Department of Mathematics

More on Estimation of Banded and Banded Toeplitz Covariance Matrices

Fredrik Berntsson Martin Singull

LiTH-MAT-R--2017/12--SE

LINKÖPING UNIVERSITY

Department of Mathematics
Linköping University
S-581 83 Linköping

More on estimation of banded and banded Toeplitz covariance matrices

Fredrik Berntsson * Martin Singull †

Abstract

In this paper we consider two different linear covariance structures, e.g., banded and bended Toeplitz, and how to estimate them using different methods, e.g., by minimizing different norms.

One way to estimate the parameters in a linear covariance structure is to use *tapering*, which has been shown to be the solution to a universal least squares problem. We know that tapering not always guarantee the positive definite constraints on the estimated covariance matrix and may not be a suitable method. We propose some new methods which preserves the positive definiteness and still give the correct structure.

More specific we consider the problem of estimating parameters of a multivariate normal $p-$dimensional random vector for (i) a banded covariance structure reflecting $m-$dependence, and (ii) a banded Toeplitz covariance structure.

1 Introduction

Many testing, estimation and confidence interval procedures discussed in the multivariate statistical literature are based on the assumption that the observation vectors are independent and normally distributed (Muirhead, 1982; Srivastava, 2002). The main reasons for this are that multivariate observations are often, at least approximately, normally distributed. Moreover, the multivariate normal distribution is mathematically tractable. Since normally distributed data can be modelled entirely in terms of their means and variances/covariances, these parameters specify the complete probability distribution of data. Estimating the mean and the covariance matrix is therefore a problem of great interest in statistics.

Patterned covariance matrices arise from a variety of contexts and have been studied by many authors. Below we mention some papers. Wilks (1946), is one of the early papers dealing with patterned structures, considered a set of measurements on k equivalent psychological tests. This led to a covariance matrix with equal diagonal elements and equal off-diagonal elements. Votaw (1948) extended this model to a set of blocks in which each block had a pattern. Olkin and Press (1969) considered a circular stationary model, where variables are thought of as being equally spaced around a circle, and the covariance between two variables depends on their distance. Olkin (1973) studied a multivariate

*Department of Mathematics, Linköping University, Sweden (fredrik.berntsson@liu.se)
†Department of Mathematics, Linköping University, Sweden (martin.singull@liu.se)

version in which each element was a matrix, and the blocks were patterned. More generally, permutation invariant covariance matrices may be of interest, see for example Nahtman (2006).

Banded covariance matrices and their inverses arise frequently in signal processing applications, including autoregressive or moving average image modelling, covariances of Gauss-Markov random processes (Woods, 1972; Moura and Balram, 1992), or numerical approximations to partial differential equations based on finite difference. Banded matrices are also used to model the correlation of cyclostationary processes in periodic time series (Chakraborty, 1998). Estimation of banded covariance structures has been studied by Ohlson et al. (2011) and Karlsson and Singull (2015).

There exist many papers on Toeplitz covariance matrices, e.g., see Marin and Dhorne (2002) and Christensen (2007), which all are banded matrices. To have a Toeplitz structure means that certain invariance conditions are fulfilled, e.g., equality of variances.

In this paper we study banded matrices with unequal elements except that certain covariances are zero and then also banded matrices that are also Toeplitz. The basic idea is that widely separated observations appear often to be uncorrelated and therefore it is reasonable to work with a banded covariance structure where all covariances more than m steps apart equal zero.

Furthermore, we focus on estimation of parameters of a multivariate random vector (with dimensionality p and sample size n). To avoid singularity, we limit ourselves to the case when the matrix dimension p is smaller than the sample size n.

Hence, let $\mathbf{X} \in \mathbb{R}^{p\times n}$ be matrix normally distributed, with independent columns that have the same mean, i.e.,

$$\mathbf{X} = (x_1, \dots, x_n) : p \times n, \quad x_i \underset{iid}{\sim} N_p(\mu, \mathbf{\Sigma}), \quad 1 \leq i \leq n,$$

where $\mu \in \mathbb{R}^p$ is the mean and $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ is the covariance matrix.

If there is no extra structure imposed on the covariance matrix $\mathbf{\Sigma}$, it can be estimated by the sample covariance matrix

$$\mathbf{S} = \frac{1}{n-1}(\mathbf{X} - \widehat{\mu}\mathbf{1}_n^T)(\mathbf{X} - \widehat{\mu}\mathbf{1}_n^T)^T, \tag{1}$$

where

$$\widehat{\mu} = \frac{1}{n}\mathbf{X}\mathbf{1}_n. \tag{2}$$

This estimator is theoretically known to produce *good results*, i.e., it is an unbiased and consistent of the covariance matrix $\mathbf{\Sigma}$.

In our work we are interested in the case when the covariance matrix $\mathbf{\Sigma}$ has a certain known structure, e.g., banded or Toeplitz. Since there is no guarantee that the sample covariance has the same structure we look at alternative estimators. The matrices produced by good estimators should also have other properties shared by all covariance matrices, e.g., be symmetric and positive definite. This is not always easy to achieve.

2 Estimating Banded Covariance Matrices

Suppose we have apriori knowledge about the structure of the covariance matrix $\mathbf{\Sigma}$, e.g. the matrix is banded with $2k+1$ non-zero diagonals. Estimation of this banded structure has earlier been studied by Ohlson et al. (2011) and Karlsson and Singull (2015). Given this structure, it is of course desirable that our estimator produces covariance matrices with the same correct banded structure. Additionally, all covariance matrices are symmetric and positive definite, and the same should hold for the estimated covariance matrices.

If $p < n$ the sample covariance matrix $\mathbf{S}$, given in (1), is symmetric and positive definite, but will not generally have the same banded structure as the covariance matrix $\mathbf{\Sigma}$. The simplest idea for imposing the correct banded structure on the estimated covariance matrix is known as *tapering*, i.e., starting from $\mathbf{S}$ we set the appropriate elements to zero to obtain the estimate $\widehat{\mathbf{\Sigma}}_{tap}$. In order to simplify the notation we also introduce an operator r $\mathcal{T}_k$ that performs tapering on a matrix and have the following definition.

Definition 2.1. Let $\mathbf{S} \in \mathbb{R}^{p\times p}$. Let $\mathcal{T}_k$ be an operator that sets the elements outside the diagonals $-k, \ldots, k$ to zero. The estimated covariance matrix by *tapering* is

$$\widehat{\mathbf{\Sigma}}_{tap} = \mathcal{T}_k(S). \tag{3}$$

The estimator $\widehat{\mathbf{\Sigma}}_{tap}$ is an unbiased and consistent estimator of the structured (banded) covariance matrix $\mathbf{\Sigma}$. However, as demonstrated by the following example $\widehat{\mathbf{\Sigma}}_{tap}$ may not be positive definite; even though $\mathbf{S}$ is positive definite.

Example 2.2. In order to illustrate our ideas we choose the following true covariance matrix and a mean vector,

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & 1/2 & 0 \\ 1/2 & 2 & 1/3 \\ 0 & 1/3 & 3 \end{pmatrix}, \quad \text{and,} \quad \mu = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

and generate $n = 10$ samples from the distribution $N_p(\mu, \mathbf{\Sigma})$ to obtain the matrix $\mathbf{X} \in \mathbb{R}^{p\times n}$. The sample mean $\widehat{\mu}$ is estimated using (2), and the sample covariance matrix $\mathbf{S}$ is calculated using the $\mathbf{QR}$ decomposition, i.e.,

$$\mathbf{QR} := (\mathbf{X} - \widehat{\mu}\mathbf{1}_n^T)^T, \quad \text{and,} \quad \mathbf{S} = \frac{1}{n}\mathbf{R}^T\mathbf{R}.$$

where $\mathbf{R} \in \mathbb{R}^{p\times p}$ is upper triangular. Specifically for this example we obtained the matrix

$$\mathbf{S} = \begin{pmatrix} 0.4197 & 0.6232 & -0.6454 \\ 0.6232 & 1.5415 & -1.4336 \\ -0.6454 & -1.4336 & 2.7330 \end{pmatrix},$$

and thus tapering would produce the estimate,

$$\widehat{\mathbf{\Sigma}}_{tap} = \begin{pmatrix} 0.4197 & 0.6232 & 0 \\ 0.6232 & 1.5415 & -1.4336 \\ 0 & -1.4336 & 2.7330 \end{pmatrix}.$$

for this particular case $\mathbf{S}$ is positive definite and its smallest eigenvalue is $\lambda_{min}(\mathbf{\Sigma}) = 0.1421$ while $\widehat{\mathbf{\Sigma}}_{tap}$ has the smallest eigenvalue $\lambda_{min}(\widehat{\mathbf{\Sigma}}_{tap}) = -0.0415$ and is therefore not positive definite. This demonstrates that tapering needs to be used with caution.

2.1 Tapering the Cholesky factor

In the previous section we observed that even though sample covariance $\mathbf{S}$ is positive definite the estimate $\widehat{\boldsymbol{\Sigma}}_{tap}$ obtained by tapering may not be. In this section we introduce a new estimator which is positive definite by construction. Before proceeding we give a useful relation between a positive definite banded matrix and its Cholesky factor.

Lemma 2.3. *Let* $\mathbf{S}$ *be a positive definite matrix and* $\mathbf{R}^T\mathbf{R} = \mathbf{S}$ *be its Cholesky decomposition. If* $\mathbf{S}$ *is banded with* $2k+1$ *non-zero diagonals then* $\mathbf{R}$ *is upper triangular with* $k+1$ *non-zero diagonals.*

Since every symetric positive definite matrix has a Cholesky decomposition; and considering Lemma 2.3 we can obtain an estimate of the covariance matrix, with the correct band structure, by instead using tapering on the Cholesky factor of $\mathbf{S}$. In this section we assume that the exact covariance matrix $\boldsymbol{\Sigma}$ is banded, with $2k+1$ non-zero diagonals, and that $\mathbf{S}$ is sample covariance matrix as defined by (1). We introduce the estimator and discuss its properties.

Definition 2.4. The estimator $\widehat{\boldsymbol{\Sigma}}_{chol}$ is given by the relations:

$$\widehat{\boldsymbol{\Sigma}}_{chol} = \mathbf{R}_k^T\mathbf{R}_k, \quad \mathbf{R}_k = \mathcal{T}_k(\mathbf{U}), \quad \mathbf{S} = \mathbf{U}^T\mathbf{U}. \quad \square$$

The estimate $\widehat{\boldsymbol{\Sigma}}_{chol}$ is positive definite and banded, with $2k+1$ non-zero diagonals. Thus the estimator has the correct structure. We also expect the estimate to be asymptotically correct and unbiased.

Example 2.5. Using the data from Example 2.2 we compute the Cholesky factor

$$\mathbf{S} = \begin{pmatrix} 0.4197 & 0.6232 & -0.6454 \\ 0.6232 & 1.5415 & -1.4336 \\ -0.6454 & -1.4336 & 2.7330 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 0.6478 & 0.9620 & -0.9963 \\ 0 & 0.7848 & -0.6054 \\ 0 & 0 & 1.1722 \end{pmatrix}$$

such that $\mathbf{S} = \mathbf{R}^T\mathbf{R}$. Tapering is applied to $\mathbf{R}$ gives

$$\mathbf{R}_1 = \begin{pmatrix} 0.6478 & 0.9620 & 0 \\ 0 & 0.7848 & -0.6054 \\ 0 & 0 & 1.1722 \end{pmatrix}$$

and the estimate is

$$\widehat{\boldsymbol{\Sigma}}_{chol} = \mathbf{R}_1^T\mathbf{R}_1 = \begin{pmatrix} 0.4197 & 0.6232 & 0 \\ 0.6232 & 1.5415 & -0.4751 \\ 0 & -0.4751 & 1.7404 \end{pmatrix},$$

which is positive definite and has the correct band structure.

We observe that even if the estimate $\widehat{\boldsymbol{\Sigma}}_{tap}$, obtained by applying tapering on $\mathbf{S}$, is positive definite we generally get a different result by instead using tapering on the Cholesky factors to obtain $\widehat{\boldsymbol{\Sigma}}_{chol}$. This is potentially not so good since the sample covariance is known have good properties.

2.2 Fitting in the Frobenius norm

Again considering Lemma 2.3. If the covariance matrix $\mathbf{\Sigma}$ is banded with $2k+1$ non-zero diagonals. Then we can write $\mathbf{\Sigma} = \mathbf{R}_k^T\mathbf{R}_k$, where $\mathbf{R}_k$ is upper triangular and has $k+1$ non-zero diagonals. Writing the covariance estimate in terms of its Cholesky factor means the estimator is, by construction, positive semi definite. This leads us to the following definition.

Definition 2.6. The estimator $\widehat{\mathbf{\Sigma}}_F$ is given by $\widehat{\mathbf{\Sigma}}_F = \mathbf{R}_k^T\mathbf{R}_k$, where $\mathbf{R}_k$ is a minimizer of

$$\min_{\mathbf{R}_k} \|\mathbf{R}_k^T\mathbf{R}_k - \mathbf{S}\|_F, \tag{4}$$

where $\|\cdot\|_F$ is the Frobenius norm, and the minimum is taken over all upper triangular banded matrices.

Since the Frobenius norm satisfies

$$\|\mathbf{A}\|_F = \|\mathrm{vec}(\mathbf{A})\|_2, \qquad \forall \mathbf{A} \in \mathbb{R}^{p\times p},$$

(see (Golub and Van Loan , 1996) for details) the estimate $\widehat{\mathbf{\Sigma}}_F$ is obtained by solving a small non-linear least squares problem (Björck , 1996). The parameters of the problem are the non-zero elements, r_{ij}, of the Cholesky factor R_k. The solution is unique up to the sign of the diagonal elements of R_k. This in turn means that the estimator $\widehat{\mathbf{\Sigma}}_F$ is unique.

The least squares problem consists of a set of simple polynomial equations and is not at all difficult to solve. In our tests we use an implementation of the Neider-Mead Simplex search algorithm for unconstrained optimization. See Lagarias et al. (1998) for more details.

Lemma 2.7. *If $\mathcal{T}_k(\mathbf{S})$ is positive definite then $\widehat{\mathbf{\Sigma}}_F = \widehat{\mathbf{\Sigma}}_{tap}$.*

Proof. If the estimate obtained by tapering $\mathbf{S}$ is positive definite then it has a Cholesky decomposition: $\widehat{\mathbf{\Sigma}}_{tap} = \mathbf{R}_k^T\mathbf{R}_k$, where R_k has $k+1$ non-zero diagonals. Since $\mathbf{R}_k^T\mathbf{R}_k - \mathbf{S}$ is zero, except for outside the main $2k+1$ diagonals, the specific choice of $\mathbf{R}_k$ as the Cholesky factor of $\widehat{\mathbf{\Sigma}}_{tap}$ is a minimizer of (4).

Example 2.8. Again we use the same sample covariance matrix $\mathbf{S}$ as in the two previous examples. By finding an upper bidiagonal matrix $\mathbf{R}_1$ that minimize (4) we obtain the estimator $\widehat{\mathbf{\Sigma}}_F$. For this particular example we obtain

$$\mathbf{R}_1 = \begin{pmatrix} 0.6685 & 0.9021 & 0 \\ 0 & 0.8618 & -1.6544 \\ 0 & 0 & 0.0000 \end{pmatrix},$$

and

$$\widehat{\mathbf{\Sigma}}_F = \begin{pmatrix} 0.4469 & 0.6031 & 0 \\ 0.6031 & 1.5564 & -1.4258 \\ 0 & -1.4258 & 2.7371 \end{pmatrix}.$$

For this case tapering gives an estimate $\mathbf{\Sigma}_{tap}$ that has one negative eigenvalue; as a result the estimator $\mathbf{\Sigma}_F$ has one zero eigenvalue and is thus positive semi-definite.

From the above example we conclude that if tapering fails to produce a positive definite estimate $\widehat{\mathbf{\Sigma}}_{tap}$ then the estimate $\widehat{\mathbf{\Sigma}}_F$ obtained by minimizing the residual in the Frobenius norm will not be strictly positive definite. This is not desirable and therefore we also introduce a regularized variant of the estimator.

Definition 2.9. The estimator $\widehat{\mathbf{\Sigma}}_F^{\alpha}$, $\alpha > 0$, is given by $\widehat{\mathbf{\Sigma}}_F^{\alpha} = \mathbf{R}_k^T\mathbf{R}_k + \alpha I$, where $\mathbf{R}_k$ is a minimizer of

$$\min_{\mathbf{R}_k} \|\mathbf{R}_k^T\mathbf{R}_k + \alpha I - \mathbf{S}\|_F. \tag{5}$$

The above construction means that the smallest eigenvalue of the estimate $\widehat{\mathbf{\Sigma}}_F^{\alpha}$ is greater than or equal to α so the estimate is strictly positive definite. Furthermore, if $\lambda_{min}(\mathbf{S}) > \alpha$ then $\widehat{\mathbf{\Sigma}}_F^{\alpha}$ and $\widehat{\mathbf{\Sigma}}_F$ coincides. For a specific case the choice of α may be done from a priori knowledge about the problem, or based on $\mathbf{S}$, e.g., $\alpha = \lambda_{min}(\mathbf{S})$.

Lemma 2.10. *The estimator $\widehat{\mathbf{\Sigma}}_F^{\alpha}$, where $\alpha = \lambda_{min}(\mathbf{S})$, is an unbiased and consistent estimator of the covariance matrix $\mathbf{\Sigma}$ with correct banded structure.*

2.3 Fitting in the Euclidean norm

In the previous section we introduced estimators based on minimizing the residual $\|\widehat{\mathbf{\Sigma}} - \mathbf{S}\|_F$ in the Frobenius norm. The minimization was carried out under the assumption that the estimate should have a certain structure, e.g., a banded with $2k+1$ non-zero diagonals. This ensures that $\widehat{\mathbf{\Sigma}}$ and S are close elementwise. Other important properties of the estimate, such as the trace, may be quite different from $\mathbf{S}$.

Lemma 2.11. *For a symmetric matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$ the Euclidean operator norm is $\|\mathbf{A}\|_2 = \max|\lambda(A)|$.*

Definition 2.12. The estimator $\widehat{\mathbf{\Sigma}}_E^{\alpha}$, $\alpha > 0$, is given by

$$\widehat{\mathbf{\Sigma}}_E^{\alpha} = \mathbf{R}_k^T\mathbf{R}_k + \alpha I, \quad \min_{\mathbf{R}_k} \|\mathbf{R}_k^T\mathbf{R}_k + \alpha I - \mathbf{S}\|_2. \tag{6}$$

From Lemma 2.11 we condlude that the estimator $\widehat{\mathbf{\Sigma}}_E^{\alpha}$ will attempt to produce estimates that are banded, positive definite, and has eigenvalues as close to those of $\mathbf{S}$ as possible.

3 Estimating Covariance Matrices with Banded Toeplitz Structure

A matrix has Toeplitz structure if its constant along diagonals (Golub and Van Loan , 1996). In applications it is sometimes known that the true covariance matrix has Toeplitz structure. In such cases it is desirable that the estimator also has the same structure.

Previously, we defined *tapering* to mean simply setting elements to zero. In the case of a Toeplitz matrix an obvious extension is to also average along diagonals.

Definition 3.1. Let the covariance matrix $\mathbf{\Sigma}$ be tridiagonal and Toeplitz. The estimator $\widehat{\mathbf{\Sigma}}_{zav}$ is the tridoagonal toeplitz matrix, with diagonal elements

$$d_0 = \frac{1}{n}\sum_{i=1}^{n} s_{ii},$$

$$d_{-1} = d_1 = \frac{1}{2n-2}\sum_{i=1}^{n-1} s_{i+1,i} + s_{i,i+1}.$$

The estimator $\widehat{\mathbf{\Sigma}}_{zav}$ is symmetric and has Toeplitz structure but, like $\widehat{\mathbf{\Sigma}}_{tap}$, it is not nessecarily positive definite. In order to ensure positive definiteness of the estimator we again formulate the estimators in terms of a least squares problem involving a parametrization of the class of symmetric, positive definite, Toeplitz matrices. In order to present the our estimates we give a couple of lemmas.

Lemma 3.2. *Suppose* $\mathbf{R}_1$ *is an upper bidiagonal Toeplitz matrix. Then*

$$\mathbf{T}_1 = \mathbf{R}_1^T\mathbf{R}_1 + r_{12}^2 e_1 e_1^T, \tag{7}$$

is a symmetric, positive definite, tridiagonal, Toeplitz matrix.

Proof. Denote by $(\mathbf{R}_1)_{ii} = r_0$ and $(\mathbf{R}_1)_{i,i+1} = r_1$ diagonal elements of the matrix $\mathbf{R}_1$. The diagonal elements of $\mathbf{T}_1$ are $(\mathbf{T}_1)_{i,i-1} = (\mathbf{T}_1)_{i,i+1} = r_0 r_1$ and $(\mathbf{T}_1)_{i,i} = r_0^2 + r_1^2$. Thus, according to the Gershgorin theorem, any eigenvalue λ of $\mathbf{T}_1$ satisfies $|\lambda - (r_0^2 + r_1^2)| \leq 2|r_0 r_1|$. Since $r_0^2 + r_1^2 - 2|r_0 r_1| \geq (|r_0| - |r_1|)^2 > 0$ we conclude that $\lambda \geq 0$ so $\mathbf{T}_1$ is positive semi-definite.

Definition 3.3. Let the covariance matrix $\mathbf{\Sigma}$ be tridiagonal and Toeplitz. The estimator $\widehat{\mathbf{\Sigma}}_{toep}$ is a minimizer of the least squares problem

$$\min_{\mathbf{T}} \|\mathbf{T} - \mathbf{S}\|_F, \quad \mathbf{T} = \mathbf{R}_1^T\mathbf{R}_1 + r_{12}^2 e_1 e_1^2, \tag{8}$$

where the minimum is taken over all upper bidiagonal Toeplitz matrices $\mathbf{R}_1$.

Example 3.4. In order to illustrate the estimator $\widehat{\mathbf{\Sigma}}_{toep}$ we pick a covariance matrix with Toeplitz structure and a mean as follows:

$$\mathbf{\Sigma} = \begin{pmatrix} 2 & 1/3 & 0 \\ 1/3 & 2 & 1/3 \\ 0 & 1/3 & 2 \end{pmatrix}, \quad \text{and,} \quad \mu = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

and generate $n = 10$ samples from the distribution $N_p(\mu, \mathbf{\Sigma})$ to obtain the matrix $\mathbf{X} \in \mathbb{R}^{p\times n}$.

The mean $\widehat{\mu}$ and the sample covariance $\mathbf{S}$ were calculated as previously. For this particular example we obtained the matrix

$$\mathbf{S} = \begin{pmatrix} 1.1979 & -0.8063 & -0.7923 \\ -0.8063 & 3.6874 & 1.1895 \\ -0.7923 & 1.1895 & 1.7608 \end{pmatrix},$$

which is not a Toeplitz matrix. The error can be calculated using a Frobenius norm as $\|\mathbf{\Sigma} - \mathbf{S}\|_F = 2.98$. If we instead use the Toeplitz estimator, given in (8), we obtain

$$\widehat{\mathbf{\Sigma}}_{toep} = \begin{pmatrix} 2.2155 & 0.1916 & 0 \\ 0.1916 & 2.2155 & 0.1916 \\ 0 & 0.1916 & 2.2155 \end{pmatrix}.$$

Here the correct structure is enforced and since more information is used a smaller error $\|\Sigma - \widehat{\Sigma}_{toep}\|_F = 0.4687$ is achived.

We remark that here we defined the estimator $\widehat{\Sigma}_{toep}$ by minimizing a residual in the Frobenius norm but it is also possible to use $\|\cdot\|_2$ or $\|\cdot\|_\infty$ instead with similar results. Further, while we only show our estimator for the case of a tridiagonal Toeplitz matrix the same construction works for Toeplitz matrices with any number of non-zero diagonals.

4 Simulations

In this section we illustrate the performance of our suggested estimators by performing several numerical simulations.

Simulation 4.1. As an initial simulation we choose the true covariance matrix to be have a banded structure with $k = 1$. The covariance matrix and mean vector used for this simulation are similar to the ones in Example 2.2 but for two different dimensions $p = 3$ and 8, i.e., $\mu = \begin{pmatrix} 1 & \dots & p \end{pmatrix}^T$ and $\Sigma = (\sigma_{ij})_{i,j=1}^p$, with

$$\sigma_{ii} = i, \quad \text{for } i = 1, \dots, p,$$
$$\sigma_{i,i+1} = \sigma_{i+1,i} = \frac{1}{i+1}, \quad \text{for } i = 1, \dots, p-1,$$
$$\sigma_{ij} = 0, \quad \text{otherwise.}$$

In order to clearly illustrate the properties of the estimators we construct a series of simulations with different sample sizes n. In each case we generate n samples from the distribution $N_p(\mu, \Sigma)$ to obtain the matrix $\mathbf{X} \in \mathbb{R}^{p \times n}$, and compute the corresponding sample covariance matrix $\mathbf{S}$. The different estimators are then used to find estimates of the true covariance matrix with the correct structure. For each sample size n we performed 20 different simulations and report both the maximum and mean errors in Figure 1. In this experiment calculate the errors using the Frobenius norm $\|\cdot\|_F$, but the results are very similar if a different norm is used. As a comparison we also include the error if the sample covariance $\mathbf{S}$ is used to estimate Σ. The results show that all the estimators behave more or less the same. Generally as long as the correct band structure is enforced the estimator produces more or less equally good results.

Simulation 4.2. In the second simulation, we choose the true covariance matrix to have a banded Toeplitz structure. The covariance matrix and mean vector used for this simulation are similar to the ones in Example 3.4 but for two different dimensions $p = 3$ and 8, i.e., $\mu = \begin{pmatrix} 1 & \dots & p \end{pmatrix}^T$ and $\Sigma = (\sigma_{ij})_{i,j=1}^p$, with

$$\sigma_{ii} = 2, \quad \text{for } i = 1, \dots, p,$$
$$\sigma_{i,i+1} = \sigma_{i+1,i} = \frac{1}{3}, \quad \text{for } i = 1, \dots, p-1,$$
$$\sigma_{ij} = 0, \quad \text{otherwise.}$$

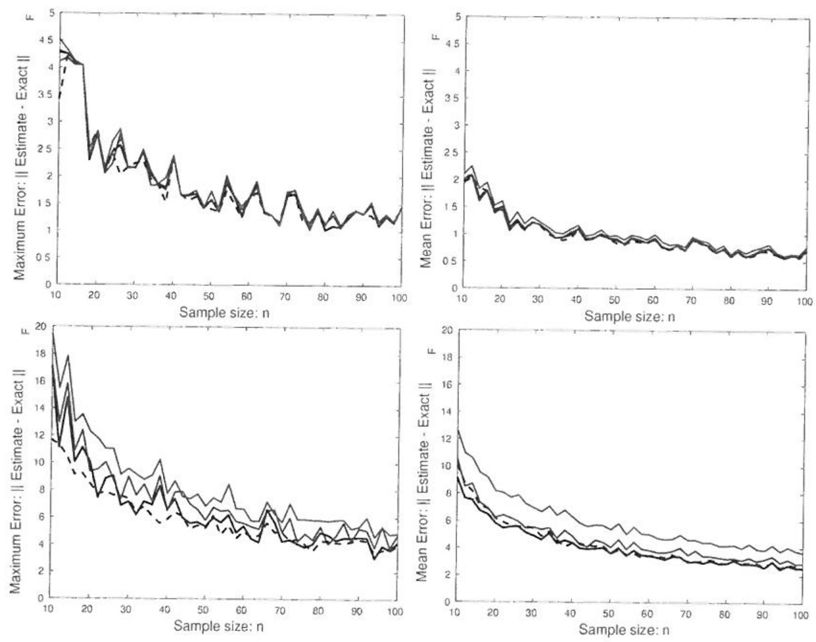

Figure 1: The errors for the different estimators as measured in the Frobenius norm, e.g., $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F$. For each sample size we performed 20 different series of simulations and we report the mean error (left) and the maximum error (right) for each sample size n. We performed the tests for dimension $p = 3$ (top) and $p = 8$ (bottom). We use the estimators $\widehat{\mathbf{\Sigma}}_{tap}$ (blue,dashed), $\widehat{\mathbf{\Sigma}}_{chol}$ (black,dashed), $\widehat{\mathbf{\Sigma}}_F$ (blue,solid) and $\widehat{\mathbf{\Sigma}}_E$ (black,solid). In all cases the estimators behave in a similar way and produce similar errors. As a comparison the error $\|\mathbf{S} - \mathbf{\Sigma}\|_F$ is also displayed (red,solid).

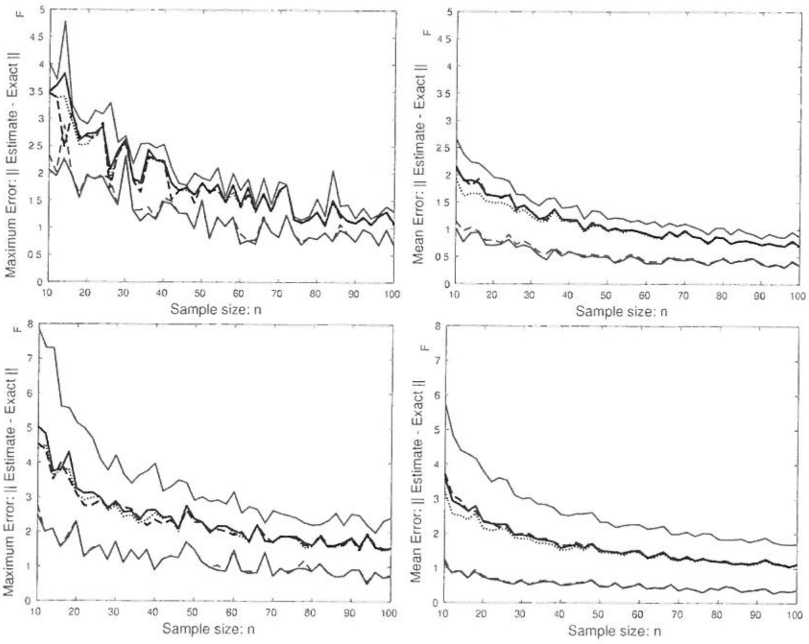

Figure 2: The errors for the different estimators as measured in the Frobenius norm, e.g., $\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_F$. For each sample size we preformed 20 different series of tests and we report the mean error (left) and the maximum error (right) for each sample size n. Again we performed the simulations for dimension $p = 3$ (top) and $p = 8$ (bottom). We display the error $\|\mathbf{S} - \mathbf{\Sigma}\|_F$ for the sample covariance matrix (red,solid). The estimators that take the band structure into acccount are $\widehat{\mathbf{\Sigma}}_{tap}$ (black,solid), $\widehat{\mathbf{\Sigma}}_{chol}$ (black,dashed), and $\widehat{\mathbf{\Sigma}}_F$ (black,dotted). Finally $\widehat{\mathbf{\Sigma}}_{zav}$ (blue,dashed) and $\widehat{\mathbf{\Sigma}}_{toep}$ (blue,solid) take the Toeplitz structure into account.

As previously we simulate the estimators for different sample sizes n. For each sample size we carry out 20 different simulations and calculate both the maximum and mean error.

The results are shown in Figure 2. In this simulation the true matrix $\mathbf{\Sigma}$ has a banded Toeplitz structure. This structure is not taken into account if $\mathbf{S}$ is used as an estimate which means large errors. The estimators $\mathbf{\Sigma}_{tap}$, $\mathbf{\Sigma}_{chol}$, and $\mathbf{\Sigma}_F$ take the band structure into account and produces better estimates. Finally the estimators $\mathbf{\Sigma}_{zav}$ and $\mathbf{\Sigma}_{toep}$ take the Toeplitz structure into account and produces the best results.

5 Concluding Remarks

In this paper we have introduced a number of different methods for estimating a covariance matrix with either banded or banded Toeplitz structure. Our estimators start from the sample covariance $\mathbf{S}$ and find the closest matrix, in a certain norm, that has the desired structure and properties.

The simplest estimator is *tapering* where certain elements of $\mathbf{S}$ are explicitly set to zero. We show that this gives accurate estimates but positive definiteness of the estimate is not ensured. By instead applying tapering on the Cholesky factor of $\mathbf{S}$ we obtain a positive semi-definite estimate. This approach is also shown to work well. Finally we give estimators that are defined by minimizing the difference between $\mathbf{S}$ and a parametrization of the class of matrices with the desired structure in either the Euclidean or Frobenius norms. This approach has the advantage that it can be regularized in the sense that a minimum bound for the eigenvalues of the estimate can be enforced, e.g., by $\min \lambda(\mathbf{S})$. This means that the estimates are positive definite.

We conclude the paper with a small simulation study, where we estimate covariance matrices with banded structure or banded Toeplitz structure. The results show that all estimators work well and the more information regarding the structure of the true covariance matrix we include, the better the results.

References

Björck, Å. (1996). *Numerical methods for least squares problems.* SIAM.

Chakraborty, M. (1998). An efficient algorithm for solving general periodic Toeplitz system. *IEEE Transactions on Signal Processing*, 46(3):784–787.

Christensen, L. P. B. (2007). An EM-algorithm for band-toeplitz covariance matrix estimation. *Submitted to IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations.* 3ed Johns Hopkins.

Karlsson, E. and Singull, M. (2015). More on explicit estimators for a banded covariance matrix. *Acta et Commentationes Universitatis Tartuensis de Mathematica* 19(1):49-62.

Lagarias, J.C. and Reeds, J. A. and Wright, M. H. and Wright, P. E. (1998). Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization* 9(1):112-147.

Marin, J. and Dhorne, T. (2002). Linear Toeplitz covariance structure models with optimal estimators of variance components. *Linear Algebra and Its Applications*, 354:195–212.

Moura, J. M. F. and Balram, N. (1992). Recursive structure of noncausal Gauss Markov random fields. *IEEE Transactions on Information Theory*, 38(2):334–354.

Muirhead, R. (1982). *Aspects of multivariate statistical theory.* John Wiley & Sons, New York, USA.

Nahtman, T. (2006). Marginal permutation invariant covariance matrices with applications to linear models. *Linear Algebra and Its Applications*, 417(1):183–210.

Ohlson, M., Andrushchenko, Z. and von Rosen, D. (2011). Explicit estimators under m-dependence for a multivariate normal distribution. *Annals of the Institute of Statistical Mathematics* 63(1):29-42.

Olkin, I. (1973). *Testing and estimation for structures which are circularly symmetric in blocks.* In D. G. Kabe and R. P. Gupta, eds, 'Multivariate Statistical Inference', pp. 183-195, North-Holland, Amsterdam.

Olkin, I. and Press, S. (1969). Testing and estimation for a circular stationary model. *The Annals of Mathematical Statistics*, 40:1358–1373.

Srivastava, M. S. (2002). *Methods of multivariate statistics.* Wiley-Interscience, New York, USA.

Votaw, D. F. (1948). Testing compound symmetry in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 19:447–473.

Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 17:257–28.

Woods, J. W. (1972). Two-dimensional discrete Markovian fields. *IEEE Transactions on Information Theory*, IT-18(2):232–240.