

Robust asymptotic tests for the equality of multivariate coefficients of variation

S. Aerts · G. Haesbroeck

the date of receipt and acceptance should be inserted later

Abstract In order to easily compare several populations on the basis of more than one feature, multivariate coefficients of variation (MCV) may be used as they allow to summarize relative dispersion in a single index. However, up to date, no test of equality of one or more MCVs has been developed in the literature. In this paper, several classical and robust Wald-type tests are proposed and studied. The asymptotic distributions of the test statistics are derived under elliptical symmetry, and the asymptotic efficiency of the robust versions is compared to the classical tests. Robustness of the proposed procedures is examined through partial influence functions of the test statistic, as well as by means of power and level influence functions. A simulation study compares the performance of the classical and robust tests under uncontaminated and contaminated schemes, and the difference with the usual covariance homogeneity test is highlighted. As a by-product, these tests may also be considered in the univariate context where they yield procedures that are both robust and easy-to-use. They provide an interesting alternative to the numerous parametric tests existing in the literature, which are, in most cases, unreliable in presence of outliers. The methods are illustrated on a real data set.

Keywords Multivariate Coefficient of Variation · Robust testing · Wald-type test · External Quality Assessment

Mathematics Subject Classification (2000) 62H15 · 62F05 · 62F35

S. Aerts
HEC-ULg, University of Liege (ULg, N1), Rue Louvrex 14, 4000 Liege, Belgium
E-mail: stephanie.aerts@ulg.ac.be

G. Haesbroeck
University of Liege (ULg, Polytech 1), Allée de la Découverte 12, 4000 Liège, Belgium

1 Introduction

The univariate coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, is a well-known measure of relative dispersion. Its use is advocated in many practical applications when interest is on the comparison of dispersion of a feature in populations having really different means. For example in finance, the inverse CV, known as the Sharpe ratio, is a measure of a portfolio performance relative to its risk. The CV is also recommended for controlling the performance of clinical laboratories, as well as comparing measurement techniques in External Quality Assessment (EQA) schemes that are organized worldwide by government health agencies or private companies.

Given the widespread use of the CV for comparison purposes, many authors have proposed some parametric tests for the equality of CV's, or equivalently, of inverse CV's in K populations. Both Bennett (1976) and later Forkman (2009) proposed some tests based on McKay (1932) approximation; Jobson and Korkie (1981), Miller (1991a), Feltz and Miller (1996), Gupta and Ma (1996), Nairy and Rao (2003) and Ledoit and Wolf (2008) provided some Wald-type tests for the homogeneity of CV's or inverse CV's in two or more populations, using the asymptotic distribution of either the MLE or the sample estimator; Pardo and Pardo (2000) proposed a family of tests based on Renyi's divergence. Likelihood ratio tests and signed likelihood ratio tests were studied by Nairy and Rao (2003), Verrill and Johnson (2007) and Krishnamoorthy and Lee (2014) while score tests and modified versions of them were proposed by Gupta and Ma (1996), Nairy and Rao (2003) and Tsou (2009). Forkman (2009) proposed a test for the equality of two or more CVs based on F -statistics.

All these tests are based either on the MLE or the sample estimator of the CV which are however known to be potentially extremely sensitive to extreme values, outliers or blunders. However, such perturbation in the data is very common in real applications, for instance in EQA schemes (Healy, 1979). Therefore, robust estimators of the CV should be used instead. An easy non-parametric approach, which has been used in the Belgian EQA program for nearly thirty years, consists in taking the ratio of either the interquartile range (IQR) or the median absolute deviation (MAD) to the median as a robust alternative to the classical CV. To our knowledge though, no testing procedure adapted to these robust versions is developed in the literature.

In the multivariate setting, i.e. when the comparison of relative variability is based on several characteristics, comparing the marginal CVs may lead to contradictions. Therefore, the use of a multivariate coefficient of variation summarizing the relative dispersion in one single index has already been advocated in several applications, for instance in the EQA context (Zhang et al., 2010) where it is used to compare protein electrophoretic techniques. Among the four MCVs proposed in the literature and reviewed by Albert and Zhang (2010), we chose to focus on Voinov and Nikulin's one (Voinov and Nikulin, 1996) because of its nice invariance properties. For a p -variate random variable \mathbf{X} , it is defined as

$$\gamma = (\boldsymbol{\mu}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})^{-1/2}$$

where $\boldsymbol{\mu}$ is the (non-null) mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix, assumed to be positive-definite. This coefficient reduces to the univariate CV when $p = 1$, provided that $\mu > 0$. In practice, using the plug-in principle, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be replaced by any pair of location and covariance estimators. As in the univariate case, the use of robust estimators is preferable in the presence of outliers, as already mentioned in Zhang et al. (2010) and Aerts et al. (2015). The properties of these coefficients have so far been little studied and no testing procedure for comparing K p -variate MCVs, either based on classical estimators or on robust ones, has been proposed in the literature. The goal of this paper is to suggest a general Wald-type test for the homogeneity of MCVs in K elliptically symmetric populations and to study its robustness and efficiency. In the particular case of a dimension $p = 1$, our proposal yields a natural robustification of the classical Wald-type test and allows to robustly compare the CVs of K populations, unlike the already existing procedures.

The paper is organized as follows. In Section 2, the test statistics are defined and their asymptotic distributions under the null, a fixed alternative or a sequence of contiguous alternatives are stated in Section 3. In Section 4, the relative efficiencies of these tests are compared. Robustness of the testing procedures is analyzed in Section 5 through partial and joint influence functions of the test statistics, as well as power and level influence functions. The results of a Monte Carlo study examining the performance of those tests in terms of both level and power are summarized in Section 6. The difference between these tests and the well-known tests of homogeneity of covariance matrices is also illustrated. Finally, the method is illustrated on a real data example in Section 7, and some concluding remarks are outlined in Section 8.

2 Framework and test statistics

In this section, four statistics are provided for testing the homogeneity of MCVs in K p -variate populations. The data consist of K mutually independent random samples of varying sizes n_1, \dots, n_K , i.e. for each k in $1, \dots, K$, the k -th random sample $(\mathbf{X}_{k1}, \dots, \mathbf{X}_{kn_k})$ is formed by n_k independent p -variate random vectors each distributed according to a distribution $F_k := F_{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k}$ with density

$$f_k := f_{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k}(\mathbf{x}) = c_{p,f} |\boldsymbol{\Sigma}_k|^{-1/2} g_k((\mathbf{x} - \boldsymbol{\mu}_k)^t \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)) \quad (1)$$

where $\boldsymbol{\mu}_k \in \mathbb{R}^p$ and $\boldsymbol{\mu}_k \neq 0$, $\boldsymbol{\Sigma}_k$ is assumed to be symmetric and positive-definite, i.e. $\boldsymbol{\Sigma}_k \in \mathcal{S}_p^+$, g_k is a known real-valued function having a strictly negative derivative g'_k , and $c_{p,f}$ is a normalizing factor. This means that each distribution $F_{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k}$ belongs to a parametric class of absolutely continuous and unimodal elliptically symmetric distributions generated by the radial density g_k . In order to properly identify $\boldsymbol{\Sigma}_k$ and g_k , g_k is defined such that $\boldsymbol{\Sigma}_k$ is the covariance matrix of the k -th population. Let us note that the K populations do not necessarily share the same radial density. In the sequel, \mathcal{F} stands for the K -fold product $\mathcal{F} = F_1 \times \dots \times F_K$.

Distribution	Additional parameter	Radial density	Notation
Normal	\backslash	$g(t) = \exp(-t/2)$	$\Phi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
Student	$\nu > 0$	$g_\nu(t) = (1 + a_\nu t/\nu)^{-(p+\nu)/2}$	$t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
Power exponential	$\beta > 0$	$g_\beta(t) = \exp(-b_{p,\beta} t^{2\beta})$	$\text{PE}_\beta(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Table 1 Radial densities of well-known elliptical distributions. The positive constants a_ν and $b_{p,\beta}$ are such that $\boldsymbol{\Sigma}$ is the covariance matrix.

Radial densities of well-known elliptical distributions used herein are given in Table 1. The suggested notations do not explicitly mention the dimension p , as it is implicitly included in the mean vector and covariance matrix.

Let us denote by γ_k the theoretical MCV of the k -th population. Here, the null hypothesis of interest is $H_0 : \gamma_k = \gamma$ for $1 \leq k \leq K$, where γ is unspecified, against the alternative $H_1 : \exists i, j : \gamma_i \neq \gamma_j$ with $i \neq j$.

Let V_{n_k} be a plug-in estimator of the MCV γ_k obtained from a pair of location and covariance estimators $(\mathbf{T}_{n_k}, \mathbf{C}_{n_k})$. It is assumed here that the estimators V_{n_k} are all related to the same functional V whatever the population, i.e. $V_{n_k} = V(\mathbb{F}_{k;n_k})$ where $\mathbb{F}_{k;n_k}$ corresponds to the empirical distribution function associated with the k -th sample $(\mathbf{X}_{k1}, \dots, \mathbf{X}_{kn_k})$. Under elliptical symmetry with finite fourth-order moments, some standard assumptions on \mathbf{T}_{n_k} and \mathbf{C}_{n_k} allow to obtain a simple expression for the asymptotic variance (ASV) of the related plug-in estimator V_{n_k} . Throughout this paper, it is assumed that:

- (A1) \mathbf{T}_{n_k} and \mathbf{C}_{n_k} are affine-equivariant
- (A2) \mathbf{T}_{n_k} and \mathbf{C}_{n_k} are asymptotically independent
- (A3) $\sqrt{n_k}(\mathbf{T}_{n_k} - \boldsymbol{\mu}_k) \xrightarrow{\mathcal{L}} N_p(0, \tau_k \boldsymbol{\Sigma}_k)$, for a constant τ_k
- (A4) $\sqrt{n_k} \text{vec}(\mathbf{C}_{n_k} - \boldsymbol{\Sigma}_k) \xrightarrow{\mathcal{L}} N_{p^2}(0, \Xi)$ with $\Xi = \sigma_{1;k}(\mathbf{I}_{p^2} + \mathbf{K}_{p^2})(\boldsymbol{\Sigma}_k \otimes \boldsymbol{\Sigma}_k) + \sigma_{2;k}(\text{vec} \boldsymbol{\Sigma}_k)(\text{vec} \boldsymbol{\Sigma}_k)^t$ for some constants $\sigma_{1;k}, \sigma_{2;k}$, where vec and \otimes stand for the usual vectorization operator and Kronecker product respectively, \mathbf{I}_{p^2} is the p^2 -dimensional identity matrix and \mathbf{K}_{p^2} is the commutation matrix of dimension $p^2 \times p^2$ defined by $\mathbf{K}_{p^2} := \sum_{i,j=1}^p (\mathbf{e}_i \mathbf{e}_j^t) \otimes (\mathbf{e}_j \mathbf{e}_i^t)$ with \mathbf{e}_i the i th vector in the canonical basis of \mathbb{R}^p .

Under these conditions, the corresponding estimator V_{n_k} is consistent and asymptotically normally distributed (details can be found in Aerts et al., 2016). Moreover, its ASV only depends on the choice of location and scatter estimators, on the dimension p , on the radial density g_k and on γ_k but not on the true parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, i.e.,

$$\text{ASV}_V(F_k; \gamma_k) := \tau_k \gamma_k^4 + (2\sigma_{1;k} + \sigma_{2;k}) \frac{\gamma_k^2}{4} \quad (2)$$

These conditions are satisfied by many location and covariance estimators under elliptical symmetry with finite fourth-order moments, in particular by the sample estimators, for which we have $\tau_k = 1$, $\sigma_{1;k} = 1 + \kappa_k$ and $\sigma_{2;k} = \kappa_k$ where κ_k is the kurtosis coefficient in the k -th population defined by

$$\kappa_k = \frac{p}{p+2} \frac{\mathbb{E}_{F_k}[D^4]}{\mathbb{E}_{F_k}[D^2]^2} - 1 \quad \text{where } D = \sqrt{(\mathbf{X} - \boldsymbol{\mu}_k)^t \boldsymbol{\Sigma}_k^{-1} (\mathbf{X} - \boldsymbol{\mu}_k)} \quad (3)$$

As already mentioned, when there are outliers in the data, the use of robust estimators is highly advisable. It is well-known that the median, the MAD and the IQR (multiplied by a consistency factor), as well as, in the multivariate setting, the robust S and MCD estimators, among others, or reweighted versions of them, satisfy (A1) to (A4). The corresponding constants $\tau_k, \sigma_{1:k}$ and $\sigma_{2:k}$ can be found in Falk (1997) for the median, IQR and MAD, in Lopuhää (1997) and Cator and Lopuhää (2010) for the S and MCD estimators respectively, or derived from Lopuhää (1999) for their reweighted version.

2.1 Wald test

The null hypothesis H_0 is equivalent to testing $H_0^1 : H\mathbf{\Gamma} = 0$ against $H_1^1 : H\mathbf{\Gamma} \neq 0$ where $\mathbf{\Gamma} = (\gamma_1, \dots, \gamma_K)^t$ and H is the matrix which constructs the $K - 1$ contrasts, i.e. $H\mathbf{\Gamma} = (\gamma_1 - \gamma_2, \dots, \gamma_1 - \gamma_K)^t$. The first proposed statistic to test H_0 or equivalently H_0^1 is a Wald-type test statistic defined by

$$W_N = \mathcal{V}_N^t H^t \mathbf{ASV}_{H\mathcal{V}}(\mathcal{F}; \mathbf{\Gamma})^{-1} H \mathcal{V}_N \quad \text{for } \mathcal{V}_N = (V_{n_1}, \dots, V_{n_K})^t$$

where $\mathbf{ASV}_{H\mathcal{V}}(\mathcal{F}; \mathbf{\Gamma}) = H \text{diag}(\text{ASV}_V(F_1; \gamma_1)/\lambda_1, \dots, \text{ASV}_V(F_K; \gamma_K)/\lambda_K) H^t$ with $\lambda_k = n_k/N$ and N is the total sample size, i.e. $N = \sum_{k=1}^K n_k$. In non-matrix form, the statistic can be expressed as a weighted sum of the squared differences between the estimators V_{n_k} and the estimator of a pooled coefficient of variation $\tilde{\gamma} = \frac{\sum_{k=1}^K w_k \gamma_k}{\sum_{k=1}^K w_k}$, i.e.

$$W_N = \sum_{l=1}^K w_l \left(V_{n_l} - \frac{\sum_{k=1}^K w_k V_{n_k}}{\sum_{k=1}^K w_k} \right)^2 \quad \text{where } w_k = \frac{\lambda_k}{\text{ASV}_V(F_k; \gamma_k)}.$$

In practice, the asymptotic variances $\text{ASV}_V(F_k; \gamma_k)$ need to be estimated. For any pair of location and covariance estimators and under an assumed underlying distribution F_k , the appropriate constants $\tau_k, \sigma_{1:k}$ and $\sigma_{2:k}$ may be plugged in (2) while the unknown γ_k is replaced by V_{n_k} . This plug-in technique leads to a statistic that will be denoted by \widehat{W}_N from now on.

When the sample estimator is used, a unique semi-parametric statistic, that is asymptotically valid under any K -fold product of elliptical distributions with finite fourth-order moments, can be defined. Indeed, as the ASV only depends on the kurtosis parameter κ_k , it suffices to replace it by a consistent estimator, e.g. $\hat{\kappa}_k = (p(p+2))^{-1} \frac{1}{n_k} \sum_{i=1}^{n_k} \hat{D}_{ki}^4 - 1$ where \hat{D}_{ki} stands for the sample version of D , as defined in (3), estimating $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ by their sample versions and evaluating the distance at \mathbf{X}_{ki} .

When the assumed distributions F_1, \dots, F_K share the same radial density, an alternative test statistic may be constructed extending the proposal of Feltz and Miller (1996). Under this particular setting, under H_0 , all the asymptotic variances $\text{ASV}_V(F_k; \gamma_k)$ reduce to the same quantity, denoted as $\text{ASV}_V(F_1; \gamma)$ to stress its dependence on the model and the unknown common MCV, γ say. An alternative test statistic consists in replacing the matrix $\mathbf{ASV}_{H\mathcal{V}}(\mathcal{F}; \mathbf{\Gamma})$

by $\text{ASV}_V(F_1; \gamma) \mathbf{H} \text{diag}(\lambda_1^{-1}, \dots, \lambda_K^{-1}) \mathbf{H}^t$. The common coefficient γ can be estimated by $\tilde{V}^0 = \sum_{k=1}^K \lambda_k V_{n_k}$. All the theoretical results in the sequel can easily be adapted to this statistic, which is asymptotically equivalent to \widehat{W}_N . This alternative test is not further considered here as it does not outperform the tests proposed in the next subsection. It is however worthwhile mentioning that it is more powerful than \widehat{W}_N when the γ_k values are big.

2.2 Wald tests for inverse MCVs

The null hypothesis is equivalent to the equality of inverse MCVs. In the univariate setting, several authors constructed tests for comparing inverse CVs and this idea can also be followed in the multivariate context. When V_{n_k} is the plug-in estimator based on some location and scatter estimators \mathbf{T}_{n_k} and \mathbf{C}_{n_k} satisfying (A1) to (A4), its inverse is also consistent and asymptotically normal with asymptotic variance $\text{ASV}_V(F_k; \gamma_k)/\gamma_k^4$. The second proposed statistic to test H_0 , which is a straightforward adaptation of the statistic W_N is

$$W_{N,\text{inv}} = (\mathcal{V}_N^{-1})^t \mathbf{H}^t \mathbf{ASV}_{\mathbf{H}\mathcal{V}^{-1}}(\mathcal{F}; \mathbf{\Gamma})^{-1} \mathbf{H}(\mathcal{V}_N^{-1})$$

where $\mathcal{V}_N^{-1} = (V_{n_1}^{-1}, \dots, V_{n_K}^{-1})^t$ and

$$\mathbf{ASV}_{\mathbf{H}\mathcal{V}^{-1}}(\mathcal{F}; \mathbf{\Gamma}) = \mathbf{H} \text{diag} \left(\frac{\text{ASV}_V(F_1; \gamma_1)}{\lambda_1 \gamma_1^4}, \dots, \frac{\text{ASV}_V(F_K; \gamma_K)}{\lambda_K \gamma_K^4} \right) \mathbf{H}^t$$

The version computed with estimated ASVs is denoted by $\widehat{W}_{N,\text{inv}}$.

3 Asymptotic distributions of the test statistics

In this section, the asymptotic distributions of the first test statistic \widehat{W}_N based on estimated asymptotic variances will be derived under the null hypothesis, a fixed alternative or a sequence of contiguous alternatives. This allows to study the asymptotic level and power of the proposed tests. The results for $\widehat{W}_{N,\text{inv}}$, which are straightforward, are not detailed here.

Proposition 1 *In the context described in Section 2, the null asymptotic distribution of the statistic $N \widehat{W}_N$ is a chi-square with $K - 1$ degrees of freedom.*

Proof The result follows readily from the asymptotic normality of V_{n_k} for each k , the plug-in estimator of the asymptotic variance $\mathbf{ASV}_{\mathcal{V}}(\mathcal{F}; \mathbf{\Gamma})$ being consistent when the V_{n_k} are consistent.

Let us now consider a fixed alternative hypothesis $\mathbf{\Gamma}^* = (\gamma_1^*, \dots, \gamma_K^*)$ with $\mathbf{\Gamma}^* \notin \Theta_0$ where $\Theta_0 \subset \mathbb{R}_+^K$ is the set of parameters satisfying the null hypothesis, i.e. $\Theta_0 = \{\mathbf{\Gamma} \in \mathbb{R}_+^K : \mathbf{H}\mathbf{\Gamma} = 0\}$. Theorem 5 in Basu et al. (2015) states that, under the true parameter value $\mathbf{\Gamma}^*$ and for w the operator defined by

$$w(\mathbf{\Gamma}_1, \mathbf{\Gamma}_2) = \mathbf{\Gamma}_1^t \mathbf{H}^t \mathbf{ASV}_{\mathbf{H}\mathcal{V}}(\mathcal{F}; \mathbf{\Gamma}_2)^{-1} \mathbf{H}\mathbf{\Gamma}_1, \quad (4)$$

we have $\sqrt{N} \left(\widehat{W}_N - w(\mathbf{\Gamma}^*, \mathbf{\Gamma}^*) \right) \xrightarrow{\mathcal{L}} N(0, \sigma_w^2(\mathbf{\Gamma}^*))$ for $\sigma_w^2(\mathbf{\Gamma}^*) = \left. \frac{\partial w(\mathbf{\Gamma}, \mathbf{\Gamma}^*)}{\partial \mathbf{\Gamma}} \right|_{\mathbf{\Gamma}=\mathbf{\Gamma}^*}^t \text{diag}(\text{ASV}_V(F_1; \gamma_1^*)/\lambda_1, \dots, \text{ASV}_V(F_K; \gamma_K^*)/\lambda_K) \left. \frac{\partial w(\mathbf{\Gamma}, \mathbf{\Gamma}^*)}{\partial \mathbf{\Gamma}} \right|_{\mathbf{\Gamma}=\mathbf{\Gamma}^*}$. Using this result, the power function of the statistic \widehat{W}_N under $\mathbf{\Gamma} = \mathbf{\Gamma}^*$ may be approximated, for a given nominal level $\alpha \in]0, 1[$, as

$$\mathbb{P}_{\mathbf{\Gamma}^*} \left[N \widehat{W}_N \geq \chi_{K-1, \alpha}^2 \right] \simeq 1 - \Phi \left(\frac{\sqrt{N}}{\sigma_w(\mathbf{\Gamma}^*)} \left(\frac{\chi_{K-1, \alpha}^2}{N} - w(\mathbf{\Gamma}^*, \mathbf{\Gamma}^*) \right) \right) \quad (5)$$

where $\mathbb{P}_{\mathbf{\Gamma}^*}[\cdot]$ denotes the probability computed under \mathcal{F} , assuming that $\mathbf{\Gamma} = \mathbf{\Gamma}^*$, and $\chi_{K-1, \alpha}^2$ is the $1 - \alpha$ quantile of the chi-square distribution with $K - 1$ degrees of freedom. It is therefore clear that the proposed tests are consistent since the limit of (5) when N goes to infinity is 1.

In the previous section, for any choice of location and scatter estimators, two test statistics have been suggested. In order to compare their performance for a given estimator or the performance of different estimators plugged into the same statistic, the *Pitman efficiency* might be useful. This measure requires to know the asymptotic distribution of the statistics under a sequence of contiguous alternatives converging to the null as N goes to infinity as

$$H_{1,N} : \mathbf{\Gamma}_N = \mathbf{\Gamma}_0 + \mathbf{\Delta}/\sqrt{N}. \quad (6)$$

Let $\mathbf{\Gamma}_0$ be an element of $\Theta_0 = \{\mathbf{\Gamma} \in \mathbb{R}_+^K : \mathbf{H}\mathbf{\Gamma} = 0\}$ and $\mathbf{\Delta} \in \mathbb{R}^K \setminus \Theta_0$ be any vector with at least two distinct components. Proposition 2 gives the asymptotic distribution of the proposed test statistics under (6).

Proposition 2 *Under the sequence $H_{1,N}$, the asymptotic distribution of the statistics $N \widehat{W}_N$ and $N \widehat{W}_{N;\text{inv}}$ is a non-central chi-square with $K - 1$ degrees of freedom and noncentrality parameter $\delta = \mathbf{\Delta}^t \mathbf{H}^t \mathbf{A} \mathbf{S} \mathbf{V}_{\mathbf{H}\mathcal{V}}(\mathcal{F}; \mathbf{\Gamma}_0)^{-1} \mathbf{H} \mathbf{\Delta}$.*

4 Pitman efficiencies

For consistent tests, an appropriate way to compare their asymptotic powers is to consider a sequence of contiguous alternatives $H_{1,N}$ under which the limiting power is bounded away from 1 and to compute their relative Pitman efficiency (ARE) defined by

$$\text{ARE}_{1,2}(H_{1,N}) = \lim_{N \rightarrow +\infty} \frac{N_2(N)}{N_1(N)}$$

where $N_2(N)$ is the number of observations required by test 2 to achieve the same power as test 1 when based on a sample size $N_1(N)$ under $H_{1,N}$. As shown in Puri and Sen (1971), when two test statistics both have an asymptotic non-central chi-square distribution with the same degree of freedom under the same sequence of contiguous alternatives, their Pitman efficiency coincides with the ratio of the noncentrality parameters. For a given choice of MCV estimator,

the statistics based either on the MCVs or on their inverses are asymptotically equivalent. Therefore, only finite-sample performance differences might be observed, as discussed at the end of Section 6.

Now, for a given test statistic, a choice has to be made for the MCV estimator. To guide this choice, the Pitman efficiency or identically, the ratio of the corresponding non-centrality parameters, can be computed. In the equi-radial density case, this ratio reduces to the asymptotic relative efficiency of the corresponding MCV estimators at the common coefficient of variation γ under H_0 . In other words, the ARE of one of the two proposed tests based on the functional V_1 to the same test based on the functional V_2 is given by

$$\text{ARE}_{1,2}(H_{1,N}) = \frac{\text{ASV}_{V_2}(F_1, \gamma)}{\text{ASV}_{V_1}(F_1, \gamma)}$$

Thus, it does not depend on the arbitrary vector Δ , nor on the number of populations K or sample proportions λ_k .

When $p = 1$, the most frequent pairs of consistent location and standard-deviation estimators used to compute the CV are the sample estimators, the pair median-IQR (IQRmed) or the pair median-MAD (MADmed), the two latter ones being asymptotically equivalent. The AREs of the tests based on these two robust CV estimators with respect to the same tests based on the sample CV are given in Table 2. In addition to the normal distribution, both heavier-tailed and lighter-tailed elliptical distributions are considered: the Student distribution with 5 degrees of freedom (t_5) and the power exponential distributions with parameter $\beta = 0.5$ (PE_{0.5}) or $\beta = 2$ (PE₂). The distributions are displayed in the table in the increasing order of the kurtosis. As expected, the robust tests are less efficient than the classical one under normality, especially for small values of γ . Under the heavier-tailed Student, the robust test is more efficient than the classical one, as long as Γ is not too big, while the contrary is true at both PE₂ and PE_{0.5}.

As noted by a referee, when $p = 1$, the order observed in Table 2 when considering the normal and exponential distributions coincides with the tail ordering $<_t$ defined by Loh (1984). For many location estimators, the AREs are order-preserving with respect to $<_t$. Determining whether this order-preserving is also verified for CV estimators is left for future-research.

In the multivariate setting, the robust and consistent estimators under consideration are the one-step reweighted MCD estimators¹ and the S estimators (with Tukey's biweight function) with 25% BDP. The AREs of the tests based on these robust estimators with respect to the classical one are listed in Table 3 for several values of p and γ and under the multivariate normal (Φ), Student (t_5) and power exponential (PE_{0.5} and PE₂) distributions. Under normal and Student distributions, the results are in line with those obtained by Croux and Haesbroeck (1999) when studying the asymptotic efficiencies of the one-step reweighted MCD and S scatter matrices. Under normality or under the

¹ at distribution F , the weight function was set to $I_{[0, q_\delta]}$ with $\delta = 0.025$ and $q_\delta = G^{-1}(1 - \delta)$ where $G(t) = \mathbb{P}_F[\mathbf{X}^t \mathbf{X} \leq t]$

γ	$p = 1$				
	0.01	0.1	0.5	1	2
PE ₂	0.277	0.280	0.326	0.370	0.398
Φ	0.368	0.371	0.428	0.512	0.589
PE _{0.5}	0.601	0.604	0.680	0.872	1.286
t_5	1.333	1.331	1.278	1.181	1.059

Table 2 Pitman's ARE of robust tests w.r.t classical tests - $p = 1$.

γ	$p = 3$			$p = 5$			$p = 10$		
	0.1	0.5	2	0.1	0.5	2	0.1	0.5	2
RMCD									
PE ₂	0.522	0.583	0.697	0.645	0.696	0.787	0.776	0.808	0.864
Φ	0.674	0.732	0.864	0.752	0.799	0.896	0.838	0.870	0.932
PE _{0.5}	0.873	0.905	1.011	0.892	0.920	0.996	0.930	0.950	0.995
t_5	2.100	1.931	1.359	2.128	1.947	1.347	2.060	1.894	1.334
S									
PE ₂	0.888	0.873	0.864	0.927	0.923	0.917	0.960	0.959	0.957
Φ	0.941	0.944	0.950	0.971	0.973	0.975	0.989	0.989	0.990
PE _{0.5}	1.095	1.114	1.172	1.092	1.099	1.118	1.060	1.061	1.064
t_5	2.444	2.233	1.539	2.487	2.270	1.560	2.508	2.288	1.568

Table 3 Pitman's ARE of robust tests w.r.t classical tests - Multivariate setting.

lighter-tailed power exponential distribution, the tests based on robust estimators are less efficient than the classical one, especially for small γ values. The tests based on the one-step reweighted MCD estimators only achieve a better power than the classical one under the Student distribution, while the use of S estimators leads to an outperformance of all the other tests at all considered heavier-tailed than normal distributions. As expected, tests based on estimators with breakdown point 0.5 (not shown here) are less efficient than those based on the 0.25 breakdown point estimators in all cases.

5 Robustness of testing procedures

As stated above, any pair of location and covariance estimators \mathbf{T}_{n_k} and \mathbf{C}_{n_k} yields a MCV estimator in the k -th population, and consequently, a test statistic. Since the classical estimators are extremely sensitive to outliers, testing procedures based on them may inherit their lack of robustness to atypical observations. In the current section, we formalize the local sensitivity of the proposed tests using the approach based on influence functions (Hampel et al., 1986), adapted to tests by Rousseeuw and Ronchetti (1981) and extended to the multidimensional parameter case by Heritier and Ronchetti (1994). In another context, Ghosh et al. (2016) already investigated the robustness of some Wald-type tests by means of influence functions and their analysis will be further adapted and transposed to the MCV multi-comparison problem.

5.1 Influence functions of the test statistics

Influence functions (IF) are local measures of robustness which describe the effect of an infinitesimal contamination on an estimator (Hampel et al., 1986). When considering several populations, the use of *partial* influence functions is advocated to measure resistance towards pointwise contamination in each population. Partial influence functions, although first introduced by Hampel et al. (1986), have been formalized by Pires and Branco (2002) in order to ensure that the usual properties of the influence function for the one-population case are still valid under the multi-populations case. Let \mathcal{S} be a sample space contained in \mathbb{R}^p and \mathcal{D} be the set of all finite-signed measures on \mathcal{S} . Let W be a statistical functional whose domain is the K -fold product $\mathcal{D} \times \dots \times \mathcal{D}$. Partial influence functions of W at the model $\mathcal{F} = F_1 \times \dots \times F_K$ are defined, when the limit exists, by

$$\text{IF}_k(\mathbf{x}, W, \mathcal{F}) = \lim_{\varepsilon \rightarrow 0} \frac{W(\mathcal{F}_{\varepsilon, \mathbf{x}}^k) - W(\mathcal{F})}{\varepsilon} = \left. \frac{\partial}{\partial \varepsilon} W(\mathcal{F}_{\varepsilon, \mathbf{x}}^k) \right|_{\varepsilon=0}$$

where $\mathcal{F}_{\varepsilon, \mathbf{x}}^k = F_1 \times \dots \times F_{k, \varepsilon, \mathbf{x}} \times \dots \times F_K$, and $F_{k, \varepsilon, \mathbf{x}} = (1 - \varepsilon)F_k + \varepsilon\Omega_{\mathbf{x}}$, $\Omega_{\mathbf{x}}$ being the Dirac distribution having all its mass at the point $\mathbf{x} \in \mathbb{R}^p$. More generally, we define the m -th order partial influence function as

$$\text{IFm}_k(\mathbf{x}, W, \mathcal{F}) = \left. \frac{\partial^m}{\partial \varepsilon^m} W(\mathcal{F}_{\varepsilon, \mathbf{x}}^k) \right|_{\varepsilon=0}$$

In this subsection, we will focus on the statistic \widehat{W}_N . The straightforward adaptation for the test based on inverse MCVs is left to the reader. The statistical functional related to the statistic \widehat{W}_N , computed with estimated ASV is given by

$$\widehat{W}(\mathcal{F}) = \mathcal{V}(\mathcal{F})^t \mathbf{H}^t \mathbf{A} \mathbf{S} \mathbf{V}_{\mathbf{H}\mathcal{V}}(\mathcal{F}; \mathcal{V}(\mathcal{F}))^{-1} \mathbf{H} \mathcal{V}(\mathcal{F}) \quad (7)$$

where $\mathcal{V}(\mathcal{F}) = (V(F_1), \dots, V(F_K))^t$, V being the MCV functional based on the location and covariance functionals \mathbf{T} and \mathbf{C} . When \mathbf{T} and \mathbf{C} are partially Fréchet differentiable and Fisher-consistent at the model \mathcal{F} , the influence function of this functional is identically null under H_0 . This comes from the fact that $\mathbf{H}\mathbf{\Gamma} = 0$ and implies that an influence analysis based on the first derivative is not adequate to quantify the robustness of the testing procedure. As in Hampel et al. (1986), the behaviour of the statistic under small amounts of contamination needs to be characterized by the second order partial influence function $\text{IF2}_k(\mathbf{x}, W, \mathcal{F})$, as given in Proposition 3 below. The computation of this influence function is straightforward (due to the simple form of the Wald statistic) and is therefore not detailed here.

Proposition 3 *Let the functional V be based on Fisher-consistent location and covariance functionals \mathbf{T} and \mathbf{C} at $F_k \forall k$, whose first and second partial order influence functions exist. Under the null hypothesis, the k -th second order partial influence function of W in $\mathbf{x} \in \mathbb{R}^p$ is given by*

$$\text{IF2}_k(\mathbf{x}, W, \mathcal{F}) = 2 (\text{IF}(\mathbf{x}, V, F_k))^2 (\mathbf{e}_k^t \mathbf{H}^t \mathbf{A} \mathbf{S} \mathbf{V}_{\mathbf{H}\mathcal{V}}(\mathcal{F}; \mathbf{\Gamma})^{-1} \mathbf{H} \mathbf{e}_k) \quad (8)$$

Expression (8) shows that, under the null, the test statistic will be resistant to a small amount of partial contamination in the k -th population as soon as the MCV estimator in the k -th population has a bounded influence function. The IFs of several MCV estimators can be found in Aerts et al. (2015) but the main message is the following: if the influence functions of the corresponding location and covariance functionals \mathbf{T} and \mathbf{C} are bounded, the influence function of the MCV will also be bounded. Usually, robust estimators have bounded influence functions.

Note also that, in the equi-radial density case, (8) reduces to $2(\text{IF}(\mathbf{x}, V, F_k))^2 \lambda_k(1 - \lambda_k)/\text{ASV}_V(F_1; \gamma)$. The effect of an infinitesimal contamination in the k -th group is thus smaller when the sample size of the group is very small or on the contrary, very big compared to the remaining ones. The influence is maximized when $\lambda_k = 0.5$. Under this most unfavorable setting, the second order partial influence functions of the Wald-type test statistic are plotted in Figure 5.1 for $K = 4$, $p = 2$, under (a) the PE_2 , (b) the normal and (c) the Student distributions with parameters $\boldsymbol{\mu}_k = \mathbf{1}_p$ and $\boldsymbol{\Sigma}_k = \mathbf{I}_p$, i.e. $\gamma = 1/\sqrt{p}$. The estimators under consideration are those introduced in Section 4. As expected, the IF computed with the classical estimators is not bounded while the other IFs are clearly bounded. The influence functions of the Wald-type test statistics based on the two robust estimators are quite similar, even if it is smoother when based on the S-estimator. They also look like the classical IF in the center of the domain. Let us note that the gain of efficiency of S estimators induces a loss in robustness since the downweighting of outliers occurs further away from the center than with the RMCD estimators.

As for the comparison between distributions, the local sensitivity of the test based on the classical and the S estimators is very similar under all the considered distributions. For the test statistics based on the RMCD estimators, under the leptokurtic Student and power exponential distributions, the downweighting happens for points somewhat further away from the center.

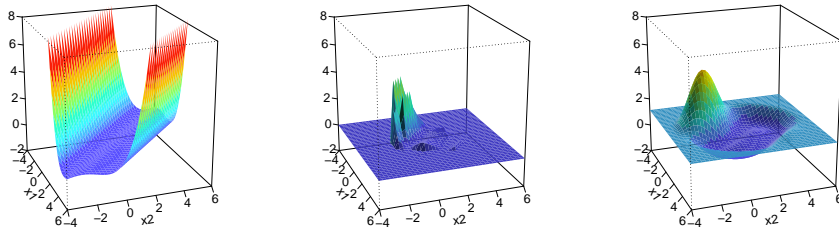
The effect of simultaneous contamination in all the populations may be investigated, following an idea introduced by Garcia-Perez (2012) or Ghosh and Basu (2013) in another context, by computing the second order *joint* influence function (in contrast with the partial IF) defined by

$$\text{JIF}2(\mathbf{x}_1, \dots, \mathbf{x}_K, W, \mathcal{F}) = \left. \frac{\partial^2}{\partial \varepsilon^2} W(F_{1,\varepsilon,\mathbf{x}_1} \times \dots \times F_{K,\varepsilon,\mathbf{x}_K}) \right|_{\varepsilon=0}$$

for $\mathbf{x}_1, \dots, \mathbf{x}_K$ being some p -dimensional contamination points in each group. Under the null hypothesis, this reduces to

$$\begin{aligned} & 2 \text{JIF}(\mathbf{x}_1, \dots, \mathbf{x}_K, \mathcal{V}, \mathcal{F})^t H^t \mathbf{A} \mathbf{S} \mathbf{V}_{\mathbf{H}\mathcal{V}}(\mathcal{F}; \mathbf{\Gamma})^{-1} H \text{JIF}(\mathbf{x}_1, \dots, \mathbf{x}_K, \mathcal{V}, \mathcal{F}) \\ &= 2 \sum_{k=1}^K w_k \left(\text{IF}(\mathbf{x}_k, V, F_k) - \frac{\sum_{l=1}^K w_l \text{IF}(\mathbf{x}_l, V, F_l)}{\sum_{l=1}^K w_l} \right)^2 \end{aligned}$$

with $\text{JIF}(\mathbf{x}_1, \dots, \mathbf{x}_K, \mathcal{V}, \mathcal{F}) = (\text{IF}(\mathbf{x}_1, V, F_1), \dots, \text{IF}(\mathbf{x}_K, V, F_K))^t$, all the other quantities being defined previously. The joint influence function is therefore

(a) PE_2 

(b) Normal

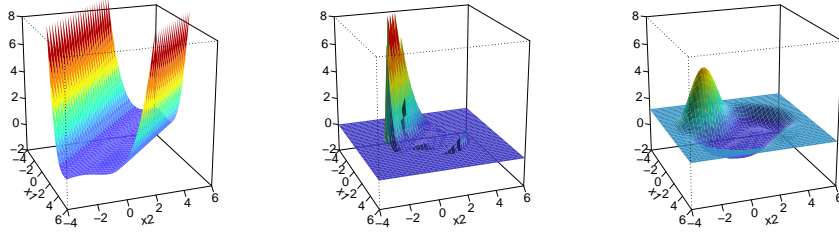
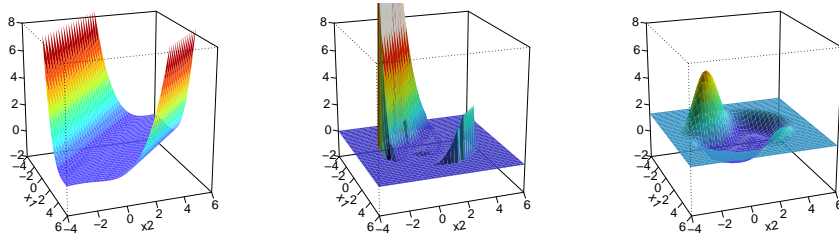
(c) t_5 

Fig. 1 Second order partial influence function of the test based on the classical (left panel), the reweighted MCD (middle) or the S estimators (right panel) with 25% breakdown point.

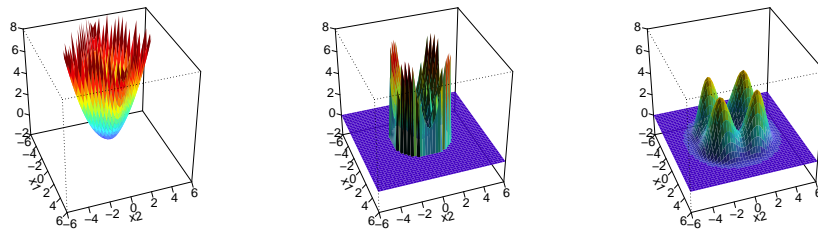


Fig. 2 Second order joint influence function of test based on the classical (left panel), the reweighted MCD (middle) or the S estimators (right panel) with 25% breakdown point.

a weighted sum of the squared differences between the IFs of V in each population and a pooled IF. The graphical representation of this JIF is more complex since it depends on the location of K p -variate points. Using the same settings as for Figure 1, the second order JIF has been plotted under normality for contamination points located in opposite directions around $\boldsymbol{\mu}_k$, i.e. $\mathbf{x}_1 = (1 + x_1, 1 + x_2)$, $\mathbf{x}_2 = (1 - x_1, 1 - x_2)$, $\mathbf{x}_3 = (1 + x_1, 1 - x_2)$ and $\mathbf{x}_4 = (1 - x_1, 1 + x_2)$, for x_1 and x_2 varying along the axis. This joint influence approach may also be applied in the next section for level and power influence functions. Since trends remain similar, i.e. the RMCD and S estimators down-weight severe outliers, whether in one population or in all the populations, the corresponding expressions, which are straightforward, will not be presented.

5.2 Level and power influence functions

According to Heritier and Ronchetti (1994), robust tests should keep a stable level under small arbitrary departures from the null (*robustness of validity*) and a good power under small arbitrary departures from specified alternatives (*robustness of efficiency*). Their idea is to compute the asymptotic level and power of the test under contamination and compare them to the nominal ones.

In order to study the performance of the test at alternatives that are close to H_0 , it is adequate to compute power functions at contiguous alternatives $H_{1,N}$, as defined in (6). The contamination must now be chosen in such a way that it converges to zero at the same rate as the sequence $H_{1,N}$ converges to the null in order to avoid any overlapping between the null neighborhoods and the alternative neighborhoods. We thus consider the sequences of partial contamination, respectively for the level and power: $\mathcal{F}_{i,\varepsilon,N,\mathbf{x}}^k = F_{i,1} \times \cdots \times F_{i,k,\varepsilon,N,\mathbf{x}} \times \cdots \times F_{i,K}$ for $i = 0, 1$ where $F_{0,k}$ stands for the model in the k -th population under H_0 , $F_{1,k}$ stands for the model under the alternative $H_{1,N}$ while $F_{i,k,\varepsilon,N,\mathbf{x}} = \left(1 - \varepsilon/\sqrt{N}\right) F_{i,k} + (\varepsilon/\sqrt{N})\Omega_{\mathbf{x}}$ with $\mathbf{x} \in \mathbb{R}^p$, $i = 0, 1$.

For $\alpha(\mathcal{F})$ and $\beta(\mathcal{F})$ being the level and power at the model \mathcal{F} , the k -th partial level and power IFs are defined, following Hampel et al. (1986), by

$$\text{LIF}_k(\mathbf{x}, W, \mathcal{F}_0) = \frac{\partial}{\partial \varepsilon} \left(\lim_{N \rightarrow +\infty} \alpha(\mathcal{F}_{0,\varepsilon,N,\mathbf{x}}^k) \right) \Big|_{\varepsilon=0} \quad (9)$$

$$\text{PIF}_k(\mathbf{x}, W, \mathcal{F}_1) = \frac{\partial}{\partial \varepsilon} \left(\lim_{N \rightarrow +\infty} \beta(\mathcal{F}_{1,\varepsilon,N,\mathbf{x}}^k) \right) \Big|_{\varepsilon=0} \quad (10)$$

Let β_0 be the asymptotic power under the sequence $H_{1,N}$ of the test based on W and α_0 be the nominal level. Provided that

$$(B1) \quad \sqrt{N} \left(\mathcal{V}_N - \mathcal{V}(\mathcal{F}_{i,\varepsilon,N,\mathbf{x}}^k) \right) \xrightarrow{\mathcal{L}} \mathcal{N}_{K-1}(0, \text{ASV}_{\mathcal{V}}(\mathcal{F}_0; \boldsymbol{\Gamma}_0)),$$

$$(B2) \quad \text{ASV}_{\mathcal{H}\mathcal{V}}(\mathcal{V}_N) \xrightarrow{\mathbb{P}} \text{ASV}_{\mathcal{H}\mathcal{V}}(\mathcal{F}_0; \boldsymbol{\Gamma}_0),$$

for $i = 0, 1$, uniformly over the ε contamination of \mathcal{F}_i , a straightforward adaptation of Proposition 4 in Heritier and Ronchetti (1994) to the multi-sample

design scheme and for statistics with estimated ASVs yields,

$$\begin{aligned} \lim_{N \rightarrow +\infty} \beta(\mathcal{F}_{1,\varepsilon,N,\mathbf{x}}^k) &= \beta_0 + \varepsilon \, 2d_1 \, \text{IF}(\mathbf{x}, V, F_{0,k}) \, \{ \mathbf{e}_k^t \mathbf{H}^t \mathbf{A} \mathbf{S} \mathbf{V}_{\mathbf{H}\mathcal{V}}(\mathcal{F}_0; \mathbf{\Gamma}_0)^{-1} \mathbf{H} \Delta \} \\ &\quad + \varepsilon^2 \, [\text{IF}(\mathbf{x}, V, F_{0,k})]^2 \left\{ d_2/2 \, (\mathbf{e}_k^t \mathbf{H}^t \mathbf{A} \mathbf{S} \mathbf{V}_{\mathbf{H}\mathcal{V}}(\mathcal{F}_0; \mathbf{\Gamma}_0)^{-1} \mathbf{H} \Delta)^2 \right. \\ &\quad \left. + d_1 \, (\mathbf{e}_k^t \mathbf{H}^t \mathbf{A} \mathbf{S} \mathbf{V}_{\mathbf{H}\mathcal{V}}(\mathcal{F}_0; \mathbf{\Gamma}_0)^{-1} \mathbf{H} \mathbf{e}_k) \right\} \end{aligned}$$

where, for G_δ being the cdf of the non-central chi-square distribution $\chi_{K-1}^2(\delta)$ and q_{α_0} the $1 - \alpha_0$ quantile of the central chi-square distribution χ_{K-1}^2 , $d_1 = -(\partial/\partial\delta)G_\delta(q_\alpha)|_{\delta=\delta^*}$ and $d_2 = -(\partial^2/\partial\delta^2)G_\delta(q_\alpha)|_{\delta=\delta^*}$, with $\delta^* = \mathbf{\Delta}^t \mathbf{H}^t \mathbf{A} \mathbf{S} \mathbf{V}_{\mathbf{H}\mathcal{V}}(\mathcal{F}_0; \mathbf{\Gamma}_0)^{-1} \mathbf{H} \mathbf{\Delta}$. The result for the asymptotic level can be obtained by replacing Δ by 0.

Plugging this expression in definitions (10), it can be seen that the partial level influence function is identically null, meaning that the asymptotic level of the test will not be affected by a contiguous contamination. As for the partial power influence function, which is proportional to the influence function of the MCV functional in population k , the use of a B-robust MCV estimator suffices to yield *efficiently robust* tests.

6 Monte Carlo Study

In this section, a simulation study is performed in order to assess the performance of the proposed tests. The goal is three-fold. First, the finite-sample performance of the tests, both in terms of level and power, is examined. Then, as the classical location and covariance estimators are known to be extremely sensitive to the presence of outliers, the stability of the tests based on them or on some robust counterparts is examined under contamination. For these first two objectives, the equi-radial density scheme will be assumed. Finally, the conceptual distinction between comparing MCVs and testing the homogeneity of covariance matrices is highlighted. Simulated data will be used to compare the results of the MCV tests with those obtained with the optimal tests of covariance homogeneity proposed by Hallin and Paindaveine (2009).

6.1 Uncontaminated scheme - Level

In order to study the convergence speed to the null asymptotic χ_{K-1}^2 distribution, we performed $M = 1000$ replications of $K = 2$ independent samples of equal size $n = N/K$ from K p -variate populations F_k , $k = 1, \dots, K$, with mean vector $\boldsymbol{\mu}_k = (1/\gamma)\mathbf{e}_1$ and covariance matrix $\Sigma_k = \mathbf{I}_{p \times p}$ where γ is the common MCV. This simple setting is not restrictive since, under elliptical symmetry, the finite-sample distribution of the MCV estimators under consideration can be shown to only depend on $\boldsymbol{\mu}_k$ and Σ_k through γ .

In addition to the normal, the Student (t_5) and the power exponential distributions ($\text{PE}_{0.5}$, PE_2) defined in Section 4 were considered. Several sample sizes, dimensions, and values of the common MCV were studied: $n =$

50, 200, 300, $p = 1, 5$ and $\gamma = 0.1, 0.5, 2$. The size of the tests was set to 5%. For each test statistic, together with the classical version, the same robust estimators as those discussed in Section 4 were examined, i.e. the IQRmed and MADmed estimators when $p = 1$, the one-step reweighted MCD and the S estimators with 25% BDP in the multivariate setting. The performance of the semi-parametric tests was also investigated.

For the sake of conciseness, the detailed simulation results are given in Online Resource 1 but the main conclusions are the following. Whatever the dimension, when the common MCV (or CV) takes small values (< 1), the classical tests have rejection percentages close to the nominal level except at the Student, under which they are too conservative especially in small samples. This extremely bad behavior of may be explained as follows: under the Student, the classical MCV estimator has a positive skewed finite-sample distribution. Given the low asymptotic efficiency of the sample covariance matrix at this distribution (the two constants $\Sigma_{1;k}$ and $\Sigma_{2;k}$, introduced in condition (A4), are large), the asymptotic variances are largely over-estimated compared to the theoretical value. Therefore, the test statistics \widehat{W}_N takes lower than expected values, leading to fewer rejections than wanted.

Under the normal or the power exponential distributions, the classical tests outperform the robust versions even though the univariate MADmed estimator turns out to be quite competitive, while the S-estimator clearly consists of an interesting challenger in the multivariate setting. However, the multivariate robust tests based on the RMCD estimator require a ratio n/p not too small to perform reasonably well but they also become much too liberal under the PE_2 distribution. It is not such a surprise as the limiting case (when β tends to infinity) of the power exponential distribution is a uniform distribution on an ellipsoid, configuration under which the MCD estimator is bound to becoming very unstable. It suffers already of this lack of efficiency under the PE_2 distribution and this, coupled with a negatively skewed finite-sample distribution for the MCV estimator, yields estimated ASVs which are very small compared to the theoretical value. This implies large values for the test statistics.

While performing quite similarly as their non-inverted counterpart for γ values strictly smaller than 1, only the tests based on the inverse MCVs perform decently for greater values of γ , especially in small dimension. The other tests completely fail to get close to the nominal level of 5% except under the $PE_{0.5}$ distribution. Indeed, for large γ values, especially in small dimension, the ASV of the non-inverted MCV is overestimated and the distribution of the ASV estimator is much more skewed than for the inverse-MCD ASV estimator. This skewness is worse as γ increases, in small dimension or under the Student distribution.

6.2 Power curves

The plots on the first row of Figures 3 and 4 allow to study the power performance of the tests. It is assumed that $K = 2$, $n = 200$ and two different

dimensions are considered, i.e. $p = 1$ (in Figure 3) or $p = 5$ (in Figure 4). The null hypothesis is $H_0 : \gamma_1 = \gamma_2$ with common MCV $\gamma_1 = 1$, while the alternative is given by $H_1 : \gamma_2 = \gamma_1 + \Delta$, for specific values of Δ . More precisely, in these plots, the observed frequencies of rejection have been plotted with respect to Δ for Δ ranging from -0.9 to 0.9 . Only the results for the tests $\widehat{W}_{N,\text{inv}}$, which showed better results in the previous subsection, are presented considering several estimators. The power curves of the other statistic are not drawn as they were really similar. A comparison between the two statistics in terms of power is however outlined at the end of this subsection.

One first notices that these observed power curves are not symmetric about 0 (i.e. not symmetric about the discrepancy Δ between the MCVs). As observed previously, the higher the MCV values, the more conservative the tests. Then, as expected, in the univariate setting, the robust tests are less powerful than the classical ones under normality or lighter-tailed distributions, while they perform similarly under heavier-tailed distributions. In higher dimension, the robust and the classical tests show similar performance under normality or the slightly heavier-tailed $\text{PE}_{0.5}$ distribution. As already mentioned before, under the PE_2 distribution, the test based on the RMCD estimators is a bit too liberal under H_0 . Its power curve shows that it is also too conservative under the alternative. Under the heavier-tailed t_5 distribution, the robust tests are more powerful than the classical one.

The disastrous effect of contamination on the performance of the classical test can be observed from the plots on the second row of Figures 3 and 4. The same simulation settings were used but this time, the data were contaminated with 5% outliers in one group. Under $F_{\boldsymbol{\mu}_k, \Sigma_k}$, the outliers were drawn from the same family of distributions with the same mean vector $\boldsymbol{\mu}_k$ but with the inflated covariance matrix $\Sigma = 100\Sigma_k$.

The case $\Delta = 0$ allows to examine the finite-sample stability of the level under contaminated scenarios. We observe that the classical and the semiparametric tests break in level since the rejection frequencies are close to 1. As for the power, the outliers succeed in shifting the power curves of the classical tests (the semi-parametric version as well) in such a way that these become almost useless. On the other hand, in dimension $p = 1$, the robust tests show similar rejection frequencies than those achieved in absence of contamination. As for the robust estimators in the multivariate setting, although much more robust than the classical ones, the tests based on S estimators seem to be more influenced by severe contamination than the one using the RMCD estimators.

For a fixed choice of estimators, it is also interesting to compare the performance of the two proposed statistics \widehat{W}_N and $\widehat{W}_{N,\text{inv}}$ under increasingly heterogeneous alternatives (increasing values of $|\Delta|$). In a few words (simulation results not shown here), for negative values of Δ , i.e. for values of γ_1 and γ_2 smaller than 1, the two statistics yields very similar rejection frequencies. Only the comparison in presence of larger γ_1 and γ_2 values allows to pinpoint a statistic over another. In this case, whatever the MCV estimator and dimension, the tests based on the statistic \widehat{W}_N are too conservative. Working with

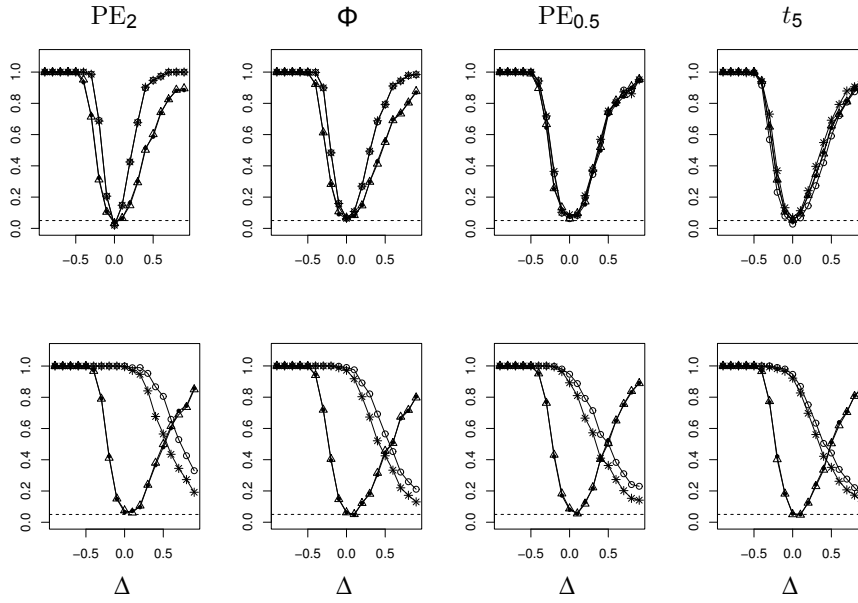


Fig. 3 Observed rejection frequencies under alternatives ($K = 2, p = 1, n = 200$) for the semi-parametric (*), classical (o), IQRmed (Δ) and MADmed (\bullet) tests in uncontaminated (top) and contaminated (5%) samples (bottom).

inverse MCVs yields more powerful tests and the discrepancy is larger under normality or under the Student distribution.

6.3 Comparison with covariance homogeneity tests

In the multivariate setting, many tests of covariance homogeneity are available in the literature, e.g. the well-known Bartlett (Bartlett, 1937) and Schott (Schott, 2001) tests under normality, or the optimal tests proposed by Hallin and Paindaveine (2009) which are valid under any K -tuple of elliptical distributions. In this subsection, we focus on the latter to briefly stress the differences and complementarities between these tests and the MCV ones.

First, simulations were conducted under settings similar to those illustrated in Hallin and Paindaveine, i.e. we generate $M = 500$ data sets consisting of $K = 2$ independent samples of respective sizes $n_1 = 200$ and $n_2 = 50$. These samples were simulated from two bivariate populations F_1 and F_2 (both Gaussian in the first case and both t_5 in the second case) with mean vectors $\mu_1 = \mu_2 = \mathbb{1}_p$ and covariance matrices $\Sigma_1 = \mathbb{I}_p$ and $\Sigma_2 = (1 + ls^2)(\Sigma_1 + l\mathbf{B})$ with $l = 1, \dots, 30$ where \mathbf{B} is a symmetric matrix such that $\text{tr } \mathbf{B} = 0$. Varying the parameter l allows to produce increasingly heterogeneous alternatives (in terms of covariance), either when l increases or decreases depending on \mathbf{B} . As far as relative dispersion is concerned, the MCV of the first population is

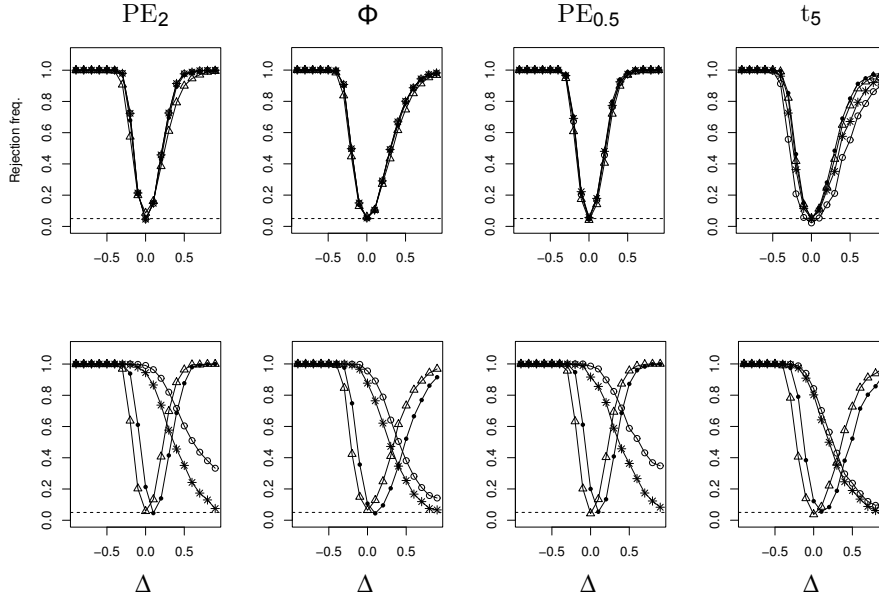


Fig. 4 Observed rejection frequencies under alternatives ($K = 2, p = 5, n = 200$) for the semi-parametric (*), classical (o), RMCD (Δ) and S (\bullet) tests in uncontaminated (top) and contaminated (5%) samples (bottom)

simply equal to $1/\sqrt{2}$ while it is easier to derive the coefficient of the second population when its covariance structure is further specified.

Three types of alternatives were investigated by Hallin and Paindaveine : (a) pure scale alternatives, for which $s^2 = 0.05$ and $s^2 = 0.08$ under Gaussian-Gaussian and $t_5 - t_5$ densities respectively, and \mathbf{B} is null ; (b) pure shape alternatives, for which $s = 0$ and \mathbf{B} is diagonal with elements -0.03 and 0.03 ; and (c) mixed alternatives, for which $s^2 = 0.04$ and $s^2 = 0.06$ for Gaussian-Gaussian and $t_5 - t_5$ densities respectively, and \mathbf{B} is diagonal with elements -0.02 and 0.02 . These schemes yield

- (a) $\gamma_2 = \sqrt{1 + ls^2}/\sqrt{2},$
- (b) $\gamma_2 = 1/\sqrt{1/(1 - 0.03l) + 1/(1 + 0.3l)}$
- (c) $\gamma_2 = \sqrt{1 + ls^2}/\sqrt{1/(1 - 0.02l) + 1 + (1 + 0.02l)}$

In the Gaussian-Gaussian case, the tests under consideration are the classical Wald test based on MCV and its inversed version as well as the Gaussian test from Hallin and Paindaveine (2009). In the Student-Student case, the semi-parametric Wald test (and its inverse-MCV version) is compared to Hallin and Paindaveine's semi-parametric homokurtic test. In Figure 5, the rejection frequencies of all the tests are plotted as a function of l . The corresponding MCV values in the second population are displayed on the top axis.

In all cases, the MCV tests and the covariance homogeneity test have a Type I risk close to the nominal level under the null ($l = 0$). Under all schemes, the covariance homogeneity tests are more powerful than the MCV ones, especially under mixed alternatives. This is mainly due to the dimension reduction resulting from the computation of MCVs: such a type of differences between covariance structures yield close MCV values that are hardly distinguishable on the basis of such a small sample size in one of the populations ($n_2 = 50$). Moreover, as observed in the previous subsection, the bigger the MCV values, the less performant the test to detect small departures from the null. Note that under pure scale alternatives (γ_2 close to 1 or larger), it is preferable to work with inverse MCVs, as noticed in the previous subsections. Under pure shape alternatives, which correspond to larger MCV deviations the proposed tests have Type I risks closer to those of the covariance homogeneity tests.

In view of the above results, when the aim is to detect differences in the (absolute) dispersion of several populations, we advocate the use of covariance homogeneity tests, which are more powerful, rather than MCV tests. However, when interest is in the comparison of *relative* dispersion, resorting to a MCV test is unavoidable since populations with distinct means and absolute dispersions may have similar MCVs. As illustration, under the Gaussian-Gaussian assumption, consider now a mean vector $\boldsymbol{\mu}_2 = (1/\sqrt{l})\boldsymbol{\mu}_1$ and covariance matrix $\Sigma_2 = l\Sigma_1$ for the second population, while keeping the same parameters as before for the first population. Therefore, in terms of relative dispersion, the two populations share the same MCV $\gamma_1 = \gamma_2 = 1/\sqrt{p}$. In Figure 6, the rejection frequencies are plotted again wrt to l with l ranging from 1 to 2. While the Gaussian covariance homogeneity test and the usual heteroscedastic Hotelling test of equality of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ (Krishnamoorthy and Yu, 2004) suggest some significant differences between the covariance or mean of the two populations (the rejection rates going up as l goes from 1 to 2), the MCV tests correctly detect homogeneity of the two populations in terms of relative dispersion.

7 Real-life application

Multivariate coefficients of variation have recently been proposed in External Quality Assessment programs for comparing the performance of assay techniques used by clinical laboratories. For instance, Zhang et al. (2010), noting that EQA datasets often entail outliers or blunders, advocate the use of robust MCVs to compare serum protein electrophoresis techniques. A measure of electrophoretic profile consists in 5 fractions, summing up to 100%, corresponding to the proportion of albumin and $\alpha_1, \alpha_2, \beta$ and *gamma*-globulins in the blood. Due to the compositional nature of the data, an isometric log-ratio (ILR) transformation is first applied to transform these five variables into four non-linearly dependent ones before computing the MCV of each technique. The MCV, as an overall measure of inter-laboratory repeatability, enables to rank the techniques. However, until now, no statistical test was available in the literature in order to highlight significant differences between the techniques.

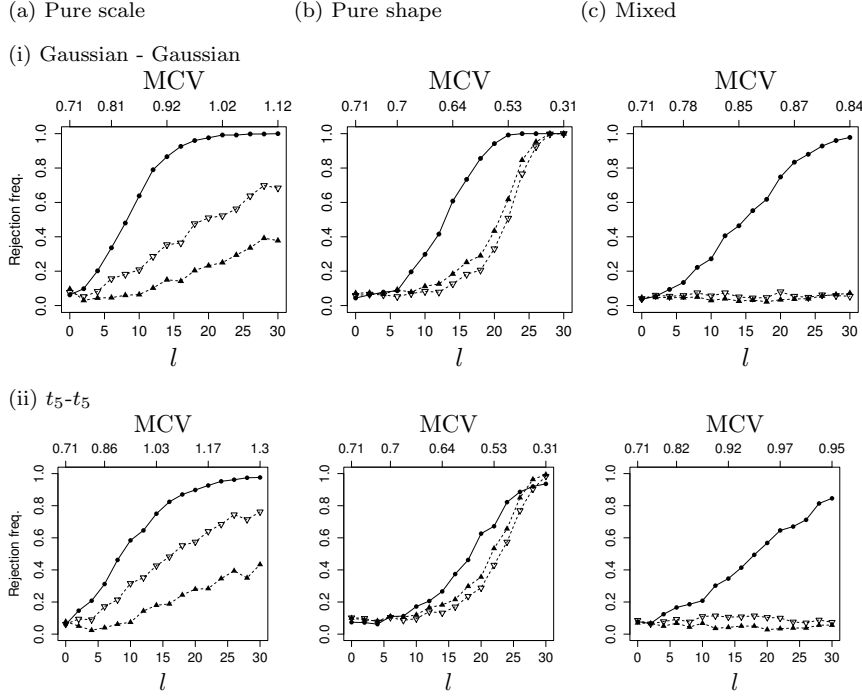


Fig. 5 Observed rejection frequencies wrt l , under the scale (a), shape (b) or mixed (c) alternatives for Hallin and Paindaveine's test (●), classical Wald test (▲) and inverse Wald test (▽)

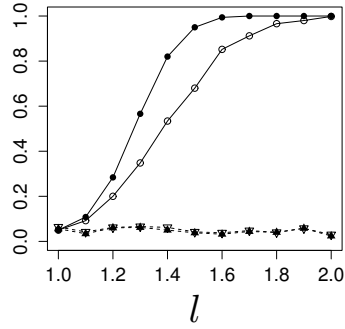


Fig. 6 Observed rejection frequencies ($\Phi - \Phi$ case) for the the heteroscedastic Hotelling test for $\mu_1 = \mu_2$ (○), Hallin and Paindaveine test (●), or classical Wald test (▲) and inverse Wald test (▽)

The data set used herein, obtained from Zhang et al. (2010), focuses on four techniques using distinct support medium, staining color or analytical principle. These techniques will be denoted simply as CH, EH, JW and GB (the full names are reported in Table 6). Computing the robust Mahalanobis distances of the observations reveals some clear outliers. Table 4 reports the MCV estimates computed using the classical, RMCD and S estimators as well

Electrophoretic technique	n	$n_{\text{out.}}$	class.	class.-out	RMCD	S
HT Cellulose acetate (CH)	141	133	.0780	.0688	.0624	.0745
HT Agarose gel (Acid blue) (EH)	121	112	.0676	.0558	.0525	.0641
HT Agarose gel (Amido black) (JH)	88	74	.0824	.0616	.0595	.0760
BCP Capillary zone (GB)	65	62	.0250	.0238	.0224	.0259

Table 4 MCV estimates - EQA application ($p = 4$)

Test	class.	Semiparam.	class.-outl.	Semiparam.-outl.	RMCD	S
H_0	.000	.000	.000	.000	.000	.000
CH vs EH	.104	.224	.021	.026	.051	.091
CH vs JH	.579	.739	.278	.318	.253	.842
CH vs GB	0	0	0	0	0	0
EH vs JH	.053	.240	.355	.386	.515	.098
EH vs GB	0	0	0	0	0	0
JH vs GB	0	0	0	0	0	0

Table 5 p-values for the multiple and pairwise comparisons tests - EQA application ($p = 4$)

as by means of the classical estimator applied to the clean data set (data set without the outliers). Although the ranking remains the same without the outliers, it can be noticed that the classical estimator is heavily influenced by their presence. The RMCD estimate and the clean-classical estimate take similar values, while the S estimate yields higher relative dispersion measures. Before turning to the comparison of the methods by means of tests, let us note that all estimated values are quite small and therefore, applying the test based on \widehat{W}_N is an appropriate option that was taken here. In Table 5, the p-values of the related statistics \widehat{W}_N and of the semi-parametric test, which is also based on the classical estimator, are reported for the test $H_0 : \gamma_{\text{CH}} = \gamma_{\text{EH}} = \gamma_{\text{JW}} = \gamma_{\text{GB}}$ versus H_1 : *at least one of these significantly differ from another one*. The p-values for the pairwise comparison tests are also given. Although all the tests agree (at level 5%) to reject the fact that all techniques are similarly performant, the choice of the estimator may change the decision of some pairwise comparison tests. Indeed, while the classical test on the complete dataset does not reject the fact that methods CH and EH have the same inter-laboratory repeatability, the decision is much more nuanced when using robust estimators or when the outliers are eliminated. Also, for the comparison of the techniques EH and JH, the outliers have an impact on the confidence with which the decision of non-rejection of the classical test is taken.

In order to assess the effect of contamination on the decision, one observation was added (before the ILR transformation) to the measurements of the best-performing technique, i.e. BCP Capillary zone (GB), at a position $0.25(100 - s, 4s, 100 - s, 100 - s, 100 - s)$ consistent with the compositional nature of the EQA measurements. The parameter s was varied from 5 to 95 by steps of five. In Figure 7, the symbol \circ corresponds to the classical MCV estimate, while symbols \triangle and \bullet correspond to the RMCD and S estimates. The MCV estimates and the p-values corresponding to the robust tests remain

stable whatever the position of the added observation. Instead, the classical estimator is highly influenced by the position of the single outlier. Consequently, the p-values of the parametric or semi-parametric tests based on this estimator fluctuate strongly with the position of the added datapoint, and the decision changes from rejection to non rejection in the three cases. Therefore, the classical test fails to detect the GB technique as the one with the highest repeatability because of the presence of only one misplaced outlier or blunder.

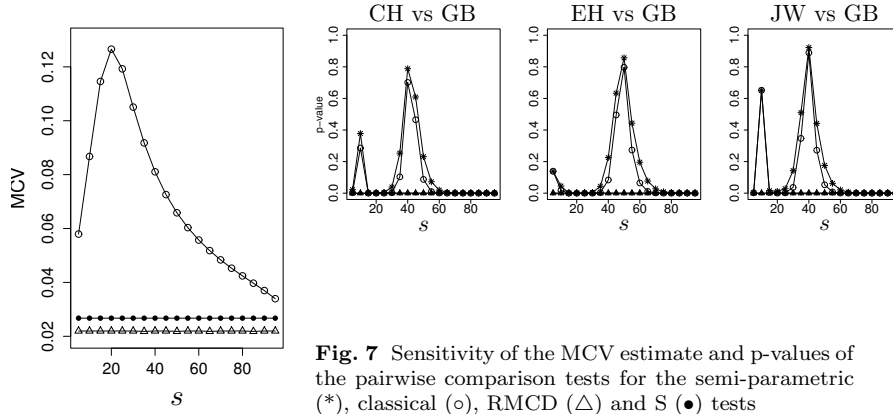


Fig. 7 Sensitivity of the MCV estimate and p-values of the pairwise comparison tests for the semi-parametric (*), classical (\circ), RMCD (\triangle) and S (\bullet) tests

8 Conclusion

Univariate and multivariate coefficients of variation are often used to compare the relative dispersion in several populations on the basis of one or several attributes. However, in the multivariate setting, as far as we know, no test for the equality of K multivariate coefficients of variation may be found in the literature. In this paper, some classical and robust Wald-type tests have been introduced and compared, as well as a semi-parametric test valid under any K -tuple of elliptical distributions. As a by-product, in the univariate case, these tests offer a robust alternative to the already existing tests which are mostly unreliable in presence of contamination.

The asymptotic distribution of the proposed test statistics was investigated under the null as well as under fixed or contiguous alternatives, which allowed to study the asymptotic relative efficiencies of the robust tests compared to the classical ones. Robustness of the testing procedures was investigated by deriving the partial and joint influence functions of the test statistic and the partial power and level influence functions.

The simulation section confirmed the expected unreliable behaviour of the classical tests under contamination while the robust tests were shown to remain stable in both level and power. In the multivariate setting, the S estimator has some interesting advantages over the RMCD estimator, i.e. it yields smooth

influence functions, it is more efficient as shown in Section 5 and it better handles situations where n/p is small under normality and under the PE_2 distribution. However, the drawback is that the corresponding test can be more affected by some severe contamination than the one based on the RMCD estimator. Knowing that, the choice of the S estimator over the RMCD should be based on the practical situation and the desired trade-off between robustness and efficiency. Future research may highlight an estimator yielding tests that are both highly robust and efficient. Then, after one finalizes the choice of an estimator, as regards the choice of the test statistic to be used among the two proposed, although similarly performant for small MCVs in all the populations, we advocate the use of tests based on inverse MCVs, which have been shown empirically more powerful in finite-samples for large γ values (> 1).

The distinction between those tests and the usual tests of covariance homogeneity has been highlighted. When the aim is to detect subtle differences in the covariance structure, it is preferable to use tests intended for this purpose, such as those proposed by Bartlett (1937), Schott (2001) or Hallin and Paindaveine (2009) as they are more powerful. However, when relative dispersion is the goal, the proposed tests were shown to be more appropriate.

Finally, through a real dataset, we illustrate how this paper meets a real practical need, raised by Zhang et al. (2010) in the context of EQA schemes, for non-arbitrary statistical testing procedures to compare several multivariate coefficients of variation.

Acknowledgements The authors would like to express their thanks to Professor A. Albert (School of Public Health, University of Liege) for making the EQA data available. This work was partially supported by the IAP Research Network P7/06 of the Belgian State.

We also thank the referees for providing constructive comments and helpful suggestions.

References

- Aerts, S., Haesbroeck, G. and Ruwet, C. (2015) Multivariate coefficients of variation: comparison and influence functions. *J Multivariate Anal*, 142, 183–198
- Aerts, S., Haesbroeck, G. and Ruwet, C. (2016) Distribution under elliptical symmetry of a distance-based multivariate coefficient of variation. *Statistical papers*, in press DOI: 10.1007/s00362-016-0777-4
- Albert, A. and Zhang, L. (2010) A novel definition of the multivariate coefficient of variation. *Biometrical J*, 52, 667–675.
- Bartlett, M.S. (1937) Properties of sufficiency and statistical tests. *P Roy Soc A- Math*, 160, 268–282.
- Basu, A., Mandal, A., Martin, A. and Pardo, L. (2015) Generalized Wald-type tests based on minimum density power divergence estimators. *Statistics*, 50, 1–26.
- Bennett, B.M. (1976) On an approximate test for homogeneity of coefficients of variation. *Ziegler WJ (ed) Contribution to applied statistics*. Birkhauser Verlag, Basel and Stuttgart, 169–171

- Cator, E.A. and Lopuhää, H.P.(2010) Asymptotic expansion of the minimum covariance determinant estimators, *J Multivariate Anal*, 101, 2372–2388.
- Croux, C. and Haesbroeck, G. (1999) Influence function and efficiency of the minimum covariance determinant estimator. *J Multivariate Anal*, 71,161–190.
- Falk, M. (1997). Asymptotic independance of median and MAD *Statistics and Probability Letters*, 34, 341–345.
- Feltz, C.J. and Miller, G.E. (1996). An asymptotic test for the equality of coefficients of variation, *Stat Med*, 15, 647–658.
- Forkman, J. (2009) Estimator and tests for common coefficients of variation in normal distributions. *Commun Stat Theory*, 38, 233–251
- Garcia-Perez, A. (2012) A linear approximation to the power function of a test. *Metrika*, 75, 855–875.
- Ghosh, A. and Basu, A. (2013) Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electron J Stat*,7, 2420–2456.
- Ghosh, A., Mandal, A., Martin, N. and Pardo, L.(2016) Influence Analysis of Robust Wald-type tests. *J Multivariate Anal*,147, 102–126
- Gupta, C.R. and Ma, S.(1996) Testing the equality of coefficients of variation in K normal populations. *Commun Stat Theory*, 25, 115–132.
- Hallin, M. and Paindaveine, D. (2009) Optimal tests for homogeneity of covariance, scale, and shape. *J Multivariate Anal*, 100, 422–444.
- Hampel, F.R., Ronchetti, E. M., Rousseeuw, P.J., and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, New-York: John Wiley.
- M.J. Healy. (1979) Outliers in clinical chemistry quality control schemes. *Clin Chem*, 25, 675–677.
- Heritier, S. and Ronchetti, E.M. (1994) Robust Bounded-Influence Tests in General Parametric Models. *JASA*, 89, 897–904.
- Jobson, J.D. and Korkie, B.M. (1981) Performance Hypothesis Testing with the Sharpe and Treynor Measures, *J Financ*, 36, 889–908
- Krishnamoorthy, K., Lee, M. (2014). Improved tests for the equality of normal coefficients of variation. *Computation Stat*, 29, 215–232.
- Krishnamoorthy, K. and Yu, J. (2004) Modified Nel and van der Merwe Test for the Multivariate Behrens-Fisher Problem. *Statistics and Probability Letters*, 66, 161–169.
- Ledoit, O. and Wolf, M. (2008) Robust performance hypothesis testing with the Sharpe ratio. *J Empir Financ*, 15, 850–859.
- Loh,W-Y. (1984) Bounds on AREs for restricted classes of distributions defined via tail-orderings, *The Annals of Statistics*, 12, 685–701.
- Lopuhää, H.P. (1997) Asymptotic expansion of S estimators of location and covariance. *Statistica Neerlandica*,51, 220–237.
- Lopuhää,H.P. (1999) Asymptotics of reweighted estimators of multivariate location and scatter, *The Annals of Statistics*, 27, 1638–1665.
- McKay, A. (1932) Distribution of the coefficient of variation and the extended t -distribution. *J R Stat Soc*, 95, 695–698

- Miller, G.E. (1991a) Asymptotic test statistics for coefficients of variation. *Commun Stat Theory*, 20, 3351–3363
- Nairy, K.S. and Rao, K.A. (2003) Tests of coefficients of variation of normal population. *Commun Stat Simulat*, 32, 641–661.
- Pardo, M.C. and Pardo, J.A. (2000). Use of Rényi's divergence to test for the equality of the coefficients of variation. *J Comput Appl Math*, 116, 93–104.
- Pires, A.M. and Branco, J.A. (2002). Partial influence functions. *J Multivariate Anal*, 83, 451–468.
- Puri, M.L. and Sen, P.K. (1971) Nonparametric methods in multivariate analysis. *John Wiley and Sons*, Inc. New-York.
- Rousseeuw, P.J. and Ronchetti, E.M. (1981). Influence Curve for General Statistics. *J Comput Appl Math*, 7, 161–166.
- Schott, J.R. (2001) Some tests for the equality of covariance matrices. *J Stat Plan Infer*, 94, 25–36.
- Tsou, T.S. (2009) A robust score test for testing several coefficients of variation with unknown underlying distributions. *Commun Stat Theory*, 38, 1350–1360
- Verrill, S.P. and Johnson, R.A. (2007) Confidence bounds and hypothesis tests for normal distribution coefficients of variation. Research Paper FPL-RP-638. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory.
- Voinov V.G. and Nikulin M.S. (1996). *Unbiased estimators and their applications*. Vol. 2, multivariate case. Dordrecht: Kluwer.
- Zhang L., Albarède S., Dumont, G., Van Campenhout C., Libeer J. and Albert, A. (2010) The multivariate coefficient of variation for comparing serum protein electrophoresis techniques in external quality assessment schemes. *Accredit Qual Assur*, 15, 351–357.