UNIVERSIDAD CARLOS III DE MADRID working papers

UC3M Working Papers
Statistics and Econometrics
16-08
ISSN 2387-0303
July 2016

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-48

D-Trace precision matrix estimator with eigenvalue control

Vahe Avagyan[a]

Abstract

The estimation of a precision matrix has an important role in several research fields. In high-dimensional settings, one of the most prominent approaches to estimate the precision matrix is the l_1 (Lasso) norm penalized convex optimization. This framework guarantees the sparsity of the estimated precision matrix. However, it does not control the eigenspectrum of the obtained estimator, and, moreover, it shrinks the largest eigenvalues of the estimated precision matrix. In this paper, we focus on D-trace precision matrix methodology. We propose imposing a negative trace penalization on the objective function of the D-trace approach, aimed to control the eigenvalues. Through extensive numerical analysis, using simulated and real datasets, we show the advantageous performance of our proposed methodology.

Keywords: Gaussian Graphical Model, Hannan-Quinn Information Criterion, D-trace, Gene expression, Trace penalization.

[a] Department of Statistics, Universidad Carlos III de Madrid.

D-Trace precision matrix estimator with eigenvalue control

Vahe Avagyan[a]

[a]*Department of Statistics, Universidad Carlos III de Madrid, c/ Madrid 126, 28903 Getafe (Madrid), Spain*

Abstract

The estimation of a precision matrix has an important role in several research fields. In high-dimensional settings, one of the most prominent approaches to estimate the precision matrix is the ℓ_1 (Lasso) norm penalized convex optimization. This framework guarantees the sparsity of the estimated precision matrix. However, it does not control the eigenspectrum of the obtained estimator, and, moreover, it shrinks the largest eigenvalues of the estimated precision matrix. In this paper, we focus on D-trace precision matrix methodology. We propose imposing a negative trace penalization on the objective function of the D-trace approach, aimed to control the eigenvalues. Through extensive numerical analysis, using simulated and real datasets, we show the advantageous performance of our proposed methodology.

Keywords: Gaussian Graphical Model, Hannan-Quinn Information Criterion, D-trace, Gene expression, Trace penalization.

1. Introduction

The estimation of high dimensional inverse covariance or precision matrix has attracted significant interest in the current age of large-scale data explosion. It is an important and ubiquitous problem in several rapidly developing fields, such as genetics (Stifanelli et al. 2013; Yin and Li 2013), medicine (Huang et al. 2010; Ryali et al. 2012), climate studies (Zerenner et al. 2014), finance (Stevens 1998; Frahm and Memmel 2010; Goto and Xu 2015). Moreover, it has a crucial role in various machine learning methodologies, such as classification and forecasting (McLachlan 2004).

The precision matrix is exceedingly useful under the assumption of multivariate normality of data. It is well known that under this assumption the zero entry at the position (i, j) of the precision

Email address: vahe.avagyan@uc3m.es (Vahe Avagyan)

matrix indicates the conditional independence between the variables X^i and X^j, given all the other variables (Dempster 1972; Lauritzen 1996). In other words, the precision matrix represents the statistical dependency among normally distributed variables. Thus, in high dimensional settings, the precision matrix is usually sparse, since some of the variables do not interact.

The sparse precision matrix is related to the Gaussian Graphical Models (GGM) proposed by Whittaker (1990) and later analysed by Lauritzen (1996). It is an undirected graph $G = (N, E)$, where the set of the nodes, $N = \{1, ..., p\}$, contains the indexes of the variables. The set of the edges, $E \subseteq N \times N$, consists of the pair indexes (i, j), that correspond to $\omega_{ij} \neq 0$, for $1 \leq i, j \leq p$. More specifically, the GGM is a useful and costless framework for illustrating the structure of the dependencies between normally distributed variables. Therefore, the sparsity of the estimated precision matrix is not only desirable but also an essential property. A prominent example of GGM application is the genetic interaction network.

In this paper, we focus on the estimation of a high-dimensional precision matrix. We also consider the estimation of the corresponding GGM (also known as covariance selection problem) due to its importance in several research fields. Throughout the paper we assume that a centred sample data matrix, $\mathbf{X}_{n\times p}$, is observed, where each row $X_i = (X_{i1}, ..., X_{ip}) \sim \mathbf{IIDN}(0, \Sigma)$ is a realization of a p-variate normal random vector that is independent and identically distributed for $i = 1, ..., n$, and has a covariance matrix Σ with the corresponding precision matrix $\Omega = \Sigma^{-1}$. Although we assume that precision matrix Ω is sparse, we do not make any particular assumption on its sparsity pattern.

In high dimensional settings, the estimation of precision matrix greatly suffers from the absence of a naive estimator. In classical statistics, the traditional precision matrix estimator is the inverse of the sample covariance matrix[1] S (i.e., maximum likelihood estimator). However, this estimator becomes completely unreliable when the number of the variables p approaches n, since its bias increases drastically (Anderson 2003). Moreover, the matrix S^{-1} does not exist when $n < p$, since the matrix S is singular.

A considerable research is devoted to the estimation of precision matrices. Haff (1980), Frahm and Memmel (2010) and Kourtis et al. (2012) considered linear combination between the MLE (i.e., S^{-1}) and a target matrix. Note that these approaches are applicable only when $n >> p$.

[1] See Section 2 for the formal definition of the matrix S.

In high dimensional settings, the regularization framework has gained a substantial attention. The most popular approach is the ℓ_1 or LASSO regularization (Tibshirani 1996), which addresses the sparsity requirement of the estimated matrix and, moreover, is convex. In this way, Banerjee et al. (2006) proposed the ℓ_1 norm penalized log-likelihood maximization approach, which is known in literature as GLASSO[2] estimator. Several other authors studied the GLASSO method (see, for example, Yuan and Lin 2007; d'Aspremont et al. 2008; Banerjee et al. 2008; Rothman et al. 2008; Yin and Li 2013). On the other hand, Fan et al. (2009) proposed to employ adaptive LASSO and SCAD (Smoothly Clipped Absolute Deviation) penalties to reduce the bias of the GLASSO estimator. Finally, van Wieringen and Peeters (2016) proposed the ridge estimation of the precision matrix through ℓ_2 norm penalized log-likelihood maximization problem.

Several authors studied non-likelihood based approaches for estimating either the precision matrix or the GGM structure. Meinshausen and Bühlmann (2006) introduced the Neighborhood Selection approach. This method based on conducting LASSO regressions and is aimed to select the GGM. Yuan (2010) proposed a similar approach to estimate each column of the precision matrix by using the Dantzig selector. Cai et al. (2011) proposed the constrained ℓ_1 norm minimization estimator known as CLIME estimator. More recently, Zhang and Zou (2014) proposed a precision estimation method through ℓ_1 norm penalized D-trace loss minimization method. This approach is an alternative to GLASSO method, since it uses a novel and more simple objective function.

In this paper, we focus on the ℓ_1 norm penalized D-trace loss minimization method (hereafter, DT estimator). As mentioned earlier, the ℓ_1 regularization controls the sparsity pattern of the estimated precision matrix. However, this framework does not control the eigenvalues of the corresponding estimated matrix. Moreover, as the penalty parameter increases (i.e., the matrix becomes sparser), the largest eigenvalues of the estimated matrix decrease considerably and the smallest eigenvalues increase insignificantly. As a result, the eigenspectrum of the estimated precision matrix shrinks. In section (2), we provide a simple example to demonstrate the shrinkage of the eigenvalues.

We propose an extension of DT estimator with an eigenvalue control. We employ an additional regularization of the objective function of the DT estimator through negative trace penalization. This penalty sustains the stability of the eigenvalues. In particular, the negative trace penalty

[2]GLASSO or Graphical LASSO is the name of a popular algorithm, proposed by Friedman et al. (2008), which solves the ℓ_1 norm penalized log-likelihood maximization problem.

diminishes the significant decrease of the largest eigenvalues. Thus, the estimated matrix becomes sparse without having to shrink its eigenspectrum significantly. Through extensive simulation study, we show that the proposed methodology outperforms the DT estimator, by considering different evaluation measures. In particular, we use the following loss functions to measure the statistical performance: Kullback-Leiblar loss, Reverse Kullback-Leiblar loss, Frobenius norm loss, ℓ_1 norm loss, operator (i.e., spectral) norm loss. Moreover, we use Specificity, Sensitivity, Matthews Correlation Coefficient (MCC) and F_1 score to measure the GGM prediction accuracy. Furthermore, we propose a novel penalty parameter calibration procedure based on the Hannan-Quinn information criterion. In the simulation study, we employ BIC and HQIC approaches to select the penalty parameters.

The rest of the manuscript is organized as follows. In section 2, after introducing the main notations, we describe the proposed methodology. In section 3, we propose a different approach for selecting the penalty parameters. In section 4, we exhaustively evaluate the statistical loss and GGM prediction performance of the proposed methodology and compare them with that of DT and GLASSO methods. In section 5, we apply the proposed methodology to an empirical application: prediction of breast cancer state using LDA. Finally, we provide the conclusions in Section 6. We provide the description of the algorithm in Appendix A and the simulation results in Appendix B.

2. Proposed methodology

Before proceeding with the proposed methodology, we introduce the following notations. For any vector $\mathbf{a} = (a_1, ..., a_p)^T \in \mathbb{R}^p$, we define the ℓ_2 or Euclidean norm by $||\mathbf{a}||_2 = \sqrt{\sum_{j=1}^{p} a_j^2}$. For any symmetric matrix $\mathbf{A} = [a_{ij}]_{1 \leq i,j \leq p}$, we denote the Frobenius norm by $||\mathbf{A}||_2 = \sqrt{\sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij}^2}$, the matrix ℓ_∞ norm by $||\mathbf{A}||_\infty = \max_{1 \leq i,j \leq p} |a_{ij}|$, the matrix ℓ_1 norm by $||\mathbf{A}||_{\ell_1} = \max_{1 \leq j \leq p} \sum_{i=1}^{p} |a_{ij}|$, the componentwise ℓ_1 norm by $||\mathbf{A}||_1 = \sum_{i=1}^{p} \sum_{j=1}^{p} |a_{ij}|$, and the spectral or operator norm by $||\mathbf{A}||_{\text{spec}} = \sup_{||x||_2 \leq 1} ||Ax||_2$. For any two symmetric $p \times p$ matrices A and B, we write $A \succeq B$ or $A \succ B$ if the matrix $A - B$ is positive semidefinite or positive definite, respectively. Finally, we assume that $\mathbf{X}$ is a centered sample data matrix with dimension $n \times p$, where each row $X_i = (X_{i1}, ..., X_{ip})$ is a realization of a p-variate normal random vector that is independent and identically distributed for $i = 1, ..., n$, with covariance matrix Σ and precision matrix $\Omega = \Sigma^{-1}$.

Zhang and Zou (2014) have proposed the D-trace loss function, which has the following definition:

$$f_{DT}(\Omega, \Sigma) = \frac{1}{2}\text{trace}(\Omega^2\Sigma) - \text{trace}(\Omega). \tag{1}$$

It is easy to check that the function $f_{DT}(\Omega, \Sigma)$ is convex in Ω, has a positive-definite Hessian matrix, and a unique minimizer at Σ^{-1}. By regularizing the $f_{DT}(\Omega, \Sigma)$ function through a ℓ_1 norm,[3] Zhang and Zou (2014) proposed the ℓ_1 penalized D-trace loss minimization estimator (hereafter, DT estimator). The DT estimator is defined as the solution of the following optimization problem:

$$\widehat{\Omega}_{\text{DT}} = \arg\min_{\Omega \succeq \epsilon I} \frac{1}{2}\text{trace}(\Omega^2 S) - \text{trace}(\Omega) + \tau||\Omega||_1, \tag{2}$$

where $S = (1/n)\sum_{i=1}^{n} X_i X_i^T$ is the sample covariance matrix, $\tau > 0$ is the associated penalty parameter and ϵ is a small positive value.[4] The constraint $\Omega \succ \epsilon I$ guarantees the positive definiteness of the matrix $\widehat{\Omega}_{\text{DT}}$. To solve the problem (2), Zhang and Zou (2014) developed an algorithm based on the alternating direction method.

As discussed in the Section 1, this paper addresses the eigenvalues control of the DT estimator. The ℓ_1 norm penalization guarantees the sparsity of the estimated precision matrix for a well-selected parameter τ. However, as τ increases (i.e., $\widehat{\Omega}_{\text{DT}}$ becomes sparser), the eigenspectrum of the estimated matrix shrinks. In other words, the largest eigenvalues decrease significantly and the smallest eigenvalues increase insignificantly.

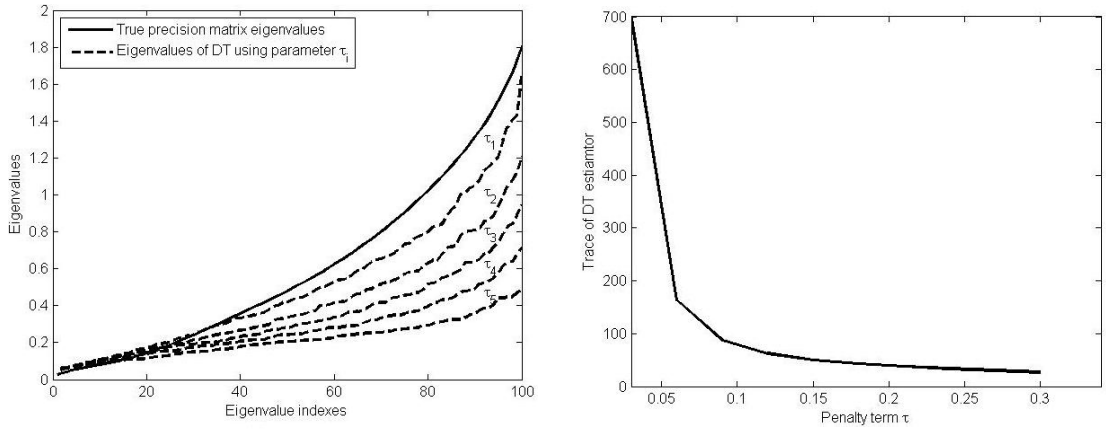
In order to illustrate the shrinkage of the eigenvalues, we consider a simple example. Assume that the true precision matrix has a known sparse structure which is given by *Model 2* described in Section 4.1. For this example we specify the values $p = 100$ and $n = 100$. In Figure 1a, the eigenvalues of the estimator DT is shown for different penalty parameters τ_i, where $\tau_1 < ... < \tau_5$. Note that, as the parameter τ increases (i.e., the matrix becomes sparser), the largest eigenvalues of the DT estimator decrease significantly towards a constant. As a result, the trace of the matrix decreases significantly due to ℓ_1 norm penalization (see Figure 1b).

In order to control the eigenspectrum of the estimated precision matrix $\widehat{\Omega}_{\text{DT}}$, an additional constraint is required. We note that the decrement of the largest eigenvalues of the estimated matrix is associated with the decrease of its trace. We propose to impose an additional penalization

[3] In this paper, we consider the ℓ_1 norm of the matrix, including the diagonal entries.

[4] In practice, we fix $\epsilon = 10^{-8}$.

Figure 1. (a) Eigenvalues of DT estimator for different penalty parameters, (b) Trace of DT estimator for $\tau \in [0.03; 0.3]$.



(a) **(b)**

in the problem (2) through a negative trace of Ω. In this way, we propose the DT estimator with Eigenvalue Control (or shortly, DTEC). Our proposed estimator is the solution of the following optimization problem.

$$\widehat{\Omega}_{\text{DTEC}} = \arg\min_{\Omega \succeq \epsilon I} \ \frac{1}{2}\text{trace}(\Omega^2 S) - \text{trace}(\Omega) + \tau||\Omega||_1 - \gamma\text{trace}(\Omega), \tag{3}$$

where $\tau > 0$ and $\gamma > 0$ are penalty parameters. More specifically, the penalty term $-\gamma\text{trace}(\Omega)$ in problem (3) endorses the trace (and, therefore, the largest eigenvalues) of the estimated precision matrix not to decrease significantly. Thus, the eigenvalues (especially the largest ones) of the estimated precision matrix $\widehat{\Omega}_{\text{DTEC}}$ will be closer to the true ones, than those of the estimated matrix $\widehat{\Omega}_{\text{DT}}$.

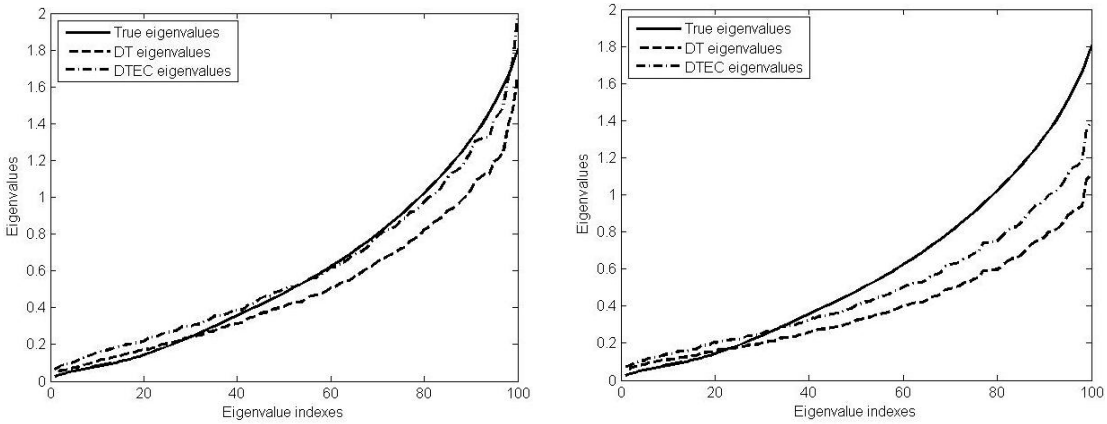
Furthermore, we can write our proposed estimator as the solution of the following optimization problem:

$$\widehat{\Omega}_{\text{DTEC}} = \arg\min_{\Omega \succeq \epsilon I} \ \frac{1}{2}\text{trace}(\Omega^2 S) - (1+\gamma)\text{trace}(\Omega) + \tau||\Omega||_1. \tag{4}$$

Note that we can solve the problem (4) using a similar algorithm based on alternating direction method, as for the problem (2). For completeness, we present the algorithm for solving the DTEC method in the Appendix A.

To better illustrate the behaviour of the proposed methodology, we show a particular example. Assume that the true precision matrix Ω has the sparse structure used in the earlier example. In Figure 2 we present the eigenvalues of true precision matrix Ω and the eigenvalues of estimators DT and DTEC. Figure 2a illustrates the eigenvalues of the optimal (i.e., oracle) estimators in terms of the Frobenius norm loss.[5] In other words, DT and DTEC estimators are obtained using the penalty parameters that minimize the Frobenius norm loss assuming that the true precision matrix is known. Figure 2b shows the eigenvalues of DT and DTEC estimators obtained through BIC[6] approach described in the Section 3. From both figures we can see that the eigenvalues of DTEC estimator are larger than those of DTEC estimator. Moreover, the eigenvalues (especially the largest ones) of DTEC estimator are much closer to the eigenvalues of the true precision matrix than those of DT. Therefore, the trace penalization diminishes the significant decrease of the DT estimator eigenvalues.

Figure 2. Eigenvalues of the true precision matrix and estimators DT, DTEC obtained through (a) optimal penalty parameters, (b) BIC selection approach.



(a) **(b)**

In Section 4, through an exhaustive empirical analysis including several sparsity patterns for the precision matrix, we show that our proposed DTEC estimator can outperform the DT under

[5] See Section 4.2 for a formal definition.

[6] For the sake of space we do not include the results obtained through the HQIC.

several statistical performance measures including those for graphical models. Moreover, we use different calibration techniques to select penalty parameters τ and γ.

3. Penalty Parameter Selection

The selection of the penalty parameter is an important problem for methodologies based on regularization. The performance of the estimated precision matrix greatly depends on the penalty parameter. Moreover, the penalty parameter controls the sparsity pattern of the estimated matrix. In this paper, we suggest the use of a criterion based on the log-likelihood function of the Gaussian model. The score function of this criterion is defined as

$$\mathrm{SF}(\tau) = -\log\det\widehat{\Omega}(\tau) + \mathrm{trace}(S\widehat{\Omega}(\tau)) + \mathrm{df}\times\mathrm{nz}, \tag{5}$$

where $\mathrm{nz} = \mathrm{card}\{(i,j) : 1 \leq i \leq j \leq p, [\widehat{\Omega}(\tau)]_{ij} \neq 0\}$ and df is degrees of freedom (i.e., penalization term). The estimated penalty parameter is obtained by $\hat{\tau} = \arg\min_{\tau} \mathrm{SF}(\tau)$. The advantage of the suggested criterion is twofold. First, it is effective in terms of the computational time than the techniques based on the cross-validation framework. Second, it accounts the sparsity characteristics of the estimated precision matrix through the term nz. We consider the following two choices of degrees of freedom:

$$\mathrm{df}_{\mathrm{BIC}} = \frac{\log n}{n}, \tag{6}$$

$$\mathrm{df}_{\mathrm{HQIC}} = \frac{2\log\log n}{n}. \tag{7}$$

The criterion (5) with degrees of freedom $\mathrm{df}_{\mathrm{BIC}}$ corresponds to the Bayesian Information Criterion (BIC). Yuan and Lin (2007) proposed BIC for precision matrix estimation methodologies. This criterion is very popular in this framework and was employed by several authors (see, for instance, Peng et al. 2009; Lian 2011; Tarr et al. 2016). On the other hand, the criterion (5) with degrees of freedom $\mathrm{df}_{\mathrm{HQIC}}$ corresponds to the Hannan-Quinn Information Criterion (HQIC) (Hannan and Quinn 1979). This is a novel criterion in the precision matrix estimation framework and a little attention has been paid to the HQIC, in general. The advantage of $\mathrm{df}_{\mathrm{HQIC}}$ is that it is stable when the number of the observations n is very large.

We note that our proposed methodology requires selection of two parameters, τ and γ. For this reason, we define the following multivariate score function to select simultaneously these parameters:

$$\mathrm{MSF}(\tau,\gamma) = -\log\det\widehat{\Omega}(\tau,\gamma) + \mathrm{trace}(S\widehat{\Omega}(\tau,\gamma)) + \mathrm{df}\times\mathrm{nz}, \tag{8}$$

where $\widehat{\Omega}(\tau,\gamma)$ is the estimated precision matrix for given values τ and γ. The estimated parameters $\hat{\tau}$ and $\hat{\gamma}$ are selected by $(\hat{\tau},\hat{\gamma}) = \arg\min_{\tau,\gamma} \mathrm{MSF}(\tau,\gamma)$ through two-dimensional grid search.

4. Simulation Study

In this section, we conduct a simulation analysis to evaluate the performance of our proposed methodology. In subsection 4.1, we detail the considered models for the precision matrix Ω, and in subsection 4.2, we describe the performance evaluation. Finally, in subsection 4.3, we provide the discussion of the results.

4.1. Considered models

We perform an exhaustive simulation study through six different patterns for the precision matrix with varying sizes. The considered models for the precision matrix Ω are the following:

(ii) Deterministic patterns

- *Model* 1. Tridiagonal structure: $\omega_{ii} = 1$, $\omega_{i,i-1} = \omega_{i-1,i} = 0.45$ and other values are 0.
- *Model* 2. Tridiagonal structure with varying entries: $\Omega = D^{1/2}\Omega_1 D^{1/2}$, where D is a diagonal matrix with entries $D_{ii} = \dfrac{4i+p-5}{5(p-1)}$, $i = 1, ..., p$ and Ω_1 is a matrix defined in the model 1.
- *Model* 3. Decay structure: $\omega_{ij} = 0.6^{|i-j|}$.
- *Model* 4. A block-diagonal matrix, with four equally sized blocks along the diagonal, with a decay model in each block.

(i) Random patterns[7]

- *Model* 5. A random p.d. matrix, containing approximately 20% of non-zero entries.
- *Model* 6. A random p.d. matrix, containing approximately 50% of non-zero entries.

For each precision matrix model, we simulate multivariate normal random samples with zero mean. In this analysis study the performance of the estimators when $n \leq p$. We set $n = 100$ and $p = 100, 200$ and 300. The number of replications is 100.

[7]The random models are generated using the MATLAB command *sprandsym*.

4.2. Performance evaluation

To evaluate the performance of a precision matrix estimator, we use several losses and measures. In particular, we consider the Entropy loss, also known as the Kullback-Leibler loss (KLL). This loss is based on the Bregman divergence and has the following definition:

$$\mathrm{KLL}(\widehat{\Omega}, \Omega) = \mathrm{trace}(\Omega^{-1}\widehat{\Omega}) - \log\det(\Omega^{-1}\widehat{\Omega}) - p. \tag{9}$$

The KL loss function is one of the popular precision matrix estimation measures in the literature and been used widely in prior research on this framework (see, for instance, Yuan and Lin 2007; Rothman et al. 2008; Fan et al. 2009; Yin and Li 2013; Lian and Fan 2015). In this paper, we propose another version of the KLL which has not been considered previously. We call it Reverse Entropy or Reverse Kullback-Leibler loss (RKLL). This loss is also based on the Bregman divergence and is defined as

$$\mathrm{RKLL}(\widehat{\Omega}, \Omega) = \mathrm{trace}(\Omega\widehat{\Omega}^{-1}) - \log\det(\Omega\widehat{\Omega}^{-1}) - p. \tag{10}$$

We note that RKLL is sensitive to the precision matrix estimator, which allows as to capture even small estimation errors.

Moreover, in line with KLL and RKLL, we also consider matrix losses: the Frobenius norm ℓ_2, the spectral norm ℓ_{spec} and the matrix ℓ_1 norm, defined respectively as:

$$\ell_2(\widehat{\Omega}, \Omega) = ||\widehat{\Omega} - \Omega||_2, \tag{11}$$

$$\ell_{\mathrm{spec}}(\widehat{\Omega}, \Omega) = ||\widehat{\Omega} - \Omega||_{\mathrm{spec}}, \tag{12}$$

$$\ell_1(\widehat{\Omega}, \Omega) = ||\widehat{\Omega} - \Omega||_{\ell_1}. \tag{13}$$

In order to evaluate the sparsity pattern of the precision matrix estimator (i.e., GGM selection performance), we compute Specificity, Sensitivity, Matthews Correlation Coefficient (MCC) and F_1 score, defined as:

$$\text{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{14}$$

$$\text{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{15}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}}, \tag{16}$$

$$\text{F}_1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{17}$$

where TP, TN, FP and FN are the numbers of true positives (number of correctly estimated non-zero entries), true negatives (number of correctly estimated zero entries), false positives (number of incorrectly estimated non-zero entries) and false negatives (number of incorrectly estimated zero entries), respectively. In (17), we define Precision = TP/(TP + FP) and Recall = TP/(TP + FN). Both MCC (Matthews 1975) and F_1 score (Powers 2011) are commonly used to evaluate the performance of binary classifiers. In our context, we consider these measures for the overall evaluation of the GGM selection. The values of MCC are in [-1,1], and the closer the MCC to one is, the better the classification is. On the other hand, the values of F_1 score are in [0,1], and the closer the F_1 score to one is, the better the classification is. We note that we put MCC above F_1 score, as it uses all the information about the classification, whereas F_1 score ignores the FP and TN parameters.

We compare our proposed estimator with DT and GLASSO. The GLASSO estimator is defined as the solution of the following optimization problem:

$$\widehat{\Omega}_{\text{GLASSO}} = \arg\min_{\Omega} \; -\log\det\Omega + \text{trace}(S\Omega) + \nu||\Omega||_1. \tag{18}$$

Several algorithms have been developed to solve the problem (18) efficiently, such as the Graphical Lasso (Friedman et al. 2008), a Project Sub-gradient Method (Duchi et al. 2008), an Alternating Linear Minimization (Scheinberg et al. 2010), an Interior Point method (Li and Toh 2010) and a Quadratic Approximation method (Hsieh et al. 2014), among others. In this paper we choose the Graphical Lasso or *glasso* algorithm.

The penalty parameters of the methods DT and GLASSO are estimated using univariate BIC and HQIC criteria. On the other hand, the parameters τ and γ of the proposed estimator DTEC are selected using multivariate BIC and HQIC criteria. Moreover, since the trace of the estimated matrix decreases with the increase of τ, we consider $\gamma = \tau$ in the proposed problem (3) as a naive case. We call the obtained estimator simplified DTEC or SDTEC, which is defined as the solution of the following optimization problem:

$$\widehat{\Omega}_{\text{SDTEC}} = \arg\min_{\Omega \succeq \epsilon I} \; \frac{1}{2}\text{trace}(\Omega^2 S) - (1+\tau)\text{trace}(\Omega) + \tau||\Omega||_1. \tag{19}$$

In this way, we can calibrate only one parameter instead of two, which leads to saving the computational time.

4.3. Discussion of results

We provide the simulation results in the Appendix B (see Tables B.3-B.8) to conserve space. Each table reports the averages over 100 replications and the standard deviations (SD) of the corresponding losses and prediction measures. We organize the discussion of our results as follows. We first compare our proposed estimator DTEC with DT and GLASSO estimators when the penalty parameters are selected using the BIC approach. We then discuss the same comparison when the penalty parameters are selected using the HQIC approach. We finally compare BIC and HQIC approaches through the corresponding losses and measures.

First, we observe that when the penalty parameters are selected through BIC approach, our proposed estimator DTEC outperforms DT method in terms of all statistical losses for almost all the models. The only scenarios when DTEC does not outperform DT are for model 1, when $p = 100$ and for model 3, when $p = 300$, in terms of KLL. However, for both cases SDTEC outperforms DT in terms of KLL. The comparison of the graphical model selection performances yields the following insights. The proposed estimator DTEC outperforms DT for models 2, 4, 5, 6 in terms of Specificity[8] and MCC and for models 2 in terms of F_1 score. On the other hand, DT method outperforms DTEC for all the models in terms of Sensitivity, for model 1 in terms of MCC and for models 1, 3, 4, 5, 6 in terms of F_1 score. Furthermore, we observe that GLASSO outperforms all the other estimators for models 5, 6 in terms of Sensitivity. However, for the other models GLASSO provides high statistical losses and low GGM prediction measures.

Second, the results show that when the penalty parameters are selected through HQIC approach, our proposed estimator DTEC outperforms DT method for models 1 (except when $p = 100$), 3 (except when $p = 100$), 4, 5, 6 (except when $p = 100$) in terms of KLL. Moreover, DTEC outperforms DT for all the models in terms of RKLL, ℓ_2, ℓ_1, ℓ_2, ℓ_{spec}. On the other hand, DTEC outperforms DT for models 1, 2, 4, 5, 6 (except of $p = 300$) in terms of Specificity and MCC, for models 1, 2 in terms of Sensitivity and F_1 score. DT method outperforms DTEC for models 3, 4, 5, 6 (except of $p = 300$) in terms of Sensitivity and F_1 score. Furthermore, we observe that GLASSO

[8]Specificity and MCC are excluded for model 3, because these measurements are not defined for dense models.

outperforms all the other estimators for model 6 in terms of Sensitivity and F_1 score. However, for the other models GLASSO provides poor results.

Finally, we conduct a comparison of the statistical losses and GGM prediction measures of the methods based on the penalty parameter selection approach (BIC or HQIC). We observe, that BIC criterion provides higher statistical losses, whereas HQIC criterion provides lower statistical losses. Moreover, estimators obtained through BIC criterion provide higher MCC than those for HQIC criterion, and higher F_1 score for models 1, 2. Estimators obtained through HQIC criterion provide higher F_1 score for models 3, 4, 5, 6.

In sum, the proposed DTEC estimation method provides better performance, including matrix losses and GGM predictions, than DT and GLASSO methods for most of the models. Note also that this conclusion holds if we consider the simplified SDTEC method (i.e., $\gamma = \tau$), since it performs similar to DTEC. This finding leads to saving significantly the computational time without sacrificing too much the performance.

5. Real Data Application

In this section, we perform an empirical analysis of the proposed DTEC approach through real-data example. In particular, we use breast cancer dataset to predict the cancer state using Linear Discriminant Analysis (LDA). The applied dataset is available in the web site of the National Center for Biotechnology Information (`http://www.ncbi.nlm.nih.gov/`).

In this application, we focus on the problem of predicting breast cancer patients (subjects) with pathological complete response (pCR). This is an important issue because after the neoadjuvant chemotherapy, according to Kuerer et al. (1999), the pCR indicates a cancer-free life with high probability. For this application we use a dataset (for the description of the dataset we refer to Shi et al. 2010) containing gene expression levels of subjects with different stages of breast cancer. The dataset consists of 22,283 gene expression levels of 271 subjects. There are 58 subjects with pCR and 213 subjects with residual disease (RD).

First, we divide the data into a training set and a testing set with sizes 227 (almost 5/6 of the observations) and 44 (almost 1/6 of the observations), respectively, and repeat this process 100 times. For the testing set, we randomly select 9 subjects with pCR and 35 subjects with RD

(roughly 1/6 of the subjects in each group).[9] The training set contains the remaining subjects. Second, based on the training set we perform two sample t-tests between the two groups in order to select the most significant 200 genes with the smallest p-values. Third, using the training set, we estimate the precision matrix Ω with the DT, DTEC, SDTEC and GLASSO methods. We obtain the penalty parameters for these methods using the BIC and HQIC criteria. Finally, we use the estimated precision matrix in the LDA score, defined as follows:

$$\delta_t(Y) = Y^T \widehat{\Omega}\widehat{\mu}_t - \frac{1}{2}\widehat{\mu}_t^T \widehat{\Omega}\widehat{\mu}_t, \tag{20}$$

where $t = 1, 2$ ($t = 1$ for pCR and $t = 2$ for RD) and $\widehat{\mu}_t = \dfrac{1}{n_t}\sum_{i \in class_t} x_i$ is the within group average, calculated using the training data. We use the LDA score $\delta_t(Y)$ to classify the subject Y from the testing set. The rule for the classification is $\widehat{t} = \arg\max \delta_t(Y)$ $(t = 1, 2)$. To measure the prediction accuracy for all the methods, we use the Specificity, Sensitivity, MCC and F_1 score, as defined in Section 4.2. We consider TP and TN as the number of correctly predicted RD and pCR, respectively, and FP and FN as the number of erroneously predicted RD and pCR, respectively. We report the average measurements over 100 replications in Table 1 and Table 2.

Table 1. Average pCR/RD classification measurements over 100 replications using BIC.

Method	Specificity	Sensitivity	MCC	F1
DT	0.692	0.749	0.378	0.818
DTEC	0.719	0.742	0.392	0.816
SDTEC	0.713	0.744	0.389	0.817
GLASSO	0.461	0.790	0.232	0.816

Our findings show that for both cases the GLASSO provides the highest Sensitivity, but it attains the lowest Specificity and MCC. On the other hand, the DTEC approach provides the highest Specificity and dominates all the other estimators in terms of MCC. We note that all methods provide very similar results in terms of the F_1 score. However, as mentioned earlier, this measure ignores FP and TN. The results also show that SDTEC performs almost as good as the DTEC estimator. Furthermore, we observe that the obtained Specificity and the MCC of the

[9] We follow Cai et al. (2011) for the analysis scheme.

Table 2. Average pCR/RD classification measurements over 100 replications using HQIC.

Method	Specificity	Sensitivity	MCC	F1
DT	0.559	0.787	0.312	0.827
DTEC	0.606	0.778	0.341	0.827
SDTEC	0.606	0.777	0.340	0.826
GLASSO	0.443	0.802	0.231	0.823

considered estimators are higher for BIC than the same for HQIC, whereas the Sensitivity and F_1 score of the estimators are higher for HQIC than the same for BIC.

In sum, for the considered application our proposed DTEC method provides better classification performance than DT and GLASSO approaches.

6. Conclusions

In this paper, we develop a new approach for estimating high-dimensional precision matrices, using the ℓ_1 penalization framework. The proposed method imposes a negative trace penalization on the recently introduced D-trace estimator. The additional penalty term controls the eigenvalues of the precision matrix estimator and diminishes the reduction of its largest eigenvalues and its trace. We conduct an extensive simulation study where we use several statistical loss functions and GGM prediction measures for the estimation evaluation. The results show that our proposed methodology outperforms DT and GLASSO methods for most of the considered scenarios. Moreover, we illustrate the advantageous performance of our proposed approach through an empirical application using breast cancer data. The application aimed to predict the patients with pCR. Furthermore, we propose a simplified version of our methodology, which leads to saving the computational time without having to sacrifice the performance significantly.

Acknowledgements

The author thanks professors Andrés M. Alonso and Francisco J. Nogales for their support and valuable comments.

Appendix A. Algorithm

In this section, we describe an algorithm for obtaining the estimator DTEC based on the alternating direction method. First, we introduce matrices Ω_0 and Ω_1. Next, we consider the following optimization problem equivalent to problem (3):

$$\begin{aligned} &\widehat{\Omega}_{\mathrm{DTEC}} = \arg\min_{\Omega_1 \succ \epsilon I} \frac{1}{2}\mathrm{trace}(\Omega^2 S) - (1+\gamma)\mathrm{trace}(\Omega) + \tau||\Omega_0||_1, \\ &\text{subject to} \quad \{\Omega, \Omega\} = \{\Omega_0, \Omega_1\} \end{aligned} \tag{A.1}$$

The Lagrangian of the problem (A.1) has the following form:

$$\begin{aligned} L(\Omega, \Omega_0, \Omega_1, \Lambda_0, \Lambda_1) &= \frac{1}{2}\mathrm{trace}(\Omega^2 S) - (1+\gamma)\mathrm{trace}(\Omega) + \tau||\Omega_0||_1 + h(\Omega_1 \succeq \epsilon I) \\ &+ \mathrm{trace}(\Lambda_0(\Omega - \Omega_0)) + \mathrm{trace}(\Lambda_1(\Omega - \Omega_1)) + \frac{\rho}{2}||\Omega - \Omega_0||_2^2 + \frac{\rho}{2}||\Omega - \Omega_1||_2^2, \end{aligned} \tag{A.2}$$

where ρ, Λ_0, Λ_1 are the multipliers and $h(\Omega_1 \succeq \epsilon I)$ is an indicator function, which returns 0 if $\Omega_1 \succeq \epsilon I$ is true and ∞ otherwise. For simplicity, we take $\rho = 1$. Assume that $(\Omega^t, \Omega_0^t, \Omega_1^t, \Lambda_0^t, \Lambda_1^t)$ is the solution at step t, for $t = 0, 1, 2, ...$ The solution is updated according to the following:

$$\Omega^{t+1} = \arg\min_{\Omega = \Omega^T} L(\Omega, \Omega_0^t, \Omega_1^t, \Lambda_0^t, \Lambda_1^t), \tag{A.3}$$

$$\{\Omega_0^{t+1}, \Omega_1^{t+1}\} = \underset{\Omega_0 = \Omega_0^T, \Omega_1 \succeq \epsilon I}{\operatorname{argmin}} L(\Omega^{t+1}, \Omega_0, \Omega_1, \Lambda_0^t, \Lambda_1^t), \tag{A.4}$$

$$\{\Lambda_0^{t+1}, \Lambda_1^{t+1}\} = \{\Lambda_0^t, \Lambda_1^t\} + \{\Omega^{t+1} - \Omega_0^{t+1}, \Omega^{t+1} - \Omega_1^{t+1}\}. \tag{A.5}$$

From the equation $(A.3)$ we have the following:

$$\Omega^{t+1} = \underset{\Omega = \Omega^T}{\operatorname{argmin}} \frac{1}{2}\mathrm{trace}(\Omega^2(S + 2I)) - \mathrm{trace}(\Omega((1+\gamma)I + \Omega_0^t + \Omega_1^t - \Lambda_0^t - \Lambda_1^t)). \tag{A.6}$$

First, for any $p \times p$ symmetric matrix $X \succ 0$ and any $p \times p$ symmetric matrix Y we define a matrix $G(X, Y)$. Assuming that $X = UVU^T$ is the eigendecomposition of matrix X and $v_1 \geq ... \geq v_p$ are its eigenvalues, we define

$$G(X, Y) = U\{(U^T Y U) \circ C\}U^T, \tag{A.7}$$

where $C_{i,j} = \dfrac{2}{v_i + v_j}$ for $1 \leq i, j \leq p$ and $\circ$ denotes the Hadamard product of matrices. We can write the solution of the problem $(A.6)$ as $\Omega^{t+1} = G(S + 2I; (1+\gamma)I + \Omega_0^t + \Omega_1^t - \Lambda_0^t - \Lambda_1^t)$.[10]

[10] We refer to Theorem 1 in Zhang and Zou (2014) for detailed proof.

From the first part of the equation $(A.4)$ it follows that

$$\Omega_0^{t+1} = \underset{\Omega_0=\Omega_0^T}{\operatorname{argmin}} \frac{1}{2}\operatorname{trace}(\Omega_0^2) - \operatorname{trace}(\Omega_0(\Omega^{t+1} + \Lambda_0^t)) + \tau||\Omega_0||_1. \tag{A.8}$$

We rewrite the problem (A.8) in the following form

$$\Omega_0^{t+1} = \underset{\Omega_0=\Omega_0^T}{\operatorname{argmin}} \frac{1}{2}\operatorname{trace}(\Omega_0^2) - \operatorname{trace}(\Omega_0 A) + \tau||\Omega_0||_1, \tag{A.9}$$

where $A = \Omega^{t+1} + \Lambda_0^t$. It is easy to check that the solution of the problem (A.9) is given as $\Omega_0^{t+1} = T(A, \tau)$, where T is the soft thresholding operator defined by:

$$[T(A,\tau)]_{ij} = \operatorname{sign}(A_{ij})\max(|A_{ij}| - \tau, 0)I_{i\neq j} + A_{ij}I_{i=j} = \begin{cases} A_{ij}, & \text{if } i = j, \\ A_{ij} - \tau, & \text{if } i \neq j, A_{ij} > \tau, \\ A_{ij} + \tau, & \text{if } i \neq j, A_{ij} < -\tau, \\ 0, & \text{if } i \neq j, -\tau \leq A_{ij} \leq \tau \end{cases} \tag{A.10}$$

for $1 \leq i, j \leq p$. From the second part of the equation $(A.4)$ it follows that

$$\Omega_1^{t+1} = \underset{\Omega_1 \succeq \epsilon I}{\operatorname{argmin}} \frac{1}{2}\operatorname{trace}(\Omega_1^2) - \operatorname{trace}(\Omega_1(\Omega^{t+1} + \Lambda_1^t)). \tag{A.11}$$

The solution of the problem $(A.11)$ is given as $\Omega_1^{t+1} = \left[\Omega^{t+1} + \Lambda_1^t\right]_+$, where for any symmetric matrix X with an eigendecomposition $X = U\operatorname{diag}(v_1, ..., v_p)U^T$ the operator $[X]_+$ is defined as $[X]_+ = U\operatorname{diag}(\max\{v_1, \epsilon\}, ..., \max\{v_p, \epsilon\})U^T$. After having all the steps of the alternating direction method provided above, we describe Algorithm 1.

Algorithm 1 Alternating direction method

Step 1. Initialization: $t = 0$, $\Lambda_0^0 = \Lambda_1^0$, $\Theta_0^0 = \Theta_1^0$.

Step 2. Repeat the following sub-steps until convergence:

(a) Set t=t+1.

(b) Compute the matrix $\Theta^{t+1} = G(S + 2I, (1 + \gamma)I + \Theta_0^t + \Theta_1^t - \Lambda_0^t - \Lambda_1^t)$.

(c) Set $\Theta_1^{t+1} = [\Theta^{t+1} + \Lambda_1^t]_+$ and compute $\Theta_0^{t+1} = T(\Theta^{t+1} + \Lambda_0^t, \tau)$.

(d) Set $\Lambda_0^{t+1} = \Lambda_0^t + (\Theta^{t+1} - \Theta_0^{t+1})$ and $\Lambda_1^{t+1} = \Lambda_1^t + (\Theta^{t+1} - \Theta_1^{t+1})$.

We can significantly reduce the computational time of Algorithm 1 by ignoring the constraint $\Omega \succeq \epsilon I$ in the initial optimization problem (3). This enables us to omit the step $\Theta_1^{t+1} = [\Theta^{t+1} + \Lambda_1^t]_+$ from 2(c), which is the most computationally expensive part of the algorithm. We can call the optimization problem without the constraint $\Omega \succeq \epsilon I$ the secondary problem, defined as:

$$\tilde{\Omega} = \arg \min_{\Omega^T = \Omega} \frac{1}{2} \text{trace}(\Omega^2 S) - (1+\gamma)\text{trace}(\Omega) + \tau ||\Omega||_1, \tag{A.12}$$

Following Zhang and Zou (2014), we also present the simplified version of Algorithm 1.

Algorithm 2 Alternating direction method (simplified)

Step 1. Initialization: $t = 0$, Λ^0, $\Theta_0^0 = \text{diag}(S)^{-1}$.

Step 2. Repeat the following sub-steps until convergence:

(a) Set t=t+1.

(b) Compute the matrix $\Theta^{t+1} = G(S + 2I, (1+\gamma)I + \Theta_0^t - \Lambda^t)$.

(c) Compute $\Theta_0^{t+1} = T(\Theta^{t+1} + \Lambda^t, \tau)$.

(d) Set $\Lambda^{t+1} = \Lambda_0^t + (\Theta^{t+1} - \Theta_0^{t+1})$.

Step 3. Consider the converged Θ^t as the solution of the secondary problem (A.12).

Step 4. If $\lambda_{\min}(\tilde{\Theta}) > \epsilon$, report $\tilde{\Theta}$ as the solution of the initial problem. Otherwise, use Algorithm 1 with $\tilde{\Theta}$ as the starting value for Θ_0^0 and Θ_1^0.

In other words, if $\tilde{\Omega} \succeq \epsilon I$, we have $\hat{\Omega} = \tilde{\Omega}$, otherwise we use Algorithm 1 to find $\hat{\Omega}$ considering $\tilde{\Omega}$ as the initial value of $\hat{\Omega}$. It is clear that Algorithm 2 is not self-contained and the implementation of Algorithm 1 may be required for some iterations. However, the introduction of Algorithm 2 may save considerably the computational time.

For both algorithms we consider convergence if the following two conditions are satisfied:

$$\frac{||\Theta^{t+1} - \Theta^t||_2}{\max(1, ||\Theta^t||_2, ||\Theta^{t+1}||_2)} < 10^{-7}, \qquad \frac{||\Theta_0^{t+1} - \Theta_0^t||_2}{\max(1, ||\Theta_0^t||_2, ||\Theta_0^{t+1}||_2)} < 10^{-7},$$

For more details we refer to Zhang and Zou (2014).

Appendix B. Performance Measures for Simulation Study

Table B.3. Average measures (with standard deviations) over 100 replications: Model 1.

		BIC				HQIC			
	p	DT	DTEC	SDTEC	GLASSO	DT	DTEC	SDTEC	GLASSO
KLL	100	10.68 (0.61)	10.82 (1.03)	10.67 (1.24)	18.95 (1.38)	8.94 (0.62)	9.43 (0.68)	9.41 (0.66)	14.14 (1.11)
	200	31.07 (1.15)	24.92 (1.22)	25.05 (1.19)	49.43 (1.90)	23.23 (0.88)	22.76 (1.05)	22.88 (1.03)	39.96 (3.29)
	300	46.62 (1.14)	38.66 (1.53)	38.12 (1.28)	81.80 (4.85)	35.41 (1.05)	35.66 (1.25)	35.51 (1.25)	68.39 4.64)
RKLL	100	13.32 (0.87)	9.18 (0.69)	8.84 (1.23)	31.91 (2.92)	10.28 (0.88)	7.66 (0.61)	7.55 (0.48)	21.58 (2.30)
	200	43.52 (1.90)	20.54 (1.35)	21.09 (0.99)	90.22 (4.41)	29.11 (1.28)	17.49 (1.10)	18.70 (0.82)	67.97 (7.58)
	300	64.93 (2.07)	33.28 (2.89)	31.64 (1.02)	154.36 (12.41)	43.36 (1.43)	29.10 (1.68)	28.47 (0.92)	121.38 (11.46)
ℓ_2	100	4.77 (0.13)	3.64 (0.24)	3.51 (0.31)	6.76 (0.18)	4.19 (0.17)	3.22 (0.23)	3.17 (0.13)	5.90 (0.25)
	200	8.21 (0.13)	5.40 (0.28)	5.54 (0.15)	10.52 (0.14)	6.99 (0.13)	4.76 (0.29)	5.13 (0.14)	9.70 (0.33)
	300	10.03 (0.11)	6.98 (0.42)	6.74 (0.13)	13.31 (0.27)	8.49 (0.12)	6.38 (0.32)	6.26 (0.13)	12.49 (0.31)
ℓ_1	100	1.18 (0.04)	1.03 (0.06)	1.02 (0.06)	1.40 (0.03)	1.15 (0.06)	0.99 (0.07)	0.98 (0.07)	1.39 (0.04)
	200	1.30 (0.03)	1.08 (0.05)	1.09 (0.05)	1.46 (0.02)	1.24 (0.05)	1.04 (0.06)	1.07 (0.06)	1.48 (0.04)
	300	1.30 (0.03)	1.11 (0.05)	1.10 (0.04)	1.49 (0.02)	1.27 (0.04)	1.10 (0.05)	1.10 (0.05)	1.52 (0.03)
ℓ_{spec}	100	0.98 (0.03)	0.82 (0.05)	0.80 (0.05)	1.21 (0.02)	0.89 (0.04)	0.75 (0.05)	0.74 (0.05)	1.10 (0.03)
	200	1.13 (0.02)	0.87 (0.04)	0.88 (0.03)	1.30 (0.01)	1.02 (0.03)	0.81 (0.04)	0.84 (0.04)	1.23 (0.03)
	300	1.13 (0.02)	0.91 (0.04)	0.89 (0.03)	1.34 (0.02)	1.03 (0.02)	0.87 (0.04)	0.86 (0.03)	1.28 (0.02)
Sensitivity	100	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
	200	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
	300	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)
Specificity	100	0.971 (0.002)	0.981 (0.005)	0.977 (0.009)	0.945 (0.009)	0.944 (0.006)	0.964 (0.006)	0.963 (0.002)	0.898 (0.015)
	200	0.995 (0.0005)	0.992 (0.001)	0.992 (0.0006)	0.978 (0.003)	0.981 (0.001)	0.984 (0.003)	0.987 (0.001)	0.957 (0.009)
	300	0.995 (0.0003)	0.993 (0.001)	0.993 (0.0003)	0.987 (0.003)	0.982 (0.0007)	0.988 (0.001)	0.987 (0.0006)	0.974 (0.005)
MCC	100	0.710 (0.016)	0.787 (0.047)	0.764 (0.080)	0.585 (0.035)	0.581 (0.031)	0.669 (0.033)	0.665 (0.015)	0.460 (0.029)
	200	0.871 (0.012)	0.795 (0.023)	0.818 (0.012)	0.640 (0.026)	0.667 (0.012)	0.697 (0.035)	0.734 (0.013)	0.508 (0.044)
	300	0.834 (0.009)	0.787 (0.030)	0.767 (0.009)	0.664 (0.042)	0.595 (0.008)	0.679 (0.025)	0.669 (0.008)	0.530 (0.046)
F1	100	0.683 (0.019)	0.772 (0.054)	0.745 (0.091)	0.532 (0.043)	0.527 (0.038)	0.634 (0.041)	0.629 (0.018)	0.381 (0.034)
	200	0.865 (0.013)	0.795 (0.027)	0.805 (0.014)	0.590 (0.033)	0.624 (0.014)	0.661 (0.043)	0.706 (0.015)	0.425 (0.054)
	300	0.823 (0.010)	0.767 (0.036)	0.745 (0.010)	0.617 (0.053)	0.530 (0.010)	0.636 (0.032)	0.623 (0.011)	0.448 (0.059)

Table B.4. Average measures (with standard deviations) over 100 replications: Model 2.

		BIC				HQIC			
	p	DT	DTEC	SDTEC	GLASSO	DT	DTEC	SDTEC	GLASSO
KLL	100	11.59 (0.56)	11.34 (0.82)	11.26 (0.72)	20.72 (1.58)	9.47 (0.87)	10.12 (0.80)	10.02 (0.69)	15.42 (1.22)
	200	30.95 (1.20)	27.00 (1.34)	27.62 (1.18)	56.26 (1.40)	23.57 (0.84)	23.80 (1.11)	23.74 (1.09)	43.11 (3.37)
	300	46.77 (1.07)	42.42 (1.44)	42.30 (1.42)	95.97 (2.85)	36.36 (2.39)	37.62 (2.10)	37.45 (2.04)	75.74 (4.01)
RKLL	100	14.80 (0.80)	9.46 (1.02)	9.36 (0.55)	37.27 (3.83)	10.77 (1.51)	8.06 (0.74)	7.87 (0.57)	24.48 (2.81)
	200	43.04 (2.09)	22.00 (1.69)	23.50 (1.13)	113.36 (3.82)	28.90 (1.21)	18.60 (1.09)	18.97 (0.79)	77.13 (9.03)
	300	64.52 (1.85)	35.95 (1.57)	35.61 (1.04)	204.31 (8.71)	43.31 (4.54)	31.38 (1.85)	29.26 (2.04)	143.87 (11.53)
ℓ_2	100	3.21 (0.09)	2.36 (0.21)	2.36 (0.10)	4.83 (0.12)	2.71 (0.19)	2.09 (0.17)	2.05 (0.11)	4.25 (0.16)
	200	5.24 (0.11)	3.57 (0.21)	3.76 (0.13)	7.53 (0.05)	4.43 (0.11)	3.18 (0.18)	3.26 (0.11)	6.85 (0.20)
	300	6.41 (0.08)	4.65 (0.14)	4.62 (0.09)	9.56 (0.08)	5.36 (0.24)	4.28 (0.23)	4.00 (0.21)	8.85 (0.15)
ℓ_1	100	1.04 (0.06)	0.86 (0.08)	0.86 (0.07)	1.44 (0.03)	0.99 (0.08)	0.84 (0.09)	0.83 (0.08)	1.36 (0.04)
	200	1.17 (0.05)	0.95 (0.07)	0.98 (0.06)	1.54 (0.01)	1.09 (0.08)	0.90 (0.09)	0.91 (0.09)	1.46 (0.03)
	300	1.18 (0.05)	0.99 (0.06)	0.99 (0.06)	1.57 (0.01)	1.12 (0.06)	0.97 (0.07)	0.95 (0.07)	1.51 (0.02)
ℓ_{spec}	100	0.88 (0.05)	0.70 (0.07)	0.70 (0.06)	1.32 (0.03)	0.78 (0.07)	0.64 (0.07)	0.63 (0.06)	1.20 (0.04)
	200	1.02 (0.05)	0.77 (0.06)	0.80 (0.05)	1.46 (0.01)	0.89 (0.06)	0.70 (0.07)	0.72 (0.06)	1.36 (0.03)
	300	1.02 (0.04)	0.81 (0.04)	0.81 (0.04)	1.51 (0.01)	0.90 (0.05)	0.77 (0.06)	0.73 (0.06)	1.43 (0.02)
Sensitivity	100	1 (0)	1 (0)	1 (0)	0.996 (0.004)	1 (0)	1 (0)	1 (0)	0.999 (0.001)
	200	1 (0)	1 (0)	1 (0)	0.990 (0.005)	1 (0)	1 (0)	1 (0)	0.998 (0.001)
	300	1 (0)	1 (0)	1 (0)	0.979 (0.007)	1 (0)	1 (0)	1 (0)	0.996 (0.003)
Specificity	100	0.972 (0.002)	0.980 (0.004)	0.980 (0.002)	0.933 (0.010)	0.936 (0.014)	0.958 (0.009)	0.955 (0.007)	0.880 (0.017)
	200	0.992 (0.001)	0.991 (0.002)	0.992 (0.001)	0.972 (0.002)	0.974 (0.001)	0.981 (0.002)	0.982 (0.001)	0.941 (0.009)
	300	0.992 (0.0004)	0.993 (0.0006)	0.993 (0.0004)	0.983 (0.002)	0.975 (0.004)	0.986 (0.003)	0.983 (0.003)	0.963 (0.004)
MCC	100	0.719 (0.017)	0.776 (0.034)	0.772 (0.018)	0.544 (0.032)	0.558 (0.060)	0.642 (0.046)	0.627 (0.041)	0.426 (0.030)
	200	0.812 (0.022)	0.797 (0.036)	0.824 (0.028)	0.581 (0.017)	0.602 (0.010)	0.657 (0.023)	0.666 (0.011)	0.443 (0.0357)
	300	0.763 (0.010)	0.785 (0.016)	0.781 (0.010)	0.596 (0.019)	0.536 (0.047)	0.650 (0.041)	0.612 (0.054)	0.456 (0.028)
F1	100	0.694 (0.020)	0.761 (0.039)	0.756 (0.021)	0.482 (0.039)	0.498 (0.073)	0.600 (0.055)	0.582 (0.049)	0.343 (0.034)
	200	0.798 (0.026)	0.780 (0.042)	0.812 (0.033)	0.518 (0.021)	0.543 (0.012)	0.611 (0.029)	0.623 (0.014)	0.346 (0.043)
	300	0.739 (0.013)	0.765 (0.018)	0.761 (0.012)	0.536 (0.025)	0.455 (0.059)	0.599 (0.051)	0.551 (0.067)	0.356 (0.036)

Table B.5. Average measures (with standard deviations) over 100 replications: Model 3.

		BIC				HQIC			
	p	DT	DTEC	SDTEC	GLASSO	DT	DTEC	SDTEC	GLASSO
KLL	100	15.75 (0.51)	15.36 (1.04)	15.38 (1.21)	22.98 (0.72)	13.74 (0.41)	14.19 (0.69)	14.17 (0.68)	20.78 (0.68)
	200	34.09 (0.73)	33.40 (1.06)	33.41 (0.99)	50.60 (1.09)	33.30 (2.05)	30.52 (1.14)	30.36 (1.13)	46.51 (1.63)
	300	51.30 (1.43)	51.72 (1.32)	50.76 (1.08)	81.73 (1.97)	51.13 (0.92)	50.20 (1.42)	50.57 (1.13)	75.77 (0.50)
RKLL	100	36.58 (1.75)	27.02 (1.87)	27.02 (3.49)	64.83 (3.02)	28.74 (1.23)	22.43 (2.12)	23.20 (2.28)	55.48 (2.78)
	200	82.40 (2.42)	60.70 (3.75)	61.39 (1.90)	150.16 (4.85)	78.86 (8.90)	52.59 (3.11)	49.66 (3.62)	131.92 (7.13)
	300	122.81 (5.61)	100.01 (6.10)	91.82 (2.17)	252.30 (9.36)	122.47 (3.12)	84.75 (4.95)	91.86 (2.36)	224.33 (2.05)
ℓ_2	100	10.34 (0.08)	9.65 (0.15)	9.63 (0.28)	11.41 (0.06)	9.84 (0.09)	9.20 (0.25)	9.29 (0.29)	11.14 (0.07)
	200	14.99 (0.07)	14.04 (0.20)	14.08 (0.08)	16.51 (0.06)	14.84 (0.38)	13.59 (0.21)	13.38 (0.22)	16.21 (0.11)
	300	18.35 (0.12)	17.57 (0.24)	17.24 (0.08)	20.53 (0.10)	18.34 (0.08)	16.89 (0.24)	17.24 (0.09)	20.23 (0.02)
ℓ_1	100	3.48 (0.04)	3.37 (0.05)	3.36 (0.05)	3.58 (0.02)	3.42 (0.04)	3.31 (0.05)	3.32 (0.05)	3.59 (0.03)
	200	3.54 (0.03)	3.43 (0.04)	3.44 (0.04)	3.64 (0.02)	3.53 (0.03)	3.41 (0.04)	3.40 (0.04)	3.65 (0.03)
	300	3.56 (0.03)	3.48 (0.03)	3.46 (0.03)	3.65 (0.01)	3.56 (0.03)	3.44 (0.03)	3.46 (0.03)	3.66 (0.02)
ℓ_{spec}	100	3.22 (0.02)	3.07 (0.04)	3.06 (0.06)	3.44 (0.01)	3.11 (0.02)	2.96 (0.06)	2.98 (0.06)	3.38 (0.01)
	200	3.29 (0.01)	3.14 (0.03)	3.15 (0.02)	3.49 (0.01)	3.26 (0.06)	3.07 (0.03)	3.03 (0.04)	3.45 (0.01)
	300	3.29 (0.02)	3.19 (0.03)	3.15 (0.02)	3.53 (0.01)	3.29 (0.01)	3.10 (0.03)	3.15 (0.02)	3.50 (0.003)
Sensitivity	100	0.047 (0.001)	0.043 (0.005)	0.045 (0.008)	0.046 (0.003)	0.070 (0.002)	0.060 (0.012)	0.059 (0.015)	0.067 (0.006)
	200	0.025 (0.0007)	0.021 (0.001)	0.021 (0.0006)	0.022 (0.001)	0.029 (0.009)	0.032 (0.004)	0.035 (0.005)	0.032 (0.003)
	300	0.020 (0.001)	0.014 (0.001)	0.016 (0.0004)	0.012 (0.001)	0.020 (0.0005)	0.018 (0.002)	0.016 (0.0004)	0.017 (0.0008)
Specificity	100	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)
	200	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)
	300	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)
MCC	100	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)
	200	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)
	300	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)	NA (NA)
F1	100	0.090 (0.003)	0.083 (0.009)	0.086 (0.015)	0.088 (0.006)	0.131 (0.004)	0.113 (0.021)	0.111 (0.026)	0.126 (0.012)
	200	0.049 (0.001)	0.042 (0.002)	0.041 (0.001)	0.043 (0.003)	0.056 (0.018)	0.062 (0.008)	0.069 (0.009)	0.062 (0.007)
	300	0.039 (0.002)	0.029 (0.002)	0.032 (0.0007)	0.025 (0.002)	0.039 (0.001)	0.036 (0.003)	0.032 (0.0008)	0.034 (0.001)

Table B.6. Average measures (with standard deviations) over 100 replications: Model 4.

		BIC				HQIC			
	p	DT	DTEC	SDTEC	GLASSO	DT	DTEC	SDTEC	GLASSO
KLL	100	16.30 (0.87)	14.95 (0.89)	15.04 (1.11)	22.60 (0.70)	13.34 (0.43)	13.76 (0.66)	13.81 (0.57)	20.30 (0.74)
	200	33.45 (0.70)	32.81 (0.81)	32.81 (0.77)	50.07 (0.33)	33.13 (1.58)	29.90 (1.00)	29.99 (1.28)	45.91 (1.17)
	300	50.87 (2.20)	50.87 (1.39)	49.88 (1.07)	80.98 (2.57)	50.40 (0.82)	49.56 (1.10)	49.77 (1.15)	75.33 (0.94)
RKLL	100	38.63 (3.35)	25.97 (1.98)	26.27 (3.16)	62.60 (2.81)	27.41 (1.23)	21.54 (1.86)	22.44 (1.77)	53.10 (2.97)
	200	79.96 (2.53)	59.01 (3.29)	59.50 (1.85)	147.07 (1.35)	78.35 (7.20)	51.97 (2.60)	48.86 (4.40)	128.79 (5.14)
	300	121.68 (8.59)	97.93 (5.57)	90.04 (2.26)	247.89 (11.87)	119.69 (2.68)	82.66 (4.39)	89.30 (3.37)	221.61 (4.06)
ℓ_2	100	10.24 (0.19)	9.37 (0.17)	9.38 (0.26)	11.15 (0.06)	9.55 (0.09)	8.93 (0.23)	9.04 (0.23)	10.86 (0.09)
	200	14.77 (0.08)	13.81 (0.18)	13.85 (0.09)	16.32 (0.01)	14.70 (0.33)	13.44 (0.19)	13.19 (0.27)	16.01 (0.08)
	300	18.20 (0.19)	17.39 (0.22)	17.06 (0.09)	20.37 (0.13)	18.16 (0.08)	16.69 (0.22)	17.03 (0.17)	20.08 (0.05)
ℓ_1	100	3.47 (0.04)	3.33 (0.05)	3.33 (0.06)	3.56 (0.02)	3.39 (0.04)	3.27 (0.06)	3.29 (0.05)	3.56 (0.03)
	200	3.53 (0.02)	3.43 (0.03)	3.43 (0.03)	3.62 (0.01)	3.53 (0.03)	3.41 (0.05)	3.39 (0.05)	3.63 (0.02)
	300	3.56 (0.03)	3.48 (0.03)	3.45 (0.04)	3.65 (0.01)	3.56 (0.03)	3.43 (0.04)	3.46 (0.04)	3.66 (0.02)
ℓ_{spec}	100	3.14 (0.04)	2.95 (0.04)	2.95 (0.06)	3.29 (0.01)	3,00 (0.03)	2.85 (0.06)	2.88 (0.06)	3.24 (0.02)
	200	3.25 (0.01)	3.11 (0.03)	3.11 (0.02)	3.45 (0.005)	3.24 (0.05)	3.05 (0.03)	3.01 (0.05)	3.41 (0.01)
	300	3.28 (0.02)	3.18 (0.03)	3.13 (0.02)	3.51 (0.01)	3.27 (0.02)	3.08 (0.04)	3.13 (0.03)	3.48 (0.008)
Sensitivity	100	0.132 (0.009)	0.133 (0.009)	0.134 (0.012)	0.130 (0.005)	0.170 (0.006)	0.155 (0.014)	0.151 (0.015)	0.150 (0.009)
	200	0.070 (0.001)	0.066 (0.001)	0.066 (0.001)	0.064 (0.001)	0.073 (0.009)	0.079 (0.005)	0.083 (0.007)	0.075 (0.003)
	300	0.050 (0.002)	0.044 (0.001)	0.046 (0.0007)	0.040 (0.001)	0.051 (0.001)	0.049 (0.002)	0.047 (0.003)	0.045 (0.005)
Specificity	100	0.988 (0.005)	0.987 (0.004)	0.987 (0.007)	0.982 (0.003)	0.962 (0.003)	0.973 (0.010)	0.976 (0.011)	0.960 (0.008)
	200	0.990 (0.0007)	0.993 (0.001)	0.993 (0.0006)	0.992 (0.001)	0.988 (0.008)	0.984 (0.003)	0.981 (0.005)	0.981 (0.003)
	300	0.990 (0.001)	0.995 (0.0009)	0.993 (0.0004)	0.996 (0.001)	0.990 (0.001)	0.992 (0.002)	0.993 (0.002)	0.991 (0.001)
MCC	100	0.262 (0.011)	0.261 (0.012)	0.262 (0.012)	0.234 (0.011)	0.224 (0.011)	0.237 (0.016)	0.240 (0.015)	0.194 (0.016)
	200	0.168 (0.005)	0.178 (0.005)	0.179 (0.004)	0.167 (0.005)	0.164 (0.013)	0.155 (0.008)	0.150 (0.013)	0.140 (0.009)
	300	0.126 (0.007)	0.143 (0.005)	0.137 (0.003)	0.142 (0.006)	0.125 (0.004)	0.130 (0.006)	0.136 (0.005)	0.124 (0.004)
F1	100	0.227 (0.011)	0.228 (0.011)	0.229 (0.015)	0.220 (0.006)	0.265 (0.008)	0.251 (0.014)	0.246 (0.014)	0.237 (0.009)
	200	0.128 (0.002)	0.122 (0.002)	0.121 (0.002)	0.119 (0.002)	0.132 (0.012)	0.139 (0.007)	0.145 (0.010)	0.133 (0.005)
	300	0.093 (0.003)	0.084 (0.002)	0.087 (0.001)	0.077 (0.002)	0.094 (0.002)	0.091 (0.003)	0.088 (0.004)	0.085 (0.002)

Table B.7. Average measures (with standard deviations) over 100 replications: Model 5.

		BIC				HQIC			
	p	DT	DTEC	SDTEC	GLASSO	DT	DTEC	SDTEC	GLASSO
KLL	100	13.31 (0.57)	12.24 (0.89)	12.31 (0.85)	23.60 (1.46)	10.85 (0.50)	10.74 (0.68)	10.70 (0.58)	18.88 (1.18)
	200	37.39 (2.16)	33.20 (1.48)	33.40 (1.63)	61.96 (2.50)	31.80 (1.57)	29.98 (0.91)	29.89 (0.83)	52.96 (2.65)
	300	60.49 (0.88)	55.62 (1.65)	55.45 (1.63)	98.60 (3.20)	52.01 (0.88)	51.28 (1.23)	51.20 (1.06)	88.28 (2.93)
RKLL	100	27.05 (1.96)	18.55 (2.18)	18.62 (2.24)	59.02 (5.66)	18.63 (1.36)	13.55 (1.25)	13.55 (1.22)	40.76 (4.44)
	200	84.32 (7.61)	52.24 (3.82)	52.55 (4.21)	175.09 (11.44)	64.22 (6.25)	41.93 (2.78)	42.13 (2.04)	135.28 (11.54)
	300	157.43 (4.16)	105.24 (6.54)	104.93 (5.40)	301.02 (15.68)	120.35 (3.04)	87.20 (5.30)	88.53 (2.66)	251.68 (13.73)
ℓ_2	100	1.83 (0.04)	1.58 (0.06)	1.58 (0.06)	2.65 (0.05)	1.63 (0.05)	1.44 (0.06)	1.44 (0.06)	2.44 (0.07)
	200	2.94 (0.08)	2.54 (0.06)	2.54 (0.07)	3.97 (0.04)	2.70 (0.09)	2.37 (0.07)	2.37 (0.06)	3.79 (0.07)
	300	3.74 (0.04)	3.39 (0.07)	3.39 (0.07)	4.94 (0.04)	3.53 (0.13)	3.31 (0.15)	3.31 (0.14)	4.79 (0.05)
ℓ_1	100	1.25 (0.03)	1.21 (0.05)	1.21 (0.05)	1.47 (0.02)	1.20 (0.04)	1.15 (0.05)	1.15 (0.05)	1.42 (0.02)
	200	1.63 (0.05)	1.56 (0.05)	1.56 (0.05)	1.74 (0.01)	1.57 (0.07)	1.50 (0.07)	1.50 (0.06)	1.72 (0.01)
	300	1.84 (0.08)	1.78 (0.14)	1.78 (0.14)	2.05 (0.01)	1.83 (0.46)	1.77 (0.39)	1.77 (0.38)	2.04 (0.01)
ℓ_{spec}	100	0.76 (0.02)	0.73 (0.03)	0.73 (0.03)	0.84 (0.01)	0.72 (0.03)	0.68 (0.03)	0.68 (0.03)	0.80 (0.01)
	200	0.90 (0.01)	0.87 (0.02)	0.87 (0.02)	0.92 (0.00)	0.87 (0.02)	0.84 (0.02)	0.84 (0.02)	0.90 (0.01)
	300	0.85 (0.06)	0.83 (0.13)	0.83 (0.12)	0.89 (0.01)	0.87 (0.24)	0.87 (0.25)	0.86 (0.25)	0.88 (0.01)
Sensitivity	100	0.144 (0.006)	0.131 (0.011)	0.131 (0.011)	0.156 (0.015)	0.183 (0.011)	0.167 (0.014)	0.167 (0.012)	0.212 (0.020)
	200	0.068 (0.007)	0.062 (0.005)	0.062 (0.006)	0.065 (0.005)	0.088 (0.010)	0.082 (0.005)	0.082 (0.004)	0.091 (0.011)
	300	0.043 (0.001)	0.037 (0.002)	0.037 (0.002)	0.046 (0.003)	0.059 (0.002)	0.049 (0.003)	0.048 (0.002)	0.060 (0.006)
Specificity	100	0.980 (0.002)	0.986 (0.005)	0.987 (0.005)	0.957 (0.009)	0.954 (0.007)	0.966 (0.008)	0.965 (0.007)	0.912 (0.017)
	200	0.991 (0.004)	0.994 (0.002)	0.994 (0.002)	0.982 (0.004)	0.979 (0.008)	0.983 (0.003)	0.983 (0.003)	0.962 (0.009)
	300	0.992 (0.0005)	0.995 (0.0009)	0.995 (0.0008)	0.989 (0.002)	0.981 (0.001)	0.989 (0.002)	0.989 (0.001)	0.980 (0.004)
MCC	100	0.240 (0.011)	0.249 (0.012)	0.249 (0.013)	0.184 (0.013)	0.210 (0.014)	0.222 (0.014)	0.221 (0.014)	0.156 (0.015)
	200	0.165 (0.008)	0.170 (0.006)	0.170 (0.007)	0.116 (0.009)	0.148 (0.011)	0.153 (0.008)	0.153 (0.007)	0.098 (0.009)
	300	0.114 (0.004)	0.121 (0.004)	0.120 (0.004)	0.105 (0.005)	0.099 (0.005)	0.108 (0.005)	0.109 (0.005)	0.095 (0.005)
F1	100	0.235 (0.008)	0.221 (0.013)	0.220 (0.013)	0.234 (0.014)	0.267 (0.011)	0.254 (0.014)	0.255 (0.013)	0.268 (0.012)
	200	0.123 (0.010)	0.115 (0.008)	0.114 (0.009)	0.115 (0.008)	0.150 (0.011)	0.143 (0.008)	0.142 (0.006)	0.145 (0.012)
	300	0.080 (0.002)	0.069 (0.004)	0.070 (0.004)	0.084 (0.005)	0.104 (0.003)	0.089 (0.005)	0.087 (0.003)	0.104 (0.008)

Table B.8. Average measures (with standard deviations) over 100 replications: Model 6.

		BIC				HQIC			
	p	DT	DTEC	SDTEC	GLASSO	DT	DTEC	SDTEC	GLASSO
KLL	100	19.67 (1.86)	19.40 (1.11)	19.50 (1.01)	30.42 (1.69)	15.71 (1.31)	16.71 (0.96)	16.78 (0.95)	24.63 (1.36)
	200	46.30 (2.29)	45.63 (1.51)	45.37 (1.41)	70.47 (2.69)	39.99 (1.71)	41.51 (1.39)	41.70 (1.36)	61.12 (2.01)
	300	84.45 (3.63)	77.81 (2.66)	77.58 (3.08)	119.27 (3.64)	74.74 (1.45)	70.72 (1.62)	71.73 (1.12)	106.39 (3.18)
RKLL	100	44.73 (7.44)	32.36 (3.44)	32.99 (3.16)	88.30 (8.13)	28.10 (5.37)	23.06 (2.65)	23.25 (3.15)	60.90 (6.18)
	200	114.93 (10.88)	83.67 (5.63)	83.04 (4.78)	217.72 (13.82)	84.49 (7.47)	67.50 (4.59)	67.04 (6.11)	170.55 (9.83)
	300	234.17 (16.41)	147.95 (9.38)	148.14 (9.74)	383.36 (20.13)	189.14 (8.01)	123.71 (8.36)	126.28 (3.56)	315.08 (16.68)
ℓ_2	100	2.24 (0.10)	2.04 (0.07)	2.05 (0.07)	2.84 (0.04)	1.93 (0.12)	1.79 (0.08)	1.80 (0.09)	2.66 (0.05)
	200	3.19 (0.07)	2.95 (0.06)	2.95 (0.06)	4.08 (0.04)	2.95 (0.07)	2.80 (0.07)	2.79 (0.08)	3.92 (0.04)
	300	4.19 (0.08)	3.72 (0.07)	3.72 (0.08)	5.11 (0.04)	3.96 (0.05)	3.53 (0.08)	3.55 (0.05)	4.96 (0.04)
ℓ_1	100	1.56 (0.05)	1.52 (0.05)	1.53 (0.04)	1.78 (0.01)	1.45 (0.06)	1.42 (0.05)	1.42 (0.05)	1.73 (0.02)
	200	1.87 (0.02)	1.85 (0.02)	1.85 (0.02)	1.97 (0.01)	1.84 (0.02)	1.83 (0.04)	1.83 (0.03)	1.95 (0.01)
	300	2.09 (0.05)	2.01 (0.03)	2.01 (0.03)	2.29 (0.01)	2.06 (0.04)	1.99 (0.03)	1.99 (0.03)	2.27 (0.01)
ℓ_{spec}	100	0.71 (0.03)	0.67 (0.03)	0.68 (0.03)	0.88 (0.01)	0.63 (0.04)	0.61 (0.03)	0.61 (0.03)	0.84 (0.01)
	200	0.82 (0.01)	0.79 (0.03)	0.79 (0.03)	0.90 (0.006)	0.79 (0.02)	0.77 (0.05)	0.77 (0.04)	0.88 (0.01)
	300	0.85 (0.01)	0.82 (0.03)	0.82 (0.03)	0.90 (0.004)	0.84 (0.01)	0.81 (0.03)	0.81 (0.01)	0.88 (0.01)
Sensitivity	100	0.094 (0.018)	0.079 (0.010)	0.078 (0.010)	0.097 (0.014)	0.155 (0.027)	0.128 (0.018)	0.128 (0.020)	0.161 (0.022)
	200	0.040 (0.004)	0.031 (0.003)	0.031 (0.003)	0.040 (0.005)	0.066 (0.007)	0.049 (0.007)	0.049 (0.009)	0.067 (0.008)
	300	0.022 (0.004)	0.021 (0.003)	0.021 (0.003)	0.025 (0.003)	0.033 (0.003)	0.035 (0.004)	0.031 (0.002)	0.040 (0.005)
Specificity	100	0.978 (0.010)	0.986 (0.005)	0.987 (0.005)	0.965 (0.009)	0.934 (0.020)	0.955 (0.012)	0.955 (0.014)	0.917 (0.018)
	200	0.984 (0.004)	0.991 (0.002)	0.991 (0.002)	0.987 (0.003)	0.963 (0.006)	0.977 (0.006)	0.977 (0.007)	0.971 (0.005)
	300	0.994 (0.003)	0.995 (0.002)	0.995 (0.002)	0.991 (0.002)	0.986 (0.003)	0.985 (0.003)	0.987 (0.002)	0.980 (0.004)
MCC	100	0.156 (0.011)	0.157 (0.010)	0.157 (0.010)	0.126 (0.010)	0.144 (0.011)	0.149 (0.010)	0.149 (0.010)	0.121 (0.010)
	200	0.075 (0.007)	0.080 (0.005)	0.080 (0.006)	0.087 (0.006)	0.065 (0.007)	0.072 (0.006)	0.072 (0.006)	0.089 (0.007)
	300	0.072 (0.004)	0.073 (0.003)	0.073 (0.004)	0.063 (0.005)	0.064 (0.004)	0.063 (0.005)	0.065 (0.004)	0.059 (0.005)
F1	100	0.167 (0.028)	0.145 (0.017)	0.143 (0.017)	0.170 (0.021)	0.251 (0.036)	0.218 (0.025)	0.216 (0.029)	0.257 (0.026)
	200	0.075 (0.008)	0.059 (0.005)	0.060 (0.005)	0.076 (0.010)	0.118 (0.012)	0.091 (0.012)	0.092 (0.016)	0.122 (0.013)
	300	0.043 (0.007)	0.040 (0.005)	0.041 (0.006)	0.048 (0.005)	0.063 (0.006)	0.066 (0.007)	0.060 (0.004)	0.076 (0.009)

References

Anderson, T., W. (2003). *An Introduction to Multivariate Statistical Analysis.* Wiley-Interscience.

Banerjee, O., El Ghaoui, L., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *The Journal of Machine Learning Research*, 9:485–516.

Banerjee, O., El Ghaoui, L., d'Aspremont, A., and Natsoulis, G. (2006). Convex optimization techniques for fitting sparse gaussian graphical models. Pittsburg. Proceedings of the 23^{rd} International Conference on Machine Learning.

Cai, T., Liu, W., and Luo, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

d'Aspremont, A., Banerjee, O., and Ghaoui, L. (2008). First-order methods for sparse covariance selection. *SIAM J. Matrix Anal Appl.*, 30:56–66.

Dempster, A. (1972). Covariance selection. *Biometrics*, 28(1):157–175.

Duchi, J., Gould, S., and Koller, D. (2008). Projected subgradient methods for learning sparse gaussians. In *Proceeding of the 24th Conference on Uncertainty in Artificial Intelligence*.

Fan, J., Feng, J., and Wu, Y. (2009). Network exploration via the adaptive lasso and scad penalties. *The Annals of Applied Statistics*, 3(2):521–541.

Frahm, G. and Memmel, C. (2010). Dominating estimator for minimum-variance portfolios. *Journal of Econometrics*, 159:289–302.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Goto, S. and Xu, Y. (2015). Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, 50(6):1415?1441.

Haff, L. R. (1980). Estimation of the inverse covariance matrix: Random mixtures of the inverse wishart matrix and the identity. *The Annals of Statistics*, 8(3):586–597.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 190–195.

Hsieh, C. J., A., S. M., S., D. I., and P., R. (2014). Quic: Quadratic approximation for sparse inverse covariance estimation. *Journal of Machine Learning Research*, 15:2911–2947.

Huang, S., Li, J., Sun, L., Ye, J., Fleisher, A., Wu, T., Chen, K., and Reiman, E. (2010). Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50:935–949.

Kourtis, A., Dotsis, G., and Markellos, N. (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking and Finance*, 36:2522–2531.

Kuerer, H. M., Newman, L. A., Smith, T. L., Ames, F. C., Hunt, K. K., Dhingra, K., Theriault, R. L., Singh, G., Binkley, S. M., Sneige, N., Buchholz, T. A., Ross, M. I., McNeese, M. D., Buzdar, A. U., Hortobagyi, G. N., and Singletary, S. E. (1999). Clinical course of breast cancer patients with complete pathologic primary tumor and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. *Journal of Clinical Oncology*, 17(2):460–469.

Lauritzen, S. (1996). *Graphical Models.* Clarendon Press. Oxford.

Li, L. and Toh, K. (2010). An inexact interior point method for ℓ_1-regurized sparse covariance selection. *Mathematical Programming Computation*, 2:291–315.

Lian, H. (2011). Shrinkage tuning parameter selection in precision matrices estimation. *Journal of Statistical Planning and Inference*, 141(8):2839–2848.

Lian, H. and Fan, Z. (2015). Estimation of a sparse and spiked covariance matrix. *Journal of Nonparametric Statistics*, 27(2):241–252.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta*, 405:442–451.

McLachlan, S. (2004). *Discriminant Analysis and Statistical Pattern Recognition.* Willey Interscience.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(2):1436–1462.

Peng, W., Wang, P., Zhou, N., and Zhu, J. (2009). Partial correlation estipation by joint sparse regression models. *Journal of the American Statistical Association*, 104(486):735–746.

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1).

Rothman, A., Bickel, P., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics*, 2:494–515.

Ryali, S., Chen, T., Supekar, K., and Menon, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *NeuroImage*, 59(4):3852–3861.

Scheinberg, K., Ma, S., and Goldfarb, D. (2010). Sparse inverse covariance selection via alternating linearization methods. In *Advances in Neural Information Processing Systems*.

Shi, L., Reid, L. H., Jones, W. D., Shippy, R., Warrington, J. A., Baker, S. C., Collins, P. J., deLongueville, F., Kawasaki, E. S., Lee, K. Y., Luo, Y., Sun, Y. A., Willey, J. C., Setterquist, R. A. Fischer, G. M., Tong, W., Dragan, Y. P., Dix, D. J., Frueh, F. W., Goodsaid, F. M., Herman, D., Jensen, R. V., Johnson, C. D., Lobenhofer, E. K., Puri, R. K., Scherf, U., Thierry-Mieg, J., Wang, C., Wilson, M., and Wolber, P. K. (2010). The microarray quality control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nature Biotechnology*, 28(8):827–838.

Stevens, G. V. G. (1998). On the inverse of the covariance matrix in portfolio analysis. *The Journal of Finance*, 53(5):1821–1827.

Stifanelli, P. F., Creanza, T. M., Anglani, R., Liuzzi, V. C., Mukherjee, S., Schena, F. P., and Ancona, N. (2013). A comparative study of covariance selection models for the inference of gene regulatory networks. *Journal of Biomedical Informatics*, 46:894–904.

Tarr, G., Müller, S., and Weber, N. C. (2016). Robust estimation of precision matrices under cellwise contamination. *Computational Statistics and Data Analysis*, 93:404–420.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1):267–288.

van Wieringen, W. N. and Peeters, C. F. W. (2016). Ridge estimation of inverse covariance matrices from high-dimensional data. *Computational Statistics and Data Analysis*.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.

Yin, J. and Li, J. (2013). Adjusting for high-dimensional covariates in sparse precision matrix estimation by ℓ_1-penalization. *Journal of Multivariate Analysis*, 116:365–381.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*, 11:2261–2286.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.

Zerenner, T., Friederichs, P., Lehnertz, K., and Hense, A. (2014). A gaussian graphical model approach to climate networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(2):023103.

Zhang, T. and Zou, H. (2014). Sparse precision matrix estimation via lasso penalized d-trace loss. *Biometrika*, 88:1–18.