PARTIAL CORRELATION ESTIMATES BASED ON SIGNS

Daniel Vogel, Claudia Köllmann and Roland Fried

Fakultät Statistik, Technische Universität Dortmund,
Vogelpothsweg 87, 44225 Dortmund, GERMANY,
daniel.vogel@udo.edu, claudia.koellmann@udo.edu, fried@statistik.uni-dortmund.de

ABSTRACT

We investigate the Oja sign covariance matrix (Oja SCM) for estimating partial correlations in multivariate data. The Oja SCM estimates directly a multiple of the precision matrix and is based on the concept of Oja signs, which generalise the univariate sign function and obey some form of affine equivariance property. We compare it to the classical MLE as well as to estimates based on two alternative multivariate signs: the marginal sign and the spatial sign.

1. INTRODUCTION: PARTIAL CORRELATION AND THE ELLIPTICAL MODEL

Let $k \geq 3$ and $\boldsymbol{X} = (\boldsymbol{Z}, \boldsymbol{Y})$ with $\boldsymbol{Z} = (Z_1, Z_2)$, $\boldsymbol{Y} = (Y_1, ..., Y_{k-2})$, be a k-dimensional random vector having distribution F and a non-singular covariance matrix Σ. Let furthermore $\hat{Z}_i(\boldsymbol{Y})$, $i = 1, 2$, be the projection of Z_i onto the space of all affine linear functions of $\boldsymbol{Y}$. Then the *partial correlation of* Z_1 *and* Z_2 *given* $\boldsymbol{Y}$ is defined as

$$\varrho_{1,2\bullet Y} = \frac{\operatorname{cov}\Big(Z_1 - \hat{Z}_1(\boldsymbol{Y}),\, Z_2 - \hat{Z}_2(\boldsymbol{Y})\Big)}{\sqrt{\operatorname{var}\Big(Z_1 - \hat{Z}_1(\boldsymbol{Y})\Big) \operatorname{var}\Big(Z_2 - \hat{Z}_2(\boldsymbol{Y})\Big)}},$$

i.e. it is the correlation between the residuals $Z_1 - \hat{Z}_1(\boldsymbol{Y})$ and $Z_2 - \hat{Z}_2(\boldsymbol{Y})$. The partial correlation $\varrho_{1,2\bullet Y}$ can be computed from the covariance matrix Σ of $\boldsymbol{X}$. It holds

$$\varrho_{1,2\bullet Y} = -\frac{k_{1,2}}{\sqrt{k_{1,1}k_{2,2}}},$$

where $k_{i,j}$, $i, j = 1, ..., k$, are the elements of $K = \Sigma^{-1}$, see e.g. [1], p. 143. K is called the *concentration matrix* (or *precision matrix*) of $\boldsymbol{X}$. The matrix

$$C = (K_D)^{-\frac{1}{2}} K (K_D)^{-\frac{1}{2}}, \qquad (1)$$

where K_D denotes the diagonal matrix having the same diagonal as K, equals 1 on the diagonal and contains the negative partial correlations as its off-diagonal elements, i.e. $\varrho_{1,2\bullet Y} = -c_{1,2}$. The correlation matrix R of $\boldsymbol{X}$ can be written as

$$R = (\Sigma_D)^{-\frac{1}{2}} \Sigma (\Sigma_D)^{-\frac{1}{2}}. \qquad (2)$$

One easily checks that

$$C = ((M^{-1})_D)^{-\frac{1}{2}} M^{-1} ((M^{-1})_D)^{-\frac{1}{2}}$$

for any $k \times k$ matrix M that is proportional to Σ or R.

Partial correlations play an important role for instance in graphical modeling, where the key notion is *conditional independence*. Roughly, a *graphical model* is a family of k-dimensional distributions for $\boldsymbol{X} = (X_1, ..., X_k)$ that satisfy some given pairwise conditional independence restrictions on the components of $\boldsymbol{X}$. One can then, based on these pairwise conditional indepence assumptions, draw inferences about conditional independencies between arbitrary disjoint subsets of $\{X_1, ..., X_k\}$. The classical theory of graphical models for continuously distributed variables is built on the normality assumption. If $\boldsymbol{X} = (Z_1, Z_2, \boldsymbol{Y})$ has a multivariate normal distribution, then Z_1 and Z_2 are conditionally independent given $\boldsymbol{Y}$ if and only if $\varrho_{1,2\bullet Y} = 0$, which is equivalent to $k_{1,2} = 0$. A Gaussian graphical model is thus specified by the concentration matrix K.

We consider the problem of estimating partial correlations, but do so in the broader situation of the elliptical model, which is a popular generalisation of the multivariate normal model. Its first and second order characteristics still provide an intuitive description of the geometry of the distribution, and it is mathematically tractable. In addition it allows to model different tail behaviours.

The density f_0 of a *spherical distribution* F_0 on $\mathbb{R}^k$ is of the form $f_0(\boldsymbol{x}) = g(\boldsymbol{x}^T\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^k$, where $g : [0,\infty) \to [0,\infty)$ is such that f_0 integrates to 1. If furthermore

$$\operatorname{med}|X_1| = u_{.75}, \qquad (3)$$

where X_1 is the first component of $\boldsymbol{X} \sim F_0$ and $u_{.75}$ the 75% quantile of the standard normal distribution, we call F_0 a *standardized spherical distribution*. In the following we assume that $\boldsymbol{X}_0 \sim F_0$, where F_0 is a standardized spherical distribution admitting the Lebesgue-density f_0. A random vector $\boldsymbol{X}$ has an elliptical distribution F if

$$\boldsymbol{X} \overset{\mathscr{L}}{=} S^{\frac{1}{2}}\boldsymbol{X}_0 + \boldsymbol{b}$$

for some $\boldsymbol{b} \in \mathbb{R}^k$ and symmetric, positive definite $k \times k$ matrix S. Then its density is

$$f(\boldsymbol{x}) = \det(S)^{-\frac{1}{2}} g\big((\boldsymbol{x}-\boldsymbol{b})^T S^{-1}(\boldsymbol{x}-\boldsymbol{b})\big). \qquad (4)$$

We use the standardisation assumption (3) in order to fix S and g in (4) without requiring the existence of any moments of F. It is a major advantage of sign methods that they usually work without any moment assumptions. The existence of partial correlations, of course, necessitates the existence of second moments. If expectation and variance of $\boldsymbol{X}$ exist, then $\mathbb{E}(\boldsymbol{X}) = \boldsymbol{b}$ and $\operatorname{Var}(\boldsymbol{X}) = \Sigma(F)$ is proportional to S. If F is normal, then $\Sigma(F) = S$. We call $\boldsymbol{b}$ the *symmetry center* and S the *shape matrix* of F, and – following [2] – denote the class of all elliptical distributions on $\mathbb{R}^k$ having these parameters by $E_k(\boldsymbol{b}, S)$. By choosing the function g we model the tail behaviour of the distribution F. The normal distribution $N_k(\boldsymbol{b}, \Sigma)$ corresponds to $g_{N_k}(y) = (2\pi)^{-\frac{k}{2}} \exp\big(-\frac{1}{2}y\big)$. Another important subclass of elliptical distributions is the multivariate $t_{\nu,k}$-family with

$$g_{t_{\nu,k}}(y) = c_\nu \frac{\Gamma(\frac{\nu+k}{2})}{(\nu\pi)^{\frac{k}{2}}\Gamma(\frac{\nu}{2})}(1 - \frac{c_\nu^2 y}{\nu})^{-\frac{\nu+k}{2}}.$$

Here the first subscript ν denotes the degrees of freedom. The constant $c_\nu = t_{\nu;.75}/u_{.75}$ is due to the standardization (3), $t_{\nu;.75}$ being the 75% quantile of the usual, univariate t_ν-distribution with ν degrees of freedom. The $t_{\nu,k}(\boldsymbol{b}, S)$ distribution converges to $N_k(\boldsymbol{b}, S)$ as $\nu \to \infty$ and is, for small ν, a popular example of a heavy-tailed distribution. Its moments are finite only up to order $(\nu - 1)$.

It is considered a shortcoming of the elliptical model that it does not include independent margins, unless the margins are normal, cf. e.g. [2], p. 51. Consequently, partial uncorrelatedness (i.e. an off-diagonal zero entry in the precision matrix K) does in general not mean conditional independence. It is, however, equivalent to *conditional uncorrelatedness*, cf. [3]. Thus partial correlation is a measure of conditional linear dependence.

2. MULTIVARIATE SIGNS

A common approach in nonparametric statistics is to replace the observations by their signs or ranks. This means in general loosing efficiency under normality, but one can hope to get robust and distribution-free methods. For reasons of simpliciy we only consider signs here. Since we analyse multivariate data we are interested in multivariate signs. There are several possible generalisations of the univariate notion *sign* to the multivariate setting, three of which we want to name here: the marginal sign, the spatial sign and the Oja sign. We start by recalling the usual, univariate sign function. Suppose we have a univariate data set $\mathbb{X} = (x_1, ..., x_n)$, $n \in \mathbb{N}$. We call $\operatorname{sgn}_{\mathbb{X}}(x) = \operatorname{sgn}(x - \operatorname{med}(\mathbb{X}))$, $x \in \mathbb{R}$, the *sign of x w.r.t. the data sample* $\mathbb{X}$, where sgn is the univariate sign function ($\operatorname{sgn}(x) = \frac{x}{|x|}$ if $x \neq 0$ and zero otherwise), and $\operatorname{med}(\mathbb{X})$ is the univariate median of $\mathbb{X}$. One obvious extension of this concept to multivariate data is the component-wise application of the univariate sign, leading to the *marginal sign*. We call

$$\mathbf{msgn}_{\mathbb{X}}(\boldsymbol{x}) = \mathbf{msgn}(\boldsymbol{x} - \mathbf{mmed}(\mathbb{X})),$$

the *marginal sign of $\boldsymbol{x} \in \mathbb{R}^k$ w.r.t. the k-variate data sample* $\mathbb{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$, $n \in \mathbb{N}$, where $\mathbf{mmed}(\mathbb{X})$ is the component-wise, *marginal median of* $\mathbb{X}$. Another fairly straightforward multivariate generalisation is obtained from the *spatial sign* function

$$\mathbf{ssgn}(\boldsymbol{x}) = \begin{cases} \frac{1}{||\boldsymbol{x}||}\,\boldsymbol{x} & \text{if } \boldsymbol{x} \neq \mathbf{0}, \\ \mathbf{0} & \text{if } \boldsymbol{x} = \mathbf{0}. \end{cases}$$

The spatial median $\mathbf{smed}(\mathbb{X})$ is the gravity point of $\underset{x \in \mathbb{R}^k}{\arg\min} \big|\big| \sum_{i=1}^{k} \mathbf{ssgn}(\boldsymbol{x}_i - \boldsymbol{x})\big|\big|$, and as before

$$\mathbf{ssgn}_{\mathbb{X}}(\boldsymbol{x}) = \mathbf{ssgn}(\boldsymbol{x} - \mathbf{smed}(\mathbb{X})), \quad \boldsymbol{x} \in \mathbb{R}^k,$$

is the *spatial sign of $\boldsymbol{x}$ w.r.t.* $\mathbb{X}$. A third possible multivariate extension is the Oja sign. For $0 \leq p \leq n$ let

$$Q_p = \big\{q = \{i_1, ..., i_p\} \,\big|\, 1 \leq i_1 < ... < i_p \leq n\big\}$$

be the system of all subsets of $\{1, ..., n\}$ of size p and $N_p = |\,Q_p| = \binom{n}{p}$. Then the *Oja median*

$\mathbf{omed}(\mathbb{X})$ *of the data sample* $\mathbb{X}$ is defined as the gravity point of

$$\underset{x\in\mathbb{R}^k}{\arg\min}\sum_{Q_k}\left|\det\begin{pmatrix}1 & \dots & 1 & 1\\ \boldsymbol{x}_{i_1} & \dots & \boldsymbol{x}_{i_k} & \boldsymbol{x}\end{pmatrix}\right|. \quad (5)$$

The *Oja sign of the point* $\boldsymbol{x}\in\mathbb{R}^k$ *w.r.t.* $\mathbb{X}$ is

$$\mathbf{osgn}_{\mathbb{X}}(\boldsymbol{x}) = \frac{1}{N_{k-1}}\sum_{Q_{k-1}}\nabla\boldsymbol{x}\left|\det(\boldsymbol{y}_{i_1}\ \dots\ \boldsymbol{y}_{i_{k-1}}\ \boldsymbol{y})\right|,$$

where $\boldsymbol{y} = \boldsymbol{x}-\mathbf{omed}(\mathbb{X})$ and $\boldsymbol{y}_i = \boldsymbol{x}_i-\mathbf{omed}(\mathbb{X})$, $i = 1,...,n$. Note that contrary to $\mathbf{msgn}_{\mathbb{X}}$ and $\mathbf{ssgn}_{\mathbb{X}}$ the Oja sign $\mathbf{osgn}_{\mathbb{X}}$ does *not only* depend upon the data sample $\mathbb{X}$ through its center point $\mathbf{omed}(\mathbb{X})$. Note that for $k = 1$ expression (5) comes down to

$$\underset{x\in\mathbb{R}}{\arg\min}\sum_{1}^{n}\left|\det\begin{pmatrix}1 & 1\\ x_i & x\end{pmatrix}\right|$$

and

$$\mathbf{osgn}_{\mathbb{X}}(x) = \frac{\partial}{\partial x}\left|\det(x-\mathbf{omed}(\mathbb{X}))\right|.$$

Thus the Oja median and the Oja sign are indeed proper multivariate generalisations of med and $\mathrm{sgn}_{\mathbb{X}}$.

It should be noted that these three multivariate signs have different invariance properties. All of them are invariant w.r.t. translations. The marginal sign is also invariant w.r.t. re-scaling. The spatial sign on the other hand is equivariant under orthogonal transformations, i.e. if we let $A\mathbb{X} = (A\boldsymbol{x}_1,...,A\boldsymbol{x}_n)$ for some orthogonal matrix A, then

$$\mathbf{ssgn}_{A\mathbb{X}}(A\boldsymbol{x}) = A\,\mathbf{ssgn}_{\mathbb{X}}(\boldsymbol{x}).$$

The Oja sign even obeys some form of affine linear equivariance:

$$\mathbf{osgn}_{A\mathbb{X}}(A\boldsymbol{x}) = \det(A)A^{-1}\mathbf{osgn}_{\mathbb{X}}(\boldsymbol{x})$$

for any full rank $k\times k$ matrix A, cf. e.g. [4] or [5], p. 330.

3. SIGN COVARIANCE MATRICES

Now we construct scatter estimates based on the multivariate signs introduced in the previous section: the *marginal sign covariance matrix (MSCM)*, the *spatial sign covariance matrix (SSCM)* and the *Oja sign covariance matrix (OSCM)*. All of these, along with some basic properties, can be found in [6]. We start with the most frequently used estimate of scatter, the *empirical covariance matrix (ECM)*

$$\mathrm{ECM}(\mathbb{X}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}}_n)(\boldsymbol{x}_i-\overline{\boldsymbol{x}}_n)^T,$$

where $\overline{\boldsymbol{x}}_n$ is the mean of $\boldsymbol{x}_1,...,\boldsymbol{x}_n$. The ECM is the (biased) MLE of the covariance matrix Σ, provided the latter exists. The sign covariance matrices follow the same construction principle as the ECM, with $\boldsymbol{x}_i-\overline{\boldsymbol{x}}_n$ being replaced by the respective signs:

$$\mathrm{MSCM}(\mathbb{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{msgn}_{\mathbb{X}}(\boldsymbol{x}_i)\mathbf{msgn}_{\mathbb{X}}(\boldsymbol{x}_i)^T,$$

$$\mathrm{SSCM}(\mathbb{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{ssgn}_{\mathbb{X}}(\boldsymbol{x}_i)\mathbf{ssgn}_{\mathbb{X}}(\boldsymbol{x}_i)^T,$$

$$\mathrm{OSCM}(\mathbb{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{osgn}_{\mathbb{X}}(\boldsymbol{x}_i)\mathbf{osgn}_{\mathbb{X}}(\boldsymbol{x}_i)^T.$$

The next lemma tells what these estimators estimate in the elliptical model. We understand the *theoretical counterpart* $\Sigma^m(F)$ of the MSCM as the functional

$$\mathbb{E}\Big[\mathbf{msgn}(\boldsymbol{X}-\mathbf{mmed}(F))\,\mathbf{msgn}(\boldsymbol{X}-\mathbf{mmed}(F))^T\Big]$$

with $\boldsymbol{X}\sim F$. If F is the empirical distribution function generated by the data $\mathbb{X}$, then $\Sigma^m(F) = MSCM(\mathbb{X})$. Similarly we define the theoretical counterparts of SSCM and OSCM, the latter is also explicitely stated in [4].

Lemma 3.1 *Let* $\boldsymbol{X}\sim F$ *and* $\boldsymbol{X}_0\sim F_0$ *with* $F\in E_k(\mathbf{0},S)$ *and* F_0 *the corresponding standardized spherical distribution. The theoretical counterparts of the MSCM, SSCM and OSCM at* F*, denoted by* $\Sigma^m(F)$, $\Sigma^s(F)$ *and* $\Sigma^O(F)$*, respectively, are given by:*

(I) $\sigma^m_{i,j}(F) = \frac{2}{\pi}\arcsin(\varrho_{i,j})$,
where $\varrho_{i,j}$, $i,j = 1,...,k$, *are the elements of* $(S_D)^{-\frac{1}{2}}S(S_D)^{-\frac{1}{2}}$, *which equals the correlation matrix* R, *provided it exists.*

(II) $$\Sigma^s(F) = \mathbb{E}\left(\frac{S^{\frac{1}{2}}\boldsymbol{X}\boldsymbol{X}^TS^{\frac{1}{2}}}{\boldsymbol{X}^TS\boldsymbol{X}}\right).$$

(III) $\Sigma^O(F) = \gamma_{F_0}\det(S)S^{-1}$,
if $\mathbb{E}||\boldsymbol{X}_0|| < \infty$. *The constant* γ_{F_0} *depends only on* $\mathbb{E}||\boldsymbol{X}_0||$ *and the dimension* k.

Parts (I) and (II) are straightforward, the proof of (III) is carried out in [4], where the constant γ_{F_0} is also made explicit. Ollila et al. [4] show furthermore that, if the second order moments of F exist, OSCM converges in probability to $\Sigma^O(F)$ and is asymptotically normal. It is intuitive that similar convergence results hold for MSCM and SSCM without any moment condition on F.

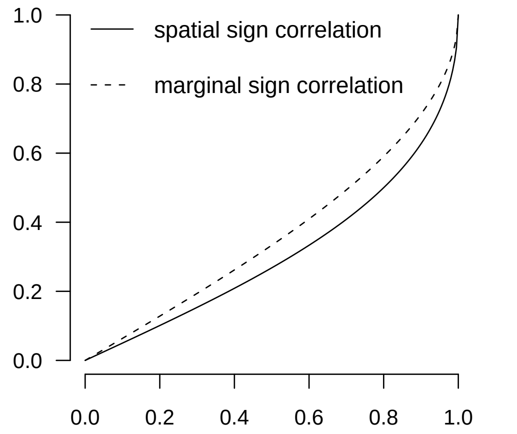

Figure 1. Functions σ^s (solid) and σ^m (dashed), defined in (6) and (7), respectively.

Unfortunately there is not such a simple relation between Σ^s and S, as there is in (I) between Σ^m and R. For example, in the very simple case of the 2×2 shape matrix

$$S = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \qquad -1 \leq \rho \leq 1,$$

we have

$$\Sigma^s = \frac{1}{2} \begin{pmatrix} 1 & \sigma^s(\rho) \\ \sigma^s(\rho) & 1 \end{pmatrix} \tag{6}$$

and

$$\Sigma^m = \begin{pmatrix} 1 & \sigma^m(\rho) \\ \sigma^m(\rho) & 1 \end{pmatrix}, \tag{7}$$

where $\sigma^m(\rho) = \frac{2}{\pi} \arcsin(\rho)$ and

$$\sigma^s(\rho) = \frac{2\sqrt{1+\rho}}{\sqrt{1+\rho} + \sqrt{1-\rho}} - 1. \tag{8}$$

Thus Figure 1 shows the relation of marginal-sign-correlation (also known as quadrant correlation) and spatial-sign-correlation to the usual Pearson-correlation at a two-dimensional standardized elliptical distribution. Theorem 1 in [7] sheds some light on the structure of Σ^s in general. There is always a one-to-one connection between Σ^s and S and both matrices share the same eigenvectors.

4. PARTIAL CORRELATION ESTIMATORS

For notational convenience we define

$$\begin{aligned} \hat{K}^e &= \text{ECM}^{-1},\ \hat{K}^O = \text{OSCM} \ \text{ and} \\ \hat{K}^m &= \big(h(\text{MSCM})\big)^{-1}, \end{aligned}$$

where the mapping h is the element-wise application of $x \mapsto \sin(\frac{\pi}{2}x)$. We call $h(\text{MSCM})$ the *modified MSCM*. From Lemma 1 we know that (if the covariance exists) $\hat{K}^e$ and $\hat{K}^O$ estimate the concentration matrix K, respectively a multiple of it, and $\hat{K}^m$ the inverse of R. From what has been said in Section 1 we can thus construct estimators of the matrix C:

$$\begin{aligned} \hat{C}^e &= \big(\hat{K}^e_D\big)^{-\frac{1}{2}} \hat{K}^e \big(\hat{K}^e_D\big)^{-\frac{1}{2}}, \\ \hat{C}^O &= \big(\hat{K}^O_D\big)^{-\frac{1}{2}} \hat{K}^O \big(\hat{K}^O_D\big)^{-\frac{1}{2}}, \\ \hat{C}^m &= \big(\hat{K}^m_D\big)^{-\frac{1}{2}} \hat{K}^m \big(\hat{K}^m_D\big)^{-\frac{1}{2}}. \end{aligned}$$

$\hat{C}^e$ is the normal MLE of C, see e.g. [8]. $\hat{C}^m$ as above is not well defined. It may happen – especially for small n – that M^m is not positive definite. The common structure of ECM and the sign covariance matrices guarantees that these matrices are always positive semi-definite, and – as long as $k < n$ and the underlying distribution F has a Lebesgue-density – ECM, OSCM and SSCM are positive definite with probability 1. This is not true for the MSCM. The additional modification step h may furthermore lead to negative eigenvalues of M^m. A remedy could be to perform an eigenvalue decomposition and set the non-positive eigenvalues to small positive values. Such a manipulation does not affect the asymptotics. We carried out a simulation study using several elliptical distributions to examine the finite-sample performance of the proposed estimators. In the examples that follow we fix the mean to **0** and the shape matrix to

$$S = \begin{pmatrix} 1 & -0.865 & 0.657 & -0.231 \\ -0.865 & 1 & -0.510 & 0.077 \\ 0.657 & -0.510 & 1 & -0.601 \\ -0.231 & 0.077 & -0.601 & 1 \end{pmatrix},$$

which corresponds to the following matrix of partial correlations

$$-C = \begin{pmatrix} -1 & -0.8 & 0.4 & 0 \\ -0.8 & -1 & 0 & -0.2 \\ 0.4 & 0 & -1 & -0.6 \\ 0 & -0.2 & -0.6 & -1 \end{pmatrix}.$$

Figure 2 shows the estimated densities of $-\hat{c}^e_{1,4}$ $-\hat{c}^O_{1,4}$ and $-\hat{c}^m_{1,4}$ (left plot) and $-\hat{c}^e_{1,3}$, $-\hat{c}^O_{1,3}$ and $-\hat{c}^m_{1,3}$ (right plot) calculated from 30 observations drawn from a normal distribution with covariance $\Sigma = S$ as above. The true values to be estimated, $\varrho_{1,4\bullet 2,3} = -c_{1,4} = 0$ and $\varrho_{1,3\bullet 2,4} = -c_{1,3} = 0.4$, respectively, are indicated by vertical lines. The

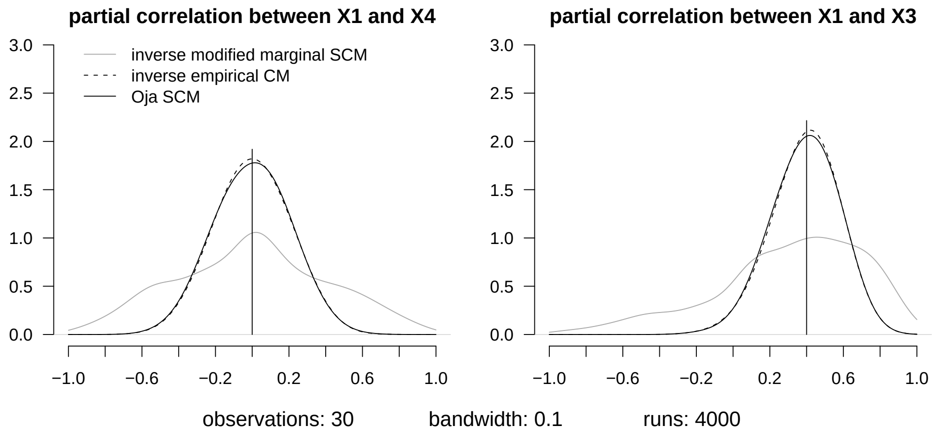

Figure 2. Densities of three partial correlation estimators at the multivariate normal distribution

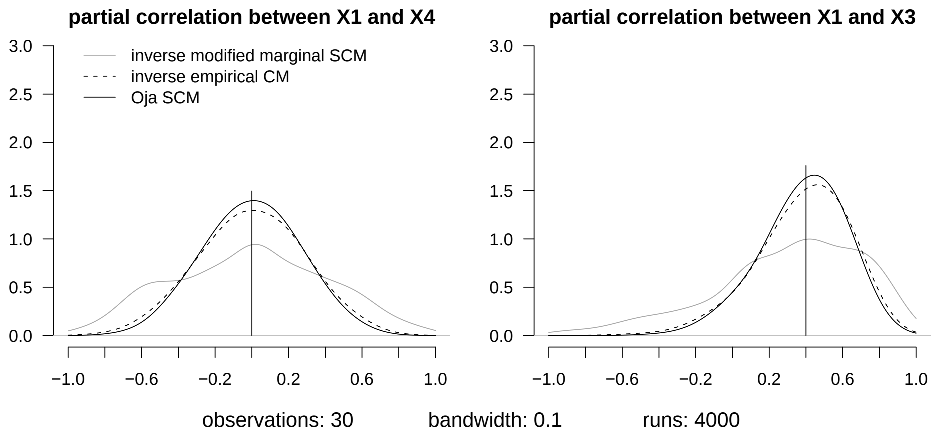

Figure 3. Densities of three partial correlation estimators at the t_3-distribution

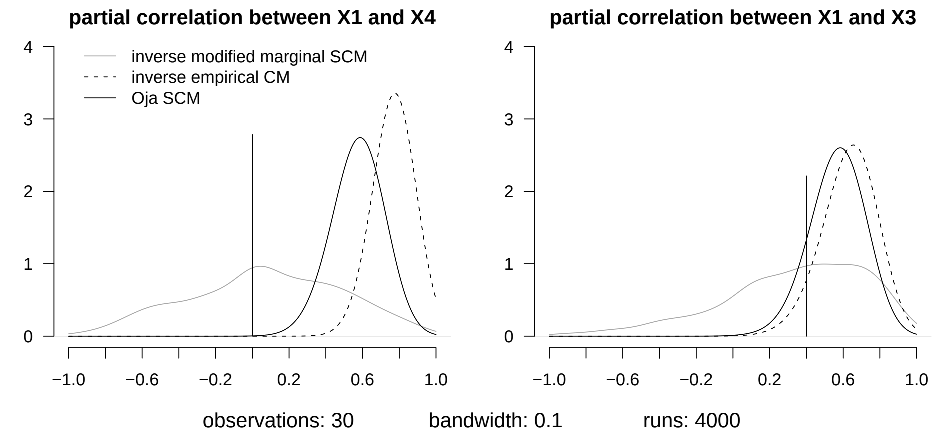

Figure 4. Densities of partial correlation estimators under outlier scenario

density estimation is based on 4000 repititions, using the R function `density()` with a Gauss kernel and bandwidth .1. There does not seem to be any relevant difference between $-\hat{c}^e_{i,j}$ and $-\hat{c}^O_{i,j}$. In fact, the asymptotic relative efficiency of $\hat{c}^O_{i,j}$ at the normal model (compared to the MLE $\hat{c}^E_{i,j}$) is more than 98%, cf. [9].

In Figure 3 we see the results of an experiment with the same parameters except that the population distribution is now $t_{3,4}(\mathbf{0}, S)$. We find that $-\hat{c}^e_{i,j}$ and $-\hat{c}^O_{i,j}$ have a higher variability (compared to the normal model), but the Oja SCM estimator $\hat{c}^O_{i,j}$ performs substantially better than the MLE $\hat{c}^e_{i,j}$. The marginal SCM estimator $\hat{c}^m_{i,j}$ is distribution-free w.r.t. g. It should be mentioned, though, that its high variability is to a large portion due to the modification by applying h.

We also examined the behaviour of the estimators under outlier scenarios. Figure 4 shows the effect of a systematic outlier. We sampled again from the multivariate normal distribution (with $S = \Sigma$ as above), but added each time $(6, 0, 0, 6)$ to the first observation. The direction of this contamination was particularly aimed at destroying the partial uncorrelatedness of the variables X_1 and X_4, suggesting instead a strong positive partial dependence. $\hat{C}^m$ is little affected by the outlier. On the other hand $\hat{C}^e$ and $\hat{C}^O$ can both be made to break down by one single outlying observation, but we also find that the impact is quantitatively smaller on $\hat{C}^O$ than on $\hat{C}^e$. These findings are in agreement with the structure of the respective influence functions. At a standardized spherical distribution F_0 the influence function $IF(\boldsymbol{x}, \hat{C}^O, F_0)$ of $\hat{C}^O$ equals

$$k\Big(1 - \frac{2||\boldsymbol{x}||}{\mathbb{E}||\boldsymbol{X}_0||}\Big)\big(\boldsymbol{u}\boldsymbol{u}^T - (\boldsymbol{u}\boldsymbol{u}^T)_D\big),$$

where $\boldsymbol{u} = \frac{\boldsymbol{x}}{||\boldsymbol{x}||}$ and $\boldsymbol{X}_0 \sim F_0$, cf. [9]. For any fixed direction $\boldsymbol{u}$ this is an affine linear function of the distance $||\boldsymbol{x}||$, whereas the influence functions of $\hat{C}^e$ and $\hat{C}^m$ are quadratic, respectively constant, in $||\boldsymbol{x}||$.

5. CONCLUSION

The Oja SCM is well suited to the task of estimating partial correlations in elliptical models, better than the related concepts MSCM and SSCM, since – contrary to these – it retains the whole shape information. It almost equals the efficiency of the MLE $\hat{C}^e$ in the Gaussian case, but behaves qualitatively better under model misspecifications. The loss of efficiency under heavy-tailed distributions is considerably smaller, and the same is true for the impact of outliers. We can recommend the Oja SCM as an estimator for partial correlations in graphical models, but – and this is the main drawback – only for data sets of moderate size. The reason is that its computation necessitates the evaluation of $\binom{n}{k-1}$ $(k-1)$-dimensional hyperplanes.

6. REFERENCES

[1] Joe Whittaker, *Graphical models in applied multivariate statistics.*, Wiley Series in Probability and Mathematical Statistics. Chichester etc.: John Wiley & Sons Ltd., 1990.

[2] Martin Bilodeau and David Brenner, *Theory of multivariate statistics.*, Springer Texts in Statistics. New York, NY: Springer, 1999.

[3] Kunihiro Baba, Ritei Shibata, and Masaaki Sibuya, "Partial correlation and conditional correlation as measures of conditional independence.," *Aust. N. Z. J. Stat.*, vol. 46, no. 4, pp. 657–664, 2004.

[4] Esa Ollila, Hannu Oja, and Christophe Croux, "The affine equivariant sign covariance matrix: Asymptotic behavior and efficiencies.," *J. Multivariate Anal.*, vol. 87, no. 2, pp. 328–355, 2003.

[5] Thomas Hettmansperger and Joseph McKean, *Robust Nonparametric Statistical methods.*, Arnold, London, 1998.

[6] Samuli Visuri, Visa Koivunen, and Hannu Oja, "Sign and rank covariance matrices.," *J. Stat. Plann. Inference*, vol. 91, no. 2, pp. 557–575, 2000.

[7] Samuli Visuri, "Array and multichannel signal processing using nonparamtric statistics.," M.S. thesis, Helsinki University of Technology, Helsinki, Finland, 2001.

[8] Steffen L. Lauritzen, *Graphical models.*, Oxford Statistical Science Series. 17. Oxford: Oxford Univ. Press, 1996.

[9] Daniel Vogel and Roland Fried, "Estimating partial correlations using the Oja sign covariance matrix.," *Proceedings of the International conference on Computational Statistics, Porto, Portugal, August 24–29, 2008*, in press.