Covariance Structure Regularization via Frobenius-Norm Discrepancy

DOI:
10.1016/j.laa.2016.08.013

Covariance Structure Regularization via Frobenius-Norm Discrepancy

Xiangzhao Cui[1], Chun Li[1], Jine Zhao[1], Li Zeng[1], Defei Zhang[1] and Jianxin Pan[2*]

[1]*School of Mathematics, University of Honghe, Yunnan, China*

[2]*School of Mathematics, University of Manchester, UK*

Abstract

In many practical problems, the underlying structure of an estimated covariance matrix is usually blurred due to random noise, particularly when the dimension of the matrix is high. Hence, it is necessary to filter the random noise or regularize the available covariance matrix in certain senses, so that the covariance structure becomes clear. In this paper, we propose a new method for regularizing the covariance structure of a given covariance matrix. By choosing an optimal structure from an available class of covariance structures, the regularization is made in terms of minimizing the discrepancy, defined by Frobenius-norm, between the given covariance matrix and the class of covariance structures. A range of potential candidate structures, including the order-1 moving average structure, compound symmetry structure, order-1 autoregressive structure, order-1 autoregressive moving average structure, are considered. Simulation studies show that the proposed new approach is reliable in regularization of covariance structures. The proposed approach is also applied to real data analysis in signal processing, showing the usefulness of the proposed approach in practice.

Keywords: Covariance estimation; Covariance structure; F-norm; Regularization.

*Corresponding author: Email: Jianxin.Pan@manchester.ac.uk; Tel: +44 161 2755864; Fax: +44 161 275 5819.

1 Introduction

In many practical fields including signal processing [11], network [13], and control problems [6], a structured covariance matrices is really important and has to be estimated. However, the underlying structure of an estimated covariance matrix is usually blurred due to random noise, especially when the dimension of the covariance matrix is large. Although the estimation of covariance matrix has been studied widely in the literature (e.g., [9; 12]), it has received little attention for regularizing an available/estimated covariance matrix into the one with a clear structure.

Specifically, suppose A is a given $m \times m$ covariance matrix, that is, it is symmetric nonnegative definite. Let $\mathcal{S}$ be the set of all $m \times m$ positive definite covariance matrices with structure s, for example, compound symmetry or uniform covariance structure. A discrepancy between the given covariance matrix A and the set $\mathcal{S}$ is defined by

$$D(A, \mathcal{S}) = \min_{B \in \mathcal{S}} L(A, B), \tag{1.1}$$

where $L(A, B)$ is a measure of the distance between the two $m \times m$ matrices A and B. Assume there is a given class of k candidate covariance structures $\{s_1, s_2, \ldots, s_k\}$. Let S_i be the set of all covariance matrices with structure s_i. Denote the set of $m \times m$ covariance matrices with the likely structures by $\Omega = \cup_{i=1}^{k} S_i$. The discrepancy between a given covariance matrix A and the set Ω is then defined by

$$D(A, \Omega) = \min_{B \in \Omega} L(A, B). \tag{1.2}$$

The idea is that, in this set Ω, the structure with which A has the smallest discrepancy can be viewed as the most likely underlying structure of A, and the minimizer B with this particular structure is considered to be the regularized covariance matrix of A.

Very recently, Lin *et al.* [7] considered the use of the entropy loss function,

$$L(A, B) = \mathrm{tr}(A^{-1}B) - \log(\det(A^{-1}B)) - m,$$

also known as the Kullback-Leibler divergence, to measure the difference between the matrices A and B. However, this measure has some drawbacks, including that (a) it is a nonsymmetric measure in the sense that $L(A, B) \neq L(B, A)$, and (b) it requires the existence of the inverse of the given matrix A. In some circumstances, the inverse of A may not always exist, or it may exist but its computation is too intensive, for example, when the dimensional of A is rather high. To conquer the difficulty, in this paper we propose to consider the distance between two matrices A and B, defined by square of the Frobenius-norm, or hereafter F-norm,

$$L(A, B) = \mathrm{tr}\{(A - B)^T (A - B)\}. \tag{1.3}$$

It is worth mentioning that the matrix A is not necessarily a sample covariance matrix. It can be any estimates of a covariance matrix, obtained by various statistical methods such as those based on modified Cholesky decomposition methods [9; 14] and thresholding principal orthogonal complements [3] among others.

Regarding the likely structures of covariance matrix, in this paper we focus on the following four candidates that are commonly used in time series, longitudinal and spatial studies. Other candidate structures of covariance matrix may be studied in a similar manner.

(1) The first-order moving average structure, MA(1), has a tri-diagonal structure of covariance matrix,

$$B = \sigma^2 \begin{bmatrix} 1 & c & \cdots & 0 & 0 \\ c & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 1 & c \\ 0 & 0 & \cdots & c & 1 \end{bmatrix}, \tag{1.4}$$

where $\sigma^2 > 0$ and $-\dfrac{1}{2\cos(\pi/(m+1))} < c < \dfrac{1}{2\cos(\pi/(m+1))}$.

(2) The covariance of compound symmetry (CS) structure assumes that the correlation coefficients of any two observations are the same. In other words, the covariance matrix has the form

$$B = \sigma^2 \begin{bmatrix} 1 & c & \cdots & c & c \\ c & 1 & \ddots & \ddots & c \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c & \ddots & \ddots & 1 & c \\ c & c & \cdots & c & 1 \end{bmatrix}, \tag{1.5}$$

where $\sigma^2 > 0$ and $-1/(m-1) < c < 1$.

(3) The first-order autoregressive structure, AR(1), has the property that the correlation between any pair of observations decays exponentially towards zero as the distance between two observations increases. The covariance matrix is of the form

$$B = \sigma^2 \begin{bmatrix} 1 & c & \cdots & c^{m-2} & c^{m-1} \\ c & 1 & \ddots & \ddots & c^{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c^{m-2} & \ddots & \ddots & 1 & c \\ c^{m-1} & c^{m-2} & \cdots & c & 1 \end{bmatrix}, \tag{1.6}$$

where $\sigma^2 > 0$ and $-1 < c < 1$.

(4) More generally, the first-order autoregressive moving average structure, ARMA(1,1), has one more parameter than AR(1), reflecting an additional decrease in

correlation for each additional lag. The covariance matrix has the form

$$B = \sigma^2 \begin{bmatrix} 1 & r & rc & \cdots & rc^{m-4} & rc^{m-3} & rc^{m-2} \\ r & 1 & r & \ddots & \ddots & rc^{m-4} & rc^{m-3} \\ rc & r & 1 & \ddots & \ddots & \ddots & rc^{m-4} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ rc^{m-4} & \ddots & \ddots & \ddots & 1 & r & rc \\ rc^{m-3} & rc^{m-4} & \ddots & \ddots & r & 1 & r \\ rc^{m-2} & rc^{m-3} & rc^{m-4} & \cdots & rc & r & 1 \end{bmatrix}, \tag{1.7}$$

where $\sigma^2 > 0$, $-1 < c < 1$ and $-1 < r < 1$.

Owing to the fact $D(A, \Omega) = \min_{1 \le i \le k}\{D(A, S_i)\}$, the main task now is to calculate the discrepancy $D(A, S_i)$ for each of the candidate covariance structures listed in (1.4)-(1.7), where the covariance matrix A is given.

The rest of this paper is organized as follows. In section 2, we transform our problem into an optimization problem in numerical analysis and explore some of its general properties. In section 3, we show that the problem of finding B with structure MA(1), CS, AR(1) or ARMA(1,1) that minimizes $L(A, B)$ is reduced to computing the zeros of a nonlinear function. In section 4, we carry out simulation studies, illustrating how our techniques of computing the structured covariance matrix that minimizes the discrepancy function in (1.3) can be used in regularizing the underlying covariance structure. In section 5, we apply the proposed approach to a real data experiment in signal processing. Some further remarks and discussions are presented in section 6.

2 Problem of interest

We start by formulating the problem of interest and exploring some of its properties. Define $f : R_+^{m\times m} \to R$, where $R_+^{m\times m}$ is the set of all $m \times m$ symmetric positive definite matrices and $f(B) := L(A, B) = \mathrm{tr}\{(A-B)^T(A-B)\}$. Obviously, $\Omega \subset R_+^{m\times m}$. Our problem now reduces to

$$\begin{array}{cc} \min & f(B) \\ \text{subject to} & B \in \Omega \end{array}. \tag{2.8}$$

Let $\nabla_B f = (\partial f / \partial b_{ij})$ be the gradient of f, where b_{ij} is the (i, j) entry of B. Ignoring the symmetry of A and B and using results from Magnus and Neudecker [8] we have

$$\nabla_B \mathrm{tr}(A^T B) = A,$$

$$\nabla_B \mathrm{tr}(B^T B) = 2B,$$

and then

$$\nabla_B f = 2(B - A).$$

Write $b = \text{vec}(B) \in R^{m^2}$, where vec denotes the vector obtained by stacking the columns of its matrix argument on top of each other from first to last. Taking f as a function from R^{m^2} to R, the Hessian of f is then given by

$$\nabla_b^2 f := \left[\frac{\partial^2 f}{\partial b_i \partial b_j}\right] = 2(I_m \otimes I_m),$$

(See, e.g., [8, 10]). Since I_m is positive definite, $2(I_m \otimes I_m)$ is obviously positive definite, thus $f(B)$ is a strictly convex function of B.

On the other hand, the sets Ω of MA(1) and CS are obviously convex. Therefore when Ω is the set of positive definite matrices having one of the two structures the problem (2.8) is convex and so has a unique solution. When Ω is the set of AR(1) or ARMA(1,1) matrices, however, the problem is not convex because Ω is not convex. We will show later that only a local minimum of the problem can be expected to be found in these cases.

Note that when $\Omega = R_+^{m \times m}$, the minimum of $f(B)$ in (2.8) is obtained at $\nabla_B f = 0$, i.e., $B = A$, provided that A is positive definite.

3 Solution of problems

We begin by considering the matrices (1.4)-(1.7) one by one, for which the problem (2.8) is reduced to computing the zeros of a nonlinear function.

3.1 MA(1)

The matrix in (1.4) can be rewritten as

$$B(c, \sigma) = \sigma^2 \begin{bmatrix} 1 & c & \cdots & 0 & 0 \\ c & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & 1 & c \\ 0 & 0 & \cdots & c & 1 \end{bmatrix} = \sigma^2(I + cT_1), \tag{3.9}$$

where T_1 is a symmetric matrix with the first superdiagonal and subdiagonal equal to 1 and all other elements equal to 0. Note that the eigenvalues of $B(c, \sigma)$ are

$$\lambda_j = \sigma^2(1 + 2cs_j), \quad j = 1, \cdots, m,$$

where $s_j = \cos(\pi j/(m+1))$, see, e.g., [4], Sec. 28.5. Assuming $m \geq 2$, we have $s_1 > s_2 > \cdots \geq 0 \geq \cdots > s_m$, $s_j = -s_{m+1-j}$ and hence $B(c, \sigma)$ is positive definite if and only if

$$-\frac{1}{2s_1} < c < \frac{1}{2s_1}.$$

Given a covariance matrix A, the discrepancy function in (1.3) is now

$$f(c,\sigma) := \mathrm{tr}(A^T A) - 2\sigma^2(\mathrm{tr}(A) + \mathrm{tr}(AT_1)c) + \sigma^4(m + 2(m-1)c^2). \tag{3.10}$$

It follows that

$$\nabla f := \begin{bmatrix} \frac{\partial f}{\partial c} \\ \frac{\partial f}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} 4\sigma^4(m-1)c - 2\sigma^2\mathrm{tr}(AT_1) \\ 4\sigma^3(m + 2(m-1)c^2) - 4\sigma(\mathrm{tr}(A) + \mathrm{tr}(AT_1)c) \end{bmatrix}$$

and

$$\nabla^2 f := \begin{bmatrix} \frac{\partial^2 f}{\partial c^2} & \frac{\partial^2 f}{\partial c \partial \sigma} \\ \frac{\partial^2 f}{\partial c \partial \sigma} & \frac{\partial^2 f}{\partial \sigma^2} \end{bmatrix}$$

$$= \begin{bmatrix} 4\sigma^4(m-1) & 16\sigma^3(m-1)c - 4\sigma\mathrm{tr}(AT_1)) \\ 16\sigma^3(m-1)c - 4\sigma\mathrm{tr}(AT_1) & 12\sigma^2(m + 2(m-1)c^2) - 4(\mathrm{tr}(A) + \mathrm{tr}(AT_1)c) \end{bmatrix}.$$

So that the stationary points (c,σ) of $f(c,\sigma)$ must satisfy following equations

$$\begin{cases} \sigma^2 = \dfrac{\mathrm{tr}(AT_1)}{2(m-1)c}, \\ h(c) := m\mathrm{tr}(AT_1) - 2(m-1)\mathrm{tr}(A)c = 0. \end{cases}$$

Thus a unique stationary point is

$$\begin{cases} \sigma^2 = \dfrac{\mathrm{tr}(A)}{m} \\ c = \dfrac{m\mathrm{tr}(AT_1)}{2(m-1)\mathrm{tr}(A)}. \end{cases} \tag{3.11}$$

Since

$$(\nabla^2 f)_{11} = 4\sigma^4(m-1) > 0$$

and

$$\begin{aligned} \det(\nabla^2 f(c,\sigma)) &= 4(m-1)\sigma^4(12m\sigma^2 + 8\mathrm{tr}(AT_1)c - 4\mathrm{tr}(A)) - (4\sigma\mathrm{tr}(AT_1))^2 \\ &= 16\sigma^2(2(m-1)\sigma^2(\mathrm{tr}(A) + \mathrm{tr}(AT_1)c) - (\mathrm{tr}(AT_1))^2) \\ &= 32\frac{m-1}{m}\sigma^2(\mathrm{tr}(A))^2 \\ &> 0 \end{aligned}$$

at the stationary points (c,σ), therefore the Hessian matrix $\nabla^2 f$ is positive definite and so the stationary point is a minimum point.

We summarize the discussion above in the following theorem.

Theorem 3.1 *Given a covariance matrix A, there exists a unique positive definite matrix $B(c,\sigma)$ of the form (3.9) that minimizes the discrepancy function $f(c,\sigma) := L(A, B(c,\sigma))$ in (3.10). Furthermore, the minimum is achieved at (c,σ) given in (3.11).*

3.2 Compound Symmetry

The matrix in (1.5) can be rewritten as

$$B(c,\sigma) = \sigma^2 \begin{bmatrix} 1 & c & \cdots & c & c \\ c & 1 & \ddots & \ddots & c \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c & \ddots & \ddots & 1 & c \\ c & c & \cdots & c & 1 \end{bmatrix} = \sigma^2(I + c(ee^T - I)), \tag{3.12}$$

where $e = [1, \ldots, 1]^T \in R^m$. The eigenvalues of $B(c,\sigma)$ are $\sigma^2(1+(m-1)c)$ and $\sigma^2(1-c)$ of multiplicities 1 and $m-1$, respectively, so that $B(c,\sigma)$ is a positive definite matrix if and only if

$$-\frac{1}{m-1} < c < 1.$$

See, for example, [2], Lem. 2.1.

Given a covariance matrix A, denoting $t := \mathrm{tr}(A^T(ee^T - I))$, the discrepancy function is now given by

$$f(c,\sigma) = \mathrm{tr}(A^T A) - 2\sigma^2 \mathrm{tr}(A^T) - 2\sigma^2 ct + \sigma^4(m + m(m-1)c^2).$$

It follows that

$$\nabla f := \begin{bmatrix} \dfrac{\partial f}{\partial c} \\ \dfrac{\partial f}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} -2\sigma^2 t + 2\sigma^4 m(m-1)c \\ -4\sigma \mathrm{tr}(A^T) - 4\sigma ct + 4\sigma^3(m + m(m-1)c^2) \end{bmatrix}$$

and

$$\nabla^2 f := \begin{bmatrix} \dfrac{\partial^2 f}{\partial c^2} & \dfrac{\partial^2 f}{\partial c \partial \sigma} \\ \dfrac{\partial^2 f}{\partial c \partial \sigma} & \dfrac{\partial^2 f}{\partial \sigma^2} \end{bmatrix}$$

$$= \begin{bmatrix} 2\sigma^4 m(m-1) & -4\sigma t + 8\sigma^3 m(m-1)c \\ -4\sigma t + 8\sigma^3 m(m-1)c & -4\mathrm{tr}(A^T) - 4ct + 12\sigma^2(m + m(m-1)c^2) \end{bmatrix}.$$

The stationary points (c,σ) of $f(c,\sigma)$ must satisfy following equations

$$\begin{cases} \sigma^2 = \dfrac{\mathrm{tr}(A^T) + ct}{m + m(m-1)c^2}, \\ h(c) := -2\sigma^2 t + 2\sigma^4 m(m-1)c = 0. \end{cases}$$

Thus a unique stationary point is

$$\begin{cases} c = \dfrac{t}{(m-1)\mathrm{tr}(A^T)} \\ \sigma^2 = \dfrac{\mathrm{tr}(A^T) + ct}{m + m(m-1)c^2}, \end{cases} \tag{3.13}$$

where $t = \mathrm{tr}(A^T(ee^T - I))$. Since

$$(\nabla^2 f)_{11} = 2\sigma^4 m(m-1) > 0$$

and

$$\begin{aligned} \det(\nabla^2 f) &= 2\sigma^4 m(m-1)(-4\mathrm{tr}(A^T) - 4ct + 12\sigma^2(m + m(m-1)c^2)) \\ &- (8\sigma^3 m(m-1)c - 4\sigma t)^2 \\ &= 16\sigma^6 m^2(m-1) \\ &> 0 \end{aligned}$$

at the stationary points (c, σ), therefore $\nabla^2 f$ is positive definite and so the stationary point is a minimum point.

We summarize the above discussion in the following theorem.

Theorem 3.2 *Given a covariance matrix A, define $f(c,\sigma) := L(A, B(c,\sigma))$, where $B(c,\sigma)$ is a positive definite covariance matrix with compound symmetry structure as in (3.12). Then the global minimum of $f(c,\sigma)$ over $\sigma > 0$ and $c \in (-1/(m-1), 1)$ is achieved at (c,σ) given in (3.13).*

3.3 AR(1)

We rewrite B in (1.6) as

$$B(c,\sigma) = \sigma^2 \begin{bmatrix} 1 & c & \cdots & c^{m-2} & c^{m-1} \\ c & 1 & \ddots & \ddots & c^{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c^{m-2} & \ddots & \ddots & 1 & c \\ c^{m-1} & c^{m-2} & \cdots & c & 1 \end{bmatrix} = \sigma^2 \sum_{i=0}^{m-1} c^i T_i, \tag{3.14}$$

where $T_0 = I$ and T_i is a symmetric matrix with ones on the ith superdiagonal and subdiagonal and zeros elsewhere. It can be shown that the $k \times k$ leading principal minor of $B(c,\sigma)$ is $\sigma^{2k}(1-c^2)^{k-1}$ for $k = 2, \cdots, m$, see, e.g., [5], Prob.7.2, P12. Therefore, $B(c,\sigma)$ is a positive definite covariance matrix if and only if

$$-1 < c < 1.$$

The discrepancy function in (1.3) is now

$$f(c,\sigma) := \mathrm{tr}(A^TA) - 2\sigma^2\sum_{i=0}^{m-1} c^i\mathrm{tr}(AT_i) + \sigma^4(m + 2\sum_{i=1}^{m-1}(m-i)c^{2i}).$$

We find that

$$\nabla f := \begin{bmatrix} \dfrac{\partial f}{\partial c} \\ \dfrac{\partial f}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} -2\sigma^2 \sum\limits_{i=1}^{m-1} ic^{i-1}\mathrm{tr}(AT_i) + 4\sigma^4 \sum\limits_{i=1}^{m-1}(m-i)ic^{2i-1}) \\ -4\sigma \sum\limits_{i=0}^{m-1} c^i\mathrm{tr}(AT_i) + 4\sigma^3(m + 2\sum\limits_{i=1}^{m-1}(m-i)c^{2i}) \end{bmatrix}.$$

So the stationary points (c,σ) of $f(c,\sigma)$ must satisfy

$$\begin{cases} -\sum\limits_{i=1}^{m-1} ic^{i-1}\mathrm{tr}(AT_i) + \dfrac{2\sum\limits_{i=0}^{m-1} c^i\mathrm{tr}(AT_i)\sum\limits_{i=1}^{m-1}(m-i)ic^{2i-1}}{m + 2\sum\limits_{i=1}^{m-1}(m-i)c^{2i}} = 0, \\ \sigma^2 = \dfrac{\sum\limits_{i=0}^{m-1} c^i\mathrm{tr}(AT_i)}{m + 2\sum\limits_{i=1}^{m-1}(m-i)c^{2i}}. \end{cases} \tag{3.15}$$

Since $m + 2\sum\limits_{i=1}^{m-1}(m-i)c^{2i} > 0$, by rearranging the first equality in (3.15) we have

$$h(c) := -\sum_{i=1}^{m-1} ic^{i-1}\mathrm{tr}(AT_i)(m+2\sum_{i=1}^{m-1}(m-i)c^{2i}) + 2\sum_{i=0}^{m-1} c^i\mathrm{tr}(AT_i)\sum_{i=1}^{m-1}(m-i)ic^{2i-1} = 0.$$

Numerical experiments show that there exists at least one root of $h(c)$ in $(-1,1)$. Equivalently, the local minima of $f(c,\sigma)$ are achieved at the points (c,σ) satisfying (3.15).

We then summarize the discussion above in the following theorem.

Theorem 3.3 *Given a covariance matrix $A \in R^{m\times m}$, define $f(c,\sigma) := L(A, B(c,\sigma))$ where $B(c,\sigma)$ is a positive definite covariance matrix of the AR(1) model as in (3.14). Then the local minima of $f(c,\sigma)$ are attained at the points (c,σ) satisfying (3.15).*

3.4 ARMA(1,1)

Now we consider the problem for covariance matrix with structure of ARMA(1,1) , for which

$$B(r,c,\sigma)=\sigma^2\begin{bmatrix} 1 & r & rc & \cdots & rc^{m-4} & rc^{m-3} & rc^{m-2} \\ r & 1 & r & \ddots & \ddots & rc^{m-4} & rc^{m-3} \\ rc & r & 1 & \ddots & \ddots & \ddots & rc^{m-4} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ rc^{m-4} & \ddots & \ddots & \ddots & 1 & r & rc \\ rc^{m-3} & rc^{m-4} & \ddots & \ddots & r & 1 & r \\ rc^{m-2} & rc^{m-3} & rc^{m-4} & \cdots & rc & r & 1 \end{bmatrix}. \tag{3.16}$$

Let $q(t) = 1 + 2r\sum_{k=1}^{m-1} c^{k-1}cos(kt)$, then $B(r,c,\sigma)$ is positive-definite if and only if $q(t) \geq 0$ and $q(t) \not\equiv 0$ for all $t \in R$ (Parter, 1962, Remark II). Now the matrix B in (3.16) can be rewritten as

$$B(r,c,\sigma)=\sigma^2(I+r\sum_{i=1}^{m-1}c^{i-1}T_i),$$

where T_i is a symmetric matrix with ones on the ith superdiagonal and subdiagonal and zeros elsewhere.

The discrepancy function in (1.3) is now

$$f(r,c,\sigma)=\mathrm{tr}(A^TA)+\sigma^4(m+2r^2\sum_{i=1}^{m-1}(m-i)c^{2(i-1)})-2\sigma^2(\mathrm{tr}(A)+r\sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1}).$$

We then have the gradient of f

$$\nabla f := \begin{bmatrix} \frac{\partial f}{\partial r} \\ \frac{\partial f}{\partial c} \\ \frac{\partial f}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} 4\sigma^4 r\sum_{i=1}^{m-1}(m-i)c^{2(i-1)} - 2\sigma^2\sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1} \\ 4\sigma^4 r^2\sum_{i=2}^{m-1}(i-1)(m-i)c^{2(i-1)-1} - 2\sigma^2 r\sum_{i=2}^{m-1}(i-1)\mathrm{tr}(AT_i)c^{i-2} \\ 4\sigma^3(m+2r^2\sum_{i=1}^{m-1}(m-i)c^{2(i-1)}) - 4\sigma(\mathrm{tr}(A)+r\sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1}) \end{bmatrix},$$

so that the stationary points (r,c,σ) must satisfy

$$\begin{cases} 2\sigma^2 r\sum_{i=1}^{m-1}(m-i)c^{2(i-1)} - \sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1} = 0, \\ 2\sigma^2 r^2\sum_{i=2}^{m-1}(i-1)(m-i)c^{2(i-1)-1} - r\sum_{i=2}^{m-1}(i-1)\mathrm{tr}(AT_i)c^{i-2} = 0, \\ \sigma^2(m+2r^2\sum_{i=1}^{m-1}(m-i)c^{2(i-1)}) - \mathrm{tr}(A) - r\sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1} = 0. \end{cases} \tag{3.17}$$

By rearranging equations in (3.17), we have

$$\begin{cases} \sigma^2 = \dfrac{\mathrm{tr}(A)}{m}, \\ 2r\mathrm{tr}(A)\displaystyle\sum_{i=1}^{m-1}(m-i)c^{2(i-1)} - m\sum_{i=1}^{m-1}\mathrm{tr}(AT_i)c^{i-1} = 0, \\ 2r^2\mathrm{tr}(A)\displaystyle\sum_{i=2}^{m-1}(i-1)(m-i)c^{2(i-1)-1} - mr\sum_{i=2}^{m-1}(i-1)\mathrm{tr}(AT_i)c^{i-2} = 0. \end{cases} \tag{3.18}$$

Numerical experiments show that there exists at least one root of equations (3.18) which ensures $B \in R_+^{m\times m}$. Equivalently, the local minima of $f(r, c, \sigma)$ are achieved at the points (r, c, σ) satisfying (3.18).

We summarize the discussion above in the following theorem.

Theorem 3.4 *Given a covariance matrix $A \in R^{m\times m}$, define $f(r, c, \sigma) := L(A, B(r, c, \sigma))$ where $B(c, \sigma)$ is a positive definite covariance matrix of the ARMA(1,1) model as in (3.16). Then the local minima of $f(r, c, \sigma)$ are achieved at the points (r, c, σ) satisfying (3.18).*

4 Simulation studies

To examine our method, in this section we carry out simulation studies. All computations are performed with MATLAB R2008b. The root-finding problem in section 3 is solved with MATLAB `fzero or fsolve`.

4.1 Assessment for Gaussian data

Let m be the dimension of the covariance matrices to be tested. First, we generate an $m \times n$ data matrix Q with columns randomly drawn from the multivariate normal distribution $N(\mu, \Sigma)$, where $\mu = \mathbf{0} \in R^m$ is a mean vector and Σ is a covariance matrix that have the structures as discussed in section 1. Second, we compute the sample covariance matrix A with the generated data Q. Finally, we find for each structure a covariance matrix that minimizes the discrepancy function in (1.3). We test with the true covariance matrix Σ, where for each structure we consider several different values for m, c, r and σ^2. We choose sample size $n = 1000$, $m \in \{100, 200\}$, $c \in \{0.25, 0.5, 0.75\}$ and $\sigma^2 \in \{0.5, 1, 2, 4\}$ for Σ having MA(1), CS, AR(1) structures. For Σ having an ARMA(1,1) structure we use the above n, m, c and σ^2 but consider different choices of r, including $\{0.1, 0.35, 0.6\}$, $\{0.2, 0.45, 0.75\}$ and $\{0.25, 0.5, 0.8\}$. We summarize the experimental results in Tables 1-3 for $m = 100$ and in Tables 4-6 for $m = 200$.

Tables 1-6 are about here

In Tables 1-6 each row stands for one experiment and for each experiment we report the results averaged over 1000 repeated simulations. The first column gives the true underlying covariance structure and the second column presents the discrepancy between the true covariance matrix Σ and the sample covariance matrix A under the F-norm measure of discrepancy function. The rest of the columns report the results from the computed matrix B with different structures. Note that we do not include a row for Σ having MA(1) with $c = 0.75$ because there does not exist such a positive definite covariance matrix in this case. The notation in Tables 1-6 is summarized as follows:

- Σ: the true covariance matrix;
- A: the sample covariance matrix;
- B: the computed covariance matrix that has a certain structure and minimizes the discrepancy function $L(A, B)$ in (1.3).
- $L_{\Sigma,A}$, $L_{A,B}$ and $L_{\Sigma,B}$: the discrepancy function $L(\Sigma, A)$, $L(A, B)$ and $L(\Sigma, B)$, respectively.

In Tables 1-6, we have the following observations.

(1) The matrix B having the minimum $L_{\Sigma,B}$ has the same structure as the true covariance matrix Σ and $L_{\Sigma,B} < L_{\Sigma,A}$. In other words, the regularized estimator B is much better than the sample covariance matrix A in terms of the F-norm discrepancy function. This shows that regularization of the sample covariance matrix A, is necessary not only for the convenient use of known structure but also for the accuracy of covariance estimation.

(2) For Σ having one of the structures of MA(1), CS or AR(1), among different minimizers B, there are two structures clearly winning out in the sense of having smaller $L_{A,B}$: the one having the same structure as Σ and the ARMA(1,1), the latter always being the best. It is not surprising for the matrix B with ARMA(1,1) structure to win out because all MA(1), CS, and AR(1) are indeed special ARMA(1,1) structures. There is no doubt that minimizing among the larger feasible set will give the smaller minimum.

    Note that it is extremely important to observe the discrepancy $L_{A,B}$, because in practice the true covariance Σ is usually unknown and so is $L_{\Sigma,B}$. Thus, the discrepancy $L_{A,B}$ can be used to identify the covariance structure.

(3) The observations above are common to all choices of the structure of Σ in the class we have considered, the various values of m, c, σ^2 and r. Therefore, the findings are reliable in this sense.

4.2 Assessment for high-dimensional data

In the above simulation studies, the sample covariance matrix with sample size $n = 1000$ and dimension $m = 100, 200$ is used to be the available matrix A, on which its covariance structure needs to be identified. The sample covariance matrix A considered above is nonsingular because the sample size n is much bigger than the dimension m. In some practical scenarios, the given matrix A may happen to be singular and it is natural to wonder if the proposed approach still works in this case. We therefore run a further simulation study for the case of Σ having a CS structure with $c = 0.5$ and $\sigma^2 = 1$. This time we draw random samples with sample size $n = 500$ from m-dimensional normal distribution $N(0, \Sigma)$ with $m = 1000$. The sample covariance matrix A becomes singular due to $n << m$. This experiment was repeated 1000 times. The F-norm discrepancy results averaged over the 1000 simulations are summarized in Table 7 and the parameter estimates in the simulations are presented in box-plot in Figure 1.

Tables 7 is about here

From Table 7, it is clear that the CS and ARMA(1,1) structures stand out, implying that the two structures are the likely structure of the covariance matrix Σ. Note that the ARMA(1,1) includes CS as a special case, the two structures identified are actually almost identical, see Figure 1. In Figure 1, (c_t, s_t), (c_c, s_c) and (c_A, s_A) represent the estimates of parameters (c, σ^2) in the cases of MA(1), CS and AR(1) structures, respectively, and (r_M, c_M, s_M) is the estimate of (r, c, σ^2) for ARMA(1,1) model. Thus the CS structure is correctly identified as the structure of the covariance matrix A and then Σ. An interesting finding from Figure 1 is that, although the ARMA(1,1) model has the almost same parameter estimates as the CS structure ($c = 0.5$ and $\sigma^2 = 1$), its estimate of variance σ^2 has more variability than the one with CS structure. On the other hand, when the AR(1) is misused, the resulting estimates of parameters can be very biased.

Note here that the sample covariance matrix A is singular. This is, the proposed regularization approach works well even if the given matrix A is not nonsingular. In this case, Lin et al.'s [7] method cannot be applied as it involves the use of the inverse of the sample covariance matrix A.

4.3 Comparison with the MLE method

Although the above simulated data are generated from Gaussian distribution, we stress that the proposed method does not require a distribution assumption. A reviewer raised the issue of making comparisons with existing standard parametric methods such as maximum likelihood estimation (MLE), method of moment/Yule-Walker, least square regression method, etc. In order to save space, below we only compare the proposed approach with the MLE method by intensive simulations.

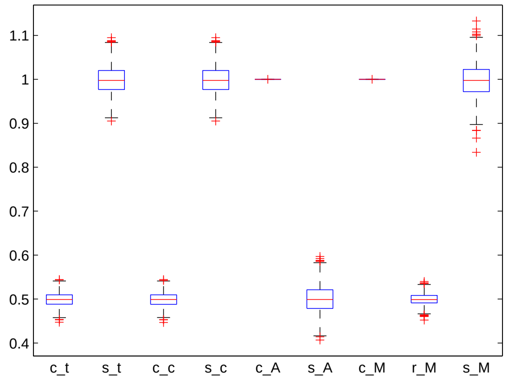

Figure 1: Box plot of parameter estimates

The comparisons are made under two assumptions, i.e., the data are Gaussian and Non-Gaussian distributions. First, for the Gaussian distribution like the above we generate an $m \times n$ data matrix with each column coming from $N(0, \Sigma)$ where Σ is of the ARMA(1,1) structure with true parameters $\sigma^2 = 1$, $c = 0.5$ and $r = 0.20, 0.45, 0.75$, respectively. Our proposed approach and the standard MLE method are used to estimate the parameters and the results for 1000 runs are summarized in Table 8, from which it is observed that the proposed approach performs almost equally well as the standard MLE method.

Table 8 is about here

Second, we carry out the similar simulations but for non-Gaussian data this time. Let Q_1 is an $m \times n$ data matrix with each column being m independent samples from χ_1^2, i.e., the chi-square distribution with one degree of freedom. Assume C is an $m \times m$ matrix of being the ARMA(1,1) structure with the same true parameters σ^2, c and r as above. Let $Q = C^{1/2} Q_1$ then each column of Q forms a multivariate sample that is not Gaussian. Obviously, $\Sigma \equiv \mathrm{Var}(q_i) = 2C$ with q_i is the ith column of Q $(i = 1, 2, ..., n)$. In other words, the covariance matrix Σ is of the ARMA(1,1) structure with $\sigma^2 = 2$, $c = 0.5$ and $r = 0.20, 0.45, 0.75$. Similarly, we compare the proposed approach to the MLE method over 1000 simulation runs and report the results in Table 9. It shows that the proposed approach is able to produce very accurate estimates for the parameters in Σ even if the data are not Gaussian. The MLE method, however, can lead to very biased estimates for the parameters in Σ when data are not Gaussian.

Table 9 is about here

4.4 Assessment for non-Gaussian data

To further investigate the performance of the proposed approach for non-Gaussian data, we conduct another simulation study, in which the simulation setup is the same as above except that this time each column of Q_1 are random samples from a Bernoulli's distribution $B(p)$ with the probability $p = 0.1, 0.3, 0.5$. Note this time $\Sigma \equiv \mathrm{Var}(q_i) = p(1-p)C$. Based on the data matrix Q we form the sample matrix A and calculate various F-norm discrepancy values, reported in Table 10. From Table 10, it is clear that even if the data are not Gaussian and actually generated by a linear transformation of Bernoulli distributions, the proposed approach still performs very well and is able to find the true structure of covariance matrix, just like what it does for Gaussian data. A slight difference in format reported in Table 10 is that we have now reported the adjusted F-norm discrepancy

$$L^*(A, B) = \mathrm{tr}\{(A-B)^T(A-B)\}/\mathrm{tr}(A^T A). \tag{4.19}$$

This is because the original F-norm discrepancy defined in (1.3) is somehow in the sense of absolute error and may result in a very large value as seen in Tables 1-7.

Table 10 is about here

5 Real data analysis

In the real data analysis, we consider the regularization of covariance matrices for the synthetic control chart time series data. This data set contains 600 examples of control charts synthetically generated by the process in [1]. The control charts were assigned to six different classes: Normal, Cyclic, Increasing trend, Decreasing trend, Upward shift and Downward shift. The data set is presented in an 600×60 matrix, with a single chart per row, and the classes are organized as follows: 1-100 are the Normal class, 101-200 are the Cyclic class, 201-300 are the Increasing trend class, 301-400 are the Decreasing trend class, 401-500 are the Upward shift class, and 501-600 are the Downward shift class.

These classes of data sets as well as their pooled data were tested using three test methods, IPS, Fisher-ADF and Fisher-PP tests, for their stationarity. It is concluded that apart from the Cyclic class, other five classes as well as the pooled data of those the five classes are all stationary, after taking the first order difference so as to remove the intercept and the time trend effects. Our analysis below is then for the newly transformed data by using the first order difference. The regularization of the covariance matrices for the new data of the five classes, as well as their pooled data, is now made using the adjusted F-norm in (4.19). The numerical results are reported in Table 11, where the column "Time" gives the time (in second) spent for finding the optimal matrix B with each candidate structure.

Table 11 is about here

Note that the true covariance matrix Σ is unknown for any real data, so that $L^*_{\Sigma,A}$ and $L^*_{\Sigma,B}$ are actually not available, where the given matrix A is chosen to be the sample covariance matrix. We then use the adjusted F-norm discrepancy $L^*_{A,B}$ in (4.19) to identify the most likely covariance structure among the possible candidate structures: MA(1), CS, AR(1) and ARMA(1,1).

From Table 11, it is clear that for the transformed data using the first order difference we have reasons to believe the five classes of the new data together with their pooled data are all of MA(1) structure. Note that the ARMA(1,1) seems to have a slightly smaller F-norm discrepancy value than the MA(1), but the difference is so small that it can be ignorable. Since the MA(1) is a special case of the ARMA(1,1), it is believed that the ARMA(1,1) almost reduces to the MA(1) in this case. Therefore, the MA(1) is preferred for the new data as it is more parsimonious than the ARMA(1,1).

6 Discussion

Given a matrix A and a class of candidate covariance structures, a new method was proposed to regularize available covariance matrix A so that its underlying structure becomes clear. In other words, random noise can be filtered in this sense. Our simulation studies demonstrate the reliability of the proposed method, which filters not only random noise in A but also reveal characteristics of the stochastic process structuring the covariance matrix.

In the simulation studies and real data analysis, the available matrix A considered is taken as the sample covariance matrix. In practice, it does not have to be the sample covariance matrix. In theory, the matrix A can be any available estimate of the covariance matrix, obtained by various statistical methods. As long as A is provided, our proposed method can be used to regularize the covariance matrix A even if the distribution of the data is unknown, the dimension of matrix A is high, or the matrix A is singular. In particular, our simulations show that by using the sample covariance matrix the proposed method works very well in identifying the true structure of the population covariance matrix Σ even for the high-dimensional case, i.e., $m >> n$. In this case, the established approaches such as the maximum likelihood estimation and moment estimation may not work properly, because the inverse of the sample covariance matrix is usually involved in such methods.

We also show that, for Gaussian data with $n > m$, the proposed approach performs almost the same as the standard MLE method in estimation of the parameters in covariance matrices. For non-Gaussian data, the proposed approach still performs very well in estimation of the parameters. In contrast, the standard MLE method that wrongly assumes normality for non-Gaussian data results in very biased estimates of the parameters in covariance matrices. In other words, our proposed method does not require any distribution assumption for the data. As long as a reasonable covariance matrix estimate A is given, the underlying structures of the population covariance matrix Σ can be captured by regularizing the estimate of A. In this sense, the proposed approach is robust against the distribution of the data.

In addition to the four likely candidate structures considered in this paper, there are a lot of other useful covariance structures in practice, such as AR(2), AR(3), factor analytic structure, general linear structure, ARMA(p, q), banded or Toeplitz structures, etc. In theory our proposed approach is applicable to any other likely structures of covariance matrix, the corresponding optimization problem and computation may become difficult, especially when the dimension of matrix A is very high and sparse. We will investigate this problem in our future work.

A referee also raised the issue of measuring the accuracy of the parameter estimates or their confidence intervals for the covariance structures we considered. Although it may be challenging, it is interesting to study the convergence rate and the asymptotic distribution of the parameter estimates based on the sample covariance matrix by assuming certain distributional conditions of the data. Alternatively, we may use Bootstrapping resampling technique to construct the confidence intervals of the parameter estimates. However, the focus here is on regularizing the covariance matrix estimate A, aiming to find the underlying structure of the population covariance matrix Σ that is usually unknown. This issue definitely deserves a further investigation.

It is worth mentioning that there are other regularization methods in the literature such as banding, tapering, thresholding (e.g., Bickel and Levina [2], 2008; Cai and Liu [4], 2011; Pourahmadi [16], 2013), and POET (Fan, et al. [6], 2016) among others. The proposed method in this paper has a clear distinction from these methods as the aim here is to find the underlying structure of the covariance matrix from a class of candidates. An interesting issue is to compare with at least some of such literature work. We will also explore such interesting topics in our future work.

Acknowledgements

We gratefully acknowledge very helpful and constructive comments and suggestions made by three anonymous referees, which leads to significant improvements to the paper.

References

[1] R. J. Alcock and Y. Manolopoulos. Time-Series Similarity Queries Employing a Feature-Based Approach. 7th Hellenic Conference on Informatics. August 27-29. Ioannina, Greece, 1999.

[2] P. J. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, 36, 2577C2604, 2008

[3] R. Borsdorf, N. J. Higham, and M. Raydan. Computing a nearest correlation matrix with factor structure. *SIAM J. Matrix Anal. Appl.*, 31(5): 2603-2622, 2010.

[4] T. Cai and W. Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494): 672C684, 2011.

[5] J. Fan, Y. Liao and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of Royal Statistical Society, Series B*, 75(4), 656-658, 2013.

[6] J. Fan, Y. Liao, and H. Liu. An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, 19(1): C1CC32, 2016.

[7] N. J. Higham. *Accuracy and Stability of Numerical Algorithms. Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA, second edition, 2002. ISBN 0-89871-521-0.

[8] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, UK, second edition, 2013. ISBN 978-0-521-83940-2.

[9] F. Lin and M. R. Jovanović. Least-squares approximation of structured covariances. *IEEE Trans. Automat. Control*, 54(7): 1643-1648, 2009.

[10] L. Lin, N. J. Higham and J. Pan. Covariance structure regularization via entropy loss function. *Computational Statistics & Data Analysis*, 72, 315327, 2014.

[11] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley, Chichester, UK, revised edition, 1999. ISBN 0-471-98633-X.

[12] J. Pan and G. Mackenzie. On modelling mean-covariance structures in longitudinal studies. *Biometrika*, 90(1): 239-244, 2003.

[13] S. V. Parter. An observation on the numerical solution of difference equations and a theorem of Szegö. Numerische Mathematik, 4: 293-295, 1962.

[14] F. Pascal, Y. Chitour, J. P. Ovarlez, P. Forster, and P. Larzabal. Covariance structure maximum-likelihood estimates in compound Gaussian noise: existence and algorithm analysis. *IEEE Trans. Signal Processing*, 56(1): 34-48, 2008.

[15] M. Pourahmadi. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3): 677-690, 1999. ISSN 0006-3444.

[16] M. Pourahmadi. *High-Dimensional Covariance Estimation: With High-Dimensional Data*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2013.

[17] V. Vinciotti and H. Hashem. Robust methods for inferring sparse network structures. *Computational Statistics & Data Analysis*, 67, 84-94, 2013.

[18] H. Ye and J. Pan. Modelling of covariance structures in generalised estimating equations for longitudinal data. *Biometrika*, 93(4): 927-941, 2006. ISSN 0006-3444.

Table 1: Simulation results with $m = 100$; $c = 0.25$.

$\sigma^2 = 0.50$		B							
		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	2.5327	2.5315	0.0012	5.5630	3.0337	2.6943	0.1638	2.5309	0.0018
CS	2.6833	154.1047	151.6041	2.3516	0.3317	15.9702	13.9584	2.3496	0.3337
AR(1)	2.5308	2.7341	0.2052	5.7147	3.1905	2.5294	0.0015	2.5288	0.0020
ARMA-r=0.1	2.5253	2.5571	0.0337	3.0325	0.5114	2.5363	0.0131	2.5237	0.0016
ARMA-r=0.35	2.5340	2.9336	0.4012	8.7838	6.2519	2.5848	0.0548	2.5316	0.0023
ARMA-r=0.6	2.5467	3.7187	1.1772	20.9075	18.3702	3.8263	1.2868	2.5430	0.0037

$\sigma^2 = 1$		B							
		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	10.1308	10.1259	0.0049	22.2522	12.1347	10.7770	0.6550	10.1235	0.0072
CS	10.7333	616.4186	606.4165	9.4064	1.3269	63.8807	55.8334	9.3983	1.3350
AR(1)	10.1233	10.9363	0.8209	22.8589	12.7620	10.1174	0.0059	10.1153	0.0080
ARMA-r=0.1	10.1012	10.2285	0.1349	12.1299	2.0457	10.1454	0.0525	10.0946	0.0065
ARMA-r=0.35	10.1359	11.7343	1.6048	35.1352	25.0077	10.3391	0.2190	10.1265	0.0094
ARMA-r=0.6	10.1869	14.8749	4.7090	83.6298	73.4807	15.3052	5.1472	10.1719	0.0150

$\sigma^2 = 2$		B							
		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	40.5230	40.5035	0.0195	89.0088	48.5388	43.1081	2.6201	40.4940	0.0290
CS	42.9331	2465.6745	2425.6661	37.6255	5.3076	255.5229	223.3337	37.5931	5.3399
AR(1)	40.4932	43.7452	3.2836	91.4355	51.0478	40.4697	0.0235	40.4613	0.0319
ARMA-r=0.1	40.4047	40.9139	0.5394	48.5194	8.1828	40.5814	0.2098	40.3786	0.0262
ARMA-r=0.35	40.5435	46.9372	6.4191	140.5408	100.0307	41.3564	0.8761	40.5060	0.0375
ARMA-r=0.6	40.7475	59.4996	18.8360	334.5192	293.9228	61.2207	20.5887	40.6876	0.0600

$\sigma^2 = 4$		B							
		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	162.0920	162.0140	0.0780	356.0350	194.1550	172.4323	10.4803	161.9761	0.1159
CS	171.7322	9862.6978	9702.6643	150.5020	21.2303	1022.0915	893.3348	150.3725	21.3597
AR(1)	161.9729	174.9810	13.1344	365.7421	204.1913	161.8790	0.0940	161.8454	0.1276
ARMA-r=0.1	161.6188	163.6558	2.1576	194.0777	32.7313	162.3257	0.8392	161.5142	0.1048
ARMA-r=0.35	162.1740	187.7489	25.6766	562.1630	400.1229	165.4256	3.5043	162.0241	0.1501
ARMA-r=0.6	162.9901	237.9984	75.3440	1338.0768	1175.6912	244.8828	82.3548	162.7502	0.2398

Table 2: Simulation results with $m = 100$; $c = 0.5$.

		B							
$\sigma^2 = 0.50$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	2.5443	2.5426	0.0017	14.6677	12.1304	4.3834	1.8420	2.5418	0.0025
CS	3.1941	608.7943	606.4146	1.8749	1.3192	7.9307	7.3818	1.8724	1.3217
AR(1)	2.5452	6.6103	4.0714	18.0074	15.4796	2.5413	0.0040	2.5405	0.0047
ARMA-r=0.2	2.5275	3.1773	0.6524	4.9969	2.4780	2.7801	0.2556	2.5251	0.0025
ARMA-r=0.45	2.5410	5.8378	3.2980	15.0778	12.5387	2.5632	0.0294	2.5367	0.0044
ARMA-r=0.75	2.5657	11.7168	9.1595	37.3788	34.8265	3.5500	1.0001	2.5574	0.0083
		B							
$\sigma^2 = 1$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	10.1773	10.1704	0.0069	58.6707	48.5216	17.5336	7.3679	10.1673	0.0100
CS	12.7764	2435.1774	2425.6584	7.4996	5.2768	31.7229	29.5270	7.4896	5.2867
AR(1)	10.1809	26.4412	16.2858	72.0295	61.9184	10.1651	0.0158	10.1621	0.0188
ARMA-r=0.2	10.1101	12.7090	2.6095	19.9875	9.9118	11.1202	1.0225	10.1003	0.0098
ARMA-r=0.45	10.1641	23.3513	13.1920	60.3113	50.1550	10.2527	0.1174	10.1467	0.0174
ARMA-r=0.75	10.2628	46.8672	36.6379	149.5151	139.3059	14.2002	4.0003	10.2296	0.0332
		B							
$\sigma^2 = 2$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	40.7091	40.6816	0.0275	234.6828	194.0864	70.1342	29.4718	40.6691	0.0400
CS	51.1055	9740.7095	9702.6337	29.9984	21.1070	126.8915	118.1081	29.9586	21.1469
AR(1)	40.7236	105.7649	65.1431	288.1179	247.6735	40.6603	0.0634	40.6483	0.0753
ARMA-r=0.2	40.4406	50.8361	10.4379	79.9501	39.6472	44.4809	4.0900	40.4013	0.0393
ARMA-r=0.45	40.6563	93.4052	52.7681	241.2454	200.6198	41.0107	0.4697	40.5867	0.0697
ARMA-r=0.75	41.0512	187.4690	146.5514	598.0605	557.2238	56.8007	16.0012	40.9186	0.1327
		B							
$\sigma^2 = 4$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	162.8363	162.7264	0.1100	938.7311	776.3455	280.5369	117.8872	162.6765	0.1599
CS	204.4220	38962.8381	38810.5348	119.9938	84.4282	507.5659	472.4326	119.8344	84.5876
AR(1)	162.8946	423.0596	260.5722	1152.4717	990.6942	162.6412	0.2534	162.5933	0.3012
ARMA-r=0.2	161.7623	203.3442	41.7517	319.8004	158.5889	177.9237	16.3600	161.6052	0.1570
ARMA-r=0.45	162.6254	373.6209	211.0724	964.9816	802.4793	164.0429	1.8788	162.3468	0.2786
ARMA-r=0.75	164.2048	749.8759	586.2058	2392.2420	2228.8952	227.2026	64.0048	163.6742	0.5306

Table 3: Simulation results with $m = 100$; $c = 0.75$.

		B							
$\sigma^2 = 0.50$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	3.9150	1365.2677	1364.4280	1.0904	2.8246	2.6037	4.3421	1.0883	2.8267
AR(1)	2.5794	37.5059	34.9775	56.9499	54.4603	2.5581	0.0214	2.5568	0.0226
ARMA-r=0.25	2.5353	6.4169	3.8873	8.5743	6.0536	10.6511	8.1262	2.5303	0.0051
ARMA-r=0.5	2.5534	18.0680	15.5462	26.6927	24.2067	3.6861	1.1573	2.5411	0.0123
ARMA-r=0.8	2.5991	42.3503	39.7962	64.5068	61.9660	2.6285	0.0783	2.5727	0.0263
		B							
$\sigma^2 = 1$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	15.6600	5461.0707	5457.7122	4.3617	11.2982	10.4146	17.3684	4.3531	11.3069
AR(1)	10.3177	150.0234	139.9098	227.7995	217.8411	10.2323	0.0855	10.2274	0.0904
ARMA-r=0.25	10.1412	25.6677	15.5493	34.2971	24.2145	42.6044	32.5047	10.1210	0.0202
ARMA-r=0.5	10.2136	72.2720	62.1848	106.7706	96.8267	14.7445	4.6293	10.1645	0.0491
ARMA-r=0.8	10.3964	169.4013	159.1847	258.0274	247.8641	10.5138	0.3134	10.2910	0.1053
		B							
$\sigma^2 = 2$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	62.6399	21844.2828	21830.8487	17.4469	45.1930	41.6585	69.4735	17.4122	45.2277
AR(1)	41.2710	600.0936	559.6393	911.1980	871.3646	40.9290	0.3421	40.9096	0.3616
ARMA-r=0.25	40.5649	102.6708	62.1972	137.1884	96.8581	170.4176	130.0189	40.4841	0.0808
ARMA-r=0.5	40.8546	289.0882	248.7394	427.0825	387.3069	58.9779	18.5174	40.6580	0.1966
ARMA-r=0.8	41.5857	677.6053	636.7388	1032.1095	991.4565	42.0552	1.2536	41.1639	0.4213
		B							
$\sigma^2 = 4$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	250.5596	87377.1312	87323.3948	69.7876	180.7720	166.6339	277.8940	69.6488	180.9108
AR(1)	165.0840	2400.3746	2238.5573	3644.7921	3485.4583	163.7160	1.3683	163.6382	1.4462
ARMA-r=0.25	162.2595	410.6833	248.7886	548.7534	387.4326	681.6706	520.0755	161.9362	0.3233
ARMA-r=0.5	163.4182	1156.3527	994.9574	1708.3300	1549.2275	235.9115	74.0694	162.6318	0.7862
ARMA-r=0.8	166.3429	2710.4212	2546.9553	4128.4380	3965.8260	168.2210	5.0143	164.6555	1.6853

Table 4: Simulation results with $m = 200$; $c = 0.25$.

		B							
$\sigma^2 = 0.50$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	10.0649	10.0637	0.0012	16.2196	6.1583	10.3926	0.3291	10.0632	0.0018
CS	10.6463	625.0203	615.6754	9.3896	1.2567	37.0791	28.9558	9.3855	1.2608
AR(1)	10.0648	10.4757	0.4135	16.5833	6.5228	10.0633	0.0014	10.0628	0.0020
ARMA-r=0.1	10.0623	10.1270	0.0670	11.1042	1.0447	10.0853	0.0253	10.0606	0.0017
ARMA-r=0.35	10.0679	10.8734	0.8096	22.8366	12.7836	10.1725	0.1088	10.0654	0.0024
ARMA-r=0.6	10.0936	12.4643	2.3773	47.6369	37.5648	12.6792	2.5914	10.0897	0.0039
		B							
$\sigma^2 = 1$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	40.2597	40.2550	0.0048	64.8785	24.6331	41.5704	1.3165	40.2527	0.0070
CS	42.5852	2500.0814	2462.7015	37.5582	5.0270	148.3163	115.8232	37.5421	5.0431
AR(1)	40.2590	41.9030	1.6540	66.3333	26.0913	40.2534	0.0056	40.2511	0.0079
ARMA-r=0.1	40.2493	40.5082	0.2680	44.4167	4.1786	40.3412	0.1013	40.2426	0.0067
ARMA-r=0.35	40.2715	43.4934	3.2385	91.3463	51.1345	40.6900	0.4351	40.2617	0.0098
ARMA-r=0.6	40.3743	49.8570	9.5093	190.5475	150.2592	50.7169	10.3654	40.3589	0.0155
		B							
$\sigma^2 = 2$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	161.0390	161.0198	0.0191	259.5139	98.5325	166.2815	5.2659	161.0109	0.0280
CS	170.3408	10000.3254	9850.8060	150.2330	20.1079	593.2651	463.2930	150.1686	20.1723
AR(1)	161.0361	167.6118	6.6162	265.3330	104.3650	161.0136	0.0225	161.0044	0.0317
ARMA-r=0.1	160.9971	162.0326	1.0721	177.6666	16.7144	161.3649	0.4053	160.9703	0.0269
ARMA-r=0.35	161.0859	173.9737	12.9539	365.3852	204.5380	162.7599	1.7404	161.0469	0.0391
ARMA-r=0.6	161.4973	199.4281	38.0372	762.1900	601.0367	202.8675	41.4616	161.4354	0.0619
		B							
$\sigma^2 = 4$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	644.1559	644.0793	0.0766	1038.0555	394.1299	665.1262	21.0635	644.0438	0.1122
CS	681.3634	40001.3017	39403.2242	600.9319	80.4314	2373.0605	1853.1720	600.6743	80.6890
AR(1)	644.1444	670.4473	26.4646	1061.3320	417.4602	644.0542	0.0902	644.0175	0.1268
ARMA-r=0.1	643.9886	648.1305	4.2885	710.6666	66.8577	645.4595	1.6213	643.8811	0.1075
ARMA-r=0.35	644.3437	695.8950	51.8157	1461.5409	818.1520	651.0397	6.9615	644.1877	0.1562
ARMA-r=0.6	645.9893	797.7125	152.1489	3048.7600	2404.1466	811.4701	165.8464	645.7417	0.2476

Table 5: Simulation results with $m = 200$; $c = 0.5$.

		B							
$\sigma^2 = 0.50$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	10.0876	10.0859	0.0017	34.7046	24.6290	13.7978	3.7121	10.0852	0.0024
CS	12.4924	2471.9393	2462.6999	7.4777	5.0147	19.7831	17.3309	7.4727	5.0197
AR(1)	10.0917	18.3236	8.2380	42.2092	32.1312	10.0877	0.0040	10.0869	0.0048
ARMA-r=0.2	10.0669	11.3824	1.3190	15.2032	5.1423	10.5789	0.5160	10.0643	0.0025
ARMA-r=0.45	10.0812	16.7480	6.6731	36.0862	26.0271	10.1301	0.0558	10.0767	0.0045
ARMA-r=0.75	10.1277	28.6442	18.5346	82.3645	72.2929	12.1208	2.0087	10.1190	0.0086
		B							
$\sigma^2 = 1$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	40.3505	40.3436	0.0068	138.8182	98.5162	55.1911	14.8485	40.3408	0.0097
CS	49.9696	9887.7572	9850.7994	29.9107	20.0589	79.1323	69.3236	29.8907	20.0789
AR(1)	40.3669	73.2942	32.9521	168.8369	128.5248	40.3506	0.0162	40.3476	0.0193
ARMA-r=0.2	40.2674	45.5297	5.2761	60.8128	20.5693	42.3155	2.0638	40.2573	0.0102
ARMA-r=0.45	40.3248	66.9918	26.6926	144.3447	104.1083	40.5203	0.2231	40.3069	0.0179
ARMA-r=0.75	40.5106	114.5767	74.1384	329.4581	289.1715	48.4832	8.0346	40.4761	0.0345
		B							
$\sigma^2 = 2$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	161.4019	161.3746	0.0273	555.2729	394.0647	220.7645	59.3939	161.3631	0.0389
CS	199.8784	39551.0287	39403.1976	119.6429	80.2355	316.5294	277.2944	119.5628	80.3156
AR(1)	161.4674	293.1770	131.8082	675.3478	514.0994	161.4026	0.0648	161.3903	0.0770
ARMA-r=0.2	161.0696	182.1187	21.1042	243.2512	82.2772	169.2620	8.2554	161.0291	0.0406
ARMA-r=0.45	161.2993	267.9674	106.7704	577.3788	416.4330	162.0813	0.8923	161.2277	0.0716
ARMA-r=0.75	162.0426	458.3068	296.5535	1317.8325	1156.6861	193.9327	32.1385	161.9044	0.1382
		B							
$\sigma^2 = 4$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
MA(1)	645.6077	645.4983	0.1094	2221.0916	1576.2587	883.0582	237.5758	645.4523	0.1554
CS	799.5137	158204.1146	157612.7905	478.5715	320.9421	1266.1176	1109.1776	478.2513	321.2624
AR(1)	645.8696	1172.7078	527.2330	2701.3912	2056.3975	645.6104	0.2592	645.5613	0.3082
ARMA-r=0.2	644.2785	728.4747	84.4170	973.0047	329.1090	677.0481	33.0215	644.1162	0.1625
ARMA-r=0.45	645.1971	1071.8695	427.0815	2309.5151	1665.7320	648.3251	3.5693	644.9110	0.2862
ARMA-r=0.75	648.1702	1833.2270	1186.2140	5271.3299	4626.7443	775.7309	128.5542	647.6174	0.5528

Table 6: Simulation results with $m = 200$; $c = 0.75$.

		B							
$\sigma^2 = 0.50$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	15.6162	5554.8375	5541.0741	4.3933	11.2229	7.4676	14.3044	4.3897	11.2265
AR(1)	10.1738	81.2141	71.1382	128.4485	118.4413	10.1514	0.0224	10.1501	0.0236
ARMA-r=0.25	10.0724	17.9662	7.9052	23.2171	13.1621	42.9548	32.9090	10.0673	0.0052
ARMA-r=0.5	10.1182	41.7266	31.6179	62.7155	52.6420	12.4233	2.3290	10.1046	0.0136
ARMA-r=0.8	10.1966	91.1004	80.9393	144.8784	134.7604	10.2791	0.1328	10.1699	0.0266
		B							
$\sigma^2 = 1$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	62.4648	22219.3499	22164.2965	17.5733	44.8914	29.8703	57.2177	17.5589	44.9058
AR(1)	40.6950	324.8565	284.5529	513.7938	473.7654	40.6056	0.0894	40.6005	0.0945
ARMA-r=0.25	40.2896	71.8649	31.6206	92.8685	52.6485	171.8192	131.6359	40.2690	0.0206
ARMA-r=0.5	40.4727	166.9065	126.4714	250.8619	210.5680	49.6933	9.3161	40.4183	0.0545
ARMA-r=0.8	40.7863	364.4014	323.7570	579.5136	539.0417	41.1164	0.5312	40.6798	0.1065
		B							
$\sigma^2 = 2$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	249.8591	88877.3996	88657.1860	70.2933	179.5658	119.4813	228.8706	70.2358	179.6233
AR(1)	162.7801	1299.4259	1138.2116	2055.1753	1895.0615	162.4223	0.3577	162.4018	0.3781
ARMA-r=0.25	161.1585	287.4597	126.4825	371.4739	210.5941	687.2768	526.5434	161.0761	0.0824
ARMA-r=0.5	161.8909	667.6259	505.8857	1003.4478	842.2721	198.7730	37.2644	161.6732	0.2179
ARMA-r=0.8	163.1453	1457.6058	1295.0282	2318.0545	2156.1670	164.4658	2.1248	162.7191	0.4259
		B							
$\sigma^2 = 4$		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	999.4362	355509.5984	354628.7442	281.1731	718.2632	477.9250	915.4825	280.9431	718.4931
AR(1)	651.1205	5197.7035	4552.8464	8220.7013	7580.2462	649.6892	1.4307	649.6073	1.5125
ARMA-r=0.25	644.6341	1149.8388	505.9300	1485.8957	842.3765	2749.1071	2106.1737	644.3042	0.3297
ARMA-r=0.5	647.5636	2670.5037	2023.5428	4013.7910	3369.0885	795.0921	149.0575	646.6929	0.8714
ARMA-r=0.8	652.5812	5830.4232	5180.1127	9272.2180	8624.6679	657.8630	8.4994	650.8763	1.7036

Table 7: Simulation results for CS with $(c, \sigma^2) = (0.5, 1)$ and $(m, n) = (1000, 500)$

		B							
		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$	$L_{A,B}$	$L_{\Sigma,B}$
CS	2507.9700	251256.3608	249253.5243	1500.7138	1007.2562	1749.7276	1256.7082	1500.5837	1007.6397

Table 8: Comparison for Gaussian data with $(m, n) = (100, 1000)$

$\sigma^2 = 1$		σ^2_mean	σ^2_std	c_mean	c_std	r_mean	r_std
r=0.2	Our approach	1.00015	0.00076	0.49967	0.01152	0.20019	0.00043
	MLE	1.01999	0.00080	0.49972	0.00849	0.20092	0.00039
r=0.45	Our approach	1.00330	0.00126	0.49856	0.00269	0.45074	0.00047
	MLE	1.02386	0.00135	0.49723	0.00204	0.45133	0.00041
r=0.75	Our approach	0.99761	0.00179	0.50528	0.00201	0.74577	0.00019
	MLE	1.03601	0.01091	0.50636	0.00088	0.77189	0.00032

Table 9: Comparison for non-Gaussian data with $(m, n) = (100, 1000)$

degree of freedom=1		σ^2_mean	σ^2_std	c_mean	c_std	r_mean	r_std
r=0.2	Our approach	2.00006	0.00085	0.50168	0.00018	0.19982	0.00001
	MLE	3.71577	0.01514	0.82444	0.00035	0.28089	0.00015
r=0.45	Our approach	1.99947	0.00058	0.50000	0.00006	0.44981	0.00001
	MLE	4.58252	0.57929	0.88007	0.00259	0.69203	0.00298
r=0.75	Our approach	1.99781	0.00064	0.49956	0.00001	0.74987	0.00000
	MLE	2.09557	0.00083	0.49442	0.00005	0.78559	0.00001

Table 10: Simulations for non-Gaussian data with $(m, n) = (100, 1000)$

		B							
Parameter p=0.1		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L^*_{\Sigma,A}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$
MA(1)	0.06663	0.06655	0.00007	0.36927	0.30286	0.11244	0.04601	0.06654	0.00009
CS	0.00524	0.94214	0.94107	0.00305	0.00218	0.01244	0.01159	0.00305	0.00219
AR(1)	0.06041	0.15275	0.09222	0.41155	0.35050	0.06028	0.00014	0.06026	0.00015
ARMA-r=0.2	0.08802	0.10912	0.02158	0.16939	0.08187	0.09613	0.00848	0.08790	0.00011
ARMA-r=0.45	0.06531	0.14559	0.08015	0.37091	0.30455	0.06578	0.00078	0.06513	0.00018
ARMA-r=0.75	0.04163	0.18330	0.14164	0.57976	0.53837	0.05665	0.01551	0.04145	0.00018
		B							
Parameter p=0.3		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L^*_{\Sigma,A}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$
MA(1)	0.06300	0.06298	0.00002	0.36733	0.30392	0.10899	0.04613	0.06297	0.00004
CS	0.00475	0.94201	0.94421	0.00282	0.00193	0.01226	0.01137	0.00281	0.00194
AR(1)	0.05711	0.14909	0.09279	0.40836	0.35279	0.05702	0.00009	0.05700	0.00011
ARMA-r=0.2	0.08294	0.10453	0.02163	0.16472	0.08222	0.09139	0.00847	0.08286	0.00007
ARMA-r=0.45	0.06165	0.14208	0.08084	0.36822	0.30737	0.06220	0.00071	0.06156	0.00009
ARMA-r=0.75	0.03919	0.18074	0.14204	0.57774	0.54009	0.05446	0.01550	0.03907	0.00012
		B							
Parameter p=0.5		MA(1)		CS		AR(1)		ARMA(1,1)	
Σ	$L^*_{\Sigma,A}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$	$L^*_{A,B}$	$L^*_{\Sigma,B}$
MA(1)	0.06268	0.06266	0.00002	0.36660	0.30441	0.10878	0.04620	0.06264	0.00004
CS	0.00450	0.94196	0.94690	0.00288	0.00162	0.01234	0.01109	0.00287	0.00162
AR(1)	0.05669	0.14935	0.09256	0.40908	0.35198	0.05660	0.00008	0.05659	0.00010
ARMA-r=0.2	0.08228	0.10385	0.02164	0.16453	0.08227	0.09064	0.00847	0.08221	0.00006
ARMA-r=0.45	0.06128	0.14192	0.08085	0.36803	0.30742	0.06181	0.00072	0.06117	0.00011
ARMA-r=0.75	0.03923	0.18107	0.14227	0.57884	0.54100	0.05449	0.01551	0.03913	0.00010

Table 11: Regularization results for the transformed control chart data

	MA(1)		CS		AR(1)		ARMA(1,1)	
	$L^*_{A,B}$	Time	$L^*_{A,B}$	Time	$L^*_{A,B}$	Time	$L^*_{A,B}$	Time
Normal	0.29221	0.00676	0.51449	0.00511	0.33031	0.01434	0.29204	0.01846
Increasing trend	0.28062	0.00023	0.51337	0.00025	0.31407	0.00933	0.28058	0.01452
Decreasing trend	0.28445	0.00023	0.51078	0.00025	0.31609	0.00925	0.28442	0.01464
Upward shift	0.29431	0.00023	0.47325	0.00025	0.32145	0.00870	0.29394	0.01129
Downward shift	0.32414	0.00023	0.50201	0.00025	0.35293	0.00927	0.32339	0.01168
pooled data	0.17392	0.00820	0.30127	0.00593	0.17342	0.01344	0.17022	0.01464