Elliptical graphical modeling in higher dimensions

D. Vogel, A. Dürre and R. Fried

***Abstract*—Simpson's famous paradox vividly exemplifies the importance of considering conditional, rather than marginal, associations for assessing the dependence structure of several variables. The study of conditional dependencies is the subject matter of graphical models. The statistical methods applied in graphical models for continuous variables rely on the assumption of normality, which leads to the term *Gaussian graphical models*. We consider *elliptical graphical models*, that is, we allow the population distribution to be elliptical instead of normal. We examine the class of affine equivariant scatter estimators and propose an adjusted version of the deviance tests, valid under ellipticity. A detailed derivation can be found in [1]. In this exposition we report the results of a simulation study, demonstrating the feasibility of our approach also in higher dimensions. Graphical models based on classical, non-robust estimators have been used, e.g., to explore successfully the partial correlation structure within high-dimensional physiological time series [2] and within high-dimensional time series describing neural oscillators [3].**

***Index Terms*—partial correlation, deviance test,**

I. Graphical models

We first introduce the basic terms and notions. Let $p \geq 3$ and $\boldsymbol{X} = (X_1, X_2, \boldsymbol{Y})$ with $\boldsymbol{Y} = (X_3, ..., X_p)$ be a p-dimensional random vector following some distribution F with non-singular covariance matrix Σ. Let $\hat{X}_i(\boldsymbol{Y})$, $i = 1, 2$, be the projection of X_i onto the space of all affine linear functions of $\boldsymbol{Y}$. Then the *partial correlation* $p_{1,2}$ *of* X_1 *and* X_2 *given* $X_3, ..., X_p$ is defined as the correlation between the residuals $X_1 - \hat{X}_1(\boldsymbol{Y})$ and $X_2 - \hat{X}_2(\boldsymbol{Y})$. The partial correlation $p_{1,2}$ can be interpreted as a measure of the *linear* association between X_1 and X_2 after the common *linear* effects of all other variables have been removed. It is a moment-based characteristic of the distribution F and can be computed from the covariance matrix Σ. It holds

$$p_{1,2} = -\frac{k_{1,2}}{\sqrt{k_{1,1}k_{2,2}}},$$

where $k_{i,j}$, $i, j = 1, ..., k$, are the elements of $K = \Sigma^{-1}$, see e.g. [4]. The matrix K is called the *concentration matrix* (or *precision matrix*) of $\boldsymbol{X}$.

The partial correlation structure of the random variable $\boldsymbol{X}$ can be coded in a graph, which originates the term *graphical model*. An undirected graph $G = (V, E)$, where V is the vertex set and E the edge set, is constructed the following way: the variables $X_1, ..., X_p$ are the vertices, and an (undirected) edge is drawn between X_i and X_j, $i \neq j$, if and only if $p_{i,j} \neq 0$. The thus obtained graph G is called the *partial correlation graph (PCG)* of $\boldsymbol{X}$. Formally we set $V = \{1, ..., p\}$ and write the elements of E as unordered pairs $\{i, j\}$, $1 \leq i < j \leq p$. The partial correlation graph is a useful data analytical tool. It concisely displays the important aspects of the interrelations of several variables. It allows furthermore to draw conclusions about the dependence between groups of variables (note that the graph is constructed from pairwise relations between individual variables) and facilitates the understanding of the underlying physiological process. We will not dwell further on the purpose and the properties of the PCG, the reader may consult the review article [5] or any of the classical textbooks [4], [6], [7], [8]. Our concern here is the statistical modeling.

The authors are with the Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany. E-mail: daniel.vogel@tu-dortmund.de.

II. Gaussian graphical models

Suppose we have a data set $\mathbb{X}_n = (\boldsymbol{x}_1^T, ..., \boldsymbol{x}_n^T)^T$ of n i.i.d. realizations of the p-dimensional random vector $\boldsymbol{X}$. In order to sensibly "estimate" the PCG of $\boldsymbol{X}$ from the data, we have to make some distributional assumption about $\boldsymbol{X}$. This assumption is usually multivariate normality, i.e. $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ for some $\boldsymbol{\mu} \in \mathbb{R}^p$ and positive definite matrix $\Sigma \in \mathbb{R}^{p \times p}$. Then the *Gaussian graphical model* $\mathscr{M}(G)$ induced by the undirected graph $G = (V, E)$ is the set of all p-dimensional Gaussian distributions satisfying the zero partial correlation restrictions specified by G. Precisely, if we denote the set of all positive definite $p \times p$ matrices by $\mathscr{S}_p^+$ and let

$$\mathscr{S}_p^+(G) = \left\{ K \in \mathscr{S}_p^+ \,\middle|\, k_{i,j} = 0 \,\forall\, i \neq j \text{ with } \{i, j\} \notin E \right\},$$

then

$$\mathscr{M}(G) = \left\{ N_p(\boldsymbol{\mu}, \Sigma) \,\middle|\, \boldsymbol{\mu} \in \mathbb{R}^p,\ K = \Sigma^{-1} \in \mathscr{S}_p^+(G) \right\}.$$

An integral part of almost any model selection scheme is the possibility to test if a model under consideration fits the data or not. In the context of Gaussian graphical models the classical tool for this purpose is the *deviance test*, which is described in the following. For any graph $G = (V, E)$ define the function $h_G : \mathscr{S}_p^+ \to \mathscr{S}_p^+ : A \mapsto A_G$ by

$$\begin{cases} [A_G]_{i,j} = a_{i,j}, & \{i, j\} \in E \quad \text{or} \quad i = j, \\ [A_G^{-1}]_{i,j} = 0, & \{i, j\} \notin E \quad \text{and} \quad i \neq j. \end{cases} \tag{1}$$

It is not trivial and a deeper result of the theory of Gaussian graphical models that a unique and positive definite solution A_G of (1) exists for any positive definite A. The solution can be found by the *iterative proportional scaling* algorithm, for which convergence has been shown, cf. [7], Chap. 5. If we let further $\hat{\Sigma}_n$ denote the sample covariance matrix, then $\hat{\Sigma}_G = h_G(\hat{\Sigma}_n)$ is a sensible estimator for Σ subject to the assumption $\Sigma^{-1} \in \mathscr{S}_p^+(G)$. It is indeed the maximum likelihood estimator in the Gaussian graphical model $\mathscr{M}(G)$. Now suppose we have two nested models $\mathscr{M}(G_0) \subsetneq \mathscr{M}(G_1)$, i.e. the edge set E_0 of G_0 is a strict subset of the edge set E_1 of G_1. Let q be the number of edges that are in E_1 but not E_0. Then, under $\mathscr{M}(G_0)$,

$$\hat{D}_n(\hat{\Sigma}_n) = n \left(\ln\det h_{G_0}(\hat{\Sigma}_n) - \ln\det h_{G_1}(\hat{\Sigma}_n) \right) \tag{2}$$

converges to a χ^2 distribution with q degrees of freedom. This the likelihood ratio test for testing $\mathscr{M}(G_0)$ against the larger model $\mathscr{M}(G_1)$. The statistic $\hat{D}_n(\hat{\Sigma}_n)$ is also referred to as *deviance*. Many model selection procedures (backward elimination, forward selection, Edwards-Havránek,...) consist of an iterative application of this test. For details see, e.g., [8].

III. Elliptical graphical models

A problem of the Gaussian graphical modeling described in the previous section is its lack of robustness, which is mainly due to the poor robustness of the estimator $\hat{\Sigma}_n$. Hence a promising way of robustifying the procedure is to replace $\hat{\Sigma}_n$ by a more robust scatter estimator. Over the last four decades many proposals of robust multivariate dispersion estimators have been made, for a review see, e.g., [9]. Indeed, it can be shown that the convergence of (2) remains true, if $\hat{\Sigma}_n$ is replaced by any scatter estimator $\hat{S}_n$ that fulfills the following regularity conditions.

(I) $\hat{S}_n$ is (at least proportionally) affine equivariant, i.e. $\hat{S}_n(\mathbb{X}_n A^T + \boldsymbol{b}) \propto A\hat{S}_n(\mathbb{X}_n)A^T$ for any $\boldsymbol{b} \in \mathbb{R}^p$ and full rank matrix $A \in \mathbb{R}^{p\times p}$, and

(II) $\hat{S}_n$ is $\sqrt{n}$*-convergent*, i.e. $\sqrt{n}(\hat{S}_n - S)$ converges in distribution, where S is some multiple of Σ.

These two conditions are very natural conditions for multivariate scatter estimators. For details see [1]. Then, under $\mathscr{M}(G_0)$,

$$\tfrac{1}{\sigma_1}\hat{D}_n(\hat{S}_n) = \tfrac{n}{\sigma_1}\Big(\ln\det h_{G_0}(\hat{S}_n) - \ln\det h_{G_1}(\hat{S}_n)\Big) \qquad (3)$$

converges to a χ^2 distribution with q degrees of freedom, where $\sigma_1 > 0$ is a suitable scalar-valued constant, which is a function of the estimator $\hat{S}_n$, but does neither depend on G_0, G_1 nor the true covariance Σ. We call $\hat{D}_n(\hat{S}_n)$ *pseudo-deviance*, and the corresponding test *adjusted deviance test*, since we have to divide the test statistic by the consistency factor σ_1.

Furthermore, $\frac{1}{\sigma_1}\hat{D}_n(\hat{S}_n)$ also converges to a χ^2_q limit, if the data $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are sampled from an elliptical distribution. Then σ_1 has to be chosen accordingly, examples are given in the next section. A continuous distribution F in $\mathbb{R}^p$ is said to be *elliptical* if it has a density f of the form

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

for some $\boldsymbol{\mu} \in \mathbb{R}^p$ and positive definite $p \times p$ matrix S. We call $\boldsymbol{\mu}$ the symmetry center and S the shape matrix of F. If the second-order moments of $\boldsymbol{X} \sim F$ exist, then $\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\mu}$, and $\mathrm{Var}(\boldsymbol{X}) = \Sigma(F)$ is proportional to S. The class of all continuous, elliptical distributions constitutes a generalization of the multivariate normal model, that allows arbitrarily heavy tails and is therefore well suited to model outlying observations. The normal distribution is obtained by $g_{N_p}(y) = (2\pi)^{-\frac{p}{2}}\exp\big(-\frac{1}{2}y\big)$. A prominent example of a heavy-tailed distribution is the $t_{\nu,p}$-distribution, specified by

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

where the index ν is referred to as the *degrees of freedom*. The moments of $t_{\nu,p}$ are finite only up to order $\nu - 1$. For $\nu \geq 3$ its covariance is $\Sigma = \frac{\nu}{\nu-2}S$, and for $\nu \geq 5$ the excess kurtosis (of each component) is $6/(\nu-4)$. Elliptical distributions do generally not possess finite moments, i.e. Σ does not necessarily exist. Provided $\hat{S}_n$ is $\sqrt{n}$-convergent, we may nevertheless use the adjusted deviance test to test (more generally) for a certain zero pattern in the inverse of the shape matrix S instead of Σ.

IV. Examples of robust scatter estimators

If the fourth-order moments of $\boldsymbol{X} \sim F$ are finite, then $\hat{\Sigma}_n$ fulfills conditions (I) and (II). The corresponding value of σ_1 is $1 + \frac{\kappa}{3}$, where κ is the excess kurtosis of the first (or any other component) of $\boldsymbol{X}$. Thus, if we assume an elliptical population distribution F (with finite fourth-order moments), we may apply the adjusted deviance test, but have to divide $\hat{D}_n(\hat{\Sigma}_n)$ by a consistent estimate of σ_1. Under a heavy-tailed distribution, i.e., if κ is large, the estimator $\hat{\Sigma}_n$ is relatively inefficient, resulting in a test with poor power. An alternative, which keeps its efficiency under heavy tails, is Tyler's scatter estimator. It is defined as the solution $\hat{T}_n = \hat{T}_n(\mathbb{X}_n)$ of

$$\frac{p}{n}\sum_{i=1}^{n}\frac{(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_n)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_n)^T}{(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_n)^T\hat{T}_n^{-1}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_n)} = \hat{T}_n \qquad (5)$$

which satisfies $\det\hat{T}_n = 1$. Here $\hat{\boldsymbol{\mu}}_n$ is an appropriate location estimator, which may be the mean, or, in light of robustness, the Hettmansperger-Randles median [10]. Existence and uniqueness of a solution of (5) and the asymptotic properties of the estimator $\hat{T}_n$ are treated in the original publication [11]. The estimator evidently satisfies condition (I) and under some mild regularity conditions on the population distribution also condition (II). The corresponding value of σ_1 is $1 + \frac{2}{p}$, irrespective of the specific elliptical distribution. This remarkable fact may be phrased as to say the test statistic $\hat{D}_n(\hat{T}_n)$ is asymptotically distribution-free within the elliptical model, a property which it inherits from the estimator $\hat{T}_n$. This has the nice practical implication that, when carrying out the adjusted deviance test, σ_1 needs not to be estimated. Furthermore, for large p, $\hat{T}_n$ is almost as efficient as the MLE $\hat{\Sigma}_n$ at the normal distribution and outperforms $\hat{\Sigma}_n$ at distributions with slightly heavier than normal tails, e.g., at the $t_{\nu,p}$ distribution, if $\nu < p + 4$.

The third example we want to mention is the RMCD, the reweighted version of Rousseeuw's minimum covariance determinant estimator [12], see also [13], Chap. 7, which has become a very popular highly robust scatter estimator. Very roughly, a subsample of size $h = \lfloor tn \rfloor$, where $\frac{1}{2} \leq t < 1$ is some fixed fraction, of the data points is chosen such that the determinant of the sample covariance matrix computed from this subsample is minimal. Afterwards a reweighting step is applied, which increases the efficiency, but maintains the high breakdown point of the initial estimator. The RMCD fulfills conditions (I) and (II). The asymptotics are treated in [14] and [15]. Values for σ_1 can be found in [15]. The RMCD is an appropriate estimator if the outlying observations are assumed to be contaminations, but the bulk of the data is well described by a Gaussian distribution. Similar to Tyler's estimator, the efficiency of the RMCD, relative to sample covariance matrix, increases with p.

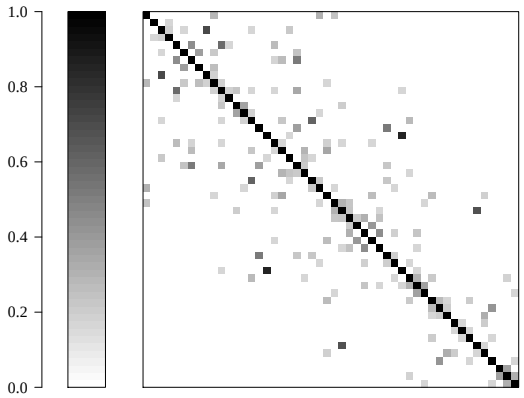

Fig. 1. Partial correlation matrix (absolute values)

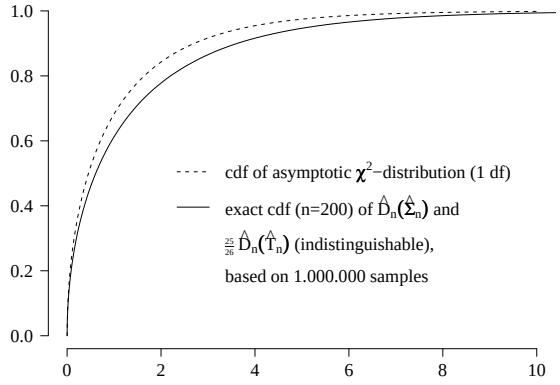

Fig. 2. Asymptotic approximation of the test statistic for $n = 200$ at the normal distribution

V. Simulation study

We want to compare the estimators mentioned in the previous section in a simulation study, to give an impression of their applicability in elliptical graphical modeling. In [1] we report the results obtained from a small toy model consisting of five nodes and five edges. The following is aimed at complementing these numerical investigations by considering a high-dimensional example, where e.g. also run-time plays a role. Our set-up is as follows. We sample 200 i.i.d. observations of a 50-dimensional random vector that follows an elliptical distribution. For each of the elliptical distributions we consider, cf. Table I, we take 1000 samples, and from each sample compute several estimates. Based on each estimate, we select a model and compare it to the true model. In all runs we use the same model (Figure 3) and the same shape matrix with identical diagonal elements. The partial correlation matrix is visualized in Figure 1: the absolute values of the matrix entries are coded by different shades of gray, ranging from 0 (white) to 1 (black). Despite the many intersecting edges in Figure 3 this is a sparse graph. Of 1225 possible edges only 94 are present, and only two nodes (11 and 18) have more than six neighbors.

We perform a very simple model selection: we carry out an edge-exclusion test for every possible edge, i.e. we test, for each pair $\{i, j\}$, the model with all edges but $\{i, j\}$ against the saturated model and exclude the edge $\{i, j\}$, if the test accepts the smaller model. More sophisticated model search procedures generally show better results, but lead to similar conclusions as far as the comparison of the estimators is concerned. Our simple one-step model selection allows to better study the properties of the adjusted deviance tests and the effects of the choice of the scatter estimator. We perform each test at the significance level $\alpha = 0.01$, which is an ad hoc choice. It is chosen rather small due to the sparsity of the graph. Since the vast majority of possible edges is absent, identifying these non-edges correctly is of greater importance for the overall performance in this example. If we view the model selection as a multiple-testing problem, i.e. we want to restrict the probability that the fitted graph is too large, an individual significance level of $\alpha = 0.01$ is already high.

Besides getting an impression of the general performance we want to examine the finite-sample behavior of the estimators, i.e. check if the asymptotic χ^2-approximation of the test statistics are useful in practice. A sample size of 200 seems large enough to expect some "validity" of the asymptotics. We therefore consider two criteria. The main criterion by which we measure the goodness of the model selection is the *relative mean edge difference (RMED)*, i.e. the average number of edges (averaged over all 1000 runs) that are wrongly specified in the selected model—may it be that an existing edge was rejected or an absent edge was wrongly included—divided by the total number of possible edges (1225). An RMED below 0.5 indicates that the model selection procedure is superior to random guessing. In a less complex situation it might be also of interest to know, how often the true model is found, but with 1225 test decisions in each trial we can not expect a positive number in only 1000 trials. Any model selection procedure that is based on testing for zero parameters aims at controlling the probability of correctly specifying the non-edges. Our second criterion is therefore the percentage of wrongly specified non-edges, which is the same as the rejection probability of the test under the null and should turn out to be about 1%.

The findings of our experiment are summarized in Table I. The benchmark is traditional graphical modeling, i.e. the performance of $\hat{\Sigma}_n$ at the normal distribution, cf. first row of Table I. We observe two things: First, the test is anti-conservative. The actual rejection probability under the null hypothesis is about 2.7%. The simulated cdf of the test statistic (for $n = 200$) and its limit for $n \to \infty$ are plotted in Figure 2. Second, the test goes wrong, if we move away from normality. We assume only ellipticity but no further knowledge about the distribution and want methods that are valid over the whole class of elliptical distributions. A possible remedy is to adjust the $\hat{\Sigma}_n$-based test statistic by an estimate of σ_1, which is here based on the component-wise sample kurtosis. The results of this adjusted deviance test are reported in the second row of Table I. This adjustment repairs the test, and does so surprisingly well—even in the case of the t_3-distribution, where the population kurtosis is not defined. In this case we do not have an "asymptotic justification" of the test, but we find it to be conservative. This effect, which we did not observe at

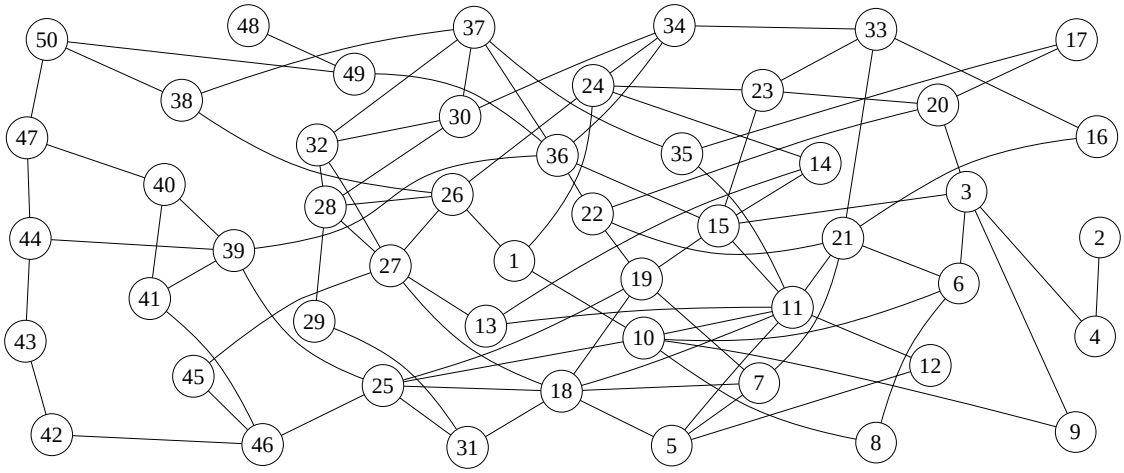

Fig. 3. Example graph

TABLE I
ONE-STEP MODEL SELECTION BASED ON DIFFERENT ESTIMATORS
RMED / WRONGLY SPECIFIED NON-EDGES (%)

distribution	normal	t_{20}	t_5	t_3
$\hat{\Sigma}$	4.9 / 2.6	5.5 / 3.2	8.1 / 6.0	11.4 / 9.5
$\hat{\Sigma}^*$	5.0 / 2.7	5.0 / 2.5	5.3 / 1.1	6.1 / 0.3
$\hat{T}$	5.1 / 2.6	5.0 / 2.6	5.0 / 2.6	5.1 / 2.6
RMCD 0.5**	6.4 / 1.0	6.1 / 0.8	6.2 / 0.9	6.3 / 1.0
RMCD 0.75**	4.2 / 1.0	4.5 / 1.3	6.0 / 2.9	8.1 / 5.2

* test statistic adjusted by estimated kurtosis
** with finite-sample correction

the low-dimensional example, certainly deserves some further investigation.

For Tyler's estimator, there are mainly two things to note. We recognize its asymptotic efficiency properties: it almost equals the performance of $\hat{\Sigma}_n$ at the normal model, but shows no loss under larger tails. On the other hand, the test statistic shows a very similar behavior as the $\hat{\Sigma}_n$-based deviance under normality, cf. Figure 3. It has in particular the same bias w.r.t. the asymptotic χ^2_1-distribution. This gives rise to the hope that finite-sample correction techniques developed for $\hat{\Sigma}_n$-based analyses, cf. e.g. [7], p. 143, can be applied to $\hat{T}_n$ as well and be brought to benefit also under ellipticity.

Finally, Table I also reports results for the RMCD, with subsample fractions $t = 0.5$ and $t = 0.75$, which both exhibit generally good efficiencies, which is in contrast to the low-dimensional example. But it must be pointed out that we did not carry out an asymptotic test in this situation. It is a known problem of the RMCD that it converges very slowly to its asymptotic distribution. The "asymptotic" σ_1-value is of no use here. The problem is usually taken care of by multiplying by a correction factor which has to be determined numerically. We have chosen σ_1 such that the test delivers the desired rejection probability of 0.01 under the null at the normal model. This makes the RMCD look unjustifiably good in comparison to the other estimators.

All calculations were done in R 2.9.1, employing routines from the packages `mvtnorm` (random sampling), `ggm` (constrained estimation, i.e. the function h_G), `ICSNP` (Tyler matrix) and `rrcov` (RMCD). The simulations were run on a 2.83 GHz Intel Core2 CPU. The computation of Tyler's estimator lasted less then a second, the RMCD less than 3 seconds, all 1225 edge-exclusion tests took about 37 seconds. Figure 3 was created using Graphviz.

REFERENCES

[1] D. Vogel and R. Fried, "Elliptical graphical modelling." Working paper, 2010.

[2] U. Gather, M. Imhoff, and R. Fried, "Graphical Models for Multivariate Time Series form Intensive Care Monitoring." *Statistics in Medicine*, vol. 21, pp. 2685–2701, 2002.

[3] B. Schelter, M. Winterhalder, B. Hellwig, B. Guschlbauer, C. H. Lücking, and J. Timmer, "Direct or Indirect? Graphical Models for Neural Oscillators." *J. Physiol.*, vol. 99, pp. 37–46, 2006.

[4] J. Whittaker, *Graphical models in applied multivariate statistics.* Wiley Series in Probability and Mathematical Statistics. Chichester etc.: Wiley, 1990.

[5] D. Vogel and R. Fried, "On robust Gaussian graphical modelling," in *Recent Developments in Applied Probability and Statistics. Dedicated to the Memory of Jürgen Lehn*, L. Devroye, B. Karasözen, M. Kohler, and R. Korn, Eds. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 155–182.

[6] D. R. Cox and N. Wermuth, *Multivariate dependencies: models, analysis and interpretation.* Monographs on Statistics and Applied Probability. 67. London: Chapman and Hall, 1996.

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

[8] D. Edwards, *Introduction to graphical modelling.* Springer Texts in Statistics. New York, NY: Springer, 2000.

[9] Y. Zuo, "Robust location and scatter estimators in multivariate analysis," in *Frontiers in statistics. Dedicated to Peter John Bickel on honor of his 65th birthday*, J. Fan and H. Koul, Eds. London: Imperial College Press, 2006, pp. 467–490.

[10] T. Hettmansperger and R. Randles, "A practical affine equivariant multivariate median." *Biometrika*, vol. 89, pp. 851–860, 2002.

[11] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter." *Ann. Stat.*, vol. 15, pp. 234–251, 1987.

[12] P. J. Rousseeuw, "Multivariate estimation with high breakdown point." in *Mathematical statistics and applications, Proc. 4th Pannonian Symp. Math. Stat., Bad Tatzmannsdorf, Austria, September 4-10, 1983, Vol. B*, W. Grossmann, G. C. Pflug, I. Vincze, and W. Wertz, Eds. Dordrecht etc.: D. Reidel, 1985, pp. 283–297.

[13] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York etc.: Wiley, 1987.

[14] R. W. Butler, P. L. Davies, and M. Jhun, "Asymptotics for the minimum covariance determinant estimator." *Ann. Stat.*, vol. 21, no. 3, pp. 1385–1400, 1993.

[15] C. Croux and G. Haesbroeck, "Influence function and efficiency of the minimum covariance determinant scatter matrix estimator." *J. Multivariate Anal.*, vol. 71, no. 2, pp. 161–190, 1999.