{n} -Contamination Balls Still in the $k = 1$ dim. normal location model 3.1.1, we continue with infinitesimal contamination nbd's 2.1.2 (b) and get for $r \in [0, \infty)$

$$\eta_r(u) = A_r u \min\{1, c_r |u|^{-1}\} \quad (3.14)$$

where

$$1 = A_r \mathbb{E}|u| \min\{|u|, c_r\}, \quad r^2 c_r = \mathbb{E}(|u| - c_r)_+ \quad (3.15)$$

and for $r = \infty$,

$$\eta_\infty(u) = \omega_c^{\min} \text{sign}(u) \quad (3.16)$$

which is the IC of the median and achieves minimum bias

$$\omega_c^{\min} = (\mathbb{E}|A|)^{-1} = \sqrt{\pi/2} \quad (3.17)$$

In Subsection A.2 we prove the following relation between the maximum risks of the optimal estimates in the two models 3.3.1 and 3.3.2,

$$(1-s) \max\text{MSE}(\eta_{r_0}, r) = \max\text{Var}(\psi_{s_0}, s) \quad (3.18)$$

where the radii $r_0, r \in [0, \infty)$ and sizes $s_0, s \in [0, 1)$ are connected by

$$s = r^2/(1+r^2), \quad s_0 = r_0^2/(1+r_0^2) \quad (3.19)$$

Consequently, by (3.18) and (3.19), the inefficiency curves coincide in the two models,

$$\text{relMSE}(\eta_{r_0}, r) = \text{relVar}(\psi_{s_0}, s) \quad (3.20)$$

3.4 One-Dimensional Normal Scale

3.4.1 r/\sqrt{n} -Contamination Balls In the normal scale model 3.1.2, we start with infinitesimal contamination nbd's 2.1.2 (b) and get for $r \in [0, \infty)$

$$\eta_r(u) = A_r(u^2 - \alpha_r^2) \min \{1, c_r/|u^2 - \alpha_r^2|\} \quad (3.21)$$

where A_r and α_r ensure that η_r is an IC and

$$r^2 c_r = \mathbb{E}(|u^2 - \alpha_r^2| - c_r)_+ \quad (3.22)$$

The parabola $u^2 - \alpha_r^2$ in (3.21) is clipped only from above for $r \leq 0.920$, and for $r \geq 0.920$ from above as well as from below. The centering constant α_r decreases from $\alpha_0 = 1$ to $\alpha_\infty = \Phi^{-1}(3/4) \approx 0.674$. For $r = \infty$,

$$\eta_\infty(u) = \omega_c^{\min} \text{sign}(|u| - \alpha_\infty) \quad (3.23)$$

which is the IC of the MAD and the IQR, attaining minimum bias

$$\omega_c^{\min} = (\mathbb{E}|u^2 - \alpha_\infty^2|)^{-1} = (4\alpha_\infty \varphi(\alpha_\infty))^{-1} \approx 1.166 \quad (3.24)$$

3.4.2 r/\sqrt{n} -Total Variation Balls Still in scale model 3.1.2, we continue with infinitesimal total variation nbd's 2.1.2 (b) and get for $r \in [0, \infty)$:

$$\eta_r(u) = A_r\{[g_r \vee u^2 \wedge (g_r + c_r)] - 1\} \quad (3.25)$$

where A_r and g_r ensure that η_r is an IC and

$$r^2 c_r = \mathbb{E}(g_r - u^2)_+ \quad (3.26)$$

For $r = \infty$, we get

$$\eta_\infty(u) = \omega_v^{\min} \{P(|u| < 1) \mathbf{I}(|u| > 1) - P(|u| > 1) \mathbf{I}(|u| < 1)\} \quad (3.27)$$

with minimum bias

$$\omega_v^{\min} = (\mathbb{E} \Lambda_+)^{-1} = \sqrt{\pi e / 2} \approx 2.066 \quad (3.28)$$

3.5 One-Dimensional Location and Scale, Contamination

We consider the one-dim. normal location and scale model 3.1.3 with infinitesimal contamination nbd's 2.1.2 (b) and get for $r \in [0, \infty)$

$$\eta_r(u) = \begin{pmatrix} \eta_r^{\text{loc}}(u) \\ \eta_r^{\text{sc}}(u) \end{pmatrix} = \begin{pmatrix} A_r^{\text{loc}} u \\ A_r^{\text{sc}}(u^2 - \alpha_r^2) \end{pmatrix} w_r(u) \quad (3.29)$$

with

$$w_r(u) = \min \left\{ 1, \frac{b_r}{|T_r|} \right\}, \quad |T_r|^2 = (A_r^{\text{loc}})^2 u^2 + (A_r^{\text{sc}})^2 (u^2 - \alpha_r^2)^2 \quad (3.30)$$

where A_r^{loc} , A_r^{sc} , and α_r ensure that η_r is an IC and

$$r^2 b_r = \mathbb{E}(|T_r| - b_r)_+ \quad (3.31)$$

The location part η_r^{loc} of the minimax IC η_r is a redescending function in u . The centering constant α_r of the scale part decreases from $\alpha_0 = 1$ to $\alpha_\infty \approx 0.610$. For $r = \infty$, we get

$$\eta_\infty(u) = \omega_c^{\min} \left(A_\infty (u^2 - \alpha_\infty) \right) |T_\infty|^{-1} \quad (3.32)$$

with

$$|T_\infty|^2 = u^2 + A_\infty^2 (u^2 - \alpha_\infty)^2 \quad (3.33)$$

attaining minimum bias

$$\omega_c^{\min} = \max \left\{ \frac{1 + A_\infty}{\mathbb{E} |T_\infty|} \mid \alpha_\infty, A_\infty \in \mathbb{R} \right\} \approx 1.618 \quad (3.34)$$

where $A_\infty \approx 0.792$.

3.6 k -Dimensional Location, Contamination

We consider the k -dim. location model 3.1.1 and infinitesimal contamination nbd's 2.1.2 (b). For $r \in [0, \infty)$, due to spherical symmetry (Lemma A.3), we get

$$\eta_r(u) = \alpha_r u \min \{1, c_r |u|^{-1}\} \quad (3.35)$$

where

$$k = \alpha_r \mathbb{E} |u| \min \{|u|, c_r\}, \quad r^2 c_r = \mathbb{E} (|u| - c_r)_+ \quad (3.36)$$

For $r = \infty$, we get

$$\eta_\infty = \omega_c^{\min} u/|u| \quad (3.37)$$

being the IC of the minimum L_1 -estimate and attains minimum bias ω_c^{\min} . In Subsection A.4 we show that

$$\omega_c^{\min} = \frac{k}{\mathbb{E} |A|} = \frac{k \Gamma(\frac{k}{2})}{\sqrt{2} \Gamma(\frac{k+1}{2})} \quad (3.38)$$

For increasing dimension, it holds that

$$\lim_{k \rightarrow \infty} \frac{\omega_c^{\min}}{\sqrt{k}} = \lim_{k \rightarrow \infty} \frac{\mathbb{E} |\eta_\infty|^2}{k} = 1 \quad (3.39)$$

Thus, the squared minimum bias is about the same as the MSE in the ideal model, in which the minimum L_1 -estimate becomes approximately efficient. Since, moreover,

$$\lim_{k \rightarrow \infty} \frac{\max \text{MSE}(\eta_{r_0}, r)}{\max \text{MSE}(\eta_\infty, r)} = 1 \quad (3.40)$$

where the convergence is uniform on bounded r_0 -, r -intervals, this most robust estimate also becomes approximately radius-minimax. Proofs for assertions (3.39) and (3.40) may be found on the web-page to this article.

3.7 Regression, Average (Square) Contamination

In the conditional nbd setting of this subsection, we read off the optimal solutions provided by HR (1994; Theorem 7.4.13 and 7.5.15) and use spherical symmetry, respectively Lemma A.3.

3.7.1 Average Contamination Nbd's (= c, $\alpha = 1$)* We consider the k -dim. normal regression model 3.1.4 and average conditional regression nbd's 2.1.2 (c) of type contamination. For $r \in [0, \infty)$, we get

$$\eta_r(x, u) = \alpha_r x u \min \{1, c_r |xu|^{-1}\} \quad (3.41)$$

where

$$k = \alpha_r \mathbf{E} |xu| \min \{|xu|, c_r\}, \quad r^2 c_r = \mathbf{E} (|xu| - c_r)_+ \quad (3.42)$$

For $r = \infty$,

$$\eta_\infty(x, u) = \omega_{c,1}^{\min} \frac{x}{|x|} \text{sign}(u) \quad (3.43)$$

which achieves minimum bias $\omega_{c,1}^{\min}$. Analogously to (3.38) we show that

$$\omega_{c,1}^{\min} = \frac{k}{\mathbf{E} |A|} = \sqrt{\frac{\pi}{2}} \frac{k}{\mathbf{E} |x|} \quad (3.44)$$

As proven on the web-page to this article, for increasing dimension $k \rightarrow \infty$, the MSE-inefficiency tends uniformly on bounded r_{0^-} , r -intervals to the MSE-inefficiency in the one-dim. location model 3.3.2,

$$\lim_{k \rightarrow \infty} \text{relMSE}(\eta_{r_0}, r) = \text{relMSE}(\eta_{r_0}^{1.\text{loc}}, r) \quad (3.45)$$

where $\eta_{r_0}^{1.\text{loc}}$ is given by (3.14) and (3.15).

3.7.2 Average Square Contamination Nbd (= c, $\alpha = 2$)* We consider the k -dim. normal regression model 3.1.4 and average square conditional regression nbd's 2.1.2 (c) of type contamination. For $r \in [0, \infty)$, we get

$$\eta_r(x, u) = \alpha_r x u \min \{1, c_r |u|^{-1}\} \quad (3.46)$$

where

$$k = \alpha_r \mathbf{E} |x|^2 \cdot \mathbf{E} |u| \min \{|u|, c_r\}, \quad r^2 c_r = \mathbf{E} (|u| - c_r)_+ \quad (3.47)$$

For $r = \infty$, we get

$$\eta_\infty(x, u) = \mathcal{K}^{-1} \frac{x}{\mathbf{E} |u|} \text{sign}(u) \quad (3.48)$$

with $\mathcal{K} = \mathbf{E} x x' = \gamma \mathbb{I}_k$ for some $\gamma \in (0, \infty)$, cf. Lemma A.2, where

$$\omega_{c,2}^{\min} = \sqrt{\text{tr} \mathcal{K}^{-1} / \mathbf{E} |u|} = \sqrt{\pi k / (2\gamma)} \quad (3.49)$$

Comparing (3.14), (3.15) and (3.46), (3.47), we obtain the following relation to $\max\text{MSE}$ in the one-dim. location model 3.3.2,

$$\max\text{MSE}(\eta_{r_0}, r) = k^2 \max\text{MSE}(\eta_{r_0}^{1.\text{loc}}, r) / \mathbb{E}|x|^2 \quad (3.50)$$

where $\eta_{r_0}^{1.\text{loc}}$ denotes the corresponding minimax IC for radius r_0 ; in fact, the constants in (3.14), (3.15) and (3.46), (3.47) are connected via

$$c_{r_0} = c_{r_0}^{1.\text{loc}}, \quad \alpha_{r_0} = k / (\mathbb{E}|x|^2 A_{r_0}^{1.\text{loc}}) \quad (3.51)$$

Hence, by relation (3.50), the MSE-inefficiencies coincide with those in one-dim. location, independently of the regressor distribution $K(dx)$.

4 Numerical Algorithms

We use R (R Development Core Team 2005) to implement the algorithms and to generate the graphical output. In detail we use the following numerical procedures:

One-Dimensional Location The results for the models 3.3.1 and 3.3.2 are obtained by the routines for the k -dim. location model 3.6 with $k = 1$. Note the coincidence (3.20) via the relation (3.19).

One-Dimensional Scale The clipping bound c_r and the centering constant α_r in (3.21)–(3.22) and the clipping constants g_r and c_r in (3.25)–(3.26) are calculated by the R-function `uniroot()`.

k -Dimensional Location (Models 3.6) We compute the constants c_r and α_r in (3.35), (3.36) by using clipped absolute moments of $\mathcal{N}(0, \mathbb{I}_k)$. Because of the boundedness and the arbitrary smoothness of these moments, we can apply a two dim. Newton method to calculate c_r and α_r simultaneously (cf. Ruckdeschel (2001; Definition D.2.4, Lemma D.2.5, and Korollar D.2.9).

Regression, Average (Square) Contamination (Models 3.7.1, 3.7.2) For average contamination neighborhoods, the determination of clipping bound c_r in (3.41), (3.42) is performed by the R-function `uniroot()`, where the integration of the outer integral is done numerically using the R-function `integrate()`, while for average square contamination neighborhoods, the procedures may be obtained from the one-dim. location case.

General Procedures In all these models, we do the three calculations:

- (a) Given $1 < \delta < \text{Var}(\eta_\infty)$, we determine $r \in (0, \infty)$ s.t. $\text{Var}(\eta_r) = \delta$.
- (b) Given $\rho \in [0, 1)$ and $r \in (0, \infty)$, we determine $r_0 \in [\rho r, r/\rho]$ such as to achieve $\text{relMSE}(\eta_{r_0}, \rho r) = \text{relMSE}(\eta_{r_0}, r/\rho)$.

The algorithms in (a) and (b) use the R-function `uniroot()`.

- (c) Then, given $\rho \in [0, 1)$, the minimax subefficiency over $[\rho r, r/\rho]$ is maximized with respect to $r \in (0, \infty)$. For reasons of monotonicity we may use the R-function `optimize()`. For $\rho = 0, \frac{1}{3}, \frac{1}{2}$, thus the least favorable radii r_0 , r_2 , and r_3 are obtained.

Plots One plot of the typical inefficiency curve is attached at the end of this paper. Some more risk- and inefficiency-plots for the models considered in this study may be looked at, using access name `radius`, password `unknown`, under <http://www.uni-bayreuth.de/departments/math/org/mathe7/radius> and downloaded. Further plots are included in Kohl (2005).

References

- Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972) *Robust estimates of location: Survey and Advances*. Princeton University Press, Princeton, New Jersey.
- Beran, R.J. (1976) Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.* **11**: 432–452
- Beran, R.J. (1981) Efficient robust estimates in parametric models. *Z. Wahrsch. verw. Gebiete* **55**: 91–108.
- Bickel, P.J. (1981) Quelques aspects de la statistique robuste. In *Ecole d'Été de Probabilités de Saint Flour IX 1979* (P.L. Hennequin, ed.), pp. 1–72. Lecture Notes in Mathematics #876. Springer-Verlag, Berlin.
- Bickel, P.J. (1982) On adaptive estimation. *Ann. Statist.* **10**: 647–671.
- Bickel, P.J. (1984) Robust regression based on infinitesimal neighborhoods. *Ann. Statist.* **12**: 1349–1368.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) *Robust Statistics—The Approach Based on Influence Functions*. Wiley, New York.
- Huber, P.J. (1964) Robust estimation of a location parameter. *Ann. Math. Statist.* **35**: 73–101.
- Huber, P.J. (1981) *Robust Statistics*. Wiley, New York.
- Huber, P.J. (1983) Minimax aspects of bounded influence regression. *J. Amer. Statist. Assoc.* **78**: 66–80.
- Huber, P.J. (1996) *Robust Statistical Procedures* (2nd ed.). CBMS-NSF Regional Conference Series in Applied Mathematics # 68. Soc. Industr. Appl. Math., Philadelphia, Pennsylvania.
- Klaassen, C. (1980) Statistical performance of location estimators. *Mathematical Centre Tract #133*. Mathematisch Centrum, Amsterdam.
- Kohl, M. (2005) *Numerical Contributions to the Asymptotic Theory of Robustness*. Dissertation, Universität Bayreuth. Also available under <http://stamats.de/ThesisMKohl.pdf>
- Kreiss, J.-P. (1987) On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15**: 112–133.
- Millar, P.W. (1981) Robust estimation via minimum distance methods. *Z. Wahrsch. verw. Gebiete* **55**: 73–89.
- Pfanzagl, J. and Wefelmeyer, W. (1982) *Contributions to a General Asymptotic Statistical Theory*. Lecture Notes in Statistics #13. Springer, Berlin.
- R Development Core Team (2005) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0; <http://www.R-project.org>.

- Rieder, H. (1994) *Robust Asymptotic Statistics*. Springer, New York.
- Rieder, H. (2000) Neighborhoods as nuisance parameters? Robustness vs. semiparametrics. Submitted to *Mathematical Methods of Statistics*.
- Rieder, H. (2001) Nonuniform convergence of adaptive estimators in linear time series models. Manuscript.
- Ruckdeschel, P. (2001) *Ansätze zur Robustifizierung des Kalman-Filters*. Dissertation, Universität Bayreuth.
- Ruckdeschel, P. and Rieder, H. (2004) Optimal influence curves for general loss functions. *Statistics and Decisions* **22**: 201–223.
- Witting, H. and Müller-Funk, U. (1995) *Mathematische Statistik II*. B.G. Teubner, Stuttgart.

A Lemmas and Proofs

A.1 Optimization

The following lemma serves to derive the solutions to the original MSE problems from the solutions given (for linear bias) in HR(1994; Theorems 7.4.11.b, 7.4.12.b, 7.4.16.b, and 7.4.18.b), if we set $\gamma(v) = v^2$.

Lemma A.1 For a convex subset A of a real vector space X , consider three convex functions $f: A \rightarrow \mathbb{R}$, $g: A \rightarrow [0, \infty)$, and $\gamma: [0, \infty) \rightarrow [0, \infty)$; γ increasing. Let $\beta_0 \in [0, \infty)$. Suppose $z_0 \in A$ minimizes the Lagrangian $L_0 = f + \beta_0 \gamma \circ g$ over A . Assume that γ is differentiable at $g_0 = g(z_0)$, and put

$$\beta_1 = \beta_0 \gamma'(g_0) \quad (\text{A.1})$$

Then z_0 also minimizes the Lagrangian $L_1 = f + \beta_1 g$ over A .

PROOF Employ the convex combinations $z_s = (1-s)z_0 + sz_1$, $0 \leq s \leq 1$, for any $z_1 \in A$. Then z_0 minimizes a convex function ℓ over A iff the right-hand derivatives $\partial \ell = \frac{d}{ds} \big|_{s=0} \ell(z_s)$ at zero are all nonnegative. But

$$\partial L_0 = \partial f + \beta_0 \gamma'(g_0) \partial g = \partial L_1 \quad (\text{A.2})$$

because $\partial(\gamma \circ g) = \gamma'(g_0) \partial g$ [chain rule]. \square

A.2 One-Dimensional Location

Proof of (3.13) To ψ_s for $s \in [0, 1)$, integration by parts applies so that

$$\int \psi'_s(u) \Phi(du) = \int u \psi_s(u) \Phi(du) \quad (\text{A.3})$$

Therefore we can rewrite (3.10) as

$$\max \text{Var}(\psi_s, s) = \frac{(1-s) \int m_s^{-2} \psi_s^2(u) \Phi(du) + s}{(1-s)^2 [\int u m_s^{-1} \psi_s(u) \Phi(du)]^2} \quad (\text{A.4})$$

Consequentially, the Var-inefficiency of the median is

$$\text{relVar}(\psi_1, s) = \frac{\pi \left[\int u m_s^{-1} \psi_s(u) \Phi(du) \right]^2}{2 \left[(1-s) \int m_s^{-2} \psi_s^2(u) \Phi(du) + s \right]} \quad (\text{A.5})$$

$\psi_s/m_s \rightarrow \psi_1 = \text{sign}$ pointwise for $s \rightarrow 1$ and $|\psi_s/m_s| \leq 1$ entail that

$$\lim_{s \rightarrow 1} \text{relVar}(\psi_1, s) = \frac{\pi}{2} (\mathbf{E}|u|)^2 = 1 \quad (\text{A.6})$$

by the dominated convergence theorem. \square

Proof of (3.18)–(3.20) By (3.9) and (3.15), $c_r = m_s$, if r and s are related by (3.19): $r^2 = s/(1-s)$. Using (3.15) we can rewrite (2.6) as

$$\begin{aligned} \max \text{MSE}(\eta_{r_0}, r) &= \frac{\mathbf{E} \min \{u^2, c_{r_0}^2\} + r^2 c_{r_0}^2}{[\mathbf{E}|u| \min \{|u|, c_r\}]^2} \\ &= \frac{\mathbf{E} \min \{u^2, m_{s_0}^2\} + s m_{s_0}^2 / (1-s)}{[\mathbf{E}|u| \min \{|u|, m_s\}]^2} = \frac{(1-s) \mathbf{E} \psi_{s_0}^2 + s m_{s_0}^2}{(1-s) [\mathbf{E} \psi'_{s_0}]^2} \quad \text{by (A.3)} \\ &= (1-s) \max \text{Var}(\psi_{s_0}, s) \end{aligned}$$

which proves (3.18) implying (3.20). \square

A.3 Invariance Under Rescaling

Inefficiency in models 3.1.1–3.1.4 is invariant under rescaling of the errors u_i and the regressors x_i , respectively. We prove this invariance for k -dim. regression and average conditional contamination neighborhoods 2.1.2 (c) ($* = c, \alpha = 1$), even allowing general error and regressor distributions $F(du)$ resp. $K(dx)$ as in HR (1994; Theorem 7.4.13). For the other models considered here, proofs are similar.

Proof of invariance under rescaling ($* = c, \alpha = 1$) According to HR (1994; Theorem 7.4.13.b and Remark 7.4.9), the minimax MSE solution is of form

$$\eta_r(x, u) = A_r x [A_f(u) - \vartheta_r(x)] \min \left\{ 1, \frac{b_r}{|A_r x [A_f(u) - \vartheta_r(x)]|} \right\} \quad (\text{A.7})$$

where A_r , ϑ_r , and b_r assure that η_r is an IC and

$$r^2 b_r = \mathbf{E} (|A_r x [A_f(u) - \vartheta_r(x)] - b|)_+ \quad (\text{A.8})$$

For regressor $z = \tau x$ and error $v = \sigma u$, rescaled by any $\tau, \sigma \in (0, \infty)$,

$$\tilde{\eta}_r(z, v) = \tilde{A}_r z [A_{\tilde{f}}(v) - \tilde{\vartheta}_r(z)] \min \left\{ 1, \frac{\tilde{b}_r}{|\tilde{A}_r z [A_{\tilde{f}}(v) - \tilde{\vartheta}_r(z)]|} \right\} \quad (\text{A.9})$$

with

$$\tilde{A}_r = \frac{\sigma^2}{\tau^2} A_r, \quad \tilde{b}_r = \frac{\sigma}{\tau} b_r, \quad \tilde{\vartheta}_r(z) = \frac{1}{\sigma} \vartheta_r\left(\frac{z}{\tau}\right) \quad (\text{A.10})$$

where $\Lambda_{\tilde{f}}(v) = \sigma^{-1}\Lambda_f(v/\sigma)$ and $\Lambda_f = -f'/f$.

Then it is easy to verify that $\tilde{\eta}_r$ is an IC in the rescaled model for which condition (A.8) holds, so $\tilde{\eta}_r$ is indeed the optimum IC there. Using the relations (A.10) we obtain

$$\max\text{MSE}(\tilde{\eta}_{r_0}, r) = \sigma^2 \max\text{MSE}(\eta_{r_0}, r) / \tau^2 \quad (\text{A.11})$$

for any $r_0, r \in [0, \infty)$. The factor σ^2/τ^2 cancels in relMSE. \square

A.4 Spherical Symmetry

We consider models with spherically symmetric scores A at P_θ ; i.e.;

$$\mathcal{L}(GA) = \mathcal{L}(A) \quad (\text{A.12})$$

for all orthogonal matrices $G \in \mathbb{R}^{k \times k}$ and obtain

Lemma A.2 *Let $A \in \mathbb{R}^{k \times k}$ be symmetric and $GAG' = A$ for all orthogonal matrices $G \in \mathbb{R}^{k \times k}$. Then $A = \alpha \mathbb{I}_k$ for some $\alpha \in \mathbb{R}$.*

PROOF Since A is symmetric, there is an orthogonal Matrix $G \in \mathbb{R}^{k \times k}$ such that $A = GAG' = \text{diag}(\alpha_1, \dots, \alpha_k)$; so A is diagonal. Now consider a permutation matrix $G \in \mathbb{R}^{k \times k}$. Such G being orthogonal, again $A = GAG'$; so necessarily $\alpha_1 = \dots = \alpha_k$. \square

Fisher information of such models satisfies

$$G\mathcal{I}G' = E(GA)(GA)' = EAA' = \mathcal{I} \quad (\text{A.13})$$

for all orthogonal $G \in \mathbb{R}^{k \times k}$, hence, by Lemma A.2, is a multiple of the identity: $\mathcal{I} = \gamma \mathbb{I}_k$; $\gamma \geq 0$ since \mathcal{I} is positive semidefinite, and $\gamma > 0$ if \mathcal{I} has full rank. The second application of Lemma A.2 is to

Optimally robust influence curves as given by HR (1994; Theorems 5.5.7, 7.4.11, 7.4.13, 7.4.15, 7.4.18, 7.4.19).

Lemma A.3 *Under assumption (A.12) the standardizing matrix A (to achieve Fisher consistency) satisfies $A = \alpha \mathbb{I}_k$ for some $\alpha \in (0, \infty)$.*

PROOF We will prove this for k -dim. regression and conditional contamination neighborhoods 2.1.2 (c); that is, for the cases $* = c$ and $\alpha = 1, 2, \infty$. The proofs in the other cases are similar.

Case ($*$ = c , $\alpha = 1$) For $r \in [0, \infty)$ define $c_r \in (0, \infty]$ and then $\alpha_r \in (0, \infty)$ by

$$r^2 c_r = \mathbb{E}(|x u| - c_r) \quad (\text{A.14})$$

$$k = \alpha_r \mathbb{E}|x u| \min\{|x u|, c_r\} \quad (\text{A.15})$$

and put

$$\eta_r(x, u) = \alpha_r x u \min\{1, c_r |x u|^{-1}\} \quad (\text{A.16})$$

As for all orthogonal $G \in \mathbb{R}^{k \times k}$: $G \mathbb{E} \eta_r A' G' = \mathbb{E} \eta_r A'$ by spherical symmetry of K , Lemma A.2 tells us that $\mathbb{E} \eta_r A' = \beta \mathbb{I}_k$. Passing to the trace, (A.15) yields that $\beta = 1$. Because of symmetry of the error distribution, $\mathbb{E} \eta_r = 0$ a.e. $K(dx)$. Thus, with $b_r := \alpha_r c_r$, η_r in fact is an IC as in HR (1994; Theorem 7.4.13.b), which form is sufficient to minimax asymptotic MSE.

Case ($*$ = c , $\alpha = 2$) The median of A_f is unique; so HR (1994; equation (134) in Theorem 7.4.15.b) applies, giving the minimax IC. Here $D = \mathbb{I}_k$; $\vartheta = 0$, by symmetry of the error distribution, and $\mathcal{K} = \mathcal{I} = \gamma \mathbb{I}_k$ for some $\gamma \in (0, \infty)$ (Lemma A.2 and $\mathcal{I}_f^{\text{loc}} = 1$). This gives (A.14)–(A.16). \square

Minimum Bias — Proof of (3.38) and (3.44) For $D = \mathbb{I}_k$, minimum bias ω_c^{\min} in model 3.6 (k -dim. location) and $\omega_{c,1}^{\min}$ in model 3.7.1 (k -dim. normal regression) according to HR (1994; Theorems 5.5.1.b and 7.4.13.c), are respectively given by

$$\omega_c^{\min} = \max \left\{ \frac{\text{tr } A}{\mathbb{E}|A\Lambda - a|} \mid a \in \mathbb{R}^k, A \in \mathbb{R}^{k \times k} \setminus \{0\} \right\} \quad (\text{A.17})$$

$$\omega_{c,1}^{\min} = \max \left\{ \frac{\text{tr } A}{\mathbb{E}|Ax| \mathbb{E}|A_f - m|} \mid A \in \mathbb{R}^{k \times k} \setminus \{0\} \right\} \quad (\text{A.18})$$

In our case, the median m of $A_f(u) = u$ under $F = \mathcal{N}(0, 1)$ is zero. Also in (A.17), we may put $a = 0$. Indeed, by triangle inequality and (spherical) symmetry of $\mathcal{L}(A)$, the zero centering vector $0 = \frac{1}{2}a + \frac{1}{2}(-a)$ would decrease the denominator $\mathbb{E}|A\Lambda - a| = \mathbb{E}|A\Lambda + a|$.

Despite of different scores functions, we can now handle both models in one proof, only drawing on the spherical symmetry of $\mathcal{L}(A)$.

By the singular value decomposition, $U'AV = \text{diag}(\alpha_1, \dots, \alpha_k) = V'A'U$ for some orthogonal matrices $U, V \in \mathbb{R}^{k \times k}$. Then

$$\mathbb{E}|A\Lambda| = \mathbb{E}|U'AV\Lambda| = \mathbb{E}|V'A'U\Lambda| = \mathbb{E}|A'\Lambda| \quad (\text{A.19})$$

Putting $A_s := \frac{1}{2}(A + A')$, the trace stays fixed, while $\mathbb{E}|A\Lambda|$ decreases (triangle inequality). So we may limit attention to symmetric matrices A . Since

$$\text{tr } GAG' = \text{tr } A, \quad \mathbb{E}|GAG'\Lambda| = \mathbb{E}|A\Lambda| \quad (\text{A.20})$$

for any orthogonal G , especially for G obtained from the spectral decomposition of A , we may further assume A diagonal, and then with all diagonal elements nonnegative. Differentiating the Lagrangian $L: [0, \infty)^k \rightarrow \mathbb{R}$,

$$L(a) = \mathbb{E}|AA| - \lambda \operatorname{tr} A \quad (\text{A.21})$$

where $A = \operatorname{diag}(\alpha_1, \dots, \alpha_k)$, it follows that $\alpha_1 = \dots = \alpha_k = \frac{1}{k}$. \square

B Tabulated Inefficiencies And Least Favorable Radii

B.1 k -Dimensional Location

relMSE over r/\sqrt{n} -contamination nbd ($k \geq 1$) and, in case $k = 1$, relVar over symmetric s -contamination nbd's, where $s = r^2/(1 + r^2)$.

k	relMSE ($\eta_\infty, 0$)	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	1.571	1.181	1.088	1.044	0.621	0.548	0.574
5	1.104	1.062	1.035	1.018	0.577	0.450	0.481
10	1.051	1.035	1.020	1.011	0.520	0.385	0.413
15	1.034	1.025	1.014	1.008	0.485	0.351	0.375

B.2 One-Dimensional Normal Scale

	relMSE ($\eta_\infty, 0$)	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
$* = c$	2.721	1.505	1.208	1.099	0.499	0.485	0.557
$* = v$	1.850	1.254	1.115	1.056	0.265	0.237	0.249

B.3 One-Dimensional Location and Scale

	relMSE ($\eta_\infty, 0$)	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
$* = c$	1.746	1.314	1.147	1.072	0.579	0.536	0.591

B.4 Regression, Average (Square) Contamination

Average Contamination Nbd ($* = c, \alpha = 1$)

$K(dx) = Ufo_k(0, m)$ see web-page.

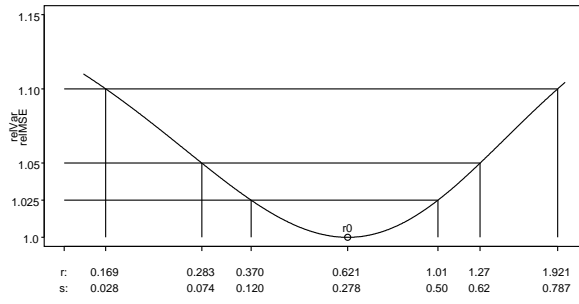
$K(dx) = \mathcal{N}(0, \sigma^2 \mathbb{I}_k)$

k	relMSE ($\eta_\infty, 0$)	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	2.467	1.347	1.146	1.070	0.515	0.474	0.496
5	1.735	1.231	1.107	1.053	0.583	0.514	0.542
10	1.651	1.207	1.098	1.049	0.598	0.526	0.553
15	1.624	1.199	1.095	1.047	0.605	0.532	0.558

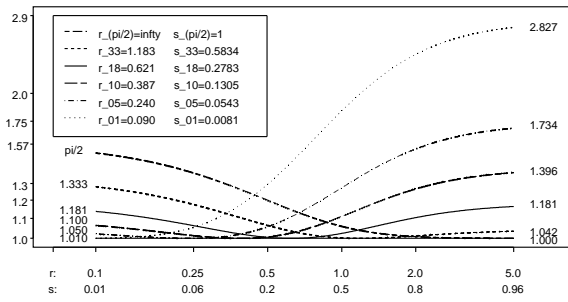
Average Square Contamination Nbd ($* = c, \alpha = 2$)

Same numbers as in one-dimensional location.

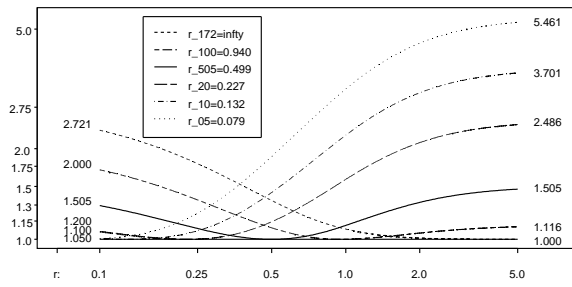
relVar, relMSE: 1-Dimensional Location (Var = 1.181 at r, s = 0)



1-Dimensional Location: relMSE, relVar vs. r, resp. $s=r^2/(1+r^2)$



relMSE: 1-Dimensional Scale (contamination)



relMSE: 1-Dimensional Scale (total variation)

