

# The Influence Function of Graphical Lasso Estimators

Citation for published version (APA):

Louvet, G., Raymaekers, J., Van Bever, G., & Wilms, I. (2023). The Influence Function of Graphical Lasso Estimators. *Econometrics and Statistics*. Advance online publication. <https://doi.org/10.1016/j.ecosta.2023.03.004>

**Document status and date:**

E-pub ahead of print: 06/04/2023

**DOI:**

[10.1016/j.ecosta.2023.03.004](https://doi.org/10.1016/j.ecosta.2023.03.004)

**Document Version:**

Publisher's PDF, also known as Version of record

**Document license:**

Taverne

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

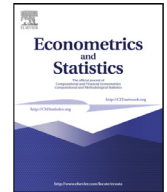
[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.



Contents lists available at ScienceDirect

## Econometrics and Statistics

journal homepage: [www.elsevier.com/locate/ecosta](http://www.elsevier.com/locate/ecosta)

## The Influence Function of Graphical Lasso Estimators

Gaëtan Louvet<sup>a,\*</sup>, Jakob Raymaekers<sup>b</sup>, Germain Van Bever<sup>a</sup>, Ines Wilms<sup>b</sup><sup>a</sup> Department of Mathematics and Namur Institute for Complex Systems (naXys), Université de Namur, Belgium<sup>b</sup> Department of Quantitative Economics, Maastricht University, the Netherlands

## ARTICLE INFO

## Article history:

Received 15 September 2022

Revised 31 March 2023

Accepted 31 March 2023

Available online xxx

## Keywords:

Asymptotic variance

Glasso

Gross-error sensitivity

Influence function

Outliers

Robustness

## ABSTRACT

The precision matrix that encodes conditional linear dependency relations among a set of variables forms an important object of interest in multivariate analysis. Sparse estimation procedures for precision matrices such as the graphical lasso (Glasso) gained popularity as they facilitate interpretability, thereby separating pairs of variables that are conditionally dependent from those that are independent (given all other variables). Glasso lacks, however, robustness to outliers. To overcome this problem, one typically applies a robust plug-in procedure where the Glasso is computed from a robust covariance estimate instead of the sample covariance, thereby providing protection against outliers. These estimators are studied theoretically, by deriving and comparing their influence function, sensitivity curve and asymptotic variance.

© 2023 EcoSta Econometrics and Statistics. Published by Elsevier B.V. All rights reserved.

## 1. Introduction

The estimation of precision (inverse covariance) matrices is indispensable in multivariate analysis with applications ranging from economics and finance to genomics, neuroscience and social networks (Fan et al., 2014). Let  $\mathbf{X} = (X_1, \dots, X_p)^\top$  be a  $p$ -variate random vector with finite second order moments, with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . We focus on estimating the precision matrix  $\boldsymbol{\Omega} := \boldsymbol{\Sigma}^{-1}$  whose entries capture the conditional linear dependencies between the components of  $\mathbf{X}$ . Importantly, when  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , the variables  $X_i$  and  $X_j$  are conditionally independent given all other variables if and only if  $(\boldsymbol{\Omega})_{i,j} = 0$ . In that case, recovering the sparsity pattern of the precision matrix is equivalent to recovering the graph structure of the Gaussian graphical model  $G = (V, E)$  where the vertex set  $V = \{1, \dots, p\}$  and the edge set  $E$  consists of the pairs  $(i, j)$  that are connected by an edge (thus having  $(\boldsymbol{\Omega})_{i,j} \neq 0$ ).

A common method for estimating sparse precision matrices is the graphical lasso (Glasso) (Yuan and Lin, 2007; Banerjee et al., 2008; Rothman et al., 2008; Friedman et al., 2008), defined as

$$\hat{\boldsymbol{\Omega}} = \arg \min_{\boldsymbol{\Omega}} \{-\log \det(\boldsymbol{\Omega}) + \text{tr}(\mathbf{S}\boldsymbol{\Omega}) + \lambda \|\boldsymbol{\Omega}^{-\text{diag}}\|_1 \text{ s.t. } \boldsymbol{\Omega} = \boldsymbol{\Omega}^\top, \boldsymbol{\Omega} \succ 0\}, \quad (1)$$

where  $\log \det(\boldsymbol{\Omega}) = \log(|\boldsymbol{\Omega}|)$ ,  $\text{tr}(\cdot)$  denotes the trace,  $\mathbf{S}$  is the sample covariance matrix,  $\boldsymbol{\Omega}^{-\text{diag}}$  is  $\boldsymbol{\Omega}$  with each diagonal element set to 0,  $\cdot \succ 0$  denotes a positive definite matrix, and  $\lambda > 0$  is a penalty parameter controlling the degree of sparsity. Throughout, we refer to the estimator in Eq. (1), with the sample covariance matrix as input, as the “standard Glasso”. By adding an  $\ell_1$ -penalty (in this case, the  $\ell_1$ -norm of the off-diagonal elements of  $\boldsymbol{\Omega}$ ) to the negative log-likelihood of a sample of multivariate normal random variables, a sparse precision matrix estimate is obtained thereby also permitting estimation in high-dimensional settings with more parameters than observations.

\* Corresponding author at: Rue de Bruxelles, 61; 5000 Namur, Belgium  
E-mail address: [gaetan.louvet@unamur.be](mailto:gaetan.louvet@unamur.be) (G. Louvet).

While sparse precision matrix estimation forms an active area of research, far less research has looked into the performance of such methods in the presence of outliers. Nonetheless, in many multivariate settings, the occurrence of outliers is to be expected. The standard *Gllasso*, however, performs poorly in presence of outliers as it relies on the sample covariance matrix  $\mathbf{S}$  in Eq. (1), which is sensitive to even a single outlying data point. Various robustification approaches of the *Gllasso* have been proposed. [Finegold and Drton \(2011\)](#) build upon the standard *Gllasso* but use the likelihood of the multivariate  $t$ -distribution instead of a normal one thereby providing better protection against data contamination. [Öllerer and Croux \(2015\)](#); [Croux and Öllerer \(2016\)](#); [Tarr et al. \(2016\)](#); [Lafit et al. \(2022\)](#) consider high-dimensional precision matrix estimation under cellwise contamination and propose to plug-in a pairwise robust covariance or correlation estimator instead of the (non-robust) sample covariance into Eq. (1). Applications of such robust sparse precision matrices to, for instance, discriminant analysis are studied in [Aerts and Wilms \(2017\)](#). However, most studies above do not provide a theoretical analysis on robust *Gllasso* estimators. Exceptions are [Öllerer and Croux \(2015\)](#); [Croux and Öllerer \(2016\)](#) who derive a finite sample breakdown point of their robust *Gllasso* under cellwise contamination, and [Loh and Tan \(2018\)](#) who analyze the statistical consistency of robust high-dimensional precision matrix estimators under cellwise contamination.

In this paper, we study the robustness of sparse *Gllasso*-based precision matrix estimators theoretically by deriving their influence function. The influence function is an important tool to measure the robustness of a statistical functional ([Hampel et al., 1986](#)); it quantifies the influence of a small amount of contamination placed at a given value  $\mathbf{z}$  on a statistical functional  $T$ , for data coming from a distribution  $F$ . More precisely, the influence function of the statistical functional  $T$  at the distribution  $F$  is, when it exists, given by

$$IF(\mathbf{z}; T, F) = \lim_{\varepsilon \rightarrow 0} \frac{T(F_{\varepsilon, \mathbf{z}}) - T(F)}{\varepsilon} = \left. \frac{\partial}{\partial \varepsilon} T(F_{\varepsilon, \mathbf{z}}) \right|_{\varepsilon=0}, \quad (2)$$

where  $F_{\varepsilon, \mathbf{z}} = (1 - \varepsilon)F + \varepsilon\Delta(\mathbf{z})$  is the contaminated distribution and  $\Delta(\mathbf{z})$  is the probability distribution that puts all its mass at  $\mathbf{z}$ . As commonly done in the study of influence functions, we thus limit ourselves to point mass contamination; different forms of outliers can be studied through other robustness concepts such as the breakdown point (see, e.g., [Öllerer and Croux \(2015\)](#); [Croux and Öllerer \(2016\)](#)).

The remainder of the article is structured as follows. In [Section 2](#), we present the influence function of the *Gllasso* for any plug-in scatter functional, then discuss the special case of the standard *Gllasso* and show that its influence function is unbounded. [Section 3.3](#) discusses how the *Gllasso* influence function can be bounded and expressions of the gross-error sensitivity are derived. We consider different robust *Gllasso* estimators and compare them in terms of their influence functions and sensitivity curves. In [Section 4](#), we compute the asymptotic variances and compare the robust *Gllasso* estimators in terms of their statistical efficiencies. [Section 5](#) concludes. All proofs, together with precise descriptions of the robust correlation measures used in [Section 3.3](#), are collected in the Appendix.

## 2. Influence Function of the *Gllasso*

We first derive the general expression of the influence function of the *Gllasso* for any plug-in scatter functional ([Section 2.1](#)). Next, we discuss the case of the standard *Gllasso* and establish its lack of robustness ([Section 2.2](#)).

### 2.1. The Influence Function

We start with deriving the influence function of the associated functional representation of the *Gllasso*. The following notation will be used throughout. Let  $p$  be fixed. For any  $p \times p$  matrix  $\mathbf{\Omega}$ , let  $\boldsymbol{\omega} := \text{vec}(\mathbf{\Omega})$  denote the  $p^2$ -dimensional vector resulting from stacking the columns of  $\mathbf{\Omega}$ . Let  $\text{mat}$  denote the inverse operator of  $\text{vec}$ , that is,  $\text{mat}(\text{vec}(\mathbf{\Omega})) = \mathbf{\Omega}$ . For a vector  $\boldsymbol{\omega} \in \mathbb{R}^{p^2}$  with  $s \leq p^2$  nonzero elements, let  $\mathbf{D}_{\boldsymbol{\omega}}$  denote a  $p^2 \times p^2$  symmetric permutation matrix putting the nonzero elements first. In particular,  $(\mathbf{D}_{\boldsymbol{\omega}})_{ij} = 0$  for all  $i > s$  and  $\mathbf{D}^2 = \mathbf{I}_{p^2}$ , the  $p^2 \times p^2$  identity matrix. Let  $\|\boldsymbol{\omega}\|_1 = \|\mathbf{\Omega}\|_1$  denote the usual  $\ell_1$ -norm and let  $\|\boldsymbol{\omega}\|_1^{-\text{diag}} = \|\text{mat}(\boldsymbol{\omega})^{-\text{diag}}\|_1$ . Let  $\otimes$  denote the Kronecker product,  $C^\infty(\mathbb{R}^{p^2})$  the set of continuous and infinitely differentiable functions on  $\mathbb{R}^{p^2}$  and let  $W^{2,2}(\mathbb{R}^{p^2})$  be the Sobolev space of functions with square integrable second weak derivatives.

For any distribution  $F$  on  $\mathbb{R}^p$  and any scatter functional  $\mathbf{S}(F)$ , we derive the influence function of the matrix-valued functional

$$T_{\mathbf{\Omega}}(F) = \arg \min_{\mathbf{\Omega} \in \mathbb{R}^{p \times p}} \{-\log \det(\mathbf{\Omega}) + \text{tr}(\mathbf{S}(F)\mathbf{\Omega}) + \lambda \|\mathbf{\Omega}^{-\text{diag}}\|_1 \text{ s.t. } \mathbf{\Omega} = \mathbf{\Omega}^\top, \mathbf{\Omega} > \mathbf{0}\}. \quad (3)$$

To do so, we rather provide the influence function of the equivalent vector-valued version

$$T_{\boldsymbol{\omega}}(F) = \arg \min_{\boldsymbol{\omega} \in \mathbb{R}^{p^2}} \{-\log \det(\text{mat}(\boldsymbol{\omega})) + (\text{vec}(\mathbf{S}(F)^\top))^\top \boldsymbol{\omega} + \lambda \|\boldsymbol{\omega}\|_1^{-\text{diag}}\}. \quad (4)$$

**Remark 1.** The equivalence between both definitions stems from the fact that, for any matrices  $\mathbf{S}$  and  $\mathbf{\Omega}$ ,  $\text{tr}(\mathbf{S}\mathbf{\Omega}) = (\text{vec}(\mathbf{S}^\top))^\top \text{vec}(\mathbf{\Omega})$ . Moreover, in (4), we do not impose  $\boldsymbol{\omega}$  to be the vectorisation of a symmetric, positive definite matrix. The solution to the minimisation problem, without loss of generality, always respects those constraints. This comes from the fact that, when looking at the first order conditions, the gradient stays symmetric and that the logdet function acts as a barrier to the positive definite cone (see [Srinivasan and Panda, 2022](#), for example). Therefore, it ensures that the solution (in matrix form) of the first order conditions is positive definite. Hence, it holds that  $T_{\mathbf{\Omega}}(F) = \text{mat}(T_{\boldsymbol{\omega}}(F))$ .

**Remark 2.** The choice of the plug-in scatter functional  $\mathbf{S}(F)$  will allow to distinguish between robust and non-robust Glasso alternatives. For example, the estimator provided in (1) is the empirical version of (3) obtained by setting

$$\mathbf{S}(F) = \mathbb{E}_F[(\mathbf{X} - \mathbb{E}_F[\mathbf{X}])(\mathbf{X} - \mathbb{E}_F[\mathbf{X}])^\top] \quad \text{and taking} \quad F = F_n.$$

Note that, unless  $\lambda = 0$  and  $S$  is the classical covariance, there is no guarantee that  $\mathbf{T}_\Omega = \mathbf{\Omega}$  in general.  $\mathbf{T}_\Omega$  inherits robustness properties of the scatter functional  $S$  and sparsity from the penalisation. Other choices of robust scatter functionals will be discussed in Section 3.3.

**Remark 3.** The expressions  $\mathbf{T}_\Omega(F)$  and  $\mathbf{T}_\omega(F)$  are short-hand for  $\mathbf{T}_{\Omega,\lambda}(F)$  and  $\mathbf{T}_{\omega,\lambda}(F)$  as they depend on the regularisation parameter  $\lambda$ . For simplicity, the former notation will be used and the dependence on  $\lambda$  will be implicit.

To derive the influence function of the Glasso, care is required due to the non-differentiability of the penalty term. Similarly to Avella-Medina (2017) for regression problems, we circumvent the problem by computing the influence function of the functional

$$\mathbf{T}_{\omega,p_m}(F) = \arg \min_{\omega \in \mathbb{R}^{p^2}} \{-\log \det(\text{mat}(\omega)) + (\text{vec}(\mathbf{S}(F)^\top))^\top \omega + \lambda p_m(\omega)\}, \quad (5)$$

where  $p_m \in C^\infty(\mathbb{R}^{p^2})$  is a convex and differentiable norm which converges to  $\|\cdot\|_1^{-\text{diag}}$  in  $W^{2,2}(\mathbb{R}^{p^2})$  as  $m \rightarrow \infty$ . Two examples of such sequences are given by  $p_m(\omega) = \|\text{mat}(\omega)\|_{k_m}^{-\text{diag}}$  or  $p'_m(\omega) = \sum_{i \neq j} |(\text{mat}(\omega))_{i,j}|^{k_m}$ , with  $1 < k_m \in \mathbb{R}$  for all  $m$  and  $k_m \searrow 1$ . We then define the influence function of  $\mathbf{T}_\omega$  as the limit for  $m \rightarrow \infty$  of the influence function of  $\mathbf{T}_{\omega,p_m}$ . The result is obtained provided the limit is independent of the chosen  $p_m$  sequence, which we state in the following lemma.

**Lemma 1.** Let  $F$  be a distribution over  $\mathbb{R}^p$  and let  $\mathbf{S}(F)$  be a scatter functional. For any two sequences  $p_m$  and  $p'_m$  converging to  $\|\cdot\|_1^{-\text{diag}}$  in  $W^{2,2}(\mathbb{R}^{p^2})$  and any  $\mathbf{z} \in \mathbb{R}^p$ , it holds that

$$\lim_{m \rightarrow \infty} (IF(\mathbf{z}; \mathbf{T}_{\omega,p_m}, F) - IF(\mathbf{z}; \mathbf{T}_{\omega,p'_m}, F)) = 0.$$

Moreover, if  $\sup_{\mathbf{z} \in \mathbb{R}^p} \|\text{vec}(IF(\mathbf{z}; S(F), F))\| < \infty$ , that is, if the gross error sensitivity of  $S(F)$  is bounded, then the convergence holds true uniformly:

$$\lim_{m \rightarrow \infty} \sup_{\mathbf{z} \in \mathbb{R}^p} (IF(\mathbf{z}; \mathbf{T}_{\omega,p_m}, F) - IF(\mathbf{z}; \mathbf{T}_{\omega,p'_m}, F)) = 0.$$

The theorem below then gives the influence function of  $\mathbf{T}_\omega(F)$ . The proof is provided in Appendix A.

**Theorem 1.** Let  $F$  be a distribution over  $\mathbb{R}^p$  and let  $\mathbf{S}(F)$  be a scatter functional. Let  $\mathbf{z} \in \mathbb{R}^p$ . Assume that  $\mathbf{T}_\omega := \mathbf{T}_\omega(F)$  has nonzero components. Let  $\mathbf{D} := \mathbf{D}_{\mathbf{T}_\omega}$ . Then, the influence function of the graphical lasso satisfies  $\mathbf{DIF}(\mathbf{z}; \mathbf{T}_\omega, F) =$

$$-\left( \begin{matrix} (\mathbf{D}(\text{mat}(\mathbf{T}_\omega)^{-1} \otimes \text{mat}(\mathbf{T}_\omega)^{-1})\mathbf{D})_{1:s,1:s}^{-1} (\mathbf{DIF}(\mathbf{z}; \text{vec}(\mathbf{S}(F)), F))_{1:s} \\ \mathbf{0} \end{matrix} \right).$$

Theorem 1 above shows that the influence of  $\mathbf{z}$  on null components of  $\mathbf{T}_\omega$  is 0. This translates to the fact that, under infinitesimal contamination, null values in the precision matrix remain zero, so that the contamination does not affect the sparsity of the solution obtained. This finding is similar to penalised regression models where the influence function is zero for the null components of the parameter vector (see, for example, Öllerer et al., 2015; Avella-Medina, 2017).

The influence of  $\mathbf{z}$  on the non-zero components of  $\mathbf{T}_\omega$  depends on the influence of  $\mathbf{z}$  on the plug-in scatter functional. A somewhat surprising conclusion at first sight is that the elements of the covariance matrix estimator corresponding to the null elements of the precision matrix do not appear in the influence function of the latter. This phenomenon originates from the fact that the contamination only appears through the covariance matrix estimator in the trace, more precisely, from the term

$$\text{tr}(\mathbf{S}(F_{\varepsilon,\mathbf{z}})\mathbf{\Omega}) = (\text{vec}(\mathbf{S}(F_{\varepsilon,\mathbf{z}})^\top))^\top \omega.$$

In turn, if  $(\mathbf{T}_\Omega(F))_{i,j} = (\mathbf{T}_\Omega(F_{\varepsilon,\mathbf{z}}))_{i,j} = 0$ , putting contamination in the corresponding elements of  $\mathbf{S}(F_{\varepsilon,\mathbf{z}})$  will not impact the scalar product, nor the estimation of the precision matrix.

## 2.2. Non-robustness of the Standard Glasso

Setting  $\mathbf{S}(F)$  in Eq. (4) to be the classical covariance functional yields the standard Glasso, which we now discuss. First, the standard Glasso has an unbounded influence function, hence is not robust to outliers, due to its reliance on the sample covariance. We illustrate its lack of robustness by computing the influence function and sensitivity curve for a setting with  $p = 3$  variables whose covariance matrix and corresponding sparse precision matrix are respectively given by

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{\Omega} = \begin{pmatrix} 4/3 & -2/3 & 0 \\ -2/3 & 5/3 & -2/3 \\ 0 & -2/3 & 4/3 \end{pmatrix}. \quad (6)$$

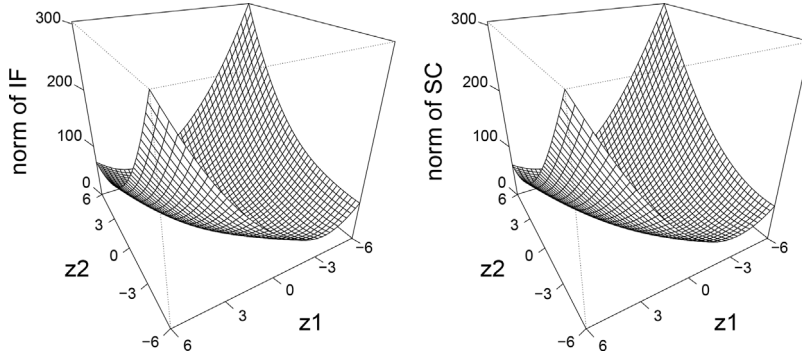


Fig. 1. Frobenius norm of the standard Glasso influence function (left) and sensitivity curve (right) for contamination at  $(z_1, z_2, 0)$ .

We introduce contamination at point  $\mathbf{z} = (z_1, z_2, 0)$  with values of  $z_1$  and  $z_2$  ranging from  $-6$  to  $6$ .

The left panel of Fig. 1 displays  $\|IF(\cdot; \mathbf{T}_\Omega, F)\|_F$ , the Frobenius norm of the influence function of the standard Glasso. The penalty parameter is set to  $\lambda = 8 \cdot 10^{-4}$  to ensure the correct sparsity pattern is recovered. The influence function of the standard Glasso is unbounded, hence it is not robust. This is to be expected as the influence function of the standard Glasso depends on that of the classical covariance, which is unbounded. The value of the norm increases fastest along the direction  $z_1 = -z_2$ .

Next, we compute the sensitivity curve introduced by Tukey (1977). The sensitivity curve can be thought of as a finite-sample equivalent of the influence function, as in many cases the sensitivity curve converges to the influence function when the sample size goes to infinity. Suppose we have a sample  $\mathbf{X}_{n-1} = \{\mathbf{x}_1, \dots, \mathbf{x}_{n-1}\}$  of size  $n - 1$  of  $p$ -variate observations. To this sample we add a variable point  $\mathbf{z}$  to obtain  $\mathbf{X}_z = \{\mathbf{x}_1, \dots, \mathbf{x}_{n-1}, \mathbf{z}\}$ . We now define the sensitivity curve by

$$SC_n(\mathbf{z}; \hat{\Omega}; \mathbf{X}_{n-1}) = n \cdot (\hat{\Omega}(\mathbf{X}_z) - \hat{\Omega}(\mathbf{X}_{n-1})). \tag{7}$$

It measures the change of the estimated precision matrix caused by adding the contaminated point  $\mathbf{z}$  to the clean sample, standardised by  $1/n$ , the amount of contamination. The right panel of Fig. 1 displays the Frobenius norm of the sensitivity curve for the standard Glasso (again with  $\lambda = 8 \cdot 10^{-4}$ ) for  $n = 1000$ . The results are averaged over 50 random samples of  $n - 1$  clean observations drawn from a multivariate normal distribution with the covariance matrix from Eq. (6) and an additional contaminated data point at location  $\mathbf{z} = (z_1, z_2, 0)$ . The sensitivity curve of the standard Glasso is unbounded, in line with the obtained results of the influence function.

Note that the Glasso is not orthogonally equivariant for  $\lambda > 0$ , and this, regardless of the initial covariance estimator used. Hence, the numerical results on the Glasso influence function and sensitivity curve presented here (and in later sections for robust plug-in estimators) are dependent on the choice of covariance  $\Sigma$ . This lack of equivariance is, however, a price one needs to pay for obtaining a regularised estimate of the precision matrix. Nonetheless, we believe that the results presented above shed light on the behavior of Glasso in the standard covariance case. Similarly, in the robust cases from Section 3.3, they provide valuable insights into the trade-off between robustness and statistical efficiency of the various Glasso estimators.

Next, we consider the gross-error sensitivity

$$\gamma^*(\mathbf{T}_\Omega, F) = \sup_{\mathbf{z} \in \mathbb{R}^p} \|IF(\mathbf{z}; \mathbf{T}_\Omega, F)\|_F,$$

which summarises the influence function in a single index by evaluating the maximal influence an observation may have. In the particular case of the standard Glasso, the gross-error sensitivity diverges since the influence function is unbounded. It still remains of interest, however, to determine in which direction the influence will increase the most, that is, to determine

$$\gamma_1^*(\mathbf{T}_\Omega, F) := \arg \max_{\mathbf{z} \in S^{p-1}} \|IF(\mathbf{z}; \mathbf{T}_\Omega, F)\|_F,$$

where  $S^{p-1}$  denotes the hypersphere in  $\mathbb{R}^p$ . The following lemma provides the result for the unpenalised case. The solution turns out to be the eigenvector corresponding with either the smallest or the largest eigenvalue, depending on a condition on the eigenvalues themselves.

**Lemma 2.** Let  $F$  be a distribution over  $\mathbb{R}^p$  and let  $\mathbf{S}(F)$  be the classical covariance functional. Fix  $\lambda = 0$ ,  $g: \mathbb{R}^+ \rightarrow \mathbb{R}^+ : x \mapsto g(x) = x^4 - 2x^3$  and let  $\lambda_1 \geq \dots \geq \lambda_p$  (resp.  $\mathbf{v}_1, \dots, \mathbf{v}_p$ ) be the eigenvalues (resp. eigenvectors) of  $\mathbf{T}_\Omega$ . It holds that

$$\gamma_1^*(\mathbf{T}_\Omega, F) = \begin{cases} \mathbf{v}_1 & \text{if } g(\lambda_1) \geq g(\lambda_p) \\ \mathbf{v}_p & \text{otherwise} \end{cases}$$

and

$$\max_{\mathbf{z} \in S^{p-1}} \|IF(\mathbf{z}; \mathbf{T}_\Omega, F)\|_F = \sum_{i=1}^p \lambda_i^2 + \max(g(\lambda_1), g(\lambda_p)).$$

Note that, in the penalised case, it is not possible to obtain an equivalent result to [Lemma 2](#) above. Indeed, in that situation, there is no guarantee that  $\mathbf{T}_\Omega = \mathbf{\Omega}$ , so that the influence function might depend on the precision matrix in more complex forms. It is, however, easy to show that [Lemma 2](#) also holds when  $\lambda$  is taken large enough so that  $\mathbf{T}_\Omega$  is diagonal.

### 3. Influence Function of Robust Glasso Estimators

In [Section 3.1](#) we discuss how the Glasso influence function can be bounded. [Section 3.2](#) then presents several plug-in covariance candidates to obtain a robust Glasso and compares them in terms of their Glasso influence function ([Section 3.3](#)) and sensitivity curve ([Section 3.4](#)).

#### 3.1. Bounding the Influence Function via Robust Plug-in Estimators

In [Section 2](#), we show that the influence function of the Glasso depends on the influence function of the plug-in scatter functional. To bound the influence function of a Glasso-type of estimator, one could plug in a robust covariance estimator instead of the sample covariance in [Eq. \(1\)](#). In [Section 3.2](#), we discuss several candidates for robust covariance plug-ins. Such robust Glasso estimators would indeed result in a bounded influence function since the gross-error sensitivity can be bounded as

$$\gamma^*(\mathbf{T}_\omega, F) \leq \left\| (\mathbf{D}(\text{mat}(\mathbf{T}_\omega)^{-1} \otimes \text{mat}(\mathbf{T}_\omega)^{-1})\mathbf{D})_{1:s,1:s}^{-1} \right\| \gamma^*((\mathbf{D}\text{vec}(\mathbf{S}))_{1:s}, F),$$

where the norm used is the operator norm. Note that, should  $(\mathbf{T}_\omega(F))_{i,j} \neq 0$  everywhere (implying  $\mathbf{D} = \mathbf{I}_p$ ), this bound simplifies to

$$\gamma^*(\mathbf{T}_\omega, F) \leq \|\mathbf{T}_\Omega\|^2 \gamma^*((\text{vec}(\mathbf{S}))_{1:s}, F).$$

#### 3.2. Robust Covariance Plug-ins

As candidates for the robust covariance estimator to be plugged into the Glasso objective function, we consider the popular, traditional row-wise robust Minimum Covariance Determinant (MCD) estimator ([Rousseeuw, 1984; 1985; Rousseeuw and Van Driessen, 1999](#)) as well as a robust covariance estimator based on pairwise robust correlation estimators (see e.g. [Öllerer and Croux, 2015; Croux and Öllerer, 2016](#)). Unlike the MCD, the latter are frequently used in the literature on Glasso robustifications as they are available in high-dimensional settings and protect against cellwise outliers ([Alqallaf et al., 2009](#)).

The influence function of the Glasso based on the MCD<sup>1</sup> can simply be obtained by plugging in the expression of the MCD influence function (see [Croux and Haesbroeck, 1999](#)) into the expression of the Glasso influence function provided in [Theorem 1](#). To arrive at the Glasso influence function based on a pairwise robust correlation estimator, we need to introduce additional notation related to the latter's functional representation.

Denote with  $S$  a scale functional and  $R$  a correlation functional. The definition of the covariance matrix functional  $\mathbf{T}_\Sigma$  based on pairwise robust correlation estimators is given by

$$(\mathbf{T}_\Sigma(F))_{j,k} = S(F_j)S(F_k)R(F_{j,k}), \quad (8)$$

where  $F_j$  and  $F_k$  denote the marginal distributions of the  $j$ th and  $k$ th components respectively, and  $F_{j,k}$  denotes the bivariate (marginal) distribution of the  $j$ th and  $k$ th components.

As a scale estimator, we use the  $Q_n$  estimator ([Rousseeuw and Croux, 1993](#)), whose functional version is given by  $S(F) = cH^{-1}(\frac{1}{4})$ , where  $c$  is a constant ensuring Fisher-consistency and  $H$  is the distribution function of the absolute difference of two independent random variables, each with the same distribution  $F$ .

As robust correlation estimator, we consider Spearman's rank correlation ([Spearman, 1904](#)), Kendall's correlation coefficient ([Kendall, 1938](#)), the Gaussian rank correlation ([Boudt et al., 2012](#)) and the Quadrant correlation ([Blomqvist, 1950](#)). As our main focus is the influence function, we will consider the functional version of these estimators below. For the sake of completeness, we give the more common finite-sample definitions of these estimators in [Appendix B](#).

The functional form of Spearman's rank correlation coefficient  $R_S$  is given by

$$R_S(F_{j,k}) = 12\mathbb{E}_{F_{j,k}}[F_j(X_j)F_k(X_k)] - 3,$$

where  $(X_j, X_k) \sim F_{j,k}$  and  $F_j$  and  $F_k$  are the marginal distributions of  $X_j$  and  $X_k$  as before. Kendall's correlation coefficient  $R_K$  is given by

$$R_K(F_{j,k}) = \mathbb{E}_{F_{j,k}}[\text{sign}((X_{j_1} - X_{j_2})(X_{k_1} - X_{k_2}))],$$

where  $(X_{j_1}, X_{k_1})$  and  $(X_{j_2}, X_{k_2})$  are independent bivariate random variables, each with distribution  $F_{j,k}$ . The Gaussian rank correlation coefficient is given by

$$R_G(F_{j,k}) = \mathbb{E}_{F_{j,k}}[\Phi^{-1}(F_j(X_j))\Phi^{-1}(F_k(X_k))],$$

<sup>1</sup> Note that, throughout this paper, MCD is defined using 75% of the data, so as to achieve a balance between robustness and efficiency. For the sake of comparison, we also considered the reweighted MCD, with the same 25% breakdown point.

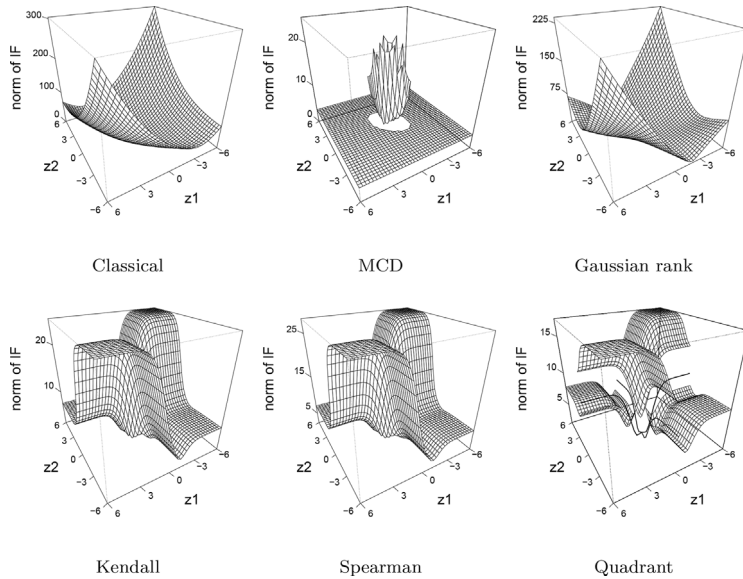


Fig. 2. Norm of the Glasso influence function based on several covariance matrix estimators for contamination at  $(z_1, z_2, 0)$ .

where  $\Phi^{-1}$  denotes the quantile function of the standard normal distribution. Note that if  $F_{j,k}$  is bivariate normal, we obtain the functional of the Pearson correlation coefficient. The Quadrant correlation coefficient is given by

$$R_Q(F_{j,k}) = \mathbb{E}_{F_{j,k}} [\text{sign}((X_j - \text{median}(X_j))(X_k - \text{median}(X_k)))].$$

For comparability of the influence functions of the Glasso based on the different plug-in covariances, it is important that they all estimate the same population quantity, i.e. are Fisher consistent. Hence, we use the Fisher consistent versions of the Spearman, Kendall and Quadrant correlation at the normal model by applying the non-linear transformations as given, for instance, in [Croux and Dehon \(2010\)](#). In particular, we use  $\tilde{R}_S = 2 \sin(\frac{\pi}{6} R_S)$ ,  $\tilde{R}_K = \sin(\frac{\pi}{2} R_K)$  and  $\tilde{R}_Q = \sin(\frac{\pi}{2} R_Q)$ . No transformation for the Gaussian rank correlation is needed as it is Fisher consistent at the normal model. Note that, in finite samples, these transformations can destroy the positive semidefiniteness (PSD) of the resulting covariance matrix. This is typically dealt with by projecting the estimate onto the space of PSD matrices, for example with respect to the Frobenius norm (as in [Öllerer and Croux, 2015](#)) or the elementwise  $\ell_\infty$  norm (as in [Loh and Tan, 2018](#)). This does not affect the discussion on the influence function.

The influence function of the resulting covariance estimator is then given by

$$\begin{aligned} IF(\mathbf{z}; (\mathbf{T}_\Sigma(F))_{j,k}, F) &= IF(z_j; S, F_j)S(F_k)R(F_{j,k}) \\ &+ S(F_j)IF(z_k; S, F_k)R(F_{j,k}) \\ &+ S(F_j)S(F_k)IF(\mathbf{z}; R, F_{j,k}). \end{aligned}$$

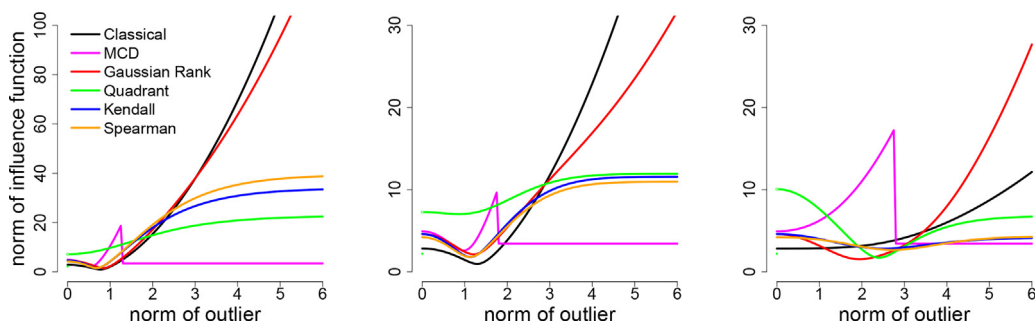
To compute this influence function, the expression of the influence function of the scale estimator, namely the  $Q_n$ , and the considered correlation estimators are thus needed. The former is available in [Rousseeuw and Croux \(1993\)](#), the latter in [Croux and Dehon \(2010\)](#) for Kendall, Spearman and Quadrant, and [Boudt et al. \(2012\)](#) for the Gaussian rank.

In the next subsections, we present the influence functions and sensitivity curves of the Glasso estimators based on these different robust covariances.

### 3.3. Influence Functions

We consider the same example as in [Section 2.2](#) and obtain the influence function of the Glasso (with the same fixed penalty parameter as before) computed from five different robust covariance plug-ins, namely the MCD, and the pairwise correlation estimators based on Spearman, Kendall, Gaussian rank and Quadrant.

[Fig. 2](#) displays the norm of the Glasso influence functions for the various covariance estimators, and includes the one for the classical covariance (from [Fig. 1](#)) to facilitate comparisons. When using the MCD as initial estimator, we see that the norm of the influence function resembles the one for the standard Glasso around the center, while the influence becomes bounded for points further away from it. The Glasso influence function based on the MCD as initial estimator is thus B-robust (see [Hampel et al., 1986](#)). However, while being highly robust, it should be noted that it is computationally demanding to obtain and, in practice, only available when  $p$  is small relative to  $n$ . Note that the same comments apply for the reweighted version of MCD, see [Fig. C.4](#) of [Appendix C](#).



**Fig. 3.** Norm of the Glasso (with  $\lambda = 8 \cdot 10^{-4}$ ) influence function based on several covariance matrix estimators for contamination in the direction of the eigenvector corresponding to the first (left), second (middle) and third (right) eigenvalue of  $T_{\Omega}$  based on the true covariance matrix.

Next, we turn to the Glasso influence function based on the Gaussian rank. At first sight, it largely resembles the results for the classical covariance, as its influence function is also unbounded. This is not surprising, as the influence function of the Gaussian rank correlation is unbounded (Boudt et al., 2012). Nonetheless, the values of the norm are smaller, thereby indicating a somewhat larger resistance to single outliers than the classical covariance. This is most likely due to the use of  $Q_n$  as an estimator of the marginal scales. This unboundedness of the influence function is somewhat misleading however, and our results on the sensitivity curve (to be discussed in Section 3.4) indicate that the Gaussian rank yields a substantially more robust estimator than the standard Glasso in finite samples.

The Glasso influence functions based on Kendall and Spearman are very similar: they are bounded and smooth, hence the Glasso computed from these initial covariance matrices is robust to outliers. Finally, the Glasso influence function based on the Quadrant is also bounded but not smooth: it has jumps at the coordinate axes due to the occurrence of the sign function and the median in its definition. Hence, small changes in data points close to the median of one of the marginals may lead to relatively large changes in its Glasso. Nonetheless, usage of the Quadrant as initial estimator results in the overall smallest values of the norm of the Glasso influence function. This is in line with findings in the setting of correlation estimation where the Quadrant is known to be the “most B-robust”, i.e. it has the lowest gross-error sensitivity among a class of correlation estimators based on product moments (hence not including the MCD) (Raymaekers and Rousseeuw, 2021).

We summarise the performance of the Glasso estimators based on the different initial covariances in Fig. 3. Fig. 3 displays the norm of the Glasso influence functions for increasing values of the outlier’s norm where the contamination is placed in the direction of the eigenvectors of  $T_{\Omega}$  based on the true covariance matrix. In line with our previous discussion, we see that the influence function for the standard Glasso and the Glasso based on the Gaussian rank are unbounded. The Glasso influence function based on the MCD, Quadrant, Spearman and Kendall are all bounded, with the former two being most robust for more severe outliers. Finally, much like the unpenalised case described in Lemma 2, the influence seems higher in the direction associated with the largest eigenvalue of  $T_{\Omega}$  (left panel Fig. 3).

### 3.4. Sensitivity Curves

We compute the Glasso sensitivity curves based on pairwise correlation estimators computed from the Gaussian rank, Spearman and Kendall correlation, analogously to the procedure described in Section 2.2 for the sample covariance<sup>2</sup>. Likewise, the Quadrant case is omitted as its discontinuity leads to unstable behaviour in finite samples. Note that, contrary to the influence functions which are always well-defined, the sensitivity curves can only be computed provided  $S(F_n)$  exists. In particular, this requires  $p < n$  in some cases (classical covariance function, MCD). Results are available in Figs. C.5 and C.6 of Appendix C.

The Glasso sensitivity curves for the Spearman and Kendall correlation are bounded and similar, in line with the results of Section 3.3. The Glasso computed from the Gaussian rank is considerably more robust than the standard Glasso in finite samples. Nonetheless, its influence function does not reflect this as it is unbounded (see Fig. 2). This apparent contradiction was discussed by Boudt et al. (2012) for the Gaussian rank correlation estimator. More precisely, for any fixed sample size, the sensitivity curves for the Gaussian rank correlation and the Pearson correlation are both bounded due to the boundedness of the correlation coefficient itself. As the sample size increases, however, the finite-sample gross-error sensitivities of both estimators diverge to infinity which matches with their unbounded influence functions. The key difference is that the Pearson correlation has a finite-sample gross-error sensitivity diverging at a rate of  $\sim n$ , whereas that of the Gaussian rank correlation diverges much slower at a rate of  $\sim \log(n)$ . This explains their difference in robustness, which translates to the Glasso setting. Still, the Glasso computed from the Spearman and Kendall correlations are considerably more robust.

<sup>2</sup> We omit results for the MCD as it suffers from approximation errors when using the classical FastMCD algorithm and is computationally demanding to compute exactly



**Table 1**

Asymptotic efficiencies of the Glasso estimator based on several robust covariance matrix estimators relative to the standard Glasso and this for three components of the precision matrix.

Component	Glasso with covariance estimator					
	Gaussian rank	Kendall	Spearman	Quadrant	MCD	Reweighted MCD
(1,1)	0.8210	0.8150	0.8097	0.4866	0.3003	0.6755
(2,2)	0.8085	0.8091	0.8027	0.4187	0.3076	0.6709
(2,1)	0.9563	0.8725	0.8491	0.3004	0.2556	0.7101

#### 4. Asymptotic Variances

Apart from robustness, it is desirable that an estimator has a high statistical efficiency. To measure the statistical efficiency of the different Glasso estimators, we first provide the expressions of their asymptotic variances, then compare the asymptotic variances to the standard Glasso.

Should a Bahadur representation of  $\mathbf{T}_\Omega(F)$  exist, one would then conclude to asymptotic normality and Fréchet differentiability of the precision matrix estimator. The asymptotic variances would then be given by

$$\text{ASV}((\mathbf{T}_\Omega)_{i,j}, F) = \int_{\mathbb{R}^p} \text{IF}^2(\mathbf{z}; (\mathbf{T}_\Omega)_{i,j}, F) dF(\mathbf{z}),$$

for any  $(i, j)$  such that  $(\mathbf{T}_\Omega(F))_{i,j} \neq 0$ . In matrix form, plugging in the expression of the influence function derived in [Theorem 1](#) yields

$$\text{ASV}((\mathbf{D}\mathbf{T}_\omega)_{1:s}, F) = \mathbf{A}(\mathbf{D}\text{ASV}(\text{vec}(\mathbf{S}(F)), F)\mathbf{D})_{1:s,1:s} \mathbf{A}^\top, \quad (9)$$

with

$$\mathbf{A} = (\mathbf{D}(\text{mat}(\mathbf{T}_\omega)^{-1} \otimes \text{mat}(\mathbf{T}_\omega)^{-1})\mathbf{D})_{1:s,1:s}^{-1}.$$

A formal verification of the Bahadur representation seems to be challenging and is still an open question. For example, in the context of scatter estimation, results include [He and Shao \(1996\)](#) (for  $M$ -estimation) or [Cator and Lopuhaä \(2010\)](#) (for the MCD estimator). Asymptotic normality results in the graphical models context are typically based on minimax bounds (see, for example, [Ren et al. \(2015\)](#) and references therein).

We now turn to the numerical example introduced in [Section 2.2](#), and compute the asymptotic variances for the Glasso based on the MCD, and pairwise correlation estimators. [Table 1](#) summarises the asymptotic efficiencies relative to the standard Glasso for three different components— diagonal elements (1,1), (2,2) and off-diagonal element (2,1)—of the precision matrix. Results on the other components are similar and therefore omitted. The values from [Table 1](#) were obtained from [\(9\)](#) using numerical integration. The asymptotic variances obtained above are consistent with simulation results conducted by the authors.

The Glasso based on Gaussian rank returns, overall, the highest efficiencies but is closely followed by the Glasso based on the Kendall and Spearman correlations. Their asymptotic efficiencies remain above 80%. We suspect that the efficiency of around 80% for the diagonal elements stems mainly from the fact that we are using the  $Q_n$  estimator to estimate the diagonal elements of our initial covariance matrix. The  $Q_n$  has an efficiency of 82% at Gaussian data ([Rousseeuw and Croux, 1993](#)). The Glasso based on Quadrant correlation and MCD, on the other hand, suffer from a severe loss of statistical efficiencies with drops to around 30%–40% for the former and to around 25–30% for the MCD. The reweighted MCD displays considerably higher efficiencies of around 70%.

#### 5. Conclusion

The standard Glasso is one of the most often used sparse estimators for precision matrices yet it may seriously be affected by the presence of even a single outlier. In this paper, we study the robustness of the Glasso by deriving expressions of its influence function for any plug-in scatter functional. We show that the influence function of the standard Glasso is unbounded, thereby proving its lack of robustness. Nonetheless, the Glasso influence function can be easily bounded by plugging in a robust covariance matrix estimate instead of the sample covariance.

We consider several robust Glasso estimators relying on different covariance matrix plug-ins (namely the MCD or pairwise correlation estimators based on Spearman, Kendall, Gaussian rank or Quadrant), and study their trade-off in robustness versus statistical efficiency. When plugging in a pairwise robust correlation estimate based on Kendall or Spearman, the resulting Glasso combines the attractive properties of (i) a bounded and smooth influence function, and (ii) high statistical efficiency at the normal model. Opting for the Glasso based on the Gaussian rank leads to a higher statistical efficiency, better protection against single outliers than the standard Glasso in finite samples but a price is paid in terms of robustness compared to the Glasso based on Spearman or Kendall. Opting for the Glasso based on the MCD or Quadrant correlation, on the other hand, results in a very strong resistance against outliers but low statistical efficiency though the efficiency of the

Glasso MCD is considerably increased by considering its reweighted version. Besides, the MCD is computationally demanding and only available for  $p < n$ , thereby limiting its applicability to low-dimensional settings.

An appropriate choice of initial covariance matrix estimate to be plugged into the Glasso thus depends on the trade-off between robustness and statistical efficiency the practitioner is willing to make. As general advice to applied statisticians, we recommend to first apply the standard Glasso as well as a Glasso estimator with a highly-robust covariance plug-in. If both yield very different results, then outliers are likely present and one could opt for a Glasso estimator with strong resistance against outliers. If both yield similar results, then one could opt for a robust Glasso estimator with high statistical efficiency.

The presented results on the influence function of Glasso allow for the study of (optimal) B-robustness within classes of estimators (Hampel et al., 1986, p. 116). For example, a relevant question would be whether we can find the estimator with the highest efficiency given a bound on the gross-error sensitivity, within the class of Glasso estimators based on pairwise initial covariance estimators. In addition to B-robustness, also V-robustness (Hampel et al., 1986, p. 128) could be studied. We consider these interesting directions for future research.

## Acknowledgements

We thank the associate editor and referees for their constructive comments which substantially improved the quality of the manuscript. We acknowledge support through the HiTEC Cost Action CA21163. The first and third author's work is supported by the Action de Recherche Concert e ARC-IMAL. The last author is supported by the Dutch Research Council (NWO) under grant number VI.Vidi.211.032.

## Appendix A. Proofs

We collect all proofs from the paper in this first appendix.

**Proof of Theorem 1.** Let  $p_m$  be a sequence of convex norms in  $C^\infty(\mathbb{R}^{p^2})$  which converges to  $\|\cdot\|_1^{\text{diag}}$  in  $W^{2,2}(\mathbb{R}^{p^2})$  as  $m \rightarrow \infty$  and let  $\mathbf{T}_{\omega, p_m}(F)$  be defined as in (5). Without loss of generality, assume that the Hessian matrix  $\nabla^2 p_m(\cdot)$  is everywhere diagonal<sup>3</sup>. Throughout, we write  $\mathbf{\Omega} = \text{mat}(\boldsymbol{\omega})$  and  $\mathbf{T}_{\omega, p_m}(F) = \mathbf{T}_{\omega, p_m}$  when there is no ambiguity to the underlying distribution. Since the function to be minimised in  $\mathbf{T}_{\omega, p_m}$  is differentiable, the first order condition writes

$$-\text{vec}(\text{mat}(\mathbf{T}_{\omega, p_m}(F))^{-\top}) + \text{vec}(\mathbf{S}(F)) + \lambda \nabla p_m(\mathbf{T}_{\omega, p_m}(F)) = 0.$$

Note that the first term comes from the fact that (see Kollo and von Rosen, 2005, page 134)

$$\frac{d}{d\omega_{i,j}}(-\log\det(\text{mat}(\boldsymbol{\omega}))) = -\frac{1}{\det(\text{mat}(\boldsymbol{\omega}))} \frac{d}{d\omega_{i,j}} \det(\text{mat}(\boldsymbol{\omega})) = -(\text{mat}(\boldsymbol{\omega})^{-1})_{j,i},$$

so that  $\nabla(-\log\det(\text{mat}(\boldsymbol{\omega}))) = -\text{vec}(\text{mat}(\boldsymbol{\omega})^{-\top})$ . Define

$$f(\varepsilon, \boldsymbol{\omega}) = -\text{vec}(\text{mat}(\boldsymbol{\omega})^{-1}) + \text{vec}(\mathbf{S}(F_{\varepsilon, \mathbf{z}})) + \lambda \nabla p_m(\boldsymbol{\omega}).$$

In particular, it holds  $f(0, \mathbf{T}_{\omega, p_m}(F)) = 0$ . The partial derivatives at  $(0, \mathbf{T}_{\omega, p_m}(F))$  are given by

$$\left. \frac{df(\varepsilon, \boldsymbol{\omega})}{d\varepsilon} \right|_{\varepsilon=0} = \text{vec}(\text{IF}(\mathbf{z}; \mathbf{S}(F), F))$$

and

$$\left. \frac{df(\varepsilon, \boldsymbol{\omega})}{d\boldsymbol{\omega}} \right|_{\boldsymbol{\omega}=\mathbf{T}_{\omega, p_m}} = \lambda \nabla^2 p_m(\mathbf{T}_{\omega, p_m}) + \mathbf{U}(\mathbf{T}_{\omega, p_m}),$$

where (see Kollo and von Rosen, 2005, pages 127-131)

$$\mathbf{U}(\boldsymbol{\omega}) := \frac{d(-\text{vec}(\mathbf{\Omega}^{-1}))}{d\text{vec}(\mathbf{\Omega})} = -\frac{d(\mathbf{\Omega}^{-1})}{d\mathbf{\Omega}} = \mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1} = \text{mat}(\boldsymbol{\omega})^{-1} \otimes \text{mat}(\boldsymbol{\omega})^{-1}.$$

The implicit function theorem yields

$$\text{IF}(\mathbf{z}; \mathbf{T}_{\omega, p_m}, F) = -(\lambda \nabla^2 p_m(\mathbf{T}_{\omega, p_m}) + \mathbf{U}(\mathbf{T}_{\omega, p_m}))^{-1} \text{vec}(\text{IF}(\mathbf{z}; \mathbf{S}(F), F)).$$

Note that this result applies here since the derivative with respect to  $\boldsymbol{\omega}$  is nowhere vanishing, a consequence of the positive semi-definiteness of  $\nabla^2 p_m(\boldsymbol{\omega})$  and of the positive definiteness of  $\mathbf{U}(\boldsymbol{\omega})$  (see Srinivasan and Panda, 2022). Taking now the limit in  $m$ ,  $\lim_{m \rightarrow \infty} \text{IF}(\mathbf{z}; \mathbf{T}_{\omega, p_m}, F)$  rewrites

$$\lim_{m \rightarrow \infty} -(\lambda \nabla^2 p_m(\mathbf{T}_{\omega, p_m}) + \mathbf{U}(\mathbf{T}_{\omega, p_m}))^{-1} \text{vec}(\text{IF}(\mathbf{z}; \mathbf{S}(F), F))$$

<sup>3</sup> This is the case for a relaxation proposed in Avella-Medina (2016), page 57-58. The proof of Lemma 1 below shows that the limit is independent of the chosen sequence.

$$\begin{aligned}
 &= \lim_{m \rightarrow \infty} -\mathbf{D}\mathbf{D}^{-1}(\lambda \nabla^2 p_m(\mathbf{T}_{\omega, p_m}) + \mathbf{U}(\mathbf{T}_{\omega, p_m}))^{-1} \mathbf{D}^{-\top} \mathbf{D}^{\top} \text{vec}(\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{S}(F), F)) \\
 &= \lim_{m \rightarrow \infty} -\mathbf{D}(\mathbf{D}^{\top} \lambda \nabla^2 p_m(\mathbf{T}_{\omega, p_m}) \mathbf{D} + \mathbf{D}^{\top} \mathbf{U}(\mathbf{T}_{\omega, p_m}) \mathbf{D})^{-1} \mathbf{D}^{\top} \text{vec}(\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{S}(F), F)) \\
 &= \lim_{m \rightarrow \infty} -\mathbf{D}(\lambda \nabla^2 p_m(\mathbf{D}\mathbf{T}_{\omega, p_m}) + \mathbf{D}^{\top} \mathbf{U}(\mathbf{T}_{\omega, p_m}) \mathbf{D})^{-1} \mathbf{D}^{\top} \text{vec}(\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{S}(F), F)) \\
 &= -\mathbf{D} \left( \lim_{m \rightarrow \infty} (\lambda \nabla^2 p_m(\mathbf{D}\mathbf{T}_{\omega, p_m}) + \mathbf{D}^{\top} \mathbf{U}(\mathbf{T}_{\omega, p_m}) \mathbf{D})^{-1} \right) \mathbf{D} \text{vec}(\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{S}(F), F)).
 \end{aligned}$$

We now show that the remaining limit converges to

$$\begin{pmatrix} (\mathbf{D}\text{mat}(\mathbf{T}_{\omega}(F))^{-1} \otimes \text{mat}(\mathbf{T}_{\omega}(F))^{-1}) \mathbf{D}^{-1}_{1:s, 1:s} & 0 \\ 0 & 0 \end{pmatrix},$$

which allows to conclude (using  $\mathbf{D}^2 = \mathbf{I}_p$ ). To show the convergence, partition

$$\lambda \nabla^2 p_m(\mathbf{D}\mathbf{T}_{\omega, p_m}) + \mathbf{D}^{\top} \mathbf{U}(\mathbf{T}_{\omega, p_m}) \mathbf{D} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

into blocks of  $s$  (resp.  $(p^2 - s)$ ) lines/columns. Provided all inverses exist, the inverse is given by

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -A_{11}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{pmatrix}.$$

Notice that, since the inverse function is continuous and the Hessian matrix  $\nabla^2 p_m(\mathbf{T}_{\omega, p_m})$  is diagonal, it holds  $(\nabla^2 p_m(\mathbf{D}\mathbf{T}_{\omega}(F)))_{jj} \rightarrow 0$  for  $j = 1, \dots, s$  and  $(\nabla^2 p_m(\mathbf{D}\mathbf{T}_{\omega}(F)))_{jj}$  diverges to  $+\infty$  for  $j = s + 1, \dots, p^2$ . Therefore,  $A_{22}^{-1} \rightarrow 0$ ,  $(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \rightarrow 0$  and  $(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} \rightarrow (\mathbf{D}^{\top} \mathbf{U}(\mathbf{T}_{\omega, p_m}) \mathbf{D})^{-1}_{1:s, 1:s}$ , allowing to conclude to the claimed limit.

Note that we formally showed that  $\lim_{m \rightarrow \infty} \mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{T}_{\omega, p_m}, F)$  converges to a limit that might be distinct from the influence function should the derivative not exist. In our situation, however, the components of the limiting result are either 0 (showing that the component of the estimated precision matrix stays at zero) or the derivative taken at a positive value (which then is differentiable), which therefore concludes the proof.

**Proof of Lemma 1.** We prove first that  $\lim_{m \rightarrow \infty} \mathbf{T}_{\omega, p_m}(F)$  is unique. Using Lemma 2 in [Avella-Medina \(2016\)](#) with  $\Lambda_{\lambda}(\omega; F, p) = -\log \det(\omega) + \text{vec}(\mathbf{S})^{\top} \omega + \lambda p_m(\omega)$  readily gives

$$\lim_{m \rightarrow \infty} \mathbf{T}_{\omega, p_m}(F) = \mathbf{T}_{\omega}(F). \tag{A.1}$$

We can now prove the independence of  $\lim_{m \rightarrow \infty} \mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{T}_{\omega, p_m}, F)$  of  $p_m$ . Let  $\{p_m\}_{m \geq 1}$  and  $\{p'_m\}_{m \geq 1}$  be two sequences in  $C^{\infty}(\mathbb{R}^{p^2})$  converging to  $p$  in  $W^{2,2}(\mathbb{R}^{p^2})$ . We have

$$\begin{aligned}
 \mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{T}_{\omega, p_m}, F) - \mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{T}_{\omega, p'_m}, F) &= \left( -(\lambda \nabla^2 p_m(\mathbf{T}_{\omega, p_m}(F)) + \mathbf{U}(\mathbf{T}_{\omega, p_m}(F)))^{-1} \right. \\
 &\quad \left. + (\lambda \nabla^2 p'_m(\mathbf{T}_{\omega, p'_m}(F)) + \mathbf{U}(\mathbf{T}_{\omega, p'_m}(F)))^{-1} \right) \text{vec}(\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{S}(F), F)).
 \end{aligned}$$

Since the inverse of a matrix is continuous and  $\nabla^2 p_m$  and  $\nabla^2 p'_m$  are sequences in  $C^{\infty}(\mathbb{R}^{p^2})$ , we find, using (A.1),

$$\lim_{m \rightarrow \infty} (\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{T}_{\omega, p_m}, F) - \mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{T}_{\omega, p'_m}, F)) = 0.$$

The convergence becomes uniform if  $\sup_{\mathbf{z} \in \mathbb{R}^p} \|\text{vec}(\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{S}(F), F))\| < \infty$ , since it is the only term depending on  $\mathbf{z}$ .  $\square$

**Proof of Lemma 2.** Without loss of generality due to the equivariance of  $\mathbf{S}$ , we assume throughout that  $\mathbb{E}_F[\mathbf{X}] = 0$ . It is easy to see that  $\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{S}, F) = \mathbf{z}\mathbf{z}^{\top} - \mathbf{\Sigma}$ . Hence, since  $\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C}) = \text{tr}(\mathbf{B}\mathbf{C}\mathbf{A})$  and using the spectral decomposition  $\mathbf{\Omega} = \mathbf{O}\mathbf{\Lambda}\mathbf{O}^{\top}$ , it follows

$$\begin{aligned}
 \|\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{T}_{\omega}, F)\|_F^2 &= \|\mathbf{\Omega}\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{S}, F)\mathbf{\Omega}\|_F^2 = \|\mathbf{\Omega}(\mathbf{z}\mathbf{z}^{\top} - \mathbf{\Sigma})\mathbf{\Omega}\|_F^2 = \|\mathbf{\Omega}^2(\mathbf{z}\mathbf{z}^{\top} - \mathbf{\Sigma})\|_F^2 \\
 &= \|\mathbf{\Omega}^2\mathbf{z}\mathbf{z}^{\top} - \mathbf{\Omega}\|_F^2 = \|\mathbf{\Lambda}^2\mathbf{x}\mathbf{x}^{\top} - \mathbf{\Lambda}\|_F^2 \\
 &= \sum_{i=1}^p \lambda_i^2 + \sum_{i=1}^p x_i^2 (\lambda_i^4 - 2\lambda_i^3),
 \end{aligned}$$

with  $\mathbf{x} = (x_1, \dots, x_p)^{\top} = \mathbf{O}^{\top} \mathbf{z}$ . Since  $x_i = \mathbf{v}_i^{\top} \mathbf{z}$  and using  $\mathbf{z} = \sum_{j=1}^p \alpha_j \mathbf{v}_j$ , we conclude

$$\|\mathbf{I}\mathbf{F}(\mathbf{z}; \mathbf{T}_{\omega}, F)\|_F^2 = \sum_{i=1}^p \lambda_i^2 + \sum_{i=1}^p \alpha_i^2 (\lambda_i^4 - 2\lambda_i^3) \leq \sum_{i=1}^p \lambda_i^2 + \max_{i \in \{1, \dots, p\}} (\lambda_i^4 - 2\lambda_i^3),$$

which is reached for the eigenvector corresponding to the eigenvalue which maximises  $(\lambda^4 - 2\lambda^3)$ . Since  $g$  is strictly decreasing on the interval  $[0, 1.5]$  then strictly increasing, the influence function is maximised at either  $\lambda_1$  or  $\lambda_p$ , which completes the proof.  $\square$

**Appendix B. Robust Correlation Estimators**

To make the paper self-contained, we hereby include an overview of the finite-sample definitions of the different correlation estimators we consider.

Let  $\{x_1, \dots, x_n\}$  be a sample, with  $x_i = (x_{i1}, \dots, x_{ip})^\top$ . First, Spearman's rank correlation  $r_S$  is the sample correlation of the ranks of the observations. Second, Kendall's correlation between variable  $j$  and  $k$  is given by

$$r_K = \frac{2}{n(n-1)} \sum_{i < \ell} \text{sign}((x_{ij} - x_{\ell j})(x_{ik} - x_{\ell k})).$$

Third, the Gaussian rank correlation between variable  $j$  and  $k$  is the sample correlation estimated from the Van Der Waerden scores of the data, as given by

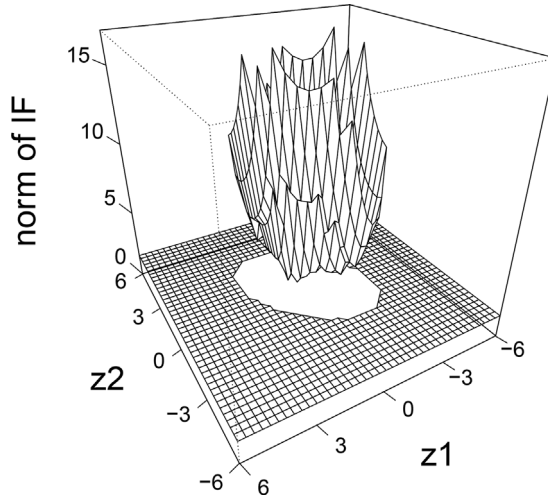
$$r_G = \frac{\sum_{i=1}^n \Phi^{-1}\left(\frac{\text{rank}(x_{ij})}{n+1}\right) \Phi^{-1}\left(\frac{\text{rank}(x_{ik})}{n+1}\right)}{\sum_{i=1}^n \left(\Phi^{-1}\left(\frac{i}{n+1}\right)\right)^2},$$

where  $\text{rank}(x_{ij})$  denotes the rank of  $x_{ij}$  among all components of the  $j$ th variable. Finally, the Quadrant correlation between variable  $j$  and  $k$  is given by

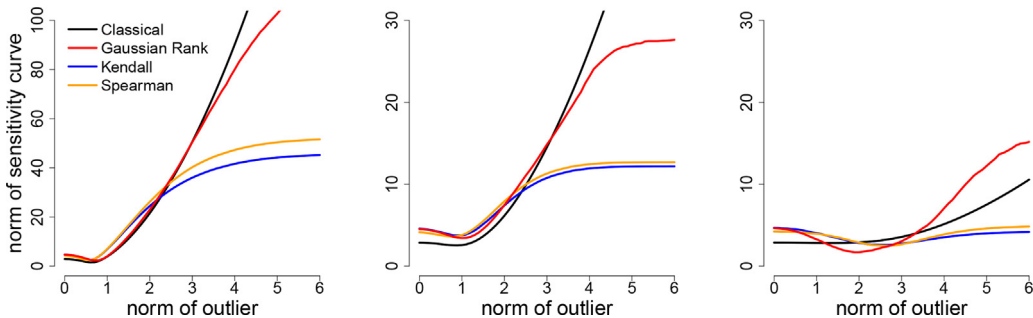
$$r_Q = \frac{1}{n} \sum_{i=1}^n \text{sign}\{(x_{ij} - \text{med}_\ell(x_{\ell j}))(x_{ik} - \text{med}_\ell(x_{\ell k}))\},$$

defined as the difference between, on the one hand, the frequency of the centered observations in the first and third quadrant and, on the other hand, the frequency of the centered observations in the second and fourth quadrant.

**Appendix C. Additional Figures**



**Fig. C1.** Norm of the Glasso influence function based on the reweighted MCD for contamination at  $(z_1, z_2, 0)$ .



**Fig. C2.** Norm of the Glasso sensitivity curve based on several covariance matrix estimators for contamination in the direction of the eigenvector corresponding to the first (left), second (middle) and third (right) eigenvalue of  $T_G$  based on the true covariance matrix.

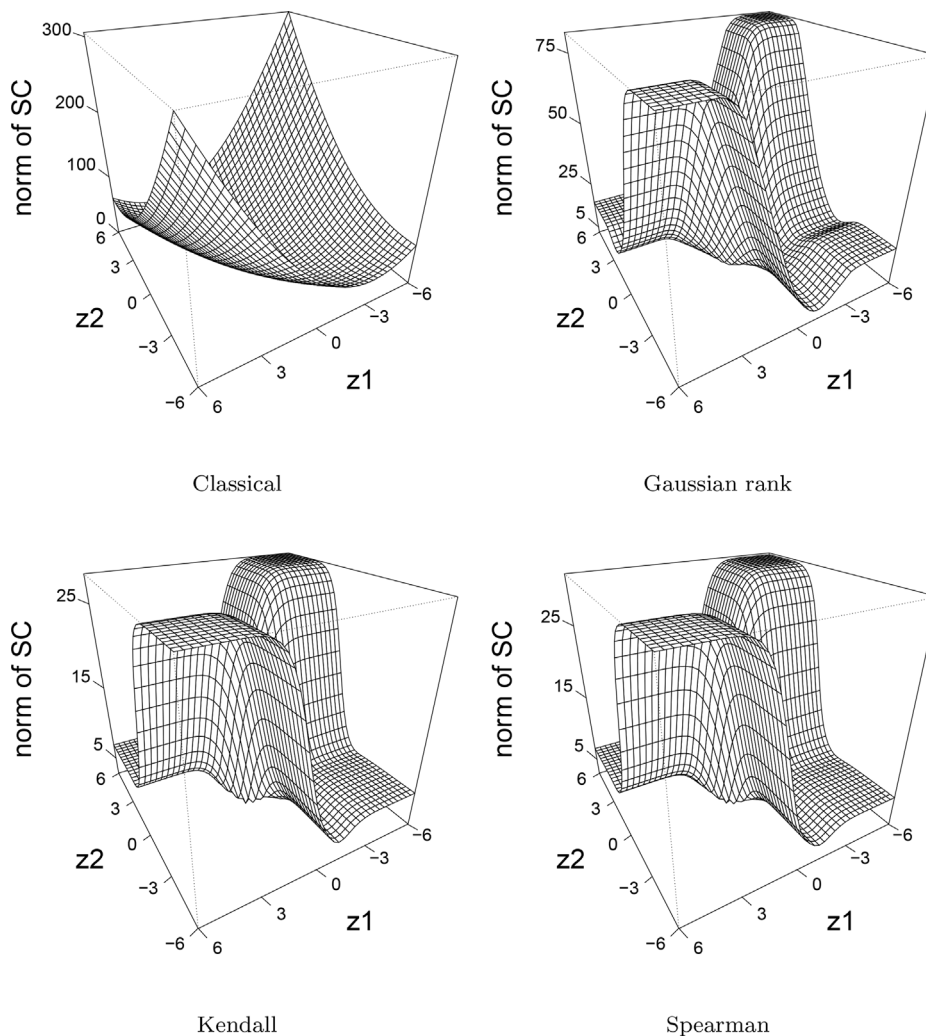


Fig. C3. Norm of the Glasso sensitivity curve based on several covariance matrix estimators for contamination at  $(z_1, z_2, 0)$ .

## References

- Aerts, S., Wilms, I., 2017. Cellwise robust regularized discriminant analysis. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 10 (6), 436–447.
- Alqallaf, F., Van Aelst, S., Yohai, V.J., Zamar, R.H., 2009. Propagation of outliers in multivariate data. *The Annals of Statistics* 37 (1), 311–331.
- Avella-Medina, M., 2016. Robust penalized M-estimators for generalized linear and additive models. Geneva University.
- Avella-Medina, M., 2017. Influence functions for penalized M-estimators. *Bernoulli* 23 (4B), 3178–3196.
- Banerjee, O., El Ghaoui, L., d’Aspremont, A., 2008. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *Journal of Machine Learning Research* 9, 485–516.
- Blomqvist, N., 1950. On a measure of dependence between two random variables. *The Annals of Mathematical Statistics* 21 (4), 593–600.
- Boudt, K., Cornelissen, J., Croux, C., 2012. The Gaussian rank correlation estimator: Robustness properties. *Statistics and Computing* 22 (2), 471–483.
- Cator, E., Lopushaa, H., 2010. Asymptotic expansion of the minimum covariance determinant estimator. *J. Mult. Anal.* 101, 2372–2388.
- Croux, C., Dehon, C., 2010. Influence functions of the Spearman and Kendall correlation measures. *Statistical methods & applications* 19 (4), 497–515.
- Croux, C., Haesbroeck, G., 1999. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis* 71 (2), 161–190.
- Croux, C., Öllerer, V., 2016. Robust and sparse estimation of the inverse covariance matrix using rank correlation measures. In: *Recent Advances in Robust Statistics: Theory and Applications*. Springer, pp. 35–55.
- Fan, J., Han, F., Liu, H., 2014. Challenges of big data analysis. *National Science Review* 1 (2), 293–314.
- Finegold, M., Drton, M., 2011. Robust graphical modeling of gene networks using classical and alternative  $t$ -distributions. *The Annals of Applied Statistics* 5 (2A), 1057–1080.
- Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9 (3), 432–441.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. *Robust statistics: the approach based on influence functions*. John Wiley & Sons.
- He, X., Shao, Q.-M., 1996. A general bahadur representation of  $m$ -estimators and its application to linear regression with nonstochastic designs. *Ann. Stat.* 24 (6), 2608–2630.
- Kendall, M.G., 1938. A new measure of rank correlation. *Biometrika* 30 (1/2), 81–93.
- Kollo, T., von Rosen, D., 2005. *Advanced Multivariate Statistics with Matrices*. Springer.

- Lafit, G., Nogales, F., Ruiz, M., Zamar, R., 2022. Robust graphical lasso based on multivariate winsorization. Working paper 1–33. arXiv:2201.03659
- Loh, P.-L., Tan, X.L., 2018. High-dimensional robust precision matrix estimation: Cellwise corruption under  $\epsilon$ -contamination. *Electronic Journal of Statistics* 12 (1), 1429–1467.
- Öllerer, V., Croux, C., 2015. Robust high-dimensional precision matrix estimation. In: *Modern nonparametric, robust and multivariate methods*. Springer, pp. 325–350.
- Öllerer, V., Croux, C., Alfons, A., 2015. The influence function of penalized regression estimators. *Statistics* 49 (4), 741–765.
- Raymaekers, J., Rousseeuw, P.J., 2021. Fast robust correlation for high-dimensional data. *Technometrics* 63 (2), 184–198.
- Ren, Z., Sun, T., Zhang, C., Zhou, H., 2015. Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics* 43 (3), 991–1026.
- Rothman, A.J., Bickel, P.J., Levina, E., Zhu, J., 2008. Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics* 2, 494–515.
- Rousseeuw, P.J., 1984. Least median of squares regression. *Journal of the American statistical association* 79 (388), 871–880.
- Rousseeuw, P.J., 1985. Multivariate estimation with high breakdown point. *Mathematical statistics and applications* 8 (283–297), 37.
- Rousseeuw, P.J., Croux, C., 1993. Alternatives to the median absolute deviation. *Journal of the American Statistical association* 88 (424), 1273–1283.
- Rousseeuw, P.J., Van Driessen, K., 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41 (3), 212–223.
- Spearman, C., 1904. General intelligence, objectively determined and measured. *The American Journal of Psychology* 15 (2), 201–292.
- Srinivasan, S., Panda, N., 2022. What is the gradient of a scalar function of a symmetric matrix? *Indian Journal of Pure and Applied Mathematics* doi:10.1007/s13226-022-00313-x.
- Tarr, G., Müller, S., Weber, N.C., 2016. Robust estimation of precision matrices under cellwise contamination. *Computational Statistics & Data Analysis* 93, 404–420.
- Tukey, J., 1977. *Exploratory Data Analysis*. Reading (Addison-Wesley), Massachusetts.
- Yuan, M., Lin, Y., 2007. Model selection and estimation in the Gaussian graphical model. *Biometrika* 94 (1), 19–35.