Supplement to the paper "Predictive estimation of a covariance matrix and its structural parameters"

Haruhiko Ogasawara

This supplement gives the proofs of the theorems and corollaries in Ogasawara (2017).

A.1 Proof of Theorem 1

We minimize the following with respect to k:

$$
\begin{aligned}
\mathrm{E}^{(\mathbf{S})}(-2\hat{\bar{l}}^{*}) &= \mathrm{E}^{(\mathbf{S})}\{\ln|\hat{\boldsymbol{\Sigma}}| + \mathrm{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_0)\} + \mathrm{E}^{(\mathbf{T})}\{C(\mathbf{T})\} \\
&= p\ln k + \mathrm{E}^{(\mathbf{S})}\{\ln|\mathbf{S}|\} + k^{-1}\mathrm{E}^{(\mathbf{S})}\{\mathrm{tr}(\mathbf{S}^{-1}\boldsymbol{\Sigma}_0)\} + \mathrm{E}^{(\mathbf{S})}\{C(\mathbf{S})\} \\
&= p\ln k + \frac{np}{k(n-p-1)} + \frac{p+1}{n}\{\ln|\boldsymbol{\Sigma}_0| + p\ln(2/n) + \psi_p(n/2)\} \\
&\quad + p\ln(2/n) + \frac{2}{n}\ln\Gamma_p(n/2),
\end{aligned}
\tag{A1.1}
$$

where under normality $\mathrm{E}^{(\mathbf{S})}(\mathbf{S}^{-1}) = \{n/(n-p-1)\}\boldsymbol{\Sigma}_0^{-1}$ (Anderson, 2003, Lemma 7.7.1) is used. Differentiating (A1.1) with respect to k and setting the result equal to 0, we obtain (2.8).

A.2 Proof of Theorem 2

Noting that $n = N-1$, the following is minimized with respect to k:

$$
\begin{aligned}
\mathrm{E}^{(\mathbf{X})}(-2\hat{\bar{l}}^{*}) \\
&= p\ln(2\pi) + p\ln k + \mathrm{E}^{(\mathbf{X})}(\ln|\hat{\boldsymbol{\Sigma}}_{\mathrm{NML}}|) \\
&\quad + \mathrm{tr}\{k^{-1}\mathrm{E}^{(\mathbf{X})}(\hat{\boldsymbol{\Sigma}}_{\mathrm{NML}}^{-1})(\boldsymbol{\Sigma}_0 + N^{-1}\boldsymbol{\Sigma}_0)\} \\
&= p\ln(2\pi) + p\ln k + \mathrm{E}^{(\mathbf{X})}(\ln|\hat{\boldsymbol{\Sigma}}_{\mathrm{NML}}|) \\
&\quad + k^{-1}\frac{N}{N-1}\frac{n}{n-p-1}(1+N^{-1})p
\end{aligned}
\tag{A2.1}
$$

$$= p \ln k + \frac{(N+1)p}{k(N-p-2)} + p \ln(2\pi) + p \ln \frac{N-1}{N}$$

$$+ \ln |\mathbf{\Sigma}_0| + p \ln(2/n) + \psi_p(n/2),$$

where the independence of $\overline{\mathbf{x}}$ and $\hat{\mathbf{\Sigma}}_{\text{NML}}$ under normality is used. Differentiating (A2.1) with respect to k, we obtain (2.17).

A.3 Proof of Theorem 3

First, we have

$$\overline{l} = -\frac{p}{2}\ln(2\pi) - \frac{1}{2}\ln |\mathbf{\Sigma}| - \frac{N^{-1}}{2}\sum_{i=1}^{N}(\mathbf{x}_i - \mathbf{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}_i - \mathbf{\mu}) \quad \text{(A3.1)}$$

which gives

$$\frac{\partial \overline{l}}{\partial \mathbf{\theta}_0} = \left(\frac{\partial \overline{l}}{\partial \mathbf{\mu}_0'}, \frac{\partial \overline{l}}{\partial \mathbf{\sigma}_0'} \right)', \quad \frac{\partial \overline{l}}{\partial \mathbf{\mu}_0} = \mathbf{\Sigma}_0^{-1}(\overline{\mathbf{x}} - \mathbf{\mu}_0),$$

$$\frac{\partial \overline{l}}{\partial \sigma_{0ij}} = -\frac{2-\delta_{ij}}{2}\sigma_0^{ij} + \frac{2-\delta_{ij}}{2}\left(N^{-1}\sum_{a=1}^{N}\mathbf{\Sigma}_0^{-1}(\mathbf{x}_a - \mathbf{\mu}_0)(\mathbf{x}_a - \mathbf{\mu}_0)'\mathbf{\Sigma}_0^{-1} \right)_{ij}$$

$$(p \geq i \geq j \geq 1),$$

$$\hat{\mathbf{\theta}}_{\text{NML}} = \left[\overline{\mathbf{x}}', \text{v}'\left\{ N^{-1}\sum_{a=1}^{N}\mathbf{\Sigma}_0^{-1}(\mathbf{x}_a - \overline{\mathbf{x}})(\mathbf{x}_a - \overline{\mathbf{x}})' \right\} \right]'$$

$$= \{\hat{\mathbf{\mu}}_{\text{NML}}', \text{v}'(\hat{\mathbf{\Sigma}}_{\text{NML}})\}' = (\hat{\mathbf{\mu}}_{\text{NML}}', \hat{\mathbf{\sigma}}_{\text{NML}}')',$$

$$\frac{\partial^2 \overline{l}}{\partial \mathbf{\theta}_0 \partial \mathbf{\theta}_0'} = \begin{bmatrix} \dfrac{\partial^2 \overline{l}}{\partial \mathbf{\mu}_0 \partial \mathbf{\mu}_0'} & \dfrac{\partial^2 \overline{l}}{\partial \mathbf{\mu}_0 \partial \mathbf{\sigma}_0'} \\ \dfrac{\partial^2 \overline{l}}{\partial \mathbf{\sigma}_0 \partial \mathbf{\mu}_0'} & \dfrac{\partial^2 \overline{l}}{\partial \mathbf{\sigma}_0 \partial \mathbf{\sigma}_0'} \end{bmatrix}, \quad \frac{\partial^2 \overline{l}}{\partial \mathbf{\mu}_0 \partial \mathbf{\mu}_0'} = -\mathbf{\Sigma}_0^{-1}, \quad \text{(A3.2)}$$

$$\frac{\partial^2 \bar{l}}{\partial \boldsymbol{\mu}_0 \partial \sigma_{0ij}} = -\frac{2-\delta_{ij}}{2} \boldsymbol{\Sigma}_0^{-1}(\mathbf{E}_{ij}+\mathbf{E}_{ji})\boldsymbol{\Sigma}_0^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)$$

$$= -\frac{2-\delta_{ij}}{2}\sum_{(i,j)}^{2}(\boldsymbol{\Sigma}_0^{-1})_{\cdot i}\{\boldsymbol{\Sigma}_0^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)\}_j \quad (p \geq i \geq j \geq 1),$$

$$\frac{\partial^2 \bar{l}}{\partial \sigma_{0ab}\partial \sigma_{0cd}} = \frac{1}{4}(2-\delta_{ab})(2-\delta_{cd})(\sigma_0^{ac}\sigma_0^{bd}+\sigma_0^{ad}\sigma_0^{bc})$$

$$-\frac{1}{4}(2-\delta_{ab})(2-\delta_{cd})$$

$$\times \sum_{(a,b)}^{2}\sum_{(c,d)}^{2}\sigma_0^{ac}\left\{N^{-1}\sum_{i=1}^{N}\boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_0)(\mathbf{x}_i-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}\right\}_{db}$$

$$(p \geq a \geq b \geq 1; \ p \geq c \geq d \geq 1),$$

where δ_{ij} is the Kronecker delta; $\sigma_{0ij}=(\boldsymbol{\Sigma}_0)_{ij}$ and $\sigma_0^{ij}=(\boldsymbol{\Sigma}_0^{-1})_{ij}$; $(\cdot)_{ij}$ is the (i,j)th element of a matrix; $\mathbf{E}_{ij}$ is the matrix of an appropriate size whose (i,j)th element is 1 with the remaining elements being 0; $\sum_{(i,j)}^{(2)}(\cdot)$ indicates the sum of two quantities exchanging i and j; $(\cdot)_{\cdot i}$ is the i-th column of a matrix; and $(\cdot)_j$ is the j-th element of a vector.

The non-zero submatrices of $\mathbf{I}_0$ are given by (A3.2) as follows:

$$(\mathbf{I}_0)_{\boldsymbol{\mu}_0,\boldsymbol{\mu}_0'} = \boldsymbol{\Sigma}_0^{-1}, \quad (\mathbf{I}_0)_{\boldsymbol{\sigma}_0,\boldsymbol{\sigma}_0'} = \frac{1}{2}\mathbf{D}_p'(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\Sigma}_0^{-1})\mathbf{D}_p, \quad \text{(A3.3)}$$

where $(\cdot)_{\mathbf{x},\mathbf{y}'}$ indicates a submatrix of a matrix whose rows and columns correspond to $\mathbf{x}$ and $\mathbf{y}'$, respectively. The elementwise expression of the last submatrix of (A3.3) is

$$(\mathbf{I}_0)_{\sigma_{0ab},\sigma_{0cd}} = \frac{1}{4}(2-\delta_{ab})(2-\delta_{cd})(\sigma_0^{ac}\sigma_0^{bd}+\sigma_0^{ad}\sigma_0^{bc})$$

$$(p \geq a \geq b \geq 1; \ p \geq c \geq d \geq 1).$$

(A3.4)

Equation (A3.3) gives

$$N \operatorname{acov}(\hat{\boldsymbol{\theta}}_{\mathrm{NML}}) = \mathbf{I}_0^{-1}, \ (\mathbf{I}_0^{-1})_{\boldsymbol{\mu}_0, \boldsymbol{\mu}_0'} = \boldsymbol{\Sigma}_0, \ (\mathbf{I}_0^{-1})_{\boldsymbol{\sigma}_0, \boldsymbol{\mu}_0'} = \mathbf{O},$$

$$(\mathbf{I}_0^{-1})_{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_0'} = 2\mathbf{D}_p^+ (\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0) \mathbf{D}_p^+ {}'. \tag{A3.5}$$

where $\operatorname{acov}(\cdot)$ is the asymptotic covariance matrix of order $O(N^{-1})$ for a stochastic vector and $\mathbf{O}$ is a zero matrix. The elementwise expression of the last submatrix is

$$(\mathbf{I}_0^{-1})_{\sigma_{0ab}, \sigma_{0cd}} = \sigma_{0ac}\sigma_{0bd} + \sigma_{0ad}\sigma_{0bc} \ (p \geq a \geq b \geq 1; \ p \geq c \geq d \geq 1). \tag{A3.6}$$

From the above results we obtain non-zero elements of $N\operatorname{E}\left(\dfrac{\partial \overline{l}}{\partial \boldsymbol{\theta}_0'} \otimes \mathbf{M}\right)$ in (2.22) as follows:

$$N\operatorname{E}\left\{\frac{\partial \overline{l}}{\partial \boldsymbol{\mu}_0'} \otimes (\mathbf{M})_{\sigma_{0ab}, \boldsymbol{\mu}_0'}\right\} = -\frac{2 - \delta_{ab}}{2} \sum_{(a,b)}^{2} (\boldsymbol{\Sigma}_0^{-1})_{a\bullet}(\boldsymbol{\Sigma}_0^{-1})_{b\bullet},$$

$$N\operatorname{E}\left\{\frac{\partial \overline{l}}{\partial \sigma_{0ab}} \otimes (\mathbf{M})_{\sigma_{0cd}, \sigma_{0ef}}\right\} = \frac{1}{8}(2 - \delta_{ab})(2 - \delta_{cd})(2 - \delta_{ef})$$

$$\times N\operatorname{E}\left[\underbrace{\left[\boldsymbol{\Sigma}_0^{-1}\left\{N^{-1}\sum_{i=1}^{N}(\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_a - \boldsymbol{\mu}_0)' - \boldsymbol{\Sigma}_0\right\}\boldsymbol{\Sigma}_0^{-1}\right]_{ab}}_{(A)}\right.$$

$$\times \left[\ 2(\sigma_0^{ce}\sigma_0^{fd} + \sigma_0^{cf}\sigma_0^{ed})\right.$$

$$\left.\left. - \sum_{(c,d)}^{2}\sum_{(e,f)}^{2}\sigma_0^{ce}\left\{N^{-1}\sum_{i=1}^{N}\boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}\right\}_{fd}\right]_{(A)}\right]$$

$$= \frac{1}{8}(2-\delta_{ab})(2-\delta_{cd})(2-\delta_{ef})\left\{ 2\sigma_0^{ab}(\sigma_0^{ce}\sigma_0^{fd}+\sigma_0^{cf}\sigma_0^{ed}) \right.$$

$$\left. -\sum_{g,h=1}^{p}\sum_{i,j=1}^{p}\sigma_0^{ag}\sigma_0^{hb}\sum_{(c,d)}^{2}\sum_{(e,f)}^{2}\sigma_0^{ce}\sigma_0^{fi}\sigma_0^{jd}(\sigma_{0gh}\sigma_{0ij}+\sigma_{0gi}\sigma_{0hj}+\sigma_{0gj}\sigma_{0ih}) \right\}$$

$$= \frac{1}{8}(2-\delta_{ab})(2-\delta_{cd})(2-\delta_{ef})\left\{ 2\sigma_0^{ab}(\sigma_0^{ce}\sigma_0^{fd}+\sigma_0^{cf}\sigma_0^{ed}) \right.$$

$$\left. -\sum_{(c,d)}^{2}\sum_{(e,f)}^{2}\sigma_0^{ce}(\sigma_0^{ab}\sigma_0^{fd}+\sigma_0^{af}\sigma_0^{bd}+\sigma_0^{ad}\sigma_0^{bf}) \right\}$$

$$= -\frac{1}{8}(2-\delta_{ab})(2-\delta_{cd})(2-\delta_{ef})\sum_{(c,d)}^{2}\sum_{(e,f)}^{2}\sigma_0^{ce}(\sigma_0^{af}\sigma_0^{bd}+\sigma_0^{ad}\sigma_0^{bf}) \right\} \tag{A3.7}$$

$(p \geq a \geq b \geq 1;\ p \geq c \geq d \geq 1;\ p \geq e \geq f \geq 1),$

where $(\cdot)_{a\bullet}$ is the a-th row of a matrix and $\underset{(A)\ (A)}{[\cdot]}$ is for ease of finding

correspondence. Note that $N\mathrm{E}\left\{ \dfrac{\partial \bar{l}}{\partial \boldsymbol{\mu}_0{}'}\otimes(\mathbf{M})_{\boldsymbol{\mu}_0,\,\boldsymbol{\sigma}_0{}'} \right\}$ is also non-zero. However,

this will not be used. Then, (2.22) becomes

$$\mathbf{k} = \mathrm{diag}^{-1}(\boldsymbol{\theta}_0)\mathbf{I}_0^{-1}\left\{ N\mathrm{E}\left(\frac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0{}'}\otimes\mathbf{M} \right)\mathrm{vec}(\mathbf{I}_0^{-1}) + \mathrm{diag}^{-1}(\boldsymbol{\theta}_0)\mathbf{1}_{(q)} \right\}$$

$$= \mathrm{diag}^{-1}(\boldsymbol{\mu}_0{}',\boldsymbol{\sigma}_0{}')'\begin{bmatrix} \boldsymbol{\Sigma}_0 & \mathbf{O} \\ \mathbf{O} & 2\mathbf{D}_p^{+}(\boldsymbol{\Sigma}_0\otimes\boldsymbol{\Sigma}_0)\mathbf{D}_p^{+}{}' \end{bmatrix}$$

$$\times\left\{ \begin{pmatrix} \mathbf{0}\ (p\times 1) \\ \mathbf{x}_{\boldsymbol{\sigma}}(p^{*}\times 1) \end{pmatrix} + \begin{pmatrix} \mathbf{0}\ (p\times 1) \\ \mathbf{y}_{\boldsymbol{\sigma}}(p^{*}\times 1) \end{pmatrix} + \mathrm{diag}^{-1}(\boldsymbol{\mu}_0{}',\boldsymbol{\sigma}_0{}')'\mathbf{1}_{(q)} \right\}, \tag{A3.8}$$

where $\mathbf{0}\ (p\times 1)$ is the $p\times 1$ zero vector; and

$$(\mathbf{x}_{\boldsymbol{\sigma}})_{(ef)} = -\frac{2-\delta_{ef}}{2}(\sigma_0^{ef} + \sigma_0^{fe}) = -(2-\delta_{ef})\sigma_0^{ef},$$

$$(\mathbf{y}_{\boldsymbol{\sigma}})_{(ef)}$$

$$= \frac{2-\delta_{ef}}{8}\sum_{a,b=1}^{p}\sum_{c,d=1}^{p}[-\{\sigma_0^{ce}(\sigma_0^{af}\sigma_0^{bd} + \sigma_0^{ad}\sigma_0^{bf}) + \sigma_0^{de}(\sigma_0^{af}\sigma_0^{bc} + \sigma_0^{ac}\sigma_0^{bf})$$

$$+ \sigma_0^{cf}(\sigma_0^{ae}\sigma_0^{bd} + \sigma_0^{ad}\sigma_0^{be}) + \sigma_0^{df}(\sigma_0^{ae}\sigma_0^{bc} + \sigma_0^{ac}\sigma_0^{be})\}] \quad (A3.9)$$

$$\times(\sigma_{0ac}\sigma_{0bd} + \sigma_{0ad}\sigma_{0bc})$$

$$= \frac{2-\delta_{ef}}{8}\{-8(p+1)\sigma_0^{ef}\} = -(2-\delta_{ef})(p+1)\sigma_0^{ef} \quad (p \ge e \ge f \ge 1),$$

where $(\cdot)_{(ef)}$ is the element of a vector corresponding to σ_{0ef} using the

double subscript notation, and $\sum_{a \ge b}\sum_{c \ge d}(2-\delta_{ab})(2-\delta_{cd})(\cdot) = \sum_{a,b=1}^{p}\sum_{c,d=1}^{p}(\cdot)$ is

used.

From (A3.9), we have $(\mathbf{x}_{\boldsymbol{\sigma}} + \mathbf{y}_{\boldsymbol{\sigma}})_{(ef)} = -(2-\delta_{ef})(p+2)\sigma_0^{ef}$, which is

rewritten as

$$\mathbf{x}_{\boldsymbol{\sigma}} + \mathbf{y}_{\boldsymbol{\sigma}} = -(p+2)\mathbf{D}_p{}'\text{vec}(\boldsymbol{\Sigma}_0^{-1}). \quad (A3.10)$$

From (A3.8) and (A3.10), $\mathbf{k}_{\boldsymbol{\mu}_0}$ and $\mathbf{k}_{\boldsymbol{\sigma}_0}$ in $\mathbf{k} = (\mathbf{k}_{\boldsymbol{\mu}_0}{}', \mathbf{k}_{\boldsymbol{\sigma}_0}{}')'$ are

$$\mathbf{k}_{\boldsymbol{\mu}_0} = \text{diag}^{-1}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0\text{diag}^{-1}(\boldsymbol{\mu}_0)\mathbf{1}_{(p)} \quad (A3.11)$$

and

$$\mathbf{k}_{\boldsymbol{\sigma}_0} = \text{diag}^{-1}(\boldsymbol{\sigma}_0)2\mathbf{D}_p^{+}(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\mathbf{D}_p^{+}{}'$$

$$\times\{-(p+2)\mathbf{D}_p{}'\text{vec}(\boldsymbol{\Sigma}_0^{-1}) + \text{diag}^{-1}(\boldsymbol{\sigma}_0)\mathbf{1}_{(p*)}\}$$

$$= -2(p+2)\text{diag}^{-1}(\boldsymbol{\sigma}_0)\mathbf{D}_p^{+}\text{vec}(\boldsymbol{\Sigma}_0) \quad (A3.12)$$

$$+ 2\text{diag}^{-1}(\boldsymbol{\sigma}_0)\mathbf{D}_p^{+}(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\mathbf{D}_p^{+}{}'\text{diag}^{-1}(\boldsymbol{\sigma}_0)\mathbf{1}_{(p*)},$$

respectively, where

$$(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\mathbf{D}_p^{+}{}'\mathbf{D}_p{}' = \mathbf{D}_p^{+}{}'\mathbf{D}_p{}'(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0), \ \mathbf{D}_p^{+}\mathbf{D}_p^{+}{}'\mathbf{D}_p{}' = \mathbf{D}_p^{+} \text{ and } (A3.13)$$

$$(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\text{vec}(\boldsymbol{\Sigma}_0^{-1}) = \text{vec}(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_0) = \text{vec}(\boldsymbol{\Sigma}_0)$$

are used. In (A3.12), noting $\mathbf{D}_p^+ \mathrm{vec}(\mathbf{\Sigma}_0) = \mathbf{\sigma}_0$, (A3.12) becomes

$$\mathbf{k}_{\mathbf{\sigma}_0} = -2(p+2)\mathbf{1}_{(p*)}$$
$$+ 2\mathrm{diag}^{-1}(\mathbf{\sigma}_0)\mathbf{D}_p^+(\mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0)\mathbf{D}_p^+{}'\mathrm{diag}^{-1}(\mathbf{\sigma}_0)\mathbf{1}_{(p*)}. \qquad \text{(A3.14)}$$

Equations (A3.11) and (A3.14) give (2.25) and (2.26), respectively.

A.4 Proof of Theorem 4

We derive

$$\mathbf{k} = \mathrm{diag}^{-1}(\mathbf{\theta}_0)\mathbf{I}_0^{-1}\left\{ n\mathrm{E}\left(\frac{\partial \overline{l}}{\partial \mathbf{\theta}_0{}'} \otimes \mathbf{M} \right)\mathrm{vec}(\mathbf{I}_0^{-1}) + \mathrm{diag}^{-1}(\mathbf{\theta}_0)\mathbf{1}_{(q)} \right\} \qquad \text{(A4.1)}$$

(see (2.22)). Noting $\overline{l} = -\frac{1}{2}\ln|\mathbf{\Sigma}| - \frac{1}{2}\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}) - \frac{1}{2}C(\mathbf{S})$ (for $C(\mathbf{S})$ see (2.4)), we have

$$\frac{\partial \overline{l}}{\partial \theta_{0i}} = \frac{1}{2}\mathrm{tr}\left\{ \mathbf{\Sigma}_0^{-1}(\mathbf{S} - \mathbf{\Sigma}_0)\mathbf{\Sigma}_0^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_{0i}} \right\},$$

$$\frac{\partial^2 \overline{l}}{\partial \theta_{0i}\partial \theta_{0j}} = -\mathrm{tr}\left\{ \mathbf{\Sigma}_0^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_{0j}}\mathbf{\Sigma}_0^{-1}(\mathbf{S} - \mathbf{\Sigma}_0)\mathbf{\Sigma}_0^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_{0i}} \right\}$$

$$- \frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}_0^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_{0j}}\mathbf{\Sigma}_0^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_{0i}} \right) \qquad \text{(A4.2)}$$

$$+ \frac{1}{2}\mathrm{tr}\left\{ \mathbf{\Sigma}_0^{-1}(\mathbf{S} - \mathbf{\Sigma}_0)\mathbf{\Sigma}_0^{-1}\frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_{0i}\partial \theta_{0j}} \right\},$$

$$(\mathbf{M})_{ij} = \frac{\partial^2 \bar{l}}{\partial \theta_{0i} \partial \theta_{0j}} - \mathrm{E}\left(\frac{\partial^2 \bar{l}}{\partial \theta_{0i} \partial \theta_{0j}}\right) = \frac{\partial^2 \bar{l}}{\partial \theta_{0i} \partial \theta_{0j}} + (\mathbf{I}_0)_{ij}$$

$$= -\mathrm{tr}\left\{ \boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0i}} \boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0j}} \boldsymbol{\Sigma}_0^{-1}(\mathbf{S} - \boldsymbol{\Sigma}_0) \right\}$$

$$+ \frac{1}{2}\mathrm{tr}\left\{ \boldsymbol{\Sigma}_0^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_{0i} \partial \theta_{0j}} \boldsymbol{\Sigma}_0^{-1}(\mathbf{S} - \boldsymbol{\Sigma}_0) \right\}$$

$(i, j = 1, ..., q)$.

Then, it follows that

$$n\mathrm{E}\left\{ \frac{\partial \bar{l}}{\partial \theta_{0a}}(\mathbf{M})_{bc} \right\} = \frac{1}{4}\mathrm{vec}'\left(\boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0a}} \boldsymbol{\Sigma}_0^{-1} \right) n\,\mathrm{cov}\{\mathrm{vec}(\mathbf{S})\}$$

$$\times \mathrm{vec}\left(-2\boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0b}} \boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0c}} \boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_0^{-1} \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_{0b} \partial \theta_{0c}} \boldsymbol{\Sigma}_0^{-1} \right)$$

$$= \frac{1}{2}\mathrm{vec}'\left(\boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0a}} \boldsymbol{\Sigma}_0^{-1} \right) \boldsymbol{\Pi}_p(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\boldsymbol{\Pi}_p$$

$$\times \mathrm{vec}\left\{ \boldsymbol{\Sigma}_0^{-1}\left(-2\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0b}} \boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0c}} + \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_{0b} \partial \theta_{0c}} \right)\boldsymbol{\Sigma}_0^{-1} \right\} \quad \text{(A4.3)}$$

$$= \frac{1}{2}\mathrm{tr}\left\{ \boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0a}} \boldsymbol{\Sigma}_0^{-1}\left(-2\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0b}} \boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0c}} + \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_{0b} \partial \theta_{0c}} \right) \right\}$$

$(a, b, c = 1, ..., q)$,

where $n\,\mathrm{cov}\{\mathrm{vec}(\mathbf{S})\} = 2\mathbf{D}_p(\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{D}_p'(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\mathbf{D}_p(\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{D}_p'$

$\equiv 2\boldsymbol{\Pi}_p(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0)\boldsymbol{\Pi}_p$ and $\boldsymbol{\Pi}_p\mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\mathbf{A})$ when $\mathbf{A} = \mathbf{A}'$, are used.

Note that $n\,\mathrm{cov}\{\mathrm{vec}(\mathbf{S})\}$ under normality is an exact result (see, e.g., Kaplan, 1952, Equation (3)). The (i, j)th element of the information matrix is given by

$$(\mathbf{I}_0)_{ij} = -\mathrm{E}\left(\frac{\partial^2 \bar{l}}{\partial \theta_{0i} \partial \theta_{0j}} \right) = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0i}} \boldsymbol{\Sigma}_0^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_{0j}} \right) \quad (i, j = 1, ..., q).$$ The

above results yield (2.29).

A.5 Proof of Corollary 1

First, we derive $n\mathrm{E}\left(\dfrac{\partial \bar{l}}{\partial \boldsymbol{\theta}_0'} \otimes \mathbf{M}\right)$. From (A4.3) with $\dfrac{\partial^2 \boldsymbol{\Sigma}}{\partial \theta_{0b} \partial \theta_{0c}} = 0$ $(b,c = 1,...,p^*)$, it follows that

$$
\begin{aligned}
n\mathrm{E}\left\{\frac{\partial \bar{l}}{\partial \sigma_{0ab}}(\mathbf{M})_{cd,ef}\right\} &= -\frac{1}{8}(2-\delta_{ab})(2-\delta_{cd})(2-\delta_{ef}) \\
&\quad \times \mathrm{tr}\{\boldsymbol{\Sigma}_0^{-1}(\mathbf{E}_{ab}+\mathbf{E}_{ba})\boldsymbol{\Sigma}_0^{-1}(\mathbf{E}_{cd}+\mathbf{E}_{dc})\boldsymbol{\Sigma}_0^{-1}(\mathbf{E}_{ef}+\mathbf{E}_{fe})\} \\
&= -\frac{1}{8}(2-\delta_{ab})(2-\delta_{cd})(2-\delta_{ef})\sum_{(cd,ef)}^4 \sigma_0^{ce}(\sigma_0^{ad}\sigma_0^{bf}+\sigma^{af}\sigma_0^{bd})
\end{aligned}
$$

$$(A5.1)$$

$$(p \ge a \ge b \ge 1;\ p \ge c \ge d \ge 1;\ p \ge e \ge f \ge 1).$$

Using (A5.1), $\mathbf{k}$ is written as

$$
\begin{aligned}
\mathbf{k} = \mathrm{diag}^{-1}(\boldsymbol{\theta}_0)\sum_{c \ge d}(\mathbf{I}_0^{-1})_{\cdot cd}\Bigg[\sum_{a \ge b}\sum_{e \ge f}\Bigg\{-\frac{1}{8}(2-\delta_{ab})(2-\delta_{cd})(2-\delta_{ef}) \\
\times \sum_{(cd,ef)}^4 \sigma_0^{ce}(\sigma_0^{ad}\sigma_0^{bf}+\sigma^{af}\sigma_0^{bd})\Bigg\}(\mathbf{I}_0^{-1})_{ab,ef}+\frac{1}{\sigma_{0cd}}\Bigg].
\end{aligned}
$$

$$(A5.2)$$

The (gh)-th element of $\mathbf{k}$ in (A5.2) using the double subscript notation is

$$
\begin{aligned}
(\mathbf{k})_{(gh)} &= \frac{1}{\sigma_{0gh}}\sum_{c \ge d}(\sigma_{0gc}\sigma_{0hd}+\sigma_{0gd}\sigma_{0hc}) \\
&\quad \times \sum_{a \ge b}\sum_{e \ge f}\Bigg\{-\frac{1}{8}(2-\delta_{ab})(2-\delta_{cd})(2-\delta_{ef}) \\
&\quad \times \sum_{(cd,ef)}^4 \sigma_0^{ce}(\sigma_0^{ad}\sigma_0^{bf}+\sigma^{af}\sigma_0^{bd})\Bigg\}(\sigma_{0ae}\sigma_{0bf}+\sigma_{0af}\sigma_{0be}) \quad (A5.3) \\
&\quad +\frac{1}{\sigma_{0gh}}\sum_{c \ge d}(\sigma_{0gc}\sigma_{0hd}+\sigma_{0gd}\sigma_{0hc})\frac{1}{\sigma_{0cd}}
\end{aligned}
$$

$$= -\frac{1}{8\sigma_{0gh}} \sum_{c,d=1}^{p} (\sigma_{0gc}\sigma_{0hd} + \sigma_{0gd}\sigma_{0hc}) \sum_{(cd,ef)}^{4} (\sigma_0^{cd} p + \sigma_0^{cd} + \sigma_0^{cd} p + \sigma_0^{cd})$$

$$+ \frac{1}{\sigma_{0gh}} \sum_{c \geq d} (\sigma_{0gc}\sigma_{0hd} + \sigma_{0gd}\sigma_{0hc}) \frac{1}{\sigma_{0cd}}$$

$$= -2(p+1) + \frac{1}{\sigma_{0gh}} \sum_{c \geq d} (\sigma_{0gc}\sigma_{0hd} + \sigma_{0gd}\sigma_{0hc}) \frac{1}{\sigma_{0cd}} \quad (p \geq g \geq h \geq 1).$$

A.6 Proof of Theorem 5

Let $E_g^{(\mathbf{S})}(\cdot)$ denote the expectation using the distribution of $\mathbf{S}$ based on the density $g(\mathbf{S}|\cdot)$, which is possibly different from the Wishart density $f(\mathbf{S}|\mathbf{\Sigma})$. When obvious a simple notation $E_g(\cdot)$ is also used. Even under non-normality, the following result holds:

$$-2\hat{\bar{l}}^* = -2E_g^{(\mathbf{T})}\{\bar{l}(\hat{\boldsymbol{\theta}}|\mathbf{T})\} = E_g^{(\mathbf{T})}\{\ln|\hat{\mathbf{\Sigma}}| + tr(\hat{\mathbf{\Sigma}}^{-1}\mathbf{T}) + C(\mathbf{T})\}$$

$$= \ln|\hat{\mathbf{\Sigma}}| + tr(\hat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}_0) + E_g^{(\mathbf{T})}\{C(\mathbf{T})\}. \tag{A6.1}$$

When $\hat{\mathbf{\Sigma}} = k\mathbf{S}$, (A6.1) gives

$$E_g^{(\mathbf{S})}(-2\hat{\bar{l}}^*) = p\ln k + E_g(\ln|\mathbf{S}|) + k^{-1}E_g\{tr(\mathbf{S}^{-1}\mathbf{\Sigma}_0)\} + E_g\{C(\mathbf{S})\}. \tag{A6.2}$$

In (A6.2), $E_g(\ln|\mathbf{S}|)$ under arbitrary distributions is given up to order $O(n^{-1})$ using the Taylor series expansion as follows:

$$E_g(\ln|\mathbf{S}|) = \ln|\mathbf{\Sigma}_0| + \frac{n^{-1}}{2}\frac{\partial^2 \ln|\mathbf{S}|}{(\partial \mathbf{s}')^{<2>}}\bigg|_{\mathbf{s}=\boldsymbol{\sigma}_0} nE_g\{(\mathbf{s}-\boldsymbol{\sigma}_0)^{<2>}\}_{\to O(n^{-1})}$$

$$+ O(n^{-2}), \tag{A6.3}$$

where $\mathbf{x}^{<i>} = \mathbf{x} \otimes \cdots \otimes \mathbf{x}$ (i times of $\mathbf{x}$);

$$\frac{\partial^2 \ln|\mathbf{S}|}{\partial s_{ab}\partial s_{cd}}\bigg|_{\mathbf{s}=\boldsymbol{\sigma}_0} = \frac{2-\delta_{ab}}{2}\frac{\partial tr\{\mathbf{S}^{-1}(\mathbf{E}_{ab}+\mathbf{E}_{ba})\}}{\partial s_{cd}}\bigg|_{\mathbf{s}=\boldsymbol{\sigma}_0}$$

$$= -\frac{1}{2}(2-\delta_{ab})(2-\delta_{cd})(\sigma_0^{ac}\sigma_0^{db} + \sigma_0^{ad}\sigma_0^{cb}) = -2(\mathbf{\Omega}_{NT}^{-1})_{ab,cd} \; ; \; s_{ab} = (\mathbf{S})_{ab} \tag{A6.4}$$

$$(p \geq a \geq b \geq 1; \; p \geq c \geq d \geq 1);$$

and $E_g\{\cdot\}_{\to O(n^{-1})}$ indicates that the expectation is taken up to order $O(n^{-1})$.

Let $\boldsymbol{\Omega} = n E_g\{(\mathbf{s} - \boldsymbol{\sigma}_0)(\mathbf{s} - \boldsymbol{\sigma}_0)'\}_{\to O(n^{-1})} = \boldsymbol{\Omega}_{\mathrm{NT}} + \mathbf{K}_{(4)}$. Then, (A6.3) becomes

$$E_g(\ln|\mathbf{S}|) = \ln|\boldsymbol{\Sigma}_0| - n^{-1}\mathrm{tr}\{\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}(\boldsymbol{\Omega}_{\mathrm{NT}} + \mathbf{K}_{(4)})\} + O(n^{-2})$$

$$= \ln|\boldsymbol{\Sigma}_0| - n^{-1}\{p^* + \mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\} + O(n^{-2}). \qquad \text{(A6.5)}$$

On the other hand, $E_g\{\mathrm{tr}(\mathbf{S}^{-1}\boldsymbol{\Sigma}_0)\}$ in (A6.2) is

$$E_g\{\mathrm{tr}(\mathbf{S}^{-1}\boldsymbol{\Sigma}_0)\} = p + \frac{n^{-1}}{2}\frac{\partial^2\mathrm{tr}(\mathbf{S}^{-1}\boldsymbol{\Sigma}_0)}{(\partial\mathbf{s}')^{<2>}}|_{\mathbf{s}=\boldsymbol{\sigma}_0} n E_g\{(\mathbf{s}-\boldsymbol{\sigma}_0)^{<2>}\}_{\to O(n^{-1})}$$

$$+ O(n^{-2}), \qquad \text{(A6.6)}$$

where

$$\frac{\partial^2\mathrm{tr}(\mathbf{S}^{-1}\boldsymbol{\Sigma}_0)}{\partial s_{ab}\partial s_{cd}}|_{\mathbf{s}=\boldsymbol{\sigma}_0}$$

$$= \frac{1}{2}(2-\delta_{ab})(2-\delta_{cd})\mathrm{tr}\{\boldsymbol{\Sigma}_0^{-1}(\mathbf{E}_{ab}+\mathbf{E}_{ba})\boldsymbol{\Sigma}_0^{-1}(\mathbf{E}_{cd}+\mathbf{E}_{dc}))\}$$

$$= (2-\delta_{ab})(2-\delta_{cd})(\sigma_0^{ac}\sigma_0^{db} + \sigma_0^{ad}\sigma_0^{cb}) \qquad \text{(A6.7)}$$

$$(p \geq a \geq b \geq 1; \ p \geq c \geq d \geq 1).$$

Consequently, we obtain

$$E_g\{\mathrm{tr}(\mathbf{S}^{-1}\boldsymbol{\Sigma}_0)\} = p + n^{-1}2\,\mathrm{tr}\{\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}(\boldsymbol{\Omega}_{\mathrm{NT}} + \mathbf{K}_{(4)})\} + O(n^{-2})$$

$$= p + n^{-1}2\{p^* + \mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\} + O(n^{-2}). \qquad \text{(A6.8)}$$

From (A6.5) and (A6.8), (A6.2) becomes

$$E_g^{(\mathbf{S})}(-2\hat{\bar{l}}^*) = p\ln k + \ln|\boldsymbol{\Sigma}_0| - n^{-1}\{p^* + \mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\}$$

$$+ k^{-1}[p + n^{-1}2\{p^* + \mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\}] + E_g\{C(\mathbf{S})\} + O(n^{-2}). \qquad \text{(A6.9)}$$

Differentiating (A6.9) with respect to k, we obtain

$$k = 1 + n^{-1}\{p + 1 + 2p^{-1}\mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\}. \qquad \text{(A6.10)}$$

Note that for (A6.10), $E_g(\ln|\mathbf{S}|)$ is unnecessary as $E_g\{C(\mathbf{S})\}$ though shown here for later use.

A.7 Proof of Theorem 6

When $k = 1 + n^{-1}\{p + 1 + 2p^{-1}\mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}} + \mathbf{K}_{(4)})\}$, it follows that

$$
\mathrm{E}_g^{(\mathbf{S})}(-2\hat{\bar{l}}^* + 2\hat{\bar{l}}_{\mathrm{WML}}^*) = p\ln k + (k^{-1} - 1)\mathrm{E}_g\{\mathrm{tr}(\mathbf{S}^{-1}\boldsymbol{\Sigma}_0)\}
$$

$$
= p\left[\ n^{-1}\{p + 1 + 2p^{-1}\mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\}\right.
$$

$$
\left. -\frac{n^{-2}}{2}\{p + 1 + 2p^{-1}\mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\}^2 + O(n^{-3})\right]
$$

$$
+\left[\ -n^{-1}\{p + 1 + 2p^{-1}\mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\}\right.
$$

$$
\left. +n^{-2}\{p + 1 + 2p^{-1}\mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\}^2 + O(n^{-3})\ \right]
$$

$$
\times[p + n^{-1}2\{p^* + \mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\} + O(n^{-2})]
$$

$$
= -\frac{n^{-2}}{2}p\{p + 1 + 2p^{-1}\mathrm{tr}(\boldsymbol{\Omega}_{\mathrm{NT}}^{-1}\mathbf{K}_{(4)})\}^2 + O(n^{-3}),
$$

(A7.1)

which gives (4.8). The result of (4.9) under normality is immediately given from (A7.1) with $\mathbf{K}_{(4)} = \mathbf{O}$. When $k = n/(n - p - 1)$, (A7.1) under normality becomes

$$
\mathrm{E}_f^{(\mathbf{S})}(-2\hat{\bar{l}}^* + 2\hat{\bar{l}}_{\mathrm{WML}}^*) = p\ln k + (k^{-1} - 1)\mathrm{E}_f\{\mathrm{tr}(\mathbf{S}^{-1}\boldsymbol{\Sigma}_0)\}
$$

$$
= p\ln\frac{n}{n - p - 1} + \left(\frac{n - p - 1}{n} - 1\right)\frac{np}{n - p - 1}
$$

(A7.2)

$$
= p\ln\frac{n}{n - p - 1} - \frac{p(p + 1)}{n - p - 1},
$$

which gives (4.10).

A.8 $\mathrm{E}_g^{(\mathbf{S})}(-2\hat{\bar{l}}^* + 2\bar{l}_0^*)$ when $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\mathrm{WML}}$

Since

$$\text{E}_g^{(\mathbf{S})}(-2\hat{\bar{l}}^* + 2\bar{l}_0^*)$$

$$= \text{E}_g(\ln|\hat{\mathbf{\Sigma}}_{\text{WML}}|) - \ln|\mathbf{\Sigma}_0| + \text{E}_g\{\text{tr}(\hat{\mathbf{\Sigma}}_{\text{WML}}^{-1}\mathbf{\Sigma}_0)\} - p, \tag{A8.1}$$

the two expectations on the right-hand side of (A8.1) up to order $O(n^{-1})$ are given as follows. Firstly,

$$\text{E}_g(\ln|\hat{\mathbf{\Sigma}}_{\text{WML}}|)$$

$$= \ln|\mathbf{\Sigma}_0| + \frac{n^{-1}}{2}\frac{\partial^2 \ln|\hat{\mathbf{\Sigma}}_{\text{WML}}|}{(\partial\mathbf{s}')^{<2>}}|_{\mathbf{s}=\mathbf{\sigma}_0}n\text{E}_g\{(\mathbf{s}-\mathbf{\sigma}_0)^{<2>}\}_{\to O(n^{-1})} + O(n^{-2}), \tag{A8.2}$$

where

$$\frac{\partial \ln|\hat{\mathbf{\Sigma}}_{\text{WML}}|}{\partial s_{ab}} = \sum_{i=1}^{q}\text{tr}\left\{\hat{\mathbf{\Sigma}}_{\text{WML}}^{-1}\frac{\partial\hat{\mathbf{\Sigma}}_{\text{WML}}}{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_i}\right\}\frac{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_i}{\partial s_{ab}}, \tag{A8.3}$$

$$\frac{\partial^2 \ln|\hat{\mathbf{\Sigma}}_{\text{WML}}|}{\partial s_{ab}\partial s_{cd}} = \sum_{i=1}^{q}\left[\text{tr}\left\{\hat{\mathbf{\Sigma}}_{\text{WML}}^{-1}\frac{\partial\hat{\mathbf{\Sigma}}_{\text{WML}}}{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_i}\right\}\frac{\partial^2(\hat{\mathbf{\theta}}_{\text{WML}})_i}{\partial s_{ab}\partial s_{cd}}\right.$$

$$+ \sum_{j=1}^{q}\text{tr}\left\{\hat{\mathbf{\Sigma}}_{\text{WML}}^{-1}\frac{\partial^2\hat{\mathbf{\Sigma}}_{\text{WML}}}{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_i\partial(\hat{\mathbf{\theta}}_{\text{WML}})_j} - \hat{\mathbf{\Sigma}}_{\text{WML}}^{-1}\frac{\partial\hat{\mathbf{\Sigma}}_{\text{WML}}}{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_i}\hat{\mathbf{\Sigma}}_{\text{WML}}^{-1}\frac{\partial\hat{\mathbf{\Sigma}}_{\text{WML}}}{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_j}\right\}$$

$$\left.\times\frac{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_i}{\partial s_{ab}}\frac{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_j}{\partial s_{ab}}\right] \quad (p \geq a \geq b \geq 1; \ p \geq c \geq d \geq 1).$$

In (A8.3), $\dfrac{\partial(\hat{\mathbf{\theta}}_{\text{WML}})_i}{\partial s_{ab}}$ and

$\dfrac{\partial^2(\hat{\mathbf{\theta}}_{\text{WML}})_i}{\partial s_{ab}\partial s_{cd}}$ $(i = 1,...,q; \ p \geq a \geq b \geq 1; \ p \geq c \geq d \geq 1)$ are given by the

formulas for partial derivatives in implicit functions (Ogasawara, 2009, Equation (3.16)).

Secondly,

$$\text{E}_g\{\text{tr}(\hat{\mathbf{\Sigma}}_{\text{WML}}^{-1}\mathbf{\Sigma}_0)\}$$

$$= p + \frac{n^{-1}}{2}\frac{\partial^2\{\text{tr}(\hat{\mathbf{\Sigma}}_{\text{WML}}^{-1}\mathbf{\Sigma}_0)\}}{(\partial\mathbf{s}')^{<2>}}|_{\mathbf{s}=\mathbf{\sigma}_0}n\text{E}_g\{(\mathbf{s}-\mathbf{\sigma}_0)^{<2>}\}_{\to O(n^{-1})} + O(n^{-2}), \tag{A8.4}$$

where

$$\frac{\partial^2 \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \boldsymbol{\Sigma}_0)}{\partial s_{ab}} = \sum_{i=1}^{q} \left[-\mathrm{tr}\left\{ \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}}{\partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_i} \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \boldsymbol{\Sigma}_0 \right\} \frac{\partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_i}{\partial s_{ab}} \right],$$

$$\frac{\partial^2 \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \boldsymbol{\Sigma}_0)}{\partial s_{ab} \partial s_{cd}} = \sum_{i=1}^{q} \left[-\mathrm{tr}\left\{ \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}}{\partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_i} \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \boldsymbol{\Sigma}_0 \right\} \frac{\partial^2 (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_i}{\partial s_{ab} \partial s_{cd}} \right.$$

$$+ \sum_{j=1}^{q} \mathrm{tr}\left\{ -\hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \frac{\partial^2 \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}}{\partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_i \partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_j} \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \boldsymbol{\Sigma}_0 \right.$$

$$\left. \left. + 2\hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}}{\partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_i} \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \frac{\partial \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}}{\partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_j} \hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \boldsymbol{\Sigma}_0 \right\} \frac{\partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_i}{\partial s_{ab}} \frac{\partial (\hat{\boldsymbol{\theta}}_{\mathrm{WML}})_j}{\partial s_{ab}} \right] \quad \text{(A8.5)}$$

$(p \geq a \geq b \geq 1;\ p \geq c \geq d \geq 1)$.

From (A8.2) with (A8.3) and (A8.4) with (A8.5), (A8.1) is written as

$$\mathrm{E}_g^{(\mathbf{S})}(-2\hat{\overline{l}}^* + 2\overline{l}_0^*)$$

$$= \frac{n^{-1}}{2} \mathrm{tr}\left\{ \left(\frac{\partial^2 \ln |\hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}|}{\partial \mathbf{s}\, \partial \mathbf{s}'} \Big|_{\mathbf{s}=\boldsymbol{\sigma}_0} + \frac{\partial^2 \mathrm{tr}(\hat{\boldsymbol{\Sigma}}_{\mathrm{WML}}^{-1} \boldsymbol{\Sigma}_0)}{\partial \mathbf{s}\, \partial \mathbf{s}'} \Big|_{\mathbf{s}=\boldsymbol{\sigma}_0} \right) \boldsymbol{\Omega} \right\} + O(n^{-2}). \quad \text{(A8.6)}$$

A.9 Proof of Theorem 7

The result of (4.13) is given by

$$-2\hat{\overline{l}}_{\mathrm{WML}}^* = -2\overline{l}_0^* - \frac{\partial^2 \overline{l}^*}{(\partial \boldsymbol{\theta}_0{}')^{<2>}} (\hat{\boldsymbol{\theta}}_{\mathrm{WML}} - \boldsymbol{\theta}_0)^{<2>} + O_p(n^{-3/2}), \quad \text{(A9.1)}$$

where $\overline{l}^* = \mathrm{E}_f^{(\mathbf{T})}\{\overline{l}(\boldsymbol{\theta}\,|\,\mathbf{T})\}$ and

$\partial^2 \overline{l}^* / (\partial \boldsymbol{\theta}_0{}')^{<2>} = \partial^2 \overline{l}^* / (\partial \boldsymbol{\theta}')^{<2>}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -\mathrm{vec}\,'(\mathbf{I}_0)$. On the other hand,

$\mathrm{E}_f^{(\mathbf{S})}(\hat{\boldsymbol{\theta}}_{\mathrm{WML}} - \boldsymbol{\theta}_0)^{<2>} = n^{-1}\mathrm{vec}(\mathbf{I}_0^{-1}) + O(n^{-2})$. These results give

$\mathrm{E}_f^{(\mathbf{S})}(-2\hat{\overline{l}}_{\mathrm{WML}}^*) = -2\overline{l}_0^* + n^{-1}\mathrm{vec}\,'(\mathbf{I}_0)\mathrm{vec}(\mathbf{I}_0^{-1}) + O(n^{-2}) = -2\overline{l}_0^* + n^{-1}q + O(n^{-2})$.

The unchanged result up to order $O(n^{-1})$ of (4.14) is given by a special case of Ogasawara (2017, Equation (3.9)).

A.10 Proof of Theorem 8

14

Since the second term on the right-hand side of (6.2) is irrelevant to k_{ij}, the expectation of the first term, i.e., $\mathrm{E}_g^{(\mathbf{S})}\left\{\sum_{i\geq j} w_{ij}(k_{ij}s_{ij}-\sigma_{0ij})^2\right\}$ is minimized. Noting that

$$\mathrm{E}_g^{(\mathbf{S})}\left\{\sum_{i\geq j} w_{ij}(k_{ij}s_{ij}-\sigma_{0ij})^2\right\}$$

$$= \mathrm{E}_g^{(\mathbf{S})}\left\{\sum_{i\geq j} w_{ij}k_{ij}^2(s_{ij}-\sigma_{0ij})^2\right\}+\sum_{i\geq j} w_{ij}(k_{ij}-1)^2\sigma_{0ij}^2$$

$$= \sum_{i\geq j} w_{ij}\,\mathrm{var}(s_{ij})k_{ij}^2+\sum_{i\geq j} w_{ij}(k_{ij}-1)^2\sigma_{0ij}^2,$$

(A10.1)

the minimizing constant k_{ij} is given by

$$k_{ij}=\frac{\sigma_{0ij}^2}{\mathrm{var}(s_{ij})+\sigma_{0ij}^2}=\frac{1}{1+c_{\mathrm{V}}^2(s_{ij})}\quad(p\geq i\geq j\geq 1),$$

(A10.2)

which gives (6.3) with (6.4).

A.11 Proof of Theorem 9

From (6.18), we obtain

$$F\equiv \mathrm{E}^{(\mathbf{X})}(-2\hat{\overline{l}}^{*})=p\ln(2\pi)+\mathrm{E}^{(\mathbf{X})}(\ln|\hat{\boldsymbol{\Sigma}}|)$$
$$+\mathrm{tr}\left[\mathrm{E}^{(\mathbf{X})}(\hat{\boldsymbol{\Sigma}}^{-1})[\boldsymbol{\Sigma}_0+\mathrm{E}^{(\mathbf{X})}\{(\mathrm{diag}(\mathbf{k})(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)+\mathrm{diag}(\mathbf{k})\boldsymbol{\mu}_0-\boldsymbol{\mu}_0)\right.$$
$$\times(\mathrm{diag}(\mathbf{k})(\overline{\mathbf{x}}-\boldsymbol{\mu}_0)+\mathrm{diag}(\mathbf{k})\boldsymbol{\mu}_0-\boldsymbol{\mu}_0)'\}]\,]$$

(A11.1)

$$= p\ln(2\pi)+\mathrm{E}^{(\mathbf{X})}(\ln|\hat{\boldsymbol{\Sigma}}|)+\mathrm{tr}\left[\mathrm{E}^{(\mathbf{X})}(\hat{\boldsymbol{\Sigma}}^{-1})\right.$$
$$\times[\boldsymbol{\Sigma}_0+N^{-1}\mathrm{diag}(\mathbf{k})\boldsymbol{\Sigma}_0\mathrm{diag}(\mathbf{k})+\{\mathrm{diag}(\mathbf{k})\boldsymbol{\mu}_0-\boldsymbol{\mu}_0\}\{\mathrm{diag}(\mathbf{k})\boldsymbol{\mu}_0-\boldsymbol{\mu}_0\}']\,],$$

where the independence of $\hat{\boldsymbol{\Sigma}}$ and $\overline{\mathbf{x}}$ under normality is used. Since when $\hat{\boldsymbol{\Sigma}}=c_N^{-1}\sum_{i=1}^N(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})'$, we have

$$\mathrm{E}(\hat{\boldsymbol{\Sigma}}^{-1})=\frac{c_N}{n-p-1}\boldsymbol{\Sigma}_0^{-1}=\frac{c_N}{N-p-2}\boldsymbol{\Sigma}_0^{-1}$$

(A11.2)

(see Subsection A.1), (A11.1) becomes

15

$$F \equiv p\ln(2\pi) + \mathrm{E}^{(\mathbf{X})}(\ln|\hat{\boldsymbol{\Sigma}}|)$$

$$+ \frac{c_N}{N-p-2}[p + N^{-1}\mathrm{tr}\{\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}(\mathbf{k})\boldsymbol{\Sigma}_0\mathrm{diag}(\mathbf{k})\}$$

$$+ \{\mathrm{diag}(\mathbf{k})\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0\}'\boldsymbol{\Sigma}_0^{-1}\{\mathrm{diag}(\mathbf{k})\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0\}]$$

$$= p\ln(2\pi) + \mathrm{E}^{(\mathbf{X})}(\ln|\hat{\boldsymbol{\Sigma}}|)$$

$$+ \frac{c_N}{N-p-2}\{p + N^{-1}\sum_{i,j=1}^{p}\sigma_{0ij}\sigma_0^{ij}k_ik_j$$

$$+ \mathbf{k}'\mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}(\boldsymbol{\mu}_0)\mathbf{k} - 2\boldsymbol{\mu}_0'\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}(\boldsymbol{\mu}_0)\mathbf{k} + \boldsymbol{\mu}_0'\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\}.$$

(A11.3)

Differentiating (A11.3) with respect to $\mathbf{k}$ and setting the result to a zero vector, we have

$$\frac{N-p-2}{c_N}\frac{\partial F}{\partial \mathbf{k}} = 2N^{-1}\mathrm{Diag}\{\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}(\mathbf{k})\boldsymbol{\Sigma}_0\}\mathbf{1}_{(p)}$$

$$+ 2\mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}(\boldsymbol{\mu}_0)\mathbf{k} - 2\mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 = \mathbf{0}.$$

(A11.4)

Since the left-hand side of the last equation of (A11.4) is

$$2\left\{N^{-1}\begin{pmatrix} \sigma_{011}\sigma_0^{11} & \sigma_{021}\sigma_0^{21} & \cdots & \sigma_{01p}\sigma_0^{1p} \\ \sigma_{021}\sigma_0^{21} & \sigma_{022}\sigma_0^{22} & \cdots & \sigma_{02p}\sigma_0^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{0p1}\sigma_0^{p1} & \sigma_{0p2}\sigma_0^{p2} & \cdots & \sigma_{0pp}\sigma_0^{pp} \end{pmatrix} + \mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}(\boldsymbol{\mu}_0)\right\}\mathbf{k}$$

$$- 2\mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0$$

(A11.5)

$$= 2\{N^{-1}\boldsymbol{\Sigma}_0 \odot \boldsymbol{\Sigma}_0^{-1} + \mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}(\boldsymbol{\mu}_0)\}\mathbf{k} - 2\mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0,$$

we obtain

$$\mathbf{k} = \{N^{-1}\boldsymbol{\Sigma}_0 \odot \boldsymbol{\Sigma}_0^{-1} + \mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\mathrm{diag}(\boldsymbol{\mu}_0)\}^{-1}\mathrm{diag}(\boldsymbol{\mu}_0)\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0.$$ (A11.6)

A.12 Proof of Corollary 3

Equation (6.23) becomes

$$\mathrm{PGMSE} = \mathrm{E}^{(\mathbf{X})}\{(\hat{\boldsymbol{\mu}}_\mathbf{k}^{(\mathrm{P})} - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\hat{\boldsymbol{\mu}}_\mathbf{k}^{(\mathrm{P})} - \boldsymbol{\mu}_0)\} + N^{-1}p$$

$$\equiv \mathrm{GMSE} + N^{-1}p,$$

(A12.1)

which shows that minimizing PGMSE is equivalent to minimizing GMSE.

From $\hat{\boldsymbol{\mu}}_\mathbf{k}^{(\mathrm{P})} = \mathrm{diag}(\mathbf{k})\bar{\mathbf{x}}$, GMSE is

$$\text{GMSE} = \text{E}^{(\mathbf{X})}[\{\text{diag}(\mathbf{k})\overline{\mathbf{x}} - \boldsymbol{\mu}_0\}'\boldsymbol{\Sigma}_0^{-1}\{\text{diag}(\mathbf{k})\overline{\mathbf{x}} - \boldsymbol{\mu}_0\}]$$

$$= N^{-1}\text{tr}\{\text{diag}(\mathbf{k})\boldsymbol{\Sigma}_0^{-1}\text{diag}(\mathbf{k})\boldsymbol{\Sigma}_0\}$$

$$+ \{\text{diag}(\mathbf{k})\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0\}'\boldsymbol{\Sigma}_0^{-1}\{\text{diag}(\mathbf{k})\boldsymbol{\mu}_0 - \boldsymbol{\mu}_0\}.$$

(A12.2)

Since F in (A11.3) is an affine transformation of (A12.2) and consequently (A12.1), the result of the corollary follows.

References

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). New York: Wiley.

Kaplan, E. L. (1952). Tensor notation and the sampling cumulants of k-statistics. *Biometrika, 39*, 319-323.

Ogasawara, H. (2009). Asymptotic expansions in mean and covariance structure analysis. *Journal of Multivariate Analysis, 100*, 902-919.

Ogasawara, H. (2017). Predictive estimation of a covariance matrix and its structural parameters. *Journal of the Japanese Society of Computational Statistics, 30,* 45-63.