British Journal of Mathematics & Computer Science
21(2): 1-15, 2017; Article no.BJMCS.31936
*ISSN: 2231-0851*

SCIENCEDOMAIN *international*
www.sciencedomain.org

# A Recommendation for Classical and Robust Factor Analysis

**Ying-Ying Zhang[1*], Teng-Zhong Rong[1] and Man-Man Li[1]**

[1] *Department of Statistics and Actuarial Science, College of Mathematics and Statistics, Chongqing University, Chongqing, China.*

***Authors' contributions***

*This work was carried out in collaboration between all authors. Author YYZ proved the theorems and did numerical simulations. Author TZR did literature searches and revised the manuscript. Author MML revised the manuscript. All authors read and approved the final manuscript.*

***Article Information***

DOI: 10.9734/BJMCS/2017/31936
*Editor(s):*
(1) Longhua Zhao, Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University, USA.
(2) Serkan Araci, Hasan Kalyoncu University, Turkey.
(3) Paul Bracken, Department of Mathematics, The University of Texas-Pan American Edinburg, TX 78539, USA.
*Reviewers:*
(1) Radosaw Jedynak, Kazimierz Pulaski University of Technology and Humanities, Poland.
(2) Milton Rosa, Universidade Federal de Ouro Preto, Brazil.
(3) Marwa Osman Mohamed, Zagazig University, Egypt.
(4) Diana Bilkova, University of Economics, Prague, Czech Republic.
Complete Peer review History: http://www.sciencedomain.org/review-history/18254

***Received: 31st January 2017***
***Accepted: 12th March 2017***
***Published: 18th March 2017***

***Original Research Article***

## Abstract

Considering the factor analysis methods (classical or robust), the data input (data or scaled data), and the running matrix (covariance or correlation) all together, there are 8 combinations. The objective of the study is to give a recommendation for classical and robust factor analysis. First, when the variables have different units, it is common to standardize the variables, and thus it is common to use the correlation matrix as the running matrix. Second, we need to explain the factors from the loading matrix. The entries of the loading matrix from the sample covariance matrix are not limited between 0 and 1, which makes the explanations of the factors hard. Third, we may not be able to compute the robust covariance matrix, and thus the robust correlation matrix of the original data, as the stocks data example illustrates. Consequently, we recommend classical and robust factor analysis using the correlation matrix of the scaled data as the running matrix for theoretical and computational reasons. The `hbk` data and the `stock611` data illustrate our recommendation.

**Corresponding author: E-mail: robertzhang@cqu.edu.cn, robertzhangyying@qq.com;*

*Keywords: Classical factor analysis; robust factor analysis; recommendation; R software.*

**2010 Mathematics Subject Classification:** 62F35, 62H25.

# 1 Introduction

Outliers exist in virtually every data set in any application domain. In order to avoid the masking effect, robust estimators are needed. The classical estimators of multivariate location and scatter are the sample mean $\bar{\boldsymbol{X}}$ and the sample covariance matrix $\boldsymbol{S}$. These estimates are optimal if the data come from a multivariate normal distribution but are extremely sensitive to the presence of even a few outliers. If outliers are present in the input data they will influence the estimates $\bar{\boldsymbol{X}}$ and $\boldsymbol{S}$ and subsequently worsen the performance of the classical factor analysis [1]. Therefore it is important to consider robust alternatives to these estimators. There are several robust estimators in the literature: MCD [2, 3, 4, 5], OGK [6], MVE [2], M [7, 8], S [9, 10, 11, 7, 12], and Stahel-Donoho [13, 14, 15, 16]. Substituting the classical location and scatter estimates by their robust analogues is the most straightforward method for robustifying many multivariate procedures [17, 18], which is our choice for robustifying the factor analysis procedure.

The rest of the paper is organized as follows. Section 2 compares the classical and the robust factor analysis methods and presents some theoretical results. Section 3 illustrates the comparisons of the two methods with a `hbk` data example and a stocks data example. Section 4 concludes.

# 2 Comparison of Classical and Robust Factor Analysis

When compare the classical and robust factor analysis methods, there are two other issues to consider. That is, whether we should use `data` or `scale(data)` as the data input; whether we should use the covariance matrix or the correlation matrix as the running matrix (`usedMatrix`)? Consider the factor analysis methods, the data input, and the running matrix all together, there are 8 combinations, i.e.,

```
<1> classical, data, covariance matrix
<2> classical, data, correlation matrix
<3> classical, scale(data), covariance matrix
<4> classical, scale(data), correlation matrix
<5> robust, data, covariance matrix
<6> robust, data, correlation matrix
<7> robust, scale(data), covariance matrix
<8> robust, scale(data), correlation matrix
```

There are 4 classical and robust factor analysis comparisons, i.e., <1> vs <5>, <2> vs <6>, <3> vs <7>, and <4> vs <8>. We recommend <4> vs <8>. The reasons are as follows. First, when the variables have different units, it is common to standardize the variables, the sample covariance matrix of the standardized variables is the correlation matrix of the original variables. Thus it is common to use the correlation matrix as the running matrix. Second, we need to explain the factors from the loading matrix. The entries of the loading matrix from the sample covariance matrix are not limited between 0 and 1, which makes the explanations of the factors hard. The first two reasons suggest us to choose <2> vs <6> and <4> vs <8>. However, <2> and <4> (<6> and <8>) have the same running matrices, eigenvalues, loadings, uniquenesses, scoring coefficients, scaled data matrices, and score matrices, see Theorem 2.3. That is, <2> vs <6> and <4> vs <8> give us the same comparisons. We can choose any pair to do the comparison. Third, we may not be able to compute the robust covariance matrix, and thus the robust correlation matrix of the original data, as the stocks data example will illustrate. Consequently, we should choose <4> vs <8>.

The running matrices of the 8 combinations are given in Table 1.

**Table 1. The running matrices. Source: Own research**

| | | Classical | Robust |
|---|---|---|---|
| data | covariance | <1> $\boldsymbol{S}^c$ | <5> $\boldsymbol{S}^r$ |
| | correlation | <2> $\boldsymbol{R}^c$ | <6> $\boldsymbol{R}^r$ |
| scale(data) | covariance | <3> $\tilde{\boldsymbol{S}}^c$ | <7> $\tilde{\boldsymbol{S}}^r$ |
| | correlation | <4> $\tilde{\boldsymbol{R}}^c$ | <8> $\tilde{\boldsymbol{R}}^r$ |

To compare the 8 combinations, we summarize some useful results in Tables 2 and 3. There are some points in Tables 2 and 3 that are proved in the supplement. Some basic definitions such as diag(), sample mean, sample covariance matrix, scaled matrix, etc. are also put into the supplement.

**Table 2. Results of $< 1 >$, $< 2 >$, $< 5 >$, and $< 6 >$. In the table, A: data matrix; B: number of used observations; C: sample used; D: $k$th observation; E: sample mean; F: sample covariance; G: sample correlation; H: $\boldsymbol{D}$; I: scaledX; J: $k$th scaled observation; K: sample mean of scaledX; L: cov(scaledX). Source: Own research**

| | $\boldsymbol{X}$, classical <1>, <2> | $\boldsymbol{X}$, robust <5>, <6> |
|---|---|---|
| A | $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{(1)}^\top \\ \boldsymbol{X}_{(2)}^\top \\ \vdots \\ \boldsymbol{X}_{(n)}^\top \end{pmatrix}$ | $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{(1)}^\top \\ \boldsymbol{X}_{(2)}^\top \\ \vdots \\ \boldsymbol{X}_{(n)}^\top \end{pmatrix}$ |
| B | $n$ | $M$ |
| C | $\boldsymbol{X}_{(1)}, \boldsymbol{X}_{(2)}, \cdots, \boldsymbol{X}_{(n)}$ | $\boldsymbol{X}_{(1)}, \boldsymbol{X}_{(2)}, \cdots, \boldsymbol{X}_{(M)}$ |
| D | $\boldsymbol{X}_{(k)}$ | $\boldsymbol{X}_{(k)}$ |
| E | $\bar{\boldsymbol{X}}^c = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{X}_{(k)}$ | $\bar{\boldsymbol{X}}^r = \frac{1}{M}\sum_{k=1}^{M} \boldsymbol{X}_{(k)}$ |
| F | $\boldsymbol{S}^c = \frac{1}{n-1}\sum_{k=1}^{n} \left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^c\right)\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^c\right)^\top$ | $\boldsymbol{S}^r = \frac{1}{M-1}\sum_{k=1}^{M} \left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^r\right)\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^r\right)^\top$ |
| G | $\boldsymbol{R}^c$ | $\boldsymbol{R}^r$ |
| H | $\boldsymbol{D}^c = \mathrm{diag}\left(\mathrm{diag}\left(\boldsymbol{S}^c\right)\right)$ | $\boldsymbol{D}^r = \mathrm{diag}\left(\mathrm{diag}\left(\boldsymbol{S}^r\right)\right)$ |
| I | $\boldsymbol{X}^{*c} = \left(\boldsymbol{X} - \mathbf{1}\left(\bar{\boldsymbol{X}}^c\right)^\top\right)\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}} = \begin{pmatrix} \left(\boldsymbol{X}_{(1)}^{*c}\right)^\top \\ \left(\boldsymbol{X}_{(2)}^{*c}\right)^\top \\ \vdots \\ \left(\boldsymbol{X}_{(n)}^{*c}\right)^\top \end{pmatrix}$ | $\boldsymbol{X}^{*r} = \left(\boldsymbol{X} - \mathbf{1}\left(\bar{\boldsymbol{X}}^r\right)^\top\right)\left(\boldsymbol{D}^r\right)^{-\frac{1}{2}} = \begin{pmatrix} \left(\boldsymbol{X}_{(1)}^{*r}\right)^\top \\ \left(\boldsymbol{X}_{(2)}^{*r}\right)^\top \\ \vdots \\ \left(\boldsymbol{X}_{(n)}^{*r}\right)^\top \end{pmatrix}$ |
| J | $\boldsymbol{X}_{(k)}^{*c} = \left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^c\right)$ | $\boldsymbol{X}_{(k)}^{*r} = \left(\boldsymbol{D}^r\right)^{-\frac{1}{2}}\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^r\right)$ |
| K | $\bar{\boldsymbol{X}}^{*c} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{X}_{(k)}^{*c} = \mathbf{0}$ | $\bar{\boldsymbol{X}}^{*r} = \frac{1}{M}\sum_{k=1}^{M} \boldsymbol{X}_{(k)}^{*r} = \mathbf{0}$ (a) |
| L | $\boldsymbol{S}^{*c} = \boldsymbol{R}^c = \frac{1}{n-1}\sum_{k=1}^{n}\left(\boldsymbol{X}_{(k)}^{*c} - \bar{\boldsymbol{X}}^{*c}\right)\left(\boldsymbol{X}_{(k)}^{*c} - \bar{\boldsymbol{X}}^{*c}\right)^\top = \frac{1}{n-1}\sum_{k=1}^{n} \boldsymbol{X}_{(k)}^{*c}\left(\boldsymbol{X}_{(k)}^{*c}\right)^\top = \left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\boldsymbol{S}^c\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}$ | $\boldsymbol{S}^{*r} = \boldsymbol{R}^r = \frac{1}{M-1}\sum_{k=1}^{M}\left(\boldsymbol{X}_{(k)}^{*r} - \bar{\boldsymbol{X}}^{*r}\right)\left(\boldsymbol{X}_{(k)}^{*r} - \bar{\boldsymbol{X}}^{*r}\right)^\top = \frac{1}{M-1}\sum_{k=1}^{M} \boldsymbol{X}_{(k)}^{*r}\left(\boldsymbol{X}_{(k)}^{*r}\right)^\top = \left(\boldsymbol{D}^r\right)^{-\frac{1}{2}}\boldsymbol{S}^r\left(\boldsymbol{D}^r\right)^{-\frac{1}{2}}$ (b) |

**Table 3. Results of $< 3 >$, $< 4 >$, $< 7 >$, and $< 8 >$. In the table, A: data matrix; B: number of used observations; C: sample used; D: $k$th observation; E: sample mean; F: sample covariance; G: sample correlation; H: $\boldsymbol{D}$; I: scaledX; J: $k$th scaled observation; K: sample mean of scaledX; L: cov(scaledX). Source: Own research.**

| | $\boldsymbol{Y}$, classical <3>, <4> | $\boldsymbol{Y}$, robust <7>, <8> |
|---|---|---|
| A | $\boldsymbol{Y} = \text{scale}(\boldsymbol{X}) = \boldsymbol{X}^{*c} = \left(\boldsymbol{X} - \mathbf{1}(\bar{\boldsymbol{X}}^c)^\top\right)(\boldsymbol{D}^c)^{-\frac{1}{2}}$ | $\boldsymbol{Y} = \text{scale}(\boldsymbol{X}) = \boldsymbol{X}^{*c} = \left(\boldsymbol{X} - \mathbf{1}(\bar{\boldsymbol{X}}^c)^\top\right)(\boldsymbol{D}^c)^{-\frac{1}{2}}$ |
| B | $n$ | $M$ |
| C | $\boldsymbol{Y}_{(1)}, \boldsymbol{Y}_{(2)}, \cdots, \boldsymbol{Y}_{(n)}$ | $\boldsymbol{Y}_{(1)}, \boldsymbol{Y}_{(2)}, \cdots, \boldsymbol{Y}_{(M)}$ |
| D | $\boldsymbol{Y}_{(k)} = (\boldsymbol{D}^c)^{-\frac{1}{2}}(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^c) = \boldsymbol{X}^{*c}_{(k)}$ | $\boldsymbol{Y}_{(k)} = (\boldsymbol{D}^c)^{-\frac{1}{2}}(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^c) = \boldsymbol{X}^{*c}_{(k)}$ |
| E | $\bar{\boldsymbol{Y}}^c = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{Y}_{(k)} = \mathbf{0}$ (c) | $\bar{\boldsymbol{Y}}^r = \frac{1}{M}\sum_{k=1}^{M}\boldsymbol{Y}_{(k)} = (\boldsymbol{D}^c)^{-\frac{1}{2}}(\bar{\boldsymbol{X}}^r - \bar{\boldsymbol{X}}^c)$ (f) |
| F | $\tilde{\boldsymbol{S}}^c = \frac{1}{n-1}\sum_{k=1}^{n}(\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^c)(\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^c)^\top = \frac{1}{n-1}\sum_{k=1}^{n}\boldsymbol{Y}_{(k)}\boldsymbol{Y}_{(k)}^\top$ | $\tilde{\boldsymbol{S}}^r = \frac{1}{M-1}\sum_{k=1}^{M}(\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^r)(\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^r)^\top$ |
| G | $\tilde{\boldsymbol{R}}^c$ | $\tilde{\boldsymbol{R}}^r$ |
| H | $\tilde{\boldsymbol{D}}^c = \text{diag}\left(\text{diag}\left(\tilde{\boldsymbol{S}}^c\right)\right) = \boldsymbol{E}$ (d) | $\tilde{\boldsymbol{D}}^r = \text{diag}\left(\text{diag}\left(\tilde{\boldsymbol{S}}^r\right)\right)$ |
| I | $\boldsymbol{Y}^{*c} = \left(\boldsymbol{Y} - \mathbf{1}(\bar{\boldsymbol{Y}}^c)^\top\right)\left(\tilde{\boldsymbol{D}}^c\right)^{-\frac{1}{2}} = \begin{pmatrix} \left(\boldsymbol{Y}^{*c}_{(1)}\right)^\top \\ \left(\boldsymbol{Y}^{*c}_{(2)}\right)^\top \\ \vdots \\ \left(\boldsymbol{Y}^{*c}_{(n)}\right)^\top \end{pmatrix}$ | $\boldsymbol{Y}^{*r} = \left(\boldsymbol{Y} - \mathbf{1}(\bar{\boldsymbol{Y}}^r)^\top\right)\left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}} = \begin{pmatrix} \left(\boldsymbol{Y}^{*r}_{(1)}\right)^\top \\ \left(\boldsymbol{Y}^{*r}_{(2)}\right)^\top \\ \vdots \\ \left(\boldsymbol{Y}^{*r}_{(n)}\right)^\top \end{pmatrix}$ |
| J | $\boldsymbol{Y}^{*c}_{(k)} = \left(\tilde{\boldsymbol{D}}^c\right)^{-\frac{1}{2}}(\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^c) = \boldsymbol{Y}_{(k)}$ (e) | $\boldsymbol{Y}^{*r}_{(k)} = \left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}}(\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^r)$ |
| K | $\bar{\boldsymbol{Y}}^{*c} = \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{Y}^{*c}_{(k)} = \mathbf{0}$ | $\bar{\boldsymbol{Y}}^{*r} = \frac{1}{M}\sum_{k=1}^{M}\boldsymbol{Y}^{*r}_{(k)} = \mathbf{0}$ (g) |
| L | $\tilde{\boldsymbol{S}}^{*c} = \tilde{\boldsymbol{R}}^c = \frac{1}{n-1}\sum_{k=1}^{n}\left(\boldsymbol{Y}^{*c}_{(k)} - \bar{\boldsymbol{Y}}^{*c}\right)\left(\boldsymbol{Y}^{*c}_{(k)} - \bar{\boldsymbol{Y}}^{*c}\right)^\top = \frac{1}{n-1}\sum_{k=1}^{n}\boldsymbol{Y}^{*c}_{(k)}\left(\boldsymbol{Y}^{*c}_{(k)}\right)^\top = \left(\tilde{\boldsymbol{D}}^c\right)^{-\frac{1}{2}}\tilde{\boldsymbol{S}}^c\left(\tilde{\boldsymbol{D}}^c\right)^{-\frac{1}{2}}$ | $\tilde{\boldsymbol{S}}^{*r} = \tilde{\boldsymbol{R}}^r = \frac{1}{M-1}\sum_{k=1}^{M}\left(\boldsymbol{Y}^{*r}_{(k)} - \bar{\boldsymbol{Y}}^{*r}\right)\left(\boldsymbol{Y}^{*r}_{(k)} - \bar{\boldsymbol{Y}}^{*r}\right)^\top = \frac{1}{M-1}\sum_{k=1}^{M}\boldsymbol{Y}^{*r}_{(k)}\left(\boldsymbol{Y}^{*r}_{(k)}\right)^\top = \left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}}\tilde{\boldsymbol{S}}^r\left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}}$ (h) |

To prove Theorem 2.2, we need the following lemma.

**Lemma 2.1.** *Let*

$$\boldsymbol{\Lambda}_1 = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{pmatrix}, \ \boldsymbol{\Lambda}_2 = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_p \end{pmatrix}, \ \boldsymbol{A} = (a_{ij})_{p\times p}.$$

*Then*

$$\text{diag}(\boldsymbol{\Lambda}_1\boldsymbol{A}\boldsymbol{\Lambda}_2) = \text{diag}(\boldsymbol{\Lambda}_1)\cdot\text{diag}(\boldsymbol{A})\cdot\text{diag}(\boldsymbol{\Lambda}_2),$$
$$\text{diag}(\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2) = \text{diag}(\boldsymbol{\Lambda}_1)\cdot\text{diag}(\boldsymbol{\Lambda}_2).$$

The proof of the lemma is easy and is put into the supplement.

For the running matrices in Table 1, we have

$$\boldsymbol{R}^c = \tilde{\boldsymbol{S}}^c = \tilde{\boldsymbol{R}}^c. \tag{2.1}$$

The proof can be found in Section 1.5.1 of [19]. Furthermore, we have the following theorem.

**Theorem 2.2.** $\boldsymbol{R}^r = \tilde{\boldsymbol{R}}^r$.

**Proof.** We have

$$\boldsymbol{R}^r = \boldsymbol{S}^{*r} = \frac{1}{M-1}\sum_{k=1}^{M}\boldsymbol{X}^{*r}_{(k)}\left(\boldsymbol{X}^{*r}_{(k)}\right)^\top,$$

$$\tilde{\boldsymbol{R}}^r = \tilde{\boldsymbol{S}}^{*r} = \frac{1}{M-1}\sum_{k=1}^{M}\boldsymbol{Y}^{*r}_{(k)}\left(\boldsymbol{Y}^{*r}_{(k)}\right)^\top.$$

To prove that $\boldsymbol{R}^r = \tilde{\boldsymbol{R}}^r$, it suffices to prove that

$$\boldsymbol{X}^{*r}_{(k)} = \boldsymbol{Y}^{*r}_{(k)}. \tag{2.2}$$

Write

$$\boldsymbol{X}^{*r}_{(k)} = (\boldsymbol{D}^r)^{-\frac{1}{2}}\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^r\right),$$

$$\boldsymbol{Y}^{*r}_{(k)} = \left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}}\left(\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^r\right),$$

$$\boldsymbol{Y}_{(k)} = (\boldsymbol{D}^c)^{-\frac{1}{2}}\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^c\right),$$

from Table 3 (f), we have

$$\bar{\boldsymbol{Y}}^r = (\boldsymbol{D}^c)^{-\frac{1}{2}}\left(\bar{\boldsymbol{X}}^r - \bar{\boldsymbol{X}}^c\right).$$

Thus

$$\begin{aligned}\boldsymbol{Y}^{*r}_{(k)} &= \left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}}\left(\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^r\right)\\ &= \left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}}\left[(\boldsymbol{D}^c)^{-\frac{1}{2}}\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^c\right) - (\boldsymbol{D}^c)^{-\frac{1}{2}}\left(\bar{\boldsymbol{X}}^r - \bar{\boldsymbol{X}}^c\right)\right]\\ &= \left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}}(\boldsymbol{D}^c)^{-\frac{1}{2}}\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^r\right).\end{aligned}$$

To prove (2.2), it suffices to prove

$$\left(\tilde{\boldsymbol{D}}^r\right)^{-\frac{1}{2}}(\boldsymbol{D}^c)^{-\frac{1}{2}} = (\boldsymbol{D}^r)^{-\frac{1}{2}},$$

which reduces to prove

$$\tilde{\boldsymbol{D}}^r\boldsymbol{D}^c = \boldsymbol{D}^r,$$

that is,

$$\operatorname{diag}\left(\operatorname{diag}\left(\tilde{\boldsymbol{S}}^r\right)\right)\cdot\operatorname{diag}\left(\operatorname{diag}\left(\boldsymbol{S}^c\right)\right) = \operatorname{diag}\left(\operatorname{diag}\left(\boldsymbol{S}^r\right)\right),$$

which is equivalent to

$$\operatorname{diag}\left(\tilde{\boldsymbol{S}}^r\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^c\right) = \operatorname{diag}\left(\boldsymbol{S}^r\right).$$

We have just proved

$$\boldsymbol{Y}_{(k)} - \bar{\boldsymbol{Y}}^r = (\boldsymbol{D}^c)^{-\frac{1}{2}}\left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}^r\right),$$

and thus

$$\begin{aligned}\tilde{\boldsymbol{S}}^r &= \frac{1}{M-1}\sum_{k=1}^{M}\left(\boldsymbol{Y}_{(k)}-\bar{\boldsymbol{Y}}^r\right)\left(\boldsymbol{Y}_{(k)}-\bar{\boldsymbol{Y}}^r\right)^\top \\ &= \frac{1}{M-1}\sum_{k=1}^{M}\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\left(\boldsymbol{X}_{(k)}-\bar{\boldsymbol{X}}^r\right)\left(\boldsymbol{X}_{(k)}-\bar{\boldsymbol{X}}^r\right)^\top\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}} \\ &= \left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\left[\frac{1}{M-1}\sum_{k=1}^{M}\left(\boldsymbol{X}_{(k)}-\bar{\boldsymbol{X}}^r\right)\left(\boldsymbol{X}_{(k)}-\bar{\boldsymbol{X}}^r\right)^\top\right]\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}} \\ &= \left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\boldsymbol{S}^r\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}.\end{aligned}$$

Consequently,

$$\begin{aligned}\operatorname{diag}\left(\tilde{\boldsymbol{S}}^r\right) &= \operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\boldsymbol{S}^r\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\right) \\ &= \operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^r\right)\cdot\operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\right)\ \text{(Lemma 2.1)} \\ &= \operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\right)\cdot\operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-\frac{1}{2}}\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^r\right) \\ &= \operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-1}\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^r\right)\ \text{(Lemma 2.1)}.\end{aligned}$$

Finally,

$$\begin{aligned}\operatorname{diag}\left(\tilde{\boldsymbol{S}}^r\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^c\right) &= \operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-1}\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^r\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^c\right) \\ &= \operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-1}\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^r\right)\cdot\operatorname{diag}\left(\boldsymbol{D}^c\right) \\ &= \operatorname{diag}\left(\left(\boldsymbol{D}^c\right)^{-1}\right)\cdot\operatorname{diag}\left(\boldsymbol{D}^c\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^r\right) \\ &= \operatorname{diag}\left(\boldsymbol{E}\right)\cdot\operatorname{diag}\left(\boldsymbol{S}^r\right)\ \text{(Lemma 2.1)} \\ &= \operatorname{diag}\left(\boldsymbol{S}^r\right).\end{aligned}$$

The proof is complete. □

The score of the $k$th observation is summarized in Table 4. The scores method can be "regression" or "Bartlett". The running matrix can be the covariance matrix ($\boldsymbol{S}$) or the correlation matrix ($\boldsymbol{R}$). See [20, 21] for details.

**Table 4. The score of the $k$th observation. Source: Own research**

| | regression | Bartlett |
|---|---|---|
| cor = FALSE | $\tilde{\boldsymbol{f}}_{(k)} = \hat{\boldsymbol{L}}^\top\boldsymbol{S}^{-1}\left(\boldsymbol{X}_{(k)}-\bar{\boldsymbol{X}}\right)$ | $\hat{\boldsymbol{f}}_{(k)} = \left(\hat{\boldsymbol{L}}^\top\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{L}}\right)^{-1}\hat{\boldsymbol{L}}^\top\hat{\boldsymbol{\Psi}}^{-1}\left(\boldsymbol{X}_{(k)}-\bar{\boldsymbol{X}}\right)$ |
| cor = TRUE | $\tilde{\boldsymbol{f}}^*_{(k)} = \hat{\boldsymbol{L}}^{*\top}\boldsymbol{R}^{-1}\boldsymbol{X}^*_{(k)}$ | $\hat{\boldsymbol{f}}^*_{(k)} = \left(\hat{\boldsymbol{L}}^{*\top}\hat{\boldsymbol{\Psi}}^{*-1}\hat{\boldsymbol{L}}^*\right)^{-1}\hat{\boldsymbol{L}}^{*\top}\hat{\boldsymbol{\Psi}}^{*-1}\boldsymbol{X}^*_{(k)}$ |

If we define the scoring coefficient $\boldsymbol{S}_c$ by

$$\boldsymbol{S}_c = \begin{cases}\hat{\boldsymbol{L}}^\top\boldsymbol{S}^{-1}, & \text{cor = FALSE, scoresMethod = "regression"}, \\ \left(\hat{\boldsymbol{L}}^\top\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{L}}\right)^{-1}\hat{\boldsymbol{L}}^\top\hat{\boldsymbol{\Psi}}^{-1}, & \text{cor = FALSE, scoresMethod = "Bartlett"}, \\ \hat{\boldsymbol{L}}^{*\top}\boldsymbol{R}^{-1}, & \text{cor = TRUE, scoresMethod = "regression"}, \\ \left(\hat{\boldsymbol{L}}^{*\top}\hat{\boldsymbol{\Psi}}^{*-1}\hat{\boldsymbol{L}}^*\right)^{-1}\hat{\boldsymbol{L}}^{*\top}\hat{\boldsymbol{\Psi}}^{*-1}, & \text{cor = TRUE, scoresMethod = "Bartlett"},\end{cases}$$

then the score of the $k$th observation simplifies to

$$\boldsymbol{f}_{(k)} = \begin{cases} \boldsymbol{S}_c \left(\boldsymbol{X}_{(k)} - \bar{\boldsymbol{X}}\right), & \text{cor} = \text{FALSE}, \\ \boldsymbol{S}_c \boldsymbol{X}^*_{(k)}, & \text{cor} = \text{TRUE}. \end{cases}$$

If cor = FALSE, then the score matrix is

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{f}_{(1)}^\top \\ \boldsymbol{f}_{(2)}^\top \\ \vdots \\ \boldsymbol{f}_{(n)}^\top \end{pmatrix} = \begin{pmatrix} \left(\boldsymbol{X}_{(1)} - \bar{\boldsymbol{X}}\right)^\top \boldsymbol{S}_c^\top \\ \left(\boldsymbol{X}_{(2)} - \bar{\boldsymbol{X}}\right)^\top \boldsymbol{S}_c^\top \\ \vdots \\ \left(\boldsymbol{X}_{(n)} - \bar{\boldsymbol{X}}\right)^\top \boldsymbol{S}_c^\top \end{pmatrix} = \begin{pmatrix} \left(\boldsymbol{X}_{(1)} - \bar{\boldsymbol{X}}\right)^\top \\ \left(\boldsymbol{X}_{(2)} - \bar{\boldsymbol{X}}\right)^\top \\ \vdots \\ \left(\boldsymbol{X}_{(n)} - \bar{\boldsymbol{X}}\right)^\top \end{pmatrix} \boldsymbol{S}_c^\top = \left(\boldsymbol{X} - \boldsymbol{1}\bar{\boldsymbol{X}}^\top\right) \boldsymbol{S}_c^\top.$$

If cor = TRUE, then the score matrix is

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{f}_{(1)}^\top \\ \boldsymbol{f}_{(2)}^\top \\ \vdots \\ \boldsymbol{f}_{(n)}^\top \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{(1)}^{*\top} \boldsymbol{S}_c^\top \\ \boldsymbol{X}_{(2)}^{*\top} \boldsymbol{S}_c^\top \\ \vdots \\ \boldsymbol{X}_{(n)}^{*\top} \boldsymbol{S}_c^\top \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{(1)}^{*\top} \\ \boldsymbol{X}_{(2)}^{*\top} \\ \vdots \\ \boldsymbol{X}_{(n)}^{*\top} \end{pmatrix} \boldsymbol{S}_c^\top = \boldsymbol{X}^* \boldsymbol{S}_c^\top,$$

where $\boldsymbol{X}^* = \left(\boldsymbol{X} - \boldsymbol{1}\bar{\boldsymbol{X}}^\top\right) \boldsymbol{D}^{-\frac{1}{2}}$. If we define

$$\text{scaledX} = \begin{cases} \boldsymbol{X} - \boldsymbol{1}\bar{\boldsymbol{X}}^\top, & \text{cor} = \text{FALSE}, \\ \boldsymbol{X}^*, & \text{cor} = \text{TRUE}, \end{cases}$$

then

$$\boldsymbol{F} = \text{scaledX} \cdot \boldsymbol{S}_c^\top. \tag{2.3}$$

With the above notations, we have the following theorem which indicates that <2> vs <6> and <4> vs <8> give us the same comparisons.

**Theorem 2.3.** *The running matrices ($\boldsymbol{R}$), eigenvalues ($\lambda$), loadings ($\boldsymbol{L}$), uniquenesses ($\boldsymbol{\Psi}$), scoring coefficients ($\boldsymbol{S}_c$), scaled data matrices (scaledX), score matrices ($\boldsymbol{F}$) are the same for*
*(a) combinations <2> and <4>,*
*(b) combinations <6> and <8>.*

**Proof.** If the running matrices ($\boldsymbol{R}$) are the same, then the eigenvalues ($\lambda$), loadings ($\boldsymbol{L}$), uniquenesses ($\boldsymbol{\Psi}$) are the same. If $\boldsymbol{R}, \boldsymbol{L}, \boldsymbol{\Psi}$ are the same, then by the definition of scoring coefficient $\boldsymbol{S}_c$, we know that $\boldsymbol{S}_c$ are the same. If the scaled data matrices (scaledX) are the same, then by (2.3), the score matrices ($\boldsymbol{F}$) are the same. Thus it suffices to show that $\boldsymbol{R}$ and scaledX are the same.

(a) The running matrices $\boldsymbol{R}^c = \tilde{\boldsymbol{R}}^c$ by (2.1). From Table 3, we see that

$$\boldsymbol{Y}_{(k)}^{*c} = \boldsymbol{Y}_{(k)} = \boldsymbol{X}_{(k)}^{*c},$$

thus the scaledX are the same.

(b) The running matrices $\boldsymbol{R}^r = \tilde{\boldsymbol{R}}^r$ by Theorem 2.2. By (2.2), we have $\boldsymbol{X}_{(k)}^{*r} = \boldsymbol{Y}_{(k)}^{*r}$, thus the scaledX are the same. □

# 3 Numerical Examples

In this section, we will illustrate (2.1) and Theorems 2.2 and 2.3 by using the R package **robustfa**. For the base functionalities of **robustfa**, we refer readers to the vignettes of [22].

## 3.1 Example: hbk data

In this subsection, a data set `hbk` is used. The Hawkins, Bradu, and Kass data set `hbk` is from the package **robustbase** consists of 75 observations in 4 dimensions (one response and three explanatory variables). The first 10 observations are bad leverage points, and the next four points are good leverage points (i.e., their $\boldsymbol{x}$ are outlying, but the corresponding $y$ fit the model quite well). We will consider only the $\boldsymbol{x}$-part of the data.

For `x = hbk.x`, we checked that $\boldsymbol{S}^r \neq \tilde{\boldsymbol{S}}^r$, and $\boldsymbol{R}^r = \tilde{\boldsymbol{R}}^r$ for `control = "mcd"`, `"ogk"`, `"m"`, `"mve"`, `"sfast"`, `"bisquare"`, `"rocke"`, small differences between $\boldsymbol{R}^r$ and $\tilde{\boldsymbol{R}}^r$ for `control = "sde"`, `"surreal"`. The results illustrate Theorem 2.2. The R codes of the checking are:

```
compute_cov_cor(x = hbk.x, control = "mcd")
compute_cov_cor(x = hbk.x, control = "ogk")
compute_cov_cor(x = hbk.x, control = "m")
compute_cov_cor(x = hbk.x, control = "mve")
compute_cov_cor(x = hbk.x, control = "sde")
compute_cov_cor(x = hbk.x, control = "sfast")
compute_cov_cor(x = hbk.x, control = "surreal")
compute_cov_cor(x = hbk.x, control = "bisquare")
compute_cov_cor(x = hbk.x, control = "rocke")
```

The eigenvalues of the running matrices of the `hbk` data of the 8 combinations are given in Table 5. From Table 5 we see that the eigenvalues of `<2>`, `<3>`, and `<4>` are the same, the eigenvalues of `<6>` and `<8>` are the same. The results illustrate (2.1) and Theorems 2.2 and 2.3. The R codes to compute the eigenvalues for Table 5 are:

```
covC = CovClassic(x = hbk.x); covC
eigen(covC@cov)$values # <1>
eigen(cov2cor(covC@cov))$values # <2>
covMcd = CovRobust(x = hbk.x, control = "mcd"); covMcd
eigen(covMcd@cov)$values # <5>
eigen(cov2cor(covMcd@cov))$values # <6>
covC = CovClassic(x = scale(hbk.x)); covC
eigen(covC@cov)$values # <3>
eigen(cov2cor(covC@cov))$values # <4>
covMcd = CovRobust(x = scale(hbk.x), control = "mcd"); covMcd
eigen(covMcd@cov)$values # <7>
eigen(cov2cor(covMcd@cov))$values # <8>
```

**Table 5. The eigenvalues of the running matrices for hbk data. Source: Own research**

| | | Classical | Robust (MCD) |
|---|---|---|---|
| hbk.x | cov | <1> 216.16 1.98 0.92 | <5> 1.94 1.59 1.37 |
| | cor | <2> 2.92 0.06 0.02 | <6> 1.19 0.96 0.85 |
| scale(hbk.x) | cov | <3> 2.92 0.06 0.02 | <7> 0.12 0.03 0.01 |
| | cor | <4> 2.92 0.06 0.02 | <8> 1.19 0.96 0.85 |

Classical and robust (MCD) scatterplots of the first two factor scores of the `hbk` data with 97.5% tolerance ellipses are plotted in Fig 1. We see that the scores of `<2>`, `<3>`, and `<4>` are the same, the scores of <6> and <8> are the same, in agree with Theorem 2.3. Note that the tolerance ellipse is very large for <1>, since the outliers severely affected the eigenvalues of the running matrix $\boldsymbol{S}^c$. While the tolerance ellipses are very small for <2>, <3>, and <4>, also due to the outliers severely affected the eigenvalues of the running matrices $\boldsymbol{R}^c = \tilde{\boldsymbol{S}}^c = \tilde{\boldsymbol{R}}^c$. The tolerance ellipse is very small for <7> and it does not cover the regular points, due to the first two eigenvalues of <7> are very small. It exemplifies that the results from robust covariance matrix of the scaled data is not very reliable. The tolerance ellipses of <6> and <8> well separate the regular points and the outliers. The R codes for the comparison <1> vs <5> are given as follows. The R codes for other comparisons are similar and thus are omitted.

```
## <1> classical, x = hbk.x, cor = FALSE (covariance matrix)
faClassic1 = FaClassic(x = hbk.x, factors = 2, method = "pca",
scoresMethod = "regression"); faClassic1
## <5> robust, x = hbk.x, cor = FALSE (covariance matrix)
faCov5 = FaCov(x = hbk.x, factors = 2, method = "pca",
scoresMethod = "regression", cov.control = CovControlMcd()); faCov5
## <1> vs <5>
usr <- par(mfrow = c(1,2))
cfaClassic <- list(center = c(0,0), cov = diag(faClassic1@eigenvalues[1:2]),
n.obs = faClassic1@n.obs)
rrcov:::.myellipse(faClassic1@scores, xcov = cfaClassic, main = "Classical",
xlab = "Factor1", ylab = "Factor2", xlim = c(-40,40), ylim = c(-5,28), id.n = 0)
abline(v = 0)
abline(h = 0)
text(5,0,labels = "1-13", cex = 0.8)
text(0.5,6,labels = "14", cex = 0.8)
cfaCov <- list(center = c(0,0), cov = diag(faCov5@eigenvalues[1:2]),
n.obs = faCov5@n.obs)
rrcov:::.myellipse(faCov5@scores, xcov = cfaCov, main = "Robust (MCD)",
xlab = "Factor1", ylab = "Factor2", xlim = c(-40,40), ylim = c(-5,28), id.n = 4)
text(22,9.5,labels = "1-10", cex = 0.8)
abline(v = 0)
abline(h = 0)
par(usr)
```

### 3.2 Example: Stocks data

In this subsection, we apply the robust factor analysis solution to a real data set `stock611`. This data set consists of 611 observations with 12 variables. The data set is from Chinese stock market in the year 2001. It is used in [23] to illustrate factor analysis methods.

For `x = stock611[,3:12]`,

```
cov_x = CovRobust(x = x, control = control)
```

gets error message for `control = "mcd", "m", "mve", "sde", "sfast"`, thus we can not compute `cov_x`, $\boldsymbol{S}^r$, and $\boldsymbol{R}^r$ for these robust estimators. That is, we can not get results for combinations <5> and <6> for these robust estimators. However, for `x = scale(stock611[,3:12])`, we can compute `cov_x`, $\tilde{\boldsymbol{S}}^r$, and $\tilde{\boldsymbol{R}}^r$ for these robust estimators, and we can get results for combinations <7> and <8>. Although <6> and <8> have the same running matrices ($\boldsymbol{R}$), eigenvalues ($\lambda$), loadings ($\boldsymbol{L}$), uniquenesses ($\boldsymbol{\Psi}$), scoring coefficients ($\boldsymbol{S}_c$), scaled data matrices (scaledX), and score matrices ($\boldsymbol{F}$), as were proved in Theorem 2.3, we may not be able to get results for <6> due to computational error for `cov_x`, while for <8> the computational error does not occur. That is why we recommend

`<4>` vs `<8>` for classical and robust factor analysis.

The first two eigenvalues of the running matrices of the `stock611` data of the 8 combinations are given in Table 6. From Table 6 we see that the eigenvalues of `<2>`, `<3>`, and `<4>` are the same, the eigenvalues of `<6>` and `<8>` are the same. The results also illustrate (2.1) and Theorems 2.2 and 2.3. The R codes to compute the eigenvalues for Table 6 are:

```
covC = CovClassic(x = stock611[,3:12]); covC
eigen(covC@cov)$values # <1>
eigen(cov2cor(covC@cov))$values # <2>
covOgk = CovRobust(x = stock611[,3:12], control = "ogk"); covOgk
eigen(covOgk@cov)$values # <5>
eigen(cov2cor(covOgk@cov))$values # <6>
covC = CovClassic(x = scale(stock611[,3:12])); covC
eigen(covC@cov)$values # <3>
eigen(cov2cor(covC@cov))$values # <4>
covOgk = CovRobust(x = scale(stock611[,3:12]), control = "ogk"); covOgk
eigen(covOgk@cov)$values # <7>
eigen(cov2cor(covOgk@cov))$values # <8>
```

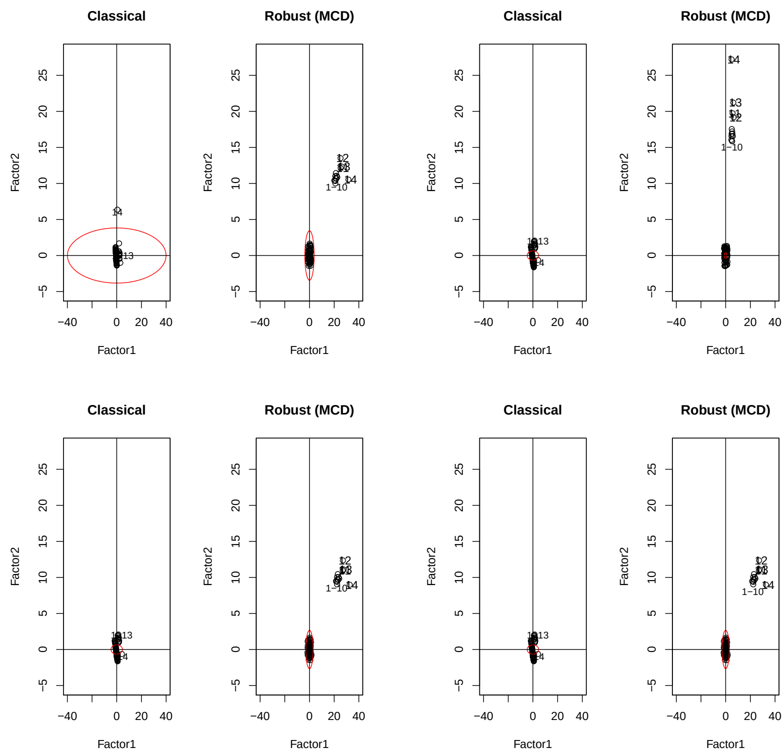


**Fig. 1. Classical and robust (MCD) scatterplots of the first two factor scores of the hbk data with 97.5% tolerance ellipses. First row: $< 1 >$ vs $< 5 >$; $< 3 >$ vs $< 7 >$. Second row: $< 2 >$ vs $< 6 >$; $< 4 >$ vs $< 8 >$**

**Table 6. The first two eigenvalues of the running matrices of the stock611 data. Source: Own research**

| | | Classical | Robust (OGK) |
|---|---|---|---|
| stock611[,3:12]) | cov | <1> 4.27e+20 1.99e+19 | <5> 3.99e+17 7.36e+16 |
| | cor | <2> 5.79 2.32 | <6> 5.16 2.41 |
| scale(stock611[,3:12])) | cov | <3> 5.79 2.32 | <7> 7.53e-01 2.97e-01 |
| | cor | <4> 5.79 2.32 | <8> 5.16 2.41 |

Classical and robust (OGK) scatterplots of the first two factor scores of the `stock611` data with 97.5% tolerance ellipses are plotted in fig 2. The scatterplots of the first two factor scores of combinations `<1>` and `<5>` are not shown, because errors occur in

```
solve.default(S): system is computationally singular.
```

To get a clearer view of the scatterplots, we zoom in the scatterplots. We see that the scores of <2>, <3>, and <4> are the same, the scores of <6> and <8> are the same, in agree with Theorem 2.3. Note that the tolerance ellipses for <2>, <3>, and <4> cover the outliers, due to the outliers severely affected the eigenvalues of the running matrices $\boldsymbol{R}^c = \tilde{\boldsymbol{S}}^c = \tilde{\boldsymbol{R}}^c$. The tolerance ellipses of <6> and <8> well separated the regular points and the outliers. The R codes for the comparison <4> vs <8> are given as follows. The R codes for other comparisons are similar and thus are omitted.

```
## <4> classical, x = scale(stock611[,3:12]), cor = TRUE (correlation matrix)
faClassic4 = FaClassic(x = scale(stock611[,3:12]), factors = 2, cor = TRUE,
method = "pca", scoresMethod = "regression"); faClassic4
## <8> robust, x = scale(stock611[,3:12]), cor = TRUE (correlation matrix)
faCov8 = FaCov(x = scale(stock611[,3:12]), factors = 2, cor = TRUE, method = "pca",
scoresMethod = "regression", cov.control = CovControlOgk()); faCov8
## ZoomIn: xlim = c(-10,10), ylim = c(-10,10)
## <4> vs <8>
usr <- par(mfrow = c(1,2))
cfaClassic <- list(center = c(0,0), cov = diag(faClassic4@eigenvalues[1:2]),
n.obs = faClassic4@n.obs)
rrcov:::.myellipse(faClassic4@scores, xcov = cfaClassic, main = "Classical",
xlab = "Factor1", ylab = "Factor2", xlim = c(-10,10), ylim = c(-10,10), id.n = 0)
abline(v = 0)
abline(h = 0)
cfaCov <- list(center = c(0,0), cov = diag(faCov8@eigenvalues[1:2]),
n.obs = faCov8@n.obs)
rrcov:::.myellipse(faCov8@scores, xcov = cfaCov, main = "Robust (OGK)",
xlab = "Factor1", ylab = "Factor2", xlim = c(-10,10), ylim = c(-10,10), id.n = 0)
abline(v = 0)
abline(h = 0)
par(usr)
```

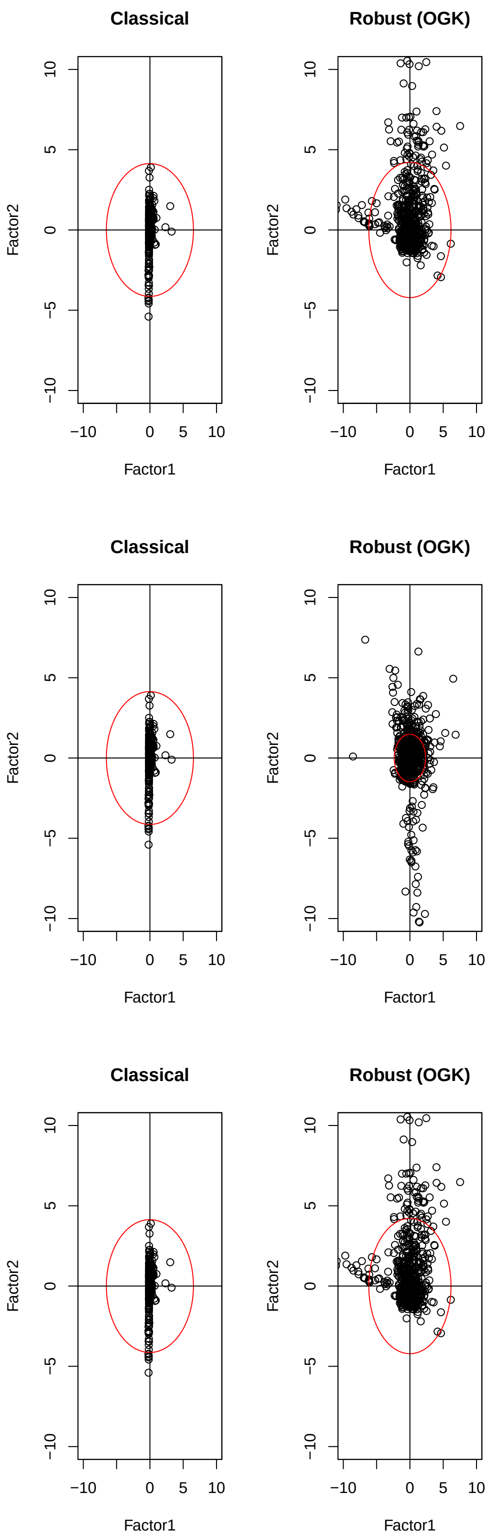


**Fig. 2. Classical and robust (OGK) scatterplots of the first two factor scores of the stock611 data with 97.5% tolerance ellipses. First row: $< 2 >$ vs $< 6 >$. Second row: $< 3 >$ vs $< 7 >$. Third row: $< 4 >$ vs $< 8 >$**

# 4 Conclusions

Consider the factor analysis methods (classical or robust), the data input (data or scaled data), and the running matrix (covariance or correlation) all together, there are 8 combinations. The objective of our study is to give a recommendation for classical and robust factor analysis. First, when the variables have different units, it is common to standardize the variables, and thus it is common to use the correlation matrix as the running matrix. Second, we need to explain the factors from the loading matrix. The entries of the loading matrix from the sample covariance matrix are not limited between 0 and 1, which makes the explanations of the factors hard. The first two reasons suggest us to choose `<2>` vs `<6>` and `<4>` vs `<8>`. Theorem 2.2 states that the robust correlation matrices of the data and the scaled data are the same. Theorem 2.3 states that `<2>` and `<4>` (`<6>` and `<8>`) have the same running matrices, eigenvalues, loadings, uniquenesses, scoring coefficients, scaled data matrices, and score matrices. That is, `<2>` vs `<6>` and `<4>` vs `<8>` give us the same comparisons. We can choose any pair to do the comparison. Third, we may not be able to compute the robust covariance matrix, and thus the robust correlation matrix of the original data, as the stocks data example illustrates. Consequently, we recommend `<4>` (classical factor analysis using the correlation matrix of the scaled data as the running matrix) vs `<8>` (robust factor analysis using the correlation matrix of the scaled data as the running matrix) for theoretical and computational reasons. The `hbk` data and the `stock611` data both illustrate the correctness of (2.1) and Theorems 2.2 and 2.3.

Finally, we give some additional suggestions for research in this area. For researchers and practitioners using factor analysis method to deal with real data with outliers, we recommend them to carry out classical and robust factor analysis using the correlation matrix of the scaled data as the running matrix.

# Acknowledgement

The authors gratefully acknowledge the constructive comments offered by the referees. Their comments improve the quality of the paper significantly. The research was supported by the Fundamental Research Funds for the Central Universities (CQDXWL-2012-004 and CDJRC10100010), China Scholarship Council (201606055028), and the MOE project of Humanities and Social Sciences on the west and the border area (14XJC910001).

# Competing Interests

Authors have declared that no competing interests exist.

# References

[1] Pison G, Rousseeuw PJ, Filzmoser P, Croux C. Robust factor analysis [J]. Journal of Multivariate Analysis. 2003;84:145-172.

[2] Rousseeuw PJ. Multivariate estimation with high breakdown point [C]. In Grossmann W, Pflug G, Vincze I, and Wertz W, editors, Mathematical Statistics and Applications Vol. B. Reidel Publishing, Dordrecht. 1985;283-297.

[3] Rousseeuw PJ, Van Driessen K. A fast algorithm for the minimum covariance determinant estimator [J]. Technometrics. 1999;41:212-223.

[4] Fritsch V, Varoquaux G, Thyreau B, Poline JB, Thirion B. Detecting outliers in high-dimensional neuroimaging datasets with robust covariance estimators [J]. Medical Image Analysis. 2012;16:1359-1370.

[5] Hou QJ, Crosser B, Mahnken JD, Gajewski BJ, Dunton N. Input data quality control for ndnqi national comparative statistics and quarterly reports: A contrast of three robust scale estimators for multiple outlier detection [J]. BMC Research Notes. 2012;5:456.

[6] Maronna RA, Zamar RH. Robust estimation of location and dispersion for high-dimensional datasets [J]. Technometrics. 2002;44:307-317.

[7] Rocke DM. Robustness properties of s-estimators of multivariate location and shape in high dimension [J]. The Annals of Statistics. 1996;24:1327-1345.

[8] Yang YN, Feng YL, Suykens JAK. Robust low-rank tensor recovery with regularized redescending m-estimator [J]. IEEE Transactions on Neural Networks and Learning Systems. 2016;27:1933-1946.

[9] Davies PL. Asymptotic behavior of s-estimators of multivariate location parameters and dispersion matrices [J]. The Annals of Statistics. 1987;15:1269-1292.

[10] Ruppert D. Computing s-estimators for regression and multivariate location/dispersion [J]. Journal of Computational and Graphical Statistics. 1992;1:253-270.

[11] Woodruff DL, Rocke DM. Computable robust estimation of multivariate location and shape in high dimension using compound estimators [J]. Journal of the American Statistical Association. 1994;89:888-896.

[12] Salibian-Barrera M,Yohai VJ. A fast algorithm for s-regression estimates [J]. Journal of Computational and Graphical Statistics. 2006;15:414-427.

[13] Stahel WA. Breakdown of covariance estimators [R]. Research Report 31, ETH Zurich; 1981.

[14] Stahel WA. Robuste schatzungen: Infinitesimale optimalitat und schatzungen von kovarianzmatrizen [D]. Phd thesis no. 6881, Swiss Federal Institute of Technology (ETH), Zurich; 1981.

[15] Donoho DL. Breakdown properties of multivariate location estimators [R]. Technical report, Harvard University, Boston; 1982.

[16] Maronna RA, Yohai VJ. The behavoiur of the Stahel-Donoho robust multivariate estimator [J]. Journal of the American Statistical Association. 1995;90:330-341.

[17] Maronna RA, Martin D, Yohai VJ. Robust Statistics: Theory and Methods [M]. John Wiley & Sons, New York; 2006.

[18] Todorov V, Filzmoser P. An object-oriented framework for robust multivariate analysis [J]. Journal of Statistical Software. 2009;32(3):1-47.

[19] Mardia KV, Kent JT, Bibby JM. Multivariate Analysis [M]. Academic Press, New York; 1976.

[20] Johnson RA and Wichern DW. Applied Multivariate Statistical Analysis [M]. Pearson, New Jersey. 6th edition; 2007.

[21] Xue Y, Chen LP. Statistical Modeling and R Software [M]. Tsinghua University Press, Beijing; 2007.

[22] Zhang YY. Robustfa: An object oriented solution for robust factor analysis [CP]. R package version 1.0-5; 2013.

[23] Wang XM. Applied multivariate analysis [M]. Shanghai University of Finance & Economics Press, Shanghai. 3rd edition; 2009.

---



***Peer-review history:***
*The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)*
*http://sciencedomain.org/review-history/18254*