Dependence in elliptical partial correlation graphs[*]

David Rossell, Piotr Zwiernik

Universitat Pompeu Fabra and Barcelona GSE
e-mail: david.rossell@upf.edu; piotr.zwiernik@upf.edu

Abstract: The Gaussian model equips strong properties that facilitate studying and interpreting graphical models. Specifically it reduces conditional independence and the study of positive association to determining partial correlations and their signs. When Gaussianity does not hold partial correlation graphs are a useful relaxation of graphical models, but it is not clear what information they contain (besides the obvious lack of linear association). We study elliptical and transelliptical distributions as middle-ground between the Gaussian and other families that are more flexible but either do not embed strong properties or do not lead to simple interpretation. We characterize the meaning of zero partial correlations in elliptical and elliptical copula models and show that they retain much of the dependence structure from the Gaussian case. Regarding positive dependence, we prove impossibility results to learn certain positive (trans)elliptical graphical models, including that an elliptical distribution that is multivariate totally positive of order two for all dimensions must be essentially Gaussian. We then show how to interpret positive partial correlations as a relaxation, and obtain important properties related to faithfulness and Simpson's paradox. We illustrate the transelliptical model potential to study tail dependence in S&P500 data, and of positivity to improve regularized inference.

MSC 2010 subject classifications: 62H05, 62H20, 62H22.
Keywords and phrases: Partial correlation graph, elliptical distribution, transelliptical distribution, graphical models, multivariate total positivity.

Received February 2021.

1. Introduction

Several papers study graphical models for elliptical and transelliptical distributions in the standard (Finegold and Drton, 2009; Vogel and Fried, 2011) and high-dimensional settings (Barber and Kolar, 2018; Bilodeau, 2014; Liu, Han and Zhang, 2012; Zhao and Liu, 2014). These models found applications in many fields, such as finance and biology (Behrouzi and Wit, 2019; Stephens, 2013; Vinciotti and Hashem, 2013), and (implicitly) wherever Gaussian graphical models were used but the underlying data-generating distribution is likely to depart from normality, e.g. be heavy-tailed or skewed. In the elliptical setting

the usual definition of graphical models mimics the Gaussian case — the model is given by zeros in the inverse covariance, or equivalently, by vanishing partial correlations. Despite this being a reasonable relaxation, the corresponding partial correlation graph (PG) cannot be interpreted in terms of conditional independence, since outside of the normal case no elliptical distributions allow for conditional independence (c.f. Proposition 2.2). It is therefore interesting to portray what type of dependence information is embedded by the PG, in particular the meaning of zero and certain positive partial correlations.

For general distributions partial correlations inform only about linear dependence. Missing edges in the PG must then be interpreted with great care and, in some cases, they can fail to capture interesting dependence information. For example, in an aircraft data set from Bowman and Foster (1993), we can model the dependence between the speed of an airplane and its wingspan. Although the sample correlation is negligible, more flexible dependence tests reveal that the variables are strongly related; see e.g. Székely and Rizzo (2009). The reason is that for very fast (military) airplanes there is a negative dependence between speed and wingspan, while this dependence is positive for regular aircrafts.

The main theme of this paper is that for (trans)elliptical distributions there is significantly more information in the partial correlation graph, and in sign constraints imposed on the partial correlations, beyond presence/absence of linear dependence. We introduce definitions and notation to aid the exposition.

Definition 1.1. A random vector $X = (X_1, \ldots, X_d)$ has an elliptical distribution if there exists $\mu \in \mathbb{R}^d$ and a positive semi-definite matrix Σ such that the characteristic function of X is of the form $\boldsymbol{t} \mapsto \phi(\boldsymbol{t}^T \Sigma \boldsymbol{t}) \exp(i \mu^T \boldsymbol{t})$ for some $\phi : [0, \infty) \to \mathbb{R}$. We write $X \sim E(\mu, \Sigma)$ making ϕ in this notation implicit.

Important examples include the multivariate normal, Laplace and multivariate t-distributions. Elliptical graphical models have been extended to transelliptical distributions (also known as elliptical copulas or meta-elliptical distributions, Fang, Fang and Kotz (2002); Liu, Han and Zhang (2012)).

Definition 1.2. A random vector Y has a transelliptical distribution with parameters (μ, Σ) if $f(Y) := (f_1(Y_1), \ldots, f_d(Y_d)) \sim E(\mu, \Sigma)$

for some fixed strictly increasing functions $f_1, \ldots, f_d$. We write $Y \sim \mathrm{TE}(\mu, \Sigma)$, making f in this notation implicit.

Here the additional challenge is that f is unknown. An elegant approach to learning partial correlations relies on directly estimating the correlation matrix of $f(Y)$ without actually learning f; see Liu, Han and Zhang (2012); Lindskog, Mcneil and Schmock (2003), and then proceed as in the elliptical case (Section 3.3).

Throughout we assume that Σ is positive definite and denote $K = \Sigma^{-1}$, the set of vertices by $V = \{1, \ldots, d\}$, by $X_{(i)}$ the $d-1$ vector obtained by removing X_i from X, and by $X_{(ij)}$ the $d-2$ vector obtained by removing (X_i, X_j) from X. Given $I, J \subseteq V$ denote by X_I and μ_I the subvectors of X and μ with coordinates in I and by Σ_{IJ} the corresponding subblock of Σ with rows in I and columns

in J. The partial correlation between (X_i, X_j) is

$$\rho_{ij \cdot V \setminus \{i,j\}} = -\frac{K_{ij}}{\sqrt{K_{ii} K_{jj}}} \qquad \text{for all } i, j \in V \qquad (1)$$

and so $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ if and only if $K_{ij} = 0$. Finally, we denote that (X_i, X_j) are independent by $X_i \perp\!\!\!\perp X_j$.

The usual interpretation of PGs in elliptical distributions is that — since the conditional expectation $\mathbb{E}(X_i | X_{(i)})$ is linear in $X_{(i)}$ and the conditional correlation is equal to the partial correlation — the condition $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ implies that (X_i, X_j) are conditionally uncorrelated. That is, zero partial correlation implies zero conditional correlation. As we show in Theorem 3.3 something *much* stronger is true. It is possible to fully characterize the PG in elliptical distributions:

> The partial correlation $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ if and only if $\operatorname{cov}(g(X_i), X_j | X_{(ij)}) = 0$ for **every** function g for which the covariance exists.

This result extends an analogous finding on marginal correlations stemming from Stein's lemma (see Lemma 3.2 in Hamada et al. (2004); Theorem 2 in Landsman and Nešlehová (2008)) to the context of graphical models. We also provide a similar characterization for transelliptical distributions. The usual interpretation of $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ is that $f_i(Y_i), f_j(Y_j)$ are conditionally uncorrelated given $f_{(ij)}(Y)$, which is not very interesting, since f is unknown. We show in Theorem 3.4 that equivalently $\operatorname{cov}(f_i(Y_i), g(Y_j) | Y_{(ij)}) = 0$ for any g, provided the covariance exists. In particular, $\operatorname{cov}(f_i(Y_i), Y_j | Y_{(ij)}) = 0$, a more explicit dependence information in terms of Y. We also show in Proposition 3.7 yet another equivalent interpretation, namely that $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ if and only if the conditional Kendall's tau correlation between $g(Y_i)$ and $h(Y_j)$ is zero for all strictly increasing g, h.

These findings are practically relevant. Recall that two variables X_i and X_j with general distribution are independent if and only if for all $L^2(\mathbb{R})$ functions g, h we have $\operatorname{cov}(g(X_i), h(X_j)) = 0$; see e.g. (Feller, 1971, page 136). That is, $X_i \perp\!\!\!\perp X_j$ if and only if there is no way to transform X_i and X_j such that the new variables are correlated. Our characterization of $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ has an analogous interpretation, in elliptical families there is no way to transform X_i such that the new variable is conditionally correlated with X_j. This rules out situations like the aircraft example above where speed is correlated with a non-linear function of wingspan. Further, in trans-elliptical families there is no way to transform X_i and X_j such that Kendall's conditional tau is non-zero.

Interpreting $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ can be important in applications, given that elliptical and transelliptical models are a popular tool to capture second-order or tail dependencies. Even though $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ such dependencies can be practically significant. For example, in Section 5 we illustrate how $\operatorname{corr}(X_i^2, X_j^2)$ can serve as a simple measure of marginal tail (or second-order) dependence related to popular models in time series analysis, and how to define a related conditional second-order dependence measure related to graphical models.

Our other main contributions relate to PGs in settings where one wishes to study positive forms of association. Two standard ways to define positive dependence are via the notions of *multivariate total positivity of order two* (MTP_2) and *conditionally increasing* (CI, Section 4). Although these concepts are different, in the Gaussian case they are equivalent and reduce to constraining partial correlations to be non-negative. It is less clear how to interpret these concepts in general elliptical families. A first contribution is showing several impossibility results: within the elliptical family with at least one partial correlation zero there exist no conditionally increasing distributions (other than the Normal) implying the same result for MTP_2 distributions. That is, if one wants to remove edges in the PG with an additional positive dependence structure, one cannot rely on the standard notions of positive dependence.

A natural relaxation is to learn a PG under the constraint that $\rho_{ij \cdot V \setminus \{i,j\}} \geq 0$, as proposed by Agrawal, Roy and Uhler (2019). We refer to this strategy as *positive partial correlation graphs* (PPG). We contribute to understanding how should one interpret missing edges in the PPG, and to characterizing embedded positivity properties such as the positive correlation of each X_i with any increasing function of the vector X. In Section 5 we illustrate how positivity constraints induce a type of regularization that can help improve inference relative to other standard forms of regularization, such as graphical LASSO. The significance of this example is that one attains a higher log-likelihood with a sparser graph by restricting attention to positive models. This is meant as a testimony that our theoretical results on positivity have practical relevance. For further examples in risk modelling see Abdous, Genest and Rémillard (2005); Rüschendorf and Witting (2017), and in psychology see Epskamp and Fried (2018); Lauritzen, Uhler and Zwiernik (2019a), for example.

This paper also contributes to recent research aimed at understanding multivariate total positivity in a wide variety of contexts; see, for example, Fallat et al. (2017); Lauritzen, Uhler and Zwiernik (2019b,a); Robeva et al. (2018); Slawski and Hein (2015). We provide in Theorem 4.9 a complete characterization of elliptical MTP_2 distributions in terms of their density generator. In Theorem 4.10 this characterisation is used to show that a density generator may induce a d-variate MTP_2 distribution for each $d \geq 2$ if and only if the underlying distribution is essentially Gaussian.

The paper is organized as follows. In Section 2 we review basic results on elliptical distributions. In Section 3 we characterize partial correlation graphs for elliptical and transelliptical distributions, giving a refined understanding of their encoded dependence information, and how one should interpret zero partial correlations. In Section 4 we study positive elliptical distributions and their properties. In Section 5 we illustrate our main results with examples.

In the rest of the article we employ the convention where the letter X is reserved for elliptical random variables, Y is reserved for transelliptical variables, Z is reserved for Gaussian variables and W for general random variables. The letter f is reserved for the monotone functions defining transelliptical distributions as in Definition 1.2, g and h denote general functions, and p denotes probability density functions.

2. Elliptical distributions

2.1. Stochastic representation

If $X \sim E(\mu, \Sigma)$ then X admits the representation

$$X = \mu + \xi \cdot \Sigma^{1/2} \cdot U, \tag{2}$$

where $\Sigma^{1/2}$ denotes the square root of the positive-definite Σ, $\xi \in \mathbb{R}^+$ is an arbitrary random variable, $U \in \mathbb{R}^d$ is uniformly distributed on the unit $(d-1)$-dimensional sphere, and $\xi \perp\!\!\!\perp U$. Elementary arguments show that, if $\mathbb{E}\xi < \infty$, then $\mathbb{E}X = \mu$ and if $\mathbb{E}\xi^2 < \infty$ then $\operatorname{cov}(X) = \frac{\mathbb{E}(\xi^2)}{d} \cdot \Sigma$. Throughout this paper we assume that $0 < \mathbb{E}\xi^2 < \infty$.

There is a useful representation equivalent to (2) in terms of normal variables. Let $D^2 \sim \chi_d^2$ with $D^2 \perp\!\!\!\perp U$. Then $Z = \sqrt{D^2}\, \Sigma^{1/2} U$ is a mean zero Gaussian variable with covariance Σ. From (2) it follows that

$$X = \mu + \frac{1}{\sqrt{\tau}} \cdot Z, \tag{3}$$

where $\tau = D^2/\xi^2$ and $Z \sim N(0, \Sigma)$. The special case where τ and Z are independent corresponds to the *scale mixture of normals* sub-family, which includes most popular elliptical distributions. The normal corresponds to $\tau \equiv 1$. If $\tau \sim \chi_k^2/k$ for $k > 2$ then X has a multivariate t-distribution with parameter k, and if $\tau \sim \operatorname{Exp}(1)$ the multivariate Laplace distribution. Equivalently, scale mixture of normals can be defined as the marginal distribution of X associated to (X, τ) when $X \mid \tau \sim N(\mu, \tau^{-1/2}\Sigma)$, and $\tau \in \mathbb{R}^+$ follows an arbitrary continuous distribution.

The elliptical family is closed under taking margins and under conditioning.

Proposition 2.1. *Let $X = (X_I, X_J) \sim E(\mu, \Sigma)$ be any split of X into subvectors X_I and X_J. Then*

(i) $X_I \sim E(\mu_I, \Sigma_{II})$,
(ii) $X_I \mid X_J = x_J \sim E(\mu_{I|J}, \Sigma_{II \cdot J})$, where $\mu_{I|J} := \mu_I + \Sigma_{IJ}\Sigma_{JJ}^{-1}(x_J - \mu_J)$ and

$$\Sigma_{II \cdot J} := \Sigma_{II} - \Sigma_{IJ}\Sigma_{JJ}^{-1}\Sigma_{JI} = K_{II}^{-1}.$$

For the proof see (Fang, 2018, Theorem 2.18). The conditional mean $\mu_{I|J}$ has the same form as in the Gaussian case, where Σ above can be replaced by $\operatorname{cov}(X)$. Moreover, the conditional correlations $\operatorname{corr}(X_i, X_j \mid X_{(ij)})$ are the normalized entries of $K = \Sigma^{-1}$ (the partial correlations $\rho_{ij \cdot V \setminus \{i,j\}}$), and do not depend on the conditioning variable $X_{(ij)}$.

2.2. Characterization of Gaussianity within the elliptical family

If X is Gaussian then each marginal distribution and each conditional distribution is Gaussian. Moreover, the conditional covariances do not depend on the

conditioning variable and independence is equivalent to zero correlations. These properties characterize the Gaussian distribution in the class of elliptical distributions. More precisely, suppose that $X \sim E(\mu, \Sigma)$. By Lemma 4, Lemma 8, and Theorem 7 in Kelker (1970) we have the following statements:

(i) If X_I is Gaussian for some $I \subseteq V$ then X is Gaussian.
(ii) If X_I given X_J is Gaussian for some $I, J \subseteq V$ then X is Gaussian.
(iii) If the conditional covariance of X_I given X_J is independent of X_J then X is Gaussian.

The standard definition of graphical models uses density factorizations that link to conditional independence through the Hammersley-Clifford theorem (Lauritzen, 1996). However, it is not possible to define conditional independence in the elliptical family outside of the Gaussian case. Again suppose that $X \sim E(\mu, \Sigma)$. By Lemma 5 in Kelker (1970), if Σ is a diagonal matrix, then the components of X are independent if and only if X has a normal distribution. This can be used to prove the following important result.

Proposition 2.2 (Theorem 3 in Baba, Shibata and Sibuya (2004)). *Suppose that $X \sim E(\mu, \Sigma)$ and $X_i \perp\!\!\!\perp X_j | X_C$ for some $i, j \in V$ and $C \subseteq \{1, \ldots, d\} \setminus \{i, j\}$. Then X is Gaussian.*

3. Graphs for (trans)elliptical distributions

3.1. Partial correlation graph and dependence

By Proposition 2.2 it is not possible to do structural learning in (non-normal) elliptical graphical models, under the conditional independence definition. It is then natural to look for relaxations that may be useful from the modelling point of view. A common strategy is to define graphs based on zeroes in the inverse covariance matrix, mimicking the Gaussian case; see Vogel and Fried (2011).

Definition 3.1. The partial correlation graph (PG) is the graph $G = G(K)$ over vertex set $V = \{1, \ldots, d\}$ with an edge between $i \neq j$ if and only if $K_{ij} \neq 0$.

Equivalently, $K_{ij} = 0$ if and only if the partial correlation $\rho_{ij \cdot V \setminus \{i,j\}} = 0$. In general, $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ does not imply conditional independence but only linear independence. The aim of this section is to understand what additional information does the PG carry in elliptical distributions. Proposition 2.1 and standard matrix algebra give

$$\mathbb{E}(X_i | X_{(i)}) = \mu_i - \tfrac{1}{K_{ii}} K_{i,(i)}(X_{(i)} - \mu_{(i)}), \tag{4}$$

hence $K_{ij} = 0$ if and only if $\mathbb{E}(X_i | X_{(i)})$ does not depend on X_j. This immediately gives the following standard result.

Proposition 3.2. *Let $X \sim E(\mu, K^{-1})$. Then $K_{ij} = 0$ if and only if $\mathrm{cov}(X_i, X_j | X_{(ij)}) = 0$, or equivalently, $\mathbb{E}(X_i | X_{(i)})$ and $\mathbb{E}(X_j | X_{(j)})$ depend on $X_{(ij)}$ only.*

As we now show, much more can be said.

Theorem 3.3. *Let $X \sim E(\mu, K^{-1})$ and $I, J \subseteq V$. Then $K_{IJ} = 0$ if and only if $\operatorname{cov}(g(X_I), X_J | X_{V \setminus (I \cup J)}) = 0$ for any function g for which this covariance exists.*

Proof.. We first show that $\Sigma_{IJ} = 0$ if and only if $\operatorname{cov}(g(X_I), X_J) = 0$ for any function g for which the covariance exists. The proof of this statement can be extracted from the first part of the proof of Lemma 3.2 in Hamada et al. (2004). Without loss of generality assume $J = \{j\}$. By the law of total expectation

$$\operatorname{cov}(g(X_I), X_j) = \mathbb{E}_{X_I} \Big[(g(X_I) - \mathbb{E}g(X_I)) \mathbb{E}(X_i - \mu_j | X_I) \Big],$$

where the expectation E_{X_I} is computed with respect to the marginal distribution of X_I. The expression inside the expectation is almost surely zero with respect to this distribution. Indeed, by Proposition 2.1,

$$\mathbb{E}(X_j - \mu_j | X_I) = \mathbb{E}(X_j | X_I) - \mu_j = \Sigma_{jI} \Sigma_{II}^{-1} (X_I - \mu_I),$$

which is zero because $\Sigma_{jI} = 0$.

To complete the proof, we apply this result to the conditional distribution of $I \cup J$ given $V \setminus (I \cup J)$. Let $A = I \cup J$ and $B = V \setminus (I \cup J)$. By Proposition 2.1(ii) the distribution of X_A given X_B is equal to the elliptical distribution $E(\mu_{A|B}, \Sigma_{A,A|B})$, where the matrix $\Sigma_{AA|B} := \Sigma_{AA} - \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BA}$ is block diagonal (with blocks corresponding to I and J) if and only if $K_{IJ} = 0$. $\square$

Theorem 3.3 is an if and only if statement, that is, it characterizes the presence of zero partial correlations and how to interpret elliptical PGs: if (X_i, X_j) are conditionally uncorrelated then so are X_j and any function of X_i. For instance, there is no linear association between X_j and higher-order moments associated to X_i^2, X_i^3, etc.

In applications it is often interesting to consider semi-parametric techniques related to copula models given by transelliptical distributions. We now extend our result to such settings.

3.2. Transelliptical distributions

Recall that Y has a transelliptical distribution, denoted $Y \sim \mathrm{TE}(\mu, \Sigma)$, if and only if $f(Y) = (f_1(Y_1), \ldots, f_d(Y_d)) \sim E(\mu, \Sigma)$ for strictly increasing deterministic functions f_i. An interesting result is that, if $f(Y)$ is Gaussian (nonparanormal sub-family) the PG gives conditional independence on Y. That is, for the nonparanormal family the PG is highly interpretable (Liu et al., 2012). More generally, a missing edge in the PG means that $\operatorname{cov}(f_i(Y_i), f_j(Y_j) | f_{(ij)}(Y)) = 0$, but this interpretation is not very interesting given that f is unknown and simply refers to linear conditional independence between the latent $(f_i(Y_i), f_j(Y_j))$. The focus should be directly on the dependence structure of Y.

Our second main result shows that a weaker version of Theorem 3.3 holds for transelliptical distributions.

Theorem 3.4. *Suppose* $Y \sim \mathrm{TE}(\mu, K^{-1})$. *Then* $K_{ij} = 0$ *if and only if the conditional covariance* $\mathrm{cov}(f_i(Y_i), g(Y_j)|Y_{(ij)})$ *is zero for every function* g *for which the covariance exists.*

Proof.. Let $X = f(Y) \sim E(\mu, K^{-1})$. Suppose that $\mathrm{cov}(f_i(Y_i), g(Y_j)|Y_{(ij)}) = 0$ for all g, taking $g = f_j$ gives $\mathrm{cov}(f_i(Y_i), f_j(Y_j) \mid Y_{(ij)}) = \mathrm{cov}(X_i, X_j \mid Y_{(ij)}) = 0$. Since each f_i is a strictly monotone fixed function, it is bijective. It follows that $X_{(ij)}$ and $Y_{(ij)}$ generate the same σ-field giving that $\mathrm{cov}(X_i, X_j \mid Y_{(ij)}) = \mathrm{cov}(X_i, X_j \mid X_{(ij)})$. Vanishing of this conditional covariance is equivalent to $K_{ij} = 0$. To prove the reverse implication, note that

$$\mathrm{cov}(f_i(Y_i), g(Y_j)|Y_{(ij)}) = \mathrm{cov}(X_i, g(f_j^{-1}(X_j))|X_{(ij)}) = 0$$

where the second equality follows from Theorem 3.3. □

Theorem 3.4 helps interpret the PG as follows. If $K_{ij} = 0$ then $f_i(Y_i)$ is conditionally uncorrelated with any function of Y_j. Hence learning a single element f_i within f (rather than the whole f) describes (local) aspects of conditional dependence of Y_i on $Y_{(i)}$ (and functions thereof). Taking g to be the identity function in Theorem 3.4 we get the following result.

Corollary 3.5. *Suppose* $Y \sim \mathrm{TE}(\mu, K^{-1})$. *If* $K_{ij} = 0$ *then* $\mathrm{cov}(g(Y_i), Y_j|Y_{(ij)}) = 0$ *for* some *strictly increasing function* g.

The function g in this corollary is precisely the function f_i in Theorem 3.4. Corollary 3.5 gives the following interpretation. If $K_{ij} = 0$ then Y_i is conditionally uncorrelated with some strictly increasing transformation of Y_j and also Y_j is conditionally uncorrelated with some strictly increasing transformation of Y_i.

3.3. Rank correlations

Theorem 3.4 is an if and only if statement, hence it fully characterizes the meaning of PG in transelliptical distributions using covariances between any function of Y_j and latent $f_i(Y_i)$. A conditional version of Kendall's tau gives an interesting alternative characterization that can be interpreted without any reference to f. Let $W = (W_1, \ldots, W_d)$ be a continuous random vector and $W' = (W'_1, \ldots, W'_d)$ be an independent copy. Kendall's tau for (W_i, W_j) is

$$\tau(W_i, W_j) := \mathrm{corr}\big(\mathrm{sign}(W_i - W'_i), \mathrm{sign}(W_j - W'_j)\big).$$

Further, define conditional Kendall's correlation as

$$\tau(W_i, W_j \mid W_{(ij)}) := \mathrm{corr}(\mathrm{sign}(W_i - \widetilde{W}'_i), \mathrm{sign}(W_j - \widetilde{W}'_j) \mid W_{(ij)}), \qquad (5)$$

where $(\widetilde{W}'_i, \widetilde{W}'_j)$ is an independent copy of (W_i, W_j) from the conditional distribution given $W_{(ij)}$. In elliptically-distributed $X \sim E(\mu, \Sigma)$ the following beautiful result relates Pearson correlations $\rho(X_i, X_j) = \mathrm{corr}(X_i, X_j) = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$ with Kendall's tau.

Lemma 3.6 (Lindskog, Mcneil and Schmock (2003))**.** *If* $X \sim E(\mu, \Sigma)$ *then*

$$\tau(X_i, X_j) \; = \; \frac{2}{\pi} \arcsin(\rho(X_i, X_j)).$$

Let $Y \sim \text{TE}(\mu, \Sigma)$, so that $X = f(Y) \sim E(\mu, \Sigma)$. Since Kendall's tau is invariant under strictly increasing transformations,

$$\tau(Y_i, Y_j) \; = \; \tau(X_i, X_j) \; = \; \frac{2}{\pi} \arcsin(\rho(X_i, X_j)).$$

Thus, if one can consistently estimate $\tau(Y_i, Y_j)$ from the observations of Y, then one can also consistently learn $\rho(X_i, X_j) = \sin(\frac{\pi}{2}\tau(Y_i, Y_j))$, the correlation matrix of $X = f(Y)$. An obvious consequence of Lemma 3.6 is that, for $Y \sim \text{TE}(0, \Sigma)$, $\text{sign}(\Sigma_{ij}) = \text{sign}(\tau(Y_i, Y_j)) \in \{-1, 0, 1\}$. For partial correlations we have the following result.

Proposition 3.7. *Let* $Y \sim \text{TE}(0, K^{-1})$. *Then* $K_{ij} = 0$ *if and only if* $\tau(Y_i, Y_j \mid Y_{(ij)}) = 0$.

Proof.. Let $X = f(Y) \sim E(\mu, \Sigma)$. The condition $\tau(Y_i, Y_j \mid Y_{(ij)}) = 0$ is equivalent to $\tau(X_i, X_j \mid Y_{(ij)}) = \tau(X_i, X_j \mid X_{(ij)}) = 0$, which by Lemma 3.6 applied to the conditional distribution of (X_i, X_j) given $X_{(ij)}$ is equivalent to $\rho_{ij \cdot V \setminus \{i,j\}} = 0$, or equivalently, $K_{ij} = 0$. □

4. Positive dependence in elliptical distributions

In this section we study PGs in elliptical distributions when one imposes positive dependence. Section 4.1 begins by recalling several important notions of multivariate positive dependence, and in Section 4.2 we show that said notions are not meaningful to learn structure in elliptical PGs. This leads to a relaxed notion of positive dependence given by elliptical distributions whose partial correlations are all nonnegative, which we refer to as positive partial correlation graphs (PPGs). In Section 4.3 we define PPGs and study their embedded positive dependence properties, which complement the interpretation offered by the characterizations in Section 3. Finally, Section 4.4 offers an extension of the impossibility results in Section 4.2, by characterizing the restrictiveness of the class of MTP_2 elliptical distributions.

4.1. Positive dependence

We start by recalling definitions of some classical notions of positive dependence. A random vector W is *associated* (A) if and only if

$$\text{cov}(g(W), h(W)) \; \geq \; 0$$

for any two coordinate non-decreasing functions $g, h : \mathbb{R}^d \to \mathbb{R}$ for which this covariance exists. A random vector W with a density function p is *multivariate totally positive of order two* (MTP_2) if and only if

$$p(w)p(w') \; \leq \; p(\min(w, w'))p(\max(w, w')) \qquad \text{for all } w, w' \in \mathbb{R}^d, \qquad (6)$$

where $\min(w, w') = (\min(w_1, w_1'), \ldots, \min(w_d, w_d'))$ is the coordinatewise minimum and $\max(w, w') = (\max(w_1, w_1'), \ldots, \max(w_d, w_d'))$ the coordinatewise maximum of w and w'. Finally, W is *conditionally increasing* (CI) if for every $i \in V$ and $C \subseteq V \setminus \{i\}$ the conditional expectation $\mathbb{E}(h(W_i)|W_C)$ is an increasing function of W_C for every increasing function $h : \mathbb{R} \to \mathbb{R}$. Good general references are Esary, Proschan and Walkup (1967) for association, Karlin and Rinott (1980) for MTP$_2$, and Müller and Scarsini (2001) for CI.

One of the first results on these model classes is that if W is A/MTP$_2$/CI then each marginal distribution is A/MTP$_2$/CI. Moreover, if W is MTP$_2$/CI then each conditional distribution is MTP$_2$/CI. Another well-known result is that these positivity notions are closed under monotone transforms and that MTP$_2$ implies CI; see Theorem 3.3 and Proposition 3.5 in Müller and Scarsini (2001) as well as Proposition 3.1 in Fallat et al. (2017).

In the Gaussian case both condition (6) and CI simplify to an explicit constraint on the inverse covariance K. A symmetric positive definite matrix K is called an M-matrix if $K_{ij} \leq 0$ for all $i \neq j$. Denote the set of inverses of M-matrices by $\mathcal{IM}$. Directly from (1), $\Sigma \in \mathcal{IM}$ if and only if all partial correlations $\rho_{ij \cdot V \setminus \{i,j\}}$ are nonnegative. By Proposition 3.6 in Müller and Scarsini (2001), for a Gaussian X we have

$$X \text{ is MTP}_2 \quad \Longleftrightarrow \quad X \text{ is } CI \quad \Longleftrightarrow \quad \Sigma \in \mathcal{IM}.$$

To help the reader, we briefly describe what is coming in the next two sections. In Sections 4.2 and 4.3 we show that that, if $Y \sim \text{TE}(\mu, \Sigma)$ then

$$Y \text{ is MTP}_2 \quad \Longrightarrow \quad Y \text{ is } CI \quad \Longrightarrow \quad \Sigma \in \mathcal{IM},$$

where the last condition corresponds to positive transelliptical distributions (Definition 4.4). In the transelliptical case all three positivity notions are closed under marginalization, conditioning, and monotone transformations. We show that both implications above can be inverted only in the Gaussian/nonparanormal case, showing that Proposition 3.6 in Müller and Scarsini (2001) characterizes the nonparanormal within the elliptical family. Moreover, in general there is no clear relation between association and positive transelliptical distributions. For modelling with zeros in Σ^{-1} we will show that neither MTP$_2$ nor CI lead to interesting models (unless in the Gaussian case), which will be the main motivation for considering positive transelliptical distributions. In this case we will also show that there is a close relation between vanishing partial correlations in various marginal distributions, which imply that this model satisfies partial faithfulness.

4.2. Positive dependence in elliptical distributions

To motivate Section 4.3 we study in this section limitations of elliptical distributions in the context of the positive dependence notions reviewed in Section 4.1. We first show in Theorem 4.2 that if K has any zeroes then X cannot be CI

(hence neither MTP$_2$, from Theorem 3.3 in Müller and Scarsini (2001)) unless X is Gaussian. The same impossibility result applies to transelliptical families (outside the nonparanormal sub-family). As a consequence, it is not possible to learn structure (remove edges) of a non-normal elliptical graphical model under these positivity constraints.

Even if one were to forsake structural learning and focus on the fully dense graph with no missing edges, it is not possible to find MTP$_2$/CI transelliptical distributions, except in very restrictive cases. For example, Proposition 4.3 shows that there are no MTP$_2$/CI t-distributions. We defer a deeper analysis to Section 4.4, where we fully characterize the elliptical MTP$_2$ class and show that it is highly restrictive, particularly as d grows.

Remark 4.1. Suppose $X \sim E(\mu, K^{-1})$. From (4) it follows that K is an M-matrix if and only if for every $i \in V$ the conditional expectation $\mathbb{E}(X_i | X_{(i)})$ is increasing in $X_{(i)}$. Note that, if X is CI then $\mathbb{E}(X_i | X_{(i)})$ must be increasing in $X_{(i)}$ and so, in particular, $\rho_{ij \cdot V \setminus \{i,j\}} \geq 0$ for all $i, j \in V$. This shows that nonnegativity of all partial correlations is a necessary condition for X to be CI and so also for X to be MTP$_2$.

Theorem 4.2. *Suppose that $X \sim E(\mu, K^{-1})$ and X is CI. If K has a zero entry then X is Gaussian. Further, suppose that $Y \sim \mathrm{TE}(\mu, K^{-1})$ and Y is CI. Let $X = f(Y) \sim E(\mu, K^{-1})$, if K has a zero entry then X is Gaussian.*

Proof.. Let $X \sim E(\mu, K^{-1})$ be CI and suppose $K_{ij} = 0$. By Proposition 3.2, the conditional covariance $\mathrm{cov}(X_i, X_j | X_{(ij)})$ is zero. Since X is CI, the conditional distribution of (X_i, X_j) given $X_{(ij)}$ is also CI. It is well known that CI distributions are also associated; c.f. Colangelo, Scarsini and Shaked (2005). By Corollary 3 in Newman (1984) applied to this conditional distribution we get that $\mathrm{cov}(X_i, X_j | X_{(ij)}) = 0$ implies $X_i \perp\!\!\!\perp X_j | X_{(ij)}$. From Proposition 2.2 we know that the latter is only possible if X is Gaussian. Consider now $Y \sim \mathrm{TE}(\mu, K^{-1})$. Since CI is closed under monotone transformations (see Proposition 3.5 in Müller and Scarsini (2001)), $X = f(Y)$ is CI and, since X is elliptical, by the first part of the proof, X must be Gaussian. □

In fact, by Proposition 4.3 below, zeros in the inverse covariance matrix are not the only obstacle for the CI property. (see Appendix C for the proof).

Proposition 4.3. *If X has a multivariate t-distribution then X is not CI.*

We remark that Proposition 3.3 in Rüschendorf and Witting (2017) states that for an elliptical distribution $\Sigma \in \mathcal{IM}$ if and only if X is CI. Unfortunately, this result is not true as illustrated both by Theorem 4.2 and Proposition 4.3.

4.3. Positive elliptical distributions

Our results show that the CI/MTP$_2$ properties are too restrictive in connection with PGs. As a natural alternative, we study the following relaxation proposed by Agrawal, Roy and Uhler (2019).

Definition 4.4. An elliptically distributed $X \sim E(\mu, \Sigma)$ is positive if $\rho_{ij \cdot V \setminus \{i,j\}} \geq 0$ for all $i, j \in V$ (equiv. $\Sigma \in \mathcal{IM}$). A transelliptically distributed $Y \sim \text{TE}(\mu, \Sigma)$ is positive if the distribution of $f(Y) \sim E(\mu, \Sigma)$ is positive elliptical.

An elementary property of inverse M-matrices implies that for positive elliptical and transelliptical distributions the covariance matrix has nonnegative entries. From Remark 4.1 this notion of positivity is weaker than both MTP_2 and CI. Similarly, positivity does not imply association: for example if K is an M-matrix with a zero entry, then it is positive but not associated in general unless Gaussian. Thus, positive elliptical distributions satisfy a fairly weak notion of positive dependence, that is also simple and useful in applied modelling.

We first collect basic properties of this family of distributions.

Proposition 4.5. *If X has a positive (trans)elliptical distribution then the same is true for each marginal and each conditional distribution. Positive transelliptical distributions are also closed under strictly increasing transformations.*

The proofs of all the results in this section are given in Appendix B.

The next proposition shows that positive elliptical distributions retain some strong properties of MTP_2 Gaussian distributions.

Proposition 4.6. *If X has a positive elliptical distribution then for all $i \in V$ and $C \subseteq V \setminus \{i\}$ the conditional mean $\mathbb{E}(X_i | X_C)$ is an increasing function of X_C. Moreover, for any two $i, j \in V$ and $C \subseteq V \setminus \{i,j\}$ it holds that*

$$\text{corr}(X_i, X_j | X_C) \geq 0$$

and

$$\text{corr}(X_i, X_j | X_C) = 0 \quad \Longrightarrow \quad \text{corr}(X_i, X_j | X_D) = 0 \quad \text{for all } D \supseteq C.$$

These properties are pivotal in the interpretation and application of the classical positive dependence measures. Briefly, the first part says that for positive elliptical distributions conditional correlations are non-negative, regardless of what subset of variables one conditions upon. The second part says that if a covariance conditional on X_C is 0, then it remains 0 when conditioning upon larger sets. In particular, zero marginal correlation implies zero partial correlation, hence Simpson's paradox cannot occur.

The following result offers an extension of Theorem 3.3 to the positive case.

Proposition 4.7. *If $X \sim E(\mu, \Sigma)$ then $\Sigma \in \mathcal{IM}$ if and only if $\text{corr}(g(X_i), X_j | X_C) \geq 0$ for every $i, j \in V$, any increasing function $g : \mathbb{R} \to \mathbb{R}$, and any conditioning set $C \subseteq V \setminus \{i,j\}$.*

Many constraint-based structure learning algorithms, like the PC algorithm (Spirtes et al., 2000), assume that the dependence structure in the data-generating distribution reflects faithfully the graph. We say that the distribution of X is faithful to a graph G if we have that $X_i \perp\!\!\!\perp X_j | X_C$ if and only if C separates vertices i and j in G, for any $C \subseteq V \setminus \{i,j\}$. In words, any independence obtained by conditioning on subsets C is reflected in the graph. Spirtes et al. (2000)

extended this definition to partial correlation graphs: the distribution of X is *linearly faithful* to an undirected graph G if we have that $\mathrm{corr}(X_i, X_j | X_C) = 0$ if and only if C separates i and j in G. Bühlmann, Kalisch and Maathuis (2010) proposed a related convenient notion unrelated to any particular graph: the distribution of X is *partially faithful* if we have that $\mathrm{corr}(X_i, X_j | X_C) = 0$ for any $C \subset V \setminus \{i, j\}$ implies that $\mathrm{corr}(X_i, X_j | X_{(ij)}) = 0$. Using partial faithfulness Bühlmann, Kalisch and Maathuis (2010) developed a simplified PC algorithm that is computationally feasible even with thousands of variables. An important property of positive elliptical distributions is given by the following result.

Proposition 4.8. *Every positive elliptical distribution is partially faithful.*

Proof.. The proof follows from Proposition 4.6. $\qquad \square$

4.4. Characterisation of MTP_2 elliptical distributions

We finish our discussion of positive dependence for elliptical distributions with a complete characterization of MTP_2 distributions. In this section we assume that X admits a density with respect to the Lebesgue measure. In this case the density necessarily takes the form

$$p(x) \;=\; |\Sigma|^{-1/2}\, \varphi_d\Big((x-\mu)^T \Sigma^{-1}(x-\mu)\Big), \tag{7}$$

where φ_d is a nonnegative function (that may depend on d) called the density generator.

Theorem 3.1 in Sampson (1983) gives a necessary and sufficient condition for bivariate elliptical distributions to be MTP_2. The author admits that the multivariate case was analytically too complicated to handle. For the bivariate case a similar characterization has been provided in Proposition 1.2 of Abdous, Genest and Rémillard (2005). This second approach has been more useful for us and with a bit of matrix algebra their proof generalizes.

Theorem 4.9. *Suppose X has a d-dimensional elliptical distribution with partial correlations $\rho_{ij \cdot V \setminus \{i,j\}} \geq 0$ and suppose X admits a twice differentiable density function with the density generator $\varphi_d(t)$ for $t \geq 0$. Let $\rho_* = \min_{i,j} \rho_{ij \cdot V \setminus \{i,j\}}$ and let $\phi(t) = \log \varphi_d(t)$. Then X is MTP_2 if and only if $\phi'(t) \leq 0$ for all $t \geq 0$; $\phi'(t) = 0$ implies $\phi''(t) = 0$ for all $t > 0$; and*

$$-\frac{\rho_*}{1 + \rho_*} \;\leq\; \frac{t\phi''(t)}{\phi'(t)} \;\leq\; \frac{\rho_*}{1 - \rho_*}. \tag{8}$$

for all $t \in \mathbb{T} = \{t : \phi'(t) < 0\}$. In particular, $\inf_{t \in \mathbb{T}} \frac{t\phi''(t)}{\phi'(t)} > -\frac{1}{2}$.

The proof is in Appendix D. To illustrate Theorem 4.9 we consider three examples. For the d-dimensional t-distribution with k degrees of freedom $\phi(t) = -\frac{k+d}{2}\log(1 + \frac{t}{k})$ and so condition (8) requires that

$$-\frac{\rho_*}{1 + \rho_*} \leq -\frac{t}{k+t} \leq \frac{\rho_*}{1 - \rho_*}$$

must be satisfied for all $t \in \mathbb{R}$, which is impossible irrespective of $\rho_* \in (-1, 1)$. Further, for a zero-mean multivariate Laplace distribution $\frac{t\phi''(t)}{\phi'(t)} \in (-1, -\frac{1}{2})$ for any d, hence Laplace distributions are never MTP$_2$. As a third example, consider $X \sim E(\mu, K^{-1})$ with density $f(x) \propto e^{-(x^T K x)^\alpha}$ (the Kotz distribution sub-family), so that $t\phi''(t)/\phi'(t) = \alpha - 1$. From Theorem 4.9, if $\alpha > 1$ (X has thinner-than-Normal tails) then MTP$_2$ holds if and only if $\rho_* > 1 - 1/\alpha$, and similarly if $1/2 < \alpha < 1$ (thicker-than-Normal tails) then MTP$_2$ also holds if and only if $\rho_* \geq 1/\alpha - 1$. If $\alpha \leq 1/2$ then X cannot be MTP$_2$.

The constraints on possible ρ_* in this third example did not take into account one more important aspect of the problem, namely that K is a $d \times d$ positive definite matrix. To illustrate this, suppose that all off-diagonal entries of K are equal, that is, $\rho_{ij \cdot V \setminus \{i,j\}} = \rho_* = -K_{ij} > 0$ for all $i \neq j$. Such K is positive definite if and only if $\rho_* < 1/(d-1)$. In the third example this gives an upper bound on ρ_* that interplays with the lower bound $\rho_* \geq |1 - 1/\alpha|$. These two bounds define a non-empty set if and only if $|1 - 1/\alpha| < 1/(d-1)$. If $d = 2$ this holds for any $\alpha > 1/2$. If $d \geq 3$ this holds if and only if

$$1 - \frac{1}{d} \; < \; \alpha \; < \; 1 + \frac{1}{d-2}.$$

Note that unless $\alpha = 1$ (the Gaussian case), this condition cannot hold for all $d \in \mathbb{N}$. This simple example generalizes and yields the following characterization of elliptical families with a fixed density generator that contain MTP$_2$ distributions.

Theorem 4.10. *Consider the family of all elliptical distributions with density generator $\varphi_d(t)$ and let $\phi(t) = \log \varphi_d(t)$. Then, there exists a scale matrix parameter Σ such that the density (7) is* MTP$_2$ *if and only if $\phi'(t) \leq 0$; $\phi'(t) = 0$ implies $\phi''(t) = 0$; and*

$$-\frac{1}{d} \; < \; \frac{t\phi''(t)}{\phi'(t)} \; < \; \frac{1}{d-2}$$

for all $t \in \mathbb{T} = \{t : \phi'(t) < 0\}$.

The proof is in Appendix E. To illustrate this result consider the elliptically symmetric logistic distribution as defined in Fang (2018), Section 3.5. The density generator satisfies

$$\varphi_d(t) = \frac{e^{-t}}{(1 + e^{-t})^2}, \qquad \frac{t\phi''(t)}{\phi'(t)} = \frac{2t}{e^t - e^{-t}} \in (0, 1].$$

Theorem 4.9 gives that a bivariate logistic distribution is MTP$_2$ if and only if $\rho_{12} \geq 1/2$. However, if $d \geq 3$, Theorem 4.10 implies that there are no MTP$_2$ distributions of this form.

To summarize, our results show that although there are some non-normal elliptical distributions that are MTP$_2$, the imposed constraints can be quite severe, particularly as the dimension d grows. We additionally showed that popular

elliptical distributions such as the t, Laplace and most Kotz-type distributions cannot be MTP$_2$. These findings highlight the need to define alternative measures of positive association, such as the PPGs in Section 4.2.

5. Examples

We illustrate the application of transelliptical PG and PPG and the interpretation afforded by our characterizations with S&P500 stock market data. The data and R code to reproduce our analyses are available at https://github.com/davidrusi/paper_examples/tree/main/2021_Rossell_Zwiernik_elliptical. Section 5.1 discusses a measure of second-order marginal dependence and extend it to conditional dependence, Section 5.2 our data analysis results without sign constraints on K, and finally Section 5.3 shows that said constraints can improve the model fit.

5.1. Second-order dependence

Elliptical models capture certain forms of tail dependence, which can be naturally illustrated by studying second-order dependence. A typical application is in time series analysis, where it is common to study how the variance (or volatility) evolves over time. For example, a centered ARCH model (Engle, 1982) poses $X_i = \sigma_i \epsilon_i$, where $\epsilon_i = \sigma_i Z_i$, $Z_i \sim N(0,1)$ independently across i and $\sigma_i^2 = \alpha_0 + \alpha_1 \epsilon_{i-1}^2$. Under this ARCH model it is easy to show that $\mathrm{corr}(X_i, X_{i-1}) = 0$, whereas $\mathrm{corr}(X_i^2, X_{i-1}^2)$ is non-zero when $\alpha_1 \neq 0$. More generally, in our context let $X \sim E(0, \Sigma)$ and consider $\theta_{ij} = \mathrm{corr}(X_i^2, X_j^2)$ as a simple measure of marginal tail (or second-order) dependence. In the special sub-family of elliptical distributions defined by scale mixtures of normals (c.f. Section 2), it is possible to show

$$\theta_{ij} = \mathrm{corr}(X_i^2, X_j^2) \;=\; \lambda + (1-\lambda)\rho_{ij}^2, \qquad \lambda \;=\; \frac{\mathrm{var}(\frac{1}{\tau})}{\mathrm{var}(\frac{1}{\tau}) + 2\mathbb{E}(\frac{1}{\tau^2})} \in [0,1], \tag{9}$$

where τ is the latent scale parameter, hence $\lambda = 0$ if and only if X is Gaussian. This measure is minimized for $\rho_{ij} = 0$, then $\mathrm{corr}(X_i^2, X_j^2) = \lambda$, which can be non-negligible, e.g. $\lambda = 1/(k-1)$ for the t-distribution with $k > 4$ degrees of freedom. Figure 1 shows this quantity and, for comparison, also the normalized mutual information (a standard measure of deviation from independence). Both measures converge to zero as $k \to \infty$ but this convergence is slow.

Similarly, and of particular relevance to graphical models, one may measure conditional tail dependence via $\theta_{ij \cdot V \setminus \{i,j\}} =$

$$\mathrm{corr}\Big((X_i - \mathbb{E}(X_i \mid X_{(ij)}))^2, (X_j - \mathbb{E}(X_j \mid X_{(ij)}))^2 \Big| X_{(ij)}\Big) = \lambda + (1-\lambda)\rho_{ij \cdot V \setminus \{i,j\}}^2, \tag{10}$$

where the right-hand side follows from Proposition 2.1. When $\rho_{ij \cdot V \setminus \{i,j\}} = 0$ the conditional tail dependence is λ.
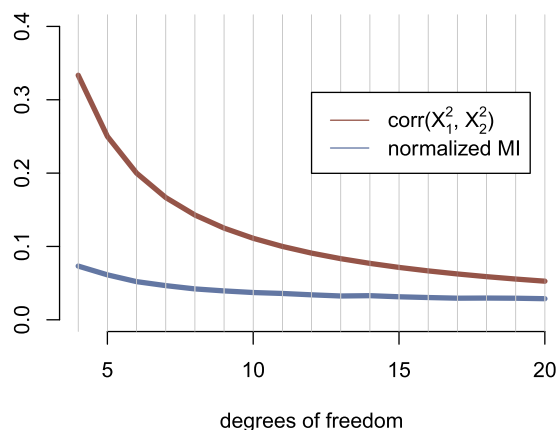
FIG 1. $\mathrm{corr}(X_i^2, X_j^2)$ *for a multivariate t-distribution with k degrees of freedom in the* $\rho_{ij} = 0$ *case*

5.2. Data analysis results

We downloaded daily log-returns of S&P500 stocks for the 10-year period ranging from 2010-04-29 to 2020-04-14 ($n = 2,514$ observations). For illustration we selected the first $d = 100$ stocks, hence the graphical model has 4,950 potential edges. We used the R package huge (Zhao et al., 2012) to apply univariate transformations aimed at improving the marginal normal fit (function `huge.npn`). Despite these transformations, we observed departures from multivariate normality. Let the observed and transformed $n \times d$ data matrices be Y and X (respectively), both with zero column means and unit variances. The empirical distribution of the Mahalanobis distances $(x_{i1}, \ldots, x_{id})^T S^{-1}(x_{i1}, \ldots, x_{id})$, where S is the sample covariance, had significantly thicker tails than the χ_d^2 expected under multivariate normal data and $S = \Sigma$ (Figure 2, top left). This departure from normality motivates considering non-Gaussian elliptical models.

Another argument for using non-Gaussian elliptical models comes from considering the second-order tail dependence measures from Section 5.1. Figure 2 (bottom left) shows that the marginal tail dependence $\hat{\theta}_{ij}$ in (9) is significantly larger than the $\hat{\rho}_{ij}^2$ expected under normality. The magnitude of these departures is practically significant. For comparison the figure also displays $\hat{\theta}_{ij}$ estimated from simulated Gaussian data, with zero mean and sample covariance matching that of X. In practical terms, θ_{ij} measures the predictability of variable i's variance (also called volatility) from that of variable j. A natural question is what predictability remains after conditioning upon other variables, i.e. what is the *conditional* tail dependence $\theta_{ij \cdot V \setminus \{i,j\}}$ in (10). To address this question for each variable pair (i, j) we computed the non-parametric estimate

$$\hat{\theta}_{ij \cdot V \setminus \{i,j\}} = \mathrm{corr}(e_i^2, e_j^2 \mid X_{(ij)})$$
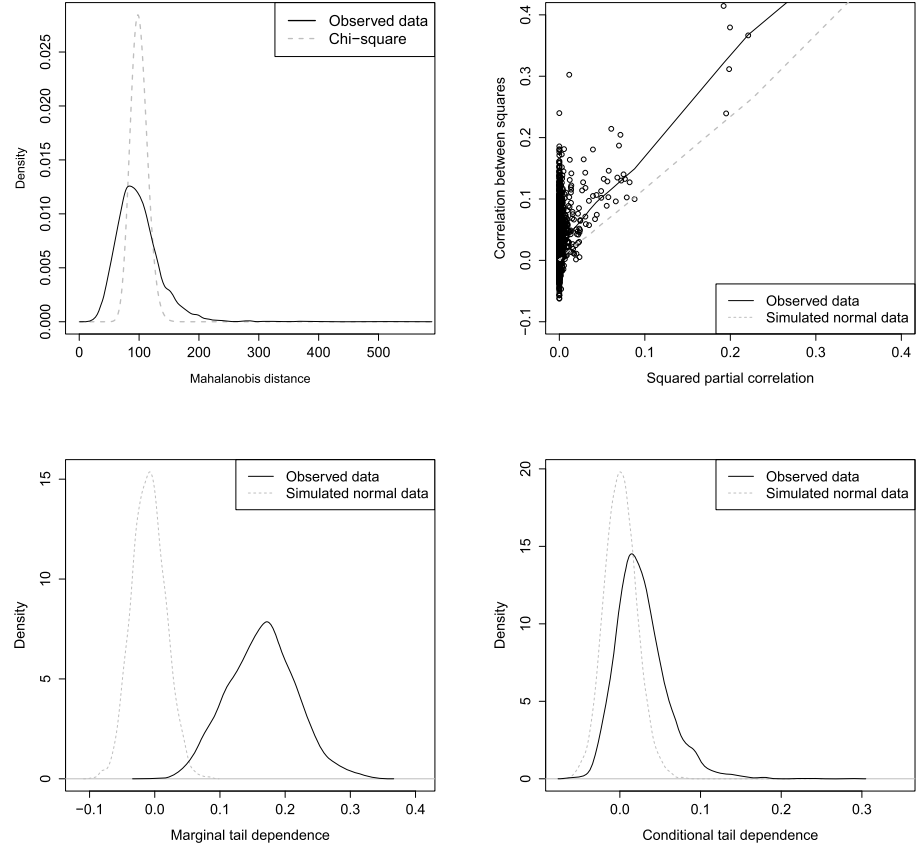
FIG 2. *S&P500 data. Top left: Mahalanobis distances and χ^2_{100} density. Top right: conditional tail dependence $\hat{\theta}_{ij \cdot V \setminus \{i,j\}}$ versus normal prediction $\hat{\rho}^2_{ij \cdot V \setminus \{i,j\}}$. Bottom: distribution of $\hat{\theta}_{ij} - \hat{\rho}^2_{ij}$ and $\hat{\theta}_{ij \cdot V \setminus \{i,j\}} - \hat{\rho}^2_{ij \cdot V \setminus \{i,j\}}$*

where $e_i = x_i - \hat{\mu}_{i|V \setminus \{i,j\}}$, x_i is the i-th column in X and $\hat{\mu}_{i|V \setminus \{i,j\}}$ the least-squares prediction given $X_{(ij)}$ (analogously for e_j). These estimates were significantly larger than the $\hat{\rho}^2_{ij \cdot V \setminus \{i,j\}}$ expected under normality (Figure 2, bottom right). As a further check, from (10) the elliptical model (more specifically, the scale mixture of normals sub-family) predicts $\theta_{ij|V \setminus \{i,j\}}$ to be linear in $\rho^2_{ij|V \setminus \{i,j\}}$. Figure 2 (top right) suggests that they are indeed roughly linearly related. Altogether, although admittedly one never expects a model to describe the data perfectly, the elliptical model appears reasonable to study volatility in these data.

We studied the dependence structure in these data using the two-step procedure outlined in Appendix A. Briefly, this *(trans)elliptical graph recovery method* obtains a sample correlation matrix S from Kendall's correlations and Lemma 3.6, and then applies a standard graphical LASSO procedure to S. In practice,

an alternative is to use a standard Gaussian approach where one estimates S with the sample correlation matrix (c.f Remark A.1), i.e. to ignore the elliptical nature of the data and trust that the Gaussian model recovers consistent estimates of Σ, even if X were truly elliptical. For completeness we considered both strategies. It turns out that in both cases we obtained similar results: the Spearman correlation between the estimated $\hat{K}_{ij}$ was 0.911, the selected PGs agreed in 93.0% of the 4,950 edges, and there were no disagreements in the signs of $\hat{K}_{ij}$ for any (i, j).

We focus discussion on the (trans)elliptical graph recovery method. The estimated partial correlation graph had 1,594 out of the 4,950 edges. Our results from Section 3 help strengthen the interpretation of the missing edges, e.g. $\hat{K}_{ij} = 0$ suggests that conditional on $X_{(ij)}$ one cannot predict the variance, asymmetry or kurtosis in x_j linearly from x_i. Further it also implies zero Kendall's conditional tau between increasing transforms of x_i and x_j, e.g. if daily returns are not conditionally positively/negative correlated (according to Kendall's tau) then neither are log-returns.

5.3. *Positivity constraints versus graphical LASSO constraints*

An interesting point is that among the 1,594 edges the estimated partial correlations were positive for $1,477$ and negative for only 117 edges. That is, the partial correlation graph was very close to being a PPG; see Agrawal, Roy and Uhler (2019) for a discussion why this may be frequently encountered in stock data, and Epskamp and Fried (2018); Lauritzen, Uhler and Zwiernik (2019a) for examples in Psychology. To compare the PPG fit with our earlier graphical LASSO fit we estimated the precision matrix under the constraint that K is an M-matrix, using the R package `golazo` available at GitHub (Lauritzen and Zwiernik, 2020)[1]. Table 1 summarizes the results. The maximized Gaussian log-likelihood (here interpreted as the negative of a loss function) under the M-matrix constraint was substantially higher than for the graphical LASSO fit ($-266{,}579$ versus $-269{,}195$) and the graph was sparser (1,216 versus 1,594 edges). Hence, the EBIC (and any other L_0 model selection criteria) strongly favored the PPG model over a model with no sign restrictions on K. These findings remained true when we replaced the graphical LASSO estimator with its refitted version, i.e. we computed the maximizer of the Gaussian log-likelihood conditional on the model chosen in the first step. In this case the Gaussian log-likelihood was $-265{,}277$, that is, higher than for the M-matrix estimate, but the BIC and EBIC criterion were worse.

Note that, from its Lagrangian interpretation, the graphical LASSO constrains the size $|\hat{K}_{ij}|$. In contrast the M-matrix constraint allows for arbitrarily large $|\hat{K}_{ij}|$, provided $\hat{K}_{ij} \leq 0$. That is, the graphical LASSO and the PPG constraints induce quite different regularization and the latter appears more

---

[1]The relevant function is `positive.golazo(S,rho=Inf)`, where S is estimated using the Kendall's τ.

TABLE 1
Comparison of the results for the graphical LASSO, refitted graphical LASSO, and the
M-matrix estimation. The model with lower value of BIC/EBIC is preferred.

	edges	log-likelihood	BIC	EBIC, $\gamma = 0.5$
gLASSO	1,594	-269,195	550,872	565,553
refitted gLASSO	1,594	-265,277	543,034	557,715
M-matrix	1,216	-266,579	542,679	553,879

appropriate for these S&P500 data, illustrating the potential value of positivity constraints in certain applications.

The selected graph being a PPG strengthens its interpretation. By Proposition 4.6, the finding suggests that all possible partial correlations are positive regardless of the conditioning set, and that Simpson's paradox does not occur in these data, i.e. stocks with zero marginal correlation also have zero partial correlation. By our earlier discussion, this implies that if $\rho_{ij} = 0$ marginally then x_i is uncorrelated with higher moments of x_j, both marginally and conditionally on $X_{(ij)}$. Further, the conditional expectation of x_i can only be increasing as a function of other variables (or increasing transformations thereof), and missing edges indicate the lack of such association.

6. Discussion

When studying multivariate dependence in applications it is often convenient to strike a balance between models that equip strong theoretical properties (e.g. Gaussian, non-paranormal, MTP_2 and CI classes) but impose potentially restrictive conditions, and models that are more flexible but do not provide such strong characterizations and/or lead to complex interpretations. We studied a natural strategy based on the transelliptical family and partial correlation graphs, including many copula models that are popular in applications. We showed that the interpretation remains simple yet goes far beyond the regular linear dependence. We also illustrated how simple tail dependence measures, like the one in (9), characterize the Gaussian distribution within the scale mixture of normals family and can help assess whether transelliptical class is useful to capture second-order dependence (variance) dependencies in the data.

A main motivation for our paper was to study the consequences of mild misspecification in Gaussian graphical models. Namely, when one assumes a Gaussian model but data are truly generated by a non-Gaussian distribution F_0. Despite misspecification, under suitable conditions one can recover the partial correlations under F_0. For general F_0 such partial correlations only inform about linear dependence but, if F_0 is trans-elliptical, we showed that much stronger interpretations are possible. Another question of interest is studying the consequences of assuming data to be trans-elliptical, when truly F_0 is not trans-elliptical. In that case, it is often still possible to consistently learn the true partial correlation matrix under F_0, for example using the log-determinant loss function as was done in Section 5; see Ravikumar et al. (2011) for some theoretical justification. However, for general F_0 such partial correlations only

inform about linear dependence.

An important part of this paper is the study of positive dependence. The notion of positivity can be quite useful in regularizing inference relative to unrestricted penalized likelihood, as we illustrated in the S&P500 example. However, we also showed that strictly speaking some standard notions of positive dependence are meaningless for structural learning in elliptical partial correlation graphs. One of our important contributions is a result that characterizes MTP_2 elliptical distributions and shows that MTP_2 becomes very restrictive in high dimensions. It is therefore important to study relaxations such as positive elliptical distributions that impose all partial correlations to be nonnegative. We showed that this family retains strong positive dependence properties that are important from the applied point of view.

In conclusion, we hope that our results help to motivate the study of other suitable relaxations of Gaussianity and positivity in graphical models, as well as strengthen the use of transelliptical graphical models in practice.

Appendix A: Graph estimation

We briefly describe a standard approach for high-dimensional elliptical data suggested by Ravikumar et al. (2011); Zhao and Liu (2014)[2]. For the elliptical model the support of the inverse covariance matrix is estimated as follows:

1. Estimate the underlying correlation matrix of X by computing sample Kendall's τ coefficients $\tau(X_i, X_j)$ and the formula in Lemma 3.6 to obtain estimated Pearson correlations $\hat{\rho}(X_i, X_j)$. Denote the resulting estimate of the correlation matrix by S, with entries $s_{ij} = \hat{\rho}(X_i, X_j)$.
2. Solve the graphical LASSO problem (Friedman, Hastie and Tibshirani, 2008; Yuan and Lin, 2007) by minimizing

$$- \log \det K + \text{tr}(SK) + \rho \sum_{i \neq j} |K_{ij}| \tag{11}$$

for a given penalty parameter $\rho > 0$.

The procedure in Step 1 can be performed with option `npn.func = ''skeptic''` in function `huge.npn`, see Liu, Han and Zhang (2012) for details. Another popular choice to obtain S in Step 1 is given by computing Tyler's M-estimator (Tyler, 1987) and normalizing it to be a correlation matrix. However, Zhao and Liu (2014) argued that Kendall's tau has better empirical behavior for heavy-tailed elliptical distributions and so we stick to this option.

We set the regularization parameter in Step 2 via the EBIC (Chen and Chen (2008); Foygel and Drton (2010), function `huge.select`). The problem (11) is convex in K and so the solution is unique. Its existence follows from (Lauritzen

---

[2]The aim of Zhao and Liu (2014) was to estimate the inverse covariance matrix and not just its support. Thus, estimating the variances was an important intermediate step. In our case this is not relevant.

and Zwiernik, 2020, Theorem 8.7). There are several coordinate descent algorithms that are guaranteed to converge and we use interchangeably the one implemented in the R package `huge` and the GOLAZO algorithm of Lauritzen and Zwiernik (2020).

It is important to observe that the same procedure can be used for transelliptical data as Kendall's τ is invariant under monotone transformations (Liu, Han and Zhang, 2012). This means that, for both the elliptical and the transelliptical model, we get exactly the same estimates and only the interpretation of zeros in the inverse covariance matrix changes. We also note that one may replace the graphical LASSO estimator in Step 2 with the CLIME estimator of Cai, Liu and Luo (2011).

Remark A.1. The standard graphical LASSO approach for Gaussian data differs from the above approach only by the fact that S in Step 1 is the sample correlation matrix.

Fitting totally positive elliptical distributions can be done in a straightforward way. We simply replace the penalized loss (11) in Step 2 and minimize $-\log\det(K) + \operatorname{tr}(SK)$ over the set of all positive definite matrices K satisfying $K_{ij} \leq 0$ for all $i \leq j$. Here, by (Slawski and Hein, 2015, Theorem 1), the optimum exists (and is unique) with probability one as long as the sample size is at least 2. This optimization can be performed iteratively using any of the procedures described in Slawski and Hein (2015); Lauritzen, Uhler and Zwiernik (2019b); Lauritzen and Zwiernik (2020). Here we use the latter because it scales better to large dimensions.

Appendix B: Proofs of Section 4.3

Proof of Proposition 4.5. If $X \sim E(\mu, \Sigma)$ then, by Proposition 2.1, for every $I \subset V$, $X_I \sim E(\mu_I, \Sigma_{II})$. If $\Sigma \in \mathcal{IM}$ then $\Sigma_{II} \in \mathcal{IM}$ by Johnson and Smith (2011), Corollary 2.3.2. Similarly, $\Sigma \in \mathcal{IM}$ then $\Sigma_{II} - \Sigma_{IJ}\Sigma_{JJ}^{-1}\Sigma_{JI} \in \mathcal{IM}$ by Johnson and Smith (2011), Corollary 2.3.1, proving that the conditional distribution of X_I given X_J is a positive elliptical distribution. The same argument after replacing X with $f(Y)$ works for transelliptical distributions. The last statement follows directly from the definition of transelliptical distributions. □

Proof of Proposition 4.6. These results are well known for Gaussian MTP_2 distributions; c.f. Fallat et al. (2017). It is convenient to translate them to equivalent statements in terms of Σ; c.f. (Drton, Sturmfels and Sullivant, 2009, Proposition 3.1.13). The statement about the conditional mean and the first statement about conditional correlations follow from the fact that $\mathcal{IM}$-matrices are closed under taking principal submatrices; c.f. Johnson and Smith (2011), Corollary 2.3.2. In consequence, for all $i, j \in V$ and $C \subseteq V \setminus \{i, j\}$ it holds that $(\Sigma_{C \cup \{i,j\}, C \cup \{i,j\}})_{ij}^{-1} \leq 0$. The last part states that if $\det \Sigma_{C \cup \{i\}, C \cup \{j\}} = 0$ for some $C \subseteq V \setminus \{i, j\}$ then $\det \Sigma_{D \cup \{i\}, D \cup \{j\}} = 0$ for every $D \supseteq C$. This statement is given in (Johnson and Smith, 2011, Theorem 3.3). □

Proof of Proposition 4.7. Since the left implication is obvious (take g identity, $C = V \setminus \{i, j\}$) we focus on the right implication. Let $A = C \cup \{i\}$. By the law of total expectation

$$\begin{aligned}
\text{cov}(g(X_i), X_j | X_C) &= \mathbb{E}_{X_i | X_C} \left[\mathbb{E} \Big((g(X_i) - \mathbb{E}g(X_i))(X_j - \mu_j) \mid X_i, X_C \Big) \right] \\
&= \mathbb{E}_{X_i | X_C} \left[(g(X_i) - \mathbb{E}g(X_i)) \mathbb{E} \Big(X_j - \mu_j \mid X_A \Big) \right] \\
&= \mathbb{E}_{X_i | X_C} \left[(g(X_i) - \mathbb{E}g(X_i)) \tilde{h}(X_A) \right].
\end{aligned}$$

where we denote $\tilde{h}(X_A) = \mathbb{E}(X_j - \mu_j | X_A) = \Sigma_{j,A} \Sigma_A^{-1}(X_A - \mu_A)$. Denote $\tilde{K} = (\Sigma_{A \cup \{j\}})^{-1}$, which is an M-matrix by Johnson and Smith (2011), Corollary 2.3.2. Using that $\tilde{K}$ is an M-matrix, expression (4) gives that all entries in the vector $\Sigma_{j,A} \Sigma_A^{-1}$ are non-negative, hence $\tilde{h}(X_A)$ is a non-decreasing function of X_A. Further, note that for any fixed X_C we can write $\tilde{h}(X_A)$ as a function of X_i only. Denote this function by $h_{X_C}(X_i)$. We thus have

$$\text{cov}(g(X_i), X_j | X_C) = \text{cov}(g(X_i), h_{X_C}(X_i) | X_C),$$

where now both g and h are nondecreasing functions of X_i. By Property 3 in Esary, Proschan and Walkup (1967) it follows that $\text{cov}(g(X_i), h_{X_C}(X_i) | X_C) \geq 0$. □

Appendix C: Proof of Proposition 4.3

Since both the CI property and t-distribution are closed under taking margins, it is enough to show that no bivariate t-distribution is conditionally increasing. Suppose (X_1, X_2) has bivariate t-distribution with k degrees of freedom. Without loss of generality assume that the mean is zero and that the scale matrix Σ satisfies $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = \rho$. By Remark 4.1, necessarily $\rho \geq 0$. Moreover, if $\rho = 0$ the statement follows from Theorem 4.2 so assume $\rho > 0$. The conditional distribution of X_1 given $X_2 = x_2$ is a t-Student distribution with $k^* = k + 1$ degrees of freedom, $\mu^* = \rho x_2$, and scale parameter

$$\sigma^* = \sqrt{\frac{1 - \rho^2}{k + 1}(k + x_2^2)}$$

(c.f. Section 5 in Roth (2012)). To show that (X_1, X_2) is not CI we provide an increasing function h for which $\mathbb{E}(h(X_1) | X_2 = x_2)$ is not increasing in x_2. Let $h(x_1) = \mathbb{1}_{[k, +\infty)}(x_1)$, so that $\mathbb{E}(h(X_1) | X_2 = x_2) = 1 - F_{X_1 | X_2}(k | x_2)$, where $F_{X_1 | X_2}$ is the conditional cumulative distribution function (c.d.f.). Using the formula (Johnson, Kotz and Balakrishnan, 1994, (28.4a)) for the c.d.f. of the t-Student distribution, if $\mu^* \leq k$ (or equiv. $x_2 \leq k/\rho$), we express $\mathbb{E}(h(X_1) | X_2 = x_2)$ in terms of the incomplete beta function

$$\mathbb{E}(h(X_1) | X_2 = x_2) = \frac{1}{2} \cdot I_{\alpha(x_2)} \left(\frac{k + 1}{2}, \frac{1}{2} \right) \qquad \text{for} \quad x_2 \leq \frac{k}{\rho},$$

where
$$\alpha(x_2) \;=\; \frac{k+1}{\left(\frac{k-\mu^*}{\sigma^*}\right)^2 + k + 1} \;=\; \frac{1}{1 + \frac{(k-\rho x_2)^2}{(k+x_2^2)(1-\rho^2)}} \;\in\; (0,1).$$

Using the definition of the incomplete beta function in terms of the beta function we get that for a positive constant C and $x_2 \le k/\rho$

$$\mathbb{E}(h(X_1)|X_2 = x_2) = C \int_0^{\alpha(x_2)} t^{(k-1)/2}(1-t)^{-1/2}\mathrm{d}t.$$

Since the integral above is strictly increasing in $\alpha(x_2)$, to show that $\mathbb{E}(h(X_1)|X_2 = x_2)$ is not increasing, it is enough to show that $\alpha(x_2)$ is not an increasing function for $x_2 \le k/\rho$. But direct calculations show

$$\alpha'(-\rho) = 0, \qquad \alpha''(-\rho) = \frac{2k(1-\rho^2)}{(k+1)^2(k+\rho^2)} > 0$$

showing that α is strictly decreasing for all $x_2 \le -\rho$ in some neighborhood of $-\rho$.

Appendix D: Proof of Theorem 4.9

Without loss of generality assume X has mean zero and $K = \Sigma^{-1}$ satisfies $K_{11} = \cdots = K_{dd} = 1$. In this case $K_{ij} = -\rho_{ij \cdot V \setminus \{i,j\}}$ for all $i \ne j$. If X admits a strictly positive and twice differentiable density function $f(x)$ then X is MTP$_2$ if and only if for every $1 \le i < j \le d$

$$\frac{\partial^2}{\partial x_i \partial x_j} \log f(x) \;\ge\; 0 \qquad \text{for all } x \in \mathbb{R}^d.$$

This result, found for example in Bach (2019) can be proved by elementary means, for example, by applying a second-order mean value theorem (Theorem 9.40 in Rudin (1964)). In our case $f(x) = |\Sigma|^{-1/2}\varphi_d(x^T K x)$ so f is MTP$_2$ if and only if for every $1 \le i < j \le d$

$$\frac{\partial^2}{\partial x_i \partial x_j}\phi(x^T K x) \;\ge\; 0 \qquad \text{for all } x \in \mathbb{R}^d. \tag{12}$$

Basic calculus gives $\nabla\phi(x^T K x) \;=\; 2\,\phi'(x^T K x)K x$ and

$$\nabla^2\phi(x^T K x) \;=\; 2\,\phi'(x^T K x)K + 4\,\phi''(x^T K x)K x x^T K.$$

If $x \ne 0$ we perform a change of coordinates to t and $v = \frac{1}{\sqrt{t}}K x$ constrained by $t > 0$ and $v^T \Sigma v = 1$. We extend it to the whole $\mathbb{R}^d$ by mapping $x = 0$ to $t = 0$ and $v = 0$. In the new coordinate system, condition (12) holds if and only if for all $1 \le i < j \le d$

$$2\phi'(t)K_{ij} + 4\phi''(t)\,t\,v_i v_j \;\ge\; 0 \qquad \text{for all } t, v. \tag{13}$$

Taking v such that $v_i = 0$ (which includes the $x = 0$ case), (13) implies that $\phi'(t) \leq 0$ for all $t \geq 0$, which gives the first condition. If $\phi'(t) = 0$ and $t > 0$, (13) cannot hold for all v unless $\phi''(t) = 0$, which gives the second condition in the theorem. Now suppose t is such that $\phi'(t) < 0$ then (13) becomes

$$2v_i v_j \frac{t\phi''(t)}{\phi'(t)} \leq -K_{ij}. \tag{14}$$

To study the bounds on $2v_i v_j$ subject to $v^T \Sigma v = 1$ we define the Lagrangian

$$L(v, \lambda) = 2v_i v_j - \lambda(v^T \Sigma v - 1).$$

Denote $A = \{i, j\}$ and $B = V \setminus A$. The Lagrangian condition $\nabla_v L(v, \lambda) = 0$ can be then reduced to $v_B = -\Sigma_{BB}^{-1} \Sigma_{BA} v_A$ and

$$\begin{bmatrix} v_j \\ v_i \end{bmatrix} = \lambda(\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}) \begin{bmatrix} v_i \\ v_j \end{bmatrix} = \lambda K_{AA}^{-1} \begin{bmatrix} v_i \\ v_j \end{bmatrix}.$$

Multiplying both sides by K_{AA} we get

$$\begin{bmatrix} 1 & K_{ij} \\ K_{ij} & 1 \end{bmatrix} \begin{bmatrix} v_j \\ v_i \end{bmatrix} = \lambda \begin{bmatrix} v_i \\ v_j \end{bmatrix}.$$

All stationary points must then satisfy $v_i^2 = v_j^2$. The maximal value of $2v_i v_j$ subject to $v^T \Sigma v = 1$ is $2\alpha^2$ obtained at a point where $v_i = v_j = \alpha$. The value of α can be found by noting that

$$v^T \Sigma v = \alpha^2 \mathbf{1}^T (\Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}) \mathbf{1} = \frac{2\alpha^2}{1 + K_{ij}},$$

where $\mathbf{1}$ is the vector of ones. Since $v^T \Sigma v = 1$, $2\alpha^2 = 1 - \rho_{ij \cdot V \setminus \{i,j\}}$. In a similar way we show that the minimal value of $2v_i v_j$ is $-(1 + \rho_{ij \cdot V \setminus \{i,j\}})$. This gives that (14) is equivalent to

$$-\frac{\rho_{ij \cdot V \setminus \{i,j\}}}{1 + \rho_{ij \cdot V \setminus \{i,j\}}} \leq \frac{t\phi''(t)}{\phi'(t)} \leq \frac{\rho_{ij \cdot V \setminus \{i,j\}}}{1 - \rho_{ij \cdot V \setminus \{i,j\}}}.$$

This inequality must be satisfied for every $i \neq j$. However the functions $\rho/(1 + \rho)$ and $\rho/(1 - \rho)$ are increasing for $\rho \in [0, 1)$ and so $\min_{ij} \rho_{ij \cdot V \setminus \{i,j\}}/(1 - \rho_{ij \cdot V \setminus \{i,j\}}) = \rho_*/(1 - \rho_*)$ and $\min_{ij} \rho_{ij \cdot V \setminus \{i,j\}}/(1 + \rho_{ij \cdot V \setminus \{i,j\}}) = \rho_*/(1 + \rho_*)$. Thus we arrive at (8).

Now suppose that ϕ is such that $\phi'(t) \leq 0$; $\phi'(t) = 0$ implies that $\phi''(t) = 0$; and (8) holds for all $t \in \mathbb{T}$. By reversing the argument above we conclude that (13) holds for all $t \in \mathbb{T}$. For all the remaining t this inequality also holds because then both sides are equal to zero. However, as we argued before (13) holds for all t if and only if X is MTP$_2$. This concludes our proof.

Appendix E: Proof of 4.10

Let $\beta_* = \inf_{t \in \mathbb{T}} \frac{t\phi''(t)}{\phi'(t)}$ and $\beta^* = \sup_{t \in T} \frac{t\phi''(t)}{\phi'(t)}$. First suppose that for some $K = \Sigma^{-1}$ the underlying elliptical distribution is MTP$_2$. By Theorem 4.9 it follows that $\phi'(t) \leq 0$ and $\phi''(t) = 0$ whenever $\phi'(t) = 0$. Assume without loss that K is normalized to have ones on the diagonal so that $K_{ij} = -\rho_{ij \cdot V \setminus \{i,j\}} \leq 0$ for $i \neq j$. The fact that partial correlations must be necessarily nonnegative follows from Remark 4.1. Since K is an M-matrix, by Proposition 6.1 in Lauritzen, Uhler and Zwiernik (2019b) we can shrink each off-diagonal entry towards zero preserving positive-definitedness. In particular, the matrix $K' = (1 + \rho_*)\mathbb{I} - \rho_* \mathbf{1}\mathbf{1}^T$ obtained from K by replacing each off-diagonal entry with $-\rho_*$, where $\rho_* = \min_{i,j} \rho_{ij \cdot V \setminus \{i,j\}}$, must be positive definite. This matrix is positive definite if and only if $\rho_* < 1/(d-1)$, which gives an upper bound on ρ_* on the top of the two upper bounds implied by Theorem 4.9, namely, $\rho_* \geq \beta^*/(1 + \beta^*)$ and $\rho_* \geq -\beta_*/(\beta_* + 1)$ (use the fact that $\beta_* > -1/2$ by the last part of Theorem 4.9). The intersection of these three constraints is non-empty if and only if $\frac{1}{d-1} > \max\{\frac{\beta^*}{1+\beta^*}, -\frac{\beta_*}{\beta_*+1}\}$. In other words, $\beta^* < 1/(d-2)$ and $\beta_* > -1/d$, which finished the proof of one implication.

Now suppose $\phi'(t) \leq 0$ and $\phi''(t) = 0$ whenever $\phi'(t) = 0$. If, in addition, the third condition in the theorem is satisfied then $\beta_* > -1/d$ and $\beta^* < 1/(d-2)$. Let K be a matrix with ones on the diagonal and $-\rho_*$ on the remaining entries. If $0 \leq \rho_* < 1/(d-1)$ then K is positive definite. If $\rho_* \nearrow 1/(d-1)$ then $\rho_*/(1 - \rho_*) \nearrow 1/(d-2)$ and $-\rho_*/(1 + \rho_*) \searrow -1/d$. Since $\beta_* > -1/d$ and $\beta^* < 1/(d-2)$ we can always find ρ_* close enough to $1/(d-1)$ so that

$$-\frac{\rho_*}{1 + \rho_*} \leq \beta_* \leq \beta^* \leq \frac{\rho_*}{1 - \rho_*}.$$

By Theorem 4.9 the corresponding distribution is MTP$_2$.

Acknowledgments

References

ABDOUS, B., GENEST, C. and RÉMILLARD, B. (2005). Dependence properties of meta-elliptical distributions. In *Statistical Modeling and Analysis for Complex Data Problems* 1–15. Springer. MR2189528

AGRAWAL, R., ROY, U. and UHLER, C. (2019). Covariance Matrix Estimation under Total Positivity for Portfolio Selection. *arXiv preprint arXiv:1909.04222*.

BABA, K., SHIBATA, R. and SIBUYA, M. (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics* **46** 657–664. MR2115961

BACH, F. (2019). Submodular functions: from discrete to continuous domains. *Mathematical Programming* **175** 419–459. MR3942896

BARBER, R. F. and KOLAR, M. (2018). Rocket: Robust confidence intervals via Kendall's tau for transelliptical graphical models. *The Annals of Statistics* **46** 3422–3450. MR3852657

BEHROUZI, P. and WIT, E. C. (2019). Detecting epistatic selection with partially observed genotype data by using copula graphical models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **68** 141–160. MR3902986

BILODEAU, M. (2014). Graphical lassos for meta-elliptical distributions. *Canadian Journal of Statistics* **42** 185–203. MR3208335

BOWMAN, A. and FOSTER, P. (1993). Density based exploration of bivariate data. *Statistics and Computing* **3** 171–177.

BÜHLMANN, P., KALISCH, M. and MAATHUIS, M. H. (2010). Variable selection in high-dimensional linear models: partially faithful distributions and the PC-simple algorithm. *Biometrika* **97** 261–278. MR2650737

CAI, T., LIU, W. and LUO, X. (2011). A constrained $\ell 1$-minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association* **106** 594–607.

CHEN, J. and CHEN, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95** 759-771. MR2443189

COLANGELO, A., SCARSINI, M. and SHAKED, M. (2005). Some notions of multivariate positive dependence. *Insurance: Mathematics and Economics* **37** 13–26. MR2156593

DRTON, M., STURMFELS, B. and SULLIVANT, S. (2009). *Lectures on Algebraic Statistics* **39**. Birkhäuser Basel. MR2723140

ENGLE, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society* 987-1007.

EPSKAMP, S. and FRIED, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological methods* **23** 617.

ESARY, J. D., PROSCHAN, F. and WALKUP, D. W. (1967). Association of random variables, with applications. *Ann. Math. Stat.* **38** 1466–1474. MR0217826

FALLAT, S., LAURITZEN, S. L., SADEGHI, K., UHLER, C., WERMUTH, N. and ZWIERNIK, P. (2017). Total positivity in Markov structures. *Annals of Statistics* **45** 1152–1184. MR3662451

FANG, K. W. (2018). *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC.

FANG, H.-B., FANG, K.-T. and KOTZ, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis* **82** 1–16. MR1918612

FELLER, W. (1971). *An Introduction to Probability Theory and Applications* **2**,

second ed. John Wiley & Sons, New York. MR0270403

Finegold, M. A. and Drton, M. (2009). Robust graphical modeling with t-distributions. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* 169–176.

Foygel, R. and Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. In *Advances in neural information processing systems* 604–612.

Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.

Hamada, M., Valdez, E. et al. (2004). *CAPM and option pricing with elliptical distributions*. School of Finance and Economics, University of Technology, Sydney.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions* **2**, Second ed. Wiley New York. MR1326603

Johnson, C. R. and Smith, R. L. (2011). Inverse M-matrices, II. *Linear Algebra and its Applications* **435** 953–983. MR2807211

Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multiv. Anal.* **10** 467–498. MR599685 (83j:60020a)

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A* 419–430. MR0287628

Landsman, Z. and Nešlehová, J. (2008). Stein's Lemma for elliptical random vectors. *Journal of Multivariate Analysis* **99** 912-927.

Lauritzen, S. L. (1996). *Graphical Models.* Clarendon Press, Oxford, United Kingdom. MR1419991

Lauritzen, S., Uhler, C. and Zwiernik, P. (2019a). Total positivity in exponential families with application to binary variables. *To appear in the Annals of Statistics.* MR4298870

Lauritzen, S., Uhler, C. and Zwiernik, P. (2019b). Maximum likelihood estimation in Gaussian models under total positivity. *Annals of Statistics* **47** 1835–1863. MR3953437

Lauritzen, S. and Zwiernik, P. (2020). Locally associated graphical models. *arXiv preprint arXiv:2008.04688.*

Lindskog, F., Mcneil, A. and Schmock, U. (2003). Kendall's tau for elliptical distributions. In *Credit Risk* 149–156. Springer.

Liu, H., Han, F. and Zhang, C.-h. (2012). Transelliptical graphical models. In *Advances in Neural Information Processing Systems* 800–808.

Liu, H., Han, F., Yuan, M., Lafferty, J. and Wasserman, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *The Annals of Statistics* **40** 2293–2326. MR3059084

Müller, A. and Scarsini, M. (2001). Stochastic comparison of random vectors with a common copula. *Mathematics of Operations Research* **26** 723–740. MR1870742

Newman, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. *Lecture Notes-Monograph*

Series 127–140. MR0789244

RAVIKUMAR, P., WAINWRIGHT, M. J., RASKUTTI, G. and YU, B. (2011). High-dimensional covariance estimation by minimizing $\ell 1$-penalized log-determinant divergence. *Electronic Journal of Statistics* **5** 935–980.

ROBEVA, E., STURMFELS, B., TRAN, N. and UHLER, C. (2018). Maximum Likelihood Estimation for Totally Positive Log-Concave Densities. *Scandinavian Journal of Statistics*.

ROTH, M. (2012). *On the multivariate t distribution*. Linköping University Electronic Press.

RUDIN, W. (1964). *Principles of Mathematical Analysis* **3**. McGraw-hill New York. MR0166310

RÜSCHENDORF, L. and WITTING, J. (2017). VaR bounds in models with partial dependence information on subgroups. *Dependence Modeling* **5** 59–74. MR3635787

SAMPSON, A. R. (1983). Positive dependence properties of elliptically symmetric distributions. *Journal of Multivariate Analysis* **13** 375–381.

SLAWSKI, M. and HEIN, M. (2015). Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. *Linear Algebra and its Applications* **473** 145–179. MR3338330

SPIRTES, P., GLYMOUR, C. N., SCHEINES, R. and HECKERMAN, D. (2000). *Causation, prediction, and search*. MIT press, Cambridge, MA, USA. MR1815675

STEPHENS, M. (2013). A unified framework for association analysis with multiple related phenotypes. *PloS one* **8**.

SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics* 1236–1265. MR2752127

TYLER, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *The annals of Statistics* 234–251. MR0885734

VINCIOTTI, V. and HASHEM, H. (2013). Robust methods for inferring sparse network structures. *Computational Statistics & Data Analysis* **67** 84–94. MR3079589

VOGEL, D. and FRIED, R. (2011). Elliptical graphical modelling. *Biometrika* **98** 935–951. MR2860334

YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35.

ZHAO, T. and LIU, H. (2014). Calibrated precision matrix estimation for high-dimensional elliptical distributions. *IEEE transactions on Information Theory* **60** 7874–7887. MR3285751

ZHAO, T., LIU, H., ROEDER, K., LAFFERTY, J. and WASSERMAN, L. (2012). The huge package for high-dimensional undirected graph estimation in R. *Journal of Machine Learning Research* **13** 1059–1062. MR2930633