Gulf Journal of Mathematics ISSN: 2309-4966
Vol 15, Issue 2 (2023) 117-132 https://doi.org/10.56947/gjom.v15i2.1603

# EXACT COMPUTATION OF THE ASYMPTOTIC VARIANCE MATRIX FOR A TUPLE OF MULTINOMIAL DISTRIBUTIONS WITH DEPENDENT PARAMETERS

ISSA CHERIF GERALDO

Abstract. In this paper, we consider a statistical model consisting of an $s-$tuple of multinomial distributions whose parameters are dependent. This model is used in road safety to study the effect of a measure implemented on a set of $s$ sites comprising a total of $r$ accident severity levels and is depending on a parameter vector of $1+sr$ components under $s$ equality constraints. The computation of the so important asymptotic variance matrix $\boldsymbol{W}$ requires some numerical matrix inversions which become very complicated for large values of $s$ and $r$. To circumvent this difficulty, we use Schur complement for the exact analytic inversion and we propose a strategy for calculating $\boldsymbol{W}$. We illustrate the effectiveness of our approach using simulated data.

## 1. Introduction

Let $\boldsymbol{\theta} \in \mathbb{R}^d$ $(d \in \mathbb{N}^*)$ be the parameter vector of a statistical model. In addition to the point estimate (whether numerical or in closed-form) of $\boldsymbol{\theta}$ from a given number of observations, statistical inference requires the calculation of the variance matrix (also called variance-covariance matrix). Indeed, any estimate of $\boldsymbol{\theta}$ obtained from observed data is actually nothing more than a realization of a certain random vector $\hat{\boldsymbol{\theta}}$ called estimator of $\boldsymbol{\theta}$. The variance matrix of $\hat{\boldsymbol{\theta}}$ is therefore of capital importance because it makes it possible to determine the probability distribution of $\hat{\boldsymbol{\theta}}$ and later to make confidence intervals and hypotheses tests.

In most cases, the estimation method used is that of maximum likelihood and only an asymptotic variance matrix can be obtained. This matrix is equal to the inverse of the Fisher information matrix (FIM) defined under some regularity conditions (see [11, Section 11.6.2]) by

$$\boldsymbol{J} = -\mathbb{E}_{\boldsymbol{\theta}}\left(\nabla^2 \ell(\boldsymbol{\theta})\right), \tag{1.1}$$

where $\mathbb{E}_{\boldsymbol{\theta}}$ denotes the expectation, $\ell(\boldsymbol{\theta})$ is the log-likelihood function and $\nabla^2\ell(\boldsymbol{\theta})$ is the Hessian matrix of $\ell(\boldsymbol{\theta})$. The square roots of the diagonal elements of $\boldsymbol{J}^{-1}$ are called standard errors and represent the respective standard deviations of the

*Date*: Received: Mar 8, 2023; Accepted: Aug 21, 2023.

* Corresponding author.

2010 *Mathematics Subject Classification.* Primary 62F12, 62F30; Secondary 62H10.

*Key words and phrases.* Variance matrix, maximum likelihood, parameter estimation, Schur complement, multinomial distribution.

components of $\hat{\boldsymbol{\theta}}$. In many problems, however, the analytical calculation of $\boldsymbol{J}$ is difficult or impossible. One may then resort to numerical approximations of $\boldsymbol{J}$ by Monte Carlo method [22], bootstrap method [3], jackknife and delta methods [19] or by exploiting some quantities obtained by running algorithms such as the Expectation-Maximization (EM) and Minorization-Maximization (MM) algorithms [9, 10].

When $\boldsymbol{\theta}$ is subject to a set of $s$ constraints collected in the form $h(\boldsymbol{\theta}) = \mathbf{0}$, where $h$ is a vector-valued function and $\mathbf{0}$ is a null vector of adequate dimension, the variance matrix of $\hat{\boldsymbol{\theta}}$ under these constraints is given by (see [1, Theorem 2, p. 824] or [12, p. 308])

$$\boldsymbol{W} = \boldsymbol{J}^{-1} - \boldsymbol{J}^{-1}\boldsymbol{H}(\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^\top\boldsymbol{J}^{-1}, \tag{1.2}$$

where

$$\boldsymbol{H} = \frac{\partial h(\boldsymbol{\theta})^\top}{\partial \boldsymbol{\theta}}$$

is the $d \times s$ matrix of partial derivatives of the vector-valued function $h(\boldsymbol{\theta})$. With the exception of some small values of $d$ ($d \leqslant 2$), the practical application of Formula (1.2) can only be done with the help of a calculation software. But two main problems may arise: (a) matrices $\boldsymbol{J}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$ may not be invertible, (b) numerical inversion of matrix $\boldsymbol{J}$ (a matrix of order $d \times d$) may be tricky for large values of $d$.

In this paper, we consider a statistical model used in road safety to study the effect of a measure implemented on a set of $s$ experimental sites paired each with a control site where the measure was not applied and such that all the accidents are classified by severity into $r$ categories. The parameter vector $\boldsymbol{\theta}$ has $1 + sr$ components and is subject to $s$ equality constraints. To circumvent the two main problems posed by the application of Formula (1.2), we use the Schur complement [18, 20] to prove the invertibility of the matrices $\boldsymbol{J}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$ involved in (1.2) and for the exact analytic computation of their respective inverses. Afterwards, we propose a strategy for calculating $\boldsymbol{W}$. The main advantage of the proposed approach is that it does not require any numerical matrix inversion computational procedure thus avoiding all the problems related to conditioning and dimension. The results obtained on simulated data prove the effectiveness of our approach.

The remainder of our paper is thus organized. In Section 2, we present the statistical model. In Section 3, we give the main contributions of the paper. We prove the invertibility of matrices $\boldsymbol{J}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$ afterwards we give the analytical expression of their respective inverses. In Section 4, we present the results of the comparison in R software [21] of the method proposed in this paper (consisting in computing $\boldsymbol{W}$ using the analytical formulas of the inverses of $\boldsymbol{J}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$) and the method of computing $\boldsymbol{W}$ using a numerical inversion procedure to invert $\boldsymbol{J}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$. Section 5 is devoted to concluding remarks.

## 2. Statistical model

The statistical model considered in this paper has been described by [14] in the context of crash data analysis. It allows to assess the mean effect of a road safety measure applied to $s$ experimental sites (also called treatment sites or treated sites) where accidents are classified by increasing severity into $r > 0$ accident types. For all $k = 1, \dots, s$, let $n_k$ be the total number of crashes recorded at the experimental site $k$,

$$\boldsymbol{X}_k = (X_{11k}, \dots, X_{1rk}, X_{21k}, \dots, X_{2rk})^\mathsf{T} \tag{2.1}$$

be a random vector with $2r$ components and

$$\boldsymbol{Z}_k = (z_{1k}, \dots, z_{rk})^\mathsf{T} \tag{2.2}$$

be a non-random vector with $r$ components where, for all $j = 1, \dots, r$, $X_{1jk}$ (respectively $X_{2jk}$), denotes the number of accidents of severity $j$ on the experimental site $k$ before (respectively after) the application of the measure and $z_{jk}$ is the ratio of the number of accidents of severity $j$ in the "after" period to the number of accidents of type $j$ in the "before" period at a site similar to the experimental site $k$ but where the measure was not applied. For all $k = 1, \dots, s$, it is assumed in [14] that

$$\boldsymbol{X}_k \rightsquigarrow \mathscr{M}(n_k, \boldsymbol{\pi}_k(\boldsymbol{\theta})), \tag{2.3}$$

where $\mathscr{M}$ denotes the multinomial distribution,

$$\boldsymbol{\pi}_k(\boldsymbol{\theta}) = (\pi_{11k}(\boldsymbol{\theta}), \dots, \pi_{1rk}(\boldsymbol{\theta}), \pi_{21k}(\boldsymbol{\theta}), \dots, \pi_{2rk}(\boldsymbol{\theta}))^\mathsf{T} \tag{2.4}$$

is a vector of class probabilities whose components (defined later in this section) depend on a common parameter vector

$$\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}_1^\mathsf{T}, \dots, \boldsymbol{\beta}_s^\mathsf{T})^\mathsf{T} \tag{2.5}$$

such that $\alpha > 0$ is the mean effect of the measure and for all $k = 1, \dots, s$, $\boldsymbol{\beta}_k = (\beta_{1k}, \dots, \beta_{rk})^\mathsf{T}$ is a vector of probabilities related to treated site $k$. The parameter vector $\boldsymbol{\theta}$ also satisfies the equation $h(\boldsymbol{\theta}) = \mathbf{0}_s$, where $\mathbf{0}_s = (0, \dots, 0)^\mathsf{T}$ denotes the null vector of $\mathbb{R}^s$, $h(\boldsymbol{\theta}) = (h_1(\boldsymbol{\theta}), \dots, h_s(\boldsymbol{\theta}))^\mathsf{T}$ and for all $k = 1, \dots, s$, $h_k$ is the function from $\mathbb{R}^{sr+1}$ to $\mathbb{R}$ defined by

$$h_k(\boldsymbol{\theta}) = \sum_{j=1}^{r} \beta_{jk} - 1. \tag{2.6}$$

For all $k = 1, \dots, s$, the components of $\boldsymbol{\pi}_k(\boldsymbol{\theta})$ are defined by

$$\pi_{1jk}(\boldsymbol{\theta}) = \frac{\beta_{jk}}{1 + \alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle}, \quad \pi_{2jk}(\boldsymbol{\theta}) = \frac{\alpha\beta_{jk}\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle}{1 + \alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle}, \quad j = 1, \dots, r, \tag{2.7}$$

where $\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle = \sum_{p=1}^{r} z_{pk}\beta_{pk}$.

In the case $s = 1$, there exist closed-form expression of the maximum likelihood estimator (MLE) $\boldsymbol{\theta}$ and the exact variance of its components except one for which the variance is approximated via the delta method [4]. In this article, we consider the general case $s \geqslant 1$.

*Remark* 2.1. The model (2.7) belongs to a general class of statistical models for crash frequencies including also [13, 15, 24]. For instance, for $r = 1$, model (2.7) is equivalent to the one of Tanner [24]. To avoid any possibility of confusion to the reader, it is worth mentioning that the model (2.7) considered in this paper differs from that of [13] for which there exist results regarding the variance-covariance matrix [16, 17]. The difference between the model of [14] considered in this paper and the one of [13] is the definition of the class probabilities $\pi_{2jk}(\boldsymbol{\theta})$.

Let $\boldsymbol{x}_k = (x_{11k}, \ldots, x_{1rk}, x_{21k}, \ldots, x_{2rk})^\top$, $k = 1, \ldots, s$, be $s$ vectors of integers such that $\sum_{i=1}^{2} \sum_{j=1}^{r} x_{ijk} = n_k$. The log-likelihood is given, up to an additive constant (see [14]), by

$$\ell(\boldsymbol{\theta}) = \left( \sum_{k=1}^{s} \sum_{j=1}^{r} x_{\bullet jk} \log \beta_{jk} \right) + x_{2\bullet\bullet} \log \alpha - \sum_{k=1}^{s} n_k \log(1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle) + \sum_{k=1}^{s} x_{2\bullet k} \log \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle, \quad (2.8)$$

where $x_{2\bullet\bullet} = \sum_{k=1}^{s} \sum_{j=1}^{r} x_{2jk}$ and for all $k = 1, \ldots, s$, $x_{\bullet jk} = x_{1jk} + x_{2jk}$ and $x_{2\bullet k} = \sum_{j=1}^{r} x_{2jk}$.

## 3. Exact asymptotic variance matrix

3.1. **Main results.** We have the following lemmas and theorems. Lemma 3.1 gives the structure of the matrix $\boldsymbol{H}$ involved in Equation (1.2). Theorem 3.2 gives the structure of $\boldsymbol{J}$, Lemma 3.3 is an intermediate lemma for Theorem 3.4 which gives the closed-form expression of $\boldsymbol{J}^{-1}$. Finally we give in Lemma 3.5 an intermediate result needed to establish Theorem 3.6. In this latter theorem, we prove that the matrix $\boldsymbol{H}^\top \boldsymbol{J}^{-1} \boldsymbol{H}$ involved in Equation (1.2) is invertible and we give the closed-form expression of its inverse.

**Lemma 3.1.** *The $(sr+1) \times s$ matrix $\boldsymbol{H}$ of partial derivatives involved in Equation (1.2) is given by*

$$\boldsymbol{H} = \left[ \begin{array}{cccc} 0 & \cdots & \cdots & 0 \\ \hline \mathbf{1}_r & \mathbf{0}_r & \cdots & \mathbf{0}_r \\ \mathbf{0}_r & \mathbf{1}_r & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_r \\ \mathbf{0}_r & \cdots & \mathbf{0}_r & \mathbf{1}_r \end{array} \right] = \left[ \begin{array}{cccc} 0 & \cdots & \cdots & 0 \\ \hline & & & \\ & \boldsymbol{H}_2 & & \\ & & & \end{array} \right], \quad (3.1)$$

*where $\mathbf{0}_r = (0, \ldots, 0)^\top$ denotes the null vector of $\mathbb{R}^r$, $\mathbf{1}_r = (1, \ldots, 1)^\top$ and $\boldsymbol{H}_2$ is the sub-matrix of order $sr \times sr$ formed of all the rows of $\boldsymbol{H}$ except the first.*

*Proof.* For all $k = 1, \ldots, s$, the $k$th column of $\boldsymbol{H}$ is composed of partial derivatives of $h_k(\boldsymbol{\theta})$ with respect to the components of $\boldsymbol{\theta}$. □

**Theorem 3.2.** *For all $k = 1, \ldots, s$, let*

$$\gamma_k = \frac{n_k}{1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}. \quad (3.2)$$

*The FIM is given by*

$$\boldsymbol{J} = \begin{bmatrix} \tau & \boldsymbol{U}^\top \\ \boldsymbol{U} & \boldsymbol{B} \end{bmatrix}, \quad (3.3)$$

*where*

$$\tau = \sum_{k=1}^{s} \tau_k, \quad \tau_k = \frac{\gamma_k^2 \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{n_k \alpha}, \quad k = 1, \dots, s,$$

$\boldsymbol{U}$ *and* $\boldsymbol{B}$ *are respectively the matrices of order* $sr \times 1$ *and* $sr \times sr$ *defined by*

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_1 \\ \vdots \\ \boldsymbol{U}_s \end{bmatrix} \quad et \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 & \bigcirc & \cdots & \bigcirc \\ \bigcirc & \boldsymbol{B}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bigcirc \\ \bigcirc & \cdots & \bigcirc & \boldsymbol{B}_s \end{bmatrix}$$

*such that for all* $k = 1, \dots, s$,

$$\boldsymbol{U}_k = \frac{\gamma_k^2}{n_k} \boldsymbol{Z}_k, \quad \boldsymbol{V}_k = \left( \frac{\gamma_k^2 \alpha}{n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} \right)^{1/2} \boldsymbol{Z}_k,$$

$$\boldsymbol{B}_k = \boldsymbol{\Delta}_k + \frac{\gamma_k^2 \alpha}{n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} \boldsymbol{Z}_k \boldsymbol{Z}_k^{\top} = \boldsymbol{\Delta}_k + \boldsymbol{V}_k \boldsymbol{V}_k^{\top}$$

*and*

$$\boldsymbol{\Delta}_k = n_k \begin{bmatrix} 1/\beta_{1k} & & \bigcirc \\ & \ddots & \\ \bigcirc & & 1/\beta_{rk} \end{bmatrix}. \tag{3.4}$$

*Proof.* The first order partial derivatives of $\ell(\boldsymbol{\theta})$ are the following:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \frac{x_{2\bullet\bullet}}{\alpha} - \sum_{k=1}^{s} \frac{n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle},$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_{jk}} = \frac{x_{\bullet jk}}{\beta_{jk}} - \frac{n_k \alpha z_{jk}}{1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} + \frac{x_{2\bullet k} z_{jk}}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}, \quad j = 1, \dots, r, \quad k = 1, \dots, s$$

and we deduce the second partial derivatives:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha^2} = -\frac{x_{2\bullet\bullet}}{\alpha^2} + \sum_{k=1}^{s} \frac{n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle^2}{\left(1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle\right)^2},$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \beta_{jk}} = -\frac{n_k z_{jk}}{\left(1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle\right)^2}, \quad j = 1, \dots, r, \quad k = 1, \dots, s,$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{im} \partial \beta_{jk}} = \begin{cases} 0 & \text{if} \quad m \neq k \\ -\dfrac{x_{\bullet jk}}{\beta_{jk}^2} + \dfrac{n_k \alpha^2 z_{jk}^2}{(1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle)^2} - \dfrac{x_{2\bullet k} z_{jk}^2}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle^2} & \text{if} \quad m = k \quad \text{and} \quad i = j \\ \dfrac{n_k \alpha^2 z_{ik} z_{jk}}{(1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle)^2} - \dfrac{x_{2\bullet k} z_{ik} z_{jk}}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle^2} & \text{if} \quad m = k \quad \text{and} \quad i \neq j \end{cases}$$

For all $j = 1, \ldots, r$ and $k = 1, \ldots, s$, we have by the properties of the multinomial distribution,

$$\mathbb{E}_{\boldsymbol{\theta}}(X_{1jk}) = n_k \pi_{1jk}(\boldsymbol{\theta}) = \frac{n_k \beta_{jk}}{1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} = \gamma_k \beta_{jk}$$

and

$$\mathbb{E}_{\boldsymbol{\theta}}(X_{2jk}) = n_k \pi_{2jk}(\boldsymbol{\theta}) = \frac{n_k \alpha \beta_{jk} \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} = \gamma_k \alpha \beta_{jk} \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle.$$

It is deduced that

$$\mathbb{E}(X_{2\bullet k}) = \sum_{j=1}^{r} \mathbb{E}_{\boldsymbol{\theta}}(X_{2jk}) = \gamma_k \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle,$$

$$\mathbb{E}(X_{\bullet jk}) = \mathbb{E}_{\boldsymbol{\theta}}(X_{1jk}) + \mathbb{E}_{\boldsymbol{\theta}}(X_{2jk}) = n_k \beta_{jk}$$

and

$$\mathbb{E}(X_{2\bullet\bullet}) = \sum_{k=1}^{s} \sum_{j=1}^{r} \mathbb{E}_{\boldsymbol{\theta}}(X_{2jk}) = \sum_{k=1}^{s} \mathbb{E}_{\boldsymbol{\theta}}(X_{2\bullet k}) = \sum_{k=1}^{s} \gamma_k \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle.$$

We have:

$$\begin{aligned}
\tau &= -\mathbb{E}_{\boldsymbol{\theta}} \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha^2} \right) \\
&= \sum_{k=1}^{s} \left\{ \frac{\gamma_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{\alpha} - \frac{\gamma_k^2 \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle^2}{n_k} \right\} \\
&= \sum_{k=1}^{s} \left\{ \frac{\gamma_k n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle - \gamma_k^2 \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle^2}{n_k \alpha} \right\} \\
&= \sum_{k=1}^{s} \frac{\gamma_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{n_k \alpha} \left( n_k - \gamma_k \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle \right) \\
&= \sum_{k=1}^{s} \frac{\gamma_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{n_k \alpha} \left( n_k - \frac{n_k \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} \right) \\
&= \sum_{k=1}^{s} \frac{\gamma_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{n_k \alpha} \left( \frac{n_k}{1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} \right) \\
&= \sum_{k=1}^{s} \frac{\gamma_k^2 \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{n_k \alpha}.
\end{aligned}$$

For all $k = 1, \ldots, s$, set $\boldsymbol{U}_k = (U_{k,1}, \ldots, U_{k,r})^\top$. For all $j = 1, \ldots, r$, we have:

$$U_{k,j} = -\mathbb{E}_{\boldsymbol{\theta}} \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \beta_{jk}} \right) = \frac{\gamma_k^2 z_{jk}}{n_k}.$$

Finally,

$$\boldsymbol{B}_k = \Big( B_{k,i,j} \Big)_{1 \leqslant i,j \leqslant r} = \left( -\mathbb{E}_{\boldsymbol{\theta}} \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{ik} \partial \beta_{jk}} \right) \right)_{1 \leqslant i,j \leqslant r}, \quad k = 1, \ldots, s,$$

where for all $i, j \in \{1, \dots, r\}$,

$$\begin{cases} B_{k,i,i} = \dfrac{n_k}{\beta_{ik}} + \dfrac{\gamma_k^2 \alpha z_{ik}^2}{n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}, \\ \\ B_{k,i,j} = -\dfrac{\gamma_k^2 \alpha^2 z_{ik} z_{jk}}{n_k} + \dfrac{\gamma_k \alpha z_{ik} z_{jk}}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} = \dfrac{\gamma_k^2 \alpha z_{ik} z_{jk}}{n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}, \qquad (i \neq j). \end{cases}$$

□

**Lemma 3.3.** *The matrix $\boldsymbol{B}$ is invertible and its inverse is given by:*

$$\boldsymbol{B}^{-1} = \begin{bmatrix} \boldsymbol{B}_1^{-1} & \bigcirc & \cdots & \bigcirc \\ \bigcirc & \boldsymbol{B}_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bigcirc \\ \bigcirc & \cdots & \bigcirc & \boldsymbol{B}_s^{-1} \end{bmatrix}, \tag{3.5}$$

*where for all $k = 1, \dots, s$,*

$$\boldsymbol{B}_k^{-1} = \boldsymbol{\Delta}_k^{-1} - \left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1} \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k \boldsymbol{V}_k^{\mathsf{T}} \boldsymbol{\Delta}_k^{-1} \tag{3.6}$$

*and $\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}} = \boldsymbol{V}_k^{\mathsf{T}} \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k$ ($\boldsymbol{\Delta}_k$ being a positive definite matrix).*

*Proof.* We know that $\det(\boldsymbol{B}) = \prod_{k=1}^s \det(\boldsymbol{B}_k)$ so it suffices to prove that for each $k = 1, \dots, s$, the matrix $\boldsymbol{B}_k$ is invertible. Let $k = 1, \dots, s$ and

$$\boldsymbol{M}_k = \begin{bmatrix} 1 & -\boldsymbol{V}_k^{\mathsf{T}} \\ \boldsymbol{V}_k & \boldsymbol{\Delta}_k \end{bmatrix}.$$

The Schur complement of $\boldsymbol{\Delta}_k$ in $\boldsymbol{M}_k$, denoted $(\boldsymbol{M}_k/\boldsymbol{\Delta}_k)$, is such that

$$(\boldsymbol{M}_k/\boldsymbol{\Delta}_k) = 1 + \boldsymbol{V}_k^{\mathsf{T}} \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k = 1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}} > 0.$$

By the properties of Schur complements [18, Theorem 2.1], we have

$$\det(\boldsymbol{M}_k) = \det(\boldsymbol{M}_k/\boldsymbol{\Delta}_k) \det(\boldsymbol{\Delta}_k) > 0$$

and the matrix $\boldsymbol{M}_k$ is invertible. But $\boldsymbol{B}_k = (\boldsymbol{M}_k/1)$ hence

$$\det(\boldsymbol{B}_k) = \det(\boldsymbol{M}_k)/1 = \det(\boldsymbol{M}_k) > 0$$

which proves that the matrix $\boldsymbol{B}_k$ is invertible. The matrix $\boldsymbol{B}$ is then invertible and

$$\boldsymbol{B}^{-1} = \begin{bmatrix} \boldsymbol{B}_1^{-1} & \bigcirc & \cdots & \bigcirc \\ \bigcirc & \boldsymbol{B}_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bigcirc \\ \bigcirc & \cdots & \bigcirc & \boldsymbol{B}_s^{-1} \end{bmatrix}. \tag{3.7}$$

For all $k = 1, \dots, s$, the Duncan inversion formula [20, p. 13] implies that

$$\boldsymbol{B}_k^{-1} = (\boldsymbol{M}_k/1)^{-1} = \boldsymbol{\Delta}_k^{-1} + \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k (\boldsymbol{M}_k/\boldsymbol{\Delta}_k)^{-1} (-\boldsymbol{V}_k^{\mathsf{T}}) \boldsymbol{\Delta}_k^{-1}$$

hence

$$\boldsymbol{B}_k^{-1} = \boldsymbol{\Delta}_k^{-1} - \left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1} \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k \boldsymbol{V}_k^{\mathsf{T}} \boldsymbol{\Delta}_k^{-1}.$$

□

**Theorem 3.4.** *The FIM $\boldsymbol{J}$ defined by (3.3) is invertible and its inverse is*

$$\boldsymbol{J}^{-1} = \begin{bmatrix} (\boldsymbol{J}/\boldsymbol{B})^{-1} & -(\boldsymbol{J}/\boldsymbol{B})^{-1}\boldsymbol{U}^{\top}\boldsymbol{B}^{-1} \\ -\boldsymbol{B}^{-1}\boldsymbol{U}(\boldsymbol{J}/\boldsymbol{B})^{-1} & \boldsymbol{B}^{-1} + \boldsymbol{B}^{-1}\boldsymbol{U}(\boldsymbol{J}/\boldsymbol{B})^{-1}\boldsymbol{U}^{\top}\boldsymbol{B}^{-1} \end{bmatrix}, \tag{3.8}$$

*where*

$$(\boldsymbol{J}/\boldsymbol{B}) = \sum_{k=1}^{s} \frac{\tau_k}{1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}}. \tag{3.9}$$

*Proof.* Since $\boldsymbol{B}$ is invertible, the Schur complement $(\boldsymbol{J}/\boldsymbol{B})$ exists and is such that

$$\begin{aligned}
(\boldsymbol{J}/\boldsymbol{B}) &= \tau - \boldsymbol{U}^{\top}\boldsymbol{B}^{-1}\boldsymbol{U} \\
&= \sum_{k=1}^{s} \tau_k - \sum_{k=1}^{s} \boldsymbol{U}_k^{\top}\boldsymbol{B}_k^{-1}\boldsymbol{U}_k \\
&= \sum_{k=1}^{s} \left\{\tau_k - \boldsymbol{U}_k^{\top}\boldsymbol{B}_k^{-1}\boldsymbol{U}_k\right\} \\
&= \sum_{k=1}^{s} \left\{\tau_k - \boldsymbol{U}_k^{\top}\left(\boldsymbol{\Delta}^{-1} - \left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1}\boldsymbol{\Delta}_k^{-1}\boldsymbol{V}_k\boldsymbol{V}_k^{\top}\boldsymbol{\Delta}_k^{-1}\right)\boldsymbol{U}_k\right\} \\
&= \sum_{k=1}^{s} \left\{\tau_k - \boldsymbol{U}^{\top}\boldsymbol{\Delta}_k^{-1}\boldsymbol{U}_k + \left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1}\boldsymbol{U}_k^{\top}\boldsymbol{\Delta}_k^{-1}\boldsymbol{V}_k\boldsymbol{V}_k^{T}\boldsymbol{\Delta}_k^{-1}\boldsymbol{U}_k\right\}.
\end{aligned}$$

Since $\boldsymbol{U}_k = \tau_k^{1/2}\boldsymbol{V}_k$, we have

$$\begin{aligned}
(\boldsymbol{J}/\boldsymbol{B}) &= \sum_{k=1}^{s} \left\{\tau_k - \tau_k\boldsymbol{V}_k^{\top}\boldsymbol{\Delta}_k^{-1}\boldsymbol{V}_k + \tau_k\left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1}\boldsymbol{V}_k^{\top}\boldsymbol{\Delta}_k^{-1}\boldsymbol{V}_k\boldsymbol{V}_k^{\top}\boldsymbol{\Delta}_k^{-1}\boldsymbol{V}_k\right\} \\
&= \sum_{k=1}^{s} \left\{\tau_k - \tau_k\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}} + \tau_k\left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1}\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right\} \\
&= \sum_{k=1}^{s} \left\{\tau_k - \tau_k\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\left(1 - \frac{\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}}{1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}}\right)\right\} \\
&= \sum_{k=1}^{s} \left\{\tau_k - \tau_k\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\left(\frac{1}{1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}}\right)\right\} \\
&= \sum_{k=1}^{s} \left\{\tau_k\left(1 - \frac{\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}}{1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}}\right)\right\} \\
&= \sum_{k=1}^{s} \frac{\tau_k}{1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}} > 0.
\end{aligned}$$

Since $(\boldsymbol{J}/\boldsymbol{B}) > 0$ and $\boldsymbol{B}$ is invertible (Lemma 3.3), we have

$$\det(\boldsymbol{J}) = (\boldsymbol{J}/\boldsymbol{B})\det(\boldsymbol{B}) \neq 0$$

which proves that $\boldsymbol{J}$ is invertible. To obtain Formula (3.8), it suffices to apply the Banachiewicz inversion formula [20, p. 13] to $\boldsymbol{J}$ as defined by (3.3). □

**Lemma 3.5.** *Let the $s \times s$ matrix $\boldsymbol{D} = \boldsymbol{H}_2^\top \boldsymbol{B}^{-1} \boldsymbol{H}_2$ where $\boldsymbol{H}_2$ is the $sr \times s$ matrix defined by Formula (3.1). The matrix $\boldsymbol{D}$ is invertible and its inverse is the diagonal matrix defined by*

$$\boldsymbol{D}^{-1} = \begin{bmatrix} 1/(\mathbf{1}_r^\top \boldsymbol{B}_1^{-1} \mathbf{1}_r) & & \bigcirc \\ & \ddots & \\ \bigcirc & & 1/(\mathbf{1}_r^\top \boldsymbol{B}_s^{-1} \mathbf{1}_r) \end{bmatrix}. \tag{3.10}$$

*Proof.* By the structure of $\boldsymbol{H}$ (Formula (3.1)), we have

$$\begin{aligned} \boldsymbol{D} &= \begin{bmatrix} \mathbf{1}_r^\top & \mathbf{0}_r^\top & \cdots & \mathbf{0}_r^\top \\ \mathbf{0}_r^\top & \mathbf{1}_r^\top & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_r^\top \\ \mathbf{0}_r^\top & \cdots & \mathbf{0}_r^\top & \mathbf{1}_r^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_1^{-1} & \bigcirc & \cdots & \bigcirc \\ \bigcirc & \boldsymbol{B}_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bigcirc \\ \bigcirc & \cdots & \bigcirc & \boldsymbol{B}_s^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1}_r & \mathbf{0}_r & \cdots & \mathbf{0}_r \\ \mathbf{0}_r & \mathbf{1}_r & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_r \\ \mathbf{0}_r & \cdots & \mathbf{0}_r & \mathbf{1}_r \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{1}_r^\top \boldsymbol{B}_1^{-1} \mathbf{1}_r & \bigcirc & \cdots & \bigcirc \\ \bigcirc & \mathbf{1}_r^\top \boldsymbol{B}_2^{-1} \mathbf{1}_r & \ddots & \vdots \\ \vdots & \ddots & \ddots & \bigcirc \\ \bigcirc & \cdots & \bigcirc & \mathbf{1}_r^\top \boldsymbol{B}_s^{-1} \mathbf{1}_r \end{bmatrix}. \end{aligned}$$

For all $k = 1, \ldots, s$,

$$\begin{aligned} \mathbf{1}_r^\top \boldsymbol{B}_k^{-1} \mathbf{1}_r &= \mathbf{1}_r^\top \left( \boldsymbol{\Delta}_k^{-1} - \left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1} \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k \boldsymbol{V}_k^\top \boldsymbol{\Delta}_k^{-1} \right) \mathbf{1}_r \\ &= \mathbf{1}_r^\top \boldsymbol{\Delta}_k^{-1} \mathbf{1}_r - \left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1} \mathbf{1}_r^\top \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k \boldsymbol{V}_k^\top \boldsymbol{\Delta}_k^{-1} \mathbf{1}_r \\ &= \mathbf{1}_r^\top \boldsymbol{\Delta}_k^{-1} \mathbf{1}_r - \left(1 + \|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}}\right)^{-1} \left(\mathbf{1}_r^\top \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k\right)^2 \end{aligned}$$

because $\mathbf{1}_r^\top \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k = \boldsymbol{V}_k^\top \boldsymbol{\Delta}_k^{-1} \mathbf{1}_r \in \mathbb{R}$. By definition of $\boldsymbol{\Delta}_k$ (see Formula (3.4)) and $\gamma_k$ (Theorem 3.2),

$$\mathbf{1}_r^\top \boldsymbol{\Delta}_k^{-1} \mathbf{1}_r = \frac{1}{n_k} \sum_{j=1}^r \beta_{jk} = \frac{1}{n_k},$$

$$\begin{aligned} \left(\mathbf{1}_r^\top \boldsymbol{\Delta}_k^{-1} \boldsymbol{V}_k\right)^2 &= \left( \frac{1}{n_k} \left( \frac{\gamma_k^2 \alpha}{n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} \right)^{1/2} \sum_{j=1}^r z_{jk} \beta_{jk} \right)^2 \\ &= \left( \frac{1}{n_k} \left( \frac{\gamma_k^2 \alpha}{n_k \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle} \right)^{1/2} \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle \right)^2 \\ &= \frac{\gamma_k^2 \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{n_k^3} \\ &= \frac{n_k^2 \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{n_k^3 (1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle)^2} \\ &= \frac{\alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle}{n_k (1 + \alpha \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle)^2} \end{aligned}$$

and

$$\begin{aligned}
\|\boldsymbol{V}_k\|^2_{\boldsymbol{\Delta}_k^{-1}} &= \frac{\gamma_k^2\alpha}{n_k\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle}\|\boldsymbol{Z}_k\|^2_{\boldsymbol{\Delta}_k^{-1}} \\
&= \frac{\gamma_k^2\alpha}{n_k\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle} \times \frac{1}{n_k}\sum_{j=1}^r z_{jk}^2\beta_{jk} \\
&= \frac{\gamma_k^2\alpha\sum_{j=1}^r \beta_{jk}z_{jk}^2}{n_k^2\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle} \\
&= \frac{n_k^2\alpha\sum_{j=1}^r \beta_{jk}z_{jk}^2}{n_k^2\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2} \\
&= \frac{\alpha\sum_{j=1}^r \beta_{jk}z_{jk}^2}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\boldsymbol{1}_r^\top \boldsymbol{B}_k^{-1}\boldsymbol{1}_r &= \frac{1}{n_k} - \frac{\dfrac{\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle}{n_k(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}}{1+\dfrac{\alpha\sum_{j=1}^r \beta_{jk}z_{jk}^2}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}} \\
&= \frac{1}{n_k}\left(1 - \frac{\dfrac{\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle}{(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}}{1+\dfrac{\alpha\sum_{j=1}^r \beta_{jk}z_{jk}^2}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}}\right) \\
&= \frac{1}{n_k}\left(\frac{1+\dfrac{\alpha\sum_{j=1}^r \beta_{jk}z_{jk}^2}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2} - \dfrac{\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle}{(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}}{1+\dfrac{\alpha\sum_{j=1}^r \beta_{jk}z_{jk}^2}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}}\right) \\
&= \frac{1}{n_k}\left(\frac{1+\dfrac{\alpha}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}\left(\sum_{j=1}^r \beta_{jk}z_{jk}^2 - \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle^2\right)}{1+\dfrac{\alpha\sum_{j=1}^r \beta_{jk}z_{jk}^2}{\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle(1+\alpha\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2}}\right).
\end{aligned}$$

But we have

$$\begin{aligned}
\sum_{j=1}^r \beta_{jk}(z_{jk} - \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle)^2 &= \sum_{j=1}^r \beta_{jk}z_{jk}^2 - 2\sum_{j=1}^r \beta_{jk}z_{jk}\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle + \sum_{j=1}^r \beta_{jk}\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle^2 \\
&= \sum_{j=1}^r \beta_{jk}z_{jk}^2 - 2\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle\sum_{j=1}^r \beta_{jk}z_{jk} + \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle^2\sum_{j=1}^r \beta_{jk} \\
&= \sum_{j=1}^r \beta_{jk}z_{jk}^2 - 2\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle^2 + \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k\rangle^2
\end{aligned}$$

because $\langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle = \sum_{j=1}^{r} \beta_{jk} z_{jk}$ and $\sum_{j=1}^{r} \beta_{jk} = 1$. Therefore,

$$\sum_{j=1}^{r} \beta_{jk}(z_{jk} - \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle)^2 = \sum_{j=1}^{r} \beta_{jk} z_{jk}^2 - \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle^2,$$

hence

$$\sum_{j=1}^{r} \beta_{jk} z_{jk}^2 - \langle \boldsymbol{Z}_k, \boldsymbol{\beta}_k \rangle^2 \geqslant 0.$$

Thus, for all $k = 1, \dots, s$, the numerator of $\mathbf{1}_r^\top \boldsymbol{B}_k^{-1} \mathbf{1}_r$ is strictly positive and therefore $\mathbf{1}_r^\top \boldsymbol{B}_k^{-1} \mathbf{1}_r > 0$. Since $\boldsymbol{D}$ is a diagonal matrix with strictly positive diagonal elements, it is invertible and its inverse is given by (3.10). □

**Theorem 3.6.** *Let the $s \times s$ matrix $\boldsymbol{C} = \boldsymbol{H}_2^\top \boldsymbol{B}^{-1} \boldsymbol{U}$ where $\boldsymbol{H}_2$ is the $sr \times s$ matrix defined by Formula (3.1). The $s \times s$ matrix $\boldsymbol{H}^\top \boldsymbol{J}^{-1} \boldsymbol{H}$ involved in (1.2) is invertible and its inverse is given by*

$$(\boldsymbol{H}^\top \boldsymbol{J}^{-1} \boldsymbol{H})^{-1} = \boldsymbol{D}^{-1} - (\boldsymbol{A}/\boldsymbol{D})^{-1} \boldsymbol{D}^{-1} \boldsymbol{C} \boldsymbol{C}^\top \boldsymbol{D}^{-1}, \tag{3.11}$$

*where $\boldsymbol{D}$ is the $s \times s$ matrix defined in Lemma 3.5 and $(\boldsymbol{A}/\boldsymbol{D}) \in \mathbb{R}^*$.*

*Proof.* By classical results of Statistics (see for example [2, p. 54] or [23, p. 165]), we know that the Fisher information matrix (FIM) is always positive semi-definite (its eigen values are all non-negative). In our case, the FIM $\boldsymbol{J}$ is also invertible (Theorem 3.4) which means that $\boldsymbol{J}$ is positive definite (its eigen values are strictly positive). Since $\boldsymbol{J}$ is positive definite and $\boldsymbol{H}$ is of full rank ($\mathrm{rank}(\boldsymbol{H}) = s$), the matrix $\boldsymbol{H}^\top \boldsymbol{J}^{-1} \boldsymbol{H}$ is invertible.

Now, by the structure of $\boldsymbol{H}$ (Formula (3.1)) and $\boldsymbol{J}^{-1}$ (Formula (3.8)), we have

$$\boldsymbol{H}^\top \boldsymbol{J}^{-1} \boldsymbol{H} = \begin{bmatrix} \mathbf{0}_s & \boldsymbol{H}_2^\top \end{bmatrix} \begin{bmatrix} (\boldsymbol{J}/\boldsymbol{B})^{-1} & -(\boldsymbol{J}/\boldsymbol{B})^{-1} \boldsymbol{U}^\top \boldsymbol{B}^{-1} \\ -\boldsymbol{B}^{-1} \boldsymbol{U} (\boldsymbol{J}/\boldsymbol{B})^{-1} & \boldsymbol{B}^{-1} + \boldsymbol{B}^{-1} \boldsymbol{U} (\boldsymbol{J}/\boldsymbol{B})^{-1} \boldsymbol{U}^\top \boldsymbol{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0}_s^\top \\ \boldsymbol{H}_2 \end{bmatrix}$$

which is equivalent to

$$\boldsymbol{H}^\top \boldsymbol{J}^{-1} \boldsymbol{H} = \boldsymbol{H}_2^\top \boldsymbol{B}^{-1} \boldsymbol{H}_2 + \boldsymbol{H}_2^\top \boldsymbol{B}^{-1} \boldsymbol{U} (\boldsymbol{J}/\boldsymbol{B})^{-1} \boldsymbol{U}^\top \boldsymbol{B}^{-1} \boldsymbol{H}_2.$$

Since $(\boldsymbol{J}/\boldsymbol{B}) \in \mathbb{R}$ and $\boldsymbol{B}^{-1}$ is symmetric, we have

$$\boldsymbol{H}^\top \boldsymbol{J}^{-1} \boldsymbol{H} = \boldsymbol{D} + (\boldsymbol{J}/\boldsymbol{B})^{-1} \boldsymbol{C} \boldsymbol{C}^\top.$$

So if we set

$$\boldsymbol{A} = \begin{bmatrix} (\boldsymbol{J}/\boldsymbol{B}) & -\boldsymbol{C}^\top \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix},$$

then $\boldsymbol{H}^\top \boldsymbol{J}^{-1} \boldsymbol{H} = \boldsymbol{A}/(\boldsymbol{J}/\boldsymbol{B})$ where $\boldsymbol{A}/(\boldsymbol{J}/\boldsymbol{B})$ is the Schur complement of $(\boldsymbol{J}/\boldsymbol{B})$ in $\boldsymbol{A}$. Since $\boldsymbol{A}/(\boldsymbol{J}/\boldsymbol{B})$ is invertible and $(\boldsymbol{J}/\boldsymbol{B}) > 0$,

$$\det(\boldsymbol{A}) = (\boldsymbol{J}/\boldsymbol{B}) \times \det(\boldsymbol{A}/(\boldsymbol{J}/\boldsymbol{B})) \neq 0$$

and $\boldsymbol{A}$ is invertible. The matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ being invertible, the Schur complement $(\boldsymbol{A}/\boldsymbol{D}) = (\boldsymbol{J}/\boldsymbol{B}) + \boldsymbol{C}^\top \boldsymbol{D}^{-1} \boldsymbol{C}$ is a real number and we have

$$\det(\boldsymbol{A}) = \det(\boldsymbol{D}) \times (\boldsymbol{A}/\boldsymbol{D}) \neq 0$$

which means that $(\boldsymbol{A}/\boldsymbol{D}) \neq 0$. So, by the Duncan inversion formula [20, p. 13],

$$\begin{aligned}(\boldsymbol{H}^\top \boldsymbol{J}^{-1}\boldsymbol{H})^{-1} &= (\boldsymbol{A}/(\boldsymbol{J}/\boldsymbol{B}))^{-1} \\ &= \boldsymbol{D}^{-1} - (\boldsymbol{A}/\boldsymbol{D})^{-1}\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{C}^\top\boldsymbol{D}^{-1}.\end{aligned}$$

□

3.2. **Computation strategy.** We have obtained the closed-form expression of the variance-covariance matrix $\boldsymbol{W}$ by explaining the structure of the inverses $\boldsymbol{J}^{-1}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$. Although possible, the manual calculation of the $(1+sr)^2$ components of $\boldsymbol{W}$ can be quite cumbersome for large values of $s$ and $r$. We propose to implement the formulas using a calculation software having functionalities to carry out basic matrix operations (multiplication by a real, sum and product of matrices). In summary, the proposed approach consists in computing $\boldsymbol{W}$ using Formula (1.2) where $\boldsymbol{J}^{-1}$ is computed by Formula (3.8) and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$ is computed by Formula (3.11). The advantage of our approach is that no numerical inversion is necessary.

## 4. Simulation study

In this section, we compare in the R software [21],

- the method proposed in this paper (hereafter referred to as **Method 1**) consisting in computing $\boldsymbol{W}$ using respectively the analytical formulas (3.8) and (3.11) to compute the inverses of $\boldsymbol{J}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$
- and the method (hereafter referred to as **Method 2**) that consists in computing $\boldsymbol{W}$ using the R software's **solve** function to numerically invert $\boldsymbol{J}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$.

We set the values of $n_k$ $(k = 1, \ldots, s)$ to two common values ($n = 50$ and $n = 5000$) and the values of $s$ and $r$ to different values in order to vary the number of parameters of the model (and therefore the dimension of matrices $\boldsymbol{J}$ and $\boldsymbol{H}^\top\boldsymbol{J}^{-1}\boldsymbol{H}$). For each combination of values of $n$, $s$ and $r$, we made one thousand (1000) replications of the process of randomly generating the components of $\boldsymbol{\theta}$ and $\boldsymbol{Z}_k$ $(k = 1, \ldots, s)$ and computing $\boldsymbol{W}$ by both methods. The comparison of the two methods was made on two aspects: the proportion of success (expressed in % out of 1000 replications) in the computation of $\boldsymbol{W}$ and the average central process unit (CPU) time (in seconds) used for the computation. The results are given in Table 1 for $n = 50$ and Table 2 for $n = 5000$.

The analysis of the results shows that the proportion of success of the method proposed in this paper (Method 1) is always 100%. And this proportion does not vary when the number of parameters increases. For Method 2 on the other hand, the proportion of success never reaches 100% even for a small number of parameters. This proportion decreases when the number of parameters increases and vanishes from a threshold depending on $n$. As for CPU times, they are relatively the same for both methods in cases where both methods successfully computed $\boldsymbol{W}$.

Table 1. Results obtained by applying the two methods to the computation of $\boldsymbol{W}$ for $n = 50$. Prop. of success represents the proportion of success (over 1000 replications) in computing $\boldsymbol{W}$.

| $s$ | $r$ | Number of parameters | Method 1 | | Method 2 | |
|---|---|---|---|---|---|---|
| | | | Prop. of success (%) | CPU time | Prop. of success (%) | CPU time |
| 2 | 3 | 7 | 100 | 0.0005 | 99.5 | 0.0005 |
| 5 | 3 | 16 | 100 | 0.0006 | 99.5 | 0.0006 |
| 6 | 4 | 25 | 100 | 0.0007 | 98.8 | 0.0007 |
| 8 | 4 | 33 | 100 | 0.0009 | 97.4 | 0.0008 |
| 8 | 5 | 41 | 100 | 0.0010 | 96.1 | 0.0010 |
| 10 | 5 | 51 | 100 | 0.0014 | 95.7 | 0.0014 |
| 10 | 6 | 61 | 100 | 0.0018 | 91.6 | 0.0017 |
| 10 | 7 | 71 | 100 | 0.0022 | 90.2 | 0.0023 |
| 12 | 7 | 85 | 100 | 0.0032 | 87.3 | 0.0033 |
| 14 | 7 | 99 | 100 | 0.0043 | 85.2 | 0.0047 |
| 15 | 7 | 106 | 100 | 0.0052 | 85 | 0.0054 |
| 15 | 8 | 121 | 100 | 0.0068 | 0 | - |
| 20 | 7 | 141 | 100 | 0.0113 | 0 | - |
| 20 | 8 | 161 | 100 | 0.0145 | 0 | - |
| 20 | 9 | 181 | 100 | 0.0182 | 0 | - |
| 20 | 10 | 201 | 100 | 0.0227 | 0 | - |

Table 2. Results obtained by applying the two methods to the computation of $\boldsymbol{W}$ for $n = 5000$. Prop. of success represents the proportion of success (over 1000 replications) in computing $\boldsymbol{W}$.

| $s$ | $r$ | Number of parameters | Method 1 | | Method 2 | |
|---|---|---|---|---|---|---|
| | | | Prop. of success (%) | CPU time | Prop. of success (%) | CPU time |
| 2 | 3 | 7 | 100 | 0.0005 | 99.5 | 0.0005 |
| 5 | 3 | 16 | 100 | 0.0007 | 98.7 | 0.0007 |
| 6 | 4 | 25 | 100 | 0.0009 | 97.8 | 0.0008 |
| 8 | 4 | 33 | 100 | 0.0010 | 97.3 | 0.0011 |
| 8 | 5 | 41 | 100 | 0.0012 | 95.9 | 0.0012 |
| 10 | 5 | 51 | 100 | 0.0017 | 95.6 | 0.0017 |
| 10 | 6 | 61 | 100 | 0.0021 | 92.4 | 0.0022 |
| 10 | 7 | 71 | 100 | 0.0026 | 0 | - |
| 12 | 7 | 85 | 100 | 0.0038 | 0 | - |
| 14 | 7 | 99 | 100 | 0.0054 | 0 | - |
| 15 | 7 | 106 | 100 | 0.0059 | 0 | - |
| 15 | 8 | 121 | 100 | 0.0077 | 0 | - |
| 20 | 7 | 141 | 100 | 0.0129 | 0 | - |
| 20 | 8 | 161 | 100 | 0.0155 | 0 | - |
| 20 | 9 | 181 | 100 | 0.0186 | 0 | - |
| 20 | 10 | 201 | 100 | 0.0231 | 0 | - |

## 5. Concluding remarks

In this paper, we proposed a method based on Schur complement for computing the asymptotic variance matrix $\boldsymbol{W}$ of the maximum likelihood estimator of the parameter vector $\boldsymbol{\theta}$ of a statistical model used in road safety. The parameter vector has $1 + sr$ components (where $s$ and $r$ are given positive integers) and is subject to $s$ equality constraints. Because of the presence of constraints, the computation of the so important asymptotic variance matrix $\boldsymbol{W}$ can only be done using the formula proposed in [1]. This formula requires some matrix inversions which become very complicated for large values of $s$ and $r$. To circumvent this difficulty, we used Schur complement for the exact analytic inversion and we proposed a method for computing $\boldsymbol{W}$. The results of the comparison of the proposed method with the traditional method which consists in computing $\boldsymbol{W}$ using a numerical inversion procedure prove the effectiveness of our method. Indeed, the analysis of the results shows that the proportion of success of the method proposed in this paper is always 100% whatever the number of parameters while the proportion of success of the traditional method never reaches 100% even for a small number of parameters, decreases when the number of parameters increases and vanishes from a threshold depending on the sample size. Our method outperforms the traditional method and can therefore be implemented without any restriction on the number of parameters of the model.

In statistics, the computation of the asymptotic variance matrix often completes the estimation of the parameters of a model by the maximum likelihood method. For the model considered in this paper, this estimation of the parameters consists in solving the following problem:

$$\text{maximize } L(\boldsymbol{\theta}) \tag{5.1}$$

under the constraints

$$\alpha > 0, \quad \beta_{jk} > 0, \quad j = 1, \ldots, r, \quad k = 1, \ldots, s, \tag{5.2}$$

and

$$\sum_{j=1}^{r} \beta_{jk} = 1, \quad k = 1, \ldots, s. \tag{5.3}$$

In the general case $s > 1$, the resolution of (5.1) under the constraints (5.2) – (5.3) can only be done numerically. In this context, it has been proven that this problem is difficult to deal with for classical optimization algorithms when the number of parameters increases and therefore, two optimization algorithms have been developed and their efficiency has be proven on simulated data [6, 8]. The results obtained in our present paper therefore complete those in [6, 8] and allow to make applications on real data in the case $s > 1$ as the authors in [7] did for $s = 1$.

In future works, it would be interesting to study the strong consistency of the maximum likelihood estimator for the model of [14] considered in this paper as the author of [5] did for the model described in [13].

## References


1. Aitchison, J., & Silvey, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. The Annals of Mathematical Statistics, 29(3), 813-828. https://doi.org/10.1214/aoms/1177706538
2. Drton, M., & Sturmfels, B., & Sullivant, S. (2009). Lectures on Algebraic Statistics. Birkhäuser Basel, Berlin. https://doi.org/10.1007/978-3-7643-8905-5
3. Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical science, 1(1), 54-75. https://doi.org/10.1214/ss/1177013815
4. Geraldo, I. C. (2020). On the maximum likelihood estimator for a discrete multivariate crash frequencies model. Afrika Statistika, 15(2), 2335-2348. https://doi.org/10.16929/as/2020.2325.161
5. Geraldo, I. C. (2023). Existence, uniqueness and strong consistency of the maximum likelihood estimator for a model of accidents frequencies. Gulf Journal of Mathematics, 14(2), 40-53. https://doi.org/10.56947/gjom.v14i2.1177
6. Geraldo, I. C., & Katchekpele, E., & Kpanzou, T. A. (2023). A fast and efficient estimation of the parameters of a model of accident frequencies via an MM algorithm. Journal of Applied Mathematics, 2023, paper ID 3377201. https://doi.org/10.1155/2023/3377201
7. Geraldo, I. C., & Kpanzou, T. A. (2022). Theoretical and empirical comparison of two discrete statistical models of crash frequencies. Malaya Journal of Matematik, 10(4):322–335. https://doi.org/10.26637/mjm1004/003
8. Geraldo, I. C., & Kpanzou, T. A., & Katchekpele, E. (2023). A new hybrid algorithm for maximum likelihood estimation in a model of accident frequencies. Malaya Journal of Matematik, 11(1), 11-24. https://doi.org/10.26637/mjm1101/002
9. Hunter, D. R., & Lange, K. (2004). A tutorial on MM algorithms. The American Statistician, 58(1), 30-37. https://doi.org/10.1198/0003130042836
10. Jamshidian, M., & Jennrich, R. I. (2000). Standard errors for EM estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62(2), 257-270. https://doi.org/10.1111/1467-9868.00230
11. Millar, R. B. (2011). Maximum likelihood estimation and inference: with examples in R, SAS and ADMB. John Wiley & Sons.
12. Neuenschwander, B. E., & Flury, B. D. (1997). A note on Silvey's (1959) Theorem. Statistics & Probability Letters, 36(3), 307-317. https://doi.org/10.1016/S0167-7152(97)00077-1
13. N'Guessan, A., Essai, A., & Langrand, C. (2001). Estimation multidimensionnelle des contrôles et de l'effet moyen d'une mesure de sécurité routière. Revue de Statistique Appliquée, 49(2), 85-102.
14. N'Guessan, A., Essai, A., & N'Zi, M. (2006). An estimation method of the average effect and the different accident risks when modelling a road safety measure: A simulation study. Computational Statistics & Data Analysis, 51(2), 1260-1277. https://doi.org/10.1016/j.csda.2005.09.002
15. N'Guessan, A., & Langrand, C. (1993). Sur la distribution asymptotique de certaines statistiques utilisées dans le domaine de la sécurité routière. C. R. Acad. Sci. Paris, Ser. I, 317(4), 401-404.
16. N'Guessan, A., & Langrand, C. (2005). A Schur complement approach for computing subcovariance matrices arising in a road safety measure modelling. Journal of Computational and Applied Mathematics, 177(2), 331-345. https://doi.org/10.1016/j.cam.2004.09.023
17. N'Guessan, A., & Langrand, C. (2005). A covariance components estimation procedure when modelling a road safety measure in terms of linear constraints. Statistics, 39(4), 303-314. https://doi.org/10.1080/02331880500108544
18. Ouellette, D. V. (1981). Schur complements and statistics. Linear Algebra and its Applications, 36, 187-295. https://doi.org/10.1016/0024-3795(81)90232-9

19. Parr, W. C. (1983). A note on the jackknife, the bootstrap and the delta method estimators of bias and variance. Biometrika, 70(3), 719-722. https://doi.org/10.2307/2336511
20. Puntanen, S., & Styan, G. P. H. (2005). Historical Introduction: Issai Schur and the Early Development of the Schur Complement. In Zhang, F. (editor), The Schur Complement and Its Applications, pages 1-16. Springer US, Boston, MA. https://doi.org/10.1007/0-387-24273-2_1
21. R Core Team (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
22. Spall, J. C. (2005). Monte Carlo computation of the Fisher information matrix in non-standard settings. Journal of Computational and Graphical Statistics, 14(4), 889-909. https://doi.org/10.1198/106186005X78800
23. Sullivant, S. (2018). Algebraic Statistics, volume 194 of Graduate Studies in Mathematics. American Mathematical Society, Providence, Rhode Island.
24. Tanner, J. C. (1958). A problem in the combination of accident frequencies. Biometrika, 45(3-4), 331-342. https://doi.org/10.1093/biomet/45.3-4.331

Département de Mathématiques, Faculté des Sciences, Université de Lomé, 1 B.P. 1515 Lomé 1, Togo.

*Email address*: cherifgera@gmail.com