

Structure identification for a linearly structured covariance matrix: part II

Adam Mieldzioc

Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60-637 Poznań, Poland,
e-mail: adam.mieldzioc@up.poznan.pl

SUMMARY

Covariance matrices with a linear structure are widely used in multivariate analysis. The choice of covariance structure can be made from a set of possible linear structures. As a result, the most appropriate structure is determined by minimizing the discrepancy function. This paper is a continuation of previous work on identifying linear structures with an entropy loss function as a discrepancy function. We present extensive simulation studies on the correctness of identification with the assumed pentagonal banded Toeplitz structure.

Key words: covariance structure, compound symmetry matrix, banded Toeplitz matrix, identification, entropy loss function

1. Introduction

In this paper we continue the research from Mieldzioc (2022) about structure identification of a linearly structured covariance matrix under matrix normal distribution. We present new results based on simulation studies, where the covariance matrix Σ has a pentadiagonal banded Toeplitz structure.

The paper is organized as follows. In section 2 we introduce the main definitions, the structures to be considered, the discrepancy function (entropy loss function), the procedure for identification, and the algorithm for minimizing the entropy loss function on a linear structure $\mathcal{B}$. The results of simulation studies are given in section 3, and conclusions in section 4.

2. Problem formulation

In this section we present the model, the structures being considered and the method of identifying a covariance structure based on minimization of the entropy loss function, together with a corresponding algorithm.

2.1. Model with linearly structured covariance matrix

Assume that we observe n independent and identically distributed m -dimensional random vectors $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ with a normal distribution given by:

$$\mathbf{X}_i \sim N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}$ represents the expected value of $\mathbf{X}_i$ for $i = 1, \dots, n$ and $\boldsymbol{\Sigma}$ is the covariance matrix. All of these random vectors can be aggregated into an observation matrix $\mathbf{X}$ as follows

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^\top.$$

Thus, $\mathbf{X}$ has a matrix normal distribution:

$$\mathbf{X} \sim N_{n,m}(\mathbf{1}_n \boldsymbol{\mu}^\top, \mathbf{I}_n, \boldsymbol{\Sigma}),$$

where $\mathbf{1}_n$ is an n -dimensional vector of ones, and $\mathbf{I}_n$ is the $n \times n$ identity matrix.

We focus on determining the appropriate linear structure of the covariance matrix $\boldsymbol{\Sigma}$. A linear structure is defined as follows (cf. Anderson (1973)).

Definition 1. *The covariance matrix $\boldsymbol{\Sigma}$ has a linear structure if it can be expressed as a linear combination of linearly independent, known symmetric matrices $\mathbf{G}_i$:*

$$\boldsymbol{\Sigma} = \sum_{i=1}^{\ell} \sigma_i \mathbf{G}_i, \tag{1}$$

where $\sigma_i \in \mathbb{R}$ are unknown parameters. Furthermore, there exists at least one combination of $\sigma_1, \dots, \sigma_\ell$ such that $\boldsymbol{\Sigma}$ is a positive definite matrix.

The linear space generated by the basic matrices $\mathbf{G}_i$, $i = 1, \dots, \ell$, is referred to as the structure space. This basis can be transformed into an orthogonal basis $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_\ell$ using the Gram–Schmidt orthogonalization method. Recall that a basis is considered orthogonal if $\text{tr}(\mathbf{P}_i \mathbf{P}_j) = 0$ for every pair of matrices $\mathbf{P}_i$ and $\mathbf{P}_j$ ($i \neq j$), where $1 \leq i \leq \ell$ and $1 \leq j \leq \ell$.

2.2. Identification of a covariance structure

The entropy loss function is defined in the following way:

$$f(\mathbf{A}, \mathbf{B}) = \text{tr}(\mathbf{A}^{-1}\mathbf{B}) - \ln \det(\mathbf{A}^{-1}\mathbf{B}) - m,$$

where $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices of dimension $m \times m$.

To identify the linear structure of the covariance matrix Σ we consider k structures $\mathcal{C}_i$, $i = 1, \dots, k$. The set of positive definite matrices with structure $\mathcal{C}_i$ we denote as $\Psi_{\mathcal{C}_i}$. For every $\Psi_{\mathcal{C}_i}$ ($i = 1, \dots, k$) and $\mathbf{B} \in \Psi_{\mathcal{C}_i}$ we calculate the minimum of $f(\Sigma, \mathbf{B})$, denoted as:

$$\xi_{\mathcal{C}_i} = \min_{\mathbf{B} \in \Psi_{\mathcal{C}_i}} f(\Sigma, \mathbf{B}).$$

The $\mathcal{C}_i$ for which the value of $\xi_{\mathcal{C}_i}$ is smallest is indicated as the most appropriate structure.

Since the true Σ is unknown, we use the maximum likelihood estimator of Σ instead, defined as follows:

$$\mathbf{S} = \frac{1}{n} \mathbf{X} \mathbf{Q}_{1_n} \mathbf{X}^\top, \quad (2)$$

with $\mathbf{X}$ being an observation matrix and $\mathbf{Q}_{1_n} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top$.

2.3. Considered linear structures of the covariance matrix

We consider the following sets of linear structures: compound symmetry (CS), banded Toeplitz structure (T_p), and a mixed structure of CS and T_1 denoted by CS_T_1 . All matrices from these sets of structures have the following forms, respectively:

- Ψ_{CS} : $\alpha_1 \mathbf{I}_m + \alpha_2 \mathbf{J}_m$, in the orthogonal basis, $\beta_1 \mathbf{J}_m + \beta_2 (\mathbf{I}_m - \mathbf{J}_m)$, where $\mathbf{J}_m = \mathbf{1}_m \mathbf{1}_m^\top$;
- Ψ_{T_p} ($p \leq m - 1$): $\alpha_1 \mathbf{I}_m + \sum_{i=1}^p \alpha_i \mathbf{D}_i$, where $\mathbf{D}_i$ is a symmetric matrix with i -th superdiagonal and subdiagonal elements equal to 1 and all other elements equal to 0; in this paper we assume the Toeplitz covariance structure with $p = 1$ or $p = 2$;
- $\Psi_{CS_T_1}$: $\alpha_1 \mathbf{I}_m + \alpha_2 \mathbf{J}_m + \alpha_3 \mathbf{D}_1 = \beta_1 \mathbf{I}_m + \beta_2 \mathbf{D}_1 + \beta_3 (\mathbf{J}_m - \mathbf{I}_m - \mathbf{D}_1)$.

2.4. Algorithm

The algorithm for the determination of the minimum entropy loss function of a linear structure $\mathcal{B}$ is given in Mieldzioc (2022) and is written as follows

Algorithm 1 (Newton's method). *For a given starting point α and tolerance ϵ , repeat:*

1. *Compute the Newton step and decrement:*
 evaluate the gradient $\mathbf{g} = \nabla f(\alpha)$ and Hessian $\mathbf{H} = \mathbf{H}(\alpha)$ at starting point $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_\ell)$;
 $\Delta\alpha := -\mathbf{H}^{-1}\mathbf{g}; \tau := \mathbf{g}^\top \mathbf{H} \mathbf{g}.$
2. *Stopping criterion:*
 quit if $\tau^2/2 \leq \epsilon$.
3. *Backtracking line search: given parameters $a \in (0, 0.5)$ and $b \in (0, 1)$,*
 $t := 1;$
 while $\sum_{i=1}^{\ell} (\alpha_i - \Delta\alpha_i) \mathbf{P}_i$ is not a positive definite matrix,
 $t := bt;$
 while $f(\alpha + t\Delta\alpha) > f(\alpha) + at\mathbf{g}^\top \Delta\alpha,$
 $t := bt.$
4. *Update: $\alpha := \alpha + t\Delta\alpha$.*

3. Results

To test the correctness of the identification procedure based on the entropy loss function, we performed simulations assuming that the covariance matrix has the structure T_2 . In the simulation we use $\Sigma = \mathbf{I}_m + x\mathbf{D}_1 + 0.3\mathbf{D}_2 \in \Psi_{T_2}$, where $m \in \{5, 10, 20\}$ and, for respective m , $x \in \{-0.7, -0.65, \dots, 0.65, 0.7\}$, $x \in \{-0.65, -0.6, \dots, 0.6, 0.65\}$ and $x \in \{-0.6, -0.55, \dots, 0.55, 0.6\}$. This choice of x guarantees that for each m considered the covariance matrix Σ is positive definite and well-conditioned.

We can express the matrix Σ as a linear combination of orthogonal matrices $\mathbf{P}_1 = \mathbf{I}_m/\sqrt{m}$, $\mathbf{P}_2 = \mathbf{D}_1/\sqrt{2(m-1)}$ and $\mathbf{P}_3 = \mathbf{D}_2/\sqrt{2(m-2)}$. In this parameterization

$$\Sigma = \sqrt{m}\mathbf{P}_1 + x\sqrt{2(m-1)}\mathbf{P}_2 + 0.3\sqrt{2(m-2)}\mathbf{P}_3.$$

For sample size $n \in \{50, 100\}$, 1000 observation matrices $\mathbf{X} \sim N_{n,m}(\mathbf{0}, \mathbf{I}_n, \Sigma)$ were generated for every covariance matrix Σ . For each $\mathbf{X}$, the estimate $\mathbf{S}$ was computed. Moreover, for every $\mathbf{S}$, we calculated $\xi_{\mathcal{C}_i}$, where $\mathcal{C}_i$ takes values from the set $\{CS, T_1, T_2, CS_T1\}$. The starting point was taken as the

orthogonal projection of matrix $\mathbf{S}$ onto the space generated by the identity matrix, that is:

$$\boldsymbol{\alpha}_0 = (\text{tr } \mathbf{S}/m, 0, \dots, 0).$$

Since the entropy loss function has no upper bound, we used an adjusted value of the discrepancy ζ_{C_i} , that is

$$\eta_{C_i}(\mathbf{S}, \mathbf{B}_0) = 1 - \frac{1}{1 + f(\mathbf{S}, \mathbf{B}_0)},$$

where $\mathbf{B}_0 \in \Psi_{C_i}$ is the argument that minimizes the value of the entropy loss function for $C_i \in \{CS, T_1, T_2, CS_T1\}$ and given $\mathbf{S}$.

The outcomes are presented in Figures 1–6. The vertical axes show the averaged values of the adjusted entropy loss function. The horizontal axes show different covariance matrices $\boldsymbol{\Sigma}$ determined by x . The four structures considered are marked with different colors.

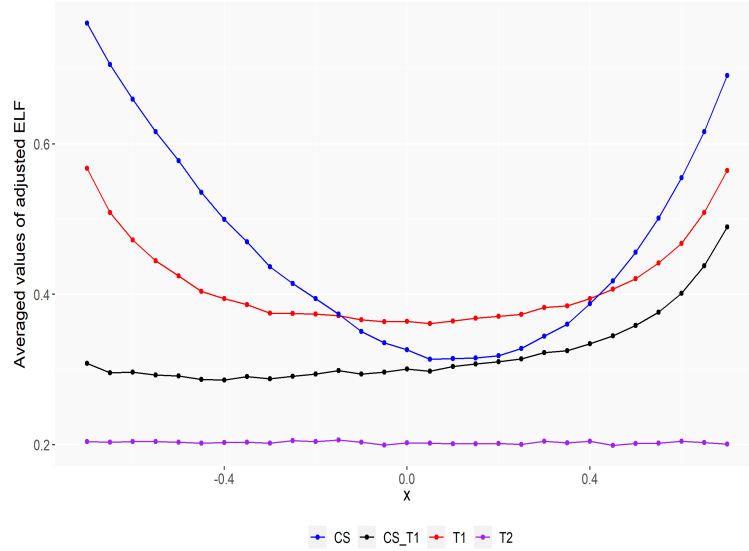


Figure 1. Averaged values of the adjusted entropy loss function for $\boldsymbol{\Sigma} = \mathbf{I}_5 + x\mathbf{D}_1 + 0.3\mathbf{D}_2$, where $x \in \{-0.7, -0.65, \dots, 0.65, 0.7\}$ and $n = 50$

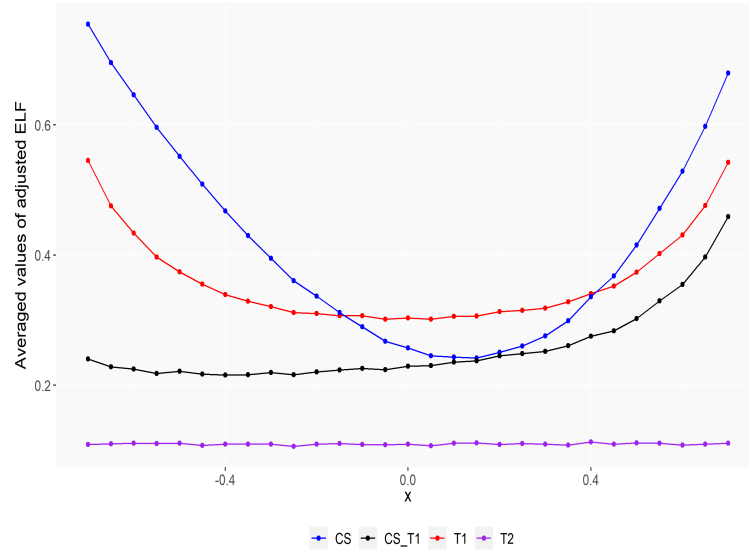


Figure 2. Averaged values of the adjusted entropy loss function for $\Sigma = \mathbf{I}_5 + x\mathbf{D}_1 + 0.3\mathbf{D}_2$, where $x \in \{-0.7, -0.65, \dots, 0.65, 0.7\}$ and $n = 100$

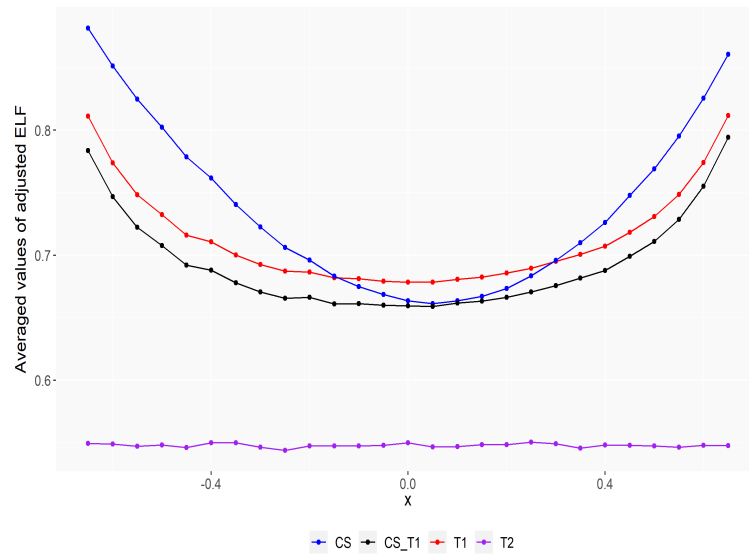


Figure 3. Averaged values of the adjusted entropy loss function for $\Sigma = \mathbf{I}_{10} + x\mathbf{D}_1 + 0.3\mathbf{D}_2$, where $x \in \{-0.65, -0.6, \dots, 0.6, 0.65\}$ and $n = 50$

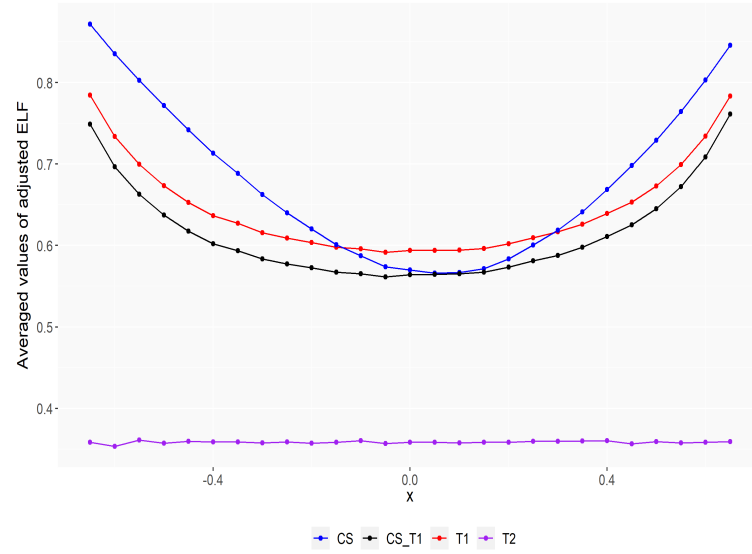


Figure 4. Averaged values of the adjusted entropy loss function for $\Sigma = \mathbf{I}_{10} + x\mathbf{D}_1 + 0.3\mathbf{D}_2$, where $x \in \{-0.65, -0.6, \dots, 0.6, 0.65\}$ and $n = 100$

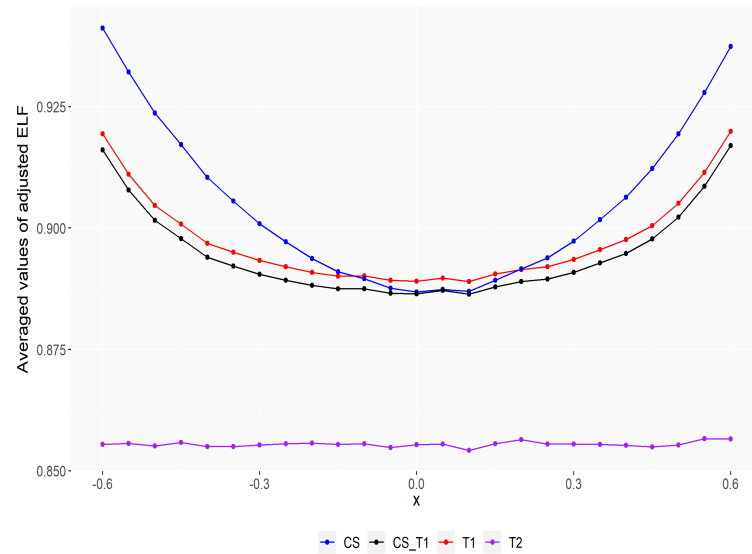


Figure 5. Averaged values of the adjusted entropy loss function for $\Sigma = \mathbf{I}_{20} + x\mathbf{D}_1 + 0.3\mathbf{D}_2$, where $x \in \{-0.6, -0.55, \dots, 0.55, 0.6\}$ and $n = 50$

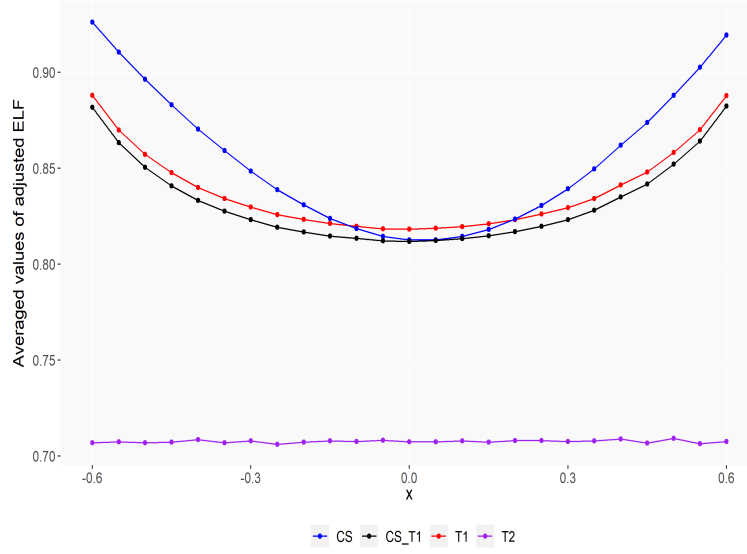


Figure 6. Averaged values of the adjusted entropy loss function for $\Sigma = \mathbf{I}_{20} + x\mathbf{D}_1 + 0.3\mathbf{D}_2$, where $x \in \{-0.6, -0.55, \dots, 0.55, 0.6\}$ and $n = 100$

It can be seen, that since the true covariance matrix in all of the cases has a structure from the set T_2 , as expected, the entropy loss function has the smallest values for T_2 . Thus, we may conclude that it detects the linear structure correctly.

We can observe that as the number of observations increases, the values of the adjusted entropy loss function decrease for given m ; and as the dimension m increases, the values of the adjusted entropy loss function increase.

4. Conclusions

In this paper we have presented new simulation study results on identifying a covariance matrix with structure T_2 using the entropy loss function. The simulation study also confirmed the correctness of the identification of structures using the entropy loss function. We can observe that the T_2 structure is not contained within any other of the structures considered (we know that $T_1 \subset T_2$ and $CS \subset CS_T_1$). In the work Mieldzioc (2022) the covariance matrix Σ had the CS_T_1 structure and the results for x close to 0 were very similar to the results for the CS structure.

REFERENCES

- Anderson T. W. (1973): Asymptotically Efficient Estimation of Covariance Matrices with Linear Structure. *The Annals of Statistics* 1(1): 135–141.
- Cui X., Li X., Zhao J., Zeng L., Zhang D., Pan J. (2016): Covariance structure regularization via Frobenius norm discrepancy. *Linear Algebra and its Application* 510: 124–145.
- Dey D.K., Srinivasan C. (1985): Estimation of a covariance matrix under Stein’s loss. *The Annals of Statistics* 13(4): 1581–1591.
- Filipiak K., Klein D., Markiewicz A., Mokrzycka M. (2021): Approximation with a Kronecker product structure with one component as compound symmetry or autoregression via entropy loss function. *Linear algebra and its Applications* 610: 625–646.
- James W., Stein C. (1961): Estimation with quadratic loss. In: Neyman, J. (ed.) *Proceedings of the Fourth Berkeley Symposium. In: Mathematical Statistics and Probability*, 1: 361–379. The Statistical Laboratory, University of California Press.
- Janiszewska M., Markiewicz A., Mokrzycka, M. (2020): Block Matrix Approximation Via Entropy Loss Function. *Applications of Mathematics* 65: 829–844.
- Ledoit O., Wolf M. (2004): A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88(2): 365–411.
- Lin L., Higham N. J., Pan J. (2014): Covariance structure regularization via entropy loss function. *Computational Statistics and Data Analysis* 72: 315–327.
- Magnus J., Neudecker H. (1986): Symmetry, 0-1 matrices and Jacobians, a review. *Econometric Theory* 2: 157–190.
- Mieldzioc A. (2022): Structure identification for a linearly structured covariance matrix. *Biometrical Letters* 59(2): 159–169.
- Mieldzioc A., Mokrzycka M., Sawikowska A. (2019): Covariance regularization for metabolomic data on the drought resistance of barley. *Biometrical Letters* 56(2): 165–181.