Estimation equations for multivariate linear models with Kronecker structured covariance matrices

Szczepańska-Álvarez[a], Chengcheng Hao[b], Yuli Liang[c], Dietrich von Rosen[d,e]

[a] Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poland,
[b] School of Business Information, Shanghai University of International Business and Economics, China,
[c] Statistics Sweden, Sweden,
[d] Department of Energy and Technology, Swedish University of Agricultural Sciences, Sweden,
[e] Department of Mathemathics, Linköping University, Sweden

2.12.2016

Model

Consider independent and identically matrix normally distributed observations

$$\mathbf{X}_i \sim N_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}), \quad i = 1, ..., n,$$

$$vec\mathbf{X}_i \sim N_{pq}(vec\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}),$$

where
$E[\mathbf{X}_i] = \boldsymbol{\mu}$ - the expected value,
$D[\mathbf{X}_i] = \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi}$ - the dispersion matrix,
$\boldsymbol{\Psi}$ - the $p \times p$ matrix describing the unknown covariance structure between the rows of $\mathbf{X}_i$,
$\boldsymbol{\Sigma}$ - the $q \times q$ matrix describing the unknown covariance structure between the columns of $\mathbf{X}_i$

Data $\mathbf{Y}$

Let

$$\mathbf{Y}_{(i)} = \mathbf{X}_i - \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i,$$

Moreover,

$$\mathbf{Y} = (\mathbf{Y}_{(1)}, \mathbf{Y}_{(2)}, ..., \mathbf{Y}_{(n)}), \quad \mathbf{Y}_{(i)} = \begin{pmatrix} \mathbf{Y}_{1i} \\ \mathbf{Y}_{2i} \end{pmatrix},$$

$\mathbf{Y}: \ p \times nq$, $\mathbf{Y}_{(i)}: \ p \times q$, $\mathbf{Y}_{1i}: \ r \times q$, $\mathbf{Y}_{2i}: \ (p-r) \times q$, $i = 1, 2, ..., n$, so

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} & ... & \mathbf{Y}_{1n} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} & ... & \mathbf{Y}_{2n} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix},$$

where
$\mathbf{Y}_1: \ r \times nq$, $\mathbf{Y}_2: \ (p-r) \times nq$.

Data $\tilde{\mathbf{Y}}'$

Let

$$\tilde{\mathbf{Y}}' = (\tilde{\mathbf{Y}}'_{(1)}, \tilde{\mathbf{Y}}'_{(2)}, ..., \tilde{\mathbf{Y}}'_{(n)}), \qquad \tilde{\mathbf{Y}}'_{(i)} = \begin{pmatrix} \tilde{\mathbf{Y}}'_{1i} \\ \tilde{\mathbf{Y}}'_{2i} \end{pmatrix}$$

$$\tilde{\mathbf{Y}}' = \begin{pmatrix} \mathbf{Y}'_{11} & \mathbf{Y}'_{12} & \cdots & \mathbf{Y}'_{1n} \\ \mathbf{Y}'_{21} & \mathbf{Y}'_{22} & \cdots & \mathbf{Y}'_{2n} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{Y}}'_{1} \\ \tilde{\mathbf{Y}}'_{2} \end{pmatrix}$$

where
$\tilde{\mathbf{Y}}' : \ q \times np,$
$\tilde{\mathbf{Y}}'_{1i} : \ q \times r,$
$\tilde{\mathbf{Y}}'_{2i} : \ q \times (p - r), \ i = 1, 2, ..., n.$

Case 1

The matrix $\mathbf{\Sigma}$ is unstructured and $\mathbf{\Psi}$ is a partitioned matrix of the form

$$\mathbf{\Psi} = \begin{pmatrix} \mathbf{A}(\theta) & \mathbf{B} \\ \mathbf{B}' & \mathbf{\Omega} \end{pmatrix},$$

where
$\mathbf{A}(\theta)$: $r \times r$, $1 < r < p$, depends on an unknown parameter θ,
$\mathbf{B}$: $r \times (p - r)$,
$\mathbf{\Omega}$: $(p - r) \times (p - r)$ - unknown matrices.

Case 1 - Theorem 1

Given that the maximum likelihood estimator for θ in $\mathbf{A}(\theta)$ can be obtained the maximum likelihood estimators of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ satisfy the following equations:

$$
\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i, \\
\frac{d\mathbf{A}(\widehat{\theta})^{-1}}{d\widehat{\theta}} vec(qn\mathbf{A}(\widehat{\theta}) - \mathbf{Y}_1(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1})\mathbf{Y}_1') &= \mathbf{0}, \\
np\widehat{\boldsymbol{\Sigma}} &= \widetilde{\mathbf{Y}}_1'(\mathbf{I}_n \otimes \mathbf{A}(\widehat{\theta})^{-1})\widetilde{\mathbf{Y}}_1 + (\widetilde{\mathbf{Y}}_2' - \widetilde{\mathbf{Y}}_1'(\mathbf{I}_n \otimes \widehat{\boldsymbol{\delta}}))(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Psi}}_{2\bullet 1}^{-1}) \\
&\quad \times(\widetilde{\mathbf{Y}}_2' - \widetilde{\mathbf{Y}}_1'(\mathbf{I}_n \otimes \widehat{\boldsymbol{\delta}}))', \\
\widehat{\boldsymbol{\delta}} &= (\mathbf{Y}_1(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1})\mathbf{Y}_1')^{-1}\mathbf{Y}_2(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1})\mathbf{Y}_1',
\end{aligned}
$$

Case 1 - Theorem 1

where

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} \mathbf{A}(\widehat{\theta}) & \widehat{\mathbf{B}} \\ \widehat{\mathbf{B}}' & \widehat{\boldsymbol{\Omega}} \end{pmatrix},$$

and

$$\widehat{\mathbf{B}} = \mathbf{A}(\widehat{\theta})\widehat{\boldsymbol{\delta}}, \qquad \widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Psi}}_{2\bullet 1} + \widehat{\boldsymbol{\delta}}'\mathbf{A}(\widehat{\theta})\widehat{\boldsymbol{\delta}},$$

$$qn\widehat{\boldsymbol{\Psi}}_{2\bullet 1} = (\mathbf{Y}_2 - \widehat{\boldsymbol{\delta}}'\mathbf{Y}_1)(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1})(\mathbf{Y}_2 - \widehat{\boldsymbol{\delta}}'\mathbf{Y}_1)'.$$

Case 1 - Corollary 1

Under the assumptions of Theorem 1, if

$$qn\mathbf{A}(\widehat{\theta}) - \mathbf{Y}_1(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1})\mathbf{Y}_1' = \mathbf{0},$$

then,

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i \\
\widehat{\boldsymbol{\Psi}} &= \frac{1}{nq}\mathbf{Y}(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1})\mathbf{Y}', \\
\widehat{\boldsymbol{\Sigma}} &= \frac{1}{np}\widetilde{\mathbf{Y}}'(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Psi}}^{-1})\widetilde{\mathbf{Y}}.
\end{aligned}$$

flip-flop algorithm

P.Dutilleul (1999)

- Since $(c\mathbf{\Sigma}) \otimes (\frac{1}{c}\mathbf{\Psi})$, all the parameters of $\mathbf{\Sigma}$ and $\mathbf{\Psi}$ are not defined uniquely.

- The direct product $\mathbf{\Sigma} \otimes \mathbf{\Psi}$ is uniquely defined.

- The convergence of the MLE algorithm may be assessed by trying various initial solution. If all of the initial solutions tried result in the same direct product $\hat{\mathbf{\Sigma}} \otimes \hat{\mathbf{\Psi}}$ and the corresponding final solutions are $\hat{\mathbf{\Sigma}}, \hat{\mathbf{\Psi}}$ satisfy the criterion of the second derivatives, then any of the final solutions $\hat{\mathbf{\Sigma}}, \hat{\mathbf{\Psi}}$ should provide maximum likelihood estimates for $\mathbf{\Sigma}$ and $\mathbf{\Psi}$; otherwise, they correspond, at the least, to local extrema of the likelihood function.

Case 1 - Corollary 2

Under the assumptions of Theorem 1, if $\mathbf{A}(\widehat{\boldsymbol{\theta}}) = 1$, then

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i \\
\widehat{\boldsymbol{\Psi}} &= \frac{1}{nq}\mathbf{Y}(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1})\mathbf{Y}', \\
\widehat{\boldsymbol{\Sigma}} &= \frac{1}{np}\widetilde{\mathbf{Y}}'(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Psi}}^{-1})\widetilde{\mathbf{Y}}.
\end{aligned}$$

Case 2

The matrix $\mathbf{\Sigma}$ is unstructured and $\mathbf{\Psi}$ is a block partitioned matrix,

$$\mathbf{\Psi} = \begin{pmatrix} \mathbf{A}(\theta) & \mathbf{B} \\ \mathbf{B}' & \mathbf{\Omega} \end{pmatrix},$$

where
$\mathbf{A}(\theta)$ - a compound symmetric structure,
i.e., $\mathbf{A}(\theta) = (1-\theta)\mathbf{I}_r + \theta \mathbf{1}_r \mathbf{1}_r'$, where $\mathbf{1}_r$ denotes the column vector of size r with all elements equal to 1.

Case 2 - Theorem 2

The maximum likelihood estimators of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ satisfy the following equations:

$$
\begin{aligned}
\widehat{\theta} &= \frac{1}{nqr(r-1)} tr(\mathbf{1}_r \mathbf{1}_r' \mathbf{Y}_1 (\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \mathbf{Y}_1') - \frac{1}{r-1}, \\
\widehat{\boldsymbol{\mu}} &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i, \\
\mathbf{A}(\widehat{\theta}) &= (1-\widehat{\theta}) \mathbf{I}_r + \widehat{\theta} \mathbf{1}_r \mathbf{1}_r', \\
np\widehat{\boldsymbol{\Sigma}} &= \widetilde{\mathbf{Y}}_1' (\mathbf{I}_n \otimes \mathbf{A}(\widehat{\theta})^{-1}) \widetilde{\mathbf{Y}}_1 + (\widetilde{\mathbf{Y}}_2' - \widetilde{\mathbf{Y}}_1' (\mathbf{I}_n \otimes \widehat{\boldsymbol{\delta}}))(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Psi}}_{2\bullet 1}^{-1}) \\
&\quad \times (\widetilde{\mathbf{Y}}_2' - \widetilde{\mathbf{Y}}_1' (\mathbf{I}_n \otimes \widehat{\boldsymbol{\delta}}))', \\
\widehat{\boldsymbol{\delta}} &= (\mathbf{Y}_1 (\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \mathbf{Y}_1')^{-1} \mathbf{Y}_2 (\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \mathbf{Y}_1',
\end{aligned}
$$

Case 2 - Theorem 2

where

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} \mathbf{A}(\widehat{\theta}) & \widehat{\mathbf{B}} \\ \widehat{\mathbf{B}}' & \widehat{\boldsymbol{\Omega}} \end{pmatrix},$$

and

$$\widehat{\mathbf{B}} = \mathbf{A}(\widehat{\theta})\widehat{\boldsymbol{\delta}}, \qquad \widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Psi}}_{2\bullet 1} + \widehat{\boldsymbol{\delta}}'\mathbf{A}(\widehat{\theta})\widehat{\boldsymbol{\delta}},$$

$$qn\widehat{\boldsymbol{\Psi}}_{2\bullet 1} = (\mathbf{Y}_2 - \widehat{\boldsymbol{\delta}}'\mathbf{Y}_1)(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Sigma}}^{-1})(\mathbf{Y}_2 - \widehat{\boldsymbol{\delta}}'\mathbf{Y}_1)'.$$

Case 3

Both matrices $\mathbf{\Sigma}$ and $\mathbf{\Psi}$ follow a compound symmetric covariance structure, i.e.

$$\begin{aligned}\mathbf{\Psi} &= (1-\rho)\mathbf{I}_p + \rho\mathbf{1}_p\mathbf{1}_p', \\ \mathbf{\Sigma} &= \sigma_1\mathbf{I}_q + \sigma_2(\mathbf{1}_q\mathbf{1}_q' - \mathbf{I}_q),\end{aligned}$$

where
ρ, σ_1 and σ_2 are unknown parameters.

Case 3 - Theorem 3

The maximum likelihood estimators of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ satisfy

$$\begin{aligned}
\widehat{\boldsymbol{\mu}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i, \\
\widehat{\boldsymbol{\Psi}} &= (1-\widehat{\rho})\mathbf{I}_p + \widehat{\rho}\mathbf{1}_p\mathbf{1}'_p, \\
\widehat{\boldsymbol{\Sigma}} &= \widehat{\sigma}_1\mathbf{I}_q + \widehat{\sigma}_2(\mathbf{1}_q\mathbf{1}'_q - \mathbf{I}_q),
\end{aligned}$$

where

$$\widehat{\sigma}_1 = \tfrac{1}{q}(\widehat{\lambda}_1 + (q-1)\widehat{\lambda}_2), \quad \widehat{\sigma}_2 = \tfrac{1}{q}(\widehat{\lambda}_1 - \widehat{\lambda}_2), \quad \widehat{\rho} = \frac{\widehat{\lambda}_3/\widehat{\lambda}_4 - 1}{\widehat{\lambda}_3/\widehat{\lambda}_4 + p - 1},$$

$\widehat{\lambda}_1$, $\widehat{\lambda}_2$ - distinct eigenvalues of $\boldsymbol{\Sigma}$
$\widehat{\lambda}_3$, $\widehat{\lambda}_4$ - distinct eigenvalues of $\boldsymbol{\Psi}$

Case 3 - Theorem 3

and

$$\hat{\lambda}_1 = \tfrac{1}{np}(\hat{\lambda}_3^{-1} t_1 + \hat{\lambda}_4^{-1} t_2), \qquad \hat{\lambda}_2 = \tfrac{1}{np(q-1)}(\hat{\lambda}_3^{-1} t_3 + \hat{\lambda}_4^{-1} t_4),$$

$$nq\hat{\lambda}_3 - nq\hat{\lambda}_3^2\hat{\lambda}_4^{-1} - \hat{\lambda}_1^{-1} t_1 - \hat{\lambda}_2^{-1} t_3 + \frac{\hat{\lambda}_1^{-1}\hat{\lambda}_4^{-2}\hat{\lambda}_3^2 t_2}{(p-1)}$$

$$+\frac{\hat{\lambda}_2^{-1}\hat{\lambda}_4^{-2}\hat{\lambda}_3^2 t_4}{(p-1)} = 0,, \qquad p = \hat{\lambda}_3 + \hat{\lambda}_4(p-1),$$

with

$$t_1 = tr\{\mathbf{P}_{1_p}\mathbf{Y}(\mathbf{I}_n \otimes \mathbf{P}_{1_q})\mathbf{Y}'\}, \quad t_2 = tr\{\mathbf{Q}_{1_p}\mathbf{Y}(\mathbf{I}_n \otimes \mathbf{P}_{1_q})\mathbf{Y}'\},$$
$$t_3 = tr\{\mathbf{P}_{1_p}\mathbf{Y}(\mathbf{I}_n \otimes \mathbf{Q}_{1_q})\mathbf{Y}'\}, \quad t_4 = tr\{\mathbf{Q}_{1_p}\mathbf{Y}(\mathbf{I}_n \otimes \mathbf{Q}_{1_q})\mathbf{Y}'\},$$

where $\mathbf{P}_{1_p} = \frac{1}{p}\mathbf{1}_p\mathbf{1}_p'$ and $\mathbf{Q}_{1_p} = \mathbf{I}_p - \mathbf{P}_{1_p}$ and the observation matrix $\mathbf{Y}$ is the centered observation matrix.

Case 4

The matrix $\mathbf{\Sigma}$ is unstructured and in $\mathbf{\Psi}$ is the matrix which all diagonal elements equal 1.

$$\mathbf{\Psi} = \mathbf{T}_d^{-\frac{1}{2}} \mathbf{T} \mathbf{T}_d^{-\frac{1}{2}},$$

where
$\mathbf{T}$: $p \times p$ - the symmetric matrix,
$\mathbf{T}_d$: $p \times p$ - the diagonal matrix with diagonal elements the same as matrix $\mathbf{T}$.

Case 4 - Theorem 4

Maximum likelihood equations are given by the following relations:

$$np\mathbf{\Sigma} = \tilde{\mathbf{Y}}'(\mathbf{I}_n \otimes (\mathbf{T}_d^{\frac{1}{2}}\mathbf{T}^{-1}\mathbf{T}_d^{\frac{1}{2}})^{-1})\tilde{\mathbf{Y}}$$

$$-2\mathbf{T}^{-1}\mathbf{T}_d^{\frac{1}{2}}\mathbf{A}\mathbf{T}_d^{\frac{1}{2}}\mathbf{T}^{-1} + (\mathbf{T}^{-1}\mathbf{T}_d^{\frac{1}{2}}\mathbf{A}\mathbf{T}_d^{\frac{1}{2}}\mathbf{T}^{-1})_d + (\mathbf{T}^{-1}\mathbf{T}_d^{\frac{1}{2}}\mathbf{A})_d\mathbf{T}_d^{-\frac{1}{2}} = \mathbf{0},$$

where

$$\mathbf{A} = 2nq\mathbf{T}_d^{\frac{1}{2}}\mathbf{T}^{-1}\mathbf{T}_d^{\frac{1}{2}} - nq\mathbf{I}_p - \mathbf{Y}(\mathbf{I}_n \otimes \mathbf{\Sigma}^{-1})\mathbf{Y}'$$

and

$(\mathbf{T}^{-1}\mathbf{T}_d^{\frac{1}{2}}\mathbf{A}\mathbf{T}_d^{\frac{1}{2}}\mathbf{T}^{-1})_d, \quad (\mathbf{T}^{-1}\mathbf{T}_d^{\frac{1}{2}}\mathbf{A})_d$ denote diagonal matrices.

Literature

Dutilleul P. (1999). The MLE algorithm for the matrix normal distribution. J. Statist. Comput. Simul, vol. 64, 105-123.

Srivastava M. S., von Rosen T. and von Rosen D.(2008). Models with a Kronecker Product Covariance Structure: Estimation and Testing. Mathematical Methods of Statistics, vol. 17, No. 4, 357–370.

Szczepańska-Álvarez A, Hao Ch., Liang Y., von Rosen D. (2016). Estimation equations for multivariate linear models with Kronecker structured covariance matrices. Communications in Statistics-Theory and Methods. DOI10.1080/03610926.2016.1165852