Chapter 15
Estimation of Covariance Matrix with ARMA Structure Through Quadratic Loss Function

Defei Zhang, Xiangzhao Cui, Chun Li, and Jianxin Pan

Abstract In this paper we propose a novel method to estimate the high-dimensional covariance matrix with an order-1 autoregressive moving average process, i.e. ARMA(1,1), through quadratic loss function. The ARMA(1,1) structure is a commonly used covariance structures in time series and multivariate analysis but involves unknown parameters including the variance and two correlation coefficients. We propose to use the quadratic loss function to measure the discrepancy between a given covariance matrix, such as the sample covariance matrix, and the underlying covariance matrix with ARMA(1,1) structure, so that the parameter estimates can be obtained by minimizing the discrepancy. Simulation studies and real data analysis show that the proposed method works well in estimating the covariance matrix with ARMA(1,1) structure even if the dimension is very high.

Keywords ARMA(1,1) structure · Covariance matrix · Quadratic loss function

15.1 Introduction

Covariance matrix estimation is a fundamental problem in multivariate analysis and time series. Especially, the estimation of high-dimensional covariance matrix is rather challenging. In the literature, many research works were proposed to tackle the problem, such as [1, 3, 8, 9] among many others. However, when the covariance matrix has a certain of structures like order-1 autoregressive moving average, i.e. ARMA(1,1) structure or others, the estimation and regularization were hardly [6]. Recently, Lin et al. [7] proposed a new method to estimate and regularize the high-dimensional covariance matrix. Their idea is summarized as follows. Suppose A is a given $m \times m$ covariance matrix, that is, it is symmetric non-negative definite.

D. Zhang · X. Cui · C. Li
Department of Mathematics, Honghe University, Mengzi 661199, China

J. Pan (✉)
Department of Mathematics, University of Manchester, Manchester M13 9PL, UK
e-mail: jianxin.pan@manchester.ac.uk

J. Fan and J. Pan (eds.), *Contemporary Experimental Design, Multivariate Analysis and Data Mining*,
https://doi.org/10.1007/978-3-030-46161-4_15

Let $\mathscr{S}$ be the set of all $m \times m$ positive definite covariance matrices with structure s, for example, compound symmetry, uniform covariance structure or AR(1). A discrepancy between the given covariance matrix A and the set $\mathscr{S}$ is defined by

$$D(A, \mathscr{S}) = \min_{B \in \mathscr{S}} L(A, B),$$

where $L(A, B)$ is a measure of the discrepancy 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 $D(A, \Omega) = \min_{B \in \Omega} L(A, B)$. The point is that, in this set Ω, the structure with which A has the smallest discrepancy can be viewed as the most likely underlying structure behind A, and the minimizer B with this particular structure is considered to be the regularized covariance matrix of A. Obviously, the bigger the class of candidate structures the better the approximation B to the underlying covariance matrix that is estimated by A. The discrepancy considered by [7] is the so-called entropy loss function and the class of the candidates of potential covariance structures they considered include order-1 moving average MA(1), compound symmetry, AR(1) and Toeplitz structures.

Motivated by this, in this paper we focus on the ARMA(1,1) covariance structure because it includes the MA(1), compound symmetry and AR(1) as its special cases. The ARMA(1,1) process is obtained by applying a recursive filter to the white noise, which is given by the model

$$X_t = \phi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (t = 1, \ldots, m),$$

where ϕ_1 and θ_1 both are parameters, and ε_t is a zero mean white noise process with variance σ_1^2. The covariance matrix of the ARMA(1,1) process (e.g., [2]) is given by

$$\Sigma(\sigma_1, \phi_1, \theta_1) = \frac{(1 + \theta_1^2 + 2\phi_1\theta_1)\sigma_1^2}{1 - \phi_1^2} \begin{bmatrix} 1 & a & a\phi_1 & a\phi_1^2 & \cdots & a\phi_1^{m-2} \\ a & 1 & a & a\phi_1 & \cdots & a\phi_1^{m-3} \\ a\phi_1 & a & 1 & a & \cdots & a\phi_1^{m-4} \\ a\phi_1^2 & a\phi_1 & a & 1 & \cdots & a\phi_1^{m-5} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ a\phi_1^{m-2} & a\phi_1^{m-3} & a\phi_1^{m-4} & \cdots & \cdots & 1 \end{bmatrix}, \tag{15.1}$$

where

$$a := a(\phi_1, \theta_1) = \frac{(1 + \phi_1\theta_1)(\phi_1 + \theta_1)}{1 + \theta_1^2 + 2\phi_1\theta_1}.$$

For simplicity, the covariance matrix in (15.1) can be written as

$$B(\sigma, c, \rho) = \sigma^2 \begin{bmatrix} 1 & c & c\rho & c\rho^2 & \cdots & c\rho^{m-2} \\ c & 1 & c & c\rho & \cdots & c\rho^{m-3} \\ c\rho & c & 1 & c & \cdots & c\rho^{m-4} \\ c\rho^2 & c\rho & c & 1 & \cdots & c\rho^{m-5} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ c\rho^{m-2} & c\rho^{m-3} & c\rho^{m-4} & \cdots & \cdots & 1 \end{bmatrix}. \tag{15.2}$$

where

$$\sigma^2 = \frac{(1+\theta_1^2+2\phi_1\theta_1)\sigma_1^2}{1-\phi_1^2}, \quad c = \frac{(1+\phi_1\theta_1)(\phi_1+\theta_1)}{1+\theta_1^2+2\phi_1\theta_1} \text{ and } \rho = \phi_1$$

It is clear that there are three special cases for the ARMA covariance matrix (15.2). When $\rho = 0$, the structure (15.2) becomes the MA(1) covariance matrix, namely

$$B(c, \sigma) = \sigma^2 \begin{bmatrix} 1 & c & 0 & \cdots & 0 \\ c & 1 & c & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 & c \\ 0 & \cdots & 0 & c & 1 \end{bmatrix}_{m\times m}, \tag{15.3}$$

where $\sigma^2 > 0$ and $-1/\cos(\pi/(m+1)) < c < 1/\cos(\pi/(m+1))$. When $\rho = 1$, it reduces to the compound symmetry structure as

$$B(c, \sigma) = \sigma^2 \begin{bmatrix} 1 & c & c & \cdots & c \\ c & 1 & c & \ddots & \vdots \\ c & \ddots & \ddots & \ddots & c \\ \vdots & \ddots & \ddots & 1 & c \\ c & \cdots & c & c & 1 \end{bmatrix}_{m\times m},$$

where $\sigma^2 > 0$ and $-1/(p-1) < c < 1$ ensure the positive definiteness of the covariance matrix. When $\rho = c$, the structure (15.2) becomes the AR(1) covariance matrix, that is

$$B(c, \sigma) = \sigma^2 \begin{bmatrix} 1 & c & c^2 & \cdots & c^{m-1} \\ c & 1 & c & \cdots & c^{m-2} \\ c^2 & c & 1 & \cdots & c^{m-3} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c^{m-1} & c^{m-2} & \cdots & c & 1 \end{bmatrix}_{m\times m}, \tag{15.4}$$

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

On the other hand, we choose the quadratic loss function rather than the entropy loss function to measure the discrepancy between two matrices. The quadratic loss function was considered by many authors including [3, 9] when estimating covariance matrix. Comparing the entropy loss function, the quadratic loss function avoids the direct calculation of eigenvalues for a likely large covariance matrix with ARMA(1,1) structure. The problem here is that for a given high-dimensional covariance matrix A we aim to find the matrix B with ARMA(1,1) structure such that the discrepancy between A and B is minimized in the domain of the parameters (σ^2, c, ρ). The resulting matrix B is considered to be an approximation to the unknown underlying covariance matrix behind A in terms of structure. The rest of this paper is organized as follows. In Sect. 15.2, we discuss the estimation process under the quadratic loss function and obtain the analytical estimation results. Simulation studies and real data analysis are considered in Sect. 15.3. Conclusions and remarks are provided in Sect. 15.4.

15.2 Estimation Process

We rewrite the covariance matrix of the ARMA(1,1) model as follows,

$$B(c, \rho, \sigma) = \sigma^2 \left(I + c \sum_{i=1}^{m-1} \rho^{i-1} T_i \right), \tag{15.5}$$

where T_i $(1 \le i \le m-1)$ is a symmetric matrix with the ith superdiagonal and subdiagonal elements equal to 1 and zeros elsewhere.

As explained in Sect. 15.1, we propose to use the following quadratic loss function

$$L(\Sigma, B) = tr\left(\Sigma^{-1}B - I_m\right)^2 \tag{15.6}$$

to measure the discrepancy between the matrices Σ and B [4, 10]. Our aim is to find the matrix B^* such that

$$L(\Sigma, B^*) = \min_{\{\sigma,c,\rho\}\in R^+\times[-1,1]^2} L(\Sigma, B)$$

for the underlying population covariance matrix Σ, where $L(\Sigma, B)$ is the function in (15.6). In general, Σ is unknown but can be estimated by an available matrix A such as the sample covariance matrix. Hence, in practice we actually calculate $L(A, B)$ by replacing Σ with A.

Now let $x_0 = \sigma^2$ and $x_i = \sigma^2 c\rho^{i-1}$, $i = 1 : m-1$. The matrix B in (15.5) can be rewritten as

$$B(x) = \sum_{i=0}^{m-1} x_i T_i,$$

where $x = [x_0, x_1, \ldots, x_{m-1}]^T \in \mathbb{R}^m$, $T_0 = I$ and T_i's $(1 \le i \le m-1)$ are already defined in (15.5). We define the set $\Omega \subset \mathbb{R}^m$ by

$$\Omega := \left\{ x \in \mathbb{R}^m : B(x) = \sum_{i=0}^{m-1} x_i T_i \text{ is positive definite} \right\} \tag{15.7}$$

and the function $f(x) : \mathbb{R}^m \mapsto \mathbb{R}$,

$$f(x) := L(\Sigma, B(x)) = tr(\Sigma^{-1}B(x) - I_m)^2.$$

Since Ω is isomorphic to the set of all positive definite matrices, the problem now reduces to minimize the function $f(B)$ over the positive definite matrices B within the set Ω in (15.7).

Since $f(B) := L(\Sigma, B)$ is a strictly convex function of B and $B(x) = \sum_{i=0}^{m-1} x_i T_i$ is an affine map of x, by the fact that a composition with an affine mapping preserves convexity, then function $f(x) := f(B(x))$ is then strictly convex in x. On the other hand, since $\nabla_{x_i} B = T_i$, by applying the chain rule [4, 10] we obtain the gradient of f as

$$\nabla_{x_i} f = 2tr(T_i(\Sigma^{-1}B - I_m)\Sigma^{-1}), \quad i = 0 : m-1,$$

and the Hessian $H = [h_{ij}] \in \mathbb{R}^{m \times m}$ of f where

$$h_{ij} = \nabla^2_{x_i x_j} f = 2tr(T_i \Sigma^{-1} T_j \Sigma^{-1}), \quad i, j = 0 : m-1.$$

Therefore, this is a convex optimization problem so that the function f has a unique minimizer.

The loss function can be now expressed as

$$\begin{aligned}
& f(\sigma, c, \rho) \\
&= tr\left(\Sigma^{-1}B - I_m\right)^2 \\
&= \sigma^4 tr\left(\Sigma^{-1} + c\sum_{i=1}^{m-1} \rho^{i-1}\Sigma^{-1}T_i\right)^2 - 2\sigma^2 tr\left(\Sigma^{-1} + c\sum_{i=1}^{m-1} \rho^{i-1}\Sigma^{-1}T_i\right) + m,
\end{aligned}$$

where

$$\begin{aligned}
& tr\left(\Sigma^{-1} + c\sum_{i=1}^{m-1} \rho^{i-1}\Sigma^{-1}T_i\right)^2 \\
&= c\sum_{i=1}^{m-1} \rho^{i-1} tr(\Sigma^{-2}T_i + \Sigma^{-1}T_i\Sigma^{-1}) + c^2\sum_{i=1}^{m-1} \rho^{2(i-1)} tr\left((\Sigma^{-1}T_i)^2\right) \\
&\quad + tr(\Sigma^{-2}) + c^2 \sum_{i,j=1, i\ne j}^{m-1} \rho^{i+j-2} tr\left(\Sigma^{-1}T_i\Sigma^{-1}T_j + \Sigma^{-1}T_j\Sigma^{-1}T_i\right).
\end{aligned}$$

Therefore, we have

$$
\begin{aligned}
& f(\sigma, c, \rho) \\
&= \sigma^4 tr(\Sigma^{-2}) + c\sigma^4 \sum_{i=1}^{m-1} \rho^{i-1} tr(\Sigma^{-2}T_i + \Sigma^{-1}T_i\Sigma^{-1}) + \sigma^4 c^2 \sum_{i=1}^{m-1} \rho^{2(i-1)} tr\left((\Sigma^{-1}T_i)^2\right) \\
&\quad + \sigma^4 c^2 \sum_{i,j=1, i\neq j}^{m-1} \rho^{i+j-2} tr\left(\Sigma^{-1}T_i\Sigma^{-1}T_j + \Sigma^{-1}T_j\Sigma^{-1}T_i\right) \\
&\quad - 2\sigma^2 tr(\Sigma^{-1}) - 2\sigma^2 c \sum_{i=1}^{m-1} \rho^{i-1} tr(\Sigma^{-1}T_i) + m \\
&= \sigma^4 tr(\Sigma^{-2}) + c\sigma^4 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(1)} + \sigma^4 c^2 \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)} \\
&\quad + \sigma^4 c^2 \sum_{i,j=1, i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)} - 2\sigma^2 tr(\Sigma^{-1}) - 2\sigma^2 c \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(4)} + m
\end{aligned}
$$

where $t_i^{(1)} := tr(\Sigma^{-2}T_i + \Sigma^{-1}T_i\Sigma^{-1})$, $t_i^{(2)} := tr((\Sigma^{-1}T_i)^2)$, $t_{ij}^{(3)} := tr\left(\Sigma^{-1}T_i \Sigma^{-1}T_j + \Sigma^{-1}T_j\Sigma^{-1}T_i\right)$, $t_i^{(4)} := tr(\Sigma^{-1}T_i)$.

Note that the first order partial derivative for $f(\sigma, c, \rho)$ is

$$
\nabla f(\sigma, c, \rho) := \begin{bmatrix} \frac{\partial f}{\partial \sigma} \\ \frac{\partial f}{\partial c} \\ \frac{\partial f}{\partial \rho} \end{bmatrix},
$$

where

$$
\begin{aligned}
\frac{\partial f}{\partial \sigma} &:= 4\sigma^3 tr(\Sigma^{-2}) + 4c\sigma^3 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(1)} + 4\sigma^3 c^2 \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)} \\
&\quad + 4\sigma^3 c^2 \sum_{i,j=1, i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)} - 4\sigma tr(\Sigma^{-1}) - 4\sigma c \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(4)}, \\
\frac{\partial f}{\partial c} &:= \sigma^4 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(1)} + 2\sigma^4 c \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)} + 2\sigma^4 c \sum_{i,j=1, i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)} - 2\sigma^2 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(4)}, \\
\frac{\partial f}{\partial \rho} &:= c\sigma^4 \sum_{i=1}^{m-1} (i-1)\rho^{i-2} t_i^{(1)} + \sigma^4 c^2 \sum_{i=1}^{m-1} (2i-2)\rho^{2i-3} t_i^{(2)} \\
&\quad + \sigma^4 c^2 \sum_{i,j=1, i\neq j}^{m-1} (i+j-2)\rho^{i+j-3} t_{ij}^{(3)} - 2\sigma^2 c \sum_{i=1}^{m-1} (i-1)\rho^{i-2} t_i^{(4)}.
\end{aligned}
$$

Let $\nabla f(\sigma, c, \rho) = 0$. We then have the estimating equations for (σ^2, c, ρ) as follows,

$$\begin{cases} \sigma^2 tr(\Sigma^{-2}) + c\sigma^2 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(1)} + \sigma^2 c^2 \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)} + \sigma^2 c^2 \sum_{i,j=1,i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)} \\ \quad = tr(\Sigma^{-1}) + c \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(4)}, \\ \sigma^2 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(1)} + 2\sigma^2 c \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)} = -2\sigma^2 c \sum_{i,j=1,i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)} + 2 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(4)}, \\ \sigma^2 \sum_{i=1}^{m-1} (i-1)\rho^{i-2} t_i^{(1)} + \sigma^2 c \sum_{i=1}^{m-1} (2i-2)\rho^{2i-3} t_i^{(2)} + \sigma^2 c \sum_{i,j=1,i\neq j}^{m-1} (i+j-2)\rho^{i+j-3} t_{ij}^{(3)} \\ \quad = 2 \sum_{i=1}^{m-1} (i-1)\rho^{i-2} t_i^{(4)}. \end{cases}$$

The Hessian matrix are given by

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

where

$$\begin{aligned} \frac{\partial^2 f}{\partial \rho^2} &:= c\sigma^4 \sum_{i=2}^{m-1} (i-1)(i-2)\rho^{i-3} t_i^{(1)} + \sigma^4 c^2 \sum_{i=1}^{m-1} (2i-2)(2i-3)\rho^{2i-4} t_i^{(2)} \\ &\quad + \sigma^4 c^2 \sum_{i,j=1,i\neq j}^{m-1} (i+j-2)(i+j-3)\rho^{i+j-4} t_{ij}^{(3)} - 2\sigma^2 c \sum_{i=2}^{m-1} (i-1)(i-2)\rho^{i-3} t_i^{(4)}, \\ \frac{\partial^2 f}{\partial \rho \partial c} &:= \sigma^4 \sum_{i=1}^{m-1} (i-1)\rho^{i-2} t_i^{(1)} + 2\sigma^4 c \sum_{i=1}^{m-1} (2i-2)\rho^{2i-3} t_i^{(2)} \\ &\quad + 2c\sigma^4 \sum_{i,j=1,i\neq j}^{m-1} (i+j-2)\rho^{i+j-3} t_{ij}^{(3)} - 2\sigma^2 \sum_{i=1}^{m-1} (i-1)\rho^{i-2} t_i^{(4)}. \\ \frac{\partial^2 f}{\partial \rho \partial \sigma} &:= 4c\sigma^3 \sum_{i=1}^{m-1} (i-1)\rho^{i-2} t_i^{(1)} + 4\sigma^3 c^2 \sum_{i=1}^{m-1} (2i-2)\rho^{2i-3} t_i^{(2)} \\ &\quad + 4\sigma^3 c^2 \sum_{i,j=1,i\neq j}^{m-1} (i+j-2)\rho^{i+j-3} t_{ij}^{(3)} - 4\sigma c \sum_{i=1}^{m-1} (i-1)\rho^{i-2} t_i^{(4)}. \\ \frac{\partial^2 f}{\partial c^2} &:= 2\sigma^4 \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)} + 2\sigma^4 \sum_{i,j=1,i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)}, \end{aligned}$$

$$\frac{\partial^2 f}{\partial c \partial \sigma} := 4\sigma^3 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(1)} + 8\sigma^3 c \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)}$$
$$+ 8\sigma^3 c \sum_{i,j=1,i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)} - 4\sigma \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(4)},$$
$$\frac{\partial^2 f}{\partial \sigma^2} := 12\sigma^2 tr(\Sigma^{-2}) + 12c\sigma^2 \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(1)} + 12\sigma^2 c^2 \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)}$$
$$+ 12\sigma^2 c^2 \sum_{i,j=1,i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)} - 4tr(\Sigma^{-1}) - 4c \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(4)}.$$

Our numerical results including simulation studies and real data analysis show that the determinant $|\nabla^2 f| > 0$ and the theoretical justification is still under investigation.

Theorem 15.1 *Given a positive definite covariance matrix Σ, there exists a unique positive definite matrix $B(\sigma, c, \rho)$ in the form (15.2) such that the quadratic loss function $L(\sigma, c, \rho) := L(\Sigma, B(\sigma, c, \rho))$ in (15.6) is minimized. Furthermore, the minimum must be attained at (σ, c, ρ) that satisfies*

$$\begin{cases} \sigma^2 tr(\Sigma^{-2}) + c\sigma^2 S_1(\rho) + c^2\sigma^2 S_2(\rho) + \sigma^2 c^2 S_3(\rho) = tr(\Sigma^{-1}) + cS_4(\rho), \\ \sigma^2 S_1(\rho) + 2c\sigma^2 S_2(\rho) + 2c\sigma^2 S_3(\rho) = 2S_4(\rho), \\ \sigma^2 S_1'(\rho) + c\sigma^2 S_2'(\rho) + c\sigma^2 S_3'(\rho) = 2S_4'(\rho), \end{cases}$$

where

$$S_1(\rho) := \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(1)}, \qquad S_2(\rho) := \sum_{i=1}^{m-1} \rho^{2i-2} t_i^{(2)},$$
$$S_3(\rho) := \sum_{i,j=1,i\neq j}^{m-1} \rho^{i+j-2} t_{ij}^{(3)}, \qquad S_4(\rho) := \sum_{i=1}^{m-1} \rho^{i-1} t_i^{(4)},$$

and $S_i'(\rho)(i = 1, \ldots, 4)$ is the derivative of $S_i(\rho)$ with respect to ρ, $t_i^{(1)} := tr(\Sigma^{-2}T_i + \Sigma^{-1}T_i\Sigma^{-1})$, $t_i^{(2)} := tr((\Sigma^{-1}T_i)^2)$, $t_{ij}^{(3)} := tr\left(\Sigma^{-1}T_i\Sigma^{-1}T_j + \Sigma^{-1}T_j \Sigma^{-1}T_i\right)$, and $t_i^{(4)} := tr(\Sigma^{-1}T_i)$.

Corollary 15.1 *Given a positive definite covariance matrix Σ, there exists a unique tridiagonal positive definite matrix, i.e. MA(1), $B(c, \sigma)$ in the form (15.3) such that the quadratic loss function $L(c, \sigma) := L(\Sigma, B(\sigma, c))$ in (15.6) is minimized. Furthermore, the minimum must be attained at (σ, c) that satisfies*

$$\begin{cases} \sigma^2 = \dfrac{tr(\Sigma^{-1})tr(\Sigma^{-2}T_1^2) - tr(\Sigma^{-2}T_1)tr(\Sigma^{-1}T_1)}{tr(\Sigma^{-2})tr(\Sigma^{-2}T_1^2) - (tr(\Sigma^{-2}T_1))^2}, \\ c = \dfrac{tr(\Sigma^{-2})tr(\Sigma^{-1}T_1) - tr(\Sigma^{-1})tr(\Sigma^{-2}T_1)}{tr(\Sigma^{-2}T_1^2)tr(\Sigma^{-1}) - tr(\Sigma^{-2}T_1)tr(\Sigma^{-1}T_1)}. \end{cases}$$

Corollary 15.2 *Given a positive definite covariance matrix Σ, there exists a unique AR(1) positive definite matrix $B(c, \sigma)$ in the form (15.4) such that the quadratic loss function $L(c, \sigma) := L(\Sigma, B(\sigma, c))$ in (15.6) is minimized. Furthermore, the minimum must be attained at (σ, c) that satisfies*

$$\begin{cases} \sigma^2 = \dfrac{\sum\limits_{i=0}^{m-1} c^i tr(\Sigma^{-1}T_i)}{\sum\limits_{i=0}^{m-1} c^{2i} tr((\Sigma^{-1}T_i)^2) + 2\sum\limits_{i=0}^{m-2} c^{2i+1} tr(\Sigma^{-1}T_i\Sigma^{-1}T_{i+1})}, \\ \dfrac{\sum\limits_{i=0}^{m-1} ic^{i-1} tr(\Sigma^{-1}T_i)}{\sum\limits_{i=0}^{m-1} c^i tr(\Sigma^{-1}T_i)} = \dfrac{\sum\limits_{i=0}^{m-1} ic^{2i-1} tr((\Sigma^{-1}T_i)^2) + \sum\limits_{i=0}^{m-2} (2i+1)c^{2i} tr(\Sigma^{-1}T_i\Sigma^{-1}T_{i+1})}{\sum\limits_{i=0}^{m-1} c^{2i} tr((\Sigma^{-1}T_i)^2) + 2\sum\limits_{i=0}^{m-2} c^{2i+1} tr(\Sigma^{-1}T_i\Sigma^{-1}T_{i+1})}. \end{cases}$$

Similar results for the compound symmetry structure can be obtained in the same manner but the details are omitted here.

15.3 Numerical Experiments

15.3.1 Simulation Studies

Let m be the dimension of the covariance matrices. We first generate an $m \times n$ data matrix R with columns randomly drawn from the multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$ with a common mean vector $\mu = \sigma^2 e$ with $e = (1, ..., 1)' \in \mathbb{R}^m$ and a common covariance matrix Σ. We then calculate the sample covariance matrix A using the generated random samples R. We assume the true covariance matrix Σ is of ARMA(1,1) structure with dimension m and the parameters (σ^2, c, ρ). We assess the performance of the estimation method by varying dimension m and and values of σ^2, c and ρ. The sample size is chosen as $n = 1000$. We summarize the experimental results in Table 15.1, which is the experiment with the covariance matrix size $m = 100$, and Table 15.2 for $m = 200$. We choose $\sigma^2 \in \{2, 4, 8\}$, $c \in \{0.2, 0.5, 0.75\}$ and $\rho \in \{-0.75, -0.5, -0.2, 0, 0.2, 0.5, 0.75\}$, meaning that Σ may have MA(1), AR(1), and ARMA(1,1) structures, respectively. The notation and abbreviation for the results reported in Tables 15.1 and 15.2 are summarized as follows.

- Σ: the true covariance matrix.
- A: the sample covariance matrix.

- B: the estimated covariance matrix with ARMA(1,1) structure, which minimizes the quadratic loss function $L(A, B)$.
- $L_{\Sigma,A}$, $L_{A,B}$ and $L_{\Sigma,B}$: the quadratic loss functions $L(\Sigma, A)$, $L(A, B)$ and $L(\Sigma, B)$, respectively.

In Tables 15.1 and 15.2, we have the following observations.

(1) When the true covariance structure for Σ is of ARMA(1,1), the resulting matrix B that has the same structure as Σ must satisfy $L_{\Sigma,B} < L_{\Sigma,A}$. It means that the regularized estimator B is much better than the sample covariance matrix A in terms of the quadratic loss function. This is because the sample covariance matrix A contains many noises so that the true ARMA(1,1) structure is blurred if only A is observed. It shows that regularization of the sample covariance matrix A into a proper structure, here ARMA(1,1), is necessary not only for the convenient use of the structure but also for the accuracy of the covariance matrix estimation.
(2) The observations above are the same for differing values of m, σ^2, c and ρ, implying that the findings are consistent and robust against the parameters (σ^2, c, ρ).
(3) Note that it is extremely important to observe the discrepancy $L_{A,B}$ because in practice the true covariance Σ is unknown, and so $L_{\Sigma,B}$ and $L_{\Sigma,A}$ are not possibly known either. The simulation studies presented here aim to assess the performance of the approximation B to the underlying covariance matrix Σ by borrowing information from the sample covariance matrix A. It is concluded that the discrepancy $L_{A,B}$ can be used to identify the true covariance structure of Σ satisfactorily.

Table 15.1 Simulation results with $m = 100$

σ^2	c	ρ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$
2	0.2	−0.75	10.19	27.65	0.23
4	0.2	−0.75	10.22	31.48	0.31
8	0.2	−0.75	10.18	83.64	0.42
2	0.2	−0.5	10.27	34.05	0.96
2	0.5	−0.2	10.01	29.75	0.27
2	0.75	−0.2	9.7	25.03	0.55
2	0.2	0	10.31	26.03	0.25
4	0.5	0	10.01	29.37	0.61
8	0.75	0	10.19	69.48	0.72
2	0.2	0.2	10.19	29.11	0.06
4	0.5	0.5	9.71	29.03	0.93
8	0.75	0.75	10.01	79.43	0.96
2	0.2	−0.2	10.11	78.31	0.98
4	0.5	−0.5	10.03	33.94	0.36
8	0.75	−0.75	10.24	84.21	0.94

Table 15.2 Simulation results with $m = 200$

σ^2	c	ρ	$L_{\Sigma,A}$	$L_{A,B}$	$L_{\Sigma,B}$
2	0.2	−0.75	40.03	46.39	0.38
4	0.2	−0.75	40.56	69.75	0.51
8	0.2	−0.75	40.61	79.01	0.71
2	0.2	−0.5	39.84	72.02	0.31
2	0.5	−0.2	40.09	83.47	0.52
2	0.75	−0.2	39.84	73.57	0.61
2	0.2	0	40.09	84.29	0.54
4	0.5	0	40.69	94.29	0.63
8	0.75	0	40.47	106.44	0.85
2	0.2	0.2	40.02	161.18	0.62
4	0.5	0.5	39.86	83.29	0.72
8	0.75	0.75	39.92	187.27	0.89
2	0.2	−0.2	40.75	92.69	0.55
4	0.5	−0.5	39.36	142.44	0.62
8	0.75	−0.75	40.87	166.73	0.63

15.3.2 Real Data Analysis

15.3.2.1 Cattle Data Analysis

We analyze the Kenward's (1987) [5] cattle data using the proposed approach. The data set involves 60 cattle assigned randomly to two treatment groups 1 and 2, each of which consists of 30 cattle, and received a certain treatment. The cattle in each group were weighed 11 times over a nineteen-week period. The weighing times for all cattle were the same, so that the cattle data is a balanced longitudinal data set. The aim of Kenward's study was to investigate treatment effects on intestinal parasites of the cattle.

Our analysis was made for the cattle data in the same way as in Sect. 15.2 and the results are reported in Table 15.3. We also record, under the column named "Time" in Table 15.3, the time (in seconds) used to find the optimal matrix B for each structure of the possible candidates MA(1), AR(1) and ARMA(1,1).

Table 15.3 Results of experiments for Kenward's cattle data

	MA(1)		AR(1)		ARMA(1,1)	
	$L_{A,B}$	Time	$L_{A,B}$	Time	$L_{A,B}$	Time
Group 1	9.91	2.91	9.46	2.86	9.33	2.82
Group 2	9.53	2.90	9.63	2.76	9.52	2.79

Table 15.4 Results of experiments on Dental data

	MA(1)		AR(1)		ARMA(1,1)	
	$L_{A,B}$	Time	$L_{A,B}$	Time	$L_{A,B}$	Time
Girl group	2.68	0.25	3.43	0.22	2.62	0.21
Boy group	3.01	0.18	3.15	0.18	2.3	0.19

Since the true covariance matrix Σ from the cattle data is unknown, the discrepancies $L_{\Sigma,A}$ and $L_{\Sigma,B}$ are not available and then only the discrepancy $L_{A,B}$ is computed and presented in Table 15.3. From Table 15.3, it is clear that the underlying covariance structures are very likely to be ARMA(1,1) structure for both groups when comparing to other possible candidate structures MA(1) and AR(1), since their discrepancy $L_{A,B}$ has smaller values than other twos.

One may argue that Group 1 is likely to have an AR(1) covariance structure as the values of $L_{A,B}$ for AR(1) and ARMA(1,1) are very close. This should not be surprised because the AR(1) is a special case of the ARMA(1,1) in the sense that c is identical to ρ, see (15.4). This is the case for the Group 1 cattle data analysis due to the fact that the estimates of c and ρ are very close. This conclusion agrees with the finding in [11, 13, 15].

15.3.2.2 Dental Data Analysis

We also did an experiment with dental data (Potthoff and Roy 1964) [12]. Dental measurements were made on 11 girls and 16 boys at ages 8, 10, 12 and 14 years. Each measurement is the distance, in millimeters, from the center of the pituitary to the pterygomaxillary fissure. Similar to the cattle data analysis, the quadratic loss function $L_{A,B}$ is computed for the dental data and presented in Table 15.4.

From Table 15.4, it is clear that the underlying covariance structures are very likely to be ARMA(1,1) for both boy and girl groups, as the discrepancy values of $L_{A,B}$ are smaller than those for both MA(1) and AR(1) structures.

15.4 Conclusions

Motivated by the work of Lin et al. [7], we estimate the underlying covariance structure by minimizing the quadratic loss function between a given covariance matrix and the covariance matrix with ARMA(1,1) structure. Differing from their method, the quadratic loss function is used to replace the entropy loss function where the latter involves the calculation of eigenvalues for a likely large covariance matrix with ARMA(1,1) structure, which is challenging especially for high-dimensional case [14]. Our numerical results including simulation studies and real data analysis show that the proposed method works well in estimating high-dimensional covariance matrices with an underlying ARMA(1,1) structure and is robust against various choices of the parameters involved in the ARMA(1,1) structure.

Acknowledgments This research is supported by the National Science Foundation of China (11761028 and 11871357). We acknowledge helpful comments and insightful suggestions made by a referee.

References

1. Fan, J., Fan, Y., Lv, J.: High dimensional covariance matrix estimation using a factor model. J. Econ. **147**(1), 186–197 (2008)
2. Francq, C.: Covariance matrix estimation for estimators of mixing weak ARMA models. J. Stat. Plan. Inference **83**(2), 369–394 (2000)
3. Haff, L.R.: Empirical bayes estimation of the multivariate normal covariance matrix. Ann. Statist. **8**(3), 586–597 (1980)
4. Horn, R.A., Johnson, C.R.: Matrix Analysis, 2nd edn. Cambridge University Press, Cambridge, UK (2013)
5. Kenward M. G.: A method for comparing profiles of repeated measurements. *Applied Statistics*, **36**(3), 296–308 (1987)
6. Lin, F. Jovanović, M.R.: Least-squares approximation of structured covariances, IEEE Trans. Automat. Control **54**(7), 1643–1648 (2009)
7. Lin, L., Higham, N.J., Pan, J.: Covariance structure regularization via entropy loss function. Comput. Stat. Data Anal. **72**(4), 315–327 (2014)
8. Ning, L., Jiang, X., Georgiou, T.: Geometric methods for structured covariance estimation. In: American Control Conference, pp. 1877–1882. IEEE (2012)
9. Olkin, I., Selliah, J.B.: Estimating covariance matrix in a multivariate normal distribution. In: Gupta, S.S., Moore, D.S. (eds.) Statistical Decision Theory and Related Topics, vol. II, pp. 313–326. Academic Press, New York (1977)
10. Pan, J., Fang, K.: Growth Curve Models and Statistical Diagnostics. Springer, New York (2002)
11. Pan, J., Mackenzie, G.: On modelling mean-covariance structures in longitudinal studies. Biometrika **90**(1), 239–244 (2003)
12. Potthoff R. F., Roy S. N.: A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika **51**(3-4), 313–326 (1964)
13. Pourahmadi M.: Joint mean–covariance models with applications to longitudinal data: unconstrained parameterisation. Biometrika **86**(3), 677–690 (1999)
14. Xiao, H., Wu, W.: Covariance matrix estimation for stationary time series. Ann. Stat. **40**(1), 466–493 (2012)
15. Ye, H., Pan, J.: Modelling of covariance structures in generalised estimating equations for longitudinal data. Biometrika **93**(4), 927–941 (2006)