J. W. LEE
(Kumoh Institute of Technology, Kumi, South Korea)
V. I. SHIN
(Gwangju Institute of Science and Technology, Gwangju, South Korea)
G. L. SHEVLYAKOV
(Gwangju Institute of Science and Technology, Gwangju, South Korea)

ROBUST ESTIMATION OF A CORRELATION COEFFICIENT FOR ε-CONTAMINATED BIVARIATE NORMAL DISTRIBUTIONS

Abstract

Robust estimators of a correlation coefficient based on: (i) direct robust counterparts of the sample correlation coefficient, (ii) nonparametric measures of correlation, (iii) robust regression, (iv) robust estimation of the variances of principal variables, (v) stable parameter estimation, and (vi) the preliminary rejection of outliers from the data with the subsequent application of the sample correlation coefficient to the rest of the observations, are considered. Their performance in ε-contaminated normal models is examined both on small and large samples, and the best of the proposed robust estimators are revealed.

[1]

1 Introduction

1.1. Preliminaries. The aim of robust methods is to ensure high stability of statistical inference under the deviations from the assumed distribution model. Far less attention is devoted in the literature to robust estimators of association and correlation as compared to robust estimators of location and scale [1, 2]. However, it is necessary to study these problems due to their widespread occurrence (estimation of the correlation and covariance matrices in regression and multivariate analysis, estimation of the correlation functions of stochastic processes, etc.), and also due to the great instability of classical methods of estimation in the presence of outliers in the data.

The simplest problem of correlation analysis is estimation of the correlation coefficient ρ between the random variables X and Y defined as

$$\rho = \mathrm{Cov}(X,Y)/\left[\mathrm{D}(X)\ \mathrm{D}(Y)\right]^{1/2}, \tag{1}$$

where $\mathrm{D}(X)$, $\mathrm{D}(Y)$ and $\mathrm{Cov}(X,Y)$ are the variances and the covariance of the r.v.'s X and Y, respectively. This definition of the correlation coefficient as the standardized covariance is only one of the possible definitions (their number is about a

[1]This work was supported by the Kumoh Institute of Technology Research Funds

dozen) [3, 4]. In the sequel, we use some of those definitions, for example, such as the standardized slope of a regression line and the geometric mean of two regression lines to construct new robust measures of correlation. Further, the correlation coefficient itself is not a unique measure of interdependence (association): there are other measures, e.g., the Spearman rank correlation [5], the quadrant (sign) correlation [6], and the Kendall's τ-correlation [7]. These measures are similar to the correlation coefficient ρ and the first two of them are used for its estimation in our paper. Thus, in general, the problem of estimation of a correlation coefficient is more complicated than, say, the problem of estimation of the center of a symmetric distribution in which at least it is clear what we really estimate.

Given the observed sample $(x_1, y_1), \ldots, (x_n, y_n)$ of a bivariate random variable (r.v.) (X, Y), the classical estimator of a correlation coefficient ρ is given by the sample correlation coefficient

$$r = \sum (x_i - \bar{x})(y_i - \bar{y}) \Big/ \left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right]^{1/2}, \tag{2}$$

where $\bar{x} = n^{-1} \sum x_i$ and $\bar{y} = n^{-1} \sum y_i$ are the sample means.

On the one hand, the sample correlation coefficient r is a statistical counterpart of the correlation coefficient ρ (1). On the other hand, it is the maximum likelihood estimator of ρ for a bivariate normal distribution density

$$\mathcal{N}(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \right.$$

$$\left. \times \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}, \tag{3}$$

where the parameters μ_1 and μ_2 are the means, σ_1 and σ_2 are the standard deviations of the r.v.'s X and Y, respectively.

To illustrate the necessity in robust counterparts of the sample correlation coefficient, consider the Tukey's gross error model [8] described by the mixture of normal densities ($0 \le \varepsilon < 0.5$)

$$f(x, y) = (1-\varepsilon)\mathcal{N}(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) + \varepsilon\mathcal{N}(x, y; \mu_1', \mu_2', \sigma_1', \sigma_2', \rho'), \tag{4}$$

where the first and the second summands generate "good" and "bad" data, respectively. In general, the characteristics of "bad" data, namely their component means $\mu_1', \mu_2',$, standard deviations σ_1', σ_2' and especially the correlation ρ' may significantly differ from their counterparts from the first summand.

Further, we are mostly interested in estimation of the correlation coefficient ρ of "good" data regarding "bad" data as the outliers. In model (4), the sample correlation coefficient is strongly biased with regard to the estimated parameter ρ, i.e., for any positive $\varepsilon > 0$ there exists $k = \sigma_1'/\sigma_1 = \sigma_2'/\sigma_2$ sufficiently large such that $\mathrm{E}(r)$ can be made arbitrarily close to ρ' [9, 10]. For instance, estimating the correlation coefficient $\rho = 0.9$ of the main bulk of the data under the contamination with

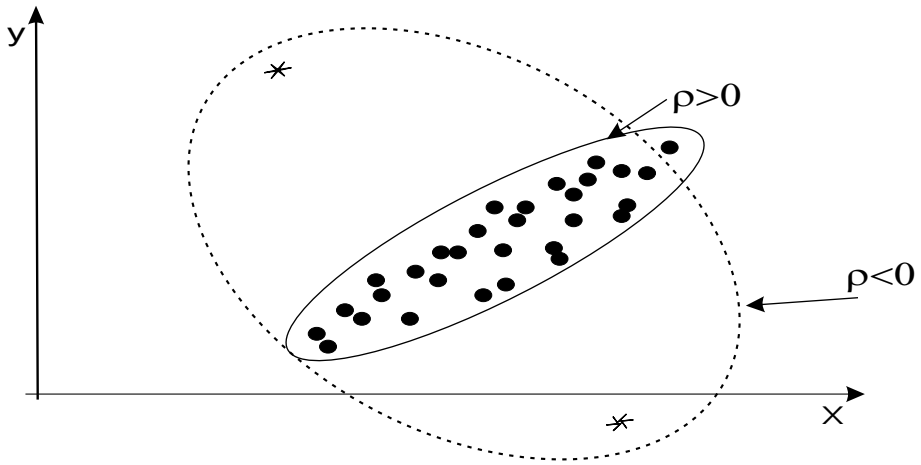

Figure 1: "Good" data (•) and outliers (∗) with their impact on correlation

$\varepsilon = 0.1$, $k = 3$ and $\rho' = -0.99$, asymptotically (as $n \to \infty$) we have $\mathrm{E}(r) = -0.055$, what means that even the sign of the sample correlation coefficient is wrong. So, the presence of even one or two outliers in the data can completely destroy the sample correlation coefficient up to the change of its sign, as it can be seen from Fig. 1. This effect is quite natural for the sample correlation coefficient, since it estimates the correlation coefficient of the whole distribution (4), but not the correlation of the "good" data. Anyway, the sample correlation coefficient is extremely sensitive to the presence of gross errors in the data, and hence it is necessary to use its robust counterparts. Now we specify what is understood under the term "a robust estimator" of the correlation coefficient.

1.2. General remarks on robustness. The field of mathematical statistics called robust statistics appeared due to the pioneer works of J. W. Tukey, P. J. Huber, and F. R. Hampel [8, 11, 12]; it has been intensively developed since 1960 and is rather definitely formed by present. The term "robust" (strong, sturdy) as applied to statistical procedures was proposed by G. E. P. Box [13].

Robustness deals with the consequences of possible deviations from the assumed statistical model and suggests the methods protecting statistical procedures against such deviations. Thus, statistical models used in robust statistics are chosen so that to account possible violations of the assumptions about the underlying distribution. For description of these violations, several neighborhoods of the underlying model based on an appropriately chosen metric are used, for example, the Kolmogorov, Prokhorov, or Lévy [1, 2, 14, 15]. Hence the initial model (basic or ideal) is enlarged up to the so-called *supermodel* that describes both the ideal model and the deviations from it.

Introducing a robust procedure, it is useful to answer the following questions: 1) Robustness of what? 2) Robustness against what? 3) Robustness in what sense?

The first answer defines the type of a statistical procedure (point or interval estimation, hypotheses testing, etc.); the second specifies the supermodel, regarding

robustness against the extension of ideal models to supermodels; and the third introduces the criterion of quality of a statistical procedure. Numerous problem settings observed in robust statistics arise due to the fact that there exist a lot of answers to each of those questions [1], [2], [16] – [20].

In this paper, we consider: 1) the point estimation of the correlation coefficient, 2) Tukey's gross error supermodel, and 3) various criteria of quality of robust estimation, items 2) and 3) further to be specified.

At present there exist two principal methods of designing robust estimators, i.e., Huber's minimax method of quantitative robustness [1, 11], and Hampel's method of qualitative robustness based on influence functions [2, 21, 22]. According to the first of these methods, we determine the least informative (favorable) distribution density minimizing Fisher information over a given class of distributions, with the subsequent construction of the maximum likelihood estimator for this density. This ensures that the asymptotic variance of an estimator will not exceed a certain threshold (namely, the supremum of the asymptotic variance as a measure of quantitative robustness) which strongly depends on the characteristics of a chosen class of distributions.

According to the second method, we construct an estimator with the assigned influence function whose type of behavior determines the qualitative robustness properties of an estimation procedure (such as its sensitivity to large outliers in the data, their rounding off, etc.). Most of robust estimators of a correlation coefficient have been obtained from heuristic considerations partially related to the desired behavior of their influence functions [9, 10, 35, 36]. To design robust estimators of correlation, in Section 2.6 we use the so-called *variational optimization approach* to stable parameter estimation [23, 24], which may be considered as a version of Hampel's method based on the change-to-variance function [2, 25]: the robust estimators designed according to that approach belong to Meshalkin's redescending exponentially weighted λ-estimators [26, 27].

1.3. Tukey's supermodel and criteria of quality of robust estimation. In our study, the observations are generated by Tukey's gross error supermodel (4). It is chosen due to the following reasons.

First, it is a standard model of the neighborhood of a normal distribution widely used in studies on robustness [1, 2], easily interpreted, since the distributions of the main bulk of data and outliers (gross errors) can be described by the the first and second terms of (4), respectively, allowing to consider different types of outliers, say, caused by the shift in distribution location, scale, or shape (correlation).

Second, the contamination parameter ε can be regarded as the probability of occurrence of outliers in the data sample with the number of outliers distributed according to the binomial law.

Third, it is the simplest neighborhood of a normal distribution being easy for modelling as compared to the other possible ε-neighborhoods of the normal distribution, say, the Kolmogorov, Prokhorov, or Lévy distances [1, 2, 14, 15].

Finally, we use the bivariate normal distribution in the first term of (4) as a conventional bivariate distribution [1, 9, 10]. Here we may also add that the nor-

mal (Gaussian) distribution shape with its properties of gravity and stability naturally arises in the description of smooth evolutionary processes with increasing entropy [28] – [31]. Thus, the normal distribution can be regarded as a good candidate for description of the main bulk of data. The second term in (4) for contamination should not necessarily be of a normal shape, but in case of its large variances, formula (4) yields a reasonable model for non-Gaussian distributions.

A survey on the applicability of Tukey's model to description of the real-life data, in particular on the frequency of gross errors, is given in [2]. It is concluded by the following words: "$1-10\%$ gross errors in routine data seem to be more the rule than the exception." ([2], pp. 25-28).

In the sequel, we use the simplified version of (4) in the following form

$$f(x,y)=(1-\varepsilon)\mathcal{N}(x,y;0,0,1,1,\rho)+\varepsilon\mathcal{N}(x,y;0,0,k,k,\rho'), \tag{5}$$

where $0\le\varepsilon<0.5,\ k>1,\ \mathrm{sgn}(\rho')=-\mathrm{sgn}(\rho)$: for the sample correlation coefficient, this choice leads to its maximum bias with respect to ρ.

Numerous applications, e.g., in communication queueing systems [37, 38] and in image processing [39, 40], show that the practice requires robust estimation of the correlations for long range dependent and essentially non-Gaussian heavy-tailed distributions, say, the Student t-distributions characterized by the low values of tail indices [41]. Robust estimators of correlation under the heavy-tailed Cauchy contamination of a bivariate normal distribution were studied in [9, 10] and it was shown that good robust estimators have qualitatively similar performances in model (5) and in models with essentially heavy-tailed contamination. The point is that, in heavy-tailed models, the maximum likelihood estimators presume some trimming of "tail" observations regardless of whether the tails are relatively or absolutely heavy, e.g., in case of estimation of location for the Tukey' mixture of normal distributions and for the Cauchy distribution: these effects were observed in Princeton's experimental study of robust estimators of location [42]. Moreover, outliers themselves can be well-defined only relatively, with respect to a given structure of the main bulk of the data [43, 44], say, for uniformly distributed data, observations from a normal distribution can be regarded as outliers and they also should be trimmed.

Thus, in this paper we use Tukey's gross error supermodel (5). Nevertheless, we underline that the problem of robust estimation of correlation for essentially heavy-tailed distributions, say, for the bivariate Student t-distribution of the data, requires a separate and thorough study.

Now we dwell on the criteria of quality of robust estimators. Those criteria are inherent to each of the aforementioned principal approaches in robustness: in Huber's approach, it is the supremum of the asymptotic variance of an estimator over a chosen class of distributions, and it is the influence and change-of-variance functions with their characteristics of estimator's sensitivity such as the breakdown and rejection points, local-shift sensitivity, etc. in Hampel's approach [2].

In what follows, we use those criteria whenever they are available and appropriate, but all the comparative study of various robust estimators of the correlation coefficient ρ in (5) is performed using the following two conventional characteristics: the bias $\mathrm{E}(\widehat{\rho})-\rho$ and the variance $\mathrm{D}(\widehat{\rho})$, asymptotic or sample.

1.4. The goals of the study. This paper pursues two main goals: first, we give an overview of various approaches to robust estimation of correlation; second, we present a comparative study of the performance of a selected subset of robust estimators generated by those approaches in ε-contaminated normal distribution models. Thus, it is mostly a survey and partially a contribution. The contribution mainly refers to the specification of the area of applicability of the minimax variance estimator of correlation in Section 2.5, to the study of a radical stable estimator of correlation in Section 2.6 and Section 3, and to the comparative analysis of selected estimators in Section 4.

The paper is organized as follows. In Section 2, we describe robust estimators of a correlation coefficient based on: direct robust counterparts of the sample correlation coefficient, nonparametric measures of correlation; robust regression, robust estimation of the variances of principal components, minimax approach, stable parameter estimation, and the preliminary rejection of outliers from the data with the subsequent application of the sample correlation coefficient to the rest of the observations. In Section 3, the performance of the typical representatives of those groups on small and large samples is studied. In Section 4, conclusions are made.

2 Robust estimators of a correlation coefficient

2.1. Robust correlation via direct robust counterparts of the sample correlation coefficient. A natural approach to robustifying the sample correlation coefficient is to replace the linear procedures of averaging by the corresponding nonlinear robust counterparts [1, 9, 10]

$$r_\alpha(\psi) = \Sigma_\alpha \psi(x_i - \widehat{x})\psi(y_i - \widehat{y}) / \left[\Sigma_\alpha \psi^2(x_i - \widehat{x}) \Sigma_\alpha \psi^2(y_i - \widehat{y})\right]^{1/2}, \tag{6}$$

where $\widehat{x}$ and $\widehat{y}$ are some robust estimators of location, for example, the sample medians $\operatorname{med} x$ and $\operatorname{med} y$; $\psi = \psi(z)$ is a monotone function, for instance, Huber's ψ-function: $\psi(z,k) = \max\left[-k, \min(z,k)\right]$; Σ_α is a robust analog of a sum.

The latter transformation is based on trimming the outer order statistics with subsequent summation of the remaining ones:

$$\Sigma_\alpha z_i = nT_\alpha(z) = n(n-2r)^{-1} \sum_{i=r+1}^{n-r} z_{(i)}, \quad 0 \le \alpha \le 0.5, \quad r = [\alpha\,(n-1)]$$

where $[\cdot]$ stands for the integer part. For $\alpha = 0$, the operations of ordinary and of robust summation coincide: $\Sigma_0 = \Sigma$.

The following version of estimator (6)

$$r_\alpha = \Sigma_\alpha (x_i - \operatorname{med} x)(y_i - \operatorname{med} y) / \left[\Sigma_\alpha (x_i - \operatorname{med} x)^2 \Sigma_\alpha (y_i - \operatorname{med} y)^2\right]^{1/2}$$

with $\alpha = 0.1, 0.2$ were used in [9, 10, 20]. For $\alpha = 0.5$, $\widehat{x} = \operatorname{med} x$, $\widehat{y} = \operatorname{med} y$, $\psi(z) = z$, formula (6) yields the correlation median estimator [35, 45]

$$r_{0.5} = r_{\mathrm{COMED}} = \frac{\operatorname{med}\{(x_1 - \operatorname{med} x)(y_1 - \operatorname{med} y), \ldots, (x_n - \operatorname{med} x)(y_n - \operatorname{med} y)\}}{\operatorname{MAD} x \operatorname{MAD} y},$$

where $\mathrm{MAD}\, z = \mathrm{med}\{|z_1 - \mathrm{med}\, z|, \ldots, |z_n - \mathrm{med}\, z|\}$ stands for the median absolute deviation.

2.2. Robust correlation via nonparametric measures. An estimation procedure can be endowed with robustness properties by the use of rank statistics. The best known of them are the quadrant (sign) correlation coefficient [6]

$$r_{\mathrm{Q}} = n^{-1} \sum \mathrm{sgn}(x_i - \mathrm{med}\, x)\mathrm{sgn}(y_i - \mathrm{med}\, y), \tag{7}$$

that is the sample correlation coefficient between the signs of deviations from medians, and the Spearman rank correlation coefficient [5]

$$r_{\mathrm{S}} = \frac{\sum [R(x_i) - \overline{R}(x)][R(y_i) - \overline{R}(y)]}{\left(\sum [R(x_i - \overline{R}(x)]^2 \sum [R(y_i) - \overline{R}(y)]^2\right)^{1/2}}, \tag{8}$$

that is the sample correlation coefficient between the observation ranks $R(x_i)$ and $R(y_i)$, where $\overline{R}(x)$ and $\overline{R}(y)$ stand for the average ranks, here equal to $n(n+1)/2$.

For computing, it is more convenient to use the transformed version of (8) [46]

$$r_S = 1 - 6\, S(d^2)/[n^3 - n], \qquad S(d^2) = \sum [R(x_i) - R(y_i)]^2.$$

2.3. Robust correlation via robust regression. The problem of estimation of the correlation coefficient is directly related to the linear regression problem of fitting the straight line of the conditional expectation [46]

$$\mathrm{E}(X \mid Y = y) = \mu_1 + \beta_1(y - \mu_2), \quad \mathrm{E}(Y \mid X = x) = \mu_2 + \beta_2(x - \mu_1).$$

For the bivariate normal distribution (3),

$$\beta_1 = \rho\sigma_1/\sigma_2, \quad \beta_2 = \rho\sigma_2/\sigma_1 \tag{9}$$

where σ_1 and σ_2 are the standard deviations of the r.v.'s X and Y, respectively [46]. Basing on (9), we propose the following robust estimator

$$r_{\mathrm{REG}} = \widehat{\beta}_1\, \widehat{\sigma}_2/\widehat{\sigma}_1, \tag{10}$$

where $\widehat{\beta}_1 = \mathrm{med}\,\{(y - \mathrm{med}\, y)/(x - \mathrm{med}\, x)\}$ is a robust estimate of slope, $\widehat{\sigma}_1$ and $\widehat{\sigma}_2$ are robust estimators of scale, namely, the median absolute deviations [35].

If the coefficients of linear regression are estimated by the least squares method

$$(\widehat{\alpha}_1, \widehat{\beta}_1) = \arg\min_{\alpha_1, \beta_1} \sum (x_i - \alpha_1 - \beta_1 y_i)^2, \quad (\widehat{\alpha}_2, \widehat{\beta}_2) = \arg\min_{\alpha_2, \beta_2} \sum (y_i - \alpha_2 - \beta_2 x_i)^2,$$

then, for the sample correlation coefficient, we get $r^2 = \widehat{\beta}_1 \widehat{\beta}_2$ [46], and basing on this dependence, we may propose a robust estimator of correlation in the form

$$\widehat{\rho}^{\,2} = \widetilde{\beta}_1 \widetilde{\beta}_2 \quad \text{or} \quad \widehat{\rho} = \sqrt{\widetilde{\beta}_1 \widetilde{\beta}_2}, \tag{11}$$

where $\widetilde{\beta}_1$ and $\widetilde{\beta}_2$ are some robust estimators of the slope [35]. For instance, we may use the least absolute values (LAV) estimators [47]

$$(\widetilde{\alpha}_1, \widetilde{\beta}_1) = \arg\min_{\alpha_1,\beta_1} \sum |x_i - \alpha_1 - \beta_1 y_i|, \quad (\widetilde{\alpha}_2, \widetilde{\beta}_2) = \arg\min_{\alpha_2,\beta_2} \sum |y_i - \alpha_2 - \beta_2 x_i|$$

or the least median squares (LMS) estimators of regression coefficients [2, 48]

$$(\widetilde{\alpha}_1, \widetilde{\beta}_1) = \arg\min_{\alpha_1,\beta_1} \operatorname{med}(x_i - \alpha_1 - \beta_1 y_i)^2, \quad (\widetilde{\alpha}_2, \widetilde{\beta}_2) = \arg\min_{\alpha_2,\beta_2} \operatorname{med}(y_i - \alpha_2 - \beta_2 x_i)^2.$$

The corresponding estimators of correlation are referred as r_{LAV} and r_{LMS}, respectively.

2.4. Robust correlation via the robust variances of principal variables. Consider the following identity for the correlation coefficient ρ [9]

$$\rho = [\mathrm{D}(U) - \mathrm{D}(V)]/[\mathrm{D}(U) + \mathrm{D}(V)], \tag{12}$$

where $U = (X/\sigma_1 + Y/\sigma_2)/\sqrt{2}$, $\quad V = (X/\sigma_1 - Y/\sigma_2)/\sqrt{2}$ are the principal variables such that

$$\mathrm{Cov}(U, V) = 0, \quad \sigma_U^2 = 1 + \rho, \quad \sigma_V^2 = 1 - \rho,$$

and σ_1 and σ_2 are the standard deviations of the r.v.'s X and Y, respectively.

Following Huber [1], introduce a scale functional $S(X) : S(aX+b) = |a|S(X)$ and write $S^2(\cdot)$ for a robust counterpart of variance. Then a corresponding counterpart for (12) is given by [1]

$$\rho^*(X, Y) = [S^2(U) - S^2(V)]/[S^2(U) + S^2(V)]. \tag{13}$$

By substituting the sample robust estimates for S into (13), we obtain robust estimates for ρ [1]

$$\widehat{\rho} = [\widehat{S}^2(U) - \widehat{S}^2(V)]/[\widehat{S}^2(U) + \widehat{S}^2(V)]. \tag{14}$$

The choice of the median absolute deviation $\widehat{S}(x) = \mathrm{MAD}\, x$ in (14) yields a remarkable estimator called the MAD-correlation coefficient [35]

$$r_{\mathrm{MAD}} = (\mathrm{MAD}^2 u - \mathrm{MAD}^2 v)/(\mathrm{MAD}^2 u + \mathrm{MAD}^2 v), \tag{15}$$

where u and v are the robust principal variables

$$u = \frac{x - \operatorname{med} x}{\sqrt{2}\ \mathrm{MAD}\, x} + \frac{y - \operatorname{med} y}{\sqrt{2}\ \mathrm{MAD}\, y}, \qquad v = \frac{x - \operatorname{med} x}{\sqrt{2}\ \mathrm{MAD}\, x} - \frac{y - \operatorname{med} y}{\sqrt{2}\ \mathrm{MAD}\, y}. \tag{16}$$

Choosing Huber's trimmed standard deviation estimators as $\widehat{S} \propto \sqrt{\sum_{n_1+1}^{n-n_2} x_{(i)}^2}$ (see [1], pp. 120-122), we obtain the *trimmed correlation coefficient*:

$$r_{\mathrm{TRIM}} = \left(\sum_{i=n_1+1}^{n-n_2} u_{(i)}^2 - \sum_{i=n_1+1}^{n-n_2} v_{(i)}^2 \right) \Bigg/ \left(\sum_{i=n_1+1}^{n-n_2} u_{(i)}^2 + \sum_{i=n_1+1}^{n-n_2} v_{(i)}^2 \right), \tag{17}$$

where $u_{(i)}$ and $v_{(i)}$ are the ith order statistics of the corresponding robust principal variables, n_1 and n_2 are the numbers of trimmed observations.

Formula (17) yields the following limit cases: (i) the sample correlation coefficient r with $n_1 = 0$, $n_2 = 0$ and with the classical estimators (the sample means for location and the standard deviations for scale) in its inner structure; (ii) the median correlation coefficient with $n_1 = n_2 = [0.5(n-1)]$

$$r_{\mathrm{MED}} = (\mathrm{med}^2\,|u| - \mathrm{med}^2\,|v|)/(\mathrm{med}^2\,|u| + \mathrm{med}^2\,|v|) \tag{18}$$

asymptotically equivalent to r_{MAD} [35].

2.5. Robust correlation via robust minimax estimation. In the literature, there are two results on applying the minimax approach to robust estimation of correlation.

In [1], it is stated that the quadrant correlation coefficient r_{Q} (7) is asymptotically minimax with respect to bias at the mixture $F = (1-\varepsilon)G + \varepsilon H$ (G and H being centrosymmetric distributions in $\mathbf{R}^2$).

In [49], it is shown that the trimmed correlation coefficient r_{TRIM} (17) is asymptotically minimax with respect to variance for ε-contaminated bivariate normal distributions

$$f(x,y) \geq (1-\varepsilon)\,\mathcal{N}(x,y;0,0,1,1,\rho), \qquad 0 \leq \varepsilon < 0.205.$$

This result holds under rather general conditions of regularity imposed on joint distribution densities $f(x,y)$ similar to the conditions under which Huber's M-estimators of scale are consistent and minimax (for details, see [1, 49], and under the following two additional conditions.

The first presumes that the parameters of location and scale of the r.v.'s X and Y are known: then we set $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$. In general, this condition is not very restrictive, as any reasonable measure of correlation should be invariant to the shifts of location and scale of the r.v.'s X and Y. Further, according to the aforementioned result, the trimmed correlation coefficient (17) should be used with the principal variables (16) in which zeros and units stand for the sample medians and median absolute deviations, respectively. However, in real-life applications, one should use formulas (16) as they are given.

The second condition is more restrictive: the underlying distribution should be independent with respect to its principal variables taking the following form

$$f(x,y) = \frac{1}{\sqrt{1+\rho}}\, g\left(\frac{u}{\sqrt{1+\rho}}\right) \frac{1}{\sqrt{1-\rho}}\, g\left(\frac{v}{\sqrt{1-\rho}}\right), \tag{19}$$

where the principal variables u, v are given by $u = (x+y)/\sqrt{2}$, $v = (x-y)/\sqrt{2}$, and $g(x)$ is a symmetric density $g(-x) = g(x)$ with unit variance [49]. In this case ρ is just the correlation coefficient of distribution (19) with $\mathrm{D}\,X = \mathrm{D}\,Y = 1$ in full correspondence with the assumption that $\sigma_1 = \sigma_2 = 1$.

The upper-bound $\overline{\varepsilon} = 0.205$ on the contamination parameter ε arises due to the requirement of a bounded variance σ_g^2. In this case, the least favorable distribution

is normal in the middle and a t-distribution in the tails (with the number of degrees of freedom defined by the value of ε), having a bounded variance just for $\varepsilon < 0.205$ (see [49] and [1], p. 121, Exhibit 5.6.1).

The idea of introducing class (19) can be formulated as follows: for any random pair (X, Y) the transformation $U = X + Y$, $V = X - Y$ gives the uncorrelated random principal variables (U, V) (actually independent for densities (19)), and estimation of their scales solves the problem of estimation of correlation between X and Y with the use of the estimators of Section 2.4. Thus, class (14) of estimators class entirely corresponds to class (19) of distribution densities, and this allows to extend Huber's results for minimax M- and L-estimators of location and scale on estimation of correlation.

Note that class (19) contains the standard bivariate normal distribution density $f(x, y) = \mathcal{N}(x, y; 0, 0, 1, 1, \rho)$ if $g(x) = (2\pi)^{-1/2} \exp(-x^2/2)$.

The levels of trimming n_1 and n_2 of the trimmed correlation coefficient r_{TRIM} depend on the contamination parameter ε: $n_1 = n_1(\varepsilon) = 0$ and $n_2 = n_2(\varepsilon) < [0.1\, n]$ for $\varepsilon < 0.205$.

The minimax variance estimator r_{TRIM} is asymptotically equivalent to the sample correlation coefficient r if $\varepsilon = 0$ with $n_1 = n_2 = 0$ and, in the particular case of heavy contamination with $\varepsilon = 0.2$, it has the levels of trimming equal to $n_1 = 0$ and $n_2 = [0.098\, n]$. Note that in the latter case, the level of trimming appears to be rather moderate as compared with the level of contamination.

2.6. Robust correlation via stable parameter estimation. In this section, the variational optimization approach to robust estimation proposed in [23, 24] is applied to stable estimation of the correlation coefficient ρ of a bivariate normal distribution.

For the observations $(x_1, y_1), \ldots, (x_n, y_n)$ from a bivariate normal distribution with zero means, unit variances and an unknown correlation coefficient ρ with density $\mathcal{N}(x, y; 0, 0, 1, 1, \rho)$ (for brevity, denote it as $\mathcal{N}(x, y; \rho)$), consider the M-estimators of ρ in the form

$$\sum \psi(x_i, y_i; \widehat{\rho}) = 0, \tag{20}$$

where $\psi(x, y; \rho)$ is a score function belonging to some class Ψ.

Under general conditions of regularity put on score functions ψ, M-estimators are consistent and asymptotically normal with the asymptotic variance $\mathrm{D}(\widehat{\rho}) = n^{-1}V(\psi) = n^{-1}\mathrm{E}_N(\psi^2)/[\mathrm{E}_N(\partial\psi/\partial\rho)]^2$ [1], where $\mathrm{E}_N(\cdot)$ denotes the operation of expectation over the underlying bivariate normal distribution $N(x, y)$ with density $\mathcal{N}(x, y; \rho)$. The minimum of the asymptotic variance is attained at the maximum likelihood score function $\psi_{\mathrm{ML}}(x, y; \rho) = \partial \log \mathcal{N}(x, y; \rho)/\partial\rho$ (here, the ML-estimator is given by $r_{\mathrm{ML}} = n^{-1}\sum x_i y_i$) with variance $V^* = V(\psi_{\mathrm{ML}}) = (1-\rho^2)^2/(1+\rho^2)$ [46], and the efficiency of an M-estimator is defined by the ratio $\mathrm{Eff}(\widehat{\rho}) = V^*/V(\psi)$.

To measure robustness of estimation, other characteristics, complementary to efficiency, are used, e.g., such as the supremum of asymptotic variance over a given class of distributions within Huber's minimax approach [1], or the influence and

change-of-variance functions within Hampel's approach [25], etc. In [23, 24], a functional called the *instability* of an M-estimator is defined as

$$W(\psi) = \int\int \psi^2(x, y; \rho)\, dx\, dy / [\mathrm{E}_N\, (\partial\psi/\partial\rho)]^2,$$

and an optimal score function $\psi_*(x, y; \rho)$ minimizing the instability is obtained: $\psi_* = \arg\min_{\psi\in\Psi} W(\psi)$. The M-estimator with this score function is called an estimator of maximum stability yielding $W_* = W(\psi_*)$, and similarly to efficiency, a new characteristic called the *stability* of an M-estimator is introduced as follows: $\mathrm{Stb}(\widehat{\rho}) = W_*/W(\psi)$, naturally lying in the $[0, 1]$ range. This characteristic of robustness measures the local sensitivity of an estimator to the variations of a model distribution, in our case, the bivariate normal distribution.

Setting different weights for efficiency and stability, various criteria of optimization of estimation were proposed in [24]. The equal weights, when $\mathrm{Eff}(\widehat{\rho}) = \mathrm{Stb}(\widehat{\rho})$, lead to a *radical* M-estimator with the score function

$$\psi_{\mathrm{RAD}}(x, y; \rho) = (\partial \log \mathcal{N}(x, y; \rho)/\partial\rho + \beta)\ \sqrt{\mathcal{N}(x, y; \rho)}, \tag{21}$$

where the constant β is obtained from the condition of consistency $\mathrm{E}_N(\psi_{\mathrm{RAD}}) = 0$.

Note that it is impossible to provide simultaneously unit efficiency and stability, as usually the stability of an efficient maximum likelihood estimator is zero.

From (21) it follows that the radical M-estimator of correlation r_{RAD}, henceforth called as *the radical correlation coefficient*, satisfies

$$\sum_i \left\{2r_{\mathrm{RAD}}^3 + [3(x_i^2 + y_i^2) - 2]r_{\mathrm{RAD}} - 3(1 + r_{\mathrm{RAD}}^2)\, x_i y_i\right\} e^{-q_i/2} = 0, \tag{22}$$

where $q_i = (x_i^2 - 2r_{\mathrm{RAD}}\, x_i y_i + y_i^2)/[2(1 - r_{\mathrm{RAD}}^2)]$. Since $\psi_{\mathrm{RAD}}(x, y; \rho) \to 0$ as $\sqrt{x^2 + y^2} \to \infty$, equation (22) defines Meshalkin's redescending λ-estimator with exponential weights [26, 27].

Finally, equation (22) was obtained in the setting, where the parameters of location and scale were assumed known. In practice, this assumption, as a rule, does not hold. Hence, in (22), one should use the data $\{(x_i, y_i)\}_1^n$ centered and standardized by the sample median and the sample median absolute deviation, respectively.

2.7. Robust correlation via rejection of outliers. The preliminary rejection of outliers from the data with the subsequent application of a classical estimator (for example, the sample correlation coefficient) to the rest of the observations defines the two-stage group of robust estimators of correlation. Their variety wholly depends on the variety of the rules for detection and/or rejection of multivariate outliers based on using discriminant, component, factor analysis, canonical correlation analysis, projection pursuit, etc. [44], [50] – [54] (note the work [53] for a deep insight into the problem).

Obviously, this topic deserves a separate consideration. However, we emphasize the following main characteristics of rejection of multivariate outliers: since multivariate outliers can distort not only location and scale, but also the orientation and

shape of the point-cloud in the space, the types of outliers are numerous and it is difficult to figure out which type the outlier belongs to [53]. Thus, it might prove to be impossible to develop just one procedure which would be a reliable guard against outliers. Hence, there must be a variety of procedures for different types of outliers with the corresponding two-stage procedures of robust estimation of correlation.

Moreover, each robust procedure of estimation inherently possesses its own rule for rejection of outliers [1, 2], and it may seem that then there is no need for any independent procedure for rejection, at least if to aim at estimation, and therefore no need for two-stage procedures of robust estimation. However, a rejection rule may be quite informal, for example, based on a prior knowledge about the nature of outliers, and, in this case, its use can improve the efficiency of estimation.

Consider a classical approach to rejection of outliers based on the use of the *Mahalanobis distances* d_i^2 between the points $\mathbf{p}_i = (x_i, y_i)^T,\ i = 1, \ldots, n$ in $\mathbf{R}^2$ and the sample mean $\mathbf{m} = n^{-1} \sum \mathbf{p}_i$: $d_i^2 = (\mathbf{p}_i - \mathbf{m})^T \mathbf{S}^{-1} (\mathbf{p}_i - \mathbf{m})$, where $\mathbf{S}$ is the sample covariance matrix $\mathbf{S} = n^{-1} \sum (\mathbf{p}_i - \mathbf{m})(\mathbf{p}_i - \mathbf{m})^T$. These distances are ranked $d_{(1)}^2 \le d_{(2)}^2 \le \cdots \le d_{(n)}^2$, and, obviously, the observations with greater Mahalanobis distances are the candidates for outliers. Further, one can use some rejection tests based on the largest Mahalanobis distances, say, the test $d_{(n)}^2 \le \lambda_\alpha$ (the threshold λ_α is determined from the test size $P[d_{(n)}^2 \le \lambda_\alpha] = 1 - \alpha$) [43, 50].

In case when there are gross errors in the data, the use of the classical sample mean and covariance matrix destroys this rejection procedure because of great sensitivity of the classical estimators to outliers. Thus the problem of rejecting outliers in the multivariate case obviously requires robust estimation of multivariate location and shape. The latter problem is one of the most difficult in robust statistics [1, 2, 24, 26, 33, 34, 44, 55, 56]. The classical procedure of rejection can be robustified by the use of robust analogs of the Mahalanobis distances with robust estimates for means and covariance matrices, e.g., the minimum volume and minimum covariance determinant combinatorial estimators [2, 44].

However, in our study, the problem of rejection of outliers is subordinate to the problem of robust estimation of correlation, so we are mainly interested in relatively simple procedures of rejection in the bivariate case.

Now we sketch a heuristic rejection procedure in principal axes based on the ellipse rule (ELL) [20, 57]. Assume that the main bulk of the data is of an elliptic shape, outliers are of a gross error nature, and the expected fraction of outliers is approximately known. Then transform the initial data $(x_1, y_1), \ldots, (x_n, y_n)$ to the principal variables $(u_1, v_1), \ldots, (u_n, v_n)$ (16) and trim all the points lying out of the ellipse contour $\left(\dfrac{u_i - \operatorname{med} u}{k \operatorname{MAD} u}\right)^2 + \left(\dfrac{v_i - \operatorname{med} u}{k \operatorname{MAD} v}\right)^2 = 1$, where k is determined iteratively so that the given fraction of the data should lie inside the ellipse ("good" data). Further, the number of outliers can be specified by the use of robust Mahalanobis distances (for details, see [20, 57]). The corresponding estimator of correlation is denoted as r_{ELL} and it is the sample correlation coefficient of the "good" data. In our study, the expected fraction of outliers was taken equal to $\varepsilon = 0.1$.

Note that the use of only the first, relatively simple, stage of this algorithm can considerably improve the performance of the sample correlation coefficient in

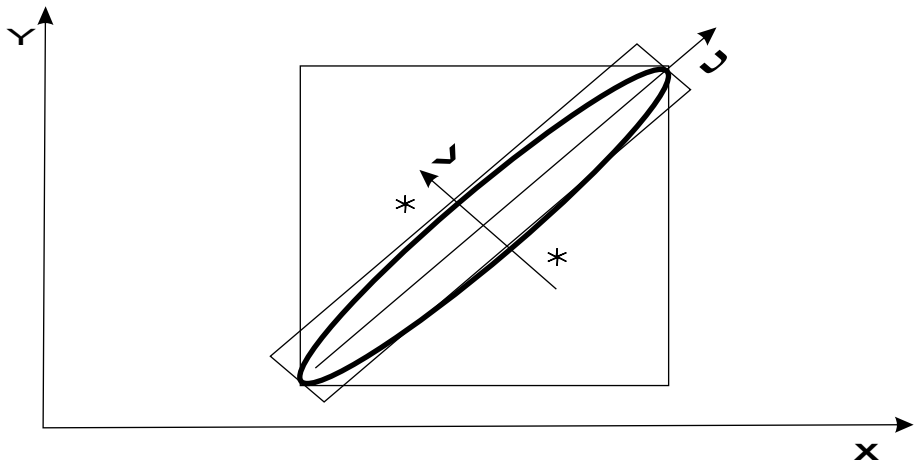

Figure 2: Outliers ($*$) in the initial and principal axes

contaminated normal models [20]. It is important to detect outliers in the principal axes u and v, not in the initial axes x and y. Fig. 2 illustrates this effect: the data regarded as outliers (marked by stars) in the principal axes should not necessarily be such in the initial axes.

3 Performance evaluation

In this section, we examine the performance of the introduced estimators of a correlation coefficient for the bivariate normal distribution density $\mathcal{N}(x, y; 0, 0, 1, 1, 0.9)$ and for the ε-contaminated bivariate normal distribution (5) with $\varepsilon = 0.1$, $k = 10$ and $\rho' = -0.9$ both on large (as $n \to \infty$) and small ($n = 20$) samples. The analytic expressions for estimator's means and asymptotic variances are exhibited whenever they are available and not cumbersome, however, their numerical values are written out for the considered models.

3.1. Performance evaluation on large samples. The performance of robust estimators on large samples was measured by their means and asymptotic variances computed partially with the use of the influence functions $IF(x, y; \widehat{\rho})$ [2]

$$\mathrm{E}(\widehat{\rho}) \approx \rho + \int IF(x, y; \widehat{\rho})\, f(x, y)\, dx\, dy, \quad \mathrm{D}(\widehat{\rho}) = n^{-1} \int IF^2(x, y; \widehat{\rho})\, f(x, y)\, dx\, dy,$$

where density $f(x, y)$ is given by (5). In the bivariate case, the influence function $IF(x, y; \rho)$ is defined as

$$IF(x, y; \widehat{\rho}) = \frac{d}{ds}\, \rho\left((1 - s)F + s\Delta_{xy}\right)\Big|_{s=0} \tag{23}$$

where $\rho = \rho(F)$ is a functional defining the correlation measure (e.g., the classical correlation coefficient, the quadrant correlation, etc.), $F = F(x, y)$ is a bivari-

ate distribution function and $\Delta_{x_0y_0}$ is a bivariate analog of the Heaviside function: $\Delta_{x_0y_0} = 1$ for $x \geq x_0$, $y \geq y_0$ and $\Delta_{x_0y_0} = 0$ otherwise.

Note that the influence function itself can serve as an important tool for the analysis of the qualitative robust properties of an estimator such as the sensitivity to gross outliers, to rounding off, etc. [2].

The sample correlation coefficient r.
For the bivariate normal distribution [46],

$$\mathrm{E}(r) = \rho\left[1 - (1-\rho^2)/(2n) + O\left(1/n^2\right)\right], \quad \mathrm{D}(r) = (1-\rho^2)^2/n;$$

and under contamination,

$$IF(x, y; r) = -\mathrm{E}(r)\,(x^2+y^2)/[2(1-\varepsilon+\varepsilon k^2)] + xy/[1-\varepsilon+\varepsilon k^2],$$

where $\mathrm{E}(r) = [(1-\varepsilon)\rho + \varepsilon k^2\rho']/[1-\varepsilon+\varepsilon k^2]$.

The direct robust counterparts of r: $r_\alpha(\psi)$, r_α, and r_{COMED}.
For the ε-contaminated bivariate normal distribution, the analytical results on the mean and variance of r_{COMED} can be found in [35, 45], some numerical results on the performance of r_α are represented in [9].

The nonparametric measures: r_{Q} and r_{S}.
For the ε-contaminated bivariate normal distribution,

$$\mathrm{E}(r_{\mathrm{Q}}) = 2\,(1-\varepsilon)\arcsin(\rho)/\pi + 2\,\varepsilon\arcsin(\rho')/\pi, \quad \mathrm{D}(r_{\mathrm{Q}}) = [1 - \mathrm{E}^2(r_{\mathrm{Q}})]/n,$$

$$IF(x, y; r_{\mathrm{Q}}) = \mathrm{sgn}(x - \mathrm{Med}\,X)\,\mathrm{sgn}(y - \mathrm{Med}\,Y) - \rho_{\mathrm{Q}},$$

where $\rho_{\mathrm{Q}} = \int \mathrm{sgn}\,(x-\mathrm{Med}\,X)\,\mathrm{sgn}\,(y-\mathrm{Med}\,Y)\,dF(x,y)$. Note that, for the bivariate normal distribution, $\mathrm{E}(r_{\mathrm{Q}}) = 2\arcsin(\rho)/\pi$, thus the quadrant correlation coefficient measures the different from ρ quantity [6].

For the Spearman rank correlation coefficient r_{S} [46],

$$\mathrm{E}(r_{\mathrm{S}}) = 6(1-\varepsilon)\arcsin\,(\rho/2)\,/\pi + 6\,\varepsilon\arcsin\,(\rho'/2)\,/\pi.$$

Similarly to r_{Q}, for the bivariate normal distribution, it measures the following quantity: $6\,\arcsin(\rho/2)/\pi$, not ρ.

Since robustness to gross errors is provided by the bounded influence function [2], we confirm this assertion having the bounded $IF(x, y; r_{\mathrm{Q}})$ and the unbounded $IF(x, y; r)$.

The regression group estimators: r_{REG}, r_{LAV} and r_{LMS}.
For regression group estimators, we represent the result on the mean for the r_{REG} based on the median of slopes [35]

$$\mathrm{E}(r_{\mathrm{REG}}) = \rho + \varepsilon\,\arctan[(\rho'-\rho)\sqrt{1-\rho^2}/\sqrt{1-\rho'^2}] + o(\varepsilon).$$

Its asymptotic variance is also given in [35]. Another good estimator of this group, the r_{LMS} based on the LMS regression, has the order of convergence $n^{-1/3}$ [44], hence, on large samples, it is inferior to all other estimators examined in this study.

Table 1. Normal distribution $\mathcal{N}(x, y; 0, 0, 1, 1, 0.9)$: asymptotics.

	r	r_{Q}	r_{S}	r_{REG}	r_{LAV}	r_{MAD}	r_{MED}	r_{TRIM}	r_{RAD}
$\mathrm{E}(\widehat{\rho})$	0.90	0.93	0.90	0.90	0.90	0.90	0.90	0.90	0.90
$n\,\mathrm{D}(\widehat{\rho})$	0.02	0.13	0.05	0.07	0.09	0.06	0.06	0.04	0.03

The estimators based on robust principal variables: r_{TRIM} and r_{MED}.
Recall that these estimators are the minimax variance estimators of a correlation coefficient for ε-contaminated bivariate normal distributions (see Section 2.5), and their asymptotic variances are written out in [49], e.g., for the bivariate normal distribution, the asymptotic variance of r_{MED} is given by

$$\mathrm{D}(r_{\mathrm{MED}}) = (1-\rho^2)^2/[8n\phi^2(\zeta_{3/4})\zeta_{3/4}^2],$$

where $\zeta_{3/4} = \Phi^{-1}(3/4)$, $\Phi(z) = (2\pi)^{-1/2}\displaystyle\int_{-\infty}^{z} e^{-t^2/2}\,dt$ is the standard normal cumulative and $\phi(z) = \Phi'(z)$. The asymptotic relative efficiency of the r_{MED} to the sample correlation coefficient r is rather low being equal to 0.367.

The radical correlation coefficient: r_{RAD}.
Its characteristics of accuracy and stability are given by

$$\mathrm{D}\,(r_{\mathrm{RAD}}) = [81(9+10\rho^2)(1-\rho^2)^2]/[512(1+\rho^2)^2\,n],$$

$$\mathrm{Eff}(r_{\mathrm{RAD}}) = \mathrm{Stb}(r_{\mathrm{RAD}}) = [512(1+\rho^2)]/[81(9+10\rho^2)]. \tag{24}$$

The efficiency and stability of the radical estimator vary in a narrow range from $\mathrm{Eff}_{\min} = \mathrm{Stb}_{\min} = 0.6654$ to $\mathrm{Eff}_{\max} = \mathrm{Stb}_{\max} = 0.7023$. Thus the radical estimator possesses reasonable levels of efficiency and stability. Note that the efficiency of the maximum likelihood estimator $r_{\mathrm{ML}} = n^{-1}\sum x_i y_i$, namely the solution of (20) with $\psi = \psi_{\mathrm{ML}}$, is unit, but its stability is zero, since the integral $\int\int \psi_{\mathrm{ML}}^2(x, y; \rho)\,dx\,dy$ does not exist.

The two-stage estimators.
Since it is difficult to describe the distribution of the data after applying any reasonable rejection procedure (there is always a nonzero probability that a small fraction of "bad" data is in the core of "good" data), there are no results on the asymptotic performance of these estimators.

Some numerical results are listed in Table 1 and Table 2. The analysis of these results is given in Section 4.

3.2. Monte Carlo results on small samples. Now we display some results on small samples when $n = 20$ for the same models which were used in Section 3.1. As a rule, the number of trials was set to 1000, and in particular cases, it was

Table 2. Contaminated normal distribution (5) with $\varepsilon = 0.1$, $\rho = 0.9$, $\rho' = -0.9$, $k = 10$: asymptotics.

	r	r_{Q}	r_{S}	r_{REG}	r_{LAV}	r_{MAD}	r_{MED}	r_{TRIM}	r_{RAD}
$\mathrm{E}(\widehat{\rho})$	-0.75	0.57	0.71	0.84	0.74	0.88	0.88	0.85	0.86
$n\,\mathrm{D}(\widehat{\rho})$	1.00	0.46	0.32	0.50	0.65	0.13	0.13	0.05	0.05

Table 3. Normal distribution $\mathcal{N}(x, y; 0, 0, 1, 1, 0.9)$: $n = 20$.

	r	r_{Q}	r_{S}	r_{REG}	r_{LMS}	r_{MAD}	r_{MED}	r_{TRIM}	r_{RAD}	r_{ELL}
$\mathrm{E}(\widehat{\rho})$	0.90	0.69	0.87	0.88	0.90	0.83	0.83	0.81	0.80	0.85
$\mathrm{D}(\widehat{\rho})$	0.00	0.03	0.01	0.05	0.04	0.02	0.02	0.01	0.00	0.01

increased to 10000 for the sake of accuracy. In our study, the performance of all estimators, in definition of which the parameters of location and scale were supposed known, was obtained using their sample estimates, namely the sample median and the median absolute deviation. The results of modelling are displayed in Tables 3, 4 and discussed in Section 4.

4 Discussion and conclusions

To the best of our knowledge, only a few works survey robust methods of estimation of correlation: namely, the pioneer works [9, 10], in which direct robust counterparts of the sample correlation coefficient, robust estimators based on nonparametric measures, principal variables and trimming outliers were examined; the book of Huber [1], in which one chapter is mainly devoted to a significantly more complicated problem of robust estimation of covariance matrices; and the works [20, 35, 36, 49], the results of which are partially represented in this paper. In particular, in the former work [35], a new group of robust estimators based on robust regression was proposed; in [49], the minimax variance approach was applied to designing robust estimators; and in the book [20], one chapter highlights the problem of robust estimation of correlation.

In our study, basing on the former and recent results in robust estimation of

Table 4. Contaminated normal distribution (5) with $\varepsilon = 0.1$, $\rho = 0.9$, $\rho' = -0.9$, $k = 10$: $n = 20$.

	r	r_{Q}	r_{S}	r_{REG}	r_{LMS}	r_{MAD}	r_{MED}	r_{TRIM}	r_{RAD}	r_{ELL}
$\mathrm{E}(\widehat{\rho})$	-0.55	0.48	0.37	0.71	0.90	0.81	0.81	0.78	0.76	0.83
$\mathrm{D}(\widehat{\rho})$	0.37	0.04	0.09	0.06	0.04	0.02	0.02	0.01	0.01	0.02

correlation, we have selected several robust estimators, the best and typical representatives of the groups of estimators introduced in Section 2, and compared their performances in ε-contaminated normal model (5). The reasons for the choice of this model were given in Section 1.3. To characterize the case of a relatively heavy contamination distinct from the main bulk of the data, the particular values of the parameters of contamination $\varepsilon = 0.1$, $\rho' = -\rho$ and $k = 10$ were chosen. Further, we discuss the influence of the choice of the parameters ε and k on the performance of optimal estimators. The choice of the sample size $n = 20$ was also made in order to consider the two distinct cases, small samples and asymptotics. The performance of robust estimators of $\rho = 0$ and $\rho = 0.5$ on samples $n = 30$ and $n = 60$ was represented in [9, 10, 20, 35], and it was shown that, on samples $n = 60$, the results of modelling practically coincided with the asymptotic recipes.

Now we analyze and discuss the results represented in Section 3.
Normal distribution. From Table 1 and Table 3 it follows that
1) on small and large samples, the best is the sample correlation coefficient r both by its bias and variance;
2) on large samples, the radical correlation coefficient r_{RAD} is the best among the rest of estimators by its variance, but on small samples it has a considerable bias.
3) the quadrant correlation coefficient r_{Q} and the rank correlation r_{S} have comparatively moderate variances, but the bias of the r_{Q} is not satisfactory;
4) the regression estimators r_{REG}, r_{LAV}, and r_{LMS} are inferior in variances to the MAD, median and trimmed correlation coefficients, the latter slightly better than the former;
5) the two-stage estimator r_{ELL} performs well on small samples being better than r_{Q}, r_{TRIM}, r_{MED} and r_{MAD} both in bias and variance;
6) the MAD- and median correlation estimators practically repeat each other in behavior;
7) on large samples, the biases of estimators can be neglected, but not their variances.

Contaminated normal distribution. From Table 2 and Table 4 it follows that
1) the sample correlation coefficient r is catastrophically bad under contamination;
2) the classical nonparametric estimators r_{Q} and r_{S} behave moderately ill;
3) the regression estimators r_{REG} and r_{LAV} are good in bias but have large variances, the exception is the estimator r_{LMS} based on the least median squares method, which performs well on small samples;
4) the best estimators are the trimmed r_{TRIM} and radical correlation r_{RAD} coefficients together with r_{LMS} and r_{ELL} which are good only on small samples;
5) the radical correlation coefficient r_{RAD} is superior in variance, but the median correlation r_{MED} is better in bias;
6) under heavy contamination, the bias of an estimator seems to be a more informative characteristic than its variance, thus the problem of designing a best robust estimator of correlation with respect to bias is still actual.

The trimmed and radical correlation coefficients: r_{TRIM} *and* r_{TRIM}.
1) These competitive estimators are rather close in performance, though they belong to different groups of estimators. The trimmed correlation coefficient r_{TRIM} is the

minimax variance estimator in the class of ε-contaminated normal distributions (in fact, a specific subclass of it) with a rather arbitrary contamination distribution, which, in particular, may be significantly more heavy-tailed than the normal distribution. The radical correlation coefficient r_{RAD} is a locally robust estimator strongly depending on the assumed parametric form of the underlying distribution, e.g., for the bivariate normal distribution, defined by (22). The r_{TRIM} is a correlation analog of such estimators as the trimmed mean for location and the trimmed standard deviation for scale [1, 49] and it inherits from them such a global robustness property as their high *breakdown points* (roughly speaking, a breakdown point defines the maximum fraction of the sample contamination still admissible for an estimator's performance [2]). Since the breakdown point of r_{TRIM} is $\varepsilon^* = \varepsilon$, it is easy to predict its performance in ε-contaminated normal models (5): e.g., if r_{TRIM} is designed for some fixed fraction of contamination ε, say, for $\varepsilon = 0.1$, then its performance will remain good for any $\varepsilon < 0.1$ regardless of the value of the contamination scale parameter k, and it may get worse for $\varepsilon > 0.1$.

2) The trimmed correlation coefficient r_{TRIM} proved its high robustness in former experimental studies [9, 10], and its optimality in ε-contaminated models explains those results. Since it is not easy to estimate the parameter of contamination ε from the sample, we recommend to use r_{TRIM} designed for a practically reasonable upper-bound value of contamination $\overline{\varepsilon} = 0.205$ with the corresponding 10% level of trimming: $n_1 = 0$ and $n_2 = [0.1\, n]$.

3) The median correlation coefficient r_{MED} has the maximum value of the breakdown point $\varepsilon^* = 1/2$ being maximally resistant against gross errors in the data. It may be regarded as a correlation analog to such well-known and widely used estimators as the sample median for location and the median absolute deviation for scale. Thus, we recommend to use r_{MED} as a fast and highly robust estimator of correlation, though it has a low efficiency (0.367) under normal models.

4) In Section 2.5, it was mentioned that the quadrant correlation coefficient r_{Q} is asymptotically minimax with respect to bias over ε-contaminated distributions. Although its bias is minimax, the quadrant correlation coefficient r_{Q} demonstrates moderate robustness in the Monte Carlo experiment. This can be explained by the choice of a relatively poor class of direct robust counterparts of the sample correlation coefficient ($r_\alpha(\psi)$-estimators of Section 2.1) for which the optimality of r_{Q} is established (see Fig. 2). Nevertheless, r_{Q} can be regarded as a moderate robust alternative to the sample correlation coefficient, for its simple structure and its sample binomial distribution [6].

5) Since the asymptotic variance of r_{TRIM} depends on the estimated correlation coefficient ρ only through the multiplier $(1-\rho^2)^2$ [49], it is possible to construct confidence intervals for it using the Fisher transformation $z = \ln[(1+\widehat{\rho})/(1-\widehat{\rho})]/2$: in this case, the variance of the transformed variable z does not depend on ρ [46]. As the asymptotic variance of r_{RAD} is not of the required form (24), its z-transform does not have the desired property, and the corresponding confidence intervals cannot be so easily constructed.

6) Computation of robust estimators of correlation requires determination of the sample medians for various data that can be performed either using some algorithm

of sorting [58] or the method of re-weighted least squares (RWLS) [47, 59]. In our computations, we usually use the latter method: to reach a practically acceptable accuracy of computing the sample median, say, with the relative error $\delta = 0.01$, it needs about $3 - 5$ iterations of the RWLS, and about $5 - 7$ iterations of Newton's method are required for computing the radical correlation coefficient r_{RAD} from (22) with the quadrant correlation r_{Q} as a starting point. Thus, with respect to computation, r_{TRIM} is preferable as compared to r_{RAD}: on average, computing of r_{TRIM} is three times faster than of r_{RAD}.
7) Finally, we definitely conclude that, on large samples ($n \geq 60$) from ε-contaminated bivariate normal distributions, the trimmed correlation coefficient r_{TRIM} is preferable as compared to the radical correlation coefficient r_{RAD} due to its optimality in a wider class of distributions and a definitely easier computation. On small samples, we cannot so definitely conclude, for, in applied statistics, recommendations for small samples are usually indefinite.

Acknowledgement: The authors are very grateful to the reviewers whose helpful comments substantially improved the paper.

References

[1] Huber, P.J., *Robust Statistics*, New York: John Wiley, 1981, Translation in Russian (1984).

[2] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., *Robust Statistics. The Approach Based on Influence Functions*, New York: John Wiley, 1986, Translation in Russian (1989).

[3] Rogers, J.L. and Nicewander, W.A., Thirteen ways to look at the correlation coefficient, *The Amer. Statistician*, 1988, vol. 42, pp. 59-66.

[4] Rovine, M.J. and von Eye, A., A 14th way to look at a correlation coefficient: correlation as the proportion of mathches, *The Amer. Statistician*, 1997, vol. 51, pp. 42-46.

[5] Spearman, C., The proof and measurement of association between two things, *Amer. J. Psychol.*, 1904, vol. 15, pp. 88-93.

[6] Blomqvist, N., On a measure of dependence between two random variables, *Ann. Math. Statist.*, 1950, vol. 21, pp. 593-600.

[7] Kendall, M.G., Rank and product-moment correlation, *Biometrika*, 1949, vol. 36, pp. 177-180.

[8] Tukey, J.W., A survey of sampling from contaminated distributions. In: *Contributions to Prob. and Statist. (Olkin, I., Ed.)*., 1960, Stanford: Stanford Univ. Press, pp. 448-485.

[9] Gnanadesikan, R. and Kettenring, J.R., Robust estimates, residuals and outlier detection with multiresponse data, *Biometrics*, 1972, vol.28, pp. 81-124.

[10] Devlin, S.J., Gnanadesikan, R. and Kettenring, J.R., Robust estimation and outlier detection with correlation coefficient, *Biometrika*, 1975, vol. 62, pp. 531-545.

[11] Huber, P.J., Robust estimation of a location parameter, *Ann. Math. Statist.*, 1964, vol 35, pp. 73-101.

[12] Hampel, F.R., *Contributions to the Theory of Robust Estimation.*, 1968, Ph.D. Thesis, University of California, Berkeley.

[13] Box, G.E.P., Non-normality and test on variances, *Biometrika*, 1953, vol. 40, pp. 318-335.

[14] Kolmogorov, A.N., Some works in the field of limit theorems of probability theory of last few years, *Bull. Moscow Univ.*, 1953, vol. 10, pp. 28-39 (in Russian).

[15] Prokhorov, Yu.V. Convergence of random processes and limit theorems in probability theory, *Theory Probab. Appl.*, 1956, vol. 1, pp. 157-214.

[16] Bickel, P.J., Another look at robustness: a review of reviews and some new developments, *Scand. J. Statist. Theory and Appl.*, 1976, vol. 3, pp. 145-168.

[17] Polyak, B.T., and Tsypkin, Ya.Z., Robust identification, *Automatica*, 1980, vol. 16, pp. 53-65.

[18] Tsypkin, Ya.Z., *Foundations of Information Theory of Identification*, Nauka, Moscow, 1984 (in Russian).

[19] Shulenin, V.P., *Introduction to Robust Statistics.* Tomsk Univ. Press, Tomsk, 1993 (in Russian).

[20] Shevlyakov, G.L. and Vilchevski, N.O., *Robustness in Data Analysis: criteria and methods*, Utrecht: VSP, 2002.

[21] Hampel, F.R., A general qualitative definition of robustness, *Ann. Math. Statist.*, 1971, vol. 42, pp. 1887-1896.

[22] Hampel, F.R., The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.*, 1974, vol. 69, pp. 383-393.

[23] Shurygin, A.M., Variational optimization of the stability of statistical estimation, *Avtomat. Telemekh.*, 1994, no. 11, pp. 73-86.

[24] Shurygin, A.M., *Applied Stochastics: robustness, estimation and forecasting*, Moscow: Finances and Statistics, 2000 (in Russian).

[25] Hampel, F.R., Rousseeuw, P.J., and Ronchetti, E. , The change-of-variance curve and optimal redescending M-estimators, *J. Amer. Statist. Assoc.*, 1981, vol. 76, pp. 643-648.

[26] Meshalkin, L.D., Some mathematical methods for the study of non-communicable diseases. In: *Proc. 6th Int. Meeting of Uses of Epidemiol. in Planning Health Services*, 1971, vol. 1, Primosten, Yugoslavia, pp. 250-256.

[27] Aivazyan, S.A., Bukhshtaber, V.M., Enyukov, I.S., and Meshalkin, L.D., *Applied Statistics: Classification and Reduction of Dimensionality*, Finansy i Statistika, Moscow, 1989 (in Russian).

[28] Maxwell, J.C., Illustration of the dynamical theory of gases. Part I. On the motion and collision of perfectly elastic spheres, *Phil. Mag.*, 1860, vol. 56.

[29] Galton, F., Family likeness in stature, *Proc. Roy. Soc. Lond.*, 1886, vol. 40, pp. 42-73.

[30] Shannon, C.E., A mathematical theory of communications. *Bell Syst. Tech. J.*, 1948, vol. 27, pp. 379-623.

[31] Jaynes, E.T., *Probability Theory. The Logic of Science*, Cambridge University Press, Cambridge, 2003.

[32] Bebbington, A. C. A method of bivariate trimming for estimation of the correlation coefficient, *Appl. Statist.*, 1978, vol. 27, pp. 221-226.

[33] Campbell, N. A., Robust procedures in multivariate analysis I: Robust covariance estimation, *Appl. Statist.*, 1980, vol. 29, pp. 231-237.

[34] Campbell, N. A., Robust procedures in multivariate analysis II: Robust canonical variate analysis, *Appl. Statist.*, 1982, vol. 31, pp. 1-8.

[35] Pasman, V.R. and Shevlyakov, G.L., Robust methods of estimation of a correlation coefficient, *Avtomat. Telemekh.*, 1987, no. 3, pp. 70-80.

[36] Shevlyakov, G. L., On robust estimation of a correlation coefficient, *J. Math. Sciences*, 1997, vol. 83, pp. 90-94.

[37] Kazakos, D. and Papantoni-Kazakos, P., *Detection and Estimation* , Rockville, MD: Computer Science Press, 1990.

[38] Delic, H. and Hocanin, A., Successive interference cancellation using robust correlation, 2002, *Proc. the First IEEE Balcan Conference on the Signal Processing, Communications, Circuits, and Systems,* Istanbul, Turkey, pp. 1221-1224.

[39] Kim, J. and Fessler, J.A., Intensity-based image registration using robust correlation coefficients, *IEEE Trans. Medical Imaging,* 2004, vol. 23, pp. 1430-1443.

[40] Trujillo, M. and Izquierdo, E., A robust correlation measure for correspondence estimation, 2004, *Proc. the 2nd Intern. Sympos. on 3D Data Processing, Visualization and Transmission*.

[41] Jurechkova, J., Statistical tests on tail index of a probability distribution, *METRON - International Journal of Statistics*, 2003, vol. LXI, no. 2, pp. 151-190.

[42] Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W., *Robust Estimates of Location*, Princeton Univ. Press, Princeton, 1972.

[43] Hawkins, D. M., *The Identification of Outliers*, London: Chapman & Hall, 1980.

[44] Rousseeuw, P. J., and Leroy, A. M., *Robust Regression and Outlier Detection*, New York: Wiley, 1987.

[45] Falk, M., A note on the correlation mediam for elliptical distributions, *J. Multivar. Analysis*, 1998, vol. 67, pp. 306-317.

[46] Kendall, M. G., and Stuart, A. , *The Advanced Theory of Statistics. Inference and Relationship*, vol. 2., London: Griffin, 1963, Translation in Russian (1973).

[47] Mudrov, V.I., and Kushko, V.L., *Methods of Data Processing: the Quasi-Likelihood Estimators*, Radio and Svyaz, Moscow, 1983 (in Russian).

[48] Rousseeuw, P.J., Least median of squares regression. *J. Amer. Statist. Assoc.*, 1984, vol. 79, pp. 871-880.

[49] Shevlyakov, G.L. and Vilchevsky, N.O., Minimax variance estimation of a correlation coefficient for epsilon-contaminated bivariate normal distributions, *Statist. Probab. Letters*, 2002, vol. 57, pp. 91-100.

[50] Barnett, V., and Lewis, T., *Outliers in Statistical Data*, New York: Wiley, 1978.

[51] Hadi, A. S., Identifying multiple outliers in multivariate data, *J. Royal Statist. Soc.*, 1992, vol. B54, pp. 761-771.

[52] Atkinson, A. C., Fast very robust methods for the detection of multiple outliers, *J. Amer. Statist. Assoc.*, 1994, vol.89, pp. 1329-1339.

[53] Rocke, D. M., and Woodruff, D. L., Identification of outliers in multivariate data, *J. Amer. Statist. Assoc.*, 1996, vol. 91, pp. 1047–1061.

[54] Atkinson, A., and Riani, M., *Robust Diagnostics Regression Analysis*, New York: Springer, 2000.

[55] Rousseeuw, P.J., Multivariate estimation with high breakdown point. In: *Mathematical Statistics and Applications (Grossman, W., Pflug, G., Vincze, I., and Wertz, W., Eds.)*, 1985, Reidel, Dodrecht, pp. 283-297.

[56] Rocke, D.M., and Woodruff, D.L., Computation of robust estimates of multivariate location and shape. *Statistica Neerlandica*, 1993, vol. 47, pp. 27-42.

[57] Shevlyakov, G.L., and Khvatova, T.Yu., On robust estimation of a correlation coefficient and correlation matrix. In: *MODA 5–Advances in Model-Oriented Data Analysis. (Atkinson, A. C.,* et al.*, Eds.)*, Physica, Heidelberg, pp. 153-162.

[58] Knuth, D.E., *The Art of Computer Programming, Vol. 1, Fundamental Algorithms*, Third edition, Moscow: Vi'iams, 2000 (in Russian).

[59] Weiszfeld, E., Sur le point par leque la somme des distansis de n points donnes est minimum, *Tôhoku Mathematics J.*, vol. 37, pp. 355-386.