Robust Multivariate Tolerance Region: Influence Function and Monte Carlo Study

Graciela Boente
Departamento de Matemática and Instituto de Cálculo,
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires and CONICET
Ciudad Universitaria, Pabellón 2, Buenos Aires, 1428, Argentina
gboente@mate.dm.uba.ar
and
Andrés Farall
Fundación Antorchas, Argentina
Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria, Pabellón 2, Buenos Aires, 1428, Argentina

Abstract

In this paper, we define a class of multivariate tolerance regions which turn out to be more resistant than the classical ones to outliers. The tolerance factors are evaluated using a simulation study under the central model and the sensitivity to deviations of the normal distribution for moderate samples is studied through a Monte Carlo study. Besides, the influence function of the coverage probability allows to compare the sensitivity of different proposals to anomalous data.

KEY WORDS: Coverage Probability, Multivariate Normal Distribution, Donoho–Stahel Estimator.

1 INTRODUCTION

Tolerance regions are widely used in industrial applications, being may be the most important one, quality control, where one seeks to guarantee that several variables satisfy given standards. See, for instance, Fuchs and Kenett (1988), where tolerance regions are applied on a quality control study of ceramic substrates used in the electronic industry. In particular, the theory of univariate tolerance regions for normal populations is well known, see Proschan (1953), Guttman (1970) and Odeh and Owen (1980). Under a normal distribution, even if there is no explicit solution for the tolerance factor, it can be obtained solving an univariate integral equation. Tables, algorithms and aproximations for this problem are described in Odeh and Owen (1980).

In the multivariate setting, the development is smaller and the computation of the tolerance factor is much more difficult since it involves, even in the normal setting, the resolution of a $(d+1)-$dimensional integral equation. The first attempt to construct a tolerance region for multivariate normally distributed observations, is due to John (1962). In the particular situation of a normal distribution the tolerance regions are ellipsoids centered at the sample mean. More precisely, given a training sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$, with distribution $P_{\boldsymbol{\theta}} = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, let $\mathbf{x}$ be a random vector independent of the sample such that $\mathbf{x} \sim P_{\boldsymbol{\theta}}$. Then, the tolerance region can be defined as $\mathcal{R} = \mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \left\{\mathbf{y} : (\mathbf{y} - \overline{\mathbf{x}})\, \mathbf{S}^{-1}\, (\mathbf{y} - \overline{\mathbf{x}}) \leq K\right\}$ where $\mathbf{S}$ denotes the sample covariance matrix and K is the tolerance factor to be choosen such that

$$P_{n,\boldsymbol{\theta}}\left[P_{\boldsymbol{\theta}}\left((\mathbf{x} - \overline{\mathbf{x}})\, \mathbf{S}^{-1}\, (\mathbf{x} - \overline{\mathbf{x}}) \leq K | \overline{\mathbf{x}}, \mathbf{S}\right) \geq q\right] \geq \delta \quad \forall \boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ ,$$

where the inner probability is measured with respect to $\mathbf{x}$ (conditional to the values $\mathbf{x}_1, \ldots, \mathbf{x}_n$) and $P_{n,\boldsymbol{\theta}}$ is the probability measure associated to $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Since the pioneer paper from John (1962), many authors have considered different approximations to compute tolerance factors. The results in John (1962), Chew (1966), Guttman (1970) and Krishnamoorthy and Mathew (1999) are based on approximations to the distribution of the arithmetic, geometric or harmonic means of the eigenvalues of a Wishart matrix to derive the values for the tolerance factors. Recently, Krishnamoorthy and Mathew (1999) gave approximations for the tolerance factor based on Monte Carlo study. These evaluations are possible even though they are computationally expensive.

In the univariate case, the sensitivity of the classical tolerance intervals was first noticed by Canavos and Koutraouvaelis (1984). See also Butler(1982) and Fernholz and Gillespie (2001) and the references therein. On the other hand, Fernholz (2002) extended the proposal of Fernholz and Gillespie (2001) to include robust statistics for the end-points in order to provide resistant tolerance intervals.

The same lack of robustness of the classical tolerance regions can be observed in the multivariate setting since they are constructed using the sample mean and the sample covariance matrix. Therefore, the content or coverage probability, $P_{\boldsymbol{\theta}}\,(\mathbf{x} \in \mathcal{R})$, and the size of the region will be strongly affected by anomalous observations. This phenomena is illustrated in Section 2.1 where the behavior of the classical tolerance region is studied when outliers or inliers are included. A robust plug–in tolerance region defined by replacing the classical estimators of location and scatter by affine equivariant robust Fisher–consistent estimators is then introduced in Section 3.

The value of the tolerance factors for the robust tolerance regions is derived through a Monte Carlo study, when the Donoho–Stahel estimators are used. The description of the algorithm used and the values obtained are given in Section 3.2. In Section 4.2 a simulation study is carried on to compare the behavior of the proposed region with that of the classical one for normal and contaminated training samples. Finally, in Section 5, the influence function of the coverage probability is computed. The influence function allows to understand the sensitivity of the classical procedure to anomalous data. On the other hand, it turns out to be bounded if the multivariate scatter estimator used has bounded influence. Proofs are given in the Appendix.

2 CLASSICAL TOLERANCE REGIONS

Given a training sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from a normal distribution, i.e, $\mathbf{x}_i \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and a new observation $\mathbf{x} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ independent of $\{\mathbf{x}_i\}$, for any $0 < q < 1$ and $0 < \delta < 1$, a tolerance region $\mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, based on the training sample, is the random region which covers with probability at least q the new observation $\mathbf{x}$ with a confidence level higher than δ. In the particular case of a normal distribution it can be written as

$$\mathcal{R} = \left\{\mathbf{y} : (\mathbf{y} - \overline{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{y} - \overline{\mathbf{x}}) \leq K\right\} \tag{1}$$

where $\overline{\mathbf{x}}$ and $\mathbf{S}$ are the sample mean and covariance matrix, respectively and the tolerance factor $K = K(q, \delta, n, d)$ depends on the size of the sample n, the dimension d, as well as the parameters q and δ.

The following result which will be used in the sequel and whose proof is straighforward shows that if we consider affine equivariant location and scatter estimates, to evaluate the tolerance factor K under elliptical distributions we can assume that the true location and scatter parameters are $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_d$. Thus, as is well known, for the classical tolerance region, one can assume that $\overline{\mathbf{x}} \sim N_d(\mathbf{0}, (1/n)\mathbf{I}_d)$ and $(n-1)\mathbf{S} \sim W_d(n-1, \mathbf{I}_d)$, where $W_d(m, \boldsymbol{\Sigma})$ denotes the $d-$dimensional Wishart distribution with m degree of freedom and scatter matrix $\boldsymbol{\Sigma}$.

Proposition 2.1. *Let $\mathbf{x}_i$, $1 \leq i \leq n$ be independent random vectors with elliptical distribution $P_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e., $\mathbf{C}^{-\frac{1}{2}}(\mathbf{x}_i - \boldsymbol{\mu}) = \mathbf{z}_i$ has an spherical distribution G with $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$. Assume that $\mathbf{t}_n = \mathbf{t}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{V}_n = \mathbf{V}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ are affine equivariant location and scatter estimators, respectively. Define the region*

$$\mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \left\{\mathbf{y} : (\mathbf{y} - \mathbf{t}_n)' \mathbf{V}_n^{-1} (\mathbf{y} - \mathbf{t}_n) \leq K\right\} .$$

Denote by K the tolerance factor related to $\mathcal{R}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$, i.e., the constant such that

$$p_n(\boldsymbol{\theta}) = P_{n,\boldsymbol{\theta}}\left[P_{\boldsymbol{\theta}}\left(\mathbf{x}\in\mathcal{R}(\mathbf{x}_1,\ldots,\mathbf{x}_n)|\mathbf{x}_1,\ldots,\mathbf{x}_n\right)\geq q\right]\geq\delta \quad \forall\,\boldsymbol{\theta}=(\boldsymbol{\mu},\boldsymbol{\Sigma})\ .$$

Then, $p_n(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$ and to compute K we can assume $\boldsymbol{\mu}=\mathbf{0}$ and $\boldsymbol{\Sigma}=\mathbf{I}_d$. More precisely, the tolerance factor is the constant K such that

$$P_{n,\boldsymbol{\theta}_0}\left[P_{\boldsymbol{\theta}_0}\left(\mathbf{x}\in\mathcal{R}(\mathbf{x}_1,\ldots,\mathbf{x}_n)|\mathbf{x}_1,\ldots,\mathbf{x}_n\right)\geq q\right]\geq\delta\ ,$$

where $\boldsymbol{\theta}_0=(\mathbf{0},\mathbf{I}_d)$.

Proof. Since the estimators are affine equivariant we have that $\mathbf{t}_n\left(\mathbf{x}_1,\ldots,\mathbf{x}_n\right) = \boldsymbol{\mu}+\mathbf{C}\,\mathbf{t}_n\left(\mathbf{z}_1,\ldots,\mathbf{z}_n\right)$ and $\mathbf{V}_n\left(\mathbf{x}_1,\ldots,\mathbf{x}_n\right)=\mathbf{C}\,\mathbf{t}_n\left(\mathbf{z}_1,\ldots,\mathbf{z}_n\right)\mathbf{C}'$. The result follows inmediately using that $\mathcal{R}(\mathbf{x}_1,\ldots,\mathbf{x}_n)=\boldsymbol{\mu}+\mathbf{C}\mathcal{R}(\mathbf{z}_1,\ldots,\mathbf{z}_n)$. □

Remark 2.1. Note that the conclusion of Proposition 2.1. still holds if the training sample and the new observation $\mathbf{x}$ have different elliptical distribution but the same location and scatter parameters.

As mentioned in the Introduction, the classical procedures are not distributionally robust since they are sensitive to even small deviations from normality. More precisely, as it will be shown just one outlying observation can modify the region obtained, due to changes in the estimation of the mean and of the covariance matrix, modifying the actual content. The question, is then if the constant q is still a reliable bound for the coverage probability of the tolerance region based on the sample mean $\overline{\mathbf{x}}$ and on the sample covariance matrix $\mathbf{S}$. In general, the answer is negative as it is in the univariate case (see, Butler(1982), Canavos and Koutraouvaelis (1984), Fernholz and Gillespie (2001)).

2.1 Sensitivity study of the classical tolerance regions

The goal of this Section is to show the sensitivity of the classical tolerance regions when the tolerance factor has been computed as the value satisfying

$$P_{n,(\mathbf{0},\mathbf{I}_d)}\left[P_{(\mathbf{0},\mathbf{I}_d)}\left(\left(\mathbf{x}-\overline{\mathbf{x}}\right)\mathbf{S}^{-1}\left(\mathbf{x}-\overline{\mathbf{x}}\right)\leq K\right)\geq q\right]\geq\delta \qquad (2)$$

with $\mathbf{x}\sim N_d\left(\mathbf{0},\mathbf{I}_d\right)$ and $\mathbf{x}_i\sim N_d\left(\mathbf{0},\mathbf{I}_d\right)$. To show the lack of robustness of the classical regions, in the multivariate setting, we have contaminated training samples of size $n=30$ with distribution $N\left(\mathbf{0},\mathbf{I}_d\right)$, by replacing one observation by an outlier $\mathbf{x}$ at distance $\|\mathbf{x}\|=2$, 4, 8 and 16 from the center of the distribution ($\mathbf{x}=\|\mathbf{x}\|\mathbf{e}_1$ with $\mathbf{e}_1$ the first vector of the canonical basis). We have performed the study in dimension $d=2$, 3, 4 and 5. For each of these samples, we have considered the region defined through (1) with the tolerance factor computed for normal samples with theoretical coverage $q=0.95$ and confidence level $\delta=0.95$.

The tolerance factors used were evaluated by simulation through the procedure described in Krishnamoorthy and Mathew (1999) and are given in Table 1. Through a Monte Carlo study, we have computed the actual coverage probability π of the region, when the training sample has distribution G and we use as location and scatter estimates $\mathbf{t}_n = \overline{\mathbf{x}}$ and $\mathbf{V}_n = \mathbf{S}$, as follows

(i) Generate n random vectors $\mathbf{x}_i \sim G$. Compute $\mathbf{t}_n$ and $\mathbf{V}_n$.

(ii) Generate R random vectors $\mathbf{y}_i \sim N_d(\mathbf{0}, \mathbf{I}_d)$ and for each of them verify if it lies or not at the region $\mathcal{R} = \left\{\mathbf{y} : (\mathbf{y} - \mathbf{t}_n)' \mathbf{V}_n^{-1} (\mathbf{y} - \mathbf{t}_n) \leq K\right\}$. If $\mathbf{y}_i \in \mathcal{R}$ define $c_i = 1$ elsewhere, $c_i = 0$. Compute the average content $\overline{c}$ as $\overline{c} = (1/R)\sum_{i=1}^{R} c_i$.

(iii) Repeat (i) and (ii) N times, preserving $\overline{c}$ at each step.

(iv) Sort the average contents $\overline{c}^{(1)} \leq \cdots \leq \overline{c}^{(N)}$ and we keep the $N(1-\delta)-$th one, $\overline{c}^{(N\,(1-\delta))} = \pi$, which approximates the actual coverage probability with confidence level δ.

We took $N = R = 1000$.

We have also computed the increase of the volume of the regions, $\mathcal{I}$, computed as % of the volume of the region for the uncontaminated distribution with respect to that of the contaminated one, i.e.,

$$\mathcal{I}^d = \frac{[\det(\mathbf{S})]^{\frac{1}{2}}}{[\det(\mathbf{S}_c)]^{\frac{1}{2}}}$$

where $\mathbf{S}_c$ denotes the sample covariance matrix evaluated over the contaminated sample. The results are shown in Table 2.

As it can be seen, the main problem is that the size of the regions increase due to an inflated estimated generalized variance ($\det(\mathbf{S})$). This explains the large values for the actual coverage probability with the contaminated sample, even when the region is centered far away from $\mathbf{0}$, the true center, as a result of the lack of robustness of the sample mean as location estimator.

To study the sensitivity to inliers we have contaminated $N(\mathbf{0}, \mathbf{I}_d)$ samples of size $n = 30$ in dimension $d = 2$, 3, 4 and 5, with 1, 2, 3, or 4 anomalous data located at zero. For each of these contaminated samples, we have considered the tolerance region defined through the tolerance factor related to $q = 0.95$ and $\delta = 0.95$ for normal data given in Table 1. Then, we evaluate the actual content asdescribed above and the increase in the volume of the region. The results are given in Table 3 which shows that including inliers lead to a reduction of the actual content as a consequence of the reduction in the volume of the region. Both effects are higher as the dimension increase.

These observations show the need of defining tolerance regions not so sensitive to a small amount of anomalous data.

3 ROBUST TOLERANCE REGIONS

3.1 Its definition

A "*plug–in*" approach can be used to define robust tolerance regions that will not be affected by the inclusion of some anomalous data. Such a procedure replaces the classical location and scatter estimates by robust counterparts. More precisely, let $\mathbf{t}_n = \mathbf{t}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{V}_n = \mathbf{V}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be robust location and scatter estimates. Define the region

$$\mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \left\{ \mathbf{y} : (\mathbf{y} - \mathbf{t}_n)' \, \mathbf{V}_n^{-1} (\mathbf{y} - \mathbf{t}_n) \leq K \right\} \ , \tag{3}$$

where the constant K is choosen such that

$$p_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = P_{n,\boldsymbol{\theta}} \left[P_{\boldsymbol{\theta}} (\mathbf{x} \in \mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_n) | \mathbf{x}_1, \ldots, \mathbf{x}_n) \geq q \right] \geq \delta \quad \forall \quad \boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \ ,$$

with $\mathbf{x} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{x}_i \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $1 \leq i \leq n$, independent, $P_{n,\boldsymbol{\theta}}$ the distribution of $(\mathbf{t}_n, \mathbf{V}_n)$ and $P_{\boldsymbol{\theta}}$ that of $\mathbf{x}$. We will then say that $\mathcal{R}$ is a *robust tolerance region*.

For the evaluation of the tolerance factor, the problem of the dependence of $p_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ on the unknown parameters araises. To avoid this problem and using Proposition 2.1, we must choose robust affine equivariant location and scatter estimates. Therefore, if $\mathbf{t}_n = \mathbf{t}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{V}_n = \mathbf{V}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ are robust affine equivariant location and scatter estimates, the tolerance factor K of the region $\mathcal{R}$ defined in (3) solves

$$P_{n,(\mathbf{0},\mathbf{I}_d)} \left[P_{(\mathbf{0},\mathbf{I}_d)} (\mathbf{x} \in \mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_n) | \mathbf{x}_1, \ldots, \mathbf{x}_n) \geq q \right] \geq \delta \, . \tag{4}$$

On the other hand, to compare the robust tolerance factors defined in (4) with the classical ones given in Table 4 from Krishnamoorthy and Mathew (1999), it is necessary that the robust location and scatter functionals will be Fisher–consistent at the multivariate normal distribution, i.e, if $\mathbf{x}_i \sim N_d(0, \mathbf{I}_d)$, $1 \leq i \leq n$, $\mathbf{V}_n \xrightarrow{c.t.p.} \mathbf{I}_d$ and $\mathbf{t}_n \xrightarrow{c.t.p.} \mathbf{0}$.

There are several proposals of robust location and scatter estimates. Possibles choices are the M–scatter estimate proposed by Maronna (1976), the minimum volume ellipsoid estimate (Rousseeuw and van Zomeren (1990)), the minimum covariance determinant (MCD, Rousseeuw (1985)), the Donoho (1982)–Stahel (1981) estimate, the S-, MM- and $\tau-$estimates (Lopuhaä (1990)). The main disadvantage of M–estimates, with monotone score functions, is that their breakdown point decreases with the dimension. All other proposals mentioned can achieve $\frac{1}{2}$ breakdown point. Among them, only the minimum volume ellipsoid estimate converges at a low rate, $n^{1/3}$, while the other ones have a root–n convergence rate. An overview of existing estimators of multivariate location and scatter is given in Maronna and Yohai (1998).

3.2 Computation of the tolerance factor for the robust tolerance regions

In this Section we will describe a method to compute the tolerance factor which is related to that given by Krishnamoorthy and Mathew (1999). It will be applied to the regions constructed using the Donoho–Stahel estimators.

A simulation algorithm similar to that introduced by Krishnamoorthy and Mathew (1999), was developped in MATLAB to approximate the value of the tolerance factor K for differewnt combinations of n, d, q and δ. This algorithm which is available upon request, is briefly described as follows

(i) For each j, generate n random vectors $\mathbf{x}_i \sim N_d(\mathbf{0}, \mathbf{I}_d)$. Evaluate robust affine equivariant location and scatter estimates $\mathbf{t}_n$ and $\mathbf{V}_n$.

(ii) Generate R random vectors $\mathbf{y}_i \sim N_d(\mathbf{0}, \mathbf{I}_d)$ and for each of them compute the Mahalanobis distance to the robust location according to $\mathbf{V}_n$, i.e., $DM_i = (\mathbf{y}_i - \mathbf{t}_n)' \mathbf{V}_n^{-1} (\mathbf{y}_i - \mathbf{t}_n)$.

(iii) Sort the quadratic forms $DM^{(1)} \leq \cdots \leq DM^{(R)}$ and search for the $q-$th percentile of the quadratic form $DM^{(R\,q)}$. Denote u_j this percentile.

(iv) Repeat (i) to (iii) N times, keeping the value u_j at each step.

(v) Sort the values u_j, $u^{(1)} \leq \cdots \leq u^{(N)}$ and search for the $\delta-$th percentile, $u^{(N\,\delta)}$ which gives an approximation to the tolerance factor K.

Tables 5 to 7 give the values of the tolerance factor K_{DS} satisfying (4) for different values of d, q, δ y n, when using the Donoho–Stahel estimator.

For the sake of completeness, we will remind the definition of the Donoho–Stahel estimates. Given a sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$, denote $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and consider all the univariate projections of the observations $\mathbf{a}'\mathbf{X}$, $\mathbf{a} \in I\!R^d$, $\mathbf{a} \neq \mathbf{0}$. Let $m(\cdot)$ and $s(\cdot)$ be univariate location and scale estimates, respectively. An outlyingness measure of $\mathbf{a}'\mathbf{x}_i$ is the standardized distance $\left|\mathbf{a}'\mathbf{x}_i - m\left(\mathbf{a}'\mathbf{X}\right)\right| / s\left(\mathbf{a}'\mathbf{X}\right)$.

The overall outlyingness measure of the point $\mathbf{x}_i \in I\!R^d$ is then defined as $r\left(\mathbf{x}_i, \mathbf{X}\right) = \sup_{a \in I\!R^d} \left|\mathbf{a}'\mathbf{x}_i - m\left(\mathbf{a}'\mathbf{X}\right)\right| / s\left(\mathbf{a}'\mathbf{X}\right)$. The Donoho–Stahel downweights each observation according to its outlyingness measure. The location estimate is then a weighted mean and the scatter matrix estimator is a weighted covariance matrix defined as

$$\mathbf{t}_n = \frac{\sum_{i=1}^{n} w_i\, \mathbf{x}_i}{\sum_{i=1}^{n} w_i}, \qquad \mathbf{V}_n = \beta\, \frac{\sum_{i=1}^{n} w_i\, \left(\mathbf{x}_i - \mathbf{t}_n\right) \left(\mathbf{x}_i - \mathbf{t}_n\right)'}{\sum_{i=1}^{n} w_i}, \tag{5}$$

where $w_i = w\left(r^2\left(\mathbf{x}_i, \mathbf{X}\right)\right)$ with w a non–negative and usually non–increasing weight function and β is a standardizing constant to obtain Fisher–consistency.

In order that $\mathbf{t}_n$ and $\mathbf{V}_n$ be affine equivariant, we have to assume that the univariate estimators m are s are equivariant. On the other hand, if we wish to obtain Fisher–consistent estimates at the $N\left(\mathbf{0}, \mathbf{I}_d\right)$ distribution, we have to calibrate the univariate scale such that $s(\Phi) = 1$ with Φ the distribution function of a $N(0,1)$ random variable and β must be choosen equal to $\beta = d\ E\left(w\left(W_d\right)\right)/E\left(w\left(W_d\right)W_d\right)$, with $W_d \sim \chi^2_d$. Table 4 gives the values of β in order to get Fisher-consistency for multivariate normal data corresponding to the Huber function, i.e., $w(t^2) = w_H\left(t\right) = \min(t^2, c^2)/t^2$ with $c = (\chi^2_{0.95,d})^{\frac{1}{2}}$.

The outlyingness measure cannot be computed exactly, then according to the proposal given in Stahel (1981) and studied in Maronna and Yohai (1995), the supremum is approximated taking the maximum over all the directions orthogonal to the linear spaces spanned by M subsamples of d points choosen among the n original ones. In dimension $d = 2$, it is possible to avoid resampling by maximizing over κ equally spaced directions (with angle $\ell \cdot 2\pi/\kappa$, $1 \leq \ell \leq \kappa$).

Using the Donoho–Stahel estimators we can define a robust tolerance region as described in Section 3.1. The region will be denoted as

$$\mathcal{R}_{\mathrm{DS}} = \left\{\mathbf{y} : \left(\mathbf{y} - \mathbf{t}_n\right)' \mathbf{V}_n^{-1} \left(\mathbf{y} - \mathbf{t}_n\right) \leq K_{\mathrm{DS}}\right\} \tag{6}$$

where $\mathbf{t}_n = \mathbf{t}_n\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right)$ and $\mathbf{V}_n = \mathbf{V}_n\left(\mathbf{x}_1, \ldots, \mathbf{x}_n\right)$ are given in (5).

Tables 5 to 7 give the values of the tolerance factor K_{DS} when the weights are computed using the Huber function. On the other hand, as location and scale estimates we have choosen the median of the observations $m(y_1, \ldots, y_n) = \underset{1\leq i\leq n}{\mathrm{median}}\,(y_i)$, and $s(y_1, \ldots, y_n) = (1/\Phi^{-1}(0.75))\,\mathrm{MAD}(y_1, \ldots, y_n)$, the median of the absolute deviations with respect to the median, calibrated such that it is Fisher–consistent when $y_i \sim N(0,1)$.

Larger regions are expected due to the smaller efficiency of the robust location and scatter estimates. To compare the volumes of both regions we calculate the $1/d$ root of the ratio between the robust region volume and the classical region volume. The comparisons for training samples of sizes $n = 20, 30, 50, 100$, dimensions $d = 2, 3, 4, 5, 8$, coverage probability $q = 0.95$ and confidence level $\delta = 0.95$ are given in Table 8. The calculations were carried out following steps i) to iv) for the computation of the actual coverage described in Section 2.1, obtained with $N = 1000$, $R = 1000$ and $M = 1000$.

It can be seen that the price paid for robustness is not excessive, except for extreme settings as $n = 20$ and $d = 8$, where the number of the observations does not allow to obtain a good approximation of the overall outlyingness measure.

4 MONTE CARLO STUDY

4.1 Sensitivity study of the robust tolerance region

In this section, we study the behavior of the robust regions when anomalous data are present in the training sample, in a similar way as it was done for the classical regions. We considered the regions $\mathcal{R}_{\mathrm{DS}}$ defined in (6) with the tolerance factors given in Tables 5, 6 and 7. As in for classical regions, we have contaminated training samples of size $n = 30$ with distribution $N(\mathbf{0}, \mathbf{I}_d)$, by replacing one observation by an outlier $\mathbf{x}$ at distance $\|\mathbf{x}\| = 2$, 4, 8 and 16 from the center of the distribution. The effect of adding this outlier is reported in Table 9, which is the counterpart of Table 2.

We see a mild increase in the volume of the tolerance region, joined with an increase of the coverage probability, as a result of the accurate centering of the region due to the robust estimate of location. In the case of the classical region the mild increase in coverage was the result of a huge increase in volume. The impact of the addition of inliers turned out to be more important in the robust tolerance region than in the classical one. This is reasonable because an inlier data being so central receives a higher weight than a typical data, reducing the estimated generalized variance through the sample weighted covariance matrix. Table 10 shows this phenomena.

4.2 Coverage study for the robust and classical region under alternative distributions

Both in the classical case as in the robust case the tolerance factors were computed with the aim of achieving actual coverage probability and actual confidence level equal to the theoretical values, when the training sample comes from a multivariate normal distribution. It seems interesting to analyze the behavior of these regions under alternative distributions G other than the normal, e.g., the multivariate t distribution with g degrees of freedom ($\mathcal{T}_g(d)$).

To study the actual coverage probability for the classical and robust regions when the training sample comes from alternative distributions (G), we used the algorithm of Section 2.1.

We will denote by π_{C} and π_{DS} the content of the classical and of the robust tolerance regions, respectively. The last one, being computed with the Donoho-Stahel estimators.

The results of actual coverage for the classical and robust regions, with $\delta = 0.95$ and $q = 0.95$, are shown in the Tables indicated between brackets for the following distributions G

- normal distribution with no contamination $G = N(\mathbf{0}, \mathbf{I}_d)$ (Table 8).
- multivariate t distribution with g degrees of freedom $G = \mathcal{T}_g(d)$, with $g = 1$, 2, 3 (Tables 11, 12 and 13, respectively).
- distribution $G = (1-\epsilon)\, N(\mathbf{0}, \mathbf{I}_d) + \epsilon\, \mathcal{C}_d$ with $\epsilon = 0.05$ y 0.10 (Tables 14 and 15, respectively).
- distribution $G = (1-\epsilon)\, N(\mathbf{0}, \mathbf{I}_d) + \epsilon\, N(\mathbf{0}, 25\, \mathbf{I}_d)$ with $\epsilon = 0.05$ y 0.10 (Tables 16 y 17, respectively).

We used training samples in dimensions $d = 2$, 3, 4, 5 y 8 and sizes $n = 20$, 30, 50 y 100.

To get a better insight of the difference of the two regions, and knowing that in the case of a Cauchy distribution the sample mean follows a Cauchy distribution too, we study the centering of the two regions. We do so by computing the ratio between the norm of the center of the classical region ($\|\overline{\mathbf{x}}\|$) and the norm of the center of the robust region ($\|\mathbf{t}_n\|$). We also show in the tables the $1/d$ root of the ratio of volumes of the two regions.

Table 11 shows that, for the Cauchy distribution, both regions the actual coverage is smaller than the theoretical coverage. Nevertheless the outstanding feature to point out is the remarkable increase in volume of the classical region related to the robust region, due to the lack of robustness of the estimates involved in the former region. The other important feature is the improvement in the centering of the robust region as a result of the robustness of the Stahel-Donoho estimate of location, as seen in the last column of table 11. Focusing on $d = 4$ and $n = 30$, we see, at a 95% confidence level that both regions cover less than 95% (79% for the classical region and 67% for the robust region). The bigger coverage of the classical region is paid with an increase of more than 81 times (3.05^4) the volume and its center almost 6 times further away from 0 than the center of the classical region.

Tables 12 and 13 give the results when the training samples are generated with distributions $\mathcal{T}_2(d)$ and $\mathcal{T}_3(d)$, respectively. As expected, increasing the degrees of freedom in the t distribution reduces the ratio of the volume of the classical region related to the volume of the robust region. The same effect is observed in the centering of both regions. This is so because when we increase the degrees of freedom we get closer to the normal distribution, where the opposite behavior occurs, though more moderate, as a result of the lower efficiency of the robust location and scatter estimates.

In a more realistic setting, in which just a minor proportion of the training sample is expected to be far form normality, we contaminate the training sample with a 5% and 10% of observations following a Cauchy distribution. Tables 14 and 15, show that,in most cases, the robust region is closer to the theoretical coverage with smaller volumes and a better centering. This phenomena becomes more important when we increase the dimension, the size of the sample and the proportion of contamination. The case $n = 20$ deserves special attention because the amount of contamination is not important enough as to reverse the effect of bigger volume and coverage probability of the robust region, already observed in Table 8. Similar conclusions are obtained when the contaminated observations come from a normal distribution, but with a bigger scatter matrix ($25\,\mathbf{I}_d$), (Tables 16 and 17, respectively).

5 INFLUENCE FUNCTION OF THE COVERAGE PROBABILITY

From now on, we will assume that the estimators, from which the tolerance regions are defined, are functionals over the space of distribution functions evaluated at the empirical distribution.

Influence functions are a measure of robustness with respect to single outliers. The importance of the influence function lies in its heuristic interpretation. It describes the effect of an infinitesimal contamination at a single point on the estimate, standarized by the amount of contamination and it is essentially the first derivative of the functional version of the estimator. Besides being of theoretical interest and helpful to calibrate the efficiency of the robust estimates measuring the influence of an observation on the classical estimates can be used as a diagnostic tool to detect influential observations.

We are interested in computing the influence function of the coverage probability (Pc), for a given confidence level δ fixed, a fixed theoretical coverage q, and a tolerance factor K, i.e,

$$\begin{aligned} Pc\,(G,F) &= P_F\left((\mathbf{x}-\mathbf{T}\,(G))'\,\mathbf{V}\,(G)^{-1}\,(\mathbf{x}-\mathbf{T}\,(G)) \leq K\right) \text{ with } \mathbf{x}\sim F \\ &= \int I_{\mathcal{R}(G)}\,(\mathbf{x})\,dF\,(\mathbf{x}) \end{aligned}$$

where

- $\mathcal{R}(G) = \{\mathbf{x} :\ (\mathbf{x}-\mathbf{T}\,(G))'\,\mathbf{V}\,(G)^{-1}\,(\mathbf{x}-\mathbf{T}\,(G)) \leq K\}$
- $\mathbf{T}\,(G)$ is the location functional at the distribution G,
- $\mathbf{V}\,(G)$ is the scatter functional at G and
- K is the tolerance factor, which is assumed to be fixed.

It is important to notice that this functional depend on two distributions, F and G. More precisely, the coverage probability $Pc\,(G,F)$ depends on the distribution with respect to which we evaluate the coverage F and on the distribution of the training sample G which allows to estimate the location and scatter parameters. We do not make assumptions on the distribution G, since the tolerance region can be evaluated for any data set, even if we expect that if we use the classical estimates the tolerance region will give poor results if the data distribution is far away from normality, as shown in Section 4. In many situations, as the one we have considered, it is assumed that $F = G = N_d(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Then, under the model, the observations from which the region is constructed, have a multivariate normal distribution and in that case, we will assume that $\mathbf{T}\,(G) = \boldsymbol{\mu}$, $\mathbf{V}\,(G) = \boldsymbol{\Sigma}$, which means that the functionals are Fisher–consistent. To evaluate the influence function we will only consider contaminations on the training sample G and not on the future data distribution F which allows to compute the coverage probability, since we are interested in knowing the effect on the coverage probability of deviations from normality in the training sample, which is the sample at hand. When, we evaluate the influence function we will assume, as in discrimination (see, Croux and Joossens (2004)) that $F = G$.

To simplify the computation of the influence function we will assume $F = N_d\,(\boldsymbol{\mu},\boldsymbol{\Sigma})$. The following Lemma gives an expression for the coverage probability, which is analogous to that given by Croux and Joossens (2004) for the total missclassification probability.

Lemma 5.1. *Let F be an ellipsoidal distribution with location parameter $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$. Let $\mathbf{T}$ be $\mathbf{V}$ multivariate location and sactter functionals. Assume that $\mathbf{x} = \boldsymbol{\mu} + \mathbf{C}\mathbf{z}$ where $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$ and $\mathbf{z} \sim F_0$ is spherically distributed. Denote $\mathbf{A}(G) = \mathbf{C}^{-1}\mathbf{V}(G)(\mathbf{C}^{-1})'$ and let us define $\boldsymbol{\beta}(G)$ as the matrix of eigenvectors of $\mathbf{A}(G)$, $\boldsymbol{\beta}(G)'\boldsymbol{\beta}(G) = \mathbf{I}_d$, $\boldsymbol{\Lambda}(G) = diag\left(\lambda_1(G), \ldots, \lambda_d(G)\right)$, the eigenvalues matrix, $\lambda_1(G) \geq \cdots \geq \lambda_d(G) \geq 0$, i.e, $\mathbf{A}(G) = \boldsymbol{\beta}(G)\boldsymbol{\Lambda}(G)\boldsymbol{\beta}(G)'$. Then,*

$$\begin{aligned} Pc(G,F) &= P_{F_0}\left((\mathbf{z}-\boldsymbol{\tau}(G))'\mathbf{A}(G)^{-1}(\mathbf{z}-\boldsymbol{\tau}(G)) \leq K\right) \\ &= P_{F_0}\left(\sum_{i=1}^{d}\left(\frac{z_i-\tau_i(G)}{\sqrt{\lambda_i(G)}}\right)^2 \leq K\right) \end{aligned} \tag{7}$$

where $\boldsymbol{\tau}(G) = \boldsymbol{\beta}(G)'\mathbf{C}^{-1}\left(\mathbf{T}(G) - \boldsymbol{\mu}\right)$.

In particular, if $F_0 = N_d\left(\mathbf{0}, \mathbf{I}_d\right)$, we have

$$Pc(G,F) = \int I_{\mathcal{S}}\left(\mathbf{y}\right)\prod_{i=1}^{d}\sqrt{\lambda_i\left(G\right)}\,\varphi\left(\sqrt{\lambda_i\left(G\right)}\,y_i + \tau_i\left(G\right)\right)\,d\mathbf{y}. \tag{8}$$

with $\mathcal{S} = \{\mathbf{y} : \sum_{i=1}^{d} y_i^2 \leq K\}$ and $\varphi(t)$ denotes the denisty of a random variable with distribution $N(0,1)$.

Using Hampel's (1974) definition, we have that the influence function for the functional Pc at the point $\mathbf{x}$ and the distribution G is defined as

$$\begin{aligned} \mathrm{IF}\left(\mathbf{x}, Pc, G\right) &= \lim_{\epsilon \to 0}\frac{Pc\left(\left(1-\epsilon\right)G + \epsilon\Delta_{\mathbf{x}}, F\right) - Pc\left(G,F\right)}{\epsilon} \\ &= \left.\frac{\partial}{\partial\epsilon}Pc\left(G_{\epsilon,\mathbf{x}}, F\right)\right|_{\epsilon=0} \end{aligned}$$

where, from now on, $G_{\epsilon,\mathbf{x}} = \left(1-\epsilon\right)G + \epsilon\Delta_{\mathbf{x}}$ denotes the distribution contaminated with a point mass $\Delta_{\mathbf{x}}$ at $\mathbf{x}$.

We will evaluate the influence function in two different situations. In the first one, the distributions G and F are such that the eigenvalues of $\mathbf{A}(G)$ are all different. In the second one, $F = G$ and $\mathbf{T}$ and $\mathbf{V}$ are Fisher–consistentes functionals, i.e., $\mathbf{T}(G) = \boldsymbol{\mu}$ y $\mathbf{V}(G) = \boldsymbol{\Sigma}$. In this case, $\boldsymbol{\tau}(G) = \mathbf{0}$ and $\mathbf{A}(G) = \mathbf{I}_d$. Thus, there is a problem to compute the influence function of the eigenvectors of $\mathbf{A}(G)$ at $G = F$ since they are not uniquely defined. To overcome this problem, we define the eigenvalues of $\mathbf{A}(G_{\epsilon,\mathbf{x}})$, for small ϵ, such that the diagonal elements of the matrix $\boldsymbol{\beta}(G_{\epsilon,\mathbf{x}})$ are positive. This ensures the continuity at $G = F$ of $\boldsymbol{\beta}(G)$.

Theorem 5.1. *Let* $\mathbf{T}$ *and* $\mathbf{V}$ *multivariate location and scatter functionals. Let* $F = N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *with* $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$ *and* G *be such that the eigenvalues,* $\lambda_1 \geq \cdots \geq \lambda_d$, *of* $\mathbf{A}(G) = \mathbf{C}^{-1}\mathbf{V}(G)(\mathbf{C}^{-1})'$ *are different. Let* $\boldsymbol{\beta}_j$ *be the eigenvector related to* λ_j. *Denote* $\boldsymbol{\tau} = \boldsymbol{\tau}(G) = \boldsymbol{\beta}'\mathbf{C}^{-1}(\mathbf{T}(G) - \boldsymbol{\mu})$ *and* F_1 *The distribution of* $\mathbf{y} = (y_1, \ldots, y_d)'$, *where* $y_1, \ldots, y_d$ *are independent and such that* $y_i \sim N\left(-\tau_i(G)\lambda_i(G)^{-\frac{1}{2}}, \lambda_i(G)^{-1}\right)$ *for* $1 \leq i \leq d$.

Assume that $IF(\mathbf{x}, \mathbf{T}, G)$ *and* $IF(\mathbf{x}, \mathbf{V}, G)$ *exist. Then, the influence function of* $Pc(G)$ *is given by*

$$\begin{aligned} IF(x, Pc, G) &= \sum_{j=1}^{d} \boldsymbol{\beta}_j' \mathbf{C}^{-1} IF(\mathbf{x}, \mathbf{V}, G)(\mathbf{C}^{-1})'\boldsymbol{\beta}_j \times \\ &\quad \times \frac{1}{2\lambda_j}\left[P_{F_1}(\mathcal{S}) - E_{F_1}\left(I_{\mathcal{S}}(\mathbf{y})y_j\left(\lambda_j\, y_j + \sqrt{\lambda_j}\tau_j\right)\right)\right] - \\ &\quad - \sum_{j=1}^{d}\left(\sqrt{\lambda_j}E_{F_1}(y_i I_{\mathcal{S}}(\mathbf{y})) + \tau_j P_{F_1}(\mathcal{S})\right) \times \\ &\quad \times \left(\sum_{i \neq j} \frac{\boldsymbol{\beta}_i' \mathbf{C}^{-1} IF(\mathbf{x}, \mathbf{V}, G)(\mathbf{C}^{-1})'\boldsymbol{\beta}_j}{\lambda_j - \lambda_i}\,\tau_i + \boldsymbol{\beta}_j'\mathbf{C}^{-1} IF(\mathbf{x}, \mathbf{T}, G)\right). \end{aligned}$$

Reamrk 5.1. When $\mathbf{T}(G) = \boldsymbol{\mu}$, $\tau_i = 0$ for all $1 \leq i \leq d$, and then

$$\mathrm{IF}(x, Pc, G) = \sum_{j=1}^{d} \boldsymbol{\beta}_j' \mathbf{C}^{-1} \mathrm{IF}(\mathbf{x}, \mathbf{V}, G)(\mathbf{C}^{-1})'\boldsymbol{\beta}_j \,\frac{P_{F_1}(\mathcal{S}) - \lambda_j E_{F_1}\left(I_{\mathcal{S}}(\mathbf{y})y_j^2\right)}{2\lambda_j}.$$

Therefore, the influence function of the coverage probability does not depend on the influence function of the location functional nor on the expression related to the influence function of the eigenvectors. The problem of having close eigenvalues seems to be overcome. This will be the expression for the influence function which will be derived in Theorem 5.2 which gives the influence function when $G = F$. In particular, if $\boldsymbol{\Sigma} = \mathbf{I}_d$, $\mathbf{T}(G) = \boldsymbol{\mu}$ and $\mathbf{V}(G) = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$, with $\lambda_1 > \cdots > \lambda_d$, we have

$$\mathrm{IF}(x, Pc, G) = \sum_{j=1}^{d} \mathrm{IF}(\mathbf{x}, \mathbf{V}, G)_{jj}\,\frac{P_{F_1}(\mathcal{S}) - \lambda_j E_{F_1}\left(I_{\mathcal{S}}(\mathbf{y})y_j^2\right)}{2\lambda_j}.$$

Theorem 5.2. *Let* $F = G = N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *with* $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$. *Let* $\mathbf{T}$ *and* $\mathbf{V}$ *be Fisher–consistent functionals at* G, *i.e.,* $\mathbf{T}(G) = \boldsymbol{\mu}$ *and* $\mathbf{V}(G) = \boldsymbol{\Sigma}$.

Assume that $IF(\mathbf{x}, \mathbf{T}, G)$ *and* $IF(\mathbf{x}, \mathbf{V}, G)$ *exist. Thn, the influence function of* $Pc(G)$ *is given by*

$$\begin{aligned} IF(x, Pc, G) &= \frac{1}{2}\left(P(W_d \leq K) - \frac{1}{d}E\left(W_d I_{(0,K]}(W_d)\right)\right) tr\left(IF(\mathbf{x}, \mathbf{V}, G)\,\boldsymbol{\Sigma}^{-1}\right) \\ &= \frac{1}{2}\left(P(W_d \leq K) - P(W_{d+2} \leq K)\right) tr\left(IF(\mathbf{x}, \mathbf{V}, G)\,\boldsymbol{\Sigma}^{-1}\right), \end{aligned}$$

where W_d denotes a random variable with distribution χ^2_d.

In particular, if $\mathbf{\Sigma} = \mathbf{I}_d$ it holds that

$$IF\,(x, Pc, G) \quad = \quad \frac{1}{2}\left(P\left(W_d \leq K\right) - P\left(W_{d+2} \leq K\right)\right) tr\left(IF\left(\mathbf{x}, \mathbf{V}, G\right)\right) \ .$$

Theorem 5.2 shows that the coverage probability will have bounded influence function if the scatter functional has bounded influence.

Corollary 5.1. *Let $F = G = N_d\left(\boldsymbol{\mu}, \mathbf{\Sigma}\right)$, $\mathbf{T}(G) = \int \mathbf{x} \ dG$ and $\mathbf{V}(G) = \int \left(\mathbf{x} - \mathbf{T}\left(G\right)\right) \left(\mathbf{x} - \mathbf{T}\left(G\right)\right)' \ dG$ be the classical location and scatter functionals. Then, the influence function of $Pc(G, F)$ is given by*

$$IF\,(x, Pc, G) \quad = \quad \frac{1}{2}\, c_d \ \left(\left(\mathbf{x} - \boldsymbol{\mu}\right)' \mathbf{\Sigma}^{-1} \left(\mathbf{x} - \boldsymbol{\mu}\right) - d\right) \ ,$$

where $c_d = P\left(W_d \leq K\right) - P\left(W_{d+2} \leq K\right)$ with W_d a random variable with χ^2_d distribution.

From the expression of the influence function, we can conclude that the influence function of the coverage probability of the classical region only depends on $\mathbf{x}$ through the Mahalanobis norm, being an increasing function of it. With respect to the constant K, the larger K the closer to 0 is the constant c_d, and so the influence function. Therefore, fixing K, the influence function is unbounded showing the sensitivity to outliers of the classical procedure.

Figure 1.(a) shows the influence of the coverage probability of the classical region when $d = 2$ for 5 tolerance factors ($K = 2$, 4, 6, 8 and 10). It is worth noticing that points with Mahalanobis norm equal to $d^{\frac{1}{2}}$ do not have any influence on the coverage probability, while contaminations with norm smaller than $d^{\frac{1}{2}}$ have a negative influence and those with norm larger than $d^{\frac{1}{2}}$ have a positive one. We can then conclude that the impact of an inlier will be a reduction on the coverage probability, corresponding $\mathbf{x} = \boldsymbol{\mu}$, where the function takes the value $-\frac{1}{2}\ d\ c_d$, to the greatest effect. On the other hand, the impact of an outlier produces an unbounded increase of the coverage probability.

For general scatter functionals, the influence function when $F = G = N_d\left(\boldsymbol{\mu}, \mathbf{\Sigma}\right)$, can be derived from Theorem 1 and Lemma 1 in Croux and Haesbroek (2000) which gives an expression for the influence function of a robust scatter functional. According to it, if $G = N(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $D^2(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$, is the Mahalanobis distance, then, for any affine equivariant scatter functional, $\mathbf{V}(G)$, such that its influence function exists, there exist two functions $\alpha_{\mathbf{V}}$ and $\gamma_{\mathbf{V}} : [0, \infty) \to I\!R$ such that $\text{IF}\left(\mathbf{x}, \mathbf{V}, G\right) = \alpha_{\mathbf{V}}\left(D(\mathbf{x})\right)(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' - \gamma_{\mathbf{V}}\left(D(\mathbf{x})\right)\mathbf{\Sigma}$. From this property, we obtain the following result.

Corollary 5.2. *Let $F = G = N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Denote* $\mathbf{T}$ *and* $\mathbf{V}$ *robust affine equivariant location and scatter functionals Fisher–consistents at G, i.e.,* $\mathbf{T}(G) = \boldsymbol{\mu}$ *and* $\mathbf{V}(G) = \boldsymbol{\Sigma}$*. Then, the influence function of $Pc(G, F)$ is given by*

$$IF(x, Pc, G) \;=\; \frac{1}{2}\, c_d\, \left[\alpha_{\mathbf{V}}\left(D(\mathbf{x})\right) D^2(\mathbf{x}) - d\,\gamma_{\mathbf{V}}\left(D(\mathbf{x})\right)\right] ,$$

where $c_d = P(W_d \le K) - P(W_{d+2} \le K)$ with W_d a random variable with χ^2_d distribution.

A review on multivariate location and scatter estimators can be found in Maronna and Yohai (1998). Among them, we can mentioned the $S-$scatter estimator (Lopuhaä (1990)). Thus, using S–estimators, the coverage probability will have bounded influence if the score function ρ and $\eta(t) = t\rho\prime(t)$ are bounded functions, which is a usual requirement to obtain robust scatter estimators with positive breakdown point. Figure 1.(c) shows the influence function of the coverage probability when an S–estimator estimator is used in dimension $d = 2$ with 5 different tolerance factors ($K = 2$, 4, 6, 8 and 10). The score function ρ was taken as the biweight Tukey's function calibrated to attain 25% breakdown point for the S–scatter estimator. As it can be seen the influence function is bounded.

The Donoho–Stahel estimator used in the previous Sections is a high–breakdown point estimator and its influence function was obtained by Gervini (2002). The following result gives the influence function of $Pc(G, F)$ when we use the median and the MAD as the univariate location and scale functionals.

Corollary 5.3. *Let $F = G = N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let* $\mathbf{T}$ *and* $\mathbf{V}$ *be the robust affine equivariant functionals related to the Donoho–Stahel estimators, such that* $\mathbf{T}(G) = \boldsymbol{\mu}$ *and* $\mathbf{V}(G) = \boldsymbol{\Sigma}$*. Assume that the following conditions hold*

W1. *$w : [0, \infty) \to [0, \infty)$ and $w(u^2)u^2$ are bounded functions.*

W2. *w is differentiable almost everywhere and $\eta(u^2) = w'(u^2)u^4$ is bounded.*

Then, if the univariate location functional is taken as the median m and $s(\cdot) = (1/\Phi^{-1}(0.75))\,\mathrm{MAD}(\cdot)$ is the univariate scale functional, the influence function of $Pc(G, F)$ is given by

$$\begin{aligned}
IF(\boldsymbol{\mu}, Pc, G) &= 0 \\
IF(\mathbf{x}, Pc, G) &= \beta\,\frac{1}{2}\,c_d \left[\frac{c_1}{c_0} g(D(\mathbf{x})) + \frac{w\left(D^2(\mathbf{x})\right) D^2(\mathbf{x}) - c_2}{c_0}\right] \quad \text{if} \quad \mathbf{x} \neq \boldsymbol{\mu} \\
&= \frac{d}{2\,c_2}\,c_d\,\left[c_1\,g(D(\mathbf{x})) + w\left(D^2(\mathbf{x})\right) D^2(\mathbf{x}) - c_2\right] \quad \text{if} \quad \mathbf{x} \neq \boldsymbol{\mu}\,,
\end{aligned}$$

with $D^2(\mathbf{x}) = (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$,

$$\begin{aligned}
\beta &= d\,\frac{c_0}{c_2} \\
c_d &= P\left(W_d \le K\right) - P\left(W_{d+2} \le K\right) \\
c_0 &= E\left(w(W_d)\right) \\
c_1 &= -2E\left(w'(W_d)W_d^2\right) = -2E\left(\eta(W_d)\right) \\
c_2 &= E\left(w(W_d)W_d\right) \\
g(t) &= \frac{0.5 - F_{\mathcal{B}\left(\frac{1}{2},\frac{d-1}{2}\right)}\left(\frac{\left[\Phi^{-1}(0.75)\right]^2}{t^2}\right)}{2\Phi^{-1}(0.75)\,\varphi\left(\Phi^{-1}(0.75)\right)}\,,
\end{aligned}$$

where $W_d \sim \chi^2_d$, φ *denotes the denisty of a random variable with distribution* $N(0,1)$ *and* $\mathcal{B}\left(1/2,(d-1)/2\right)$ *stands for the Beta distribution with parameters* $1/2$ *and* $(d-1)/2$.

A similar result can be obtained for general univariate location and scale functionals using Theorem 3 in Gervini (2002).

Condition **W1** ensures that the influence function of the covergae probability is bounded. In particular, when $w(t) = w_H\left(t^{\frac{1}{2}}\right)$, we have $w'(t) = w'_H\left(t^{\frac{1}{2}}\right)/(2t^{\frac{1}{2}})$. Then, if $d \neq 2$

$$\begin{aligned}
c_0 &= P\left(W_d < c^2\right) + c^2\, E\left(\frac{1}{W_d} I_{(c^2,\infty)}(W_d)\right) \\
&= P\left(W_d < c^2\right) + c^2\,\frac{1}{2\,(d-2)}\,\left(1 - P\left(W_{d-2} < c^2\right)\right) \\
c_1 &= 2\,c^2\, E\left(I_{(c^2,\infty)}(W_d)\right) = 2\,c^2\,\left(1 - P\left(W_d < c^2\right)\right) \\
c_2 &= E\left(W_d I_{(0,c^2)}(W_d)\right) + c^2\, E\left(I_{(c^2,\infty)}(W_d)\right) \\
&= d\, P\left(W_{d+2} < c^2\right) + c^2\,\left(1 - P\left(W_d < c^2\right)\right)
\end{aligned}$$

with $c = (\chi^2_{0.95,d})^{\frac{1}{2}}$.

Figure 1.(b) shows the influence function of the coverage probability when the Donoho–Stahel estimators are used in dimension $d = 2$ with 5 different tolerance factors ($K = 2$, 4, 6, 8 and 10). As it can be seen the influence function is bounded, with a discontinuity at 0 due to the discontinuity of the influence of the univariate funcionals. Effectively, the influence at $\mathbf{x} = \boldsymbol{\mu}$ is 0 but

$$\lim_{\mathbf{x}\to\boldsymbol{\mu}} \mathrm{IF}\left(\mathbf{x}, Pc, G\right) = -\,\frac{d}{2\,c_2}\,c_d\,\left[c_1\,\frac{1}{4\Phi^{-1}(0.75)\,\varphi\left(\Phi^{-1}(0.75)\right)} + c_2\right]$$

showing the discontinuity at $\mathbf{0}$.

Since the scale in Figure 1 does not allow to distinguish the effect of inliers on the classical estimator, in Figure 2 we have plotted the influence function on a reduce range $0 \leq D(\mathbf{x}) \leq 2$, which allows to describe that effect. As shown by Figure 2, the three estimators are sensitive to this king of anomalous data, even if the effect of them is bounded and is larger for the robust estimators as observed in Section 4.1 for the Donoho–Stahel estimator.

It is worth noticing that the diagnostic measure related to the coverage probability is defined as

$$DM(\mathbf{x}) = \frac{1}{2}\, c_d \, \left((\mathbf{x}-\widehat{\boldsymbol{\mu}})' \, \widehat{\boldsymbol{\Sigma}}^{-1} \, (\mathbf{x}-\widehat{\boldsymbol{\mu}}) - d \right) \, ,$$

where $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are robust location and scatter estimators. This measure equals except for a constant the usual diagnostic measure used to detect anomalous multivariate data and is simply a robust version of the Mahalanobis distance considered in Rousseeuw and van Zomeren (1990).

AKNOWLEDGEMENTS

This research was partially supported by Grant PICT # 03-00000-006277 from ANPCYT, Grant X-094 from the Universidad de Buenos Aires and by a grant of the Fundación Antorchas at Buenos Aires, Argentina. The research of Andrés Farall was supported by a scholarship of the Fundación Antorchas.

6 APPENDIX: PROOFS

Proof of Lemma 5.1. Since $\mathbf{x}=\boldsymbol{\mu}+\mathbf{C}\mathbf{z}$ and $\mathbf{A}(G)=\mathbf{C}^{-1}\mathbf{V}(G)(\mathbf{C}^{-1})'=\boldsymbol{\beta}(G)\boldsymbol{\Lambda}(G)\boldsymbol{\beta}(G)'$ it follows easily that

$$\begin{aligned} Pc(G,F) &= P_F\left((\mathbf{x}-\mathbf{T}(G))'\mathbf{V}(G)^{-1}(\mathbf{x}-\mathbf{T}(G)) \leq K\right) \\ &= P_{F_0}\left((\mathbf{z}-\mathbf{C}^{-1}(\mathbf{T}(G)-\boldsymbol{\mu}))'\mathbf{A}(G)^{-1}(\mathbf{z}-\mathbf{C}^{-1}(\mathbf{T}(G)-\boldsymbol{\mu})) \leq K\right) \\ &= P_{F_0}\left((\boldsymbol{\beta}(G)'\mathbf{z}-\boldsymbol{\tau}(G))'\boldsymbol{\Lambda}(G)^{-1}(\boldsymbol{\beta}(G)'\mathbf{z}-\boldsymbol{\tau}(G)) \leq K\right) \, . \end{aligned}$$

Using that F_0 is an spherical distribution we get that $\boldsymbol{\beta}(G)'\mathbf{z}$ has the same distribution as $\mathbf{z}$ then

$$\begin{aligned} Pc(G,F) &= P_{F_0}\left((\mathbf{z}-\boldsymbol{\tau}(G))'\boldsymbol{\Lambda}(G)^{-1}(\mathbf{z}-\boldsymbol{\tau}(G)) \leq K\right) \\ &= P_{F_0}\left(\sum_{i=1}^{d}\frac{1}{\lambda_i\,(G)}\,(z_i-\tau_i\,(G))^2 \leq K\right) \\ &= P_{F_0}\left(\sum_{i=1}^{d}\left(\frac{z_i-\tau_i\,(G)}{\sqrt{\lambda_i\,(G)}}\right)^2 \leq K\right) \, . \end{aligned}$$

If $F_0=N_d\,(\mathbf{0},\mathbf{I}_d)$, let $y_i=(z_i-\tau_i\,(G))\,\lambda_i\,(G)^{-\frac{1}{2}}$, then $y_i \sim N\left(-\tau_i\,(G)\lambda_i\,(G)^{-\frac{1}{2}}, \lambda_i\,(G)^{-1}\right)$ and y_i are independent. Denote F_1

the distribution of $\mathbf{y}$. Using (7), we obtain

$$\begin{aligned} Pc\,(G,F) &= P_{F_1}\left(\sum_{i=1}^{d} y_i^2 \le K\right) = E_{F_1}\left(I_{\mathcal{S}}\left(\mathbf{y}\right)\right) \\ &= \int I_{\mathcal{S}}\left(\mathbf{y}\right) \prod_{i=1}^{d} f_{Y_i}\left(y_i\right)\, d\mathbf{y} \\ &= \int I_{\mathcal{S}}\left(\mathbf{y}\right) \prod_{i=1}^{d} \sqrt{\lambda_i\left(G\right)}\, \varphi\left(\sqrt{\lambda_i\left(G\right)}\, y_i + \tau_i\left(G\right)\right)\, d\mathbf{y}\ .\ \square \end{aligned}$$

To prove Theorems 5.1 and 5.2, we will need the following Lemma.

Lemma 6.1. *Let* $\mathbf{\Lambda} = diag\left(\lambda_1, \ldots, \lambda_d\right)$ *and* $\boldsymbol{\tau} = \left(\tau_1, \ldots, \tau_d\right)'$. *Denote*

$$H(\mathbf{\Lambda}, \tau) = \int I_{\mathcal{S}}\left(\mathbf{y}\right) \prod_{i=1}^{d} \sqrt{\lambda_i}\, \varphi\left(\sqrt{\lambda_i}\, y_i + \tau_i\right)\, d\mathbf{y}$$

where φ *stands for the standard gaussian density and* $\mathcal{S} = \{\mathbf{y} :\ \|\mathbf{y}\|^2 \le K\}$. *Let* F_1 *be the distribution of* $\mathbf{y} = (y_1, \ldots, y_d)'$, *where* $y_1, \ldots, y_d$ *are independent random variables such that* $y_i \sim N\left(-\tau_i\left(G\right)\lambda_i\left(G\right)^{-\frac{1}{2}}, \lambda_i\left(G\right)^{-1}\right)$, $1 \le i \le d$. *Then, the function* H *is differentiable and*

$$\begin{aligned} \frac{\partial H}{\partial \tau_i} &= -\left[\sqrt{\lambda_i}\, E_{F_1}\left(I_{\mathcal{S}}(\mathbf{y}) y_i\right) + \tau_i P_{F_1}\left(\mathcal{S}\right)\right] \\ \frac{\partial H}{\partial \lambda_i} &= \frac{1}{2\lambda_i}\left[P_{F_1}\left(\mathcal{S}\right) - E_{F_1}\left(I_{\mathcal{S}}(\mathbf{y}) y_i\left(\lambda_i y_i + \sqrt{\lambda_i}\, \tau_i\right)\right)\right]\ . \end{aligned}$$

Proof. The proof follows the same ideas as the proof of Lemma 1 in Croux and Joossens (2004). Using that $\varphi'(t) = -t\varphi(t)$, the definition of H and the fact that we can differentiate under the integral we get

$$\begin{aligned} \frac{\partial H}{\partial \tau_i} &= \int I_{\mathcal{S}}\left(\mathbf{y}\right) \prod_{j\neq i} \sqrt{\lambda_j}\, \varphi\left(\sqrt{\lambda_j}\, y_j + \tau_j\right) \sqrt{\lambda_i}\, \frac{\partial \varphi\left(\sqrt{\lambda_i}\, y_i + \tau_i\right)}{\partial \tau_i}\, d\mathbf{y} \\ &= -\int I_{\mathcal{S}}\left(\mathbf{y}\right) \prod_{j=1}^{d} \sqrt{\lambda_j}\, \varphi\left(\sqrt{\lambda_j}\, y_j + \tau_j\right) \left(\sqrt{\lambda_i}\, y_i + \tau_i\right)\, d\mathbf{y} \\ &= -\int I_{\mathcal{S}}\left(\mathbf{y}\right) \left(\sqrt{\lambda_i}\, y_i + \tau_i\right) f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}\ . \end{aligned}$$

Besides,

$$\begin{aligned} \frac{\partial H}{\partial \lambda_i} &= \int I_{\mathcal{S}}\left(\mathbf{y}\right) \prod_{j\neq i} \sqrt{\lambda_j}\, \varphi\left(\sqrt{\lambda_j}\, y_j + \tau_j\right) \frac{\partial\left\{\sqrt{\lambda_i}\, \varphi\left(\sqrt{\lambda_i}\, y_i + \tau_i\right)\right\}}{\partial \lambda_i}\, d\mathbf{y} \\ &= \int I_{\mathcal{S}}\left(\mathbf{y}\right) \prod_{j=1}^{d} \sqrt{\lambda_j}\, \varphi\left(\sqrt{\lambda_j}\, y_j + \tau_j\right) \left[\frac{1}{2\lambda_i} - \frac{1}{2\lambda_i}\left(\lambda_i\, y_i + \sqrt{\lambda_i}\tau_i\right) y_i\right]\, d\mathbf{y} \\ &= \frac{1}{2\lambda_i}\left[P_{F_1}(\mathcal{S}) - \int I_{\mathcal{S}}\left(\mathbf{y}\right) y_i \left(\lambda_i\, y_i + \sqrt{\lambda_i}\tau_i\right) f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}\right]\ , \end{aligned}$$

which concludes the proof. □

PROOF OF THEOREM 5.1. To evaluate the influence function we will use (8). Since $Pc(G,F) = H(\mathbf{\Lambda}(G), \tau(G))$, using the chain rule we obtain

$$\begin{aligned}
\text{IF}\,(x, Pc, G) &= \left.\frac{\partial Pc\,(G_{\epsilon,\mathbf{x}}, F)}{\partial \epsilon}\right|_{\epsilon=0} \\
&= \sum_{i=1}^{d} \left.\frac{\partial H}{\partial \lambda_i}\right|_{(\mathbf{\Lambda},\boldsymbol{\tau})} \left.\frac{\partial \lambda_i\,(G_{\epsilon,\mathbf{x}})}{\partial \epsilon}\right|_{\epsilon=0} \\
&+ \sum_{i=1}^{d} \left.\frac{\partial H}{\partial \tau_i}\right|_{(\mathbf{\Lambda},\boldsymbol{\tau})} \left.\frac{\partial \tau_i\,(G_{\epsilon,\mathbf{x}})}{\partial \epsilon}\right|_{\epsilon=0}, \qquad \text{(A.1)}
\end{aligned}$$

where $(\mathbf{\Lambda}, \boldsymbol{\tau}) = (\mathbf{\Lambda}(G), \boldsymbol{\tau}(G))$. Since the eigenvalues of $\mathbf{A}(G) = \mathbf{C}^{-1}\mathbf{V}(G)(\mathbf{C}^{-1})'$ have multiplicity 1, Lemma 3 in Croux and Haesbroeck (2000) entails

$$\begin{aligned}
\text{IF}\left(\mathbf{x}, \boldsymbol{\beta}_j, G\right) &= \sum_{i \neq j} \frac{\boldsymbol{\beta}_i' \text{IF}\,(\mathbf{x}, \mathbf{A}, G)\, \boldsymbol{\beta}_j}{\lambda_j - \lambda_i}\, \boldsymbol{\beta}_i \\
&= \sum_{i \neq j} \frac{\boldsymbol{\beta}_i' \mathbf{C}^{-1} \text{IF}\,(\mathbf{x}, \mathbf{V}, G)\, (\mathbf{C}^{-1})' \boldsymbol{\beta}_j}{\lambda_j - \lambda_i}\, \boldsymbol{\beta}_i \qquad \text{(A.2)} \\
\text{IF}\,(\mathbf{x}, \lambda_j, G) &= \boldsymbol{\beta}_j' \text{IF}\,(\mathbf{x}, \mathbf{A}, G)\, \boldsymbol{\beta}_j \\
&= \boldsymbol{\beta}_j' \mathbf{C}^{-1} \text{IF}\,(\mathbf{x}, \mathbf{V}, G)\, (\mathbf{C}^{-1})' \boldsymbol{\beta}_j\,. \qquad \text{(A.3)}
\end{aligned}$$

Thus, using that $\tau_j\,(G_{\epsilon,\mathbf{x}}) = \boldsymbol{\beta}_j\,(G_{\epsilon,\mathbf{x}})'\, \mathbf{C}^{-1}\,(\mathbf{T}\,(G_{\epsilon,\mathbf{x}}) - \boldsymbol{\mu})$, where $\boldsymbol{\beta}_j\,(G_{\epsilon,\mathbf{x}})$ is the $j-$th eigenvector of the matrix $\mathbf{A}\,(G_{\epsilon,\mathbf{x}})$, we get

$$\begin{aligned}
\left.\frac{\partial \tau_j\,(G_{\epsilon,\mathbf{x}})}{\partial \epsilon}\right|_{\epsilon=0} &= \left.\frac{\partial}{\partial \epsilon} \left\{\boldsymbol{\beta}_j\,(G_{\epsilon,\mathbf{x}})'\, \mathbf{C}^{-1}\,(\mathbf{T}\,(G_{\epsilon,\mathbf{x}}) - \boldsymbol{\mu})\right\}\right|_{\epsilon=0} \\
&= \text{IF}\left(\mathbf{x}, \boldsymbol{\beta}_j, G\right)'\, \mathbf{C}^{-1}\,(\mathbf{T}\,(G) - \boldsymbol{\mu}) + \boldsymbol{\beta}_j'\, \mathbf{C}^{-1} \text{IF}\,(\mathbf{x}, \mathbf{T}, G) \\
&= \sum_{i \neq j} \frac{\boldsymbol{\beta}_i' \mathbf{C}^{-1} \text{IF}\,(\mathbf{x}, \mathbf{V}, G)\, (\mathbf{C}^{-1})' \boldsymbol{\beta}_j}{\lambda_j - \lambda_i} \tau_i \\
&+ \boldsymbol{\beta}_j'\, \mathbf{C}^{-1} \text{IF}\,(\mathbf{x}, \mathbf{T}, G)\;. \qquad \text{(A.4)}
\end{aligned}$$

Therefore, using (A.3), (A.4), from (A.1) we obtain

$$\begin{aligned}
\text{IF}\,(\mathbf{x}, Pc, G) &= \sum_{j=1}^{d} \left.\frac{\partial H}{\partial \lambda_j}\right|_{(\mathbf{\Lambda},\boldsymbol{\tau})} \boldsymbol{\beta}_j' \mathbf{C}^{-1} \text{IF}\,(\mathbf{x}, \mathbf{V}, G)\, (\mathbf{C}^{-1})' \boldsymbol{\beta}_j \\
&+ \sum_{j=1}^{d} \left.\frac{\partial H}{\partial \tau_j}\right|_{(\mathbf{\Lambda},\boldsymbol{\tau})} \left[\sum_{i \neq j} \frac{\boldsymbol{\beta}_i' \mathbf{C}^{-1} \text{IF}\,(\mathbf{x}, \mathbf{V}, G)\, (\mathbf{C}^{-1})' \boldsymbol{\beta}_j}{\lambda_j - \lambda_i} \tau_i \right. \\
&+ \left. \boldsymbol{\beta}_j'\, \mathbf{C}^{-1} \text{IF}\,(\mathbf{x}, \mathbf{T}, G) \right]
\end{aligned}$$

which together with Lemma 6.1, concludes the proof. $\square$

PROOF OF THEOREM 5.2. Is similar to that of Theorem 5.1. Notice that using (A.1) we only need to evaluate $\partial\tau_j\left(G_{\epsilon,\mathbf{x}}\right)/\partial\epsilon|_{\epsilon=0}$ and $\partial\lambda_j\left(G_{\epsilon,\mathbf{x}}\right)/\partial\epsilon|_{\epsilon=0}$. Using that $\tau_j\left(G\right)=0$ since $\mathbf{T}(G)=\boldsymbol{\mu}$ and that we have define the eigenvectors of $\mathbf{A}(G)$ in such a way that its diagonal elements are positive, from the continuity of $\boldsymbol{\beta}\left(G_{\epsilon,\mathbf{x}}\right)$ we get

$$\begin{aligned}\left.\frac{\partial\tau_j\left(G_{\epsilon,\mathbf{x}}\right)}{\partial\epsilon}\right|_{\epsilon=0} &= \lim_{\epsilon\to 0}\boldsymbol{\beta}_j\left(G_{\epsilon,\mathbf{x}}\right)'\mathbf{C}^{-1}\frac{\mathbf{T}\left(G_{\epsilon,\mathbf{x}}\right)-\mathbf{T}(G)}{\epsilon}\\ &= \boldsymbol{\beta}_j\left(G\right)'\mathbf{C}^{-1}\mathrm{IF}\left(\mathbf{x},\mathbf{T},G\right)\ . \qquad (A.5)\end{aligned}$$

On the other hand, since $\mathbf{A}(G)=\mathbf{I}_p$, $\lambda_j(G)=1$ and $\boldsymbol{\beta}_j(G)=\mathbf{e}_j$ the $j-$th canonical vector, using the orthogonality of the eigenvectors we obtain

$$\begin{aligned}\left.\frac{\partial\lambda_j\left(G_{\epsilon,\mathbf{x}}\right)}{\partial\epsilon}\right|_{\epsilon=0} &= \lim_{\epsilon\to 0}\frac{\boldsymbol{\beta}_j\left(G_{\epsilon,\mathbf{x}}\right)'\mathbf{A}(G_{\epsilon,\mathbf{x}})\boldsymbol{\beta}_j\left(G_{\epsilon,\mathbf{x}}\right)-1}{\epsilon}\\ &= \lim_{\epsilon\to 0}\boldsymbol{\beta}_j\left(G_{\epsilon,\mathbf{x}}\right)'\frac{\mathbf{A}(G_{\epsilon,\mathbf{x}})-\mathbf{I}_d}{\epsilon}\boldsymbol{\beta}_j\left(G_{\epsilon,\mathbf{x}}\right)\\ &= \boldsymbol{\beta}_j'\mathrm{IF}\left(\mathbf{x},\mathbf{A},G\right)\boldsymbol{\beta}_j=\mathrm{IF}\left(\mathbf{x},\mathbf{A},G\right)_{jj}\ . \qquad (A.6)\end{aligned}$$

Finally, from (A.5) and (A.6) using (A.1) we derive

$$\begin{aligned}\mathrm{IF}\left(\mathbf{x},Pc,G\right) &= \sum_{j=1}^{d}\left.\frac{\partial H}{\partial\lambda_j}\right|_{(\mathbf{I}_d,0)}\left(\mathbf{C}^{-1}\mathrm{IF}\left(\mathbf{x},\mathbf{V},G\right)\left(\mathbf{C}^{-1}\right)'\right)_{jj}\\ &+ \sum_{j=1}^{d}\left.\frac{\partial H}{\partial\tau_j}\right|_{(\mathbf{I}_d,0)}\boldsymbol{\beta}_j'\,\mathbf{C}^{-1}\mathrm{IF}\left(\mathbf{x},\mathbf{T},G\right)\ .\end{aligned}$$

When $F_0=N_d\left(\mathbf{0},\mathbf{I}_d\right)$, using Lemma 6.1 and the fact that $\tau_i(G)=0$, $\lambda_i=1$ we conclude that

$$\begin{aligned}\left.\frac{\partial H}{\partial\tau_i}\right|_{(\mathbf{I}_d,0)} &= -\,E_{F_0}\left(I_{\mathcal{S}}(\mathbf{y})y_i\right)=0\\ \left.\frac{\partial H}{\partial\lambda_i}\right|_{(\mathbf{I}_d,0)} &= \frac{1}{2}P_{F_0}\left(\mathcal{S}\right)-\frac{1}{2}\,E_{F_0}\left(I_{\mathcal{S}}(\mathbf{y})y_i^2\right)\\ &= \frac{1}{2}P_{F_0}\left(\mathcal{S}\right)-\frac{1}{2d}\,E_{F_0}\left(I_{\mathcal{S}}(\mathbf{y})\sum_{i=1}^{d}y_i^2\right)\ .\end{aligned}$$

On the other hand, $P_{F_0}\left(\mathcal{S}\right)=P\left(W_d\le K\right)$ and $E_{F_0}\left(I_{\mathcal{S}}(\mathbf{y})\sum_{i=1}^{d}y_i^2\right)=E\left(W_dI_{(0,K]}\left(W_d\right)\right)$, entail the desired result since $E\left(W_dI_{(0,K]}\left(W_d\right)\right)=d\,P\left(W_{d+2}\le K\right)$. $\square$

PROOF OF COROLLARY 5.3.. Let $F_0=N_d\left(\mathbf{0},\mathbf{I}_d\right)$, then if$\boldsymbol{\Sigma}=\mathbf{C}\mathbf{C}'$, we have that $\mathrm{IF}(\mathbf{x},\mathbf{V},G)=\mathbf{C}\mathrm{IF}\left(\mathbf{C}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right),\mathbf{V},F_0\right)\mathbf{C}'$. Hence,

$$\mathrm{tr}\left(\mathrm{IF}\left(\mathbf{x},\mathbf{V},G\right)\boldsymbol{\Sigma}^{-1}\right)=\mathrm{tr}\left(\mathrm{IF}\left(\mathbf{C}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right),\mathbf{V},F_0\right)\right)\ .$$

The result follows now from Theorem 3 in Gervini (2002) and for the expression given in its Appendix for the situation of the median and the MAD since $\text{IF}(\mathbf{0}, \mathbf{V}, F_0) = 0$ and if $\mathbf{z} \neq \mathbf{0}$

$$\begin{aligned} \text{IF}(\mathbf{z}, \mathbf{V}, F_0) &= \beta \left\{ \alpha(\|\mathbf{z}\|) \left(\frac{\mathbf{z}\mathbf{z}'}{\|\mathbf{z}\|^2} - \frac{\mathbf{I}_d}{d} \right) \right. \\ &+ \left. \left[\frac{c_1}{c_0} g(\|\mathbf{z}\|) + \frac{w\left(\|\mathbf{z}\|^2\right)\|\mathbf{z}\|^2 - c_2}{c_0} \right] \frac{\mathbf{I}_d}{d} \right\} \end{aligned}$$

for a function α. Taking trace we get the result since the first term of the right hand side equals zero and $\beta = d\,(c_0/c_2)$. □

References

[1] Butler, R. (1982), "Nonparametric interval and point prediction using data trimmed by a Grubbs–type outlier rule," *Annals of Statistics*, **10**, 197-204

[2] Canavos, G. and Koutraouvaelis, I. (1984), "The robustness of two sided tolerance limits for normal distributions,"*Journal od Quality Technology*, **16**, 144-149.

[3] Chew, V. (1966), "Confidence, prediction and tolerance regions for the multivariate normal distribution,"*Journal of the American Statistical Association*, **61**, 605-617.

[4] Croux, C. and Haesbroeck, G. (2000), "Principal Component Analysis based on Robust estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies, "*Biometrika*, **87**, 603-18.

[5] Croux, C. and Joossens, K. (2004), "Influence of observations on the misclassification probability in quadratic discriminant analysis,"In Press in *Journal of Multivariate Analysis.*

[6] Cuevas, A. and Baillo, A. (2003), "Parametric versus nonparametric tolerance regions in detection problems,"*Working Paper 03-70, Statistics and Econometrics Series 17.*

[7] Donoho, D. L. (1982), "Breakdown Properties of Multivariate Location Estimators,"Ph.D. qualifying paper, Harvard University.

[8] Fernholz, L. and Gillespie, J. (2001), "Content-corrected tolerance limits based on bootstrap,"*Technometrics*, **43**, 147-155.

[9] Fernholz, L.T. (2002), "Robustness Issues regarding Content Corrected Tolerance Limits,"*Metrika*, **55**, 53-66.

[10] Fraser, D. (1951), "Nonparametric tolerance regions,"University of Toronto.

[11] Fraser, D. and Guttman, I. (1956), "Tolerance regions,"*The Annals of Mathematical Statistics*, **27**, 162-179.

[12] Fuchs, C. and Kenett, S. (1987), "Multivariate tolerance regions and $F-$tests,"*Journal of Quality Technology*, **19**, 122-131.

[13] Fuchs, C. and Kenett, S. (1988), "Appraisal of ceramic substrates by multivariate tolerance regions,"*The Statician*, **37**, 401-411.

[14] Gervini, D. (2002), "The influence function of the Donoho–Stahel estimator of multivariate location and scale,"*Statistics and Probability Letters*, **60**, 425-435.

[15] Guttman, I. (1970), "Construction of $\beta-$content tolerance regions at confidence level γ for large samples for k-variate normal distribution,"*The Annals of Mathematical Statistics*, **41**, 376-400.

[16] John, S. (1962), "A tolerance region for multivariate normal distributions,"*Sankhya*, Serie A, **25**, 363-368.

[17] Krishnamoorthy, K. and Mathew, T. (1999), "Comparison of approximation methods for computing tolerance factors for a multivariate normal population,"*Technometrics*, **41**, 234-249.

[18] Lopuhaä, H. P. (1990), "Estimation of Location and Covariance with High Breakdown Point,"Ph. D. Thesis. Delft University of Technology, Netherlands.

[19] Maronna, R. A. (1976), "Robust M–Estimators of Multivariate Location and Scatter,"*Annals of Statistics,* **4**, 51-67.

[20] Maronna, R. A. and Yohai, V. J. (1995), "The Behavior of the Stahel–Donoho Robust Multivariate Estimator,"*Journal of the American Statistical Association,* **90**, 330-341.

[21] Maronna, R. A. and Yohai, V. J. (1998), "Robust Estimation of Multivariate Location and Scatter,"In *Encyclopedia of Statistical Sciences Update Volume 2,* eds. S. Kotz, C. Read and D. Banks, New York, Wiley, 589-596.

[22] Muirhead, R. (1982), "*Aspects of Multivariate Statistical Theory,"*. Wiley: New York.

[23] Murphy, R. (1948), "Non–parametric tolerance limits,"*The Annals of Mathematical Statistics*, **19**, 551-589.

[24] Odeh, R. and Owen, D. (1980), "Tables for Normal Tolerance Limits, Sampling Plans and Screening ".New York: Marcel Dekker.

[25] Proschan, F. "Confidence and tolerance intervals for the normal distribution,"*Journal of the American Statistical Association*, **48**, 550-564.

[26] Rousseeuw, P. (1985), "Multivariate estimation with High Breakdown Point,"*Mathematical Statistics and Applications,* (W. Grossmann, G. Pflug, I. Vincze and W. Werz, eds.), Dordrecht: Reidel, 283-297.

[27] Rousseeuw, P. J. and van Zomeren, B. C. (1990), "Unmasking multivariate outliers and leverage points."*Journal of the American Statistical Association*, **85**, 633-639.

[28] Stahel, W. (1981), "Robust estimation: Infinitesimal Optimality and Covariance Matrix Estimation."Thesis (in German), ETH, Zurich.

[29] Yohai, V. and Maronna, R. (1995), "The behavior of the Stahel-Donoho multivariate estimators."*Journal of the American Statistical Association*, **85**, 330-341.

7 TABLES

d	2	3	4	5
K	9.8752	13.1367	16.4330	20.0726

Table 1: Classical tolerance factors K when $q = 0.95$ and $\delta = 0.95$ for dimension d.

Δ	$d = 2$		$d = 3$		$d = 4$		$d = 5$	
	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)
2	0.9570	1.0333	0.9570	1.0219	0.9557	1.0164	0.9534	1.0129
4	0.9697	1.1143	0.9641	1.0748	0.9614	1.0554	0.9606	1.0455
8	0.9783	1.3374	0.9715	1.2131	0.9685	1.1570	0.9658	1.1222
16	0.9806	1.7687	0.9752	1.4657	0.9708	1.3300	0.9676	1.2564

Table 2: Actual coverage (π) and volume increment ($\mathcal{I}$) by the inclusion of an anomalous data with norm $\|\mathbf{x}\|$.

Number of inliers	$d = 2$		$d = 3$		$d = 4$		$d = 5$	
	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)
1	0.9439	0.9832	0.9427	0.9813	0.9411	0.9815	0.9407	0.9810
2	0.9359	0.9608	0.9333	0.9627	0.9303	0.9606	0.9295	0.9610
3	0.9293	0.9465	0.9276	0.9448	0.9204	0.9431	0.9180	0.9431
4	0.9206	0.9256	0.9137	0.9247	0.9082	0.9238	0.9064	0.9227

Table 3: Actual content (π) and size ratio ($\mathcal{I}$) for the classical tolerance regions by the replacement of one data with m inliers, when $q = 0.95$, $\delta = 0.95$ and $n = 30$.

d	2	3	4	5	6	7	8	9	10	15
β	1.0414	1.0070	1.0000	0.9957	0.9928	0.9907	0.9890	0.9878	0.9867	0.9835

Table 4: Calibrating constants β for the Donoho–Stahel estimator in dimensions d.

Dimension $d = 2$									
	$q = 0.90$			$q = 0.95$			$q = 0.99$		
n	δ= 0.90	δ= 0.95	δ= 0.99	δ= 0.90	δ= 0.95	δ= 0.99	δ= 0.90	δ= 0.95	δ= 0.99
20	10.8320	12.3821	17.5341	14.5522	16.7106	24.5253	23.1711	27.8769	40.9653
25	9.3069	10.2947	13.2224	12.3985	13.9580	18.1720	19.9671	22.3422	29.4218
30	8.2326	9.2963	11.3772	10.9862	12.2417	15.4112	17.4303	19.8460	25.5177
35	7.9080	8.6992	10.1594	10.4656	11.4732	14.0209	16.5040	18.2180	22.6908
40	7.2711	7.8236	9.3063	9.7533	10.4301	12.1758	15.3556	16.8527	19.6146
45	7.1068	7.5837	8.6729	9.3606	10.0652	11.5020	14.7729	16.3259	19.1649
50	6.7816	7.2303	8.1386	8.8651	9.4666	11.0186	14.1042	15.2943	17.4805
55	6.5337	6.9445	7.7602	8.5696	9.1817	10.3032	13.6083	14.6165	17.0323
60	6.4745	7.0355	7.8174	8.6042	9.1507	10.3930	13.4796	14.5166	16.6398
65	6.3193	6.6702	7.5251	8.2748	8.7425	9.8564	12.8604	13.8477	15.9314
70	6.1470	6.6540	7.3719	8.1761	8.6563	9.8496	12.7697	13.5505	15.6173
75	6.1106	6.4738	7.2753	8.0202	8.5651	9.5449	12.7843	13.4542	15.2029
80	6.0064	6.2933	6.9345	7.8461	8.2518	9.1011	12.3369	13.0866	14.7179
85	5.9857	6.2590	6.8931	7.8599	8.1117	9.1025	12.2277	12.9145	14.3738
90	5.9770	6.3030	7.0116	7.8260	8.2017	9.2849	12.2239	12.8636	14.7869
95	5.9036	6.1823	6.7342	7.7652	8.1079	8.8956	12.1545	12.7148	13.8342
100	5.8052	6.0258	6.6421	7.6159	7.9355	8.6379	11.9852	12.4499	13.5771

Table 5: Robust tolerance factors K_{DS} in dimension $d = 2$, for contents $q = 0.90$, 0.95 and 0.99, and confidence levels $\delta = 0.90$, 0.95 and 0.99.

Dimension $d = 3$									
	$q = 0.90$			$q = 0.95$			$q = 0.99$		
n	δ= 0.90	δ= 0.95	δ= 0.99	δ= 0.90	δ= 0.95	δ= 0.99	δ= 0.90	δ= 0.95	δ= 0.99
20	18.4888	21.1397	27.4670	24.2531	27.9458	36.5284	38.3513	44.9395	58.5057
25	14.5141	16.2015	21.2053	18.8563	21.4465	28.2591	29.2044	34.0270	45.4828
30	12.4591	13.6268	17.5833	16.0860	17.4503	23.0519	24.5519	27.5803	36.2613
35	11.3296	12.1778	15.2488	14.5182	15.6294	19.9587	22.2818	24.6301	30.5704
40	10.6179	11.4505	14.0012	13.5417	14.6996	17.9198	20.4660	22.4652	29.1666
45	10.1765	10.7696	12.3920	12.9928	13.7569	16.0137	19.5297	21.1643	24.4763
50	9.7695	10.2317	11.2996	12.3856	13.0009	14.5262	18.7228	19.8433	22.9663
55	9.3373	9.9091	11.1179	11.8649	12.5228	14.4079	17.8622	18.9811	22.5006
60	9.1459	9.8136	10.7406	11.6818	12.4917	13.7287	17.5797	19.0078	21.9252
65	8.8516	9.2392	10.7574	11.1660	11.7582	13.6343	16.7690	17.6115	20.9971
70	8.6261	9.1126	9.9451	10.9276	11.5222	12.5902	16.3966	17.4556	18.8966
75	8.6470	9.0168	9.6662	10.9072	11.5021	12.3449	16.3814	17.0689	18.5640
80	8.3896	8.8135	9.3718	10.7288	11.1830	11.9359	15.8844	16.5346	18.8273
85	8.3207	8.7319	9.4782	10.5832	11.0636	12.0174	15.8495	16.5357	17.7586
90	8.1590	8.5613	9.3359	10.3339	10.8568	11.9504	15.4399	16.0606	17.8178
95	8.1348	8.5243	9.1853	10.3238	10.6891	11.6131	15.3301	16.0695	18.0892
100	7.9569	8.2413	8.9158	10.0528	10.4174	11.2309	15.1451	15.7830	17.7933

Table 6: Robust tolerance factors K_{DS} in dimension $d = 3$, for contents $q = 0.90$, 0.95 and 0.99, and confidence levels $\delta = 0.90$, 0.95 and 0.99.

Dimension $d = 4$									
	$q = 0.90$			$q = 0.95$			$q = 0.99$		
n	δ= 0.90	δ= 0.95	δ= 0.99	δ= 0.90	δ= 0.95	δ= 0.99	δ= 0.90	δ= 0.95	δ= 0.99
20	27.4182	31.5431	44.5483	35.4487	41.4550	57.3808	55.7374	66.4408	92.8933
25	21.4547	24.6844	31.8190	27.4777	31.5780	42.3042	42.9968	49.9301	65.0356
30	16.9623	18.3120	21.6688	21.5732	23.2288	28.3680	32.7613	35.3599	44.2545
35	15.1949	16.6681	20.3717	19.0612	21.2592	26.2953	28.8685	32.1940	41.4370
40	13.5806	14.2707	16.4577	16.9452	18.0497	21.1071	24.9294	27.1523	33.2087
45	12.6335	13.3466	15.3875	15.8920	16.6360	19.7838	23.3600	25.0936	29.0160
50	12.3287	13.0614	14.1136	15.4366	16.2419	17.9883	22.6729	24.0837	27.8435
55	11.6009	12.1972	13.8174	14.5084	15.2489	17.5803	21.2475	22.5057	25.8496
60	11.4914	11.9824	13.1846	14.1753	14.9044	16.5707	20.7960	22.0699	24.7157
65	11.1283	11.6137	12.5504	13.7444	14.3994	15.5499	19.8585	21.0003	22.9426
70	10.7022	11.1047	12.1436	13.1528	13.7946	15.1036	19.0215	19.8970	22.1491
75	10.6953	11.0952	11.8944	13.1943	13.7044	14.7181	19.0660	20.0513	21.7096
80	10.3104	10.7255	11.3242	12.7346	13.3265	14.0519	18.5037	19.1277	20.5951
85	10.3025	10.6632	11.5255	12.6941	13.1263	14.0832	18.3106	19.0707	20.5999
90	10.2673	10.6312	11.4723	12.6865	13.1973	14.2593	18.2493	19.1498	20.6151
95	10.1159	10.4224	11.0732	12.4098	12.8850	13.6946	17.9134	18.7162	20.1619
100	9.9515	10.2544	10.7488	12.2488	12.6392	13.4576	17.7022	18.4893	19.9511

Table 7: Robust tolerance factors K_{DS} in dimension $d = 4$, for contents $q = 0.90$, 0.95 and 0.99, and confidence levels $\delta = 0.90$, 0.95 and 0.99.

		Classical		Robust		
d	n	K_{C}	π_{C}	K_{DS}	π_{DS}	$\sqrt[d]{\frac{\nu_{\text{C}}}{\nu_{\text{DS}}}}$
2	20	12.1744	0.9525	16.7106	0.9540	1.0754
2	30	9.8752	0.9460	12.2417	0.9500	1.0455
2	50	8.3989	0.9440	9.4666	0.9470	1.0206
2	100	7.4187	0.9470	7.9355	0.9500	1.0137
3	20	16.6939	0.9480	27.9458	0.9520	1.1030
3	30	13.2222	0.9485	17.4503	0.9470	1.0440
3	50	10.4174	0.9485	13.0009	0.9530	1.0240
3	100	9.6736	0.9490	10.1994	0.9475	1.0023
4	20	21.3464	0.9480	41.4550	0.9475	1.1409
4	30	16.9176	0.9510	23.2288	0.9525	1.0453
4	50	13.4051	0.9480	16.2419	0.9535	1.0341
4	100	11.6219	0.9490	12.6392	0.9490	1.0075
5	20	27.2366	0.9445	64.0125	0.9400	1.2097
5	30	20.0054	0.9490	29.1497	0.9440	1.0571
5	50	15.9508	0.9510	19.1330	0.9525	1.0224
5	100	13.6632	0.9480	14.8441	0.9505	1.0085
8	20	56.1119	0.9520	240.0697	0.9375	1.5496
8	30	32.9768	0.9455	57.5583	0.9490	1.1159
8	50	23.9586	0.9475	28.8758	0.9500	1.0228
8	100	20.8338	0.9480	20.5229	0.9490	1.0062

Table 8: Classical and Robust tolerance factors (K_{C} and K_{DS}) and ratio of the volumes among both regions. Normal observations, $q = 0.95$ and $\delta = 0.95$.

Δ	$d=2$		$d=3$		$d=4$		$d=5$	
	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)
2	0.9583	1.0309	0.9516	1.0183	0.9499	1.0119	0.9575	1.0077
4	0.9607	1.0467	0.9575	1.0321	0.9542	1.0247	0.9583	1.0226
8	0.9633	1.0508	0.9592	1.0383	0.9566	1.0311	0.9616	1.0263
16	0.9608	1.0535	0.9616	1.0362	0.9600	1.0295	0.9608	1.0230

Table 9: Actual content (π) and size ratio ($\mathcal{I}$) for the robust tolerance regions by the replacement of one data with an observation with norm $\|\mathbf{x}\|$, when $q = 0.95$, $\delta = 0.95$ and $n = 30$.

m	$d=2$		$d=3$		$d=4$		$d=5$	
	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)	π	$\mathcal{I}$ (%)
1	0.9455	0.9727	0.9355	0.9669	0.9310	0.9648	0.9250	0.9628
2	0.9285	0.9434	0.9120	0.9314	0.9090	0.9228	0.8930	0.9185
3	0.9050	0.9106	0.8865	0.8878	0.8625	0.8701	0.8465	0.8540
4	0.8800	0.8701	0.8440	0.8422	0.8025	0.8176	0.7600	0.7872

Table 10: Actual content (π) and size ratio ($\mathcal{I}$) for the robust tolerance regions by the replacement of one data with m inliers, when $q = 0.95$, $\delta = 0.95$ and $n = 30$.

		Classical		Robust		Relations	
d	n	K_{C}	π_{C}	K_{DS}	π_{DS}	$\sqrt[d]{\frac{\nu_{\text{C}\cdot}}{\nu_{\text{DS}}}}$	$\frac{\lVert\overline{\mathbf{x}}\rVert}{\lVert\mathbf{t}_n\rVert}$
2	20	12.1744	1.0000	16.7106	0.9960	2.7778	4.3498
2	30	9.7920	1.0000	12.2417	0.9980	3.1331	4.9244
2	50	8.3989	1.0000	9.4666	0.9980	4.0020	6.1716
2	100	7.4187	1.0000	7.9355	0.9990	5.9750	9.4369
3	20	16.6939	1.0000	27.9458	0.9940	2.4637	4.7695
3	30	13.2222	1.0000	17.4503	0.9970	3.0593	5.7492
3	50	10.9711	1.0000	13.0009	0.9980	3.9036	6.6528
3	100	9.6736	1.0000	10.4174	0.9990	5.3589	8.8928
4	20	21.3464	0.9990	41.4550	0.9930	2.3600	5.1893
4	30	16.9176	1.0000	23.2288	0.9940	2.9915	6.0277
4	50	13.4051	1.0000	16.2419	0.9980	3.6996	7.4199
4	100	11.6219	1.0000	12.6392	0.9990	4.9280	9.1556
5	20	27.2366	0.9985	64.0125	0.9860	2.1726	5.0921
5	30	20.0054	1.0000	29.1497	0.9935	2.7466	5.8958
5	50	15.9508	1.0000	19.1330	0.9980	3.4553	6.9667
5	100	13.6632	1.0000	14.8441	1.0000	4.8052	9.7124
8	20	56.1119	0.9960	240.0697	0.9425	1.7047	4.7189
8	30	32.9768	1.0000	57.5583	0.9865	2.4991	6.9566
8	50	23.9586	1.0000	28.8758	0.9970	3.0707	7.6931
8	100	19.5602	1.0000	20.8338	1.0000	4.0777	10.6633

Table 11: Actual content of the classical and robust tolerance regions (π_{C} and π_{DS}, respectively) related to the tolerance factors K_{C} and K_{DS}, obtained when the observations have a Cauchy $d-$ dimensional distribution. The tolerance factors are those corresponding to a normal distribution with $\delta = q = 0.95$.

		Classical		Robust		Relations	
d	n	K_{C}	π_{C}	K_{DS}	π_{DS}	$\sqrt[d]{\frac{\mathcal{V}_{\mathrm{C}\cdot}}{\mathcal{V}_{\mathrm{DS}}}}$	$\frac{\Vert\overline{\mathbf{x}}\Vert}{\Vert\mathbf{t}_n\Vert}$
2	20	12.1744	0.9950	16.7106	0.9805	1.2756	1.5943
2	30	9.7920	0.9960	12.2417	0.9850	1.3770	1.6817
2	50	8.3989	0.9980	9.4666	0.9880	1.5240	1.7894
2	100	7.4187	1.0000	7.9355	0.9930	1.6028	1.8072
3	20	16.6939	0.9920	27.9458	0.9820	1.2870	1.6639
3	30	13.2222	0.9960	17.4503	0.9840	1.4265	1.7984
3	50	10.9711	0.9990	13.0009	0.9905	1.5235	1.8929
3	100	9.6736	1.0000	10.4174	0.9920	1.6523	1.9467
4	20	21.3464	0.9920	41.4550	0.9715	1.2753	1.7438
4	30	16.9176	0.9970	23.2288	0.9825	1.4347	1.7523
4	50	13.4051	0.9990	16.2419	0.9890	1.5223	1.8904
4	100	11.6219	1.0000	12.6392	0.9930	1.6210	1.8976
5	20	27.2366	0.9860	64.0125	0.9750	1.2133	1.6914
5	30	20.0054	0.9940	29.1497	0.9785	1.3951	1.8109
5	50	15.9508	0.9980	19.1330	0.9860	1.5054	1.8997
5	100	13.6632	1.0000	14.8441	0.9940	1.6286	1.9849
8	20	56.1119	0.9850	240.0697	0.9455	0.9877	1.5191
8	30	32.9768	0.9920	57.5583	0.9720	1.3502	1.8526
8	50	23.9586	0.9980	28.8758	0.9860	1.4794	1.9352
8	100	19.5602	1.0000	20.8338	0.9950	1.5556	1.9178

Table 12: Actual content of the classical and robust tolerance regions (π_{C} and π_{DS}, respectively) related to the tolerance factors K_{C} and K_{DS}, obtained when the observations have a $\mathcal{T}_2(d)$ distribution. The tolerance factors are those corresponding to a normal distribution with $\delta = q = 0.95$.

		Classical		Robust		Relations	
d	n	$K_{\rm C}$	$\pi_{\rm C}$	$K_{\rm DS}$	$\pi_{\rm DS}$	$\sqrt[d]{\frac{\nu_{\rm C\cdot}}{\nu_{\rm DS}}}$	$\frac{\lVert\overline{\mathbf{x}}\rVert}{\lVert\mathbf{t}_n\rVert}$
2	20	12.1744	0.9880	16.7106	0.9760	1.1260	1.2941
2	30	9.7920	0.9880	12.2417	0.9765	1.1477	1.2558
2	50	8.3989	0.9935	9.4666	0.9810	1.2231	1.3065
2	100	7.4187	0.9960	7.9355	0.9835	1.2583	1.3278
3	20	16.6939	0.9875	27.9458	0.9780	1.0994	1.2331
3	30	13.2222	0.9890	17.4503	0.9770	1.1938	1.3162
3	50	10.9711	0.9940	13.0009	0.9815	1.2263	1.3501
3	100	9.6736	0.9970	10.4174	0.9850	1.2856	1.3747
4	20	21.3464	0.9860	41.4550	0.9690	1.0709	1.2550
4	30	16.9176	0.9910	23.2288	0.9770	1.1946	1.3047
4	50	13.4051	0.9930	16.2419	0.9805	1.2134	1.3738
4	100	11.6219	0.9970	12.6392	0.9860	1.2690	1.3459
5	20	27.2366	0.9795	64.0125	0.9575	1.0263	1.2297
5	30	20.0054	0.9875	29.1497	0.9700	1.1920	1.3743
5	50	15.9508	0.9950	19.1330	0.9830	1.2418	1.3721
5	100	13.6632	0.9980	14.8441	0.9880	1.2659	1.4181
8	20	56.1119	0.9765	240.0697	0.9525	0.8367	1.1343
8	30	32.9768	0.9840	57.5583	0.9705	1.1424	1.3359
8	50	23.9586	0.9930	28.8758	0.9780	1.2349	1.4239
8	100	19.5602	0.9980	20.8338	0.9890	1.2502	1.3946

Table 13: Actual content of the classical and robust tolerance regions ($\pi_{\rm C}$ and $\pi_{\rm DS}$, respectively) related to the tolerance factors $K_{\rm C}$ and $K_{\rm DS}$, obtained when the observations have a $\mathcal{T}_3(d)$ distribution. The tolerance factors are those corresponding to a normal distribution with $\delta = q = 0.95$.

		Classical		Robust		Relations	
d	n	K_{C}	π_{C}	K_{DS}	π_{DS}	$\sqrt[d]{\frac{\nu_{\mathrm{C}\cdot}}{\nu_{\mathrm{DS}}}}$	$\frac{\lVert\overline{\mathbf{x}}\rVert}{\lVert\mathbf{t}_n\rVert}$
2	20	12.1744	0.9610	16.7106	0.9600	0.9599	1.0483
2	30	9.7920	0.9570	12.2417	0.9580	0.9986	1.1352
2	50	8.3989	0.9530	9.4666	0.9530	1.0328	1.1408
2	100	7.4187	0.9610	7.9355	0.9530	1.1008	1.3178
3	20	16.6939	0.9545	27.9458	0.9495	0.9235	1.0183
3	30	13.2222	0.9590	17.4503	0.9545	1.0018	1.1108
3	50	10.9711	0.9545	13.0009	0.9530	1.0324	1.1956
3	100	9.6736	0.9620	10.4174	0.9540	1.0973	1.3717
4	20	21.3464	0.9520	41.4550	0.9480	0.8991	0.9963
4	30	16.9176	0.9580	23.2288	0.9450	0.9923	1.0702
4	50	13.4051	0.9585	16.2419	0.9570	1.0219	1.1730
4	100	11.6219	0.9600	12.6392	0.9545	1.1054	1.2708
5	20	27.2366	0.9530	64.0125	0.9500	0.8489	0.8987
5	30	20.0054	0.9540	29.1497	0.9500	0.9834	1.0300
5	50	15.9508	0.9580	19.1330	0.9525	1.0322	1.1578
5	100	13.6632	0.9630	14.8441	0.9560	1.0940	1.2787
8	20	56.1119	0.9555	240.0697	0.9605	0.6659	0.7926
8	30	32.9768	0.9520	57.5583	0.9445	0.9262	0.9483
8	50	23.9586	0.9570	28.8758	0.9545	1.0175	1.1092
8	100	19.5602	0.9600	20.8338	0.9550	1.0896	1.2826

Table 14: Actual coverage probability for the classical and robust tolerance regions (π_{C} and π_{DS}, respectively) related to the tolerance factors K_{C} and K_{DS}, for a sample coming from a dsitribution $0.95\, N(\mathbf{0}, \mathbf{I}_d) + 0.05\, \mathcal{C}_d$. The tolerance factors are those corresponding to a normal distribution with $\delta = q = 0.95$.

		Classical		Robust		Relations	
d	n	K_{C}	π_{C}	K_{DS}	π_{DS}	$\sqrt[d]{\frac{\nu_{\mathrm{C}\cdot}}{\nu_{\mathrm{DS}}}}$	$\frac{\lVert\overline{\mathbf{x}}\rVert}{\lVert\mathbf{t}_n\rVert}$
2	20	12.1744	0.9620	16.7106	0.9595	1.0092	1.2260
2	30	9.7920	0.9610	12.2417	0.9595	1.0547	1.3079
2	50	8.3989	0.9630	9.4666	0.9560	1.1900	1.4578
2	100	7.4187	0.9700	7.9355	0.9590	1.3981	1.8004
3	20	16.6939	0.9610	27.9458	0.9580	0.9582	1.1489
3	30	13.2222	0.9620	17.4503	0.9540	1.0727	1.2395
3	50	10.9711	0.9670	13.0009	0.9605	1.1840	1.4368
3	100	9.6736	0.9750	10.4174	0.9580	1.3780	1.7698
4	20	21.3464	0.9510	41.4550	0.9445	0.9386	1.0944
4	30	16.9176	0.9670	23.2288	0.9610	1.0745	1.2703
4	50	13.4051	0.9650	16.2419	0.9595	1.1698	1.4181
4	100	11.6219	0.9715	12.6392	0.9580	1.3426	1.8048
5	20	27.2366	0.9510	64.0125	0.9470	0.8749	1.0057
5	30	20.0054	0.9590	29.1497	0.9485	1.0385	1.1910
5	50	15.9508	0.9650	19.1330	0.9580	1.1630	1.4284
5	100	13.6632	0.9730	14.8441	0.9600	1.2988	1.6645
8	20	56.1119	0.9580	240.0697	0.9485	0.6895	0.8779
8	30	32.9768	0.9590	57.5583	0.9530	0.9795	1.0942
8	50	23.9586	0.9645	28.8758	0.9565	1.1337	1.3902
8	100	19.5602	0.9710	20.8338	0.9600	1.2704	1.7727

Table 15: Actual coverage probability for the classical and robust tolerance regions (π_{C} and π_{DS}, respectively) related to the tolerance factors K_{C} and K_{DS}, for a sample coming from a dsitribution $0.90\,N(\mathbf{0},\mathbf{I}_d)+0.10\,\mathcal{C}_d$. The tolerance factors are those corresponding to a normal distribution with $\delta = q = 0.95$.

		Classical		Robust		Relations	
d	n	K_{C}	π_{C}	K_{DS}	π_{DS}	$\sqrt[d]{\frac{\nu_{\mathrm{C}}\cdot}{\nu_{\mathrm{DS}}}}$	$\frac{\Vert\overline{\mathbf{x}}\Vert}{\Vert\mathbf{t}_n\Vert}$
2	20	12.1744	0.9645	16.7106	0.9595	1.0519	1.2873
2	30	9.7920	0.9670	12.2417	0.9580	1.1511	1.3146
2	50	8.3989	0.9720	9.4666	0.9610	1.2108	1.3870
2	100	7.4187	0.9800	7.9355	0.9620	1.3020	1.3716
3	20	16.6939	0.9610	27.9458	0.9625	1.0102	1.1857
3	30	13.2222	0.9680	17.4503	0.9605	1.1352	1.3027
3	50	10.9711	0.9705	13.0009	0.9625	1.2299	1.2912
3	100	9.6736	0.9830	10.4174	0.9630	1.3061	1.3384
4	20	21.3464	0.9640	41.4550	0.9550	0.9911	1.1543
4	30	16.9176	0.9720	23.2288	0.9610	1.1435	1.2454
4	50	13.4051	0.9700	16.2419	0.9630	1.2111	1.3921
4	100	11.6219	0.9815	12.6392	0.9630	1.2903	1.3609
5	20	27.2366	0.9580	64.0125	0.9610	0.9253	1.0750
5	30	20.0054	0.9650	29.1497	0.9580	1.1070	1.2912
5	50	15.9508	0.9730	19.1330	0.9630	1.2069	1.4072
5	100	13.6632	0.9810	14.8441	0.9660	1.2875	1.3973
8	20	56.1119	0.9595	240.0697	0.9475	0.7188	0.9663
8	30	32.9768	0.9610	57.5583	0.9615	1.0201	1.1362
8	50	23.9586	0.9700	28.8758	0.9610	1.1644	1.3456
8	100	19.5602	0.9800	20.8338	0.9660	1.2477	1.3736

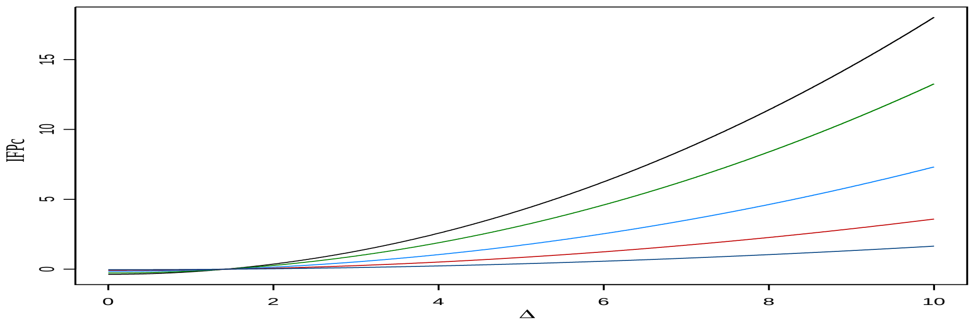
Table 16: Actual coverage probability for the classical and robust tolerance regions (π_{C} and π_{DS}, respectively) related to the tolerance factors K_{C} and K_{DS}, for a sample coming from a dsitribution $0.95\,N(\mathbf{0},\mathbf{I}_d) + 0.05\,N(\mathbf{0},25\,\mathbf{I}_d)$. The tolerance factors are those corresponding to a normal distribution with $\delta = q = 0.95$.

		Classical		Robust		Relations	
d	n	K_{C}	π_{C}	K_{DS}	π_{DS}	$\sqrt[d]{\frac{\nu_{\mathrm{C}\cdot}}{\nu_{\mathrm{DS}}}}$	$\frac{\lVert\overline{\mathbf{x}}\rVert}{\lVert\mathbf{t}_n\rVert}$
2	20	12.1744	0.9770	16.7106	0.9715	1.2718	1.5478
2	30	9.7920	0.9820	12.2417	0.9750	1.3477	1.6342
2	50	8.3989	0.9890	9.4666	0.9730	1.4440	1.6286
2	100	7.4187	0.9960	7.9355	0.9750	1.5147	1.6865
3	20	16.6939	0.9780	27.9458	0.9665	1.1768	1.4892
3	30	13.2222	0.9820	17.4503	0.9660	1.3355	1.5747
3	50	10.9711	0.9880	13.0009	0.9725	1.4306	1.6146
3	100	9.6736	0.9960	10.4174	0.9750	1.5457	1.6508
4	20	21.3464	0.9715	41.4550	0.9660	1.1350	1.4800
4	30	16.9176	0.9790	23.2288	0.9660	1.3155	1.5700
4	50	13.4051	0.9870	16.2419	0.9730	1.4421	1.6532
4	100	11.6219	0.9955	12.6392	0.9750	1.5272	1.6767
5	20	27.2366	0.9765	64.0125	0.9630	1.0739	1.3234
5	30	20.0054	0.9770	29.1497	0.9670	1.2821	1.5510
5	50	15.9508	0.9860	19.1330	0.9730	1.4193	1.6428
5	100	13.6632	0.9960	14.8441	0.9770	1.5274	1.7042
8	20	56.1119	0.9690	240.0697	0.9640	0.7895	1.2064
8	30	32.9768	0.9740	57.5583	0.9665	1.1723	1.4651
8	50	23.9586	0.9840	28.8758	0.9700	1.3610	1.6582
8	100	19.5602	0.9960	20.8338	0.9770	1.4863	1.6779

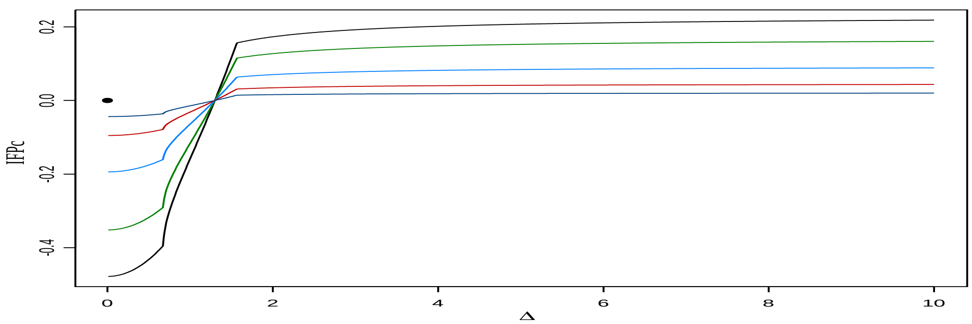
Table 17: Actual coverage probability for the classical and robust tolerance regions (π_{C} and π_{DS}, respectively) related to the tolerance factors K_{C} and K_{DS}, for a sample coming from a dsitribution $0.90\,N(\mathbf{0},\mathbf{I}_d) + 0.10\,N(\mathbf{0},25\,\mathbf{I}_d)$. The tolerance factors are those corresponding to a normal distribution with $\delta = q = 0.95$.

8 FIGURES

(a)



(b)



(c)

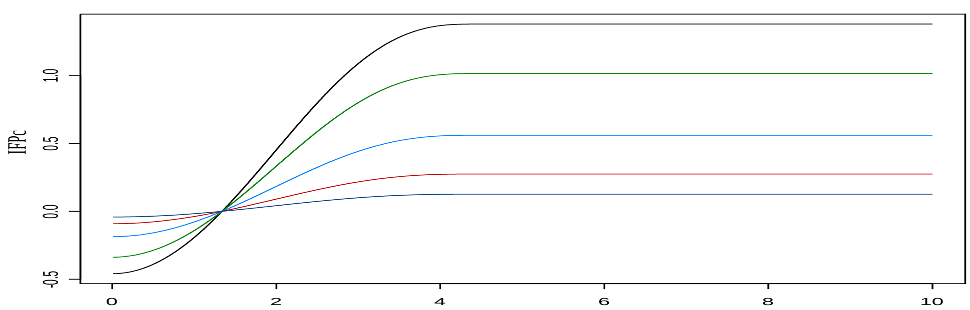
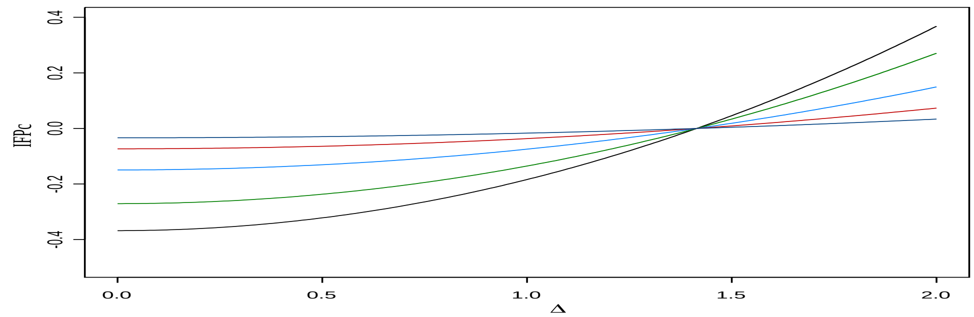
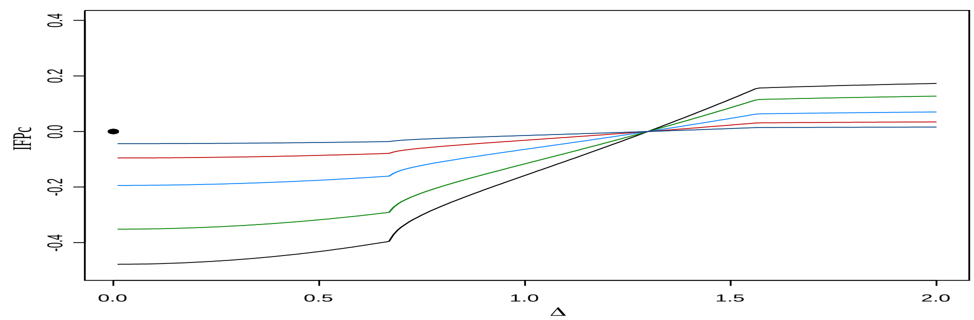


Figure 1: Influence function of the coverage probablity when we use the classical estimators (a), the Donoho–Stahel estimators (b) and the S–estimator (c). The lines in black correspond to $K = 2$, while those in green, light blue, red and dark blue correspond to $K = 4$, $K = 6$, $K = 8$ and $K = 10$, respectively.

(a)



(b)



(c)

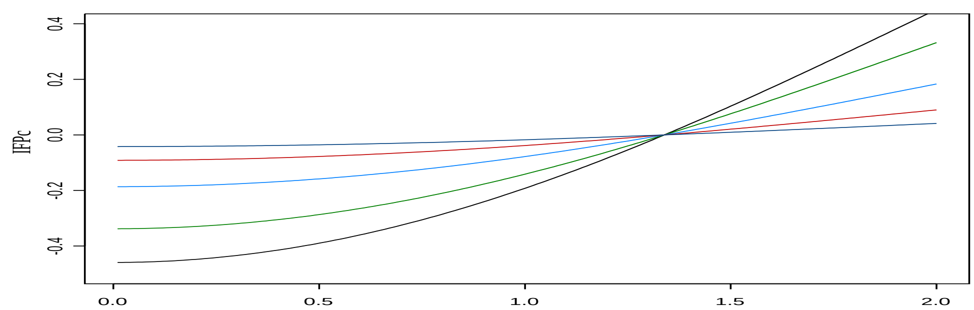


Figure 2: Influence function of the coverage probablity when we use the classical estimators (a) and the Donoho–Stahel estimators (b) and the S–estimator (c). The lines in black correspond to $K = 2$, while those in green, light blue, red and dark blue correspond to $K = 4$, $K = 6$, $K = 8$ and $K = 10$, respectively.