The Cost Of Not Knowing The Radius

Helmut Rieder
Matthias Kohl, Peter Ruckdeschel

University of Bayreuth, Germany

25 April 2006

Abstract

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

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 of our models, 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 and Phrases: Symmetric location and contamination; infinitesimal asymmetric neighborhoods; Hellinger, total variation, contamination; asymptotically linear estimators; influence curves; maximum asymptotic variance and mean square error; relative risk; inefficiency; subefficiency; least favorable radius; radius–minimax robust estimator; location, scale, Binomial, Poisson, Gamma, regression, AR(1), MA(1), ARMA(1,1), ARCH(1) models.

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

1 Introduction and Summary

1.1 Statistical Folklore

has it 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-dimensional nuisance parameter of robust neighborhood models. More abstractly, the model deviations may be treated as values of an infinite dimensional nuisance parameter; confer Rieder[1](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. Witting and Müller–Funk (1995; Anmerkung 6.44) declare the choice of the 'clipping constant c' to be of 'decisive importance' and continue: 'If c is large, the efficiency at the ideal model is large but robustness is bad, and the other way round for small c.' In Beispiel 7.4.5, they declare the radius to be unknown in practice, and hence a dubious 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; confer Bickel (1981; Théorème 8) and HR (1994; Example 6.1.1).

1.2 In Our Approach

the maximum risk of the estimator which is optimally robust for a neighborhood of radius r_0 will be evaluated over a neighborhood of radius r, and related 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 (compare 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.

[1] HR, henceforth

Remark 1.1 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. ////

Remark 1.2 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. ////

To appreciate these new notions, we take up the evaluations of Subsection 3.2.1 and consider the simplest possible example, one-dimensional 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).

1.3 Three 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 (loc.cit.), 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.

1.4 Comparison With Semiparametrics

Although the radius is a one-dimensional quantity, in connection with the robust neighborhoods it has infinite dimensional features. Therefore, a comparison with a basic semiparametric case suggests itself.

We assume the classical univariate location model with unknown symmetric error distribution F and density f of finite Fisher information $\mathcal{I}_F^{\mathrm{loc}} = \int (\Lambda_F^{\mathrm{loc}})^2 \, dF$,

where $\Lambda_F^{\mathrm{loc}} = -f'/f$, and consider the location M-estimate defined by some odd function $\psi_0 \colon \mathbb{R} \to \mathbb{R}$; for example, $\psi_0 = \Lambda_{F_0}^{\mathrm{loc}}$ 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. $\sim F$ will be asymptotically normal with asymptotic variance

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

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

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

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

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, confer HR (2001b)

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)

1.5 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-dimensional 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. Pfanzagel and Wefelmeyer (1982; Sec. 9.4) derive an asymptotic version for real-valued smooth functionals. The following extension to k-dimensional 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 $\Lambda_\theta(x, y) = \Lambda_F^{\mathrm{loc}}(y - x'\theta)\, x$ and Fisher information $\mathcal{I}_\theta = \mathcal{I}_F^{\mathrm{loc}}\, \mathcal{K}$.

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),

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

weakly, as $n \to \infty$, 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,

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

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

Remark 1.3 The result is contained in HR (2001b), where it is proved for more general i.i.d. models of location or scale structure. It is shown to hold also for MA(q)-models with innovation distribution F. An extension to AR(p)- and ARMA(p,q)-models with innovation distribution F is proved there with the bound $\geq 1-2^{-k}$ weakened to >0, and in addition assuming continuity of each S_n. The weaker result suffices to render the convergence of the adaptive estimators of Beran (1976) and Kreiss (1987) nonuniform as above. ////

On second look not so much the estimators are to be blamed for (1.2) and (1.4). Actually, the law of any estimator S_n is uniformly continuous in total variation since the distance decreases under a transformation of the measures. Rather the standardization by Fisher information in (1.3) should be questioned because of discontinuity in this strong metric: Fisher information of location/scale is vaguely lower semicontinuous, hence lower semicontinuous in total variation d_v, but not d_v-upper semicontinuous.

In robust statistics on the contrary, risk is evaluated uniformly, replacing asymptotic variance by its maximum, and Fisher information by its minimum, over symmetric contamination neighborhoods; likewise, asymptotic mean square error is maximized over shrinking neighborhoods. But, by simple set inclusions, the passage to the supremum $g(x,r) = \sup\big\{\, f(y) \;\big|\; y \in B(x,r) \,\big\}$ of any function f over balls $B(x,r)$ already implies continuity (relative to these balls) of $g(.\,,r)$ 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 with bounded radius has been established by Beran (1981), Millar (1981), Bickel (1981; Théorème 5), Huber (1981), and HR (1994; Chap. 6). This uniformity also underlies the present investigation of asymptotic risk.

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

The paper now proceeds as follows. In Section 2, the theoretical setup is formulated and the concepts are formally defined, and some general results are gathered from other references in Theorem 2.1.

In Section 3, for a comprehensive list of ideal models (see Subsection 3.1) and the neighborhood systems from Subsection 2.1.2, the optimally robust estimators and their risk functions are determined by specialization of the general results in HR (1994; Chaps. 5 and 7) and by suitable complements. The results of this

section are summarized in an introductory Subsection 3.2. Section 4 contains the mathematical proofs. The numerical algorithms are described in Section 5. The computed numbers are tabulated in Section 6. A selection of plots is attached.

2 General concept

We now introduce the notions of *inefficiency*, *subefficiency*, *least favorable radius*, and *radius–minimax procedure/IC* more formally.

2.1 Setup

2.1.1 Ideal model

For a measurable space $(\Omega, \mathcal{A})$ 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 $\mathcal{I}_\theta$ 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 $t\colon \Theta \to \mathbb{R}^p$ with $dt(\theta) = D_\theta$ for some matrix $D_\theta \in \mathbb{R}^{p\times k}$ with $\operatorname{rk} D_\theta = p \le k$.
For estimators, we assume asymptotically linear estimators (ALE's); that is, sequences $S = (S_n)$ of estimators $\big(S_n(X_1, \ldots, X_n)\big)$ such that, as $n \to \infty$

$$S_n - t(\theta) = \frac{1}{n}\sum_{i=1}^{n} \eta_\theta(X_i) + \mathrm{o}_{P_\theta^n}\big(\frac{1}{\sqrt{n}}\big) \tag{2.5}$$

for some (partial) influence curve (pIC) $\eta_\theta \in \Psi_\theta^{D_\theta}$, where the set $\Psi_\theta^{D_\theta}$ is defined as

$$\eta_\theta \in L_2^p(P_\theta), \qquad \mathrm{E}_\theta\, \eta_\theta = 0, \qquad \mathrm{E}_\theta\, \eta_\theta \Lambda_\theta' = D_\theta \tag{2.6}$$

As for this setup and definition, confer HR (1994; section 4.2).
In the somewhat more restricted one-dimensional location setup of Huber (1964); i.e., symmetric P_θ and monotone Λ_θ, the procedures are further specialized to location M-estimators.
The i.i.d. setup suitably extends to linear time series models, compare HR (2001a).

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

As general in Robust Statistics, these ideal models are enlarged to neighborhoods; more specifically, we consider

(a) symmetric contamination neighborhoods $\mathcal{V}$ of fixed size $s \in [0,1)$ about the ideal P, assumed symmetric about zero, which consist of the convex combinations

$$Q = (1-s)P + s\,H \tag{2.7}$$

with arbitrary unknown probability H, symmetric about 0.

These fixed neighborhoods, whose size does not depend on the sample size, are bound to one-dimensional location and Huber's (1964) minimax asymptotic variance approach.

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

$$Q_n = (1-r_n)P + r_n\,H_n \tag{2.8}$$

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

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

$$d_*(Q_n, P) \le r_n \tag{2.9}$$

where

$$d_v(Q,P) = \frac{1}{2}\int |dQ - dP| = \sup_A \big|Q(A) - P(A)\big| \tag{2.10}$$

$$d_h^2(Q,P) = \frac{1}{2}\int \big|\sqrt{dQ} - \sqrt{dP}\,\big|^2 \tag{2.11}$$

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

(c) conditional regression neighborhoods $\mathcal{U}_{*,\alpha}$ on the contrary, 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 neighborhood about the ideal $\Phi(\sigma_u^{-1}du)$ of radius $r\varepsilon(x)/\sqrt{n}$. The function $\varepsilon\colon \mathbb{R}^k \to [0,\infty)$, which weights the radius depending on the regressor, is called radius curve.

We employ conditional, or error-free-variables, neighborhoods with varying radius curves ε subject to $L_\alpha(K)$-norm $\|\varepsilon\|_\alpha \le 1$ for $\alpha = 1, 2, \infty$, respectively. The cases $\alpha = 1, 2$ are named average, respectively average square, conditional neighborhoods. The case $\alpha = \infty$ reduces to the fixed radius curve $\varepsilon_1 \equiv 1$.

Special treatments of error-free-variables regression neighborhoods go back to Huber (1983) and Bickel (1984). In general, confer HR (1994; Chap 7), where also the required MSE-optimality is obtained. HR (1987) derives a finite-sample minimax estimator for this type of regression neighborhoods.

With the past of the observations process serving as regressor, the conditional neighborhoods extend from regression to neighborhoods of transition probabilities in time series models; confer HR (2001a).

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 neighborhood of another (starting) radius $r \in [0,\infty)$ is

$$\operatorname{maxMSE}(\eta_{r_0}, r) = \mathrm{E}\,|\eta_{r_0}|^2 + r^2\omega_{*,\alpha}^2(\eta_{r_0}) \tag{2.12}$$

where $* = c, v, h$ and $\alpha = 1, 2, \infty$. 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,

$$\operatorname{relMSE}(\eta_{r_0}, r) = \frac{\operatorname{maxMSE}(\eta_{r_0}, r)}{\operatorname{maxMSE}(\eta_r, r)} \tag{2.13}$$

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

$$R(S) = \lim_{M\to\infty} \lim_{n} \sup_{Q_n \in \mathcal{U}_n} \int M \wedge \ell\Big(\sqrt{n}\,(S_n - \theta)\Big)\, dQ_n^n \tag{2.14}$$

With the usual identification of S and η, let $\sigma^2(\eta) = \mathrm{E}\,|\eta|^2$ 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

$$R_G(\eta) = G(r\omega(\eta), \sigma(\eta)) \tag{2.15}$$

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

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

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 (loc.cit.; (3.4)) determining the clipping height.

2.1.4 *G*-inefficiency

Corresponding to relMSE, we define the G-inefficiency ρ_G as

$$\rho_G(r', r) := \frac{G(r\omega_{r'}, \sigma_{r'})}{G(r\omega_r, \sigma_r)} \tag{2.16}$$

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

$$\bar{\rho}_G := \inf_{r' \in (r_l, r_u)} \sup_{r \in (r_l, r_u)} \rho_G(r', r) \tag{2.17}$$

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

$$\sup_{r \in (r_l, r_u)} \rho_G(r_0, r) = \bar{\rho}_G \tag{2.18}$$

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[2]. 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

The fact, that in our algorithms, we generally may obtain the least favorable radius $r_0 \in [r_l, r_u]$ as zero of the mapping $r \mapsto \rho_G(r, r_l) - \rho_G(r, r_u)$ is based upon the following properties of the mapping $(r, s) \mapsto \rho_G(r, s)$, ensuring the bowl-shape of the ρ_G curves as plotted in figures attached to this paper:

- for $s < r$, $s \mapsto \rho_G(r, s)$ is decreasing in s, and for $s > r$, $s \mapsto \rho_G(r, s)$ is increasing in s
- for each $r \in (0, \infty)$, $s \mapsto \rho_G(r, s)$ is continuous
- for $s < r$, $r \mapsto \rho_G(r, s)$ is increasing in r, and for $s > r$, $r \mapsto \rho_G(r, s)$ is decreasing in r

The following theorem guarantees these properties, as well as the risk-independence alluded to in Subsection 1.3. In particular, it covers the case of the MSE in all its assertions.

Theorem 2.1 (a) *Assume a risk of form (2.15) wtih G isotone in both arguments and totally differentiable, and that in addition G be homogeneous in the*

[2]Using access name `radius`, password `unknown`, the interested reader may try out our computer program under `http://www.uni-bayreuth.de/departments/math/org/mathe7/radius`.

sense that there is some function $\mu_G\colon \mathbb{R}_+ \to \mathbb{R}_+$ *such that for all* $\nu > 0$, $s > 0$, $w > 0$

$$G(\nu w, \nu s) = \mu_G(\nu)\, G(w, s) \tag{2.19}$$

Then for each $r' > 0$,

$$\sup_{r\in(r_l,r_u)} \rho_G(r', r) \le \mu_G\big(\frac{\omega_{r'}}{\omega_{r_u}}\big) \vee \mu_G\big(\frac{\sigma_{r'}}{\sigma_{r_l}}\big) \tag{2.20}$$

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

$$\mu_G\big(\frac{\omega_{r_1}}{\omega_{r_u}}\big) = \mu_G\big(\frac{\sigma_{r_1}}{\sigma_{r_l}}\big) \tag{2.21}$$

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

$$\sup_{r>0} \rho_G(r_0, r) = \mu_G\big(\frac{\omega_{r_0}}{\omega_{\min}}\big) = \mu_G\big(\frac{\sigma_{r_0}}{\sigma_{\min}}\big) = \bar\rho_G \tag{2.22}$$

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

$$\eta_b = (A\Lambda - a)\, \min\{1, \frac{b}{|A\Lambda - a|}\} \tag{2.23(c)}$$

$$\eta_b = c \vee A\Lambda \wedge (c + b) \tag{2.23(v)}$$

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*

$$\frac{\omega_b}{\omega_{\min}} = \frac{\sigma_b}{\sigma_{\min}} \tag{2.24}$$

Here $\omega_b = \omega(\eta_b)$, $\sigma_b = \sigma(\eta_b)$.

(c) *Assume in addition, that* G *be twice continuously differentiable. Then for any* $0 < r_l < r_u < \infty$

$$\sup_{r\in(r_l,r_u)} \frac{G(r\omega_{r'}, \sigma_{r'})}{G(r\omega_r, \sigma_r)} = \frac{G(r_l\omega_{r'}, \sigma_{r'})}{G(r_l\omega_{r_l}, \sigma_{r_l})} \vee \frac{G(r_u\omega_{r'}, \sigma_{r'})}{G(r_u\omega_{r_u}, \sigma_{r_u})} \tag{2.25}$$

and there is an $r_0 \in (r_l, r_u)$ *such that*

$$\frac{G(r_l\omega_{r_0}, \sigma_{r_0})}{G(r_l\omega_{r_l}, \sigma_{r_l})} = \frac{G(r_u\omega_{r_0}, \sigma_{r_0})}{G(r_u\omega_{r_u}, \sigma_{r_u})} \tag{2.26}$$

If in addition

$$\begin{array}{rcll} \tilde g(r; r') & = & G(r'\omega_r, \sigma_r) & \text{is increasing in } r \text{ for } r > r' \\ & & & \text{is decreasing in } r \text{ for } r < r' \end{array} \tag{2.27}$$

then

$$\sup_{r\in(r_l,r_u)} \rho_G(r_0, r) = \frac{G(r_l\omega_{r_0}, \sigma_{r_0})}{G(r_l\omega_{r_l}, \sigma_{r_l})} = \frac{G(r_u\omega_{r_0}, \sigma_{r_0})}{G(r_u\omega_{r_u}, \sigma_{r_u})} = \inf_{r\in(r_l,r_u)} \sup_{r\in(r_l,r_u)} \rho_G(r', r) \tag{2.28}$$

Assertions (a) and (b) are Ruckdeschel and Rieder (2004; Theorem 6.1); (c) is Kohl (2005; Lemma 2.2.3(b)) for MSE risk and $D = \mathbb{I}_p$, and, in the general case, is contained in Ruckdeschel (2005); the proof to this theorem uses the following

Proposition 2.2 *Let η_r be the pIC minimizing G-risk on some neighborhood of radius r. Then*

(a) *The Lagrange multipliers A_r, a_r, b_r, and c_r contained in η_r are uniformly bounded on bounded radius intervals $r \in (r_l, r_u)$, $0 < r_l < r_u < \infty$.*

(b) *For $r_n \in (0,\infty)$ such that $r_n \to r$ as $n \to \infty$,*

$$\operatorname{tr} A_{r_n} \to \operatorname{tr} A_r, \qquad b_{r_n} \to b_r, \qquad c_{r_n} \to c_r \tag{2.29}$$

If A_r and a_r are unique, also

$$A_{r_n} \to A_r, \qquad a_{r_n} \to a_r \tag{2.30}$$

(c) *Let ω_r, σ_r^2 denote bias and variance of η_r. Then σ_r is increasing and ω_r decreasing in r.*

For $D = \mathbb{I}_k$, this is Kohl (2005; Propositions 2.1.7, 2.1.9, Lemma 2.2.1). For general D, this proposition is proven in Ruckdeschel (2005).

3 Optimally Robust Estimates and Their Inefficiency Curves

3.1 List of 11 Ideal Models Considered

In this paper, we consider the following list of ideal models, which is neither too small, nor of course exhaustive, but somehow representative for our results:

3.1.1 k-dimensional normal location:

$$y_i = \theta + u_i \tag{3.1}$$

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:

$$y_i = \theta\, u_i \tag{3.2}$$

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 exponential scale:

$$y_i = \theta\, u_i \tag{3.3}$$

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

3.1.4 One-dimensional normal location and scale:

$$y_i = \mu + \sigma u_i \tag{3.4}$$

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

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

3.1.5 Binomial model:

$$y_i \text{ i.i.d. } \sim \text{Binomial}(n, \theta) \tag{3.5}$$

with parameter $\theta \in (0,1)$. The scores and Fisher information are given by $\Lambda_\theta(y) = (y - n\theta)/(\theta(1-\theta))$ and $\mathcal{I}_\theta = n/(\theta(1-\theta))$.

3.1.6 Poisson model:

$$y_i \text{ i.i.d. } \sim \text{Poisson}(\theta) \tag{3.6}$$

with parameter $\theta \in (0, \infty)$. The scores and Fisher information are given by $\Lambda_\theta(y) = \theta^{-1} y - 1$ and $\mathcal{I}_\theta = \theta^{-1}$.

3.1.7 Gamma model:

$$y_i \text{ i.i.d. } \sim \text{Gamma}(\sigma, \alpha) \tag{3.7}$$

with density

$$f_{\sigma,\alpha}(y) = (\sigma\Gamma(\alpha))^{-1} (\sigma^{-1}y)^{\alpha-1} \exp(-\sigma^{-1}y)\, \mathbf{I}(y \geq 0) \tag{3.8}$$

and parameter $(\sigma, \alpha)' \in (0, \infty) \times (0, \infty)$. We prefer to work with $\theta' = (\tau, \alpha) \in \mathbb{R} \times (0, \infty)$, where $\tau = \log(\sigma)$, as in this parametrization the scores depend on τ respectively, σ only via $z := \sigma^{-1}y$. Moreover the Fisher information is independent of τ. The scores and Fisher information are

$$\Lambda_\theta(y) = \begin{pmatrix} z - \alpha \\ \log(z) - \text{di}\Gamma(\alpha) \end{pmatrix} \qquad \mathcal{I}_\theta = \begin{pmatrix} \alpha & 1 \\ 1 & K_\alpha - \text{di}\Gamma(\alpha)^2 \end{pmatrix}$$

where $\text{di}\Gamma(\cdot)$ denotes the Digamma function, $\text{di}\Gamma(\alpha) = \frac{d}{d\alpha} \log \Gamma(\alpha)$, and $K_\alpha := \text{E} \log(z)^2$.
The scores and the Fisher information for the parametrization $(\sigma, \alpha)'$ then read $\Lambda_{\sigma,\alpha}(y) = D'^{-1} \Lambda_\theta(y)$ and $\mathcal{I}_{\sigma,\alpha} = D'^{-1} \mathcal{I}_\theta D^{-1}$, where $D = \partial g / \partial(\tau, \alpha)'$ and $g: \mathbb{R} \times (0, \infty) \to (0, \infty) \times (0, \infty)$, $(\tau, \alpha) \mapsto (e^\tau, \alpha)$.

3.1.8 k-dimensional normal linear regression:

$$y_i = x_i'\theta + u_i \tag{3.9}$$

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 $\Lambda_\theta(x,y) = \sigma_u^{-2}(y - x'\theta)\,x$ and $\mathcal{I}_\theta = \sigma_u^{-2}\,\mathrm{E}_K\,xx'$.

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

3.1.9 Order one autoregression and moving average:

$$y_i = \theta y_{i-1} + u_i \tag{3.10}$$

respectively

$$y_i = u_i - \theta u_{i-1} \tag{3.11}$$

with parameter $|\theta| < 1$, innovations u_i i.i.d. $\sim \mathcal{N}(0,\sigma_u^2)$, scale $\sigma_u \in (0,\infty)$ known. The scores are $\Lambda_{\theta,i} = (\pm)\sigma_u^{-2}u_i\sum_{j\geq 0}\theta^j u_{i-1-j}$ and $\mathcal{I}_\theta = (1-\theta^2)^{-1}$ the Fisher information, in the two models.

3.1.10 Order (1,1) autoregressive-moving average:

$$y_i + \phi y_{i-1} = u_i + \xi u_{i-1} \tag{3.12}$$

with parameter $\theta = (\phi,\xi)'$ and $|\phi|, |\xi| < 1$, innovations u_i i.i.d. $\sim \mathcal{N}(0,1)$. The scores and Fisher information are $\Lambda_{\theta,i} = u_i H_{\theta,i}$ with

$$H_{\theta,i} := \textstyle\sum_{j=1}^{\infty}(-1)^{j-1}(-\phi^{j-1},\xi^{j-1})'V_{i-j} \tag{3.13}$$

and

$$\mathcal{I}_\theta \;=\; \mathcal{K}_\theta = \mathrm{Cov}_\theta\, H_{\theta,1} = \begin{pmatrix} (1-\phi^2)^{-1} & -(1-\phi\xi)^{-1} \\ -(1-\phi\xi)^{-1} & (1-\xi^2)^{-1} \end{pmatrix} \tag{3.14}$$

where $H_{\theta,1} \sim \mathcal{N}_2(0,\mathcal{K}_\theta)$; i.e., $H_{\theta,1}$ is elliptical symmetric.

3.1.11 Order one auto-regressive conditional heteroscedastic:

$$y_i = (1 + \theta y_{i-1}^2)^{1/2} u_i \tag{3.15}$$

with parameter $\theta \in [0, 3.562)$, innovations u_i i.i.d. $\sim \mathcal{N}(0,1)$. The scores and Fisher information are $\Lambda_{\theta,i} = (u_i^2 - 1)H_{\theta,i}$ with

$$H_{\theta,i} := \frac{y_{i-1}^2}{2(1+\theta y_{i-1}^2)} \in \Big[0,\; \frac{1}{2\theta}\Big) \tag{3.16}$$

and $\mathcal{I}_\theta = 2\mathcal{K}_\theta = 2\mathrm{Cov}_\theta\, H_{\theta,1}$.

Remark 3.1 (Simplifications as to the choice of the parameter) In models 3.1.1–3.1.8, 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

$$\sigma_u = 1\,, \qquad \sigma_x = 1\,, \qquad m_x = 1 \tag{3.17}$$

Moreover, $\theta = 0$ may be fixed in models (a) and (h), $\theta = 1$ in models 3.1.2 and 3.1.3, $\theta = (0,1)'$ in model 3.1.4 and $\theta = (0,\alpha)'$ in model 3.1.6, due to equivariance of these models.

In models 3.1.8, the normal $\mathcal{N}(0,\sigma_x^2)$ with $\sigma_x^2 = \sigma_u^2/(1-\theta^2)$ plays the role of the regressor distribution. Therefore, by the invariance stated for model 3.1.7, the inefficiencies turn out the same for all values $|\theta| < 1$ and $\sigma_u \in (0,\infty)$. ////

For all considered models, we also spell out the minimax robust estimators and the minimum bias explicitly, which are given in general form by HR (1994; Chap. 5 and 7).

3.2 Evaluation in particular models — a summary

3.2.1 One-Dimensional Robust Location

Neighborhoods of fixed size: In Huber's (1964, 1981) approach, the ideal standard normal location model is enlarged to symmetric contamination neighborhoods $\mathcal{V}$ 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 $s \to 100\%$), and it even more improves on the 90.8% subefficiency (attained for $s \to 1$, 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 $.12 \le s \le .50$, below 5% for $.074 \le s \le .62$, and still below 10% for $.028 \le s \le .78$.

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 neighborhood of whatever size $s \in [0,1)$ but, at the same time, within 2.5% of the minimax risk for any size $12\% \le s \le 50\%$, within 5% for any size $7.4\% \le s \le 62\%$, and still within 10% of the minimax risk for arbitrary size $2.8\% \le s \le 78\%$.

Remark 3.2 Via relation (3.18) below, $s_0 = .278$ corresponds to $r_0 = .62$, and the corresponding radius intervals about r_0 read:

$$.37 \le r \le 1.01 \ (2.5\%), \ .29 \le r \le 1.27 \ (5\%), \ .17 \le r \le 1.92 \ (10\%). \qquad ////$$

Shrinking Neighborhoods: 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:

$$s = \frac{r^2}{1 + r^2} \tag{3.18}$$

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.

3.2.2 One-Dimensional Robust Scale

centered at the standard normal already demonstrates the limitations of the minimax asymptotic variance approach; confer Huber (1981; Sec. 5.7, p 124).

From now on, therefore, including scale, the infinitesimal robust setup is used, employing neighborhoods of radius $r/\sqrt{n}$ at size n of the i.i.d. laws, asymptotically linear estimators, and asymptotic mean square error. In the scale model, the neighborhoods 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 Our numbers obtained in the asymptotic minimax MSE approach refer to contamination neighborhoods. For univariate location 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-dimensional location and regression models of our study. ////

Due to symmetry of the scale scores, the relation between the infinitesimal contamination and total variation systems does not extend to the scale model. 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; confer Subsection 3.4.

Nevertheless, the $1:2$ relation seems to hold at least approximately for the least favorable radii; by numerical evaluation, they are $r_0 = .27$, $r_3 = .24$, and $r_2 = .25$. 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 $\rho = \frac{1}{3}, \frac{1}{2}$; confer Subsection 6.2.

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.

A summary of the results in the selected k-dimensional location and regression models follows.

3.2.3 k-Dimensional Robust Location

about the k-variate standard normal enlarged by $r/\sqrt{n}$-contamination neighborhoods 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

$$\omega_c^{\min} = \frac{k\,\Gamma(\frac{k}{2})}{\sqrt{2}\,\Gamma(\frac{k+1}{2})} \approx \sqrt{k} \qquad \text{as } k \to \infty \tag{3.19}$$

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 $k \to \infty$; confer Subsections 4.6 and 6.1.

3.2.4 Infinitesimal Neighborhood Regression

will be about the classical k-dimensional linear regression model, in which the normally distributed errors and the regressors are stochastically independent. The regressor distribution K is assumed spherically symmetric; especially we choose

the uniform $\mathrm{Ufo}_k(0,m)$ on a centered ball of radius m in $\mathbb{R}^k$, and $\mathcal{N}(0,\sigma^2\mathbb{I}_k)$, a scalar multiple of the k-dimensional standard normal.

Unconditional, or errors-in-variables, neighborhoods are about the joint law of regressor and error; in particular, the regressor distribution may be distorted, too. Conditional, or error-free-variables, regression neighborhoods, 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 neighborhoods, unconditional and conditional, radius curves, confer Subsubsection 2.1.2 and HR (1994; Chap. 7), who also provides the required MSE-optimality.

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

Average (Square) Conditional Contamination: 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-dimensional location model with infinitesimal contamination neighborhoods; confer Subsubsection 3.11.2. Minimax inefficiencies and least favorable radii, therefore, are the same as for one-dimensional location.

The Hampel–Krasker estimates are optimally robust in the 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-dimensional location) as $k \to \infty$. 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-dimensional 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; confer the tables in Subsubsection 6.8.1. The convergence of the inefficiency curves to those of the one-dimensional location model (limit in case $\alpha = 2$ attained for each k) seems to hold also in the case $\alpha = \infty$, and is visible at least in the case K uniform (first table of Subsection 6.10.1).

Constant Conditional Neighborhoods: The regression neighborhood models with $\alpha = \infty$ of either contamination or Hellinger type may be reduced to the constant radius curve $\varepsilon_1 \equiv 1$. They appear atypical in several respects: (i) nonattainability of the infimum bias, (ii) infimum bias zero in the case of normal regressor distribution, (iii) unbounded L_2-norm of order $\mathrm{o}(r^2)$ as $r \to \infty$ of the influence curve which is optimally robust for radius r, (iv) slow convergence of the inefficiency curves as $k \to \infty$, and (v) relatively large least favorable radii; confer Subsections 3.13 and 6.10.

Average (Square) Conditional Hellinger Neighborhoods: as already mentioned, are essentially smaller than gross-errors neighborhoods of the same radius. They lead to a different type of robust influence curves and estimators (regressor clipped, residual unchanged). Contrary to the scale model (with total variation vs. contamination balls), estimation in the smaller Hellinger neighborhood system is more stable under radius misspecification in comparison with contamination neigh-

borhoods, as the inefficiency numbers and least favorable radii we compute are smaller. For increasing dimension, the limit (with respect to inefficiency) is that of k-dimensional location with contamination neighborhoods.

For robust regression based on average conditional infinitesimal Hellinger balls, the inefficiencies are identical (K normal), respectively (K uniform) tend, to those for k-dimensional location and infinitesimal contamination neighborhoods (rescaled by $\sqrt{8}$), as $k \to \infty$. The convergence also holds true with conditional Hellinger neighborhoods of type $\alpha = \infty$, though at a slower rate. For average square conditional Hellinger balls, the minimax subefficiency is zero, which, as in the case of average square conditional contamination neighborhoods and one-dimensional location as the corresponding limit, already is the limiting case of k-dimensional location with contamination neighborhoods for $k \to \infty$.

Remark 3.4 The two limits for increasing dimension (different for the Hellinger and contamination systems) depend on our choice of regressor distributions, and may explained by the fact that $|x| \approx 1$ in the case $K(dx) = \mathrm{Ufo}_k(0, m)$, 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 Subsections 6.8–6.10, the first lines would be valid for all other dimensions as well. ////

3.3 One-Dimensional Location

3.3.1 Minimax Asymptotic Variance

We consider the $k = 1$ dimensional standard normal location model 3.1.1 first with symmetric contamination neighborhoods 2.1.2 (a).

The minimax M-estimate for size $s \in [0, 1)$ given by Huber (1964) is defined by

$$\psi_s(u) = (-m_s) \vee u \wedge m_s\,, \qquad \tfrac{s}{1-s}\, m_s = \mathrm{E}\left(|u| - m_s\right)_+ \tag{3.20}$$

For size $s = 1$, we take ψ_1 from the median,

$$\psi_1(u) = \mathrm{sign}\,(u) \tag{3.21}$$

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 neighborhood of fixed size $s \in [0, 1)$ is

$$\mathrm{maxVar}\,(\psi_{s_0}, s) = \frac{(1-s)\,\mathrm{E}\,\psi_{s_0}^2 + s\, m_{s_0}^2}{[(1-s)\,\mathrm{E}\,\psi'_{s_0}]^2} \tag{3.22}$$

respectively, in the case of the median,

$$\mathrm{maxVar}\,(\psi_1, s) = \frac{\pi}{2(1-s)^2} \tag{3.23}$$

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

$$\mathrm{relVar}\,(\psi_{s_0}, s) = \frac{\mathrm{maxVar}\,(\psi_{s_0}, s)}{\mathrm{maxVar}\,(\psi_s, s)} \tag{3.24}$$

Although $\mathrm{maxVar}\,(\psi_1, 1) = \infty$, the median is approximately optimal for neighborhood size $s \to 1$, as not only $\psi_s/m_s \to \psi_1$ pointwise but, more conclusively, we show that

$$\lim_{s\to 1} \mathrm{relVar}\,(\psi_1, s) = 1 \tag{3.25}$$

3.3.2 Minimax Asymptotic MSE for $r/\sqrt{n}$ -Contamination Balls

We consider the $k = 1$ dimensional standard normal location model 3.1.1 secondly with infinitesimal contamination neighborhoods 2.1.2 (b).

The minimax IC η_r for radius $r \in [0,\infty)$ is

$$\eta_r(u) = A_r u \min\left\{1,\, c_r|u|^{-1}\right\} \tag{3.26}$$

where

$$1 = A_r\,\mathrm{E}\,|u| \min\left\{|u|,\, c_r\right\}, \qquad r^2\, c_r = \mathrm{E}\left(|u| - c_r\right)_+ \tag{3.27}$$

as given by HR (1994; Theorem 5.5.7). For $r = \infty$, HR (1994; Theorem 5.5.1.b) supplies

$$\eta_\infty\,(u) = \omega_c^{\min}\,\mathrm{sign}\,(u) \tag{3.28}$$

which is the IC of the median and achieves minimum bias

$$\omega_c^{\min} = \left(\mathrm{E}\,|\Lambda|\,\right)^{-1} = \sqrt{\tfrac{\pi}{2}} \tag{3.29}$$

In Subsection 4.2 we prove the following relation between the maximum risks of the optimal estimates in the two models 3.3.1 and 3.3.2,

$$(1-s)\,\mathrm{maxMSE}\,(\eta_{r_0}, r) = \mathrm{maxVar}\,(\psi_{s_0}, s) \tag{3.30}$$

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

$$s = r^2/(1+r^2), \qquad s_0 = r_0^2/(1+r_0^2) \tag{3.31}$$

Consequentially, by (3.30) and (3.31), the inefficiency curves coincide in the two models,

$$\mathrm{relMSE}\,(\eta_{r_0}, r) = \mathrm{relVar}\,(\psi_{s_0}, s) \tag{3.32}$$

3.4 One-Dimensional Normal Scale

3.4.1 $r/\sqrt{n}$ -Contamination Balls

We consider the one-dimensional standard normal scale model 3.1.2 first with infinitesimal contamination neighborhoods 2.1.2 (b).

The minimax IC η_r of HR (1994; Theorem 5.5.7) for radius $r \in [0,\infty)$ is

$$\eta_r(u) = A_r(u^2 - \alpha_r^2)\min\left\{1, \frac{c_r}{|u^2-\alpha_r^2|}\right\} \tag{3.33}$$

where

$$0 = \mathrm{E}(u^2 - \alpha_r^2)\min\left\{1, \frac{c_r}{|u^2-\alpha_r^2|}\right\} \tag{3.34}$$

$$1 = A_r\,\mathrm{E}\,|u^2-\alpha_r^2|\min\left\{|u^2-\alpha_r^2|,\, c_r\right\} \tag{3.35}$$

and

$$r^2\, c_r = \mathrm{E}\left(|u^2-\alpha_r^2| - c_r\right)_+ \tag{3.36}$$

The parabola $u^2-\alpha_r^2$ in (3.33) is clipped only from above for radius $r \le 0.920$, and for $r \ge 0.920$ 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 $r = \infty$, from HR (1994; Theorem 5.5.1.b) we take

$$\eta_\infty(u) = \omega_c^{\min}\,\mathrm{sign}\,(|u| - \alpha_\infty) \tag{3.37}$$

which is the IC of the median absolute deviation $\mathrm{med}\,(|u_i|)/\alpha_\infty$, attaining minimum bias

$$\omega_c^{\min} = \left(\mathrm{E}\,|u^2-\alpha_\infty^2|\right)^{-1} = \left(4\alpha_\infty\varphi(\alpha_\infty)\right)^{-1} \approx 1.166 \tag{3.38}$$

3.4.2 $r/\sqrt{n}$ -Total Variation Balls

We consider the one-dimensional standard normal scale model 3.1.2 secondly with infinitesimal total variation neighborhoods 2.1.2 (b).

As minimax IC for radius $r \in [0,\infty)$, HR (1994; Theorem 5.5.7) supplies

$$\eta_r(u) = A_r\{[\,g_r \vee u^2 \wedge (g_r + c_r)] - 1\} \tag{3.39}$$

where

$$0 = \mathrm{E}(g_r - u^2)_+ - \mathrm{E}(u^2 - g_r - c_r)_+ \tag{3.40}$$

$$1 = A_r\,\mathrm{E}\,u^2\big\{[\,g_r \vee u^2 \wedge (g_r + c_r)] - 1\big\} \tag{3.41}$$

and

$$r^2\, c_r = \mathrm{E}(g_r - u^2)_+ \tag{3.42}$$

For $r = \infty$, HR (1994; Theorem 5.5.5.b) provides

$$\eta_\infty(u) = \omega_v^{\min}\big\{P(|u|<1)\,\mathbf{I}(|u|>1) - P(|u|>1)\,\mathbf{I}(|u|<1)\big\} \tag{3.43}$$

with minimum bias

$$\omega_v^{\min} = (\mathrm{E}\,\Lambda_+)^{-1} = \sqrt{\tfrac{\pi}{2}\,e} \approx 2.066 \tag{3.44}$$

3.5 One-Dimensional Exponential Scale

3.5.1 $r/\sqrt{n}$ -Contamination Balls

We consider the one-dimensional exponential scale model 3.1.3 first with infinitesimal contamination neighborhoods 2.1.2 (b).

The minimax IC η_r of HR (1994; Theorem 5.5.7) for radius $r \in [0,\infty)$ is

$$\eta_r(u) = A_r(u-\alpha_r)\min\left\{1,\ \frac{c_r}{|u-\alpha_r|}\right\}\mathbf{I}(u \geq 0) \tag{3.45}$$

where

$$0 = \mathrm{E}(u-\alpha_r)\min\left\{1,\ \frac{c_r}{|u-\alpha_r|}\right\} \tag{3.46}$$

$$1 = A_r\,\mathrm{E}\,|u-\alpha_r|\min\{|u-\alpha_r|,\ c_r\} \tag{3.47}$$

and

$$r^2\ c_r = \mathrm{E}\left(|u-\alpha_r| - c_r\right)_+ \tag{3.48}$$

The centering constant α_r decreases from $\alpha_0 = 1$ to $\alpha_\infty = \log(2) \approx 0.693$.

For $r = \infty$, from HR (1994; Theorem 5.5.5.1.b) we take

$$\eta_\infty(u) = \omega_c^{\min}\,\mathrm{sign}\,(u-\alpha_\infty)\,\mathbf{I}(u \geq 0) \tag{3.49}$$

attaining minimum bias

$$\omega_c^{\min} = \left(\mathrm{E}\,|u-\alpha_\infty|\right)^{-1} = \log(2)^{-1} \approx 1.443 \tag{3.50}$$

3.5.2 $r/\sqrt{n}$ -Total Variation Balls

We consider the one-dimensional exponential scale model 3.1.3 secondly with infinitesimal total variation neighborhoods 2.1.2 (b).

As minimax IC for radius $r \in [0,\infty)$, HR (1994; Theorem 5.5.7) supplies

$$\eta_r(u) = A_r\{[g_r \vee u \wedge (g_r + c_r)] - 1\}\,\mathbf{I}(u \geq 0) \tag{3.51}$$

where

$$0 = \mathrm{E}(g_r - u)_+ - \mathrm{E}(u - g_r - c_r)_+ \tag{3.52}$$

$$1 = A_r\,\mathrm{E}\,u\big\{[g_r \vee u \wedge (g_r + c_r)] - 1\big\} \tag{3.53}$$

and

$$r^2\ c_r = \mathrm{E}(g_r - u)_+ \tag{3.54}$$

For $r = \infty$, HR (1994; Theorem 5.5.5.b) provides

$$\eta_\infty(u) = \omega_v^{\min}\big\{P(u<1)\,\mathbf{I}(u>1) - P(u>1)\,\mathbf{I}(u<1)\big\} \tag{3.55}$$

with minimum bias

$$\omega_v^{\min} = (\mathrm{E}\,\Lambda_+)^{-1} = e^{-1} \approx 2.718 \tag{3.56}$$

3.6 One-Dimensional Location and Scale, Contamination

We consider the one-dimensional standard normal location and scale model 3.1.4 with infinitesimal contamination neighborhoods 2.1.2 (b).

The minimax IC η_r of HR (1994; Theorem 5.5.7) for radius $r \in [0,\infty)$ is

$$\eta_r(u) = \begin{pmatrix}\eta_r^{\mathrm{loc}}(u)\\ \eta_r^{\mathrm{sc}}(u)\end{pmatrix} = \begin{pmatrix}A_r^{\mathrm{loc}}u\\ A_r^{\mathrm{sc}}(u^2-\alpha_r^2)\end{pmatrix} w(u) \tag{3.57}$$

with

$$w(u) = \min\Big\{1, \frac{b}{\big[(A_r^{\mathrm{loc}})^2 u^2 + (A_r^{\mathrm{sc}})^2 (u^2 - \alpha_r^2)^2\big]^{1/2}}\Big\} \tag{3.58}$$

where

$$0 = \mathrm{E}(u^2 - \alpha_r^2)\, w(u) \tag{3.59}$$

$$1 = A_r^{\mathrm{loc}}\, \mathrm{E}\, u^2\, w(u) \tag{3.60}$$

$$1 = A_r^{\mathrm{sc}}\, \mathrm{E}(u^2 - \alpha_r^2)^2\, w(u) \tag{3.61}$$

and

$$r^2 b = \mathrm{E}\Big(\big[(A_r^{\mathrm{loc}})^2 u^2 + (A_r^{\mathrm{sc}})^2 (u^2 - \alpha_r)^2\big]^{1/2} - b\Big)_+ \tag{3.62}$$

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 $r = \infty$, from HR (1994; Theorem 5.5.1.b) we get

$$\eta_\infty(u) = \omega_c^{\min} \begin{pmatrix} u \\ A_\infty(u^2 - \alpha_\infty) \end{pmatrix} |T|^{-1} \tag{3.63}$$

with

$$|T| := \big[u^2 + A_\infty^2 (u^2 - \alpha_\infty)^2\big]^{1/2} \tag{3.64}$$

attaining minimum bias

$$\omega_c^{\min} = \max\Big\{\frac{1 + A_\infty}{\mathrm{E}\,|T|}\,\Big|\,\alpha_\infty, A_\infty \in \mathbb{R}\Big\} \approx 1.618 \tag{3.65}$$

where $A_\infty \approx 0.792$.

3.7 k-Dimensional Location, Contamination

We consider the k-dimensional normal location model 3.1.1 and infinitesimal contamination neighborhoods 2.1.2 (b).

The minimax IC η_r for radius $r \in [0, \infty)$ given by HR (1994; Theorem 5.5.7), due to spherical symmetry (Lemma 4.3 below), is

$$\eta_r(u) = \alpha_r u \min\big\{1,\, c_r |u|^{-1}\big\} \tag{3.66}$$

where

$$k = \alpha_r\, \mathrm{E}\,|u| \min\big\{|u|,\, c_r\big\}, \qquad r^2\, c_r = \mathrm{E}\,\big(|u| - c_r\big)_+ \tag{3.67}$$

For $r = \infty$, we put

$$\eta_\infty = \omega_c^{\min} \frac{u}{|u|} \tag{3.68}$$

which is the IC of the minimum L_1–estimate, and attains minimum bias $\omega_c^{\min}$; confer HR (1994; Theorem 5.5.1.b). In Subsection 4.4 we show that

$$\omega_c^{\min} = \frac{k}{\mathrm{E}\,|\Lambda|} = \frac{k\,\Gamma(\frac{k}{2})}{\sqrt{2}\,\Gamma(\frac{k+1}{2})} \tag{3.69}$$

For increasing dimension, we prove in Subsection 4.6 that

$$\lim_{k\to\infty} \frac{\omega_c^{\min}}{\sqrt{k}} = \lim_{k\to\infty} \frac{\mathrm{E}\,|\eta_\infty|^2}{k} = 1 \tag{3.70}$$

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,

$$\lim_{k\to\infty} \frac{\mathrm{maxMSE}\,(\eta_{r_0}, r)}{\mathrm{maxMSE}\,(\eta_\infty, r)} = 1 \tag{3.71}$$

where the convergence is uniform on bounded r_0-, r-intervals, this most robust estimate also becomes approximately radius–minimax.

3.8 Binomial model

3.8.1 $r/\sqrt{n}$ -Contamination Balls

We consider the Binomial model 3.1.5 first with infinitesimal contamination neighborhoods 2.1.2 (b).

The minimax IC η_r of HR (1994; Theorem 5.5.7) for radius $r \in [0,\infty)$ is

$$\eta_r(y) = A_r(y - a_r) \min\left\{1, \frac{c_r}{|y - a_r|}\right\} \tag{3.72}$$

where

$$0 = \mathrm{E}(y - a_r) \min\left\{1, \frac{c_r}{|y - a_r|}\right\} \tag{3.73}$$

$$1 = A_r\,\mathrm{E}\,|x - a_r| \min\{|y - a_r|,\, c_r\} \tag{3.74}$$

and

$$r^2\, c_r = \mathrm{E}\,\big(|y - a_r| - c_r\big)_+ \tag{3.75}$$

For $r = \infty$, from HR (1994; Theorem 5.5.1.b) we take

$$\eta_\infty\,(u) = \omega_c^{\min}\big[\,\mathbf{I}(x > m) - \mathbf{I}(x < m)\big] + b\,\mathbf{I}(x = m) \tag{3.76}$$

where $m = \mathrm{med}(x)$ and

$$b\,P(x = m) = \omega_c^{\min}\big[P(x < m) - P(x > m)\big] \tag{3.77}$$

attaining minimum bias

$$\omega_c^{\min} = \theta(1-\theta)\big(\,\mathrm{E}\,|x - m|\big)^{-1} \tag{3.78}$$

For $n = 1$ the minimax IC η_r is

$$\eta_r(y) = \mathcal{I}^{-1}\Lambda(y) \tag{3.79}$$

for all $r \in [0,\infty]$. Consequentially, $\mathrm{relMSE}\,(\eta_{r_0}, r) \equiv 1$ for all $r_0, r \in [0,\infty]$.

For $n \to \infty$ we have

$$\lim_{n\to\infty} \mathrm{relMSE}\,(\eta_{r_0}, r) = \mathrm{relMSE}\,(\eta_{r_0}^{1.\mathrm{loc}}, r) \tag{3.80}$$

where $\eta_{r_0}^{1.\mathrm{loc}}$ is given by (3.26) and (3.27). Moreover, if we consider $n\theta_n \to \vartheta$, we obtain

$$\lim_{n\to\infty} \mathrm{relMSE}\,(\eta_{r_0}, r) = \mathrm{relMSE}\,(\eta_{r_0}^{\mathrm{Pois}(\vartheta)}, r) \tag{3.81}$$

with $\eta_{r_0}^{\mathrm{Pois}(\vartheta)}$ given by (3.91)-(3.94).

3.8.2 $r/\sqrt{n}$ -Total Variation Balls

We consider the Binomial model 3.1.5 secondly with infinitesimal total variation neighborhoods 2.1.2 (b).

As minimax IC for radius $r \in [0,\infty)$, HR (1994; Theorem 5.5.7) supplies

$$\eta_r(y) = A_r\{[g_r \vee y \wedge (g_r + c_r)] - n\theta\} \tag{3.82}$$

where

$$0 = \mathrm{E}(g_r - y)_+ - \mathrm{E}(y - g_r - c_r)_+ \tag{3.83}$$

$$1 = A_r\,\mathrm{E}\,y\big\{[g_r \vee y \wedge (g_r + c_r)] - n\theta\big\} \tag{3.84}$$

and

$$r^2\, c_r = \mathrm{E}(g_r - y)_+ \tag{3.85}$$

For $n = 1$ we get $\eta_r(y) \equiv \mathcal{I}_\theta^{-1}\Lambda_\theta(y) = y - \theta$.

For $r = \infty$, HR (1994; Theorem 5.5.5.b) provides

$$\eta_\infty\,(y) = \frac{\omega_v^{\min}}{P(y \neq \theta)}\left\{P(y < \theta)\,\mathbf{I}(y > \theta) - P(y > \theta)\,\mathbf{I}(y < \theta)\right\} \tag{3.86}$$

with minimum bias

$$\omega_v^{\min} = \theta(1-\theta)\,(\mathrm{E}(y - n\theta)_+)^{-1} \tag{3.87}$$

For $n = 1$ the minimax IC η_r is

$$\eta_r(y) = \mathcal{I}^{-1}\Lambda(y) \tag{3.88}$$

for all $r \in [0,\infty]$. Consequentially, $\mathrm{relMSE}\,(\eta_{r_0}, r) \equiv 1$ for all $r_0, r \in [0,\infty]$.

For $n \to \infty$ we have

$$\lim_{n\to\infty} \mathrm{relMSE}\,(\eta_{r_0}, r) = \mathrm{relMSE}\,(\eta_{r_0}^{1.\mathrm{loc}}, r) \tag{3.89}$$

where $\eta_{r_0}^{1.\mathrm{loc}}$ is given by (3.26) and (3.27). Moreover, if we consider $n\theta_n \to \vartheta$, we obtain

$$\lim_{n\to\infty} \mathrm{relMSE}\,(\eta_{r_0}, r) = \mathrm{relMSE}\,(\eta_{r_0}^{\mathrm{Pois}(\vartheta)}, r) \tag{3.90}$$

with $\eta_{r_0}^{\mathrm{Pois}(\vartheta)}$ given by (3.99)-(3.102).

3.9 Poisson model

3.9.1 $r/\sqrt{n}$ -Contamination Balls

We consider the Poisson model 3.1.6 first with infinitesimal contamination neighborhoods 2.1.2 (b).

The minimax IC η_r of HR (1994; Theorem 5.5.7) for radius $r \in [0,\infty)$ is

$$\eta_r(y) = A_r(y-a_r)\min\left\{1, \frac{c_r}{|y-a_r|}\right\} \tag{3.91}$$

where

$$0 = \mathrm{E}(y-a_r)\min\left\{1, \frac{c_r}{|y-a_r|}\right\} \tag{3.92}$$

$$1 = A_r\,\mathrm{E}\,|x-a_r|\min\{|y-a_r|,\, c_r\} \tag{3.93}$$

and

$$r^2\, c_r = \mathrm{E}\,\big(|y-a_r| - c_r\big)_+ \tag{3.94}$$

For $r=\infty$, from HR (1994; Theorem 5.5.1.b) we take

$$\eta_\infty(u) = \omega_c^{\min}\big[\,\mathbf{I}(x>m) - \mathbf{I}(x<m)\big] + b\,\mathbf{I}(x=m) \tag{3.95}$$

where $m = \mathrm{med}(x)$ and

$$b\,P(x=m) = \omega_c^{\min}\big[P(x<m) - P(x>m)\big] \tag{3.96}$$

attaining minimum bias

$$\omega_c^{\min} = \theta\big(\,\mathrm{E}\,|x-m|\big)^{-1} \tag{3.97}$$

For $n\to\infty$ we have

$$\lim_{n\to\infty} \mathrm{relMSE}\,(\eta_{r_0}, r) = \mathrm{relMSE}\,(\eta_{r_0}^{1.\mathrm{loc}}, r) \tag{3.98}$$

where $\eta_{r_0}^{1.\mathrm{loc}}$ is given by (3.26) and (3.27).

3.9.2 $r/\sqrt{n}$ -Total Variation Balls

We consider the Poisson model 3.1.6 secondly with infinitesimal total variation neighborhoods 2.1.2 (b).

As minimax IC for radius $r \in [0,\infty)$, HR (1994; Theorem 5.5.7) supplies

$$\eta_r(y) = A_r\{[g_r \vee y \wedge (g_r + c_r)] - \theta\} \tag{3.99}$$

where

$$0 = \mathrm{E}(g_r - y)_+ - \mathrm{E}(y - g_r - c_r)_+ \tag{3.100}$$

$$1 = A_r\,\mathrm{E}\,y\big\{[g_r \vee y \wedge (g_r + c_r)] - \theta\big\} \tag{3.101}$$

and

$$r^2\, c_r = \mathrm{E}(g_r - y)_+ \tag{3.102}$$

For $r = \infty$, HR (1994; Theorem 5.5.7) provides

$$\eta_\infty(y) = \frac{\omega_v^{\min}}{P(y \neq \theta)} \left\{ P(y < \theta)\, \mathbf{I}(y > \theta) - P(y > \theta)\, \mathbf{I}(y < \theta) \right\} \tag{3.103}$$

with minimum bias

$$\omega_v^{\min} = \theta \left(\mathrm{E}(y - \theta)_+ \right)^{-1} \tag{3.104}$$

For $n \to \infty$ we have

$$\lim_{n \to \infty} \mathrm{relMSE}\,(\eta_{r_0}, r) = \mathrm{relMSE}\,(\eta_{r_0}^{1.\mathrm{loc}}, r) \tag{3.105}$$

where $\eta_{r_0}^{1.\mathrm{loc}}$ is given by (3.26) and (3.27).

3.10 Gamma model, Contamination

We consider the Gamma model 3.1.7 with infinitesimal contamination neighborhoods 2.1.2 (b).

The minimax IC η_r of HR (1994; Theorem 5.5.7) for $(\sigma, \alpha)' = g((\tau, \alpha)')$ and radius $r \in [0, \infty)$ is

$$\eta_r(y) = A_r(s(y) - a_r)w(y) \qquad s(y) = (y, \log(y))' \tag{3.106}$$

with

$$w(y) = \min\{1,\ b\,|A_r(s(y) - a_r)|^{-1}\} \tag{3.107}$$

where

$$0 = \mathrm{E}(s(y) - a_r)\, w(y) \tag{3.108}$$

$$D = A_r\, \mathrm{E}(s(y) - a_r)s(y)'w(y) \tag{3.109}$$

and

$$r^2 b = \mathrm{E}\left(|A_r(s(y) - a_r)| - b\right)_+ \tag{3.110}$$

Here expectation may be taken under Gamma$(1, \alpha)$. Thus the optimal robust IC η_r depends on σ respectively, τ only via $D = \partial g / \partial(\tau, \alpha)'$.

For $r = \infty$, from HR (1994; Theorem 5.5.1.b) we get

$$\eta_\infty(y) = \omega_c^{\min} \frac{A_\infty(s(y) - a_\infty)}{|A_\infty(s(y) - a_\infty)|} \tag{3.111}$$

attaining minimum bias

$$\omega_c^{\min} = \max \left\{ \frac{\mathrm{tr}\, AD'}{\mathrm{E}\left|A_\infty(s(y) - a_\infty)\right|} \,\middle|\, a_\infty \in \mathbb{R}^2, A_\infty \in \mathbb{R}^{2\times 2} \right\} \tag{3.112}$$

3.11 Regression, Average (Square) Contamination

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

We consider the k-dimensional normal regression model 3.1.8 and average conditional regression neighborhoods 2.1.2 (c) of type contamination.

The minimax IC η_r for radius $r \in [0,\infty)$ given by HR (1994; Theorem 7.4.13 and 7.5.15), and using spherical symmetry (Lemma 4.3), is

$$\eta_r(x,u) = \alpha_r x\, u \min\left\{1,\, c_r|xu|^{-1}\right\} \tag{3.113}$$

where

$$k = \alpha_r\, \mathrm{E}\,|xu| \min\left\{|xu|,\, c_r\right\}, \qquad r^2\, c_r = \mathrm{E}\left(|xu| - c_r\right)_+ \tag{3.114}$$

For $r = \infty$, HR (1994; Theorem 7.4.13.c) supplies

$$\eta_\infty\,(x,u) = \omega_{c,1}^{\min} \frac{x}{|x|} \operatorname{sign}(u) \tag{3.115}$$

which achieves minimum bias $\omega_{c,1}^{\min}$. Analogously to (3.69) we show that

$$\omega_{c,1}^{\min} = \frac{k}{\mathrm{E}\,|\Lambda|} = \sqrt{\frac{\pi}{2}}\,\frac{k}{\mathrm{E}\,|x|} \tag{3.116}$$

In Subsection 4.6, for increasing dimension $k \to \infty$ we prove that the MSE-inefficiency tends uniformly on bounded r_0-, r-intervals to the MSE-inefficiency in the one-dimensional location model 3.3.2,

$$\lim_{k\to\infty} \mathrm{relMSE}\,(\eta_{r_0}, r) = \mathrm{relMSE}\,(\eta_{r_0}^{1.\mathrm{loc}}, r) \tag{3.117}$$

where $\eta_{r_0}^{1.\mathrm{loc}}$ is given by (3.26) and (3.27).

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

We consider the k-dimensional normal regression model 3.1.8 and average square conditional regression neighborhoods 2.1.2 (c) of type contamination.

The minimax IC η_r for radius $r \in [0,\infty)$ given by HR (1994; Theorem 7.4.15, Corollary 7.5.14) and Lemma 4.3 below is

$$\eta_r(x,u) = \alpha_r x\, u \min\left\{1,\, c_r|u|^{-1}\right\} \tag{3.118}$$

where

$$k = \alpha_r\, \mathrm{E}\,|x|^2 \cdot \mathrm{E}\,|u| \min\left\{|u|,\, c_r\right\}, \qquad r^2\, c_r = \mathrm{E}\left(|u| - c_r\right)_+ \tag{3.119}$$

For $r = \infty$, HR (1994; Theorem 7.4.15.c) provides the IC of minimum bias,

$$\eta_\infty\,(x,u) = \mathcal{K}^{-1} \frac{x}{\mathrm{E}\,|u|} \operatorname{sign}(u) \tag{3.120}$$

with $\mathcal{K} = \mathrm{E}\,xx' = \gamma\,\mathbb{I}_k$ for some $\gamma \in (0,\infty)$, confer Lemma 4.2, where

$$\omega_{c,2}^{\min} = \frac{\sqrt{\operatorname{tr}\mathcal{K}^{-1}}}{\mathrm{E}\,|u|} = \sqrt{\frac{\pi\, k}{2\gamma}} \tag{3.121}$$

Comparing (3.26), (3.27) and (3.118), (3.119), we obtain the following relation to maxMSE in the one-dimensional location model 3.3.2,

$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = \frac{k^2}{\mathrm{E}\,|x|^2} \mathrm{maxMSE}(\eta_{r_0}^{1.\mathrm{loc}}, r) \tag{3.122}$$

where $\eta_{r_0}^{1.\mathrm{loc}}$ denotes the corresponding minimax IC for radius r_0; in fact, the constants in (3.26), (3.27) and (3.118), (3.119) are connected via

$$c_{r_0} = c_{r_0}^{1.\mathrm{loc}}, \qquad \alpha_{r_0} = \frac{k}{\mathrm{E}\,|x|^2} A_{r_0}^{1.\mathrm{loc}} \tag{3.123}$$

Consequentially, by relation (3.122), the MSE-inefficiencies coincide with those in one-dimensional location, independently of the regressor distribution $K(dx)$.

3.12 Regression, Average (Square) Hellinger Balls

3.12.1 Average Hellinger Neighborhoods ($* = h$, $\alpha = 1$)

We consider the k-dimensional normal regression model 3.1.8 and average conditional regression neighborhoods 2.1.2 (c) of type Hellinger.

The minimax IC η_r for radius $r \in [0, \infty)$ given by HR (1994; Theorem 7.4.19 and 7.5.7), using spherical symmetry (Lemma 4.3) and $\mathrm{E}\,u^2 = 1$, is

$$\eta_r(x,u) = \alpha_r x\, u \min\left\{1, c_r |x|^{-1}\right\} \tag{3.124}$$

where

$$k = \alpha_r\ \mathrm{E}\,|x| \min\left\{|x|,\, c_r\right\}, \qquad 8\, r^2\, c_r = \mathrm{E}\left(|x| - c_r\right)_+ \tag{3.125}$$

For $r = \infty$, HR (1994; Theorem 7.4.19.c) provides the minimum bias IC

$$\eta_\infty\,(x,u) = \frac{1}{\sqrt{8}}\, \omega_{h,1}^{\min}\, \frac{x}{|x|}\, u \tag{3.126}$$

where

$$\omega_{h,1}^{\min} = \sqrt{8}\, \frac{k}{\mathrm{E}\,|x|} \tag{3.127}$$

On rescaling r_0, r by $1/\sqrt{8}$, in the case $K = \mathcal{N}(0, \sigma^2 \mathbb{I}_k)$, a look on k-dimensional location 3.7 reveals that (3.125) agrees with (3.67) if x and u are exchanged. Consequentially, the $\mathrm{maxMSE}\,(\eta_{r_0}, r)$ and the $\mathrm{relMSE}\,(\eta_{r_0}, r)$ are the same in both models. In particular, the convergence result (3.71) is available for the present model ($* = h$, $\alpha = 1$) if $K = \mathcal{N}(0, \sigma^2 \mathbb{I}_k)$; but we prove (3.71) for model ($* = h$, $\alpha = 1$) also in case $K = \mathrm{Ufo}_k(0, m)$.

3.12.2 Average Square Hellinger Neighborhoods ($* = h$, $\alpha = 2$)

We consider the k-dimensional normal regression model 3.1.8 and average square conditional regression neighborhoods 2.1.2 (c) of type Hellinger.

According to HR (1994; p 277), the minimax IC η_r for radius $r \in [0, \infty)$ invariably is

$$\eta_r(x,u) = \mathcal{I}^{-1} \Lambda \tag{3.128}$$

Consequentially, $\mathrm{relMSE}\,(\eta_{r_0}, r) \equiv 1$ for all $r_0, r \in [0, \infty)$.

3.13 Constant Conditional Neighborhoods ($\alpha = \infty$)

3.13.1 Contamination Neighborhoods ($* = c$, $\alpha = \infty$)

We consider the k-dimensional normal regression model 3.1.8 and conditional regression neighborhoods 2.1.2 (c) of type contamination with $\alpha = \infty$.

The minimax IC η_r for radius $r \in [0, \infty)$ given by HR (1994; Theorems 7.4.11 and 7.5.10), and using spherical symmetry (Lemma 4.3), is

$$\eta_r(x,u) = \alpha_r x\, u \min\left\{1,\, c_r(x)|u|^{-1}\right\} \tag{3.129}$$

where

$$k = \alpha_r\, \mathrm{E}\,|x|^2\, \mathrm{E}_\cdot\, |u| \min\left\{|u|,\, c_r(x)\right\} \tag{3.130}$$

and

$$\mathrm{E}_\cdot\left(|u| - c_r(x)\right)_+ = \frac{r^2}{|x|}\,\mathrm{E}\,|x| c_r(x) \tag{3.131}$$

with $c_r(x) = 0$, if the RHS in (3.131) is larger than $\mathrm{E}_\cdot\,|u| = \sqrt{2/\pi}$. As for (3.131) confer Lemma 4.1. By $\mathrm{E}_\cdot$ we denote integration over $u \sim \mathcal{N}(0,1)$, with x fixed.

Concerning $r = \infty$, we show that the infimum bias is

$$\omega^{\min}_{c,\infty} = 0 \qquad \text{if } K = \mathcal{N}(0, \sigma^2 \mathbb{I}_k) \tag{3.132}$$

respectively

$$\omega^{\min}_{c,\infty} = \sqrt{\tfrac{\pi}{2}}\, k \qquad \text{if } K = \mathrm{Ufo}_k(0,m) \tag{3.133}$$

and, in both cases, cannot be attained.

In Subsection 4.6, for $k \to \infty$, we sketch an (incomplete) argument for the convergence (3.117) of the MSE-inefficiencies in the present model ($* = c$, $\alpha = \infty$) to the corresponding ones in the one-dimensional location model 3.3.2.

3.13.2 Hellinger Neighborhoods ($* = h$, $\alpha = \infty$)

We consider the k-dimensional normal regression model 3.1.8 and conditional regression neighborhoods 2.1.2 (c) of type Hellinger with $\alpha = \infty$.

The minimax IC for radius $r \in [0, \infty)$ given by HR (1994; Theorems 7.4.18 and 7.5.3), using Lemma 4.3 and $\mathrm{E}\,u^2 = 1$, is

$$\eta_r(x,u) = \alpha_r x\, u\big(1 - c_r|x|^{-1}\big)_+ \tag{3.134}$$

where

$$k = \alpha_r\, \mathrm{E}\,|x|\big(|x| - c_r\big)_+ \tag{3.135}$$

and

$$c_r = 8\, r^2\, \mathrm{E}\left(|x| - c_r\right)_+ \tag{3.136}$$

As for (3.136), confer Lemma 4.1.

Concerning $r = \infty$, we show that the infimum bias is

$$\omega^{\min}_{h,\infty} = 0 \qquad \text{if } K = \mathcal{N}(0, \sigma^2 \mathbb{I}_k) \tag{3.137}$$

respectively

$$\omega^{\min}_{h,\infty} = \sqrt{8}\, k \qquad \text{if } K = \mathrm{Ufo}_k(0,m) \tag{3.138}$$

and, in both cases, cannot be attained.

In Subsection 4.6, the convergence (3.71) of the MSE-inefficiencies to one, as the dimension increases, is proved also for this model ($* = h$, $\alpha = \infty$).

Remark 3.5 In our models—except scale—the results for $r/\sqrt{n}$-total variation neighborhoods ($* = v$) agree with the results for $2r/\sqrt{n}$-contamination neighborhoods ($* = c$); confer also Remark 3.3. ////

3.14 ARMA(1,1), Average (Square) Contamination

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

We consider the order (1,1) autoregressive-moving average model 3.1.10 with standard normal innovations and average conditional regression neighborhoods 2.1.2 (c) of type contamination.

The minimax IC η_r for radius $r \in [0, \infty)$ given by HR (1994; Theorems 7.4.13 and 7.5.15), and using the elliptical symmetry of H is

$$\eta_r(H, u) = V' \psi_r(Z, u) \tag{3.139}$$

with

$$\psi_r(Z, u) = A_r Z \, u \min\left\{1, \, b_r |A_r Z \, u|^{-1}\right\} \qquad A_r = \operatorname{diag}(a_{r,1}, a_{r,2}) \tag{3.140}$$

where

$$1 = a_{r,j} \, \mathrm{E} \, Z_j^2 \, \mathrm{E}_\bullet \, u^2 \min\left\{1, \, b_r |A_r Z \, u|^{-1}\right\} \qquad j = 1, 2 \tag{3.141}$$

and

$$r^2 \, b_r = \mathrm{E}\left([a_{r,1}^2 Z_1^2 + a_{r,2}^2 Z_2^2]^{1/2} \, |u| - b_r \right)_+ \tag{3.142}$$

For $r = \infty$, HR (1994; Theorems 7.4.13.c), and elliptical symmetry of H supply

$$\eta_\infty \, (H, u) = V' \psi_\infty \, (Z, u) \tag{3.143}$$

with

$$\psi_\infty \, (Z, u) = \omega_{c,1}^{\min} \begin{pmatrix} Z_1 & 0 \\ 0 & a_\infty Z_2 \end{pmatrix} [Z_1^2 + a_\infty^2 Z_2^2]^{-1/2} \tag{3.144}$$

which achieves minimum bias

$$\omega_{c,1}^{\min} = \sqrt{\tfrac{\pi}{2}} \, \max\left\{ \frac{1 + a_\infty}{\mathrm{E}[Z_1^2 + a_\infty^2 Z_2^2]^{1/2}} \right\} \tag{3.145}$$

For $\phi = -0.7, \xi = 0.35$ we get $\omega_{c,1}^{\min} \approx 2.054$, where $a_\infty \approx 3.569$.

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

We consider the order (1,1) autoregressive-moving average model 3.1.10 with standard normal innovations and average square conditional regression neighborhoods 2.1.2 (c) of type contamination.

The minimax IC η_r for radius $r \in [0,\infty)$ given by HR (1994; Theorem 7.4.15, Corollary 7.5.14), is

$$\eta_r(H,u) = A_r H\, u \min\left\{1,\, c_r|u|^{-1}\right\} \tag{3.146}$$

where

$$1 = \rho_r\, \mathrm{E}\,|u| \min\left\{|u|,\, c_r\right\} \qquad A_r = \rho_r \mathcal{K}_\theta^{-1} \tag{3.147}$$

and

$$r^2\, c_r = \mathrm{E}\left(|u| - c_r\right)_+ \tag{3.148}$$

For $r = \infty$, HR (1994; Theorem 7.4.15.c) provides

$$\eta_\infty\,(H,u) = \sqrt{\tfrac{\pi}{2}}\, \mathcal{K}_\theta^{-1} H \,\mathrm{sign}\,(u) \tag{3.149}$$

which attains minimum bias

$$\omega_{c,2}^{\min} = \sqrt{\tfrac{\pi}{2}\,\mathrm{tr}\,\mathcal{K}_\theta^{-1}} = \sqrt{\tfrac{\pi}{2}}\,\frac{1-\phi\xi}{|\phi-\xi|}\,\sqrt{2-\phi^2-\xi^2} \tag{3.150}$$

For $\phi = -0.7, \xi = 0.35$ we get $\omega_{c,2}^{\min} \approx 1.750$.

Comparing (3.26), (3.27) and (3.146)-(3.148), we obtain the following relation to maxMSE in the one-dimensional location model 3.3.2,

$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = \mathrm{tr}\,\mathcal{K}_\theta^{-1} \mathrm{maxMSE}(\eta_{r_0}^{1.\mathrm{loc}}, r) \tag{3.151}$$

where $\eta_{r_0}^{1.\mathrm{loc}}$ denotes the corresponding minimax IC for radius r_0; in fact, the constants in (3.26), (3.27) and (3.146), (3.147) are connected via

$$c_{r_0} = c_{r_0}^{1.\mathrm{loc}}, \qquad \rho_{r_0} = A_{r_0}^{1.\mathrm{loc}} \tag{3.152}$$

Consequentially, by relation (3.151), the MSE-inefficiencies coincide with those in one-dimensional location, independently of ϕ and ξ.

3.15 ARCH(1), Average (Square) Contamination

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

We consider the order one autoregressive conditional heteroscedastic model 3.1.11 with standard normal innovations and average conditional regression neighborhoods 2.1.2 (c) of type contamination.

The minimax IC η_r for radius $r \in [0,\infty)$ given by HR (1994; Theorems 7.4.13 and 7.5.15) is

$$\eta_r(H,u) = A_r H\big(u^2 - \alpha(H)^2\big) \min\left\{1, \frac{c_r}{H\,|u^2 - \alpha(H)^2|}\right\} \tag{3.153}$$

where

$$0 = \mathrm{E}_\bullet(u^2 - \alpha(H)^2) \min\left\{1, \frac{c_r}{H\,|u^2 - \alpha(H)^2|}\right\} \tag{3.154}$$

$$1 = A_r\, \mathrm{E}\, H^2\, \mathrm{E}_\bullet(u^2 - \alpha(H)^2)^2 \min\left\{1, \frac{c_r}{H\,|u^2 - \alpha(H)^2|}\right\} \tag{3.155}$$

and

$$r^2\, c_r = \mathrm{E}\left(H\,|u^2 - \alpha(H)^2| - c_r\right)_+ \tag{3.156}$$

For $r = \infty$, HR (1994; Theorems 7.4.13.c) supplies

$$\eta_\infty\,(H,u) = \omega_{c,1}^{\min}\,\mathrm{sign}\,(|u| - \alpha_\infty) \tag{3.157}$$

with centering constant $\alpha_\infty = \Phi^{-1}(3/4) \approx 0.674$, which achieves minimum bias

$$\omega_{c,1}^{\min} = \big(\mathrm{E}\,|u^2 - \alpha_\infty^2|\,\mathrm{E}\,H\big)^{-1} = \big(4\alpha_\infty\varphi(\alpha_\infty)\,\mathrm{E}\,H\big)^{-1} \approx 1.166/\,\mathrm{E}\,H \tag{3.158}$$

For $\theta = 1$ we get $\omega_{c,1}^{\min} \approx 4.637$.

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

We consider the order one autoregressive conditional heteroscedastic model 3.1.11 with standard normal innovations and average square conditional regression neighborhoods 2.1.2 (c) of type contamination.

The minimax IC η_r for radius $r \in [0,\infty)$ given by HR (1994; Theorem 7.4.15, Corollary 7.5.14) is

$$\eta_r(H,u) = A_r H(u^2 - \alpha_r^2)\min\big\{1,\, c_r|u^2 - \alpha_r^2|^{-1}\big\} \tag{3.159}$$

where

$$0 = \mathrm{E}(u^2 - \alpha_r^2)\min\big\{1,\, c_r|u^2 - \alpha_r^2|^{-1}\big\} \tag{3.160}$$

$$1 = \rho_r\,\mathrm{E}(u^2 - \alpha_r)^2\min\big\{1,\, c_r|u^2 - \alpha_r^2|^{-1}\big\} \qquad A_r = \rho_r\mathcal{K}_\theta^{-1} \tag{3.161}$$

and

$$r^2\,c_r = \mathrm{E}\,\big(|u^2 - \alpha_r^2| - c_r\big)_+ \tag{3.162}$$

For $r = \infty$, HR (1994; Theorem 7.4.15.c) provides the IC of minimum bias,

$$\eta_\infty\,(H,u) = \frac{\mathcal{K}_\theta^{-1}H}{\mathrm{E}\,|u^2 - \alpha_\infty|}\,\mathrm{sign}\,(|u| - \alpha_\infty) \tag{3.163}$$

with centering constant $\alpha_\infty = \Phi^{-1}(3/4) \approx 0.674$, which achieves minimum bias

$$\omega_{c,2}^{\min} = \mathcal{K}_\theta^{-1/2}\big(\mathrm{E}\,|u^2 - \alpha_\infty^2|\big)^{-1} \approx 1.166\,\mathcal{K}_\theta^{-1/2} \tag{3.164}$$

For $\theta = 1$ we get $\omega_{c,2}^{\min} \approx 3.898$.

Comparing (3.33)-(3.36) and (3.159)-(3.162), we obtain the following relation to maxMSE in the one-dimensional scale model 3.4.1,

$$\mathrm{maxMSE}\,(\eta_{r_0}, r) = \mathcal{K}_\theta^{-1}\mathrm{maxMSE}(\eta_{r_0}^{\mathrm{sc}}, r) \tag{3.165}$$

where $\eta_{r_0}^{\mathrm{sc}}$ denotes the corresponding minimax IC for radius r_0; in fact, the constants in (3.33)-(3.36) and (3.159)-(3.162) are connected via

$$c_{r_0} = c_{r_0}^{\mathrm{sc}}, \qquad \alpha_{r_0} = \alpha_{r_0}^{\mathrm{sc}}, \qquad \rho_{r_0} = A_{r_0}^{\mathrm{sc}} \tag{3.166}$$

Consequentially, by relation (3.165), the MSE-inefficiencies coincide with those in one-dimensional scale, independently of θ.

3.15.3 Average Total Variation Neighborhoods ($* = v$, $\alpha = 1$)

We consider the order one autoregressive conditional heteroscedastic model 3.1.11 with standard normal innovations and average conditional regression neighborhoods 2.1.2 (c) of type total variation.

As minimax IC for radius $r \in [0, \infty)$, HR (1994; Theorem 7.4.17) supplies

$$\eta_r(H, u) = g_r(H) \vee A_r H(u^2 - 1) \wedge (g_r(H) + c_r) \tag{3.167}$$

where

$$\mathrm{E}_\bullet(g_r(H) - A_r H(u^2 - 1))_+ = \mathrm{E}_\bullet(A_r H(u^2 - 1) - g_r(H) - c_r)_+ \tag{3.168}$$

$$1 = \mathrm{E}\, H\, \mathrm{E}_\bullet\, u^2[g_r(H) \vee A_r H(u^2 - 1) \wedge (g_r(H) + c_r)] \tag{3.169}$$

and

$$r^2\, c_r = \mathrm{E}_\bullet(g_r(H) - A_r H(u^2 - 1))_+ \tag{3.170}$$

For $r = \infty$, HR (1994; Theorem 5.5.5.b) provides

$$\eta_\infty(u) = \omega_{v,1}^{\min}\big\{P(|u| < 1)\, \mathbf{I}(|u| > 1) - P(|u| > 1)\, \mathbf{I}(|u| < 1)\big\} \tag{3.171}$$

with minimum bias

$$\omega_{v,1}^{\min} = (\mathrm{E}\, H\, \mathrm{E}(u^2 - 1)_+)^{-1} = \sqrt{\tfrac{\pi}{2}\, e}\,\big/\, \mathrm{E}\, H \tag{3.172}$$

For $\theta = 1$ we get $\omega_{v,1}^{\min} \approx 8.220$.

3.15.4 Average Square Total Variation Neighborhoods ($* = v$, $\alpha = 2$)

We consider the order one autoregressive conditional heteroscedastic model 3.1.11 with standard normal innovations and average square conditional regression neighborhoods 2.1.2 (c) of type total variation.

As minimax IC for radius $r \in [0, \infty)$, Kohl (2005; Section 9.3.2.3) supplies

$$\eta_r(u) = A_r H\{[g_r \vee u^2 \wedge (g_r + c_r)] - 1\} \tag{3.173}$$

where

$$0 = \mathrm{E}(g_r - u^2)_+ - \mathrm{E}(u^2 - g_r - c_r)_+ \tag{3.174}$$

$$1 = \rho_r\, \mathrm{E}\, u^2\big\{[g_r \vee u^2 \wedge (g_r + c_r)] - 1\big\} \qquad A_r = \rho_r \mathcal{K}_\theta^{-1} \tag{3.175}$$

and

$$r^2\, c_r = \mathrm{E}(g_r - u^2)_+ \tag{3.176}$$

For $r = \infty$, one can show

$$\eta_\infty(u) = \frac{\mathcal{K}_\theta^{-1} H}{\mathrm{E}(u^2 - 1)_+}\big\{P(|u| < 1)\, \mathbf{I}(|u| > 1) - P(|u| > 1)\, \mathbf{I}(|u| < 1)\big\} \tag{3.177}$$

which achieves minimum bias

$$\omega_{v,2}^{\min} = \mathcal{K}_\theta^{-1/2}(\mathrm{E}(u^2 - 1)_+)^{-1} = \sqrt{\tfrac{\pi}{2}\, e\, \mathcal{K}_\theta^{-1}} \tag{3.178}$$

For $\theta = 1$ we get $\omega_{v,2}^{\min} \approx 6.907$.

Comparing (3.39)-(3.42) and (3.173)-(3.176), we obtain the following relation to maxMSE in the one-dimensional scale model 3.4.2,

$$\text{maxMSE}\,(\eta_{r_0}, r) = \mathcal{K}_\theta^{-1}\text{maxMSE}(\eta_{r_0}^{\text{sc}}, r) \tag{3.179}$$

where $\eta_{r_0}^{\text{sc}}$ denotes the corresponding minimax IC for radius r_0; in fact, the constants in (3.39)-(3.42) and (3.173)-(3.176) are connected via

$$c_{r_0} = c_{r_0}^{\text{sc}}, \qquad \alpha_{r_0} = \alpha_{r_0}^{\text{sc}}, \qquad \rho_{r_0} = A_{r_0}^{\text{sc}} \tag{3.180}$$

Consequentially, by relation (3.179), the MSE-inefficiencies coincide with those in one-dimensional scale, independently of θ.

4 Lemmas and Proofs

4.1 Optimization

With the help of the following lemma, we can derive the solutions to the original MSE problems (with bias squared) 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 4.1 *Given a real vector space X, a convex subset A of X, consider three convex functions $f\colon A \to \mathbb{R}$, $g\colon A \to [0,\infty)$, and $\gamma\colon [0,\infty) \to [0,\infty)$; 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*

$$\beta_1 = \beta_0\,\gamma'(g_0) \tag{4.1}$$

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

Proof Employ the convex combinations $z_s = (1-s)z_0 + sz_1$, $0 \le s \le 1$, 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

$$\partial L_0 = \partial f + \beta_0\,\gamma'(g_0)\,\partial g = \partial L_1 \tag{4.2}$$

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

4.2 One-Dimensional Location

Proof of (3.25) To ψ_s for $s \in [0,1)$, integration by parts applies so that

$$\int \psi_s'(u)\,\Phi(du) = \int u\psi_s(u)\,\Phi(du) \tag{4.3}$$

Therefore we can rewrite (3.22) as

$$\text{maxVar}\,(\psi_s, s) = \frac{(1-s)\int m_s^{-2}\psi_s^2(u)\,\Phi(du) + s}{(1-s)^2\big[\int u\,m_s^{-1}\psi_s(u)\,\Phi(du)\big]^2} \tag{4.4}$$

Consequentially, the Var-inefficiency of the median is

$$\mathrm{relVar}\,(\psi_1, s) = \frac{\pi\left[\int u\, m_s^{-1}\psi_s(u)\,\Phi(du)\right]^2}{2\left[(1-s)\int m_s^{-2}\psi_s^2(u)\,\Phi(du) + s\right]} \tag{4.5}$$

As $\psi_s/m_s \to \psi_1 = \mathrm{sign}$ pointwise for $s \to 1$ and $|\psi_s/m_s| \le 1$, it follows that

$$\lim_{s\to 1} \mathrm{relVar}\,(\psi_1, s) = \frac{\pi}{2}\big(\,\mathrm{E}\,|u|\,\big)^2 = 1 \tag{4.6}$$

by the dominated convergence theorem. ////

Proof of (3.30)–(3.32) By (3.21) and (3.27), $c_r = m_s$, if r and s are related by (3.31): $r^2 = s/(1-s)$. Using (3.27) we can rewrite (2.12) as

$$\begin{aligned}
\mathrm{maxMSE}\,(\eta_{r_0}, r) &= \frac{\mathrm{E}\min\{u^2,\, c_{r_0}^2\} + r^2 c_{r_0}^2}{[\mathrm{E}\,|u|\min\{|u|,\, c_r\}]^2} \\
&= \frac{\mathrm{E}\min\{u^2,\, m_{s_0}^2\} + s\, m_{s_0}^2/(1-s)}{[\mathrm{E}\,|u|\min\{|u|,\, m_s\}]^2} \\
&= \frac{(1-s)\,\mathrm{E}\,\psi_{s_0}^2 + s m_{s_0}^2}{(1-s)[\mathrm{E}\,\psi'_{s_0}]^2} \qquad \text{by (4.3)} \\
&= (1-s)\mathrm{maxVar}\,(\psi_{s_0}, s)
\end{aligned}$$

which proves (3.30) implying (3.32). ////

4.3 Invariance Under Rescaling

As mentioned in Subsubsection 3.1, the inefficiency in models 3.1 (a)–(d) is invariant under rescaling of the errors u_i and the regressors x_i, respectively. We prove this invariance for k-dimensional regression and average conditional contamination neighborhoods 2.1.2 (c) ($* = c$, $\alpha = 1$), even allowing general error distribution $F(du)$ and regressor distribution $K(dx)$ as in HR (1994; Theorem 7.4.13). The proofs for the other models considered here 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

$$\eta_r(x,u) = A_r x\,[\Lambda_f(u) - \vartheta_r(x)]\min\left\{1, \frac{b_r}{|A_r x[\Lambda_f(u) - \vartheta_r(x)]|}\right\} \tag{4.7}$$

$$0 = \mathrm{E}_{\cdot}[\Lambda_f(u) - \vartheta_r(x)]\min\left\{1, \frac{b_r}{|A_r x[\Lambda_f(u) - \vartheta_r(x)]|}\right\} \tag{4.8}$$

$$\mathbb{I}_k = A_r\,\mathrm{E}\,xx'\Lambda_f^2(u)\min\left\{1, \frac{b_r}{|A_r x[\Lambda_f(u) - \vartheta_r(x)]|}\right\} \tag{4.9}$$

$$r^2\, b_r = \mathrm{E}\left(|A_r x[\Lambda_f(u) - \vartheta_r(x)] - b|\right)_+ \tag{4.10}$$

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

$$\tilde{\eta}_r(z, v) = \tilde{A}_r z \, [\Lambda_{\tilde{f}}(v) - \tilde{\vartheta}_r(z)] \min \Big\{ 1, \frac{\tilde{b}_r}{\big|\tilde{A}_r z [\Lambda_{\tilde{f}}(v) - \tilde{\vartheta}_r(z)]\big|} \Big\} \tag{4.11}$$

with

$$\tilde{A}_r = \tfrac{\sigma^2}{\tau^2} A_r \,, \qquad \tilde{b}_r = \tfrac{\sigma}{\tau} b_r \,, \qquad \tilde{\vartheta}_r(z) = \tfrac{1}{\sigma} \vartheta_r\big(\tfrac{z}{\tau}\big) \tag{4.12}$$

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

Then it is easy to verify conditions (4.8)–(4.10) for $\tilde{\eta}_r$ in the rescaled model, so $\tilde{\eta}_r$ is indeed the optimum IC there. Using the relations (4.12) we obtain

$$\operatorname{maxMSE}(\tilde{\eta}_{r_0}, r) = \frac{\sigma^2}{\tau^2} \operatorname{maxMSE}(\eta_{r_0}, r) \tag{4.13}$$

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

4.4 Spherical Symmetry

We consider models whose scores function Λ at P_θ is spherically symmetric; that is,

$$\mathcal{L}(G\Lambda) = \mathcal{L}(\Lambda) \tag{4.14}$$

for all orthogonal matrices $G \in \mathbb{R}^{k \times k}$. Fisher information of such models satisfies

$$G \mathcal{I} G' = \mathrm{E}(G\Lambda)(G\Lambda)' = \mathrm{E}\, \Lambda\Lambda' = \mathcal{I} \tag{4.15}$$

for all orthogonal $G \in \mathbb{R}^{k \times k}$, hence, by the following lemma, is a multiple of the identity: $\mathcal{I} = \gamma \mathbb{I}_k$; $\gamma \in [0, \infty)$ since $\mathcal{I}$ is positive semidefinite, and $\gamma > 0$ if $\mathcal{I}$ has full rank (which is the case in our models).

Lemma 4.2 *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 $A = GAG' = \operatorname{diag}(\alpha_1, \ldots, \alpha_k)$; so A is diagonal. Now consider a permutation matrix $G \in \mathbb{R}^{k \times k}$ (any matrix with a single one and otherwise zero entries). Such G being orthogonal, again $A = GAG'$; so necessarily $\alpha_1 = \ldots = \alpha_k$. ////

The second application of Lemma 4.2 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 4.3 *Under assumption* (4.14) *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-dimensional 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.

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

$$r^2\, c_r = \mathrm{E}\,(|x\,u| - c_r) \tag{4.16}$$

$$k = \alpha_r\, \mathrm{E}\,|x\,u| \min\{|x\,u|, c_r\} \tag{4.17}$$

and put

$$\eta_r(x,u) = \alpha_r x\, u \min\{1, c_r |x\,u|^{-1}\} \tag{4.18}$$

As for all orthogonal $G \in \mathbb{R}^{k\times k}$: $G\,\mathrm{E}\,\eta_r\Lambda' G' = \mathrm{E}\,\eta_r\Lambda'$ by spherical symmetry of K, Lemma 4.2 tells us that $\mathrm{E}\,\eta_r\Lambda' = \beta\,\mathbb{I}_k$. Passing to the trace, (4.17) yields that $\beta = 1$. Because of symmetry of the error distribution, $\mathrm{E}_{\cdot}\,\eta_r = 0$ 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.

$(* = c, \alpha = 2)$ Since the median of Λ_f is unique, the minimax IC is given by HR (1994; equation (134) in Theorem 7.4.15.b), with $D = \mathbb{I}_k$, where $\vartheta = 0$, by symmetry of the error distribution, and $\mathcal{K} = \mathcal{I} = \gamma\,\mathbb{I}_k$ for some $\gamma \in (0,\infty)$ (Lemma 4.2 and $\mathcal{I}_f^{\mathrm{loc}} = 1$). This gives (4.16)–(4.18).

$(* = c, \alpha = \infty)$ An argument as in the case $(* = c, \alpha = 1)$ and in addition using Lemma 4.1 shows that η_r of form (3.129)–(3.131) is an IC of the form of HR (1994; Theorem 7.4.11.b) and satisfies condition (4.1) of Lemma 4.1 above. Thus, η_r is the (unique) minimax IC. ////

Minimum Bias

Proof of (3.69) and (3.116) According to HR (1994; Theorems 5.5.1.b and 7.4.13.c), minimum bias $\omega_c^{\min}$ in model 3.7 (k-dimensional location) and $\omega_{c,1}^{\min}$ in model 3.11.1 (k-dimensional normal regression) are, with $D = \mathbb{I}_k$, given by

$$\omega_c^{\min} = \max\left\{ \frac{\mathrm{tr}\,A}{\mathrm{E}\,|A\Lambda - a|} \;\middle|\; a \in \mathbb{R}^k,\, A \in \mathbb{R}^{k\times k} \setminus \{0\} \right\} \tag{4.19}$$

respectively

$$\omega_{c,1}^{\min} = \max\left\{ \frac{\mathrm{tr}\,A}{\mathrm{E}\,|Ax|\,\mathrm{E}\,|\Lambda_f - m|} \;\middle|\; A \in \mathbb{R}^{k\times k} \setminus \{0\} \right\} \tag{4.20}$$

In our case, the median m of $\Lambda_f(u) = u$ under $F = \mathcal{N}(0,1)$ is zero. Also in (4.19), we may put $a = 0$. Indeed, by triangle inequality and (spherical) symmetry of $\mathcal{L}(\Lambda)$, the zero centering vector $0 = \frac{1}{2}a + \frac{1}{2}(-a)$ would decrease the denominator $\mathrm{E}\,|A\Lambda - a| = \mathrm{E}\,|A\Lambda + a|$.

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

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

$$\mathrm{E}\,|A\Lambda| = \mathrm{E}\,|U'AV\Lambda| = \mathrm{E}\,|V'A'U\Lambda| = \mathrm{E}\,|A'\Lambda| \tag{4.21}$$

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

$$\mathrm{tr}\,GAG' = \mathrm{tr}\,A, \qquad \mathrm{E}\,|GAG'\Lambda| = \mathrm{E}\,|A\Lambda| \tag{4.22}$$

for any orthogonal matrix G, and especially for G obtained from the spectral decomposition of A, we may further suppose A diagonal, and then with all diagonal elements nonnegative.

To complete the proof, we show that $\frac{1}{k}\mathbb{I}_k$ minimizes $\mathrm{E}\,|A\Lambda|$ among all such diagonal matrices of trace 1. Consider the Lagrangian $L\colon [0,\infty)^k \to \mathbb{R}$,

$$L(a) = \mathrm{E}\,|A\Lambda| - \lambda \operatorname{tr} A \tag{4.23}$$

where $a = (\alpha_1, \ldots, \alpha_k)'$ and $A = \operatorname{diag}(a')$. The multiplier λ is chosen as

$$\lambda = \mathrm{E}\left\{|\Lambda|^{-1}\Lambda_i^2\right\} \tag{4.24}$$

which, by spherical symmetry, is the same for all coordinates $i = 1, \ldots, k$.

The function L is convex on $[0,\infty)^k$. Applying the mean value theorem and the bound

$$\frac{(\alpha_i + \tau h)\Lambda_i^2}{|A_\tau \Lambda|} \le \frac{2\,\alpha_i \Lambda_i^2}{\frac{1}{2}|\alpha_i \Lambda_i|} = 4|\Lambda_i| \in L_1(P) \tag{4.25}$$

with some $\tau \in (0,1)$, some intermediate A_τ, and sufficiently small increment h, the dominated convergence theorem applies. Thus we obtain the partials

$$\frac{\partial L}{\partial \alpha_i} = \mathrm{E}\left\{\frac{\alpha_i \Lambda_i^2}{|A\Lambda|}\right\} - \lambda \tag{4.26}$$

which vanish at $\alpha_1 = \ldots = \alpha_k = \frac{1}{k}$. ////

4.5 Constant Conditional Neighborhoods ($\alpha = \infty$)

Proof of (3.132), (3.133) and (3.137), (3.138) The solutions are given in HR (1994; Theorems 7.4.11.c and 7.4.18.c), we only have to determine

$$\sigma_A = \inf_{K(dx)} |Ax|^{-1} = \frac{1}{\sup_{K(dx)} |Ax|} \tag{4.27}$$

for $K = \mathcal{N}(0, \mathbb{I}_k)$, respectively $K = \mathrm{Ufo}_k(0,1)$.

In the normal case, we have $\sup_{K(dx)} |Ax| = \infty$, thus $\sigma_A = 0$ and consequentially both $\omega_{c,\infty}^{\min}$ and $\omega_{h,\infty}^{\min}$ are zero. Then, since

$$\omega_{c,\infty}(\eta) = \mathrm{E} \sup_{\cdot} |\eta|\,, \qquad \omega_{h,\infty}(\eta) = \sqrt{8}\,\mathrm{E}\left(\mathrm{E}_{\cdot}\,|\eta|^2\right)^{1/2} \tag{4.28}$$

an IC η achieving zero bias would have to vanish a.e.; thus, the infimum bias cannot be attained.

In the uniform case, we obtain that

$$\sup_{K(dx)} |Ax|^2 = \sup_{|x|\le 1} |Ax|^2 = \|A\|_{\mathrm{op}} \tag{4.29}$$

Hence we have to find the minimum of $\|A\|_{\mathrm{op}}$ under the side condition $\operatorname{tr} A = 1$. Applying the triangle inequality to $\frac{1}{2}(A + A')$, A may be assumed symmetric. Then, since

$$\|A\|_{\mathrm{op}}^2 = \sup_{|x|\le 1} |Ax| = \sup_{|x|\le 1} |G'AGx| \tag{4.30}$$

for any orthogonal matrix G, it suffices to consider A diagonal (spectral decomposition). Thus, $\|A\|_{\mathrm{op}}^2 = \max_{i=1,\ldots,k} \alpha_i^2$, and consequentially $A_{\min} = \frac{1}{k}\,\mathbb{I}_k$ and $\|A_{\min}\|_{\mathrm{op}} = \frac{1}{k}$, which yields $\sigma_A = k$. According to HR (1994; Theorems 7.4.11.c and 7.4.18.c), an IC $\bar\eta$ achieving the minimum bias would necessarily have to be of form (75), respectively (220), there; in particular, $\bar\eta$ could be nonzero only for $1 = \sigma_A|Ax| = |x|$. This however, is a set of measure zero in the present cases. Therefore, the infimum bias cannot be attained. ////

4.6 Increasing Dimension $k \to \infty$

Proof of (3.70) We have $\Lambda \sim \mathcal{N}(0, \mathbb{I}_k)$, so $\frac{1}{k}|\Lambda|^2 = \frac{1}{k}\sum_{i=1}^k \Lambda_i^2 \to \mathrm{E}\,\Lambda_1^2 = 1$ a.e., hence also $|\Lambda|/\sqrt{k} \to 1$ a.e., as $k \to \infty$ (SLLN). $\left(\frac{1}{k}|\Lambda|^2\right)$ is uniformly integrable (Vitali, $\mathrm{E}\,|\Lambda|^2 = k$). Because $|\Lambda|/\sqrt{k} \le 1 + \frac{1}{k}|\Lambda|^2$, also $\left(|\Lambda|/\sqrt{k}\right)$ is uniformly integrable. Consequentially, $\mathrm{E}\,|\Lambda| \sim \sqrt{k}$ and $\mathrm{E}\,|\eta_\infty|^2 = \left(k/\,\mathrm{E}\,|\Lambda|\right)^2 \approx k$. ////

Proof of (3.71) We first give the proof for the k-dimensional location model 3.7.

For $\Lambda(u) = u \sim \mathcal{N}(0, \mathbb{I}_k)$, both $\frac{1}{k}|u|^2$ and $|u|/\sqrt{k}$ tend to 1 in L_1, as shown. Putting $\tilde c_{k,r} = c_{k,r}/\sqrt{k}$, the second equation of (3.67) reads

$$r^2\,\tilde c_{k,r} = \mathrm{E}\left(|u|/\sqrt{k} - \tilde c_{k,r}\right)_+ \tag{4.31}$$

In the case $r = 0$, we have $c_{k,0} = \infty$. Assume that $r > 0$, and suppose that $\tilde c_{k,r} \to \gamma \in [0,\infty]$ along some subsequence. Since the RHS in (4.31) is bounded, necessarily $\gamma < \infty$. Then the noted L_1-convergence implies that

$$r^2\,\gamma = \mathrm{E}(1-\gamma)_+ = (1-\gamma)_+ \tag{4.32}$$

from which it follows that $r^2\gamma = 1-\gamma$. Hence $(1+r^2)^{-1}$ is the unique accumulation point of the sequence $(\tilde c_{k,r})$, which therefore converges,

$$\lim_{k\to\infty} \tilde c_{k,r} = \frac{1}{1+r^2} \tag{4.33}$$

The first equation of (3.67) reads

$$\alpha_{k,r}^{-1} = \mathrm{E}\,k^{-1}|u|^2 \min\left\{1,\ \frac{\tilde c_{k,r}}{|u|/\sqrt{k}}\right\} \tag{4.34}$$

In the case $r = 0$ we have $\alpha_{k,0} = 1$. Now let $r > 0$. Obviously, the integrands in (4.34) converge to $(1+r^2)^{-1}$ a.e. and are uniformly integrable (being dominated by $\frac{1}{k}|u|^2$). Thus, and consistently with $\alpha_{k,0} = 1$ in case $r = 0$,

$$\lim_{k\to\infty} \alpha_{k,r} = 1 + r^2 \tag{4.35}$$

The arguments leading to (4.33) and (4.35) obtain if the fixed $r \in (0,\infty)$ is replaced by any sequence r_k with limit $r \in (0,\infty)$. In addition, we can argue in a similar way in case $r_k \downarrow r = 0$ to obtain that

$$\liminf_{k\to\infty} \tilde c_{k,r_k} \ge 1 = \lim_{k\to\infty} \alpha_{k,r_k} \tag{4.36}$$

Therefore, (4.33) and (4.35) hold uniformly on bounded r-intervals. Now (4.33) and (4.35) entail convergence of the risk (with the previous r replaced by r_0),

$$k^{-1}\,\mathrm{maxMSE}\,(\eta_{r_0}, r) = \alpha_{k,r_0}^2\,\mathrm{E}\min\left\{k^{-1}|u|^2,\,\tilde{c}_{k,r_0}^2\right\} + r^2\alpha_{k,r_0}^2\tilde{c}_{k,r_0}^2 \longrightarrow 1 + r^2 \tag{4.37}$$

Consequentially, and in addition using (3.70), we get

$$\frac{\mathrm{maxMSE}\,(\eta_{r_0}, r)}{\mathrm{maxMSE}\,(\eta_\infty, r)} = \frac{\mathrm{E}\,|\Lambda|^2}{k}\;\frac{\mathrm{maxMSE}\,(\eta_{r_0}, r)}{k(1+r^2)} \longrightarrow 1 \tag{4.38}$$

And this convergence is uniform on bounded r_0- and r-intervals.

By the coincidence mentioned in the k-dimensional regression model 3.12.1 for $* = h$ and $\alpha = 1$, the convergence (3.71) automatically holds for this model, too, with r_0 and r multiplied by $\sqrt{8}$, if the regressor distribution is $K = \mathcal{N}(0, \mathbb{I}_k)$.

In the second part of the proof we shall show (3.71) for the k-dimensional regression model 3.12.1 ($* = h$, $\alpha = 1$) with $K = \mathrm{Ufo}_k(0,1)$.

In this case, we have $\mathrm{E}\,|x| = k(k+1)^{-1}$ and $\mathrm{E}\,|x|^2 = k(k+2)^{-1}$ which implies that $\mathrm{Var}\,(x) \to 0$ as $k \to \infty$; consequentially, $|x| \to 1$ in L_2 and L_1. The second part of (3.125) reads

$$8r^2\,c_{k,r} = \mathrm{E}\left(|x| - c_{k,r}\right)_+ \tag{4.39}$$

Suppose that $c_{k,r} \to \gamma \in [0,\infty]$ along some subsequence; since $|x| \le 1$ a.e., necessarily $\gamma \le 1$. Then the noted L_1-convergence implies that

$$8r^2\,\gamma = \mathrm{E}(1-\gamma)_+ = 1 - \gamma \tag{4.40}$$

Hence $(1+8r^2)^{-1}$ is the unique accumulation point, therefore,

$$\lim_{k\to\infty} c_{k,r} = \frac{1}{1+8r^2} \tag{4.41}$$

Plugged into the first equation of (3.125), this yields

$$\frac{k}{\alpha_{k,r}} = \mathrm{E}\,|x|\min\{|x|, c_{k,r}\} \longrightarrow \frac{1}{1+8r^2} \tag{4.42}$$

The arguments leading to (4.41) and (4.42) obtain if we replace r by a bounded sequence r_k. Thus (4.41) and (4.42) hold true uniformly on bounded r-intervals. The convergences (4.41) and (4.42) now entail convergence of the risk (with the previous r replaced by r_0),

$$k^{-2}\,\mathrm{maxMSE}\,(\eta_{r_0}, r) = k^{-2}\alpha_{k,r_0}^2\,\mathrm{E}\min\{|x|^2, c_{k,r_0}^2\} + 8r^2k^{-2}\alpha_{k,r_0}^2 c_{k,r_0}^2 \longrightarrow 1 + 8r^2 \tag{4.43}$$

Consequentially,

$$\frac{\mathrm{maxMSE}\,(\eta_{r_0}, r)}{\mathrm{maxMSE}\,(\eta_\infty, r)} = \mathrm{E}\,|x|^2\;\frac{\mathrm{maxMSE}\,(\eta_{r_0}, r)}{k^2(1+8r^2)} \longrightarrow 1 \tag{4.44}$$

And this convergence is uniform on bounded r_0-, r-intervals.

In the third part of the proof, we shall show (3.71) for the k-dimensional regression model 3.13.2 ($* = h$, $\alpha = \infty$).

In case $K = \mathcal{N}(0, \mathbb{I}_k)$, the L_1-convergence $|x|/\sqrt{k} \to 1$ inserted in equations (3.136) and (3.135) by previous arguments imply that

$$\lim_{k\to\infty} \frac{c_{k,r}}{\sqrt{k}} = \frac{8r^2}{1+8r^2}, \qquad \lim_{k\to\infty} \alpha_{k,r} = 1 + 8r^2 \tag{4.45}$$

Consequentially,

$$\lim_{k\to\infty} \frac{1}{k} \operatorname{maxMSE}(\eta_{r_0}, r) = 1 + 8r^2 \tag{4.46}$$

In case $K = \mathrm{Ufo}_k(0,1)$, the L_1-convergence of $|x| \to 1$ inserted in equations (3.136) and (3.135) similarly imply that

$$\lim_{k\to\infty} c_{k,r} = \frac{8r^2}{1+8r^2}, \qquad \lim_{k\to\infty} \frac{\alpha_{k,r}}{k} = 1 + 8r^2 \tag{4.47}$$

Consequentially,

$$\lim_{k\to\infty} k^{-2} \operatorname{maxMSE}(\eta_{r_0}, r) = 1 + 8r^2 \tag{4.48}$$

Both convergences (4.46) and (4.48) hold uniformly on bounded r_0-, r-intervals (though convergence in the K normal case seems slow; Subsubsection 6.10.2). ////

Proof of (3.117) The second equation of (3.67), in case $K = \mathcal{N}(0, \mathbb{I}_k)$, reads

$$r^2\, \tilde{c}_{k,r} = \mathrm{E}\left(|u||x|/\sqrt{k} - \tilde{c}_{k,r}\right)_+ \tag{4.49}$$

where $\tilde{c}_{k,r} = c_{r,k}/\sqrt{k}$, and in case $K = \mathrm{Ufo}_k(0,1)$,

$$r^2\, c_{k,r} = \mathrm{E}\left(|x|\,|u| - c_{k,r}\right)_+ \tag{4.50}$$

If $r = 0$, we have $c_{k,0} = \infty$. Now let $r > 0$, and suppose that, along some subsequence, $\tilde{c}_{k,r}$, respectively $c_{r,k}$, tend to some $\gamma \in [0, \infty]$. Since the RHS in (4.49) and (4.50) is bounded, necessarily $\gamma < \infty$. Then the noted L_1-convergence in the proof of (3.71) implies that (in both cases)

$$r^2\, \gamma = \mathrm{E}\left(|u| - \gamma\right)_+ \tag{4.51}$$

from which it follows that $\gamma = c_r^{1.\mathrm{loc}}$ from (3.27). Therefore, respectively,

$$\lim_{k\to\infty} \tilde{c}_{k,r} = c_k^{1.\mathrm{loc}} = \lim_{k\to\infty} c_{k,r} \tag{4.52}$$

The first equation of (3.67) for $K = \mathcal{N}(0, \mathbb{I}_k)$ reads

$$\alpha_{k,r}^{-1} = \mathrm{E}\, k^{-1}|x|^2u^2 \min\left\{1,\ \frac{\tilde{c}_{k,r}}{|u||x|/\sqrt{k}}\right\} \tag{4.53}$$

and for $K = \mathrm{Ufo}_k(0,1)$

$$\frac{k}{\alpha_{k,r}} = \mathrm{E}\,|x|^2u^2 \min\left\{1,\ \frac{c_{k,r}}{|u||x|}\right\} \tag{4.54}$$

For $r=0$, we have $\alpha_{k,0}=1$ (K normal) and $\alpha_{k,0}=k+2$ (K uniform). Suppose that $r>0$. Obviously, the integrands in (4.53) and (4.54) are uniformly integrable (being dominated by $\frac{1}{k}|x|^2u^2$, respectively by $|x|^2u^2$). In the normal case, we get

$$\lim_{k\to\infty}\alpha_{k,r}^{-1}=\mathrm{E}\,u^2\min\Big\{1,\ \frac{c_r^{1.\mathrm{loc}}}{|u|}\Big\}=\big(A_r^{1.\mathrm{loc}}\big)^{-1} \tag{4.55}$$

with $A_r^{1.\mathrm{loc}}=A_r$ from (3.67), which is consistent with $\alpha_{k,0}=1=A_0^{1.\mathrm{loc}}$ in case $r=0$. In the uniform regressor case, again consistently with $r=0$,

$$\lim_{k\to\infty}\frac{k}{\alpha_{k,r}}=\mathrm{E}\,u^2\min\Big\{1,\ \frac{c_r^{1.\mathrm{loc}}}{|u|}\Big\}=\big(A_r^{1.\mathrm{loc}}\big)^{-1} \tag{4.56}$$

The arguments leading to (4.52) and (4.55), (4.56) obtain if we replace the fixed $r\in(0,\infty)$ by any sequence r_k with limit $r\in(0,\infty)$. If $r_k\downarrow r=0$, a similar argument yields $\tilde c_{k,r},c_{k,r}\to c_0^{1.\mathrm{loc}}$, respectively. Thus (4.52) and (4.55), (4.56) hold true uniformly on bounded r-intervals. (4.52) and (4.55), (4.56) now entail convergence of the risk (with the previous r replaced by r_0),

$$\begin{aligned}k^{-1}\,\mathrm{maxMSE}\,(\eta_{r_0},r)&=\alpha_{k,r_0}^2\,\mathrm{E}\min\big\{k^{-1}|x|^2u^2,\ \tilde c_{k,r_0}^2\big\}+r^2\alpha_{k,r_0}^2\tilde c_{k,r_0}^2\\&\longrightarrow\mathrm{maxMSE}\,(\eta_{r_0}^{1.\mathrm{loc}},r)\end{aligned} \tag{4.57}$$

$$\begin{aligned}k^{-2}\,\mathrm{maxMSE}\,(\eta_{r_0},r)&=k^{-2}\alpha_{k,r_0}^2\,\mathrm{E}\min\big\{|x|^2u^2,\ c_{k,r_0}^2\big\}+r^2k^{-2}\alpha_{k,r_0}^2c_{k,r_0}^2\\&\longrightarrow\mathrm{maxMSE}\,(\eta_{r_0}^{1.\mathrm{loc}},r)\end{aligned} \tag{4.58}$$

in the normal and uniform case, respectively. Hence, the inefficiencies converge accordingly,

$$\lim_{k\to\infty}\mathrm{relMSE}\,(\eta_{r_0},r)=\mathrm{relMSE}\,(\eta_{r_0}^{1.\mathrm{loc}},r) \tag{4.59}$$

uniformly on bounded r_0-, r-intervals, in both cases.

We shall sketch an argument for (4.57) and (4.58), hence (4.59), to hold also in model 3.13.1 ($*=c$, $\alpha=\infty$).

Employing the L_1-convergence $|x|/\sqrt{k}\to 1$ and $|x|\to 1$ for $K=\mathcal{N}(0,\mathbb{I}_k)$ and $K=\mathrm{Ufo}_k(0,1)$, respectively, equation (3.131) determing $c_{k,r}(x)$ uniquely may be solved by $c_{k,r}(x)\to c_r^{1.\mathrm{loc}}$ in probability, where $c_r^{1.\mathrm{loc}}$ is taken from (3.27). At this instance, we assume but do not prove that the integrals $\mathrm{E}\,|x|c_{k,r}(x)/\sqrt{k}$ and $\mathrm{E}\,|x|c_{k,r}(x)$ converge correspondingly; that is, to $\mathrm{E}\,1\cdot c_r^{1.\mathrm{loc}}=c_r^{1.\mathrm{loc}}$. Under this asumption, however, equation (3.130) now entails (4.55) and (4.56), respectively. Then (4.57) and (4.58) follow as before.

Due to variable $c_{k,r}(x)$ (and matching the gap in the proof), the tables in Subsubsection 6.10.1 indicate only slow convergence in (4.57)–(4.59), but in the K uniform case, convergence is confirmed. ////

5 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:

5.1 One-Dimensional Location

The results for the models 3.3.1 and 3.3.2 are obtained by the routines for the k-dimensional location model 3.7 with $k = 1$. Note the coincidence (3.32) via the relation (3.31).

5.2 One-Dimensional Scale

5.2.1 $r/\sqrt{n}$ -Contamination Balls (Model 3.4.1)

The clipping bound c_r and the centering constant α_r in (3.33)–(3.36) are calculated by the R-function `uniroot()`.

5.2.2 $r/\sqrt{n}$-Total Variation Balls (Model 3.4.2)

We evaluate the clipping constants g_r and c_r in (3.39)–(3.42) by the R-function `uniroot()`.

5.3 k-Dimensional Location (Model 3.7)

We compute the constants c_r and α_r in (3.66), (3.67) 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 dimensional 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).

5.4 Regression, Average (Square) Contamination

5.4.1 Average Contamination Neighborhoods (Model 3.11.1)

The determination of clipping bound c_r in (3.113), (3.114) is performed by by the R-function `uniroot()`, where the integration of the outer integral is done numerically using the R-function `integrate()`.

5.4.2 Average Square Contamination Neighborhoods (Model 3.11.2)

The procedures may be obtained from the one-dimensional location case.

5.5 Regression, Average (Square) Hellinger

5.5.1 Average Hellinger Neighborhoods (Model 3.12.1)

For $K = \mathrm{Ufo}_k(0, m)$, in view of (3.124), (3.125), we have to find the zero of

$$c_r^{k+1} - (1 + 8r^2)(k + 1)c_r + k \tag{5.1}$$

to determine c_r in the interval $[0, 1]$. Because of the boundedness of the above expression (with r fixed in $(0, \infty)$) and differentiability in c_r, we can do this by a Newton method. In the normal case we can apply the routines from the k-dimensional location model 3.7 by substituting r_0 by $r_0/\sqrt{8}$ and r by $r/\sqrt{8}$.

5.5.2 Average Square Hellinger Neighborhoods (Model 3.12.2)

Nothing to calculate.

5.6 Constant Conditional Neighborhoods ($\alpha = \infty$)

5.6.1 Contamination Neighborhoods (Model 3.13.1)

We introduce the further parameter

$$\tau := \frac{r^2}{2}\,\mathrm{E}\,|x|c_r(x) \tag{5.2}$$

and determine $c_r(x)$ for fixed $\tau \in \big(0, 1/\sqrt{2\pi}\,\big)$ in case $K = \mathrm{Ufo}_k(0,1)$, respectively, for fixed $\tau \in \big(0,\infty\big)$ in case $K = \mathcal{N}(0, \mathbb{I}_k)$, from the equation

$$2\,\frac{\tau}{|x|} = \mathrm{E}_{\cdot}(|u| - c_r(x)) \tag{5.3}$$

Then r may be easily calculated back from (5.2).

We use the R-function `integrate()`, which also performs the evaluation of $c_r(x)$, where the computation of $c_r(x)$ is implemented as a vector valued function using the R-function `splinefun()`.

5.6.2 Hellinger Neighborhoods (Model 3.13.2)

Similar to the average Hellinger case we have to find the zero of

$$c_r^{k+1} - (1 + \tfrac{1}{8}r^2)(k+1)c_r + k \tag{5.4}$$

on the interval $[0,1]$, where r is fixed in $(0,\infty)$. So this computation can again be done by a Newton method.

Remark 5.1 In the case $\alpha = \infty$ (model 3.13), for both contamination and Hellinger neighborhoods, the MSE of η_{r_0} at $r = 0$ (ideal model) is unbounded as $r_0 \to \infty$.

In case $K = \mathcal{N}(0, \sigma^2\mathbb{I}_k)$, the least favorable radius r_0 cannot be determined over the unrestricted interval $[0,\infty)$, because $\mathrm{relMSE}\,(\eta_{r_0}, r) \to \infty$ as $r \to \infty$, for each $r_0 \in [0,\infty)$. This effect is connected with infimum bias 0. Therefore, we instead compute the least favorable radius r_0 for the bounded interval $[0, 2\sqrt{k}\,]$ (increasing with the dimension k). ////

5.7 General Procedures

In all these models, we do the following three calculations:

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

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.

5.8 Plots

The complete collection of 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. A small sample of the plots is attached at the end of this paper.

6 Tabulated Inefficiencies And Least Fav. Radii

6.1 k-Dimensional Location

relMSE over $r/\sqrt{n}$-contamination neighborhoods ($k \geq 1$) and, in case $k = 1$, relVar over symmetric s-contamination neighborhoods, 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
2	1.273	1.121	1.063	1.032	0.627	0.527	0.558
3	1.178	1.091	1.049	1.026	0.611	0.496	0.529
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

6.2 One-Dimensional Normal Scale

6.2.1 $r/\sqrt{n}$-Contamination Neighborhoods

relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
2.721	1.505	1.208	1.099	0.499	0.485	0.557

6.2.2 $r/\sqrt{n}$-Total Variation Neighborhoods

relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1.850	1.254	1.115	1.056	0.265	0.237	0.249

6.3 One-Dimensional Exponential Scale

6.3.1 $r/\sqrt{n}$-Contamination Neighborhoods

relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
2.081	1.381	1.167	1.082	0.495	0.417	0.438

6.3.2 $r/\sqrt{n}$-Total Variation Neighborhoods

relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1.718	1.222	1.104	1.051	0.285	0.252	0.265

6.4 One-Dimensional Location and Scale, Contamination

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

6.5 Binomial Model

6.5.1 $r/\sqrt{n}$ -Contamination Neighborhoods

n	θ	relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	arbit.	1.000	1.000	1.000	1.000	arbit.	arbit.	arbit.
2	0.1	1.056	1.053	1.036	1.022	0.108	0.083	0.085
	0.2	1.125	1.112	1.075	1.046	0.237	0.185	0.189
	0.293[4]	1.207	1.174	1.116	1.070	0.381	0.303	0.309
	0.5	1.000	1.000	1.000	1.000	arbit.	arbit.	arbit.
10	0.067[4]	1.393	1.280	1.157	1.096	0.496	0.421	0.438
	0.2	1.212	1.143	1.081	1.051	0.455	0.381	0.417
	0.355[4]	1.491	1.193	1.104	1.061	0.792	0.717	0.928
	0.5	1.245	1.149	1.086	1.050	0.558	0.445	0.542

6.5.2 $r/\sqrt{n}$ -Total Variation Neighborhoods

n	θ	relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	arbit.	1.000	1.000	1.000	1.000	arbit.	arbit.	arbit.
2	0.1	1.056	1.052	1.036	1.022	0.098	0.077	0.078
	0.2	1.125	1.105	1.072	1.045	0.193	0.158	0.161
	0.293[4]	1.207	1.144	1.100	1.063	0.284	0.245	0.250
	0.5	1.000	1.000	1.000	1.000	arbit.	arbit.	arbit.
10	0.067[4]	1.393	1.186	1.110	1.070	0.385	0.378	0.389
	0.2	1.189	1.128	1.078	1.045	0.232	0.180	0.200
	0.355[4]	1.491	1.191	1.102	1.059	0.391	0.359	0.462
	0.5	1.245	1.149	1.086	1.050	0.279	0.222	0.271

6.6 Poisson Model

6.6.1 $r/\sqrt{n}$ -Contamination Neighborhoods

λ	relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
0.1	1.052	1.049	1.032	1.020	0.071	0.055	0.056
0.5	1.297	1.227	1.130	1.080	0.368	0.310	0.317
0.693 [4]	1.443	1.302	1.164	1.099	0.517	0.417	0.468
1.0	1.222	1.160	1.088	1.050	0.372	0.252	0.278
1.5	1.355	1.183	1.089	1.052	0.617	0.504	0.677
1.678[4]	1.517	1.232	1.120	1.075	0.738	0.672	0.802
10.0	1.419	1.178	1.089	1.044	0.605	0.557	0.606
10.5	1.474	1.181	1.088	1.044	0.599	0.520	0.550
10.669[4]	1.563	1.186	1.088	1.045	0.620	0.501	0.533
∞	1.571	1.181	1.088	1.044	0.621	0.548	0.574

[4]median not uniquely defined

6.6.2 $r/\sqrt{n}$ -Total Variation Neighborhoods

λ	relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
0.1	1.052	1.049	1.032	1.020	0.068	0.053	0.054
0.5	1.297	1.187	1.110	1.070	0.300	0.271	0.278
0.693[5]	1.443	1.188	1.106	1.068	0.404	0.383	0.429
1.0	1.136	1.098	1.058	1.035	0.215	0.167	0.196
1.5	1.468	1.211	1.122	1.070	0.363	0.325	0.383
1.678[5]	1.517	1.189	1.097	1.050	0.306	0.305	0.228
10.0	1.394	1.173	1.089	1.044	0.302	0.271	0.312
10.5	1.556	1.183	1.087	1.044	0.314	0.269	0.279
10.669[5]	1.563	1.182	1.088	1.045	0.316	0.283	0.299
∞	1.571	1.181	1.088	1.044	0.311	0.274	0.287

6.7 Gamma Model, contamination

6.7.1 $r/\sqrt{n}$ -Contamination Neighborhoods

σ	α	relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	2	2.611	1.495			0.510		

6.8 Regression, Average (Square) Contamination

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

$\boldsymbol{K(dx) = \mathrm{Ufo}_k(0, m)}$

k	relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	2.094	1.271	1.122	1.060	0.566	0.517	0.540
2	1.767	1.227	1.107	1.053	0.595	0.532	0.558
3	1.677	1.209	1.100	1.049	0.604	0.536	0.562
5	1.616	1.194	1.094	1.047	0.611	0.540	0.565
10	1.584	1.185	1.090	1.045	0.617	0.545	0.570
15	1.577	1.183	1.089	1.044	0.619	0.546	0.572

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

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
2	2.000	1.287	1.127	1.062	0.555	0.499	0.525
3	1.851	1.258	1.117	1.057	0.569	0.506	0.534
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

[5] median not uniquely defined

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

Same numbers as in one-dimensional location.

6.9 Regression, Average (Square) Hellinger

6.9.1 Average Hellinger Neighborhoods ($* = h$, $\alpha = 1$)

$\boldsymbol{K(dx) = \mathrm{Ufo}_k(0, m)}$

k	relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	1.333	1.101	1.055	1.029	0.255	0.231	0.238
2	1.125	1.055	1.032	1.017	0.247	0.211	0.220
3	1.067	1.035	1.021	1.011	0.232	0.191	0.199
5	1.029	1.018	1.011	1.006	0.207	0.162	0.169
10	1.008	1.006	1.004	1.002	0.170	0.124	0.129
15	1.004	1.003	1.002	1.000	0.149	0.104	arbitrary

$\boldsymbol{K(dx) = \mathcal{N}(0, \sigma^2 \mathbb{I}_k)}$
Same numbers as in k-dimensional location, but $r_\rho = r_\rho^{\mathrm{k.loc}}/\sqrt{8}$.

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	0.220	0.194	0.203
2	1.273	1.121	1.063	1.032	0.222	0.186	0.197
3	1.178	1.091	1.049	1.026	0.216	0.175	0.187
5	1.104	1.062	1.035	1.018	0.204	0.159	0.170
10	1.051	1.035	1.020	1.011	0.184	0.136	0.146
15	1.034	1.025	1.014	1.008	0.171	0.124	0.133

6.9.2 Average Square Hellinger Neighborhoods ($* = h$, $\alpha = 2$)

relMSE $\equiv 1$.

6.10 Constant Conditional Neighborhoods

6.10.1 Constant Contamination Neighborhoods ($* = c$, $\alpha = \infty$)

$\boldsymbol{K(dx) = \mathrm{Ufo}_k(0, m)}$

k	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	1.577	1.185	1.085	1.579	1.621	1.579
2	1.481	1.159	1.074	1.310	1.383	1.355
3	1.420	1.141	1.065	1.198	1.270	1.253
5	1.348	1.117	1.054	1.071	1.064	1.010
10	1.271	1.096	1.047	0.894	0.691	0.652
15	1.241	1.092	1.045	0.807	0.607	0.606

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

k	$[0, 2\sqrt{k}\,]$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	1.224	1.343	1.144	0.790	4.805	4.254
2	1.320	1.323	1.137	0.923	3.895	3.463
3	1.367	1.307	1.131	0.994	3.604	3.164
5	1.409	1.281	1.121	1.075	3.410	2.982
10	1.428	1.240	1.104	1.164	3.434	2.990
15	1.421	1.214	1.094	1.200	3.533	3.093

6.10.2 Constant Hellinger Neighborhoods ($* = h$, $\alpha = \infty$)

$\boldsymbol{K(dx) = \mathrm{Ufo}_k(0, m)}$

k	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	1.378	1.144	1.068	0.839	0.924	0.862
2	1.300	1.120	1.058	0.737	0.857	0.793
3	1.250	1.102	1.050	0.713	0.866	0.797
5	1.189	1.080	1.039	0.715	0.924	0.848
10	1.120	1.051	1.026	0.768	1.091	1.004
15	1.088	1.038	1.019	0.822	1.240	1.147

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

k	$[0, 2\sqrt{k}\,]$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
1	1.312	1.332	1.140	0.672	2.198	1.938
2	1.427	1.313	1.133	0.720	1.784	1.559
3	1.474	1.297	1.127	0.743	1.636	1.420
5	1.505	1.273	1.118	0.772	1.530	1.317
10	1.497	1.233	1.102	0.813	1.499	1.277
15	1.472	1.208	1.092	0.841	1.528	1.296

6.11 ARMA(1,1), Average (Square) Contamination

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

$\boldsymbol{\phi = -0.7,\ \xi = 0.35}$ **:**

relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
2.163	1.314	1.136	1.065	0.541	0.494	0.520

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

Same numbers as one-dimensional location independent from ϕ and ξ.

6.12 ARCH(1), Average (Square) Contamination

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

$\boldsymbol{\theta = 1}$

relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
3.847[6]	1.583[7]	1.230[8]	1.117[8]	0.424[7]	0.381[8]	0.405[8]

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

Same numbers as one-dimensional scale (contamination neighborhoods).

6.13 ARCH(1), Average (Square) Total Variation

6.13.1 Average Total Variation Neighborhoods ($* = v$, $\alpha = 1$)

$\boldsymbol{\theta = 1}$

relMSE $(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	r_0	r_3	r_2
2.616[9]	1.332[9]	1.140[10]	1.067[10]	0.235[9]	0.215[10]	0.225[10]

6.13.2 Average Square Total Variation Neighborhoods ($* = v$, $\alpha = 2$)

Same numbers as one-dimensional scale (total variation neighborhoods).

[7] mean of 100 Monte Carlo simulations, standard error < 0.0005
[8] mean of 10 Monte Carlo simulations, standard error < 0.0007
[9] mean of 100 Monte Carlo simulations, standard error < 0.0005
[10] mean of 10 Monte Carlo simulations, standard error < 0.0007

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'Eté de Probabilités de Saint Flour IX 1979* (P.L. Hennequin, ed.), 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. (1987): Robust regression estimators and their least favorable contamination curves. *Stat. Decis.* **5** 307–336.

Rieder, H. (1989): A finite-sample minimax regression estimator. *Statistics* **20** 211–221.

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

Rieder, H. (2000): Neighborhoods as nuisance parameters? Robustness vs. semiparametrics. Submitted to Ann. Statist.

Rieder, H. (2001 a): Robust time series based on infinitesimal neighborhoods. Manuscript.

Rieder, H. (2001 b): Nonuniform convergence of adaptive estimators in linear time series models. Manuscript.

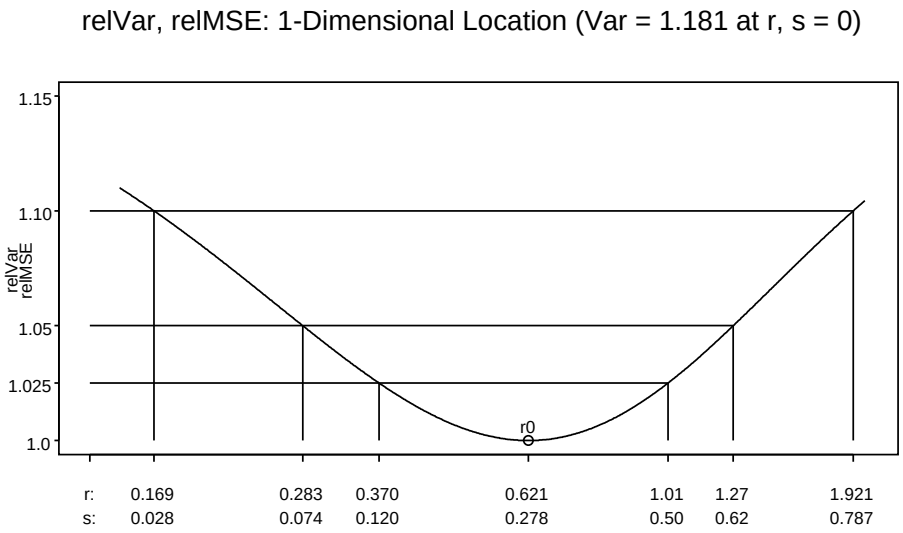
Ruckdeschel, P. (2001): *Ansätze zur Robustifizierung des Kalman–Filters*. Dissertation, Universität Bayreuth.

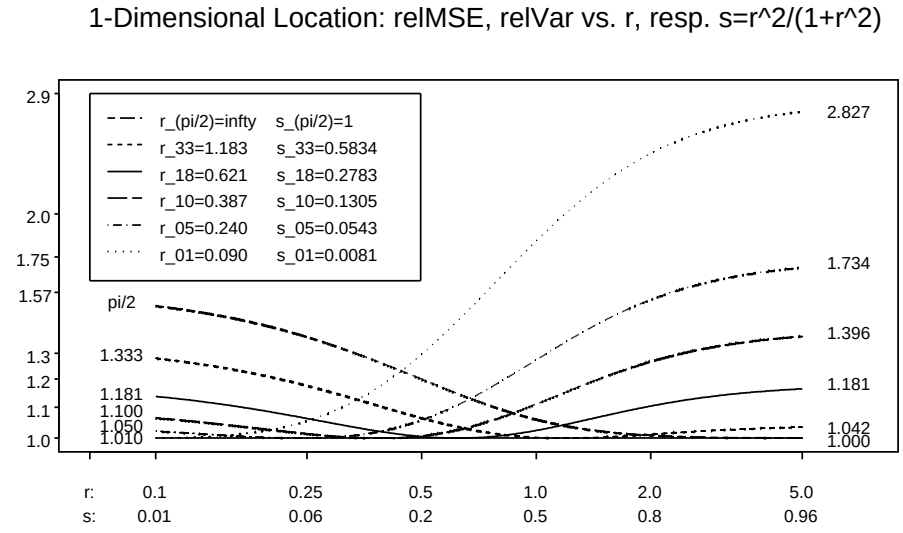
Ruckdeschel, P. (2005): Some more properties of Lagrange multipliers. Technical report, Universität Bayreuth; available upon request.

Ruckdeschel, P. and Rieder, H. (2004): Optimal influence curves for general loss functions. *Statistics and Decisions* **22** 201–223.

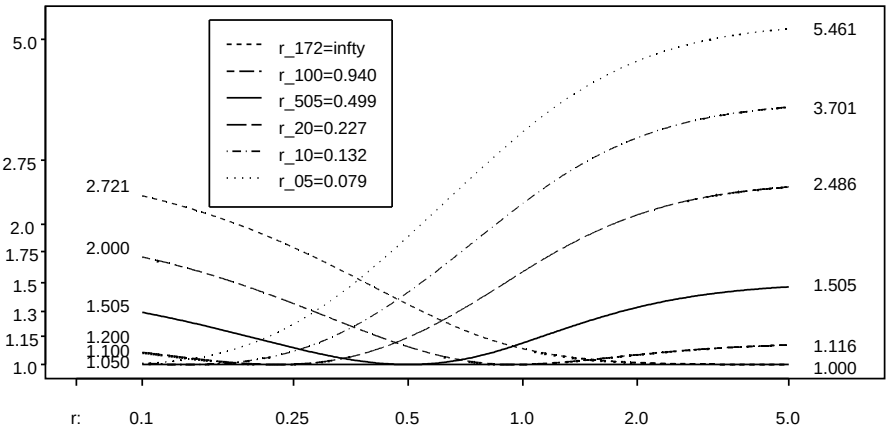
Witting, H. and Müller–Funk, U. (1995): *Mathematische Statistik II*. B.G. Teubner, Stuttgart.

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

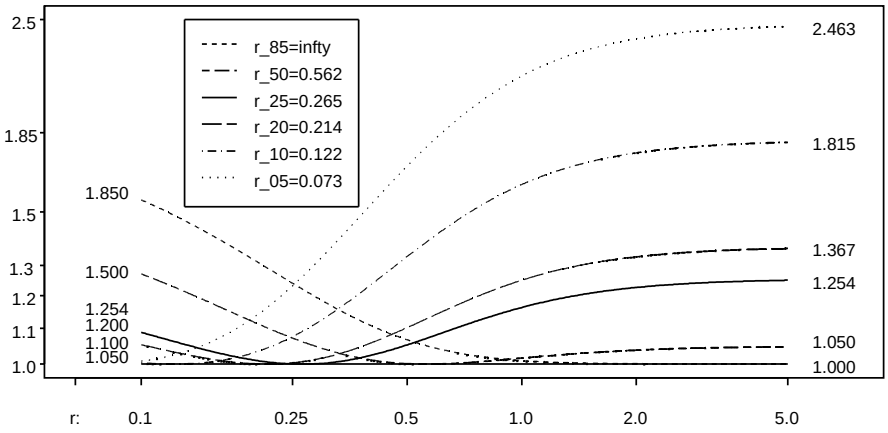

relVar, relMSE: 1-Dimensional Location (Var = 1.181 at r, s = 0)
1.15
1.10
1.05
1.025
1.0
relVar
relMSE
r0
r: 0.169 0.283 0.370 0.621 1.01 1.27 1.921
s: 0.028 0.074 0.120 0.278 0.50 0.62 0.787


1-Dimensional Location: relMSE, relVar vs. r, resp. s=r^2/(1+r^2)
r_(pi/2)=infty s_(pi/2)=1
r_33=1.183 s_33=0.5834
r_18=0.621 s_18=0.2783
r_10=0.387 s_10=0.1305
r_05=0.240 s_05=0.0543
r_01=0.090 s_01=0.0081
2.9
2.0
1.75
1.57
1.3
1.2
1.1
1.0
pi/2
1.333
1.181
1.100
1.050
1.010
2.827
1.734
1.396
1.181
1.042
1.000
r: 0.1 0.25 0.5 1.0 2.0 5.0
s: 0.01 0.06 0.2 0.5 0.8 0.96

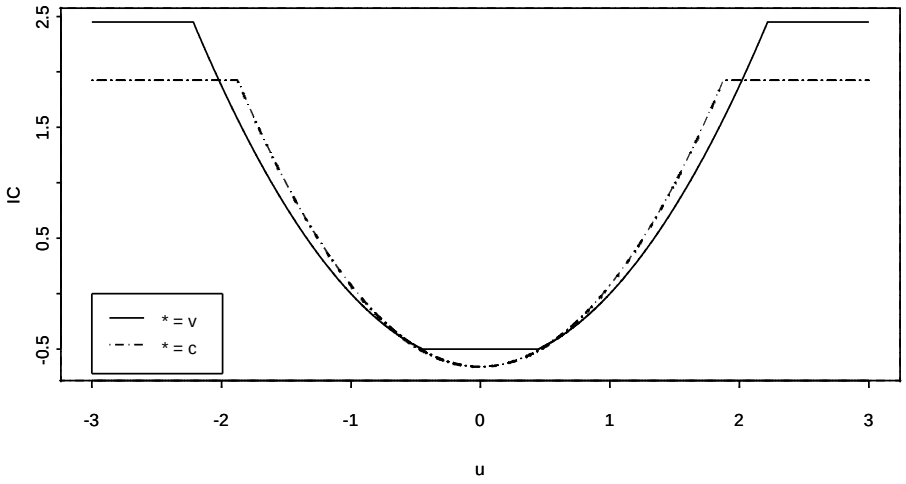
relMSE: 1-Dimensional Scale (contamination)



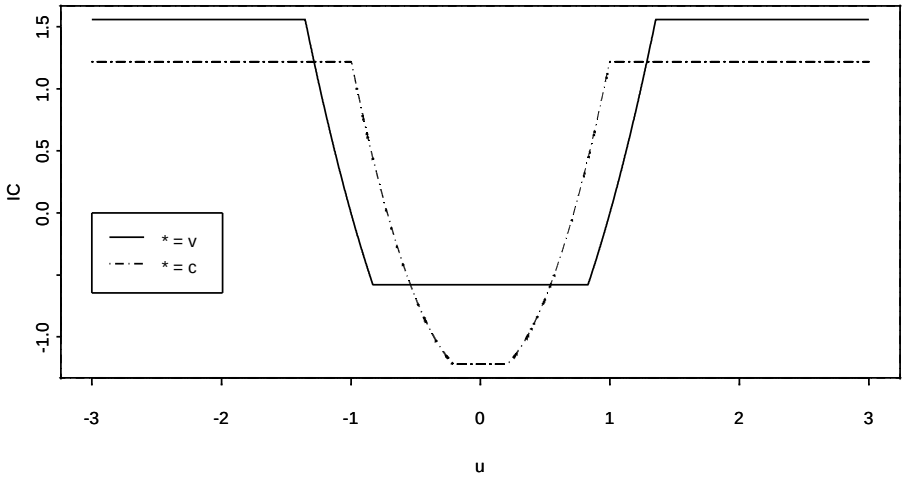
relMSE: 1-Dimensional Scale (total variation)



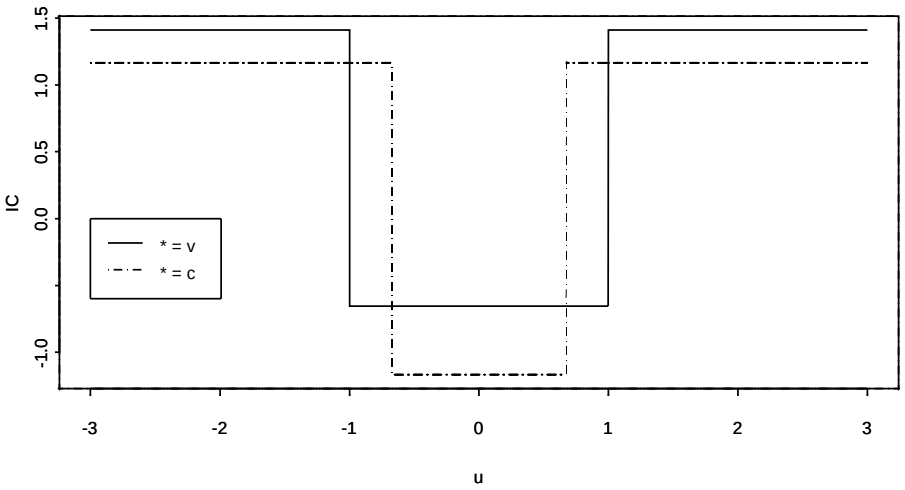
1-Dimensional Scale: IC-comparison for r=0.2 (*=c) with r=0.1 (*=v)



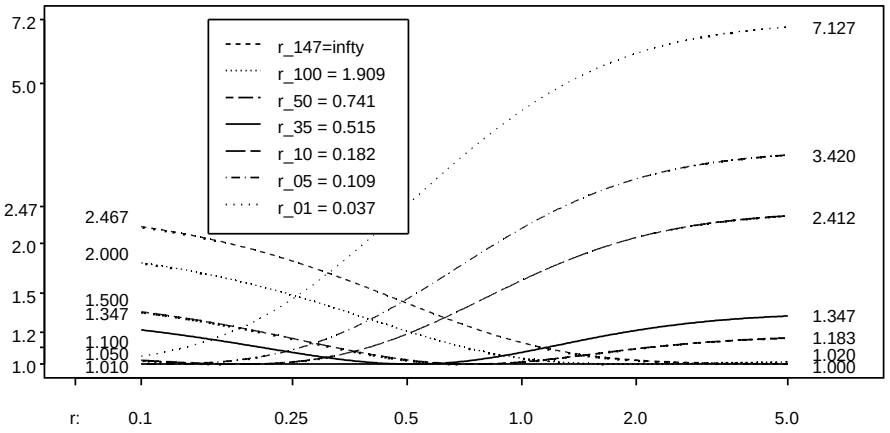
1-Dimensional Scale: IC-comparison for r=1.0 (*=c) with r=0.5 (*=v)



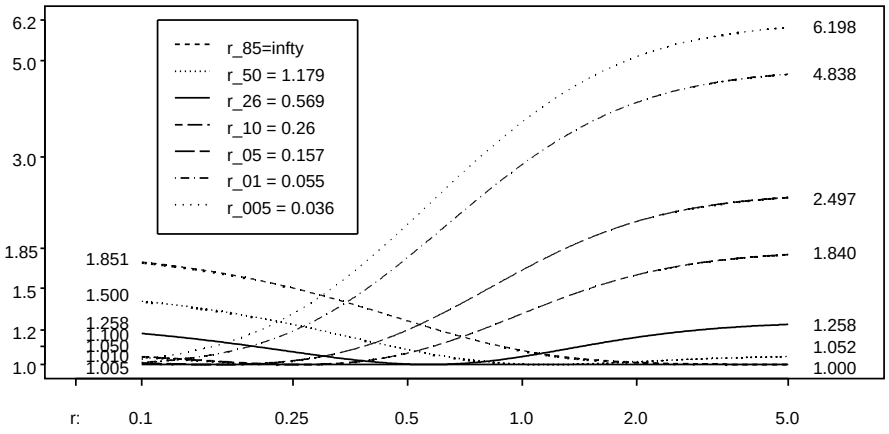
1-Dimensional Scale: most robust IC (*=v,c)



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



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



relMSE: Regression (*=c, alpha=1, K uniform, dim=2)