Likelihood, first and second order derivatives in a LVM

Brice Ozenne

In this document, we show the expression of the likelihood, its first two derivatives, the information matrix, and the first derivative of the information matrix.

1 Likelihood

At the individual level, the measurement and structural models can be written:

$$\boldsymbol{Y}_i = \nu + \boldsymbol{\eta}_i\Lambda + \boldsymbol{X}_iK + \boldsymbol{\varepsilon}_i$$
$$\boldsymbol{\eta}_i = \alpha + \boldsymbol{\eta}_iB + \boldsymbol{X}_i\Gamma + \boldsymbol{\zeta}_i$$

with Σ_ϵ the variance-covariance matrix of the residuals $\boldsymbol{\varepsilon}_i$
Σ_ζ the variance-covariance matrix of the residuals $\boldsymbol{\zeta}_i$.

By combining the previous equations, we can get an expression for $\boldsymbol{Y}_i$ that does not depend on $\boldsymbol{\eta}_i$:

$$\boldsymbol{Y}_i = \nu + (\boldsymbol{\zeta}_i + \alpha + \boldsymbol{X}_i\Gamma)(I-B)^{-1}\Lambda + \boldsymbol{X}_iK + \boldsymbol{\varepsilon}_i$$

Since $\mathbb{V}ar[Ax] = A\mathbb{V}ar[x]A^\intercal$ we have $\mathbb{V}ar[xA] = A^\intercal\mathbb{V}ar[x]A$, we have the following expressions for the conditional mean and variance of $\boldsymbol{Y}_i$:

$$\boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i) = E[\boldsymbol{Y}_i|\boldsymbol{X}_i] = \nu + (\alpha + \boldsymbol{X}_i\Gamma)(1-B)^{-1}\Lambda + \boldsymbol{X}_iK$$
$$\Omega(\boldsymbol{\theta}) = Var[\boldsymbol{Y}_i|\boldsymbol{X}_i] = \Lambda^t(1-B)^{-t}\Sigma_\zeta(1-B)^{-1}\Lambda + \Sigma_\varepsilon$$

where $\boldsymbol{\theta}$ is the collection of all parameters. The log-likelihood can be written:

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{X}) &= \sum_{i=1}^n l(\boldsymbol{\theta}|\boldsymbol{Y}_i, \boldsymbol{X}_i) \\ &= \sum_{i=1}^n -\frac{p}{2}log(2\pi) - \frac{1}{2}log|\Omega(\boldsymbol{\theta})| - \frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i))\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i))^\intercal \end{aligned}$$

2 Partial derivative for the conditional mean and variance

In the following, we denote by $\delta_{\sigma\in\Sigma}$ the indicator matrix taking value 1 at the position of σ in the matrix Σ. For instance:

$$\Sigma = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{1,2} & \sigma_{2,2} & \sigma_{2,3} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_{3,3} \end{bmatrix} \qquad \delta_{\sigma_{1,2}\in\Sigma} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The same goes for $\delta_{\lambda\in\Lambda}$, $\delta_{b\in B}$, and $\delta_{\psi\in\Psi}$.

First order derivatives:

$$\begin{aligned}
\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \nu} &= 1 \\
\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial K} &= \boldsymbol{X}_i \\
\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \alpha} &= (1-B)^{-1}\Lambda \\
\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \Gamma} &= \boldsymbol{X}_i(1-B)^{-1}\Lambda \\
\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \lambda} &= (\alpha + \boldsymbol{X}_i\Gamma)(1-B)^{-1}\delta_{\lambda\in\Lambda} \\
\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial b} &= (\alpha + \boldsymbol{X}_i\Gamma)(1-B)^{-1}\delta_{b\in B}(1-B)^{-1}\Lambda
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \psi} &= \Lambda^t(1-B)^{-t}\delta_{\psi\in\Psi}(1-B)^{-1}\Lambda \\
\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \sigma} &= \delta_{\sigma\in\Sigma} \\
\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \lambda} &= \delta^t_{\lambda\in\Lambda}(1-B)^{-t}\Psi(1-B)^{-1}\Lambda + \Lambda^t(1-B)^{-t}\Psi(1-B)^{-1}\delta_{\lambda\in\Lambda} \\
\frac{\partial \Omega(\boldsymbol{\theta})}{\partial b} &= \Lambda^t(1-B)^{-t}\delta^t_{b\in B}(1-B)^{-t}\Psi(1-B)^{-1}\Lambda + \Lambda^t(1-B)^{-t}\Psi(1-B)^{-1}\delta_{b\in B}(1-B)^{-1}\Lambda
\end{aligned}$$

Second order derivatives:

$$
\begin{aligned}
\frac{\partial^2 \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \alpha \partial b} &= \delta_\alpha (1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \Lambda \\
\frac{\partial^2 \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \alpha \partial \lambda} &= \delta_\alpha (1-B)^{-1} \delta_{\lambda \in \Lambda} \\
\frac{\partial^2 \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \Gamma \partial b} &= \boldsymbol{X}_i (1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \Lambda \\
\frac{\partial^2 \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \Gamma \partial \lambda} &= \boldsymbol{X}_i (1-B)^{-1} \delta_{\lambda \in \Lambda} \\
\frac{\partial^2 \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial \lambda \partial b} &= (\alpha + \boldsymbol{X}_i \Gamma)(1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \delta_{\lambda \in \Lambda} \\
\frac{\partial^2 \boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{X}_i)}{\partial b \partial b'} &= (\alpha + \boldsymbol{X}_i \Gamma)(1-B)^{-1} \delta_{b' \in B} (1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \Lambda \\
&\quad + (\alpha + \boldsymbol{X}_i \Gamma)(1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \delta_{b' \in B} (1-B)^{-1} \Lambda
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \Omega(\boldsymbol{\theta})}{\partial \psi \partial \lambda} &= \delta^t_{\lambda \in \Lambda} (1-B)^{-t} \delta_{\psi \in \Psi} (1-B)^{-1} \Lambda \\
&\quad + \Lambda^t (1-B)^{-t} \delta_{\psi \in \Psi} (1-B)^{-1} \delta_{\lambda \in \Lambda} \\
\frac{\partial^2 \Omega(\boldsymbol{\theta})}{\partial \psi \partial b} &= \Lambda^t (1-B)^{-t} \delta^t_{b \in B} (1-B)^{-t} \delta_{\psi \in \Psi} (1-B)^{-1} \Lambda \\
&\quad + \Lambda^t (1-B)^{-t} \delta_{\psi \in \Psi} (1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \Lambda \\
\frac{\partial^2 \Omega(\boldsymbol{\theta})}{\partial \lambda \partial b} &= \delta^t_{\lambda \in \Lambda} (1-B)^{-t} \delta^t_{b \in B} (1-B)^{-t} \Psi (1-B)^{-1} \Lambda \\
&\quad + \delta^t_{\lambda \in \Lambda} (1-B)^{-t} \Psi (1-B)^{-1} \delta^t_{b \in B} (1-B)^{-1} \Lambda \\
&\quad + \Lambda^t (1-B)^{-t} \delta^t_{b \in B} (1-B)^{-t} \Psi (1-B)^{-1} \delta_{\lambda \in \Lambda} \\
&\quad + \Lambda^t (1-B)^{-t} \Psi (1-B)^{-1} \delta^t_{b \in B} (1-B)^{-1} \delta_{\lambda \in \Lambda} \\
\frac{\partial^2 \Omega(\boldsymbol{\theta})}{\partial \lambda \partial \lambda'} &= \delta^t_{\lambda \in \Lambda} (1-B)^{-t} \Psi (1-B)^{-1} \delta_{\lambda' \in \Lambda} \\
&\quad + \delta^t_{\lambda' \in \Lambda} (1-B)^{-t} \Psi (1-B)^{-1} \delta_{\lambda \in \Lambda} \\
\frac{\partial^2 \Omega(\boldsymbol{\theta})}{\partial b \partial b'} &= \Lambda^t (1-B)^{-t} \delta^t_{b' \in B} (1-B)^{-t} \delta^t_{b \in B} (1-B)^{-t} \Psi (1-B)^{-1} \Lambda \\
&\quad + \Lambda^t (1-B)^{-t} \delta^t_{b \in B} (1-B)^{-t} \delta^t_{b' \in B} (1-B)^{-t} \Psi (1-B)^{-1} \Lambda \\
&\quad + \Lambda^t (1-B)^{-t} \delta^t_{b \in B} (1-B)^{-t} \Psi (1-B)^{-1} \delta_{b' \in B} (1-B)^{-1} \Lambda \\
&\quad + \Lambda^t (1-B)^{-t} \delta^t_{b' \in B} (1-B)^{-t} \Psi (1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \Lambda \\
&\quad + \Lambda^t (1-B)^{-t} \Psi (1-B)^{-1} \delta_{b' \in B} (1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \Lambda \\
&\quad + \Lambda^t (1-B)^{-t} \Psi (1-B)^{-1} \delta_{b \in B} (1-B)^{-1} \delta_{b' \in B} (1-B)^{-1} \Lambda
\end{aligned}
$$

3 First derivative: score

The individual score is obtained by derivating the log-likelihood:

$$\begin{aligned}\mathcal{U}(\theta|\boldsymbol{Y}_i,\boldsymbol{X}_i) =& \frac{\partial l_i(\boldsymbol{\theta}|\boldsymbol{Y}_i,\boldsymbol{X}_i)}{\partial\theta}\\ =& -\frac{1}{2}tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\right)\\ &+\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\\ &+\frac{1}{2}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\end{aligned}$$

4 Second derivative: Hessian and expected information

The individual Hessian is obtained by derivating twice the log-likelihood:

$$\begin{aligned}\mathcal{H}_i(\theta,\theta') =& -\frac{1}{2}tr\left(-\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}+\Omega(\boldsymbol{\theta})^{-1}\frac{\partial^2\Omega(\boldsymbol{\theta})}{\partial\theta\partial\theta'}\right)\\ &+\frac{\partial^2\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\\ &-\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\\ &-\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta'}^\intercal\\ &-\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\\ &-(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\\ &+\frac{1}{2}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))\Omega(\boldsymbol{\theta})^{-1}\frac{\partial^2\Omega(\boldsymbol{\theta})}{\partial\theta\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i-\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\end{aligned}$$

Using that $\boldsymbol{\mu}(\theta, \boldsymbol{X}_i)$ and $\Omega(\boldsymbol{\theta})$ are deterministic quantities, we can then take the expectation to obtain:

$$\begin{aligned}
\mathbb{E}\left[\mathcal{H}_i(\theta,\theta')\right] = &-\frac{1}{2}tr\left(-\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \Omega(\theta)^{-1}\frac{\partial^2\Omega(\boldsymbol{\theta})}{\partial\theta\partial\theta'}\right)\\
&+ \frac{\partial^2\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\,\mathbb{E}\left[(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\right]\\
&- \frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\,\mathbb{E}\left[(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\right]\\
&- \frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta})}{\partial\theta'}^\intercal\\
&- \frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}\,\mathbb{E}\left[(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\right]\\
&- \mathbb{E}\left[(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\right]\\
&+ \mathbb{E}\left[\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))\Omega(\boldsymbol{\theta})^{-1}\frac{\partial^2\Omega(\boldsymbol{\theta})}{\partial\theta\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i))^\intercal\right]
\end{aligned}$$

The last two expectations can be re-written using that $\mathbb{E}[x^\intercal Ax] = tr\left(A\mathbb{V}ar[x]\right) + \mathbb{E}[x]^\intercal A\mathbb{E}[x]$:

$$\begin{aligned}
\mathbb{E}\left[\mathcal{H}_i(\theta,\theta')\right] = &-\frac{1}{2}tr\left(-\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta} + \Omega(\boldsymbol{\theta})^{-1}\frac{\partial^2\Omega(\boldsymbol{\theta})}{\partial\theta\partial\theta'}\right)\\
&- \frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta'}^\intercal\\
&- tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}(\mathbb{V}ar\left[\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)\right])^\intercal\right)\\
&+ \frac{1}{2}tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial^2\Omega(\boldsymbol{\theta})}{\partial\theta\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}(\mathbb{V}ar\left[\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)\right])^\intercal\right)
\end{aligned}$$

where we have used that $\mathbb{V}ar\left[\boldsymbol{Y}_i - \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)\right] = \mathbb{V}ar\left[\boldsymbol{Y}_i|\boldsymbol{X}_i\right] = \Omega(\boldsymbol{\theta})$. Finally we get:

$$\begin{aligned}
\mathbb{E}\left[\mathcal{H}_i(\theta,\theta')\right] = &-\frac{1}{2}tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\right)\\
&- \frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta'}^\intercal
\end{aligned}$$

So we can deduce from the previous equation the expected information matrix:

$$\mathcal{I}(\theta,\theta') = \frac{n}{2}tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta'}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\Omega(\boldsymbol{\theta})}{\partial\theta}\right) + \sum_{i=1}^{n}\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial\theta'}^\intercal$$

5 First derivatives of the information matrix

$$
\begin{aligned}
\frac{\partial \mathcal{I}(\theta,\theta')}{\partial \theta''} =& -\frac{n}{2} tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta''}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta'}\right) \\
&+\frac{n}{2} tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial^2 \Omega(\boldsymbol{\theta})}{\partial \theta \partial \theta''}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta'}\right) \\
&-\frac{n}{2} tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta''}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta'}\right) \\
&+\frac{n}{2} tr\left(\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial^2 \Omega(\boldsymbol{\theta})}{\partial \theta' \partial \theta''}\right) \\
&+\sum_{i=1}^{n}\frac{\partial^2 \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial \theta \partial \theta''}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial \theta'}^{\intercal} \\
&+\sum_{i=1}^{n}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial \theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial^2 \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial \boldsymbol{\theta}' \partial \theta''}^{\intercal} \\
&-\sum_{i=1}^{n}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial \theta}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \Omega(\boldsymbol{\theta})}{\partial \theta''}\Omega(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta},\boldsymbol{X}_i)}{\partial \theta'}^{\intercal}
\end{aligned}
$$