Chilean Journal of Statistics
Vol. 3, No. 2, September 2012, 181–192

Probabilistic and Inferential Aspects of Skew-Symmetric Models
Special Issue: "IV International Workshop in honour of Adelchi Azzalini's 60th Birthday"

Selection of conditional independence graph models when the distribution is extended skew-normal

Simona Pacillo[1,*]
[1]Department SEGIS, University of Sannio, Benevento, Italy

(Received: 11 August 2011 · Accepted in final form: 26 June 2012)

Abstract

The extended skew-normal family of distributions is a slight extension of the skew-normal one, which achieves closure under conditioning. In this paper, we discuss its application in conditional independence graphs selection. First, we derive a test for a single edge exclusion/inclusion based on a Wald-type statistic. Then, we show how the asymptotic null distribution of the Wald test changes when some regularity conditions of the parameter space fail to hold. Finally, we propose an alternative test and carry out numerical experiments to assess the performances, in finite samples, of the two methods.

Keywords: Edge inclusion/exclusion test · Graphical model · Regularity conditions · Skew normal.

Mathematics Subject Classification: Primary 62F03 · Secondary 60E05.

1. Introduction

Graphical models (see Lauritzen, 1996; Whittaker, 1990) are a family of probability distributions for a d-dimensional multivariate random variable Y whose independence structure is characterised by a conditional independence graph $G = (V, E)$. Here, each variable of Y is identified with a vertex $V = \{1, \ldots, d\}$ and the absence of an edge between two vertices represents conditional independence.

When $Y \sim \mathrm{N}_d(\mu, \Sigma)$, we have a Gaussian graphical model or a covariance-selection model; see Dempster (1972). In this context, the covariance matrix Σ is restricted by its Markov properties and the variables Y_i and Y_j are conditionally independent given the others, $Y_i \perp\!\!\!\perp Y_j |\text{rest}$, if and only if, $\Sigma^{ij} = 0$, where Σ^{ij} indicates the (i, j)th entry of the concentration matrix Σ^{-1}. Moreover, Σ^{-1} satisfies also the linear restriction

$$e^{ij} = 0 \quad \Rightarrow \quad \Sigma^{ij} = 0, \tag{1}$$

where e^{ij} represents the (i, j) link in the edge set E.

*Corresponding author. Simona Pacillo, Statistics Department SEGIS, University of Sannio, Piazza Arechi II (palazzo De Simone) 82100, Benevento, Italy. Email: simona.pacillo@unisannio.it

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

An interesting application of the graphical models is the possibility of selecting, through an appropriate procedure, the graph underlying the data. The standard approach to model selection, the backward stepwise method (see Edwards, 2000) poses an implicit problem in multiple hypothesis testing, because it considers the exclusion of a single edge at time from the complete graph. More precisely, for each of the possible couples of vertices $\{i, j\}$, a test for conditional independency is performed and the procedure is repeated until no further edge is removed.

As in the normal context the relation given in Equation (1) implies also that the corresponding partial correlation coefficient, $\rho_{ij|\text{rest}}$, is equal to zero, the edge deletion can be verified by testing

$$H_0\colon \rho_{ij|\text{rest}} = 0 \quad \text{versus} \quad H_1\colon \rho_{ij|\text{rest}} \neq 0.$$

When carrying out a statistical analysis of data, departures from normality may be encountered, for example due to lack of symmetry. In this case, the previous model selection procedure is not feasible because 'out of the Gaussian context' non-correlation does not imply independence.

In this paper, we propose the adoption of the extended skew-normal (ESN) distribution introduced by Azzalini (1985) and studied by Capitanio et al. (2003) to cope with skewness in the data. Within this context, we present a test for a single edge inclusion/exclusion in conditional independence graphs. The properties of the ESN distribution are briefly reviewed in Section 2. Section 3 extends the model selection procedure based on the Wald test methodology and described in Capitanio and Pacillo (2008), by providing additional aspects about the study of the derivative of the scoring function. Section 4 shows some problems which may arise when some regularity conditions of the parameter space fail to hold and proposes an alternative method to deal with this specific circumstance. Results from numerical simulations in finite samples are presented in Section 5. Section 6 offers some conclusions and the Appendix provides some results for the ESN log-likelihood function.

2. Skew Normal Conditional Independence Graphs

When departures from normality are due to a lack of symmetry, the class of skew-normal (SN) distributions defined by Azzalini and Dalla Valle (1996) provides a useful model to represent the data. This family extends the Gaussian distribution by adding a vector parameter α which regulates the skewness: the normal model is obtained as a special case when $\alpha = 0$. The SN distribution allows us to carry out inference based on the likelihood function, while dealing with skewness, and enjoys many nice properties of the normal one, such as closure under marginalization and linear transformations, but unfortunately not under conditioning.

The ESN model is a slight extension of the SN distribution, which achieves also closure under conditioning, and hence it is suitable for the analysis of conditional independence relationships in the context of graphical models. The density of a d-dimensional ESN random variable is

$$f(y) = \frac{1}{\Phi(\tau)}\phi_d(y-\xi;\Omega)\Phi\left(\tau(1+\alpha^\top\overline{\Omega}\alpha)^{1/2} + \alpha^\top\omega^{-1}(y-\xi)\right), \tag{2}$$

where $\phi_d(y;\Omega)$ is the density of a d-dimensional $\mathrm{N}_d(0,\Omega)$ random variable, $\Phi(\cdot)$ is the $\mathrm{N}(0,1)$ distribution function, Ω is a full rank covariance matrix, ω is a diagonal matrix such that $\overline{\Omega} = \omega^{-1}\Omega\omega^{-1}$ is the corresponding correlation matrix, α is the parameter regulating skewness, ξ is the location parameter, and $\tau \in \Re$ is an additional shape parameter. When $\tau = 0$, the SN density is recovered.

The mean vector and the covariance matrix of Y are

$$\mathrm{E}[Y] = \xi + \zeta_1(\tau)\omega\delta \quad \text{and} \quad \mathrm{Var}[Y] = \Omega + \zeta_2(\tau)\omega\delta\delta^\top\omega,$$

where $\zeta_m(\cdot)$ is the mth derivative of $\log(2\Phi(\cdot))$ and $\delta = (1 + \alpha^\top\overline{\Omega}\alpha)^{-1/2}\overline{\Omega}\alpha$.

A density having form given in Equation (2) arises in Azzalini and Capitanio (1999) (Section 4, Equation 13) from a conditioning operation on a SN random variable. These authors stated the conditions for independence among blocks of linear transformations of skew-normal random variables (see their Proposition 6), and show (see Section 6.3) how they can be extended to the case of the conditional SN density. Arnold and Beaver (2000) also examined densities of the type given in Equation (2) and remarked the closure under conditioning. Capitanio et al. (2003) investigated the relationships of conditional independence among the components of an ESN random variable, as well as other issues related to the use of this distribution in the context of graphical models. Actually, if $Y = (Y_1, \ldots, Y_d)$ has density as given in Equation (2), pairwise conditional independence between Y_i and Y_j occurs if and only if the following two conditions hold simultaneously

$$\Omega^{ij} = 0, \tag{3}$$

$$\alpha_i\alpha_j = 0, \tag{4}$$

where Ω^{ij} denotes the (i,j)th entry of Ω^{-1} and α_i is the ith component of α.

Condition given in Equation (3) shows that, for an ESN random variable, the matrix Ω^{-1} plays the same role as the concentration matrix Σ^{-1} used in the normal context. Nevertheless, a further condition, given by Equation (4), on the elements of the shape parameter α is required as well. Some algebra yields the inverse of the covariance matrix of Y, which is

$$\Omega^{-1} - \zeta_2(\tau)\frac{\omega^{-1}\alpha\alpha^\top\omega^{-1}}{(1 + \zeta_2(\tau)\alpha^\top\overline{\Omega}\alpha + \alpha^\top\overline{\Omega}\alpha)}. \tag{5}$$

It is evident that if Y_i and Y_j fulfill both conditions in Equations (3) and (4) then the (i,j)th entry of expression given in (5) is equal to zero, whilst the fact that (5) has a zero entry does not imply that a pairwise conditional independence relationship occurs between the corresponding components of Y. Consequently the approach described in Section 1 for testing conditional independence in the Gaussian context needs to be adapted by taking into account the conditions stated in Equations (3) and (4).

The conditional independence graph $G = (V, E)$ used to specify the association structure among the components of Y is obtained by connecting two vertices $\{i\}$ and $\{j\}$ if Ω^{ij} and/or the product $\alpha_i\alpha_j$ are different from zero. On the contrary, there is a missing edge between the vertices $\{i\}$ and $\{j\}$ if conditions given in Equations (3) and (4) hold simultaneously.

Suppose that, for example, the three–dimensional random variable Y has an ESN distribution with scale and skewness parameters given by

$$\Omega^{-1} = \begin{pmatrix} \Omega^{11} & \Omega^{12} & 0 \\ \Omega^{21} & \Omega^{22} & \Omega^{23} \\ 0 & \Omega^{32} & \Omega^{33} \end{pmatrix}, \quad \text{and} \quad \alpha = \begin{pmatrix} 0 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}.$$

Since $\Omega^{13} = 0$ and $\alpha_1\alpha_3 = 0$, Y_1 and Y_3 are independent conditionally on Y_2, and the edge that connects vertices $\{1\}$ and $\{3\}$ is missing. Furthermore, since Ω^{12} and Ω^{23} are different from zero, there are two edges which connect vertices $\{1\}$ and $\{2\}$, and $\{2\}$ and $\{3\}$, respectively. The resulting conditional independence graph is shown in Figure 1.

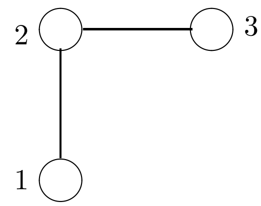

Figure 1. Conditional independence graph for the ESN_3.

3. The Wald Test's Methodology

We present the test used for model selection in conditional independence graphs when the random variables have an ESN distribution. By exploiting the pairwise conditional independence proposition given in Equations (3) and (4), the model selection procedure can be based on a test for a single edge exclusion/inclusion whose null hypothesis is

$$H_0^w\colon g(\theta) = \begin{pmatrix} \Omega^{ij} \\ \alpha_i \alpha_j \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{6}$$

Let $y_1, \ldots, y_n$ be an observed random sample drawn from the ESN distribution. The log-likelihood function for $\theta = (\xi, \Omega^{-1}, \alpha, \tau)$ is $l(\theta) = \sum_{i=1}^n \ell_i(\theta)$, where

$$\begin{aligned} \ell_i(\theta) = {} & -\frac{d}{2}\log(2\pi) + \frac{1}{2}\log\left(\left|\Omega^{-1}\right|\right) - \frac{1}{2}(y_i - \xi)^\top \Omega^{-1}(y_i - \xi) \\ & + \log\left(\Phi\left(\tau(1 + \alpha^\top \overline{\Omega}\alpha)^{1/2} + \alpha^\top \omega^{-1}(y_i - \xi)\right)\right) - \log\left(\Phi(\tau)\right). \end{aligned}$$

Let $S(y) = \{S_\xi(y), S_{\Omega^{-1}}(y), S_\alpha(y), S_\tau(y)\}$ be the score function of the ESN model based on a single observation y and thus $S_\xi(y) = \partial\left(\ell(\theta)\right)/\partial\xi$, and so forth.

The components of the score vector can be computed by considering the reparameterization $\Lambda = \Omega^{-1}$ and $\eta = \omega^{-1}\alpha$, which simplifies the derivatives of the four terms of $S(y)$. Consequently, the score functions become

$$\begin{aligned} S_\xi(y) &= \Lambda(y_i - \xi) - r(\gamma)\eta, \\ S_\Lambda(y) &= \frac{1}{2}D^\top \mathrm{vec}\left(\Lambda^{-1} - (y_i - \xi)(y_i - \xi)^\top - \tau_* r(\gamma)\Lambda^{-1}\eta\eta^\top\Lambda^{-1}\right), \\ S_\alpha(y) &= r(\gamma)\ \left\{\tau_*\omega^{-1}\Lambda^{-1}\eta + \omega^{-1}(y_i - \xi)\right\}, \\ S_\tau(y) &= r(\gamma)\sqrt{1 + \eta^\top\Lambda^{-1}\eta} - \frac{\phi(\tau)}{\Phi(\tau)}, \end{aligned} \tag{7}$$

where $r(\gamma) = \phi(\gamma)/\Phi(\gamma)$, $\gamma = \tau(1 + \eta^\top\Lambda^{-1}\eta)^{1/2} + \eta^\top(y_i - \xi)$, $\tau_* = \tau/\sqrt{1 + \eta^\top\Lambda^{-1}\eta}$, and D is the duplication matrix of dimension $d^2 \times [d(d+1)/2]$; see the Appendix for details on the derivation of Equation (7).

A natural way to approach the test of hypothesis given in (6) is to implement the likelihood-ratio test. Unfortunately, constrained optimization is quite hard to deal under the computational viewpoint, because of the additional parameter τ and the numerical instability of the procedure, which arise when one or more elements of the α vector take large values. To get a deeper insight into the problem, we focus on the asymptotic behaviour of the score function of α in the one-dimensional case given by

$$S_\alpha(y) = r(\gamma)\ \left\{\frac{\tau\alpha}{(1+\alpha^2)^{1/2}} + \omega^{-1}\left(y_i - \xi\right)\right\},$$

where

$$\gamma = \tau(1+\alpha^2)^{1/2} + \alpha z_i \quad \text{and} \quad z_i = \omega^{-1}\left(y_i - \xi\right).$$

Whatever the sign of τ, since the argument in curly brackets is $O(1)$ with respect to α, as $\alpha \to \pm\infty$, the behaviour of the score function $S_\alpha(y)$ depends only on the quantity $r(\gamma)$.

As $\alpha \to \infty$ and $z > 0$, or as $\alpha \to -\infty$ and $z < 0$, $r(\gamma)$ tends to zero. When α and z have a different sign the limit gives an indeterminate form. Consequently, if $r(\gamma)$ is considerably larger than the other term, the score function of α does not change with the observations and becomes constant. Moreover, when α diverges, the probability of z having different sign converges to zero.

The problem related to the occurrence of infinite estimates of α, even though α is finite, has been pointed out and investigated by several authors; see, e.g., Azzalini and Capitanio (1999), Rotnitzky et al. (2000), Canale (2011), and references therein.

The aspects just mentioned make the constrained estimates almost impossible to obtain numerically, so that the likelihood ratio test become unfeasible. Consequently a Wald approach is a necessary remedy.

Let $\hat{\theta}$ be the maximum likelihood estimator (MLE) of θ and $g(\hat{\theta})$ the estimator of $g(\theta)$. Furthermore, let $\Sigma_g(\theta)$ be the covariance matrix of $g(\hat{\theta})$. A Taylor expansion of $g(\hat{\theta})$ around the value θ_0 yields

$$\sqrt{n}\left\{g(\hat{\theta}) - g(\theta_0)\right\} = g'(\theta_0)\sqrt{n}(\hat{\theta} - \theta_0) + O_p\left(\frac{1}{n}\right).$$

By applying the delta method, the asymptotic variance of $g(\hat{\theta})$ turns out to be

$$\begin{aligned}\Sigma_g(\theta) &= \mathrm{Var}\left[g(\theta_0) + g'(\theta_0)(\hat{\theta} - \theta_0)\right] = \mathrm{Var}\left[g'(\theta_0)\hat{\theta}\right] \\ &= g'(\theta_0)\mathrm{Var}[\hat{\theta}]g'(\theta_0)^\top = g'(\theta_0)I^{-1}(\theta)g'(\theta_0)^\top.\end{aligned}$$

Since the MLE are asymptotically normal, we have $\sqrt{n}\left\{g(\hat{\theta}) - g(\theta_0)\right\} \xrightarrow{d} Z$, where Z is a bivariate $\mathrm{N}\left(0, g'(\theta_0)I^{-1}(\theta)g'(\theta_0)^\top\right)$ random variable. Here, $I(\theta)$ is the information matrix of $\hat{\theta}$ and $g'(\cdot)$ denotes the matrix of the first derivatives of g with respect to θ_0. In obvious notation, we have

$$\Sigma_g(\theta) = \begin{bmatrix} \mathrm{Var}[\hat{\Omega}^{ij}] & \mathrm{Cov}(\hat{\Omega}^{ij}, \hat{\alpha}_i\hat{\alpha}_j) \\ \mathrm{Cov}(\hat{\alpha}_i\hat{\alpha}_j, \hat{\Omega}^{ij}) & \mathrm{Var}[\hat{\alpha}_j\hat{\alpha}_i] \end{bmatrix},$$

where, up to first order,

$$\mathrm{Cov}(\hat{\Omega}^{ij}, \hat{\alpha}_i\hat{\alpha}_j) = \alpha_j\mathrm{Cov}(\hat{\Omega}^{ij}, \hat{\alpha}_i) + \alpha_i\mathrm{Cov}(\hat{\Omega}^{ij}, \hat{\alpha}_j), \tag{8}$$

and

$$\mathrm{Var}[\hat{\alpha}_j\hat{\alpha}_i] = \alpha_i^2\mathrm{Var}[\hat{\alpha}_j] + \alpha_j^2\mathrm{Var}[\hat{\alpha}_i] + 2\alpha_i\alpha_j\mathrm{Cov}(\hat{\alpha}_i\hat{\alpha}_j).$$

The elements of $\Sigma_g(\theta)$ can be estimated by replacing the parameters by their estimates and by replacing the variances and covariances by the corresponding elements of the inverse observed information matrix. We indicate the sample version of $\Sigma_g(\theta)$ by $\hat{\Sigma}_g$. Consequently, the test concerning the null hypothesis given in (6) can be based on the Wald–type statistic

$$W_n(Y) = \left(g(\hat{\theta}) - g(\theta)\right)^\top \hat{\Sigma}_g^{-1} \left(g(\hat{\theta}) - g(\theta)\right). \tag{9}$$

Under the null hypothesis, which implies that variables Y_i and Y_j are conditionally independent given the remaining ones, the statistic $W_n(Y)$ has an asymptotic chi-square distribution with two degrees of freedom, denoted by $\chi^2(2)$.

4. Statement and Background for the Alternative Proposal

The approach proposed in Section 3 can be applied to test the null hypothesis given in (6) only when the regularity conditions of the parameter space hold under H_0. Actually, irregularities may arise with respect to the second statement of the testing problem given in (6). The parameter space of H_0: $\alpha_i\alpha_j = 0$ is the union of two manifolds given by $\{\alpha_i = 0, \alpha_j \in R\}$ and $\{\alpha_i \in R, \alpha_j = 0\}$. When their intersection provides the single point $\{\alpha_i = 0, \alpha_j = 0\}$, the parameter space lacks its regularity.

Suppose that (6) needs to be tested and $\alpha_i = \alpha_j = 0$. Such situation affects the asymptotic behaviour of the Wald statistic since, under this specific circumstance, its asymptotic null distribution fails to be a $\chi^2(2)$ distribution; see Glonek (1993). Here, we modify the procedure for testing conditional independence to deal with the case $\alpha_i = \alpha_j = 0$ and propose an alternative formulation of the test statistic. To this purpose, we initially introduce the limiting distribution of $(\hat{\alpha}_i\hat{\alpha}_j, \hat{\Omega}^{ij})^\top$. Recall that the asymptotic distribution of the vector $(\hat{\alpha}_i, \hat{\alpha}_j, \hat{\Omega}^{ij})^\top$ is normal and therefore the statistic $(\hat{\alpha}_i\hat{\alpha}_j, \hat{\Omega}^{ij})^\top$ is also asymptotically normal (see Serfling, 1980, Section 3.3, pp. 122–125), that is,

$$\begin{pmatrix} \hat{\alpha}_i\hat{\alpha}_j \\ \hat{\Omega}^{ij} \end{pmatrix} \xrightarrow{d} \mathrm{N}_2\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{Var}[\hat{\alpha}_i\hat{\alpha}_j] & 0 \\ 0 & \mathrm{Var}[\hat{\Omega}^{ij}] \end{pmatrix}\right].$$

The covariance matrix is diagonal since Equation (8) takes value zero when $\alpha_i = \alpha_j = 0$. Hence, the test for a single edge inclusion/exclusion can be carried out by testing separately the two hypotheses reported below.

The first one refers to the components of the α vector

$$H_0^\alpha\colon \alpha_i = \alpha_j = 0, \tag{10}$$

and the test statistic is

$$T_\alpha = \frac{\hat{\alpha}_i\hat{\alpha}_j}{\sqrt{\hat{\mathrm{Var}}[\hat{\alpha}_i\hat{\alpha}_j]}}.$$

The second one considers the elements of Ω^{-1}

$$H_0^{\Omega^{-1}}\colon \Omega^{ij} = 0, \tag{11}$$

and the test statistic is

$$T_{\Omega^{-1}} = \frac{\hat{\Omega}^{ij}}{\sqrt{\hat{\mathrm{Var}}[\hat{\Omega}^{ij}]}}.$$

When $\alpha_i = \alpha_j = 0$, it is possible, under mild conditions about $f = \alpha_i\alpha_j$, to prove that $\mathrm{Var}[\hat{\alpha}_i\hat{\alpha}_j] \neq 0$; see Glonek (1993, Section 2, p. 752). Furthermore, as pointed out within the general framework of Glonek (1993), one can check that T_α is asymptotically distributed as $(1/2)Z$, where $Z \sim \mathrm{N}(0,1)$. Consequently, we have

$$2T_\alpha = 2\frac{\hat{\alpha}_i\hat{\alpha}_j}{\sqrt{\hat{\mathrm{Var}}[\hat{\alpha}_i\hat{\alpha}_j]}} \xrightarrow{d} Z.$$

Instead, according to the standard theory, the test statistic $T_{\Omega^{-1}}$ has an asymptotically $\mathrm{N}(0,1)$ distribution. Moreover the two test statistics are also asymptotically independent, $T_\alpha \overset{\text{as}}{\perp\!\!\!\perp} T_{\Omega^{-1}}$. Therefore, the overall significance level of the joint test is given by

$$\begin{aligned}
\upsilon_1 &= \mathrm{P}\left[(T_\alpha \in R_\alpha) \cup (T_{\Omega^{-1}} \in R_{\Omega^{-1}})\right] = 1 - \mathrm{P}\left(T_\alpha \notin R_\alpha\right)\mathrm{P}\left(T_{\Omega^{-1}} \notin R_{\Omega^{-1}}\right) \\
&= 1 - (1-\upsilon_\alpha)(1-\upsilon_{\Omega^{-1}}),
\end{aligned}$$

where R_α and $R_{\Omega^{-1}}$ are the rejection regions for H_0^α and $H_0^{\Omega^{-1}}$, respectively, $\upsilon_\alpha = \mathrm{P}\left(T_\alpha \in R_\alpha\right)$ and $\upsilon_{\Omega^{-1}} = \mathrm{P}\left(T_{\Omega^{-1}} \in R_{\Omega^{-1}}\right)$.

To carry out the single edge selection, we consider the following procedure. We initially test the null hypothesis given in (10) on each couples of vertices $\{i,j\}$ of the graph G. If the hypothesis $\alpha_i = \alpha_j = 0$ is rejected, we apply the approach proposed in Section 3 and carry out test given in (6). Otherwise, if (10) is not rejected, the null hypothesis given in (11) is tested. The overall significance level of the test is given by

$$\begin{aligned}
\upsilon_2 &= \mathrm{P}\left[(T_\alpha \in R_\alpha \cap W_n \in R_w) \cup (T_\alpha \notin R_\alpha \cap T_{\Omega^{-1}} \in R_{\Omega^{-1}})\right] \\
&= \mathrm{P}\left(T_\alpha \in R_\alpha \cap W_n \in R_w\right) + \mathrm{P}\left(T_\alpha \notin R_\alpha\right)\mathrm{P}(T_{\Omega^{-1}} \in R_{\Omega^{-1}}) \\
&= \mathrm{P}\left(T_\alpha \in R_\alpha\right)\mathrm{P}(W_n \in R_w | T_\alpha \in R_\alpha) + \mathrm{P}\left(T_\alpha \notin R_\alpha\right)\mathrm{P}(T_{\Omega^{-1}} \in R_{\Omega^{-1}}),
\end{aligned}$$

where W_n is the Wald statistic $W_n(Y)$ used in Equation (9) and R_w is the rejection region for H_0^w. The conditional probability $\mathrm{P}\left(W_n \in R_w | T_\alpha \in R_\alpha\right)$ is unknown. However numerical evidence reported in Section 5, shows that it is pretty close to $\mathrm{P}\left(W_n \in R_w\right) = \mathrm{P}\left(W_n > \chi^2(2,\eta)\right)$ where $\chi^2(2,\eta)$ is the ηth quantile of a $\chi^2(2,\eta)$ random variable.

Consequently, the overall significance level υ_2 can be approximated as follows

$$\upsilon_2 \cong \mathrm{P}\left(T_\alpha \in R_\alpha\right)\mathrm{P}(W_n \in R_w) + \mathrm{P}\left(T_\alpha \notin R_\alpha\right)\mathrm{P}(T_{\Omega^{-1}} \in R_{\Omega^{-1}}) \cong \upsilon_\alpha\ \upsilon_w + (1-\upsilon_\alpha)\ \upsilon_{\Omega^{-1}},$$

where $\upsilon_w = \mathrm{P}\left(W_n \in R_w\right)$.

5. Simulation Study

Here, we explore the finite sample performance of the tests proposed in Sections 3 and 4 via Monte Carlo experiments. The study is divided in two parts and in each one we investigate the properties of the corresponding tests. To this purpose, we carry out two experiments organized as follows.

We generate 10,000 random samples of size $n = 100, 200$ and 500 from each of two three–dimensional ESN random variables with parameters

$$ {}_{(1)}\Omega^{-1} = {}_{(2)}\Omega^{-1} = \begin{pmatrix} 2.0760 & -0.7920 & 0 \\ -0.7920 & 1.7424 & 0.6633 \\ 0 & 0.6633 & 1.2636 \end{pmatrix}, $$

$$ {}_{(1)}\alpha = \begin{pmatrix} 0 \\ -3 \\ 3.5 \end{pmatrix}, \quad {}_{(2)}\alpha = \begin{pmatrix} 0 \\ -3 \\ 0 \end{pmatrix}, \quad {}_{(1)}\xi = {}_{(2)}\xi = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \text{and} \quad {}_{(1)}\tau = {}_{(2)}\tau = 0.4, $$

where the subscript on the left hand side indicates the experiment in which they are used. Note that the parameters values are such that in both the experiments the conditional independence relationship $Y_1 \perp Y_3 | Y_2$ holds, so that the corresponding graph is the one shown in Figure 1, where the edge set is $E = \{(1, 2), (2, 3)\}$.

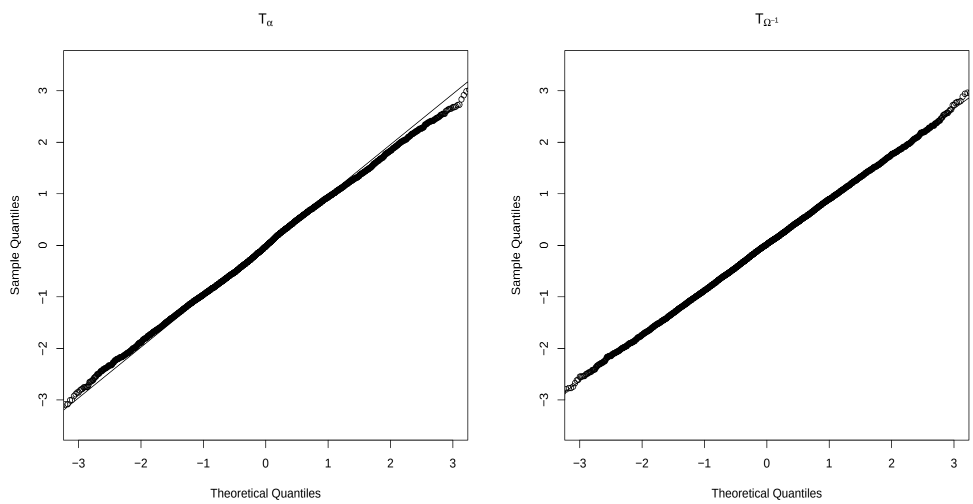

Figure 2. Asymptotic distributions of T_α and $T_{\Omega^{-1}}$.

In both parts of this study, the parameters estimates have been obtained using the reparameterization described in Capitanio et al. (2003), whereas the observed information matrix $I(\hat{\theta})$ has been evaluated numerically. The null hypothesis is always true for the edge connecting vertices $\{1\}$ and $\{3\}$ so that the corresponding entries in Tables 1, 2 and 3 approximate the actual level. On the contrary, the null hypothesis is false for the edges connecting vertices $\{1\}$ and $\{2\}$, and $\{2\}$ and $\{3\}$ respectively. Hence, the corresponding entries are to be interpreted as approximations to the power of the test.

In the first experiment, we compute the percentage of rejections of the null hypothesis given in (6) on each edge, when the nominal level of the test is 0.10 and 0.05, respectively. The results are displayed in Table 1. In this study, the Wald test seems to be slightly conservative. However, when n increases, the actual level becomes closer to the nominal one and the power is remarkably satisfactory; see Capitanio and Pacillo (2008).

Table 1. Experiment 1: percentage of rejections of the null hypothesis given in Equation (6) for two nominal levels, 0.10 and 0.05.

Nominal level	0.10 Edges			0.05 Edges		
n	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
100	85.92%	2.87%	73.45%	77.01%	1.09%	64.28%
200	99.68%	3.85%	97.14%	99.41%	1.70%	95.97%
500	100.00%	5.30%	99.95%	100.00%	2.40%	99.94%

In the second experiment, we investigate the asymptotic behaviour of the testing procedure proposed in Section 4. Figure 2 highlights, for the sample size $n = 100$, the accuracy of the approximation to the distribution of the two test statistics, T_α and $T_{\Omega^{-1}}$, provided by the asymptotic $\mathrm{N}(0, 1)$ distribution. The two graphs in this figure show that their limiting distributions are very close to the normal one even for such a pretty small sample size in the context of graphical models. Moreover, the two scatter plots of Figure 3 point out that the independence assumption for T_α and $T_{\Omega^{-1}}$ is indeed reasonable also in finite sample.

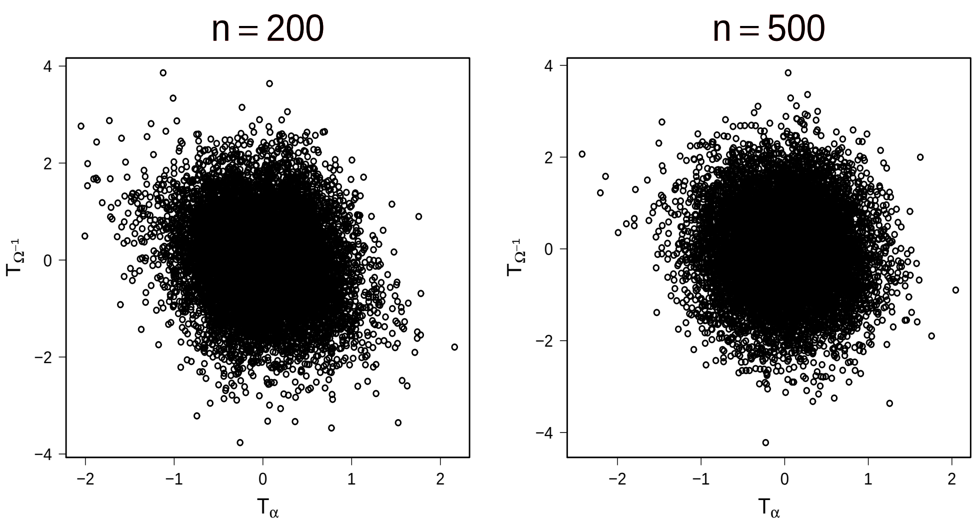

Figure 3. Independence of the test statistics T_α and $T_{\Omega^{-1}}$

Table 2 shows the percentage of rejections of the null hypothesis H_0^α: $\alpha_i = \alpha_j = 0$ and $H_0^{\Omega^{-1}}$: $\Omega^{ij} = 0$, on each edge, for the two nominal levels 0.10 and 0.05. The simulated level is close the nominal one even for sample sizes as small as 200. Moreover, the power of the test, as in the previous experiment, is definitely satisfactory.

Table 2. Experiment 2: percentage of rejections of the null hypothesis H_0^α and $H_0^{\Omega^{-1}}$ for two nominal levels, 0.10 and 0.05.

Nominal level	0.10 Edges			0.05 Edges		
n	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
100	60.94%	6.15%	64.23%	47.63%	2.51%	51.86%
200	92.61%	8.02%	94.03%	86.03%	3.87%	89.76%
500	99.96%	9.12%	99.95%	99.90%	4.19%	99.92%

Table 3 records the percentage of rejections, in experiment 2, when the complete procedure is implemented, that is, H_0^α is initially tested, and then either (6) or (11) are considered, according to the outcome of test given in (10). The overall nominal level of the test is 0.10. The results provide simulated values for the actual significance level very close to the nominal one and an appreciable power.

Table 3. Experiment 2: percentage of rejections of the null hypothesis in the complete procedure, for the nominal level 0.10.

Nominal level		0.10 Edges	
n	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$
100	67%	5.76%	69.95%
200	94.05%	7.63%	95.13%
500	99.96%	8.74%	99.97%

Table 4 provides, for experiment 2, the probabilities of $W_n > \chi^2(2,\eta)$ given $T_\alpha \in R_\alpha$. The results show that it may be reasonable to approximate such probabilities with those of a $\chi^2(2)$ random variable in the evaluation of the overall significance level.

Table 4. Simulated probabilities of $W_n > \chi^2(2,\eta)$ given $T_\alpha \in R_\alpha$.

		n	
Percentile	100	200	500
80%	0.230	0.21	0.170
90%	0.093	0.110	0.067
95%	0.044	0.062	0.027
97.5%	0.022	0.031	0.010

6. Conclusions

We have considered testing procedures for model selection in conditional independence graphs when the variables have an ESN distribution. Under regularity conditions, a Wald-type statistic, with an asymptotic chi-squared distribution, has been obtained. A simulation study have shown that when the sample size increases, the actual level of the test is reasonably close to the nominal one and the power is satisfactory. We have proposed a test to handle the situation when the parameter space lacks the usual regularity conditions. Results from numerical simulations have shown that this second test is also accurate and powerful even for small sample sizes in the context of graphical models. The main advantage of proposed procedures is the possibility to work with a model able to fit distributions of data affected by skewness which, nevertheless, shares many nice properties of the normal one. However, it would be interesting to investigate whether the proposed model selection procedure can be generalized to a conditional independence graph when the variables follow a multivariate extended skew-Student-t distribution. This class of distributions enjoys the feature of closure under conditioning, as well as, the ability to model lighter tails than the normal distribution. For a complete discussion of multivariate unified skew-elliptical distributions, see, for example, Arellano-Valle and Genton (2010b).

Acknowledgements

The author is grateful to Mathias Drton for stimulating discussion and to the anonymous referees for their helpful suggestions, which greatly improved the original manuscript. Work partially supported by PRIN 2008, grant 2008AHWTJ4 003, from MIUR, Italy.

Appendix: Score Functions

The log-likelihood function for the reparameterization $\theta = (\xi, \Lambda, \alpha, \tau)$, with $\Lambda = \Omega^{-1}$, is

$$l(\theta) = \sum_{i=1}^{n} \ell_i(\theta) = \sum_{i=1}^{n} \Big\{ -\frac{d}{2}\log(2\pi) + \frac{1}{2}\log(|\Lambda|) - \frac{1}{2}(y_i - \xi)^\top \Lambda (y_i - \xi) \\ + \log(\Phi(\gamma)) - \log(\Phi(\tau)) \Big\},$$

where $\gamma = \tau(1+\eta^\top\Lambda^{-1}\eta)^{1/2} + \eta^\top(y_i - \xi) = \tau(1+\alpha^\top\overline{\Omega}\alpha)^{1/2} + \alpha^\top\omega^{-1}(y_i-\xi)$ and $\eta = \omega^{-1}\alpha$.

The score function for ξ is

$$S_\xi(y) = \frac{\partial(\ell(\theta))}{\partial\xi} = \frac{\partial}{\partial\xi}\left(-\frac{1}{2}(y_i-\xi)^\top\Lambda(y_i-\xi) + \log(\Phi(\gamma))\right) \\ = \Lambda(y_i - \xi) - \eta\, r(\gamma),$$

where $r(\gamma) = \phi(\gamma)/\Phi(\gamma)$.

In order to obtain the score function for Λ, we compute the score vector associated with the $d(d+1)/2$ different parameters in $\Lambda = \Omega^{-1}$ (see, e.g. Magnus and Neudecker, 1999) given by

$$\frac{\partial(\ell(\theta))}{\partial\lambda} = \frac{1}{2}\ \mathrm{vec}(\Lambda)^\top\ \mathrm{vec}\left(\Lambda^{-1} - (y_i-\xi)(y_i-\xi)^\top - \frac{\phi(\gamma)}{\Phi(\gamma)}\tau_*\Lambda^{-1}\eta\eta^\top\lambda^{-1}\right).$$

Hence, the score function for Λ is

$$S_\Lambda(y) = \frac{1}{2}D^\top\mathrm{vec}\left(\Lambda^{-1} - (y_i-\xi)(y_i-\xi)^\top - \tau_* r(\gamma)\Lambda^{-1}\eta\eta^\top\Lambda^{-1}\right),$$

where D is the duplication matrix of dimension $d^2 \times [d(d+1)/2]$. The score function for α is

$$S_\alpha(y) = \frac{\partial(\ell(\theta))}{\partial\alpha} = \frac{\partial}{\partial\alpha}(\log(\Phi(\gamma))) = \frac{\phi(\gamma)}{\Phi(\gamma)}\,\frac{\partial\gamma}{\partial\alpha} \\ = r(\gamma)\ \left\{\tau_*\ \omega^{-1}\Lambda^{-1}\eta + \omega^{-1}(y_i-\xi)\right\},$$

where $\tau_* = \tau/(1+\eta^\top\Lambda^{-1}\eta)^{1/2}$. The score function for τ is

$$S_\tau(y) = \frac{\partial(\ell(\theta))}{\partial\tau} = \frac{\partial}{\partial\tau}\ (\log(\Phi(\gamma)) - \log(\phi(\tau))) \\ = r(\gamma)(1+\eta^\top\Lambda^{-1}\eta)^{1/2} - \frac{\phi(\tau)}{\Phi(\tau)}.$$

References

Arellano-Valle, R.B., Genton, M.G., 2010a. Multivariate extended skew-t distributions and related families. Metron, 68, 201–234.

Arellano-Valle, R.B., Genton, M.G., 2010b. Multivariate unified skew-elliptical distributions. Chilean Journal of Statistics, 1, 17–33.

Arnold, B.C., Beaver, R.J., 2000. Hidden truncation models. Sankhyā, 62, 23–35.

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12, 171–178.

Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate skew normal distribution. Journal of The Royal Statistical Society Series B - Statistical Methodology, 61, 579–602.

Azzalini, A., Dalla Valle, A., 1996. The multivariate skew-normal distribution. Biometrika, 83, 715–726.

Canale, A., 2011. Statistical aspects in the extended skew-normal distribution. Metron (in press).

Capitanio, A., Azzalini, A., Stanghellini, E., 2003. Graphical models for skew-normal variates. Scandinavian Journal of Statistics, 30, 129–144.

Capitanio, A., Pacillo, S., 2008. A Wald's test for conditional independence skew normal graphs. In Proceedings in Computational Statistics - Compstat 2008, Heidelberg, Physica-Verlag, pp. 421–428.

Dempster, A., 1972. Covariance selection. Biometrics, 28, 157–175.

Edwards, D., 2000. Introduction to Graphical Modelling. Springer, New York.

Glonek, G.F.V., 1993. On the behaviour of Wald statistics for the disjunction of two regular hypotheses. Journal of The Royal Statistical Society Series B - Statistical Methodology, 55, 749–755.

Lauritzen, S.L., 1996. Graphical Models. Oxford Science Publications, Oxford.

Magnus, J.R., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York.

Rotnitzky, A., Cox, D.R., Bottai, M., Robins, J., 2000. Likelihood-based inference with singular information matrix. Bernoulli, 6, 243–284.

Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.

Whittaker, J., 1990. Graphical Models in Applied Multivariate Statistics. Wiley, New York.