

Rieder, Helmut; Kohl, Matthias; Ruckdeschel, Peter

Working Paper

The costs of not knowing the radius

SFB 373 Discussion Paper, No. 2001,81

Provided in Cooperation with:

Collaborative Research Center 373: Quantification and Simulation of Economic Processes,
Humboldt University Berlin

Suggested Citation: Rieder, Helmut; Kohl, Matthias; Ruckdeschel, Peter (2001) : The costs of not knowing the radius, SFB 373 Discussion Paper, No. 2001,81, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, <https://nbn-resolving.de/urn:nbn:de:kobv:11-10050581>

This Version is available at:

<https://hdl.handle.net/10419/62707>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Costs of Not Knowing the Radius

Helmut Rieder

Matthias Kohl, Peter Ruckdeschel

University of Bayreuth, Germany

5 October 2001

Abstract

We determine the increase of the maximum risk over the minimax risk in the case that the optimally robust estimator for the false radius is used. This is done by numerical solution of the implicit equations which determine optimal robustness, for location, scale, and linear regression models, and by evaluation of maximum asymptotic variance and mean square error over fixed-size symmetric contamination and infinitesimal asymmetric neighborhoods, respectively. The maximum increase of the relative risk is minimized in the case that the radius is known only to belong to some interval $[\rho r, r/\rho]$. The effect of increasing parameter dimension is studied for these models.

The minimax increase of relative risk in case $\rho = 0$, 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, regression, AR(1), MA(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, model deviations may be treated as values of an infinite dimensional nuisance parameter; confer Rieder¹(2000). But the more elementary case of the radius as a nuisance parameter has not been considered by mathematical and semiparametric statistics.

Some textbooks even 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 .’ Despite of numerous quotations of the ‘occurrence of outliers’, Anmerkung 6.44 and Beispiel 7.4.5 (declaring the radius to be unknown in practice) are the only more mathematical appreciations of robustness in that book. Linked up with one other, they shed a dubious light of arbitrariness on robust procedures. But, apparently, the authors do not distinguish L_2 - and L_∞ -norms in this context.

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 the original 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 $r = r_0$).

Numerical evaluations in all our models establish the inefficiency curves as bowl-shaped, smoothly increasing from the value 1 at $r = r_0$ towards both sides to two relative maxima at the interval boundaries. Determination of r_0 so as to equate both boundary values will minimize the maximal subefficiency over r in the respective estimator class (M -estimates, asymptotically linear estimators).

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

Remark 1.1 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. ///

¹HR, henceforth

In addition to the true radius r being completely unknown (unrestricted radius interval), 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 case². 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.

In the course of these derivations, for the models considered, 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).

1.3 One-Dimensional Robust Location

The results are summarized first for one-dimensional location and secondly scale.

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

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

The subefficiency of the size-minimax M -estimate is the maximal 18.1% only at the unrealistic size boundaries 0 and 1. On more realistic size intervals (about $s_0 = .278$), it stays well below 18.1%: below 2.5% for $.12 \leq s \leq .50$, below 5% for $.074 \leq s \leq .62$, and still below 10% for $.028 \leq s \leq .78$.

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

Remark 1.2 Via relation (1.1) below, $s_0 = .278$ corresponds to $r_0 = .62$, and the corresponding radius intervals about r_0 read:

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

²The interested reader may use access name **radius** and password **unknown** to draw our computer program from <http://www.uni-bayreuth.de/departments/math/org/mathe7/radius>.

1.4 The Infinitesimally Robust MSE-Setup

is the other and in fact more flexible approach. It includes possibly asymmetric contamination of radius r/\sqrt{n} at sample size n (of i.i.d. laws or stationary transition probabilities), and employs asymptotically linear estimators, which are judged by maximum asymptotic mean square error; confer HR (1994; Sec. 5.5).

In the one-dimensional location model about the standard normal with i.i.d. observations—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:

$$s = \frac{r^2}{1+r^2} \quad (1.1)$$

Thus, the least favorable (starting) radius is $r_0 = .62$, 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 1.2). 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.

1.5 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 also 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 20.7% and 9.9%, respectively. The corresponding least favorable radii $r_3 = .48$ and $r_2 = .55$ define least favorable contamination ranges 1.6%–14.4% and 2.7%–11.0%, at sample size $n = 100$, respectively.

Remark 1.3 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. The inefficiency curves thus agree and the radius-minimax robust procedure stays the same. The coincidence extends to the k -dimensional location and regression models of our study. ////

Due to asymmetry 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 2.3.

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 5.2.

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.

1.6 Two Conclusions

may be drawn from these results, which extend to higher dimensional location and regression (whose summary is postponed to Subsections 1.9 and 1.10):

(i) **The minimax subefficiency is small.** Small in comparison with the most robust estimator, and small for practical purposes. Adaptive estimation of the radius hence seems neither necessary nor worthwhile—provided however 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.

1.7 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 $\mathcal{I}_F^{\text{loc}} = \int (\Lambda_F^{\text{loc}})^2 dF$, where $\Lambda_F^{\text{loc}} = -f'/f$, and consider the location M -estimate defined by some odd function $\psi_0: \mathbb{R} \rightarrow \mathbb{R}$; for example, $\psi_0 = \Lambda_{F_0}^{\text{loc}}$ for some other such law F_0 .

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

$$\text{Var}_{\text{loc}}(\psi_0, F) = \frac{\int \psi_0^2 dF}{\left(\int \psi_0 \Lambda_F^{\text{loc}} dF\right)^2} \in (0, \infty) \quad (1.2)$$

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

$$\sup_{F \in U_c^{s,i}(F_0, \varepsilon)} \text{Var}_{\text{loc}}(\psi_0, F) \cdot \mathcal{I}_F^{\text{loc}} = \infty \quad \forall \varepsilon \in (0, 1) \quad (1.3)$$

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

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

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).

1.8 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 adaptive, fully efficient estimators cannot achieve the desirable uniformity. Using equivariance, Klaassen (1980) derives such a finite-sample result for the one-dimensional location model. Bickel (1981; Note, p 51) asks for extensions. Consequences are noted by Bickel (1982; Remark 5.5) and Huber (1996; Sec. 28). Pfanzagl 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 $\Lambda_\theta(x, y) = \Lambda_F^{\text{loc}}(y - x'\theta) x$ and Fisher information $\mathcal{I}_\theta = \mathcal{I}_F^{\text{loc}} \mathcal{K}$.

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

$$\mathcal{V}_{\theta, F}^n := \mathcal{L}_{\theta, F} \{ \sqrt{n} \mathcal{I}_{\theta, F}^{1/2} (S_n - \theta) \} \xrightarrow{w} \mathcal{N}(0, \mathbb{I}_k) \quad (1.4)$$

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

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

$$\liminf_{n \rightarrow \infty} \sup_{F \in U_c^{s,i}(F_0, \varepsilon_n)} d_\kappa(\mathcal{V}_{\theta, F}^n, \mathcal{N}(0, \mathbb{I}_k)) \geq 1 - \frac{1}{2^k} \quad (1.5)$$

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

Remark 1.4 The result is contained in HR (2001 b), where it is proved for more general i.i.d. models of location or scale structure, and shown to hold also for MA(q)-models. An extension to AR(p)- and ARMA(p, q)-models we prove with

the bound $\geq 1 - 2^{-k}$ weakened to > 0 and 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 (1.3) and (1.5). 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 (1.4) 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 asymmetric shrinking neighborhoods. But, by simple set inclusions, the passage to the supremum $g(x, r) = \sup\{f(y) \mid y \in B(x, r)\}$ of any function f over balls $B(x, r)$ already implies continuity of $g(\cdot, r)$ at x for almost all radii; namely, for those r such that $g(x, r - 0) = g(x, r + 0)$.

Uniform weak convergence of optimally robust estimators over neighborhoods with bounded radius has been established by Beran (1981), Millar, P.W. (1981), Bickel (1981; Théorème 5), Huber (1981; Sec. 3.2, p 51), 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.

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

1.9 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

$$\omega_c^{\min} = \frac{k\Gamma(\frac{k}{2})}{\sqrt{2}\Gamma(\frac{k+1}{2})} \approx \sqrt{k} \quad \text{as } k \rightarrow \infty \quad (1.6)$$

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, only one out of k spherical coordinates (length) is sacrificed

by its influence curve. Consequentially, the minimum L_1 -estimate becomes the nearly optimal choice for larger dimension. For $k \geq 5$, its maximum subefficiency over the full radius range is less than 10.4%, for $k \geq 10$ less than 5.1%, and it stays within a factor of 2 of the minimax value, both subefficiencies decreasing to 0 as $k \rightarrow \infty$; confer Subsections 3.6 and 5.1.

1.10 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 $\text{Ufo}_k(0, m)$ 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 $\|\varepsilon\|_\alpha \leq 1$ for $\alpha = 1, 2, \infty$. These cases obtain the attributes average, average square, and constant conditional, respectively.

1.10.1 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 (unconditional) infinitesimal contamination neighborhoods, thus, inefficiencies and least favorable radii are the same as for one-dimensional location; confer Subsubsection 2.5.2.

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 $k \rightarrow \infty$. Related numbers, e.g. the minimax subefficiency in case the radius can be specified up to factor 3 or 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 5.3.1. 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 5.5.1).

1.10.2 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 $o(r^2)$ as $r \rightarrow \infty$ of the estimator which is optimally robust for radius r , (iv) slow convergence of the inefficiency curves as $k \rightarrow \infty$, and (v) relatively large least favorable radii; confer Subsections 2.7 and 5.5.

1.10.3 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 neighborhoods, 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 $k \rightarrow \infty$. 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 $k \rightarrow \infty$.

Remark 1.5 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 the norm of the spherically symmetric regressor is about \sqrt{k} in the case of the k -variate standard normal, and about 1 in the case of the uniform distribution on the unit ball in \mathbb{R}^k , if k is large. ///

The paper proceeds as follows. In Section 2, the theoretical setup is formulated: For a selection of ideal models and neighborhood systems, the optimally robust estimators and their risk functions are determined by specialization of the general results in HR (1994; Chaps. 5 and 7). Some relations between the inefficiency curves in different models are derived.

Section 3 contains the mathematical proofs. The numerical algorithms to evaluate the theoretical formulas are described in Section 4.

The computed numbers are presented in the tables of Section 5. A selection of plots is attached.

2 Optimally Robust Estimates and Their Inefficiency Curves

2.1 Robust Setup

2.1.1 Ideal Models With The Normal Distribution

(a) **k -dimensional normal location:**

$$y_i = \theta + u_i \quad (2.1)$$

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.

(b) **one-dimensional normal scale:**

$$y_i = \theta u_i \quad (2.2)$$

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 $\mathcal{I}_\theta = 2 \theta^{-2}$.

(c) **k -dimensional normal linear regression:**

$$y_i = x_i' \theta + u_i \quad (2.3)$$

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 $\mathcal{I}_\theta = \sigma_u^{-2} \mathbb{E}_K x x'$.

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

(d) **order one autoregression and moving average:**

$$y_i = \theta y_{i-1} + u_i \quad (2.4)$$

respectively

$$y_i = u_i - \theta u_{i-1} \quad (2.5)$$

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.

In models (a)–(c), the observations are i.i.d.. The inefficiencies turn out invariant under rescaling of the u_i and x_i , respectively. So we may fix

$$\sigma_u = 1, \quad \sigma_x = 1, \quad m_x = 1 \quad (2.6)$$

Moreover, $\theta = 0$ may be fixed in models (a) and (c), and $\theta = 1$ in model (b), due to equivariance of these models.

In models (d), 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 (c), the inefficiencies turn out the same for all values $|\theta| < 1$ and $\sigma_u \in (0, \infty)$.

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

2.1.2 Neighborhoods

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

$$Q = (1 - s)P + sH \quad (2.7)$$

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** 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

$$Q_n = (1 - r_n)P + r_n H_n \quad (2.8)$$

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

$$d_*(Q_n, P) \leq r_n \quad (2.9)$$

where

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

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

Infinitesimal neighborhoods are employed in the location, scale, and regression models (a), (b), and (c). 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** 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 $\varepsilon: \mathbb{R}^k \rightarrow [0, \infty)$, which weights the radius depending on the regressor, is called radius curve.

We employ conditional, or error-free-variables, neighborhoods with varying radius curves ε subject to $L_\alpha(K)$ -norm $\|\varepsilon\|_\alpha \leq 1$ 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 (1982) 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 (2001 a).

2.1.3 Risk and Inefficiency

The asymptotic maximum MSE of the asymptotic linear estimator with influence curve (IC)³ 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

$$\max\text{MSE}(\eta_{r_0}, r) = \mathbb{E} |\eta_{r_0}|^2 + r^2 \omega_{*,\alpha}^2(\eta_{r_0}) \quad (2.12)$$

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 ,

$$\text{relMSE}(\eta_{r_0}, r) = \frac{\max\text{MSE}(\eta_{r_0}, r)}{\max\text{MSE}(\eta_r, r)} \quad (2.13)$$

2.2 One-Dimensional Location

2.2.1 Minimax Asymptotic Variance

We consider the $k = 1$ dimensional standard normal location model 2.1.1 (a) 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

$$\psi_s(u) = (-m_s) \vee u \wedge m_s, \quad \frac{s}{1-s} m_s = \mathbb{E} (|u| - m_s)_+ \quad (2.14)$$

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

$$\psi_1(u) = \text{sign}(u) \quad (2.15)$$

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

$$\max\text{Var}(\psi_{s_0}, s) = \frac{(1-s) \mathbb{E} \psi_{s_0}^2 + s m_{s_0}^2}{[(1-s) \mathbb{E} \psi'_{m_{s_0}}]^2} \quad (2.16)$$

respectively, in case of the median,

$$\max\text{Var}(\psi_1, s) = \frac{\pi}{2(1-s)^2} \quad (2.17)$$

³In the following, influence curve is abbreviated by IC.

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

$$\text{relVar}(\psi_{s_0}, s) = \frac{\max \text{Var}(\psi_{s_0}, s)}{\max \text{Var}(\psi_s, s)} \quad (2.18)$$

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

$$\lim_{s \rightarrow 1} \text{relVar}(\psi_1, s) = 1 \quad (2.19)$$

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

We consider the $k = 1$ dimensional standard normal location model 2.1.1 (a) secondly with infinitesimal contamination neighborhoods 2.1.2 (b).

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

$$\eta_r(u) = A_r u \min \{1, c_r |u|^{-1}\} \quad (2.20)$$

where

$$1 = A_r \mathbb{E} |u| \min \{|u|, c_r\}, \quad r^2 c_r = \mathbb{E} (|u| - c_r)_+ \quad (2.21)$$

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

$$\eta_\infty(u) = \omega_c^{\min} \text{sign}(u) \quad (2.22)$$

which is the IC of the median and achieves minimum bias

$$\omega_c^{\min} = (\mathbb{E} |\Lambda|)^{-1} = \sqrt{\frac{\pi}{2}} \quad (2.23)$$

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\max \text{MSE}(\eta_{r_0}, r) = A_{r_0}^2 \mathbb{E} \min \{u^2, c_{r_0}^2\} + r^2 A_{r_0}^2 c_{r_0}^2 \quad (2.24)$$

In Subsection 3.2 we prove the following relation between the maximum risks of the optimal estimates in the two models 2.2.1 and 2.2.2,

$$(1 - s) \max \text{MSE}(\eta_{r_0}, r) = \max \text{Var}(\psi_{s_0}, s) \quad (2.25)$$

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

$$s = r^2 / (1 + r^2), \quad s_0 = r_0^2 / (1 + r_0^2) \quad (2.26)$$

Consequently, by (2.25) and (2.26), the inefficiency curves coincide in the two models,

$$\text{relMSE}(\eta_{r_0}, r) = \text{relVar}(\psi_{s_0}, s) \quad (2.27)$$

2.3 One-Dimensional Scale

2.3.1 r/\sqrt{n} -Contamination Balls

We consider the one-dimensional standard normal scale model 2.1.1 (b) 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

$$\eta_r(u) = A_r(u^2 - \alpha_r^2) \min \left\{ 1, \frac{c_r}{|u^2 - \alpha_r^2|} \right\} \quad (2.28)$$

where

$$0 = E(u^2 - \alpha_r^2) \min \left\{ 1, \frac{c_r}{|u^2 - \alpha_r^2|} \right\} \quad (2.29)$$

$$1 = A_r E |u^2 - \alpha_r^2| \min \{ |u^2 - \alpha_r^2|, c_r \} \quad (2.30)$$

and

$$r^2 c_r = E (|u^2 - \alpha_r^2| - c_r)_+ \quad (2.31)$$

The parabola $u^2 - \alpha_r^2$ in (2.28) is clipped only from above for radius $r \leq 0.920$, and for $r \geq 0.920$ from above as well as from below. The centering constant α_r decreases from $\alpha_0 = 1$ to $\alpha_\infty = \Phi^{-1}(3/4) \approx 0.674$.

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

$$\eta_\infty(u) = \omega_c^{\min} \text{sign}(u^2 - \alpha_\infty^2) \quad (2.32)$$

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

$$\omega_c^{\min} = (E |u^2 - \alpha_\infty^2|)^{-1} = (4\alpha_\infty \varphi(\alpha_\infty))^{-1} \approx 1.166 \quad (2.33)$$

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\max \text{MSE}(\eta_{r_0}, r) = A_{r_0}^2 E \min \{ |u^2 - \alpha_{r_0}^2|^2, c_{r_0}^2 \} + r^2 A_{r_0}^2 c_{r_0}^2 \quad (2.34)$$

2.3.2 r/\sqrt{n} -Total Variation Balls

We consider the one-dimensional standard normal scale model 2.1.1 (b) 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

$$\eta_r(u) = A_r \{ [g_r \vee u^2 \wedge (g_r + c_r)] - 1 \} \quad (2.35)$$

where

$$0 = E(g_r - u^2)_+ - E(u^2 - g_r - c_r)_+ \quad (2.36)$$

$$1 = A_r E u^2 \{ [g_r \vee u^2 \wedge (g_r + c_r)] - 1 \} \quad (2.37)$$

and

$$r^2 c_r = E(g_r - u^2)_+ \quad (2.38)$$

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

$$\eta_\infty(u) = \omega_v^{\min} \{ P(u^2 < 1) \mathbf{I}(u^2 > 1) - P(u^2 > 1) \mathbf{I}(u^2 < 1) \} \quad (2.39)$$

with minimum bias

$$\omega_v^{\min} = (\mathbb{E} \Lambda_+)^{-1} = \sqrt{\frac{\pi}{2}} e \approx 2.066 \quad (2.40)$$

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\max \text{MSE}(\eta_{r_0}, r) = A_{r_0}^2 \mathbb{E} \{ [g_{r_0} \vee u^2 \wedge (g_{r_0} + c_{r_0})] - 1 \}^2 + r^2 A_{r_0}^2 c_{r_0}^2 \quad (2.41)$$

2.4 k -Dimensional Location, Contamination

We consider the k -dimensional normal location model 2.1.1 (a) 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 3.3 below), is

$$\eta_r(u) = \alpha_r u \min \{1, c_r |u|^{-1}\} \quad (2.42)$$

where

$$k = \alpha_r \mathbb{E} |u| \min \{|u|, c_r\}, \quad r^2 c_r = \mathbb{E} (|u| - c_r)_+ \quad (2.43)$$

For $r = \infty$, we put

$$\eta_\infty = \omega_c^{\min} \frac{u}{|u|} \quad (2.44)$$

which is the IC of the minimum L_1 -estimate, and attains minimum bias ω_c^{\min} ; confer HR (1994; Theorem 5.5.1.b). In Subsection 3.4 we show that

$$\omega_c^{\min} = \frac{k}{\mathbb{E} |\Lambda|} = \frac{k \Gamma(\frac{k}{2})}{\sqrt{2} \Gamma(\frac{k+1}{2})} \quad (2.45)$$

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\max \text{MSE}(\eta_{r_0}, r) = \alpha_{r_0}^2 \mathbb{E} \min \{|u|^2, c_{r_0}^2\} + r^2 \alpha_{r_0}^2 c_{r_0}^2 \quad (2.46)$$

For increasing dimension, we prove in Subsection 3.6 that

$$\lim_{k \rightarrow \infty} \frac{\omega_c^{\min}}{\sqrt{k}} = \lim_{k \rightarrow \infty} \frac{\mathbb{E} |\eta_\infty|^2}{k} = 1 \quad (2.47)$$

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

$$\lim_{k \rightarrow \infty} \frac{\max \text{MSE}(\eta_{r_0}, r)}{\max \text{MSE}(\eta_\infty, r)} = 1 \quad (2.48)$$

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

2.5 Regression, Average (Square) Contamination

2.5.1 Average Contamination Neighborhoods ($\ast = c$, $\alpha = 1$)

We consider the k -dimensional normal regression model 2.1.1 (c) 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 spherical symmetry (Lemma 3.3), is

$$\eta_r(x, u) = \alpha_r x u \min \{1, c_r |xu|^{-1}\} \quad (2.49)$$

where

$$k = \alpha_r \mathbb{E} |xu| \min \{|xu|, c_r\}, \quad r^2 c_r = \mathbb{E} (|xu| - c_r)_+ \quad (2.50)$$

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

$$\eta_\infty(x, u) = \omega_{c,1}^{\min} \frac{x}{|x|} \text{sign}(u) \quad (2.51)$$

which achieves minimum bias $\omega_{c,1}^{\min}$. Analogously to (2.45) we show that

$$\omega_{c,1}^{\min} = \frac{k}{\mathbb{E} |\Lambda|} = \sqrt{\frac{\pi}{2}} \frac{k}{\mathbb{E} |x|} \quad (2.52)$$

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\max \text{MSE}(\eta_{r_0}, r) = \alpha_{r_0}^2 \mathbb{E} \min \{|x|^2 u^2, c_{r_0}^2\} + r^2 \alpha_{r_0}^2 c_{r_0}^2 \quad (2.53)$$

In Subsection 3.6, for increasing dimension $k \rightarrow \infty$ we prove that the MSE-inefficiency tends uniformly on bounded r_0 -, r -intervals to the MSE-inefficiency in the one-dimensional location model 2.2.2,

$$\lim_{k \rightarrow \infty} \text{relMSE}(\eta_{r_0}, r) = \text{relMSE}(\eta_{r_0}^{1.\text{loc}}, r) \quad (2.54)$$

where $\eta_{r_0}^{1.\text{loc}}$ is given by (2.20) and (2.21).

2.5.2 Average Square Contamination Neighborhoods ($\ast = c$, $\alpha = 2$)

We consider the k -dimensional normal regression model 2.1.1 (c) 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 3.3 below is

$$\eta_r(x, u) = \alpha_r x u \min \{1, c_r |u|^{-1}\} \quad (2.55)$$

where

$$k = \alpha_r \mathbb{E} |x|^2 \cdot \mathbb{E} |u| \min \{|u|, c_r\}, \quad r^2 c_r = \mathbb{E} (|u| - c_r)_+ \quad (2.56)$$

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

$$\eta_\infty(x, u) = \mathcal{K}^{-1} \frac{x}{\mathbb{E} |u|} \text{sign}(u) \quad (2.57)$$

with $\mathcal{K} = \mathbb{E} x x' = \gamma \mathbb{I}_k$ for some $\gamma \in (0, \infty)$ (Lemma 3.2), where

$$\omega_{c,2}^{\min} = \frac{\sqrt{\text{tr } \mathcal{K}^{-1}}}{\mathbb{E} |u|} = \sqrt{\frac{\pi k}{2\gamma}} \quad (2.58)$$

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\max \text{MSE}(\eta_{r_0}, r) = \mathbb{E} |x|^2 \alpha_{r_0}^2 (\mathbb{E} \min\{u^2, c_{r_0}^2\} + r^2 c_{r_0}^2) \quad (2.59)$$

Comparing (2.20), (2.21) and (2.55), (2.56), we obtain the following relation to $\max \text{MSE}$ in the one-dimensional location model 2.2.2,

$$\max \text{MSE}(\eta_{r_0}, r) = \frac{k^2}{\mathbb{E} |x|^2} \max \text{MSE}(\eta_{r_0}^{1.\text{loc}}, r) \quad (2.60)$$

where $\eta_{r_0}^{1.\text{loc}}$ denotes the corresponding minimax IC for radius r_0 ; in fact, the constants in (2.20), (2.21) and (2.55), (2.56) are connected via

$$c_{r_0} = c_{r_0}^{1.\text{loc}}, \quad \alpha_{r_0} = \frac{k}{\mathbb{E} |x|^2} A_{r_0}^{1.\text{loc}} \quad (2.61)$$

Consequently, by relation (2.60), the MSE-inefficiencies coincide with those in one-dimensional location, independently of the regressor distribution $K(dx)$.

2.6 Regression, Average (Square) Hellinger Balls

2.6.1 Average Hellinger Neighborhoods ($\ast = h, \alpha = 1$)

We consider the k -dimensional normal regression model 2.1.1 (c) 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; Theorems 7.4.19 and 7.5.7), using spherical symmetry (Lemma 3.3) and $\mathbb{E} u^2 = 1$, is

$$\eta_r(x, u) = \alpha_r x u \min\{1, c_r |x|^{-1}\} \quad (2.62)$$

where

$$k = \alpha_r \mathbb{E} |x| \min\{|x|, c_r\}, \quad 8 r^2 c_r = \mathbb{E} (|x| - c_r)_+ \quad (2.63)$$

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

$$\eta_\infty(x, u) = \frac{1}{\sqrt{8}} \omega_{h,1}^{\min} \frac{x}{|x|} u \quad (2.64)$$

where

$$\omega_{h,1}^{\min} = \sqrt{8} \frac{k}{\mathbb{E} |x|} \quad (2.65)$$

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\max \text{MSE}(\eta_{r_0}, r) = \alpha_{r_0}^2 \mathbb{E} \min\{|x|^2, c_{r_0}^2\} + 8 r^2 \alpha_{r_0}^2 c_{r_0}^2 \quad (2.66)$$

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 2.4 reveals that (2.63), (2.66) agree with (2.43), (2.46) if x and u are exchanged. Consequentially, the $\text{maxMSE}(\eta_{r_0}, r)$ and the $\text{relMSE}(\eta_{r_0}, r)$ are the same in both models. In particular, the convergence result (2.48) is available for the present model ($* = h, \alpha = 1$) if $K = \mathcal{N}(0, \sigma^2 \mathbb{I}_k)$; but we prove (2.48) for model ($* = h, \alpha = 1$) also in case $K = \text{Ufo}_k(0, m)$.

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

We consider the k -dimensional normal regression model 2.1.1 (c) 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

$$\eta_r(x, u) = \mathcal{I}^{-1} \Lambda \quad (2.67)$$

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

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

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

We consider the k -dimensional normal regression model 2.1.1 (c) 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 3.3), is

$$\eta_r(x, u) = \alpha_r x u \min \{1, c_r(x) |u|^{-1}\} \quad (2.68)$$

where

$$k = \alpha_r \mathbb{E} |x|^2 \mathbb{E}_\cdot |u| \min \{|u|, c_r(x)\} \quad (2.69)$$

and

$$\mathbb{E}_\cdot (|u| - c_r(x))_+ = \frac{r^2}{|x|} \mathbb{E} |x| c_r(x) \quad (2.70)$$

with $c_r(x) = 0$, if the RHS in (2.70) is larger than $\mathbb{E}_\cdot |u| = \sqrt{2/\pi}$. As for (2.70) confer Lemma 3.1. By \mathbb{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

$$\omega_{c, \infty}^{\min} = 0 \quad \text{if } K = \mathcal{N}(0, \sigma^2 \mathbb{I}_k) \quad (2.71)$$

respectively

$$\omega_{c, \infty}^{\min} = \sqrt{\frac{\pi}{2}} k \quad \text{if } K = \text{Ufo}_k(0, m) \quad (2.72)$$

and, in both cases, cannot be attained.

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\alpha_{r_0}^2 \mathbb{E} |x|^2 \mathbb{E}_\cdot \min \{u^2, c_{r_0}^2(x)\} + r^2 \alpha_{r_0}^2 [\mathbb{E} |x| c_{r_0}(x)]^2 \quad (2.73)$$

In Subsection 3.6, for $k \rightarrow \infty$, we sketch an (incomplete) argument for the convergence (2.54) of the MSE-inefficiencies in the present model ($* = c, \alpha = \infty$) to the corresponding ones in the one-dimensional location model 2.2.2.

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

We consider the k -dimensional normal regression model 2.1.1 (c) 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 3.3 and $\mathbb{E}u^2 = 1$, is

$$\eta_r(x, u) = \alpha_r x u (1 - c_r |x|^{-1})_+ \quad (2.74)$$

where

$$k = \alpha_r \mathbb{E} |x| (|x| - c_r)_+ \quad (2.75)$$

and

$$c_r = 8 r^2 \mathbb{E} (|x| - c_r)_+ \quad (2.76)$$

As for (2.76), confer Lemma 3.1.

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

$$\omega_{h, \infty}^{\min} = 0 \quad \text{if } K = \mathcal{N}(0, \sigma^2 \mathbb{I}_k) \quad (2.77)$$

respectively

$$\omega_{h, \infty}^{\min} = \sqrt{8} k \quad \text{if } K = \text{Ufo}_k(0, m) \quad (2.78)$$

and, in both cases, cannot be attained.

For $r_0, r \in [0, \infty)$, the maximal risk (2.12) is

$$\max \text{MSE}(\eta_{r_0}, r) = \alpha_{r_0}^2 \mathbb{E} (|x| - c_{r_0})_+^2 + \frac{r^2}{8 r_0^4} \alpha_{r_0}^2 c_{r_0}^2 \quad (2.79)$$

In Subsection 3.6, the convergence (2.48) of the MSE-inefficiencies to one, as the dimension increases, is proved also for this model ($*$ = h , $\alpha = \infty$).

Remark 2.1 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 1.3. ///

3 Lemmas and Proofs

3.1 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 $\gamma(u) := u^2$.

Lemma 3.1 *Given a real vector space X , a convex subset A of X , consider three convex functions $f: A \rightarrow \mathbb{R}$, $g: A \rightarrow [0, \infty)$, and $\gamma: [0, \infty) \rightarrow [0, \infty)$; γ increasing. Let $\beta_0 \in [0, \infty)$. Suppose $z_0 \in A$ minimizes the Lagrangian $L_0 = f + \beta_0 \gamma \circ g$ over A . Assume that γ is differentiable at $g_0 = g(z_0)$, and put*

$$\beta_1 = \beta_0 \gamma'(g_0) \quad (3.1)$$

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

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

$$\partial L_0 = \partial f + \beta_0 \gamma'(g_0) \partial g = \partial L_1 \quad (3.2)$$

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

////

3.2 One-Dimensional Location

Proof of (2.19) To ψ_s for $s \in [0, 1)$, integration by parts applies so that

$$\int \psi'_s(u) \Phi(du) = \int u \psi_s(u) \Phi(du) \quad (3.3)$$

Therefore we can rewrite (2.16) as

$$\max \text{Var}(\psi_s, s) = \frac{(1-s) \int m_s^{-2} \psi_s^2(u) \Phi(du) + s}{(1-s)^2 [\int u m_s^{-1} \psi_s(u) \Phi(du)]^2} \quad (3.4)$$

Consequently, the Var-inefficiency of the median is

$$\text{relVar}(\psi_1, s) = \frac{\pi [\int u m_s^{-1} \psi_s(u) \Phi(du)]^2}{2 [(1-s) \int m_s^{-2} \psi_s^2(u) \Phi(du) + s]} \quad (3.5)$$

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

$$\lim_{s \rightarrow 1} \text{relVar}(\psi_1, s) = \frac{\pi}{2} (\mathbb{E}|u|)^2 = 1 \quad (3.6)$$

by the dominated convergence theorem.

////

Proof of (2.25)–(2.27) By (2.15) and (2.21), $c_r = m_s$, if r and s are related by (2.26): $r^2 = s/(1-s)$. Using (2.21) we can rewrite (2.24) as

$$\begin{aligned} \max \text{MSE}(\eta_{r_0}, r) &= \frac{\mathbb{E} \min\{u^2, c_{r_0}^2\} + r^2 c_{r_0}^2}{[\mathbb{E}|u| \min\{|u|, c_r\}]^2} \\ &= \frac{\mathbb{E} \min\{u^2, m_{s_0}^2\} + s m_{s_0}^2 / (1-s)}{[\mathbb{E}|u| \min\{|u|, m_s\}]^2} \\ &= \frac{(1-s) \mathbb{E} \psi_{s_0}^2 + s m_{s_0}^2}{(1-s) [\mathbb{E} \psi_{s_0}'^2]} \quad \text{by (3.3)} \\ &= (1-s) \max \text{Var}(\psi_{s_0}, s) \end{aligned}$$

which proves (2.25) implying (2.27).

////

3.3 Invariance Under Rescaling

As mentioned in Subsubsection 2.1.1, the inefficiency in models 2.1.1 (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

$$\eta_r(x, u) = A_r x [\Lambda_f(u) - \vartheta_r(x)] \min \left\{ 1, \frac{b_r}{|A_r x [\Lambda_f(u) - \vartheta_r(x)]|} \right\} \quad (3.7)$$

$$0 = E. [\Lambda_f(u) - \vartheta_r(x)] \min \left\{ 1, \frac{b_r}{|A_r x [\Lambda_f(u) - \vartheta_r(x)]|} \right\} \quad (3.8)$$

$$\mathbb{I}_k = A_r E x x' \Lambda_f^2(u) \min \left\{ 1, \frac{b_r}{|A_r x [\Lambda_f(u) - \vartheta_r(x)]|} \right\} \quad (3.9)$$

$$r^2 b_r = E (|A_r x [\Lambda_f(u) - \vartheta_r(x)] - b|)_+ \quad (3.10)$$

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

$$\tilde{\eta}_r(z, v) = \tilde{A}_r z [\Lambda_{\tilde{f}}(v) - \tilde{\vartheta}_r(z)] \min \left\{ 1, \frac{\tilde{b}_r}{|\tilde{A}_r z [\Lambda_{\tilde{f}}(v) - \tilde{\vartheta}_r(z)]|} \right\} \quad (3.11)$$

with

$$\tilde{A}_r = \frac{\sigma^2}{\tau^2} A_r, \quad \tilde{b}_r = \frac{\sigma}{\tau} b_r, \quad \tilde{\vartheta}_r(z) = \frac{1}{\sigma} \vartheta_r\left(\frac{z}{\tau}\right) \quad (3.12)$$

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

Then it is easy to verify conditions (3.8)–(3.10) for $\tilde{\eta}_r$ in the rescaled model, so $\tilde{\eta}_r$ is indeed the optimum IC there. Using the relations (3.12) we obtain

$$\max \text{MSE}(\tilde{\eta}_{r_0}, r) = \frac{\sigma^2}{\tau^2} \max \text{MSE}(\eta_{r_0}, r) \quad (3.13)$$

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

3.4 Spherical Symmetry

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

$$\mathcal{L}(G\Lambda) = \mathcal{L}(\Lambda) \quad (3.14)$$

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

$$G \mathcal{I} G' = E(G\Lambda)(G\Lambda)' = E \Lambda \Lambda' = \mathcal{I} \quad (3.15)$$

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 3.2 *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 there exist some $\alpha \in \mathbb{R}$ such that $A = \alpha \mathbb{I}_k$.*

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

The second application of Lemma 3.2 is to

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

Lemma 3.3 *Under assumption (3.14) 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

$$r^2 c_r = E(|xu| - c_r) \quad (3.16)$$

$$k = \alpha_r E|xu| \min\{|xu|, c_r\} \quad (3.17)$$

and put

$$\eta_r(x, u) = \alpha_r x u \min\{1, c_r |xu|^{-1}\} \quad (3.18)$$

As for all orthogonal $G \in \mathbb{R}^{k \times k}$: $GE\eta_r\Lambda'G' = E\eta_r\Lambda'$ by spherical symmetry of K , Lemma 3.2 tells us that $E\eta_r\Lambda' = \beta \mathbb{I}_k$. Passing to the trace, (3.17) yields that $\beta = 1$. Because of symmetry of the error distribution, $E\eta_r = 0$ a.e. $K(dx)$. Thus, with $b_r := \alpha_r c_r$, η_r in fact is an IC as in HR (1994; Theorem 7.4.13.b), which form is sufficient to minimax asymptotic MSE.

($* = c, \alpha = 2$) Since the median of Λ_f is unique, the minimax IC is given by equation (134) in HR (1994; 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 3.2 and $\mathcal{I}_f^{\text{loc}} = 1$). This gives (3.16)–(3.18).

($* = c, \alpha = \infty$) An argument as in the case ($* = c, \alpha = 1$) and in addition using Lemma 3.1 shows that η_r of form (2.68)–(2.70) is an IC of the form of HR (1994; Theorem 7.4.11.b) and satisfies condition (3.1) of Lemma 3.1 below. Thus, η_r is the (unique) minimax IC. ////

Minimum Bias

Proof of (2.45) and (2.52) According to HR (1994; Theorems 5.5.1.b and 7.4.13.c), minimum bias ω_c^{\min} in model 2.4 (k -dimensional location) and $\omega_{c,1}^{\min}$ in model 2.5.1 (k -dimensional normal regression) are, with $D = \mathbb{I}_k$,

$$\omega_c^{\min} = \max \left\{ \frac{\text{tr } A}{E|A\Lambda - a|} \mid a \in \mathbb{R}^k, A \in \mathbb{R}^{k \times k} \setminus \{0\} \right\} \quad (3.19)$$

respectively

$$\omega_{c,1}^{\min} = \max \left\{ \frac{\text{tr } A}{E|Ax|E|\Lambda_f - m|} \mid A \in \mathbb{R}^{k \times k} \setminus \{0\} \right\} \quad (3.20)$$

In our case, the median m of $\Lambda_f(u) = u$ under $F = \mathcal{N}(0, 1)$ is zero. Also in (3.19), 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 $\mathbb{E}|A\Lambda - a| = \mathbb{E}|A\Lambda + a|$.

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

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

$$\mathbb{E}|A\Lambda| = \mathbb{E}|U'AV\Lambda| = \mathbb{E}|V'A'U\Lambda| = \mathbb{E}|A'\Lambda| \quad (3.21)$$

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

$$\text{tr } GAG' = \text{tr } A, \quad \mathbb{E}|GAG'\Lambda| = \mathbb{E}|A\Lambda| \quad (3.22)$$

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 $\mathbb{E}|A\Lambda|$ among all such diagonal matrices of trace 1. Consider the Lagrangian $L: [0, \infty)^k \rightarrow \mathbb{R}$,

$$L(a) = \mathbb{E}|A\Lambda| - \lambda \text{tr } A \quad (3.23)$$

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

$$\lambda = \mathbb{E}\{|\Lambda|^{-1}\Lambda_i^2\} \quad (3.24)$$

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

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

$$\frac{(\alpha_i + \tau h)\Lambda_i^2}{|A_\tau \Lambda|} \leq \frac{2\alpha_i\Lambda_i^2}{\frac{1}{2}|\alpha_i\Lambda_i|} = 4|\Lambda_i| \in L_1(P) \quad (3.25)$$

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

$$\frac{\partial L}{\partial \alpha_i} = \mathbb{E}\left\{\frac{\alpha_i\Lambda_i^2}{|A\Lambda|}\right\} - \lambda \quad (3.26)$$

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

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

Proof of (2.71), (2.72) and (2.77), (2.78) The solutions are given in HR (1994; Theorems 7.4.11.c and 7.4.18.c), we only have to determine

$$\sigma_A = \inf_{K(dx)} |Ax|^{-1} = \frac{1}{\sup_{K(dx)} |Ax|} \quad (3.27)$$

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

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

$$\omega_{c,\infty}(\eta) = \mathbb{E} \sup_{\cdot} |\eta|, \quad \omega_{h,\infty}(\eta) = \sqrt{8} \mathbb{E} (\mathbb{E} |\eta|^2)^{1/2} \quad (3.28)$$

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

$$\sup_{K(dx)} |Ax|^2 = \sup_{|x| \leq 1} |Ax|^2 = \|A\|_{\text{op}}^2 \quad (3.29)$$

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

$$\|A\|_{\text{op}}^2 = \sup_{|x| \leq 1} |Ax|^2 = \sup_{|x| \leq 1} |G'AGx|^2 \quad (3.30)$$

for any orthogonal matrix G , it suffices to consider A diagonal (spectral decomposition). Thus, $\|A\|_{\text{op}}^2 = \max_{i=1,\dots,k} \alpha_i^2$, and consequentially $A_{\min} = \frac{1}{k} \mathbb{I}_k$ and $\|A_{\min}\|_{\text{op}} = \frac{1}{k}$, 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. ////

3.6 Increasing Dimension $k \rightarrow \infty$

Proof of (2.47) We have $\Lambda \sim \mathcal{N}(0, \mathbb{I}_k)$, so $\frac{1}{k} |\Lambda|^2 = \frac{1}{k} \sum_{i=1}^k \Lambda_i^2 \rightarrow \mathbb{E} \Lambda_1^2 = 1$ a.e., hence also $|\Lambda|/\sqrt{k} \rightarrow 1$ a.e., as $k \rightarrow \infty$ (SLLN). $(\frac{1}{k} |\Lambda|^2)$ is uniformly integrable (Vitali, $\mathbb{E} |\Lambda|^2 = k$). Because $|\Lambda|/\sqrt{k} \leq 1 + \frac{1}{k} |\Lambda|^2$, also $(|\Lambda|/\sqrt{k})$ is uniformly integrable. Consequentially, $\mathbb{E} |\Lambda| \sim \sqrt{k}$ and $\mathbb{E} |\eta_\infty|^2 = (k/\mathbb{E} |\Lambda|)^2 \sim k$. ////

Proof of (2.48) We first give the proof for the k -dimensional location model 2.4.

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 (2.43) reads

$$r^2 \tilde{c}_{k,r} = \mathbb{E} (|u|/\sqrt{k} - \tilde{c}_{k,r})_+ \quad (3.31)$$

In the case $r = 0$, we have $c_{k,0} = \infty$. Assume that $r > 0$, and suppose that $\tilde{c}_{k,r} \rightarrow \gamma \in [0, \infty]$ along some subsequence. Since the RHS in (3.31) is bounded, necessarily $\gamma < \infty$. Then the noted L_1 -convergence implies that

$$r^2 \gamma = \mathbb{E} (1 - \gamma)_+ = (1 - \gamma)_+ \quad (3.32)$$

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,

$$\lim_{k \rightarrow \infty} \tilde{c}_{k,r} = \frac{1}{1 + r^2} \quad (3.33)$$

The first equation of (2.43) reads

$$\alpha_{k,r}^{-1} = \mathbb{E} k^{-1} |u|^2 \min \left\{ 1, \frac{\tilde{c}_{k,r}}{|u|/\sqrt{k}} \right\} \quad (3.34)$$

In the case $r = 0$ we have $\alpha_{k,0} = 1$. Now let $r > 0$. Obviously, the integrands in (3.34) 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 $r = 0$,

$$\lim_{k \rightarrow \infty} \alpha_{k,r} = 1 + r^2 \quad (3.35)$$

The arguments leading to (3.33) and (3.35) 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

$$\liminf_{k \rightarrow \infty} \tilde{c}_{k,r_k} \geq 1 = \lim_{k \rightarrow \infty} \alpha_{k,r_k} \quad (3.36)$$

Therefore, (3.33) and (3.35) hold uniformly on bounded r -intervals. Now (3.33) and (3.35) entail convergence of the risk (with the previous r replaced by r_0),

$$\begin{aligned} k^{-1} \max \text{MSE}(\eta_{r_0}, r) &= \alpha_{k,r_0}^2 \mathbb{E} \min \{ k^{-1} |u|^2, \tilde{c}_{k,r_0}^2 \} + r^2 \alpha_{k,r_0}^2 \tilde{c}_{k,r_0}^2 \\ &\longrightarrow 1 + r^2 \end{aligned} \quad (3.37)$$

Consequently, and in addition using (2.47), we get

$$\frac{\max \text{MSE}(\eta_{r_0}, r)}{\max \text{MSE}(\eta_{\infty}, r)} = \frac{\mathbb{E} |\Lambda|^2}{k} \frac{\max \text{MSE}(\eta_{r_0}, r)}{k(1+r^2)} \longrightarrow 1 \quad (3.38)$$

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

By the coincidence mentioned in the k -dimensional regression model 2.6.1 for $* = h$ and $\alpha = 1$, the convergence (2.48) 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 (2.48) for the k -dimensional regression model 2.6.1 ($* = h, \alpha = 1$) with $K = \text{Ufo}_k(0, 1)$.

In this case, we have $\mathbb{E} |x| = k(k+1)^{-1}$ and $\mathbb{E} |x|^2 = k(k+2)^{-1}$ which implies that $\text{Var}(x) \rightarrow 0$ as $k \rightarrow \infty$; consequentially, $|x| \rightarrow 1$ in L_2 and L_1 . The second part of (2.63) reads

$$8r^2 c_{k,r} = \mathbb{E} (|x| - c_{k,r})_+ \quad (3.39)$$

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

$$8r^2 \gamma = \mathbb{E}(1 - \gamma)_+ = 1 - \gamma \quad (3.40)$$

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

$$\lim_{k \rightarrow \infty} c_{k,r} = \frac{1}{1 + 8r^2} \quad (3.41)$$

Plugged into the first equation of (2.63), this yields

$$\frac{k}{\alpha_{k,r}} = \mathbb{E}|x| \min\{|x|, c_{k,r}\} \longrightarrow \frac{1}{1+8r^2} \quad (3.42)$$

The arguments leading to (3.41) and (3.42) obtain if we replace r by a bounded sequence r_k . Thus (3.41) and (3.42) hold true uniformly on bounded r -intervals. The convergences (3.41) and (3.42) now entail convergence of the risk (with the previous r replaced by r_0),

$$\begin{aligned} k^{-2} \max \text{MSE}(\eta_{r_0}, r) &= k^{-2} \alpha_{k,r_0}^2 \mathbb{E} \min\{|x|^2, c_{k,r_0}^2\} + 8r^2 k^{-2} \alpha_{k,r_0}^2 c_{k,r_0}^2 \\ &\longrightarrow 1 + 8r^2 \end{aligned} \quad (3.43)$$

Consequently,

$$\frac{\max \text{MSE}(\eta_{r_0}, r)}{\max \text{MSE}(\eta_{\infty}, r)} = \mathbb{E}|x|^2 \frac{\max \text{MSE}(\eta_{r_0}, r)}{k^2(1+8r^2)} \longrightarrow 1 \quad (3.44)$$

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

In the third part of the proof, we shall show (2.48) for the k -dimensional regression model 2.7.2 ($\ast = h, \alpha = \infty$).

In case $K = \mathcal{N}(0, \mathbb{I}_k)$, the L_1 -convergence $|x|/\sqrt{k} \rightarrow 1$ inserted in equations (2.76) and (2.75) by previous arguments imply that

$$\lim_{k \rightarrow \infty} \frac{c_{k,r}}{\sqrt{k}} = \frac{8r^2}{1+8r^2}, \quad \lim_{k \rightarrow \infty} \alpha_{k,r} = 1+8r^2 \quad (3.45)$$

Consequently,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \max \text{MSE}(\eta_{r_0}, r) = 1+8r^2 \quad (3.46)$$

In case $K = \text{Ufo}_k(0, 1)$, the L_1 -convergence of $|x| \rightarrow 1$ inserted in equations (2.76) and (2.75) similarly imply that

$$\lim_{k \rightarrow \infty} c_{k,r} = \frac{8r^2}{1+8r^2}, \quad \lim_{k \rightarrow \infty} \frac{\alpha_{k,r}}{k} = 1+8r^2 \quad (3.47)$$

Consequently,

$$\lim_{k \rightarrow \infty} k^{-2} \max \text{MSE}(\eta_{r_0}, r) = 1+8r^2 \quad (3.48)$$

Both convergences (3.46) and (3.48) hold uniformly on bounded r_0 -, r -intervals (though convergence in the K normal case seems slow; Subsubsection 5.5.2). $////$

Proof of (2.54) The second equation of (2.43), in case $K = \mathcal{N}(0, \mathbb{I}_k)$, reads

$$r^2 \tilde{c}_{k,r} = \mathbb{E}(|u||x|/\sqrt{k} - \tilde{c}_{k,r})_+ \quad (3.49)$$

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

$$r^2 c_{k,r} = \mathbb{E}(|x||u| - c_{k,r})_+ \quad (3.50)$$

If $r = 0$, we have $c_{k,0} = \infty$. Now let $r > 0$, and suppose that, along some subsequence, $\tilde{c}_{k,r}$, respectively $c_{k,r}$, tend to some $\gamma \in [0, \infty]$. Since the RHS in (3.49) and (3.50) is bounded, necessarily $\gamma < \infty$. Then the noted L_1 -convergence in the proof of (2.48) implies that (in both cases)

$$r^2 \gamma = \mathbb{E} (|u| - \gamma)_+ \quad (3.51)$$

from which it follows that $\gamma = c_r^{1,\text{loc}}$ from (2.21). Therefore, respectively,

$$\lim_{k \rightarrow \infty} \tilde{c}_{k,r} = c_k^{1,\text{loc}} = \lim_{k \rightarrow \infty} c_{k,r} \quad (3.52)$$

The first equation of (2.43) for $K = \mathcal{N}(0, \mathbb{I}_k)$ reads

$$\alpha_{k,r}^{-1} = \mathbb{E} k^{-1} |x|^2 u^2 \min \left\{ 1, \frac{\tilde{c}_{k,r}}{|u||x|/\sqrt{k}} \right\} \quad (3.53)$$

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

$$\frac{k}{\alpha_{k,r}} = \mathbb{E} |x|^2 u^2 \min \left\{ 1, \frac{c_{k,r}}{|u||x|} \right\} \quad (3.54)$$

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 (3.53) and (3.54) are uniformly integrable (being dominated by $\frac{1}{k}|x|^2 u^2$, respectively by $|x|^2 u^2$). In the normal case, we get

$$\lim_{k \rightarrow \infty} \alpha_{k,r}^{-1} = \mathbb{E} u^2 \min \left\{ 1, \frac{c_r^{1,\text{loc}}}{|u|} \right\} = (A_r^{1,\text{loc}})^{-1} \quad (3.55)$$

with $A_r^{1,\text{loc}} = A_r$ from (2.43), which is consistent with $\alpha_{k,0} = 1 = A_0^{1,\text{loc}}$ in case $r = 0$. In the uniform regressor case, again consistently with $r = 0$,

$$\lim_{k \rightarrow \infty} \frac{k}{\alpha_{k,r}} = \mathbb{E} u^2 \min \left\{ 1, \frac{c_r^{1,\text{loc}}}{|u|} \right\} = (A_r^{1,\text{loc}})^{-1} \quad (3.56)$$

The arguments leading to (3.52) and (3.55), (3.56) 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 $\tilde{c}_{k,r}, c_{k,r} \rightarrow c_0^{1,\text{loc}}$, respectively. Thus (3.52) and (3.55), (3.56) hold true uniformly on bounded r -intervals. (3.52) and (3.55), (3.56) now entail convergence of the risk (with the previous r replaced by r_0),

$$\begin{aligned} k^{-1} \max \text{MSE}(\eta_{r_0}, r) &= \alpha_{k,r_0}^2 \mathbb{E} \min \{ k^{-1} |x|^2 u^2, \tilde{c}_{k,r_0}^2 \} + r^2 \alpha_{k,r_0}^2 \tilde{c}_{k,r_0}^2 \\ &\rightarrow \max \text{MSE}(\eta_{r_0}^{1,\text{loc}}, r) \end{aligned} \quad (3.57)$$

$$\begin{aligned} k^{-2} \max \text{MSE}(\eta_{r_0}, r) &= k^{-2} \alpha_{k,r_0}^2 \mathbb{E} \min \{ |x|^2 u^2, c_{k,r_0}^2 \} + r^2 k^{-2} \alpha_{k,r_0}^2 c_{k,r_0}^2 \\ &\rightarrow \max \text{MSE}(\eta_{r_0}^{1,\text{loc}}, r) \end{aligned} \quad (3.58)$$

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

$$\lim_{k \rightarrow \infty} \text{relMSE}(\eta_{r_0}, r) = \text{relMSE}(\eta_{r_0}^{1,\text{loc}}, r) \quad (3.59)$$

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

We shall sketch an argument for (3.57) and (3.58), hence (3.59), to hold also in model 2.7.1 ($\ast = c$, $\alpha = \infty$).

Employing the L_1 -convergence $|x|/\sqrt{k} \rightarrow 1$ and $|x| \rightarrow 1$ for $K = \mathcal{N}(0, \mathbb{I}_k)$ and $K = \text{Ufo}_k(0, 1)$, respectively, equation (2.70) determining $c_{k,r}(x)$ uniquely may be solved by $c_{k,r}(x) \rightarrow c_r^{1.\text{loc}}$ in probability, where $c_r^{1.\text{loc}}$ is taken from (2.21). At this instance, we assume but do not prove that the integrals $\mathbb{E}|x|c_{k,r}(x)/\sqrt{k}$ and $\mathbb{E}|x|c_{k,r}(x)$ converge correspondingly; that is, to $\mathbb{E}1 \cdot c_r^{1.\text{loc}} = c_r^{1.\text{loc}}$. Under this assumption, however, equation (2.69) now entails (3.55) and (3.56), respectively. Then (3.57) and (3.58) follow as before.

Due to variable $c_{k,r}(x)$ (and matching the gap in the proof), the tables in Subsubsection 5.5.1 indicate only slow convergence in (3.57)–(3.59), but in the K uniform case, convergence is confirmed. ////

4 Numerical Algorithms

We use S-Plus 2000 to implement the algorithms and to generate the graphical output. In detail we use the following numerical procedures:

4.1 One-Dimensional Location

The results for the models 2.2.1 and 2.2.2 are obtained by the routines for the k -dimensional location model 2.4 with $k = 1$. Note the coincidence (2.27) via the relation (2.26).

4.2 One-Dimensional Scale

4.2.1 r/\sqrt{n} -Contamination Balls (Model 2.3.1)

The clipping bound c_r and the centering constant α_r in (2.28)–(2.31) are calculated by bisection methods.

4.2.2 r/\sqrt{n} -Total Variation Balls (Model 2.3.2)

We evaluate the clipping constants g_r and c_r in (2.35)–(2.38) by bisection methods.

4.3 k -Dimensional Location (Model 2.4)

We compute the constants c_r and α_r in (2.42), (2.43) 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).

4.4 Regression, Average (Square) Contamination

4.4.1 Average Contamination Neighborhoods (Model 2.5.1)

The determination of clipping bound c_r in (2.49), (2.50) is performed by a bisection method, where the integration of the outer integral is done numerically. In case of the uniform distribution we use the Simpson rule, whereas in the normal case we make use of the S-plus function `integrate()` due to the more complicated integrands.

4.4.2 Average Square Contamination Neighborhoods (Model 2.5.2)

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

4.5 Regression, Average (Square) Hellinger

4.5.1 Average Hellinger Neighborhoods (Model 2.6.1)

For $K = \text{Ufo}_k(0, m)$, in view of (2.62), (2.63), we have to find the zero of

$$c_r^{k+1} - (1 + 8r^2)(k + 1)c_r + k \quad (4.1)$$

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 2.4 by substituting r_0 by $r_0/\sqrt{8}$ and r by $r/\sqrt{8}$.

4.5.2 Average Square Hellinger Neighborhoods (Model 2.6.2)

Nothing to calculate.

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

4.6.1 Contamination Neighborhoods (Model 2.7.1)

We introduce the further parameter

$$\tau := \frac{r^2}{2} \mathbb{E}|x|c_r(x) \quad (4.2)$$

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

$$2 \frac{\tau}{|x|} = \mathbb{E}(|u| - c_r(x)) \quad (4.3)$$

Then r may be easily calculated back from (4.2).

In case of the uniform distribution we evaluate $c_r(x)$ on a grid of x values by using a bisection method and do the outer integration on the same grid applying the Simpson rule. For normal regressors we use the S-Plus function `integrate()`, which also performs the evaluation of $c_r(x)$, where the computation of $c_r(x)$ by a bisection method is implemented as a vector valued function.

4.6.2 Hellinger Neighborhoods (Model 2.7.2)

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

$$c_r^{k+1} - (1 + \frac{1}{8}r^2)(k+1)c_r + k \quad (4.4)$$

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

Remark 4.1 In the case $\alpha = \infty$ (model 2.7), for both contamination and Hellinger neighborhoods, the MSE of η_{r_0} at $r = 0$ (ideal model) is unbounded as $r_0 \rightarrow \infty$.

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 $\text{relMSE}(\eta_{r_0}, r) \rightarrow \infty$ as $r \rightarrow \infty$, 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). ////

4.7 General Procedures

In all these cases, we use bisection methods to calculate the least favorable radius r_0 for a given MSE-inefficiency $\text{relMSE}(\eta_{r_0}, r)$ at $r = 0$, and to determine the least favorable radius r_ρ for a given interval $[\rho r, r/\rho]$, where $\rho \in [0, 1]$ (in particular, $\rho = 0, \frac{1}{3}, \frac{1}{2}$) and $r \in [0, \infty]$. To find the least favorable radii r_2 and r_3 , we use the preceding procedure and search for the maximum on a increasingly finer grid of r values.

4.8 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.

5 Tabulated Results

5.1 k -Dimensional Location

maxMSE over r/\sqrt{n} -contamination neighborhoods ($k \geq 1$) and, in case $k = 1$,
maxVar 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

5.2 One-Dimensional Scale

5.2.1 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	1.207	1.099	0.499	0.481	0.551

5.2.2 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

5.3 Regression, Average (Square) Contamination

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

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

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

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

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

5.3.2 Average Square Contamination Neighborhoods ($\ast = c, \alpha = 2$)

Same numbers as in one-dimensional location.

5.4 Regression, Average (Square) Hellinger

5.4.1 Average Hellinger Neighborhoods ($\ast = h, \alpha = 1$)

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

k	$\text{relMSE}(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	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

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

Same numbers as in k -dimensional location, but $r_\rho = r_\rho^{\text{k.loc}}/\sqrt{8}$.

k	$\text{relMSE}(\eta_\infty, 0)$	$\rho = 0$	$\rho = \frac{1}{3}$	$\rho = \frac{1}{2}$	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

5.4.2 Average Square Hellinger Neighborhoods ($\ast = h, \alpha = 2$)

$\text{relMSE} \equiv 1$.

5.5 Constant Conditional Neighborhoods

5.5.1 Constant Contamination Neighborhoods ($\ast = c, \alpha = \infty$)

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

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

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

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

5.5.2 Constant Hellinger Neighborhoods ($\ast = h$, $\alpha = \infty$)

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

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

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

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

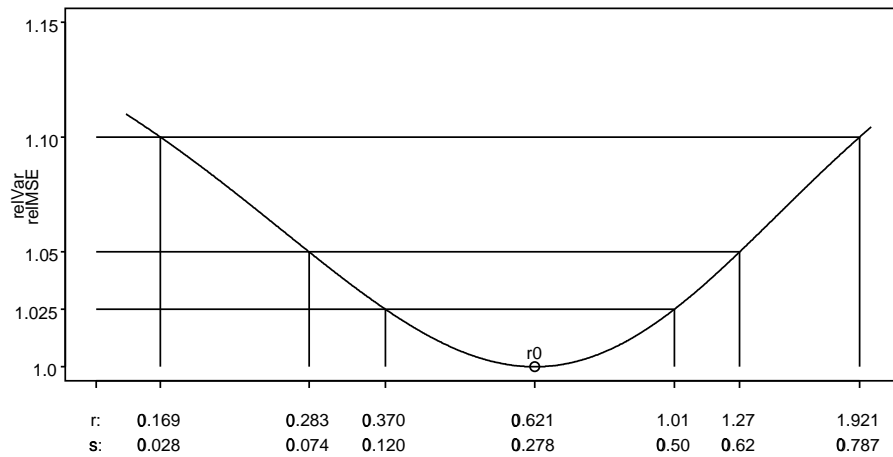
References

- [1] 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.
- [2] Beran, R.J. (1976): Adaptive estimates for autoregressive processes. *Ann. Inst. Statist. Math.* **11** 432–452
- [3] Beran, R.J. (1981): Efficient robust estimates in parametric models. *Z. Wahrsch. verw. Gebiete* **55** 91–108.
- [4] 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.
- [5] Bickel, P.J. (1982): On adaptive estimation. *Ann. Statist.* **10** 647–671.
- [6] Bickel, P.J. (1984): Robust regression based on infinitesimal neighborhoods. *Ann. Statist.* **12** 1349–1368.
- [7] 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.
- [8] Huber, P.J. (1964): Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.
- [9] Huber, P.J. (1981): *Robust Statistics*. Wiley, New York.
- [10] Huber, P.J. (1983): Minimax aspects of bounded influence regression. *J. Amer. Statist. Assoc.* **78** 66–80.
- [11] Huber, P.J. (1996): *Robust Statistical Procedures* (2nd ed.). CBMS-NSF Regional Conference Series in Applied Mathematics # 68. Soc. Industr. Appl. Math., Philadelphia, Pennsylvania.
- [12] Klaassen, C. (1980): Statistical performance of location estimators. Mathematical Centre Tract #133. Mathematisch Centrum, Amsterdam.
- [13] Kreiss, J.-P. (1987): On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.
- [14] Millar, P.W. (1981): Robust estimation via minimum distance methods. *Z. Wahrsch. verw. Gebiete* **55** 73–89.
- [15] Pfanzagl, J. and Wefelmeyer, W. (1982): *Contributions to a General Asymptotic Statistical Theory*. Lecture Notes in Statistics #13. Springer, Berlin.
- [16] Rieder, H. (1989): A finite-sample minimax regression estimator. *Statistics* **20** 211–221.

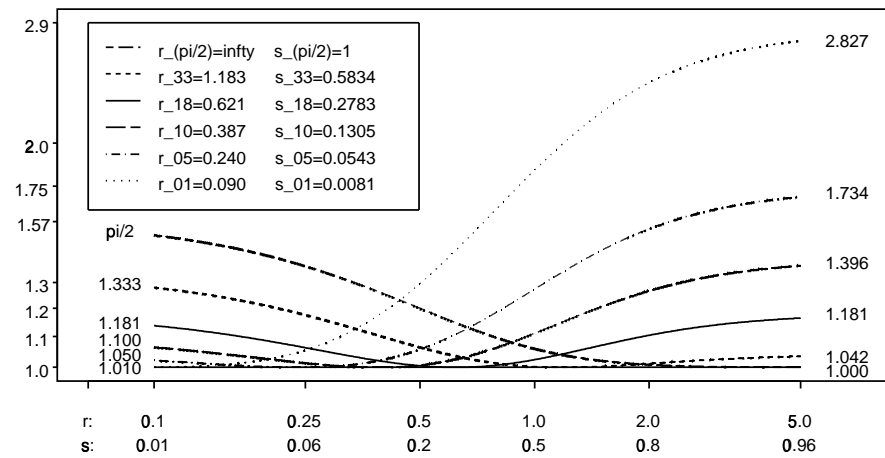
- [17] Rieder, H. (1994): *Robust Asymptotic Statistics*. Springer, New York.
- [18] Rieder, H. (2000): Neighborhoods as nuisance parameters? Robustness vs. semiparametrics. Submitted to Ann. Statist.
- [19] Rieder, H. (2001a): Robust time series based on infinitesimal neighborhoods. Manuscript.
- [20] Rieder, H. (2001b): Nonuniform convergence of adaptive estimators in linear time series models. Manuscript.
- [21] Ruckdeschel, P. (2001): *Ansätze zur Robustifizierung des Kalman-Filters*. Dissertation, Universität Bayreuth.
- [22] 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

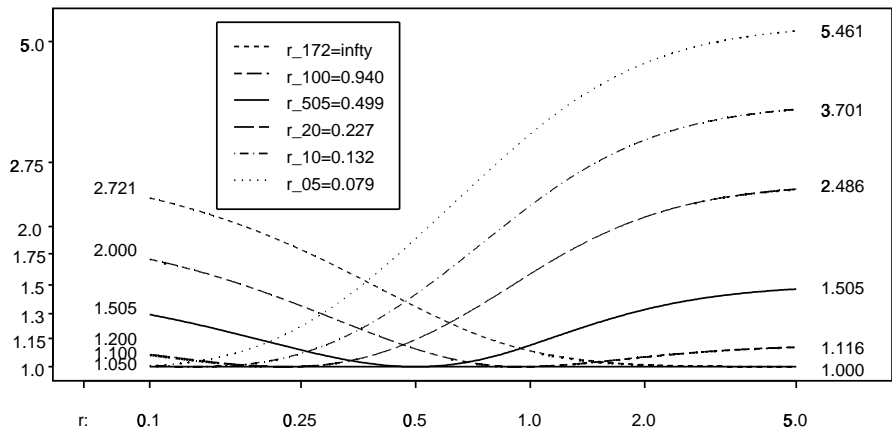
relVar, relMSE: 1-Dimensional Location (Var = 1.181 at $r, s = 0$)



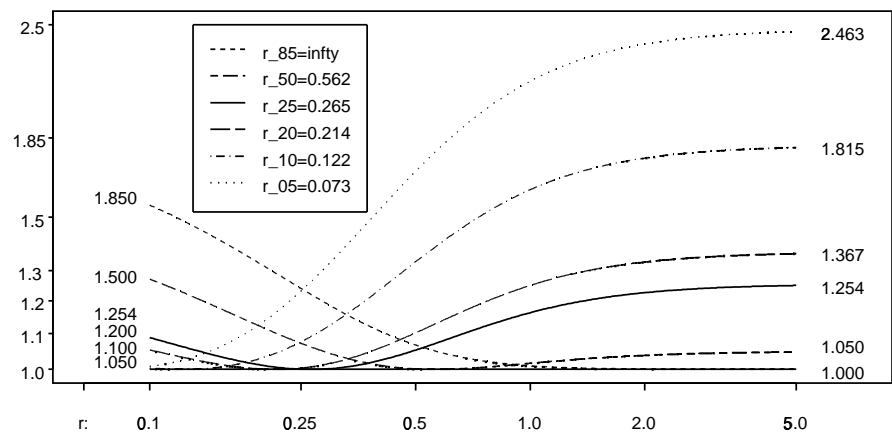
1-Dimensional Location: relMSE, relVar vs. r , resp. $s=r^2/(1+r^2)$

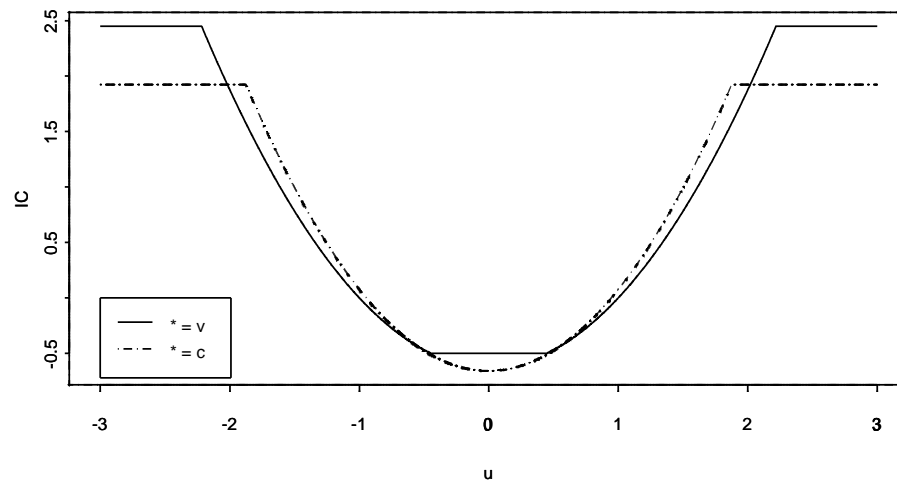
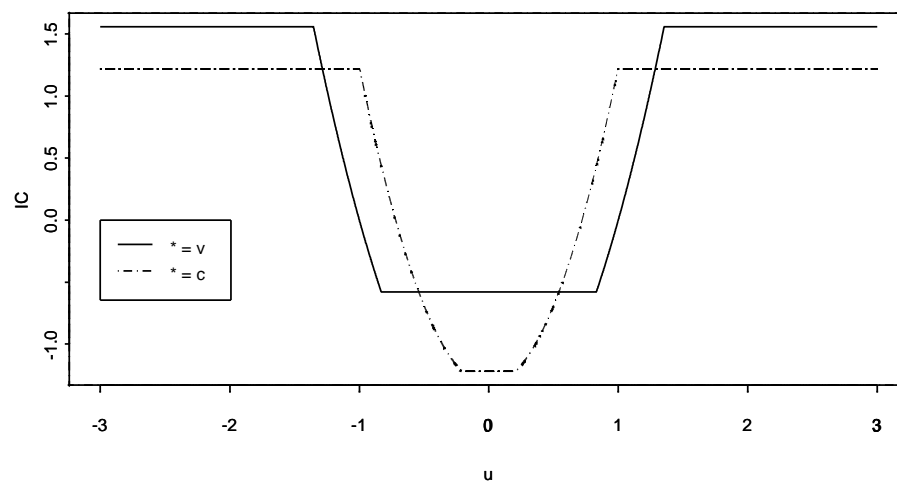


reIMSE: 1-Dimensional Scale (contamination)

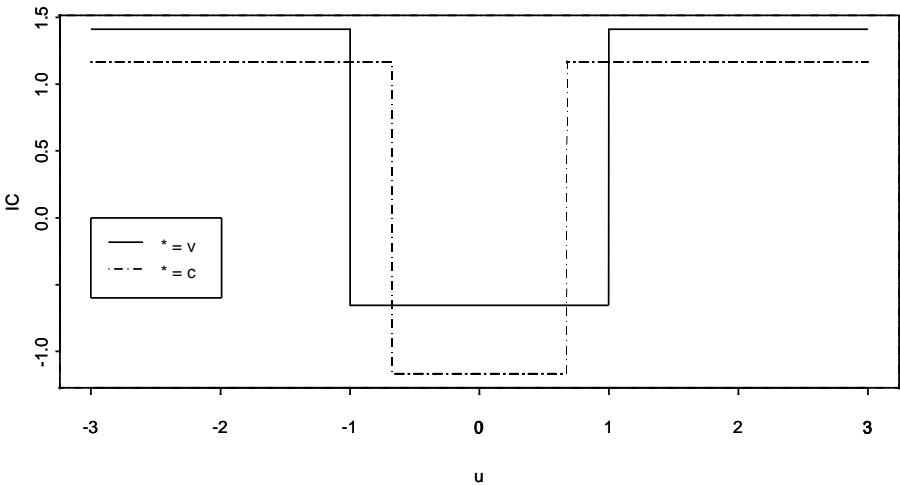


reIMSE: 1-Dimensional Scale (total variation)

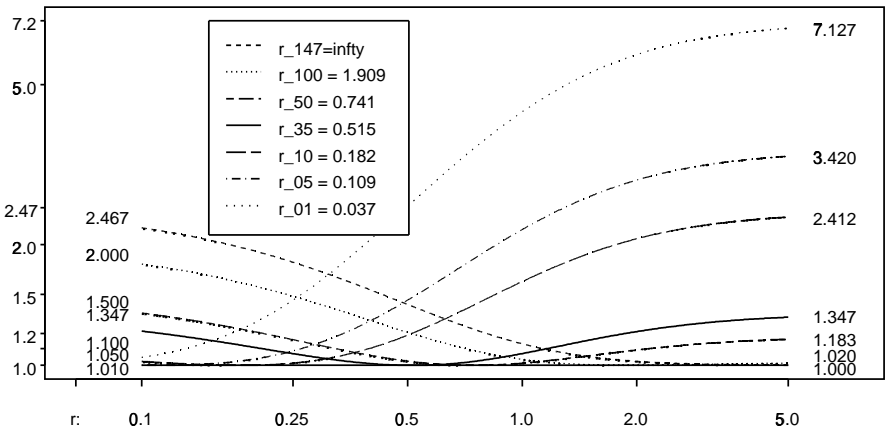


1-Dimensional Scale: IC-comparison for $r=0.2$ ($*$ =c) with $r=0.1$ ($*$ =v)1-Dimensional Scale: IC-comparison for $r=1.0$ ($*$ =c) with $r=0.5$ ($*$ =v)

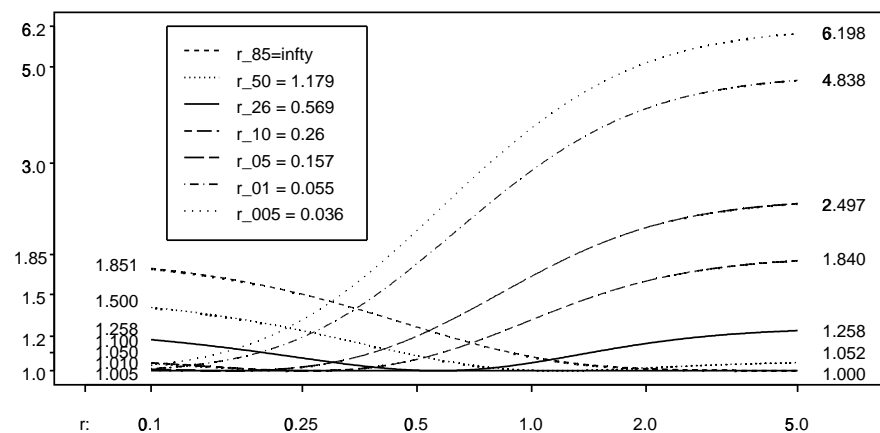
1-Dimensional Scale: most robust IC (*=v,c)



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)

