REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

Fisher information matrix for two-way random effects model with heteroscedasticity

Authors: Patrice Takam Soh ✉
– Dept. of Mathematics, University of Yaoundé 1, Yaoundé, Cameroon
patricetakam@gmail.com

Eugene Kouassi
– Dept. of Economics, University FHB, Abidjan, Cote-d'Ivoire
eugene.kouassi@gmail.com

Jean Marcelin Bosson Brou
– Dept. of Economics, University FHB, Abidjan, Cote-d'Ivoire
jmbbrou@gmail.com

Saralees Nadarajah
– Dept. of Mathematics, University of Manchester, Manchester, UK
mbbsssn2@manchester.ac.uk

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- Model specification and selection are important aspects of modeling exercises. In this context, the Fisher Information Matrix (FIM) plays an essential role. In this paper, we derive the Fisher Information Matrix (FIM) for the two way random effects panel data models in general as well as in some specific cases of heteroscedasticity. Some computational issues are then raised and discussed. In addition, some real data examples are reported and thoroughly discussed.

Keywords:

- *Computational issues; Heteroscedasticity; Information matrix; Real data applications; Two way models.*

AMS Subject Classification:

- 49A05, 78B26.

✉ Corresponding author

1 Introduction

Does the structure of the Fisher Information Matrix (FIM, hereafter) matter in model specification and selection? Does the FIM matter in testing and/or in inference analysis particularly in the two way random effects panel data model? Which FIM results are more relevant or appealing? Observed or exact FIM results?

This paper seeks to get the FIM of the two-way random effects panel data model in the absence or presence of heteroscedasticity. The approach developed follows that in [1, 4, 3, 17, 18, 20, 24] and [28]. The FIM crucially depends on the variance covariance matrix. This matrix is obtained in four cases (homoscedasticity, heteroscedasticity on the unobservable individual effect, heteroscedasticity on the composite term and heteroscedasticity on both individual and composite terms); the paper does not consider groupwise heteroscedasticity or group membership heteroscedasticity as well as cluster heteroscedasticity (see [10, 15] and [19]) for example on the unobservable time effect.

This paper rather focuses on an alternative simple procedure for obtaining the FIM that accounts for homoscedasticity and/or various heteroscedasticity schemes in the two-way random effect model. It proposes a case-by-case approach rather than an elaborated sequence of steps and built-in functions used by earlier researchers.

The contributions of this paper are therefore twofold: (i) the derivation of the FIM based on different forms of homoscedasticity and/or heteroskedasticity; an important aspect in model specification; (ii) and thereby the exploration of how to choose the correct model specification in this context.

To do that, we develop a new and efficient procedure for computing the FIM in the two-way random effect model; the new procedure is obtained under homoscedasticity as well as well as various cases of heteroscedasticity.

The remainder of the paper is organized as follows: Section 2 describes the mathematical problem to be addressed. Section 3 sets out some preliminary results. Section 4 presents the main results while Section 5 discusses some computational issues. Section 6 provides two real data examples with discussions. Section 7 concludes the paper.

2 The mathematical problem

This section deals with the mathematical problem to address and some background information.

2.1 The FIM and related problem

We are interested in the derivation and computation of the FIM for the two-way random effects model (commonly encountered in theoretical as well as empirical studies) in the presence of various forms of heteroskedasticity.

Let $Y = (Y_1, ..., Y_n)$ be a random sample, and let $f(Y|\theta)$ denote the probability density function for some model of the data, which has parameter vector $\theta = (\theta_1, ..., \theta_r)'$. Then the FIM $I_n(\theta)$ of sample size n is given by the $r \times r$ symmetric matrix whose ij-th element is

$$I_n(\theta)_{i,j} = -\mathbb{E}\left[\frac{\partial^2 \ln f(Y \mid \theta)}{\partial\theta_i \partial\theta_j}\right]. \tag{2.1}$$

This definition strictly corresponds to the expected FIM. If no expectation is taken we obtain a data-dependent quantity that is called the observed FIM. We are interested in the derivation and computation of the expected FIM for the two-way random effects panel data model in the presence of heteroskedasticity.

2.2 The two way error components model

We consider the following two way error components model

$$y_{it} = \alpha + X'_{it}\beta + u_{it}; \qquad i = 1, \cdots, N; \;\; t = 1, \cdots, T \tag{2.2}$$

with i denoting households, individuals, firms, countries, etc., and t denoting time. The subscript i, therefore, denotes the cross-section dimension whereas t denotes the time-series dimension. y_{it} is the dependent variable for i at time t. β is a $K \times 1$ scalar, X_{it} is it-th observation on K explanatory variables. In this paper, we deal with two-way error components disturbances

$$u_{it} = \mu_i + \lambda_t + \nu_{it}, \tag{2.3}$$

where μ_i denotes the unobservable individual effect, λ_t denotes the unobservable time effect and ν_{it} is the remainder stochastic disturbance term. μ_i, λ_t account for any individual specific effect or time-specific effect not included in the regression. In vector form, (2.2) can be written as

$$u = Z_\mu \mu + Z_\lambda \lambda + \nu, \tag{2.4}$$

where $Z_\mu = I_N \otimes i_T$, I_N is an identity matrix of dimension N, i_T is a vector of ones of dimension T and $\otimes$ denotes Kronecker product. Z_μ is a selector matrix of ones and zeros, or simply the matrix of individual dummies included in the regression to estimate the μ_i (assuming they are fixed parameters to be estimated). Likewise, $Z_\lambda = i_N \otimes I_T$ is the matrix of time dummies of ones which may be included in the regression to estimate the λ_t (assuming they are fixed parameters to be estimated). μ, λ and ν are defined as in [1].

In vector form (2.2) can be written as

$$y = \alpha i_{NT} + X\beta + u = Z\gamma + u, \tag{2.5}$$

where y is $NT \times 1$, X is $NT \times K$, $Z = [i_{NT}, X]$, $\gamma' = (\alpha', \beta')$ and i_{NT} is a vector of ones.

2.3 Variance-covariance matrix of u

To obtain the variance-covariance matrix of the overall error term u, we assume the following.

Assumption $\mathcal{A}_1$. (**General case**): The vectors λ, ν and μ are pairwise independent. Each of them is identically and independently normally distributed with mean 0 and variances $\sigma_\lambda^2 I_T$, $\sigma_\nu^2 \mathrm{diag}(h_\nu(w_i'\theta_\nu))$ and $\sigma_\mu^2 \mathrm{diag}(h_\mu(z_i'\theta_\mu))$. h_ν and h_μ are differentiable functions from $\mathbb{R}$ to $\mathbb{R}^+$, $w_i = (w_{1i}, ..., w_{pi})' \in \mathbb{R}^p$ and $z_i = (z_{1i}, ..., z_{qi})' \in \mathbb{R}^q$ are defined as in [3].

In the following, we set $D_\nu = \mathrm{diag}(h_\nu(w_i'\theta_\nu))$ and $D_\mu = \mathrm{diag}(h_\mu(z_i'\theta_\mu))$. Based on the general assumption $\mathcal{A}_1$, the variance-covariance matrix of the composite disturbance u is defined by

$$\Omega = \mathbb{E}(uu') = \sigma_\mu^2 Z_\mu D_\mu Z_\mu' + \sigma_\lambda^2 Z_\lambda Z_\lambda' + \sigma_\nu^2 D_\nu \otimes I_T.$$

Therefore,

$$\Omega = \sigma_\mu^2 D_\mu \otimes i_T i_T' + \sigma_\lambda^2 i_N i_N' \otimes I_T + \sigma_\nu^2 D_\nu \otimes I_T$$

which can be simplified to

$$\Omega = \sigma_\nu^2 (D_\nu \otimes I_T) + \sigma_\mu^2 (D_\mu \otimes J_T) + \sigma_\lambda^2 (J_N \otimes I_T) \tag{2.6}$$

with $J_T = i_T i_T'$ and $J_N = i_N i_N'$.

2.4 Inverse of error variance-covariance matrix: Ω^{-1}

In order to get Ω^{-1}, we use the spectral decomposition in [32]. After replacing J_N by $N\overline{J}_N$, I_N by $E_N + \overline{J}_N$, J_T by $T\overline{J}_T$ and I_T by $E_T + \overline{J}_T$ and collecting terms with the same matrices, we obtain

$$\Omega = \left[\sigma_\nu^2 D_\nu + \sigma_\lambda^2 J_N\right] \otimes E_T + \left[\sigma_\nu^2 D_\nu + T\sigma_\mu^2 D_\mu + \sigma_\lambda^2 J_N\right] \otimes \overline{J}_T. \tag{2.7}$$

The spectral decomposition allows us to write

$$\Omega^{-1} = C_1 \otimes E_T + C_2 \otimes \overline{J}_T \tag{2.8}$$

with $C_1 = \zeta(\sigma_\nu^2 D_\nu, 0)$ and $C_2 = \zeta(\sigma_\nu^2 D_\nu, \sigma_\mu^2 D_\mu)$, where

$$\begin{aligned}\zeta(X_1, X_2) &= \left[X_1 + TX_2 + \sigma_\lambda^2 \left(i_N i_N'\right)\right]^{-1} \\ &= (X_1 + TX_2)^{-1} - \frac{\sigma_\lambda^2 (X_1 + TX_2)^{-1} J_N (X_1 + TX_2)^{-1}}{\left(1 + \sigma_\lambda^2 i_N' (X_1 + TX_2)^{-1} i_N\right)}.\end{aligned}$$

The formula used to obtain the inverse of $X_1 + TX_2 + \sigma_\lambda^2 (i_N i_N')$ is provided by [9].

3 Some preliminary results

This section deals with the derivation of $\mathbb{E}\left[-d^2\ell(\theta \mid u)\right]$, where $\ell(\theta \mid y)$ is the log-likelihood of observations. The relationship between this expectation and the FIM is given by

$$\begin{aligned}\mathbb{E}\left[-d^2\ell(\theta \mid y)\right] &= \sum_{i=1}^{r}\sum_{j=1}^{r}\mathbb{E}\left[-\frac{\partial^2\ell(\theta \mid y)}{\partial\theta_i\partial\theta_j}\right] d\theta_i d\theta_j \\ &= \sum_{i=1}^{r}\sum_{j=1}^{r} I_n(\theta)_{i,j} d\theta_i d\theta_j \\ &= (d\theta)' I_n(\theta)(d\theta), \end{aligned} \tag{3.1}$$

where $d\theta = (d\theta_1, ..., d\theta_r)'$.

3.1 First order derivatives of the log-likelihood function

If μ_i, λ_t and ν_{it} are independent and identically normally distributed (from assumption $\mathcal{A}_1$), the joint distribution of $y = (y_{11}, ..., y_{1T}, y_{21}, ..., y_{2T}, ..., y_{N1}, ..., y_{NT})'$ is the NT-multivariate normal distribution and the likelihood of the observations is

$$L(\theta \mid y) = \frac{1}{(2\pi)^{\frac{NT}{2}} |\Omega|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y - X\beta)'\Omega^{-1}(y - X\beta)\right).$$

Since $u = y - X\beta$,

$$L(\theta \mid u) = \frac{1}{(2\pi)^{\frac{NT}{2}} |\Omega|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}u'\Omega^{-1}u\right).$$

By taking the logarithm of the likelihood of observations

$$\ell(\theta \mid u) = \ln L(\theta \mid u) = C - \frac{1}{2}\ln|\Omega| - \frac{1}{2}u'\Omega^{-1}u, \tag{3.2}$$

where $\theta = (\sigma_\nu^2, \sigma_\mu^2, \sigma_\lambda^2, \theta_\nu', \theta_\mu', \beta')' \in \mathbb{R}^r$ (with $r = p + q + k + 3$) is the vector of parameters and $C = -\frac{NT}{2}\ln(2\pi)$ is a constant.

We observe that the log-likelihood is continuous and at least twice differentiable with respect to each parameter. The first and second order differentials are given by the following.

Lemma 3.1. *The first order differential of L is*

$$d\ell(\theta \mid u) = -\frac{1}{2}tr\left(\Omega^{-1}d\Omega\right) - u'\Omega^{-1}du + \frac{1}{2}u'\Omega^{-1}d\Omega \cdot \Omega^{-1}u. \tag{3.3}$$

The differential of u is $\mathrm{d}u = -Xd\beta$. *The differential of Ω is*

$$\begin{aligned} d\Omega \;=\; & (D_\nu \otimes I_T)d\sigma_\nu^2 + (D_\mu \otimes J_T)d\sigma_\mu^2 + (J_N \otimes I_T)d\sigma_\lambda^2 \\ & + \sigma_\nu^2 \sum_{j=1}^{p} (D_{\nu,j}^* \otimes I_T)d\theta_{\nu,j} + \sigma_\mu^2 \sum_{j'=1}^{q} (D_{\mu,j'}^* \otimes J_T)d\theta_{\mu,j'}, \end{aligned}$$

where $D_{\nu,j}^* = \dfrac{\partial D_\nu}{\partial \theta_{\nu,j}}$ *for* $j = 1, ..., p$ *and* $D_{\mu,j'}^* = \dfrac{\partial D_\mu}{\partial \theta_{\mu,j'}}$ *for* $j' = 1, ..., q$.

Proof of Lemma 3.1: We have

$$\begin{aligned} d\ell(\theta \mid u) \;&=\; 0 - \frac{1}{2}d\ln|\Omega| - \frac{1}{2}d\left(u'\Omega^{-1}u\right) \\ &= -\frac{1}{2}tr\left(\Omega^{-1}d\Omega\right) - \frac{1}{2}d\left(u'\Omega^{-1}u\right), \end{aligned}$$

where we used the formula $d\ln|\Omega| = tr\left(\Omega^{-1}d\Omega\right)$. We also have

$$\begin{aligned} d\left(u'\Omega^{-1}u\right) \;&=\; d\left(u'\Omega^{-1}u\right)_u + u'd\left(\Omega^{-1}\right)u \\ &= \frac{\partial}{\partial u}\left(u'\Omega^{-1}u\right)du + u'd\Omega^{-1}u \\ &= 2u'\Omega^{-1}du - u'\Omega^{-1}d\Omega \cdot \Omega^{-1}u, \end{aligned}$$

where we used the fact that $dX^{-1} = -X^{-1}dX \cdot X^{-1}$. □

3.2 Second order derivatives of the log likelihood function

Lemma 3.2. *The second order differential of L is*

$$\begin{aligned} d^2\ell(\theta \mid u) \;=\; & \frac{1}{2}tr\left(\Omega^{-1}d\Omega \cdot \Omega^{-1}d\Omega\right) - \frac{1}{2}tr\left(\Omega^{-1}d^2\Omega\right) \\ & -u'\Omega^{-1}d\Omega \cdot \Omega^{-1}d\Omega \cdot \Omega^{-1}u + \frac{1}{2}u'\Omega^{-1}d^2\Omega \cdot \Omega^{-1}u \\ & +u'\Omega^{-1}d\Omega \cdot \Omega^{-1}du - du'\Omega^{-1}du. \end{aligned} \tag{3.4}$$

Proof of Lemma 3.2: We have

$$\begin{aligned} d^2\ell(\theta \mid u) &= -\frac{1}{2}tr\left(d\left[\Omega^{-1}d\Omega\right]\right) - d\left(u'\Omega^{-1}du\right) + \frac{1}{2}d\left(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\right) \\ &= -\frac{1}{2}\gamma_1 - \gamma_2 + \frac{1}{2}\gamma_3, \end{aligned}$$

where

$$\begin{aligned} \gamma_1 &= tr\left(d(\Omega^{-1})d\Omega + \Omega^{-1}d^2\Omega\right) = tr\left(-\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega + \Omega^{-1}d^2\Omega\right) \\ &= -tr\left(\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\right) + tr\left(\Omega^{-1}d^2\Omega\right), \end{aligned}$$

$$\begin{aligned} \gamma_2 &= d\left(u'\Omega^{-1}du\right) = d\left(u'\Omega^{-1}du\right)_u + d\left(u'\Omega^{-1}du\right)_\Omega \\ &= d\left(u'\Omega^{-1}\right)_u du + u'\Omega^{-1}d\left(du\right)_u + u'd\left(\Omega^{-1}\right)_\Omega du \\ &= du'\Omega^{-1}du + 0 - u'\Omega^{-1}d\Omega\cdot d\Omega^{-1}du \end{aligned}$$

and

$$\begin{aligned} \gamma_3 &= d\left(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\right) = d\left(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\right)_u + d\left(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\right)_\Omega \\ &= 2u'\Omega^{-1}d\Omega\cdot\Omega^{-1}du + \gamma_3^a, \end{aligned}$$

where

$$\begin{aligned} \gamma_3^a &= d\left(u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u\right)_\Omega = d\left(u'\Omega^{-1}\right)_\Omega d\Omega\cdot\Omega^{-1}u + u'\Omega^{-1}d\left(d\Omega\cdot\Omega^{-1}u\right)_\Omega \\ &= d\left(u'\Omega^{-1}\right)_\Omega d\Omega\cdot\Omega^{-1}u + \gamma_3^b = -u'\Omega^{-1}d\Omega.\Omega^{-1}d\Omega.\Omega^{-1}u + \gamma_3^b \end{aligned}$$

with

$$\begin{aligned} \gamma_3^b &= u'\Omega^{-1}d\left(d\Omega\cdot\Omega^{-1}u\right)_\Omega = u'\Omega^{-1}\left[d^2\Omega\cdot\Omega^{-1}u + d\Omega\cdot d(\Omega^{-1})u\right] \\ &= u'\Omega^{-1}d^2\Omega\cdot\Omega^{-1}u - u'\Omega^{-1}d\Omega\cdot\Omega^{-1}d\Omega\cdot\Omega^{-1}u. \end{aligned}$$

We deduce the final result (3.4). □

3.3 Expectation of $d^2\ell$

By taking the expectation of $-d^2\ell(\theta \mid u)$, we obtain after some algebra the following Lemma.

Lemma 3.3. *Assuming that $|d^2\ell(\theta \mid u)|$ is integrable,*

$$\mathbb{E}\left(-d^2\ell(\theta \mid u)\right) = \frac{1}{2}tr\left(\Omega^{-1}d\Omega.\Omega^{-1}d\Omega\right) + d\beta'X'\Omega^{-1}Xd\beta. \tag{3.5}$$

Proof of Lemma 3.3: The proof uses the equality

$$\mathbb{E}(u'Au) = \mathbb{E}\left(\mathrm{tr}(u'Au)\right) = \mathbb{E}\left[\mathrm{tr}(Auu')\right] = \mathrm{tr}\left(\mathbb{E}\left[(Auu')\right]\right) = A\mathrm{tr}(\Omega),$$

where u is a random vector and A is a matrix of constant terms. The expectation of $d^2\ell$ is then given by

$$\mathbb{E}\left(d^2\ell(\theta \mid u)\right) = \frac{1}{2}\kappa_0 - \frac{1}{2}\kappa_1 - \kappa_2 + \frac{1}{2}\kappa_3 + \kappa_4 - \kappa_5,$$

where

$$\kappa_0 = tr\left(\Omega^{-1}d\Omega \cdot \Omega^{-1}d\Omega\right),\ \kappa_1 = tr\left(\Omega^{-1}d^2\Omega\right),$$

$$\begin{aligned}\kappa_2 &= \mathbb{E}\big[u'\Omega^{-1}d\Omega \cdot \Omega^{-1}d\Omega \cdot \Omega^{-1}u\big] = tr\big[\Omega^{-1}d\Omega \cdot \Omega^{-1}d\Omega \cdot \Omega^{-1}\mathbb{E}(uu')\big] \\ &= tr\big[\Omega^{-1}d\Omega \cdot \Omega^{-1}d\Omega \cdot \Omega^{-1}\Omega\big] = tr\big[\Omega^{-1}d\Omega \cdot \Omega^{-1}d\Omega\big],\end{aligned}$$

$$\kappa_3 = \mathbb{E}\big[u'\Omega^{-1}d^2\Omega \cdot \Omega^{-1}u\big] = tr\big[\Omega^{-1}d^2\Omega\big],$$

$$\begin{aligned}\kappa_4 &= \mathbb{E}\big[u'\Omega^{-1}d\Omega \cdot \Omega^{-1}du\big] = -\mathbb{E}\big[u'\Omega^{-1}d\Omega \cdot \Omega^{-1}X\mathrm{d}\beta\big] \\ &= -\mathbb{E}\big[u\big]'\Omega^{-1}d\Omega \cdot \Omega^{-1}X\mathrm{d}\beta = 0,\end{aligned}$$

and

$$\kappa_5 = \mathbb{E}\big[du'\Omega^{-1}du\big] = \mathbb{E}\big[d\beta'X'\Omega^{-1}Xd\beta\big].$$

We obtain that $\kappa_1 = \kappa_3$, $\kappa_4 = 0$ and $\kappa_0 = \kappa_2$. We deduce that $\mathbb{E}\left(d^2\ell(\theta \mid u)\right) = -\frac{1}{2}\kappa_0 - \kappa_5$ and $\mathbb{E}\left(-d^2\ell(\theta \mid u)\right) = \frac{1}{2}\kappa_0 + \kappa_5$ which is the desired result. □

The following is an important Lemma based on symmetric matrices.

Lemma 3.4. *If A is a symmetric and square matrix of order $p+q$, for a given vector $z = (x, y) \in \mathbb{R}^{p+q}$, where $x = (x_1, ..., x_p)' \in \mathbb{R}^p$, $y = (y_1, ..., y_q)' \in \mathbb{R}^q$, we have the following equality*

$$z'Az = x'A_px + 2x'A_{p,q}y + y'A_qy,$$

where

$$A = \begin{bmatrix} A_p & A_{p,q} \\ A'_{p,q} & A_q \end{bmatrix},$$

where A_p, A_q and $A_{p,q}$ are matrices of dimensions of (p,p), (q,q) and (p,q) respectively.

Proof of Lemma 3.4: The proof of this Lemma is straightforward. □

4 Main results

We now turn to the main results obtained in this paper. The vector of parameters is denoted by $\theta = (\theta_1, ..., \theta_r)'$, where $\theta_1 = \sigma^2_\nu$; $\theta_2 = \sigma^2_\mu$; $\theta_3 = \sigma^2_\lambda$; $\left(\theta_{j+3} = \theta_{\nu,j}\right)_{1\le j\le p}$; $\left(\theta_{j+p+3} = \theta_{\mu,j}\right)_{1\le j\le q}$ and $\left(\theta_{q+p+3+j} = \beta_j\right)_{1\le j\le k}$. We denote by $\overline{\theta}$ the vector of dimension $(p+q+3)$ defined by $\overline{\theta}_j = \theta_j$ for $j = 1, ..., p+q+3$. The following proposition gives the relationship between the FIM at θ, the FIM at $\overline{\theta}$ and the FIM at β.

Proposition 4.1. *If $d\theta = (d\overline{\theta}', d\beta')'$ with $d\beta = (d\beta_1, ..., d\beta_k)' \in \mathbb{R}^k$ and $d\overline{\theta} = (d\sigma^2_\nu, d\sigma^2_\mu, d\sigma^2_\lambda, \{d\theta_{\nu,j}\}_{j=1,...,p}, \{d\theta_{\mu,j}\}_{j=1,...,q})' \in \mathbb{R}^{p+q+3}$, we have*

$$(d\theta)' I_n(\theta)(d\theta) = (d\overline{\theta})' I_n(\overline{\theta})(d\overline{\theta}) + 2(d\overline{\theta})' I_n(\overline{\theta}, \beta)(d\beta) + (d\beta)' I_n(\beta)(d\beta), \tag{4.1}$$

where

$$I_n(\theta) = \begin{bmatrix} I_n(\overline{\theta}) & I_n(\overline{\theta}, \beta) \\ I_n(\overline{\theta}, \beta) & I_n(\beta) \end{bmatrix}$$

and

$$I_n(\overline{\theta}) = \mathbb{E}\left[-\frac{\partial^2 \ell(u \mid \theta)}{\partial\overline{\theta}_i \partial\overline{\theta}_j}\right]_{1\le i,j\le p+q+3}, \quad I_n(\beta) = \mathbb{E}\left[-\frac{\partial^2 \ell(u \mid \theta)}{\partial\beta_i \partial\beta_j}\right]_{1\le i,j\le k},$$
$$I_n(\overline{\theta}, \beta) = \mathbb{E}\left[-\frac{\partial^2 \ell(u \mid \theta)}{\partial\overline{\theta}_i \partial\beta_j}\right]_{\substack{1\le j\le k \\ 1\le i\le p+q+3}}.$$

Proof of Proposition 4.1: Using Lemma 3.4, the proof is straightforward. □

At this stage, computing the FIM requires the derivation of tr $\left(\Omega^{-1}d\Omega.\Omega^{-1}d\Omega\right)$. By multiplying $d\Omega$ (given in (3.4)) with the expression of Ω^{-1} from equation (2.8), we obtain

$$\begin{aligned}
\Omega^{-1}d\Omega \;=\;& \underbrace{[C_1 D_\nu \otimes E_T + C_2 D_\nu \otimes \overline{J}_T]}_{\Omega_1} d\sigma^2_\nu + \underbrace{\left(C_2 D_\mu \otimes T\overline{J}_T\right)}_{\Omega_2} d\sigma^2_\mu \\
&+ \underbrace{[C_1 J_N \otimes E_T + C_2 J_N \otimes \overline{J}_T]}_{\Omega_3} d\sigma^2_\lambda + \sum_{j_1=1}^{p} \underbrace{\left(\sigma^2_\nu (C_1 D^*_{\nu,j_1} \otimes E_T + C_2 D^*_{\nu,j_1} \otimes \overline{J}_T)\right)}_{\Omega_{3+j_1}} d\theta_{\nu,j_1} \\
&+ \sum_{j_2=1}^{q} \underbrace{\left(\sigma^2_\mu (C_2 D^*_{\mu,j_2} \otimes T\overline{J}_T)\right)}_{\Omega_{p+3+j_2}} d\theta_{\mu,j_2} = \sum_{j=1}^{p+q+3} \Omega_j d\overline{\theta}_j = \sum_{j=1}^{r-k} \Omega_j d\overline{\theta}_j.
\end{aligned}$$

We deduce that the trace of $\Omega^{-1}d\Omega.\Omega^{-1}d\Omega$ is

$$\mathrm{tr}\left(\Omega^{-1}d\Omega.\Omega^{-1}d\Omega\right)=\sum_{i=1}^{r-k}\sum_{j=1}^{r-k}\mathrm{tr}\left(\Omega_i\Omega_j\right)d\overline{\theta}_i d\overline{\theta}_j.$$

From Lemma 3.3, the expectation of $-d^2\ell(u\mid\theta)$ can be written as

$$\mathbb{E}_\theta\left[-d^2\ell(u\mid\theta)\right]=\sum_{i=1}^{r-k}\sum_{j=1}^{r-k}\frac{1}{2}\mathrm{tr}\left(\Omega_i\Omega_j\right)d\overline{\theta}_i d\overline{\theta}_j+(d\beta)'(X'\Omega^{-1}X)(d\beta). \tag{4.2}$$

By comparing equations (3.1), (4.1) and (4.2), we deduce that

$$I_n(\overline{\theta},\beta)=\mathrm{O}\text{ with }I_n(\beta)=X'\Omega^{-1}X.$$

The derivation of the FIM $I_n(\theta)$ is then based on the derivation of $I_n(\overline{\theta})$. From equation (4.2), the terms of $I_n(\overline{\theta})$ are given by $I_n(\overline{\theta})_{i,j}=a_{i,j}=\frac{1}{2}\mathrm{tr}\left(\Omega_i\Omega_j\right)$ or in matrix form as

$$I_n(\overline{\theta})=\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \boxed{\mathbf{a}_{1,4}}^{1\times p} & \boxed{\mathbf{a}_{1,5}}^{1\times q}\\ a_{2,1} & a_{2,2} & a_{2,3} & \boxed{\mathbf{a}_{2,4}}^{1\times p} & \boxed{\mathbf{a}_{2,5}}^{1\times q}\\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{\mathbf{a}_{3,4}}^{1\times p} & \boxed{\mathbf{a}_{3,5}}^{1\times q}\\ \boxed{\mathbf{a}_{4,1}}^{p\times 1} & \boxed{\mathbf{a}_{4,2}}^{p\times 1} & \boxed{\mathbf{a}_{4,3}}^{p\times 1} & \boxed{\mathbf{a}_{4,4}}^{p\times p} & \boxed{\mathbf{a}_{4,5}}^{p\times q}\\ \boxed{\mathbf{a}_{5,1}}^{q\times 1} & \boxed{\mathbf{a}_{5,2}}^{q\times 1} & \boxed{\mathbf{a}_{5,3}}^{q\times 1} & \boxed{\mathbf{a}_{5,4}}^{q\times p} & \boxed{\mathbf{a}_{5,5}}^{q\times q} \end{pmatrix}.$$

$I_n(\overline{\theta})$ is a symmetric block matrix where the terms in boxes are also matrices with the dimensions indicated at the top of the boxes. We need to evaluate the components of $\overline{\theta}$:

$$\begin{aligned}
&a_{i,j}=a_{j,i}=\frac{1}{2}\mathrm{tr}(\Omega_i\Omega_j)\text{ for }i,j\in\{1,2,3\};\\
&\mathbf{a}_{i,4}(j_1)=\mathbf{a}'_{4,i}(j_1)=\frac{1}{2}\mathrm{tr}(\Omega_i\Omega_{j_1+3})\text{ for }i=1:3\text{ and for }j_1=1:p;\\
&\mathbf{a}_{i,5}(j_2)=\mathbf{a}'_{5,i}(j_2)=\frac{1}{2}\mathrm{tr}(\Omega_i\Omega_{p+3+j_2})\text{ for }i=1:3\text{ and for }j_2=1:q;\\
&\mathbf{a}_{4,4}(j_1,j_2)=\frac{1}{2}\mathrm{tr}(\Omega_{j_1+3}\Omega_{j_2+3})\text{ for }1\le j_1,j_2\le p;\\
&\mathbf{a}_{4,5}(j_1,j_2)=\mathbf{a}'_{5,4}(j_1,j_2)=\frac{1}{2}\mathrm{tr}(\Omega_{j_1+3}\Omega_{j_2+p+3})\text{ for }1\le j_1\le p,1\le j_2\le q;\\
&\mathbf{a}_{5,5}(j_1,j_2)=\frac{1}{2}\mathrm{tr}(\Omega_{j_1+p+3}\Omega_{j_2+p+3})\text{ for }1\le j_1,j_2\le q.
\end{aligned}$$

Remark 4.1. In order to evaluate the components of $\overline{\theta}$, we will use the fact that $E_T\overline{J}_T=\mathrm{O}$, $E_T^2=E_T$, $\mathrm{tr}\big(\overline{J}_T\big)=1$ and $\mathrm{tr}\big(E_T\big)=T-1$.

From the Remark 4.1 we will show how the components of $\overline{\theta}$ can be written in terms of C_1, C_2, J_N, D_ν and D_μ.

To illustrate the computation of $a_{i,j}$, we give, for example, the details of calculating $a_{1,1}$. By the definition, $a_{1,1} = \frac{1}{2}\text{tr}(\Omega_1^2)$ with $\Omega_1^2 = C_1 D_\nu C_1 D_\nu \otimes E_T^2 + C_1 D_\nu C_2 D_\nu \otimes E_T \overline{J}_T + C_2 D_\nu C_1 D_\nu \otimes \overline{J}_T E_T + C_2 D_\nu C_2 D_\nu \otimes \overline{J}_T^2$. Now using the fact that $E_T \overline{J}_T =$ O, $E_T^2 = E_T$, $\overline{J}_T^2 = \overline{J}_T$, $\text{tr}\big(\overline{J}_T\big) = 1$, $\text{tr}\big(E_T\big) = T - 1$ and taking the trace of Ω_1^2, we obtain $a_{1,1} = \frac{(T-1)}{2}\text{tr}\big[C_1 D_\nu C_1 D_\nu\big] + \frac{1}{2}\text{tr}\big[C_2 D_\nu C_2 D_\nu\big]$. The other coefficients are obtained in the same way. The coefficients are given in Appendix A.

Remark 4.2. A quick inspection of the elements of $I_n(\overline{\theta})$ in Appendix B reveals that they are all written as linear combinations of $\text{tr}\big(C_1 M_1 C_2 M_2\big)$, where $M_1 \in \{\text{diag}(\eta_i), J_N\}$ and $M_2 \in \{\text{diag}(\psi_i), J_N\}$ [i.e. either diagonal matrices or square matrices of ones].

To have the final expressions of $a_{i,j}$, it suffices to evaluate the quantity $\text{tr}\big(C_1 M_1 C_2 M_2\big)$ in the following cases:

- **Case 1**: $\text{tr}\big(C_1 M_1 C_2 M_2\big)$, where $M_1 = diag(\eta_i)$ and $M_2 = diag(\psi_i)$ are diagonal matrices.
- **Case 2**: $\text{tr}\big(C_1 M_1 C_2 M_2\big)$, where $M_1 = J_N$ and $M_2 = diag(\psi_i)$ or $M_1 = diag(\eta_i)$ and $M_2 = J_N$, that is one is a diagonal matrix while the other is a square matrix of ones.
- **Case 3**: $\text{tr}\big(C_1 M_1 C_2 M_2\big)$, where $M_1 = J_N$ and $M_2 = J_N$ are all square matrices of ones.

Before evaluating the three previous quantities above, we make the following important remark to ease the calculation of traces.

Remark 4.3. Using the expressions of C_1 and C_2 given by equation (2.9), we can prove that

$$C_l = D_l^{-1} - \frac{\sigma_\lambda^2 D_l^{-1} J_N D_l^{-1}}{1 + \sigma_\lambda^2 i_N' D_l^{-1} i_N} = \varsigma(D_l, 0), l = 1, 2$$

with $D_1 = \text{diag}(\vartheta_i^{-1})$ and $D_2 = \text{diag}(\phi_i^{-1})$.

In fact, by definition of C_1 and C_2, we have $C_1 = \varsigma(\sigma_\nu^2 D_\nu, 0)$ and $C_2 = \varsigma(\sigma_\nu^2 D_\nu, \sigma_\mu^2 D_\mu)$ which can also be written as $C_2 = \varsigma(\sigma_\nu^2 D_\nu + T\sigma_\mu^2 D_\mu, 0)$. We then deduce that

$$\vartheta_i^{-1} = \sigma_\nu^2 h_\nu(w_i'\theta_\nu) \text{ and } \phi_i^{-1} = \sigma_\nu^2 h_\nu(w_i'\theta_\nu) + T\sigma_\mu^2 h_\mu(z_i'\theta_\mu). \tag{4.3}$$

We can also observe that in the case of homokedasticity, $\vartheta_i = 1/\sigma_\nu^2$ and $\phi_i = 1/(\sigma_\nu^2 + T\sigma_\mu^2)$.

5 Some computational issues

If we derive a specific formula for the calculation of $tr\big(C_1M_1C_2M_2\big)$, it will definitively ease the computation of the FIM $I_n(\overline{\theta})$ and thereby that of the FIM $I_n(\theta)$. We consider matrices D_1, D_2, C_1, C_2, M_1 and M_2 as defined earlier. According to the definition of M_1 and M_2, $\mathrm{tr}\big(C_1M_1C_2M_2\big)$ is obtained through the real valued functions $\Psi_i(i=1,2,3)$ given by the following propositions.

Proposition 5.1. *If $M_1 = diag(\eta_i) \neq i_Ni'_N$ and $M_2 = diag(\psi_i) \neq i_Ni'_N$ then*

$$tr\big(C_1M_1C_2M_2\big) = \Psi_1(\vartheta,\eta,\phi,\psi) = <\eta\odot\vartheta,\psi\odot\phi> - \frac{\sigma_\lambda^2 <\eta\odot\vartheta,\psi\odot\phi\odot\phi>}{1+\sigma_\lambda^2 i'_N\phi} - \frac{\sigma_\lambda^2 <\eta\odot\vartheta\odot\vartheta,\psi\odot\phi>}{1+\sigma_\lambda^2 i'_N\vartheta} + \frac{\sigma_\lambda^4 <\eta,\vartheta\odot\phi><\psi,\vartheta\odot\phi>}{(1+\sigma_\lambda^2 i'_N\vartheta)(1+\sigma_\lambda^2 i'_N\phi)} \tag{5.1}$$

with $<,>$ and $\odot$ denoting the inner and Hardamar products, respectively.

Proof of Proposition 5.1: The proof follows since

$$\begin{aligned}
\mathrm{tr}\big(C_1M_1C_2M_2\big) &= tr\left[\left(D_1^{-1}M_1 - \frac{\sigma_\lambda^2 D_1^{-1}i'_Ni_ND_1^{-1}M_1}{1+\sigma_\lambda^2 i'_ND_1^{-1}i_N}\right)\left(D_2^{-1}M_2 - \frac{\sigma_\lambda^2 D_2^{-1}i'_Ni_ND_2^{-1}M_2}{1+\sigma_\lambda^2 i'_ND_2^{-1}i_N}\right)\right]\\
&= tr\left(D_1^{-1}M_1D_2^{-1}M_2\right) - \frac{\sigma_\lambda^2 i'_ND_2^{-1}M_2D_1^{-1}M_1D_2^{-1}i_N}{1+\sigma_\lambda^2 i'_ND_2^{-1}i_N}\\
&\quad -\frac{\sigma_\lambda^2 i'_ND_1^{-1}M_1D_2^{-1}M_2D_1^{-1}i_N}{1+\sigma_\lambda^2 i'_ND_1^{-1}i_N} + \frac{\sigma_\lambda^4\left(i'_ND_1^{-1}M_1D_2^{-1}i_N\right)\left(i'_ND_2^{-1}M_2D_1^{-1}i_N\right)}{\left(1+\sigma_\lambda^2 i'_ND_1^{-1}i_N\right)\left(1+\sigma_\lambda^2 i'_ND_2^{-1}i_N\right)}\\
&= \sum_{i=1}^N \vartheta_i\eta_i\psi_i\phi_i - \frac{\sigma_\lambda^2\sum_{i=1}^N \eta_i\vartheta_i\psi_i\phi_i^2}{1+\sigma_\lambda^2\sum_{i=1}^N\phi_i} - \frac{\sigma_\lambda^2\sum_{i=1}^N \eta_i\vartheta_i^2\psi_i\phi_i}{1+\sigma_\lambda^2\sum_{i=1}^N\vartheta_i}\\
&\quad + \frac{\sigma_\lambda^4\left(\sum_{i=1}^N\vartheta_i\eta_i\phi_i\right)\left(\sum_{i=1}^N\vartheta_i\psi_i\phi_i\right)}{\left(1+\sigma_\lambda^2\sum_{i=1}^N\vartheta_i\right)\left(1+\sigma_\lambda^2\sum_{i=1}^N\phi_i\right)}\\
&= <\eta\odot\vartheta,\psi\odot\phi> - \frac{\sigma_\lambda^2 <\eta\odot\vartheta,\psi\odot\phi\odot\phi>}{1+\sigma_\lambda^2 i'_N\phi}\\
&\quad - \frac{\sigma_\lambda^2 <\eta\odot\vartheta\odot\vartheta,\psi\odot\phi>}{1+\sigma_\lambda^2 i'_N\vartheta} + \frac{\sigma_\lambda^4 <\eta,\vartheta\odot\phi><\psi,\vartheta\odot\phi>}{(1+\sigma_\lambda^2 i'_N\vartheta)(1+\sigma_\lambda^2 i'_N\phi)}.
\end{aligned}$$

□

Proposition 5.2. *If $M_1 = diag(\eta_i)$ and $M_2 = i_Ni'_N$, then*

$$tr\left(C_1M_1C_2M_2\right) = \Psi_2(\vartheta,\eta,\phi) = \frac{<\eta,\vartheta\odot\phi>}{\left(1+\sigma_\lambda^2 i'_N\vartheta\right)\left(1+\sigma_\lambda^2 i'_N\phi\right)}. \tag{5.2}$$

For $M_1 = i_Ni'_N$ and $M_2 = diag(\psi_i)$, we just need to interchange the subscripts 1 and 2 leading to $\Psi_2(\eta,\vartheta,\phi)$.

Proof of Proposition 5.2: We have

$$\begin{aligned}\mathrm{tr}\big(C_1M_1C_2M_2\big) \;=\; & tr\left(i_N'D_1^{-1}M_1D_2^{-1}i_N\right) - \frac{\sigma_\lambda^2(i_N'D_2^{-1}i_N)(i_N'D_1^{-1}M_1D_2^{-1}i_N)}{1+\sigma_\lambda^2 i_N'D_2^{-1}i_N} \\ & - \frac{\sigma_\lambda^2(i_N'D_1^{-1}M_1D_2^{-1}i_N)(i_N'D_1^{-1}i_N)}{1+\sigma_\lambda^2 i_N'D_1^{-1}i_N} \\ & + \frac{\sigma_\lambda^4\left((i_N'D_1^{-1}M_1D_2^{-1}i_N)\right)\left(i_N'D_2^{-1}i_N\right)\left(i_N'D_1^{-1}i_N\right)}{\left(1+\sigma_\lambda^2 i_N'D_1^{-1}i_N\right)\left(1+\sigma_\lambda^2 i_N'D_2^{-1}i_N\right)}.\end{aligned}$$

Now if $B = i_N'D_1^{-1}M_1D_2^{-1}i_N$ (which is a scalar), observing that $i_N'D_1^{-1}i_N = i_N'\vartheta$ and $i_N'D_2^{-1}i_N = i_N'\phi$, we have

$$\begin{aligned}\mathrm{tr}\big(C_1M_1C_2M_2\big) \;=\; & B - \frac{\sigma_\lambda^2(i_N'\phi)B}{1+\sigma_\lambda^2 i_N'\phi} - \frac{\sigma_\lambda^2 B(i_N'\vartheta)}{1+\sigma_\lambda^2 i_N'\vartheta} + \frac{\sigma_\lambda^4 B\,(i_N'\phi)\,(i_N'\vartheta)}{\left(1+\sigma_\lambda^2 i_N'\vartheta\right)\left(1+\sigma_\lambda^2 i_N'\phi\right)} \\ =\; & \frac{B}{\left(1+\sigma_\lambda^2 i_N'\vartheta\right)\left(1+\sigma_\lambda^2 i_N'\phi\right)} = \frac{<\eta,\vartheta\odot\phi>}{\left(1+\sigma_\lambda^2 i_N'\vartheta\right)\left(1+\sigma_\lambda^2 i_N'\phi\right)}.\end{aligned}$$

We deduce that for $M_1 = i_Ni_N'$ and $M_2 = diag(\psi_i)$,

$$\begin{aligned}tr\left(C_1M_1C_2M_2\right) \;=\; & tr\left(C_1M_2C_2M_1\right) \\ =\; & \frac{<\psi,\vartheta\odot\phi>}{(1+\sigma_\lambda^2 i_N'\phi)(1+\sigma_\lambda^2 i_N'\vartheta)}\end{aligned}$$

since $tr(IJ) = tr(JI)$ for any matrices I and J where the matrix product holds.

□

Proposition 5.3. *If $M_1 = i_Ni_N' = M_2$, then*

$$tr\left(C_1M_1C_2M_2\right) = \Psi_3(\vartheta,\phi) = \frac{(i_N'\phi)(i_N'\vartheta)}{(1+\sigma_\lambda^2 i_N'\phi)(1+\sigma_\lambda^2 i_N'\vartheta)}. \tag{5.3}$$

Proof of Proposition 5.3: Since for the case $M_2 = i_Ni_N'$ the result depends on M_1 through B (B is the matrix used in the proof of Proposition 5.2), replacing M_1 by i_Ni_N' in the expression of B, we immediately obtain

$$\begin{aligned}tr\left(C_1M_1C_2M_2\right) \;=\; & \frac{i_N'D_2^{-1}M_1D_1^{-1}i_N}{(1+\sigma_\lambda^2 i_N'D_2^{-1}i_N)(1+\sigma_\lambda^2 i_N'D_1^{-1}i_N)} \\ =\; & \frac{(i_N'D_2^{-1}i_N)(i_N'D_1^{-1}i_N)}{(1+\sigma_\lambda^2 i_N'D_2^{-1}i_N)(1+\sigma_\lambda^2 i_N'D_1^{-1}i_N)} = \frac{(i_N'\phi)(i_N'\vartheta)}{(1+\sigma_\lambda^2 i_N'\phi)(1+\sigma_\lambda^2 i_N'\vartheta)}.\end{aligned}$$

□

Before stating the important result that gives the final expression of the FIM, we introduce the following notations:

$$\begin{aligned} D_\nu &= \operatorname{diag}(h_\nu(w_i'\theta_\nu)) = \operatorname{diag}(\eta_i) \text{ with } \eta_i = h_\nu(w_i'\theta_\nu);\\ D_\mu &= \operatorname{diag}(h_\mu(z_i'\theta_\mu)) = \operatorname{diag}(\eta_i) \text{ with } \psi_i = h_\mu(z_i'\theta_\mu);\\ D^*_{\nu,j} &= \operatorname{diag}(h'_\nu(\omega_i'\theta_\nu)w_{i,j}) = \operatorname{diag}(\eta_i^j) \text{ with } \eta_i^j = h'_\nu(\omega_i'\theta_\nu)\omega_{i,j};\\ D^*_{\mu,j} &= \operatorname{diag}(h'_\mu(z_i'\theta_\mu)z_{i,j}) = diag(\psi_i^j) \text{ with } \psi_i^j = h'_\mu(z_i'\theta_\mu)z_{i,j};\\ C_1 &= \varsigma(D_1, 0) \text{ with } D_1 = \operatorname{diag}(\vartheta_i^{-1});\\ C_2 &= \varsigma(D_2, 0) \text{ with } D_2 = \operatorname{diag}(\phi_i^{-1}); \end{aligned}$$

where $\vartheta_i^{-1} = \sigma_\nu^2 h_\nu(w_i'\theta_\nu)$ and $\phi_i^{-1} = \sigma_\nu^2 h_\nu(w_i'\theta_\nu) + T\sigma_\mu^2 h_\mu(z_i'\theta_\mu)$. We also set $\eta^j = (\eta_i^j)_{i=1}^N$; $\eta = (\eta_i)_{i=1}^N$; $\psi^j = (\psi_i^j)_{i=1}^N$; $\psi = (\psi_i)_{i=1}^N$; $\phi = (\phi_i)_{i=1}^N$ and $\vartheta = (\vartheta_i)_{i=1}^N$.

Proposition 5.4. *If y is a two-way error components model in the form $y = X\beta + u$, where u satisfies assumption $\mathcal{A}_1$, the FIM evaluated at θ is*

$$I_n(\theta) = \begin{bmatrix} I_n(\overline{\theta}) & O \\ \\ O & X'\Omega^{-1}X \end{bmatrix},$$

where the coefficients of $I_n(\overline{\theta})$ described in Appendix B are now written in terms of $\Psi_1(\vartheta, \eta, \phi, \psi)$, $\Psi_2(\vartheta, \eta, \phi)$ and $\Psi_3(\vartheta, \phi)$. These expressions are given in Appendix C.

Proof of Proposition 5.4: The proof is straightforward, see Appendix B. □

We can summarize the use of our method in practice. Given a two way error components model, the computation of the FIM is based on the following steps:

Step 1: Identify $\eta = (\eta_i)$, $\psi = (\psi_i)$ which are the vectors derived from $D_\nu = \operatorname{diag}(\eta_i)$ and $D_\mu = \operatorname{diag}(\psi_i)$;

Step 2: Compute Ω (from equation (3.4)) and then Ω^{-1};

Step 3: Deduce values of ϑ_i and ϕ_i from equation (4.3);

Step 4: Identify $\eta^* = (\eta_i^j)$ and $\psi^* = (\psi_i^j)$ which are the matrices obtained from $D^*_{\nu,j} = \operatorname{diag}(\eta_i^j)$ for $j = 1, ..., p$ and $D^*_{\mu,j} = \operatorname{diag}(\psi_i^j)$ for $j = 1, ..., q$;

Step 5: Obtain values of the parameters $\theta \in \mathbb{R}^{p+q+3+k}$ (which can be estimated using some observations);

Step 6: Observations $x_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$ are needed to compute $X'\Omega^{-1}X$.

The above information is input in our code named **FIM.FUN**. The code has been written in the R language and is available as supplementary material.

6 Two real data examples

We now consider two real data examples to illustrate the above analysis. For consistency and comparative purposes, asymptotic as well as exact results are obtained in the homoscedasticity case while only exact results are obtained in the heteroscedasticity case.

6.1 Example 1: Public capital productivity (Homoscedasticity case)

6.1.1 Model specification

Following [23, 6], we re-consider the following Cobb-Douglas production relationship investigating the productivity of public capital in private production,

$$\ln(y_{it}) = \beta_0 + \beta_1 \ln(Pc_{it}) + \beta_2 \ln(PS_{it}) + \beta_3 \ln(L_{it}) + \beta_4 U_{\mathrm{emp}_{it}} + \mu_i + \lambda_t + \varepsilon_{it}, \tag{6.1}$$

where y_{it} = gross state product; Pc_{it} = public capital; PS_{it} = private capital; L_{it} = labour input as payrolls; $U_{\mathrm{emp}_{it}}$ = unemployment rate; $\beta_0, ..., \beta_4$ = coefficients to be estimated; μ_i = the unobservable individual effect; ε_{it} = the rest of the perturbation. Data are from 48 US states (i.e $N = 48$) observed over the period 1970 to 1986, (i.e $T = 7$). The data are obtained from the Wiley web site at www.wiley.com/go/baltagi3e. Following a common unjustified practice [1],[6, 15, 23, 30] and [34], we assume that errors are homoskedastic. We consider estimating the above model based on four estimators: (i) Swamy and Arora (residuals obtained from solving a system of 3 equations); (ii) Wallace and Hussain (OLS residuals); (iii) Wansbeek and Kapteyn (LSDV residuals); (iv) maximum likelihood (ML residuals). Results based on the following restrictions (i) ($\beta_1 = 0$)(one restriction); (ii)($\beta_1 = 0, \beta_2 = 0$); ($\beta_1 = 0, \beta_3 = 0$); ($\beta_1 = 0, \beta_4 = 0$) (two restrictions); and (iii) ($\beta_1 = 0, \beta_2 = 0, \beta_3 = 0$); ($\beta_1 = 0, \beta_2 = 0, \beta_4 = 0$) and ($\beta_1 = 0, \beta_3 = 0, \beta_4 = 0$) (three restrictions) are reported below.

6.1.2 Results

In Table 1a estimations are done based on the FIM based on the observed information matrix. Rather, in Table 1b we used the FIM which relies on exact information matrix developed by the authors.

The results are based on the FIM with homoscedastic errors. In Table 1a, results clearly indicate that irrespective of the estimator used and no matter how many restrictions are used, public capital remains important and productive in private production. The results are consistent with the study by [6] and [23]. In Table 1b the FIM is now based on exact information matrix. Comparing Table 1a and 1b, we notice that results are close. This means that asymptotic results based on the Central Limit Theorem adequately approximate our results in many cases.

6.1.3 Discussion

Some inconsistencies still exist when applying the linear restriction tests. For example, the number of times some test results are not available (i.e., NA) remain relatively high. Possible reason could be that the sample size is not very large. This could explain why the Wald, Likelihood ratio (LR) and Lagrange multiplier (LM) tests give negative values.

One restriction($\beta_1 = 0$)			Two restrictions($\beta_1 = 0,\ \beta_2 = 0$)			Two restrictions($\beta_1 = 0,\ \beta_3 = 0$)			Two restrictions($\beta_1 = 0,\ \beta_4 = 0$)		
λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}
Swamy and Arora											
26.074^a	25.662^a	25.266^a	1443.184^a	803.969^a	521.267^a	8103.542^a	1951.533^a	741.348^a	105.407^a	99.134^a	93.348^a
Wansbeek and Kapteyn											
NA	NA	NA	1189.133^a	733.625^a	483.924^a	10132.516^a	2187.806^a	755.183^a	NA	NA	NA
Wallace and Hussain											
35.625^a	34.866^a	34.134^a	1457.202^a	836.017^a	523.084^a	9583.345^a	2076.788^a	751.971^a	84.680^a	80.569^a	76.719^a
Maximum likelihood											
36.574^a	35.778^a	35.005^a	1505.446^a	853.161^a	529.174^a	10441.093^a	2141.460^a	756.850^a	87.307^a	82.945^a	78.868^a

Three restrictions($\beta_1 = 0,\ \beta_2 = 0, \beta_3 = 0$)			Three restrictions($\beta_1 = 0,\ \beta_2 = 0, \beta_4 = 0$)			Three restrictions($\beta_1 = 0,\ \beta_3 = 0, \beta_4 = 0$)		
λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}
Swamy and Arora								
92545.656^a	3867.694^a	808.867^a	1395.961^a	813.732^a	514.974^a	9656.094^a	2082.476^a	752.416^a
Wansbeek and Kapteyn								
69599.274^a	3637.524^a	806.543^a	1000.809^a	653.145^a	449.502^a	8278.811^a	1967.413^a	742.787^a
Wallace and Hussain								
91481.904^a	3858.343^a	808.785^a	1411.693^a	819.515^a	517.100^a	9316.481^a	2055.575^a	750.285^a
Maximum likelihood								
89097.000^a	3836.985^a	808.594^a	1402.137^a	816.007^a	515.813^a	10012.449^a	2109.782^a	754.508^a

Table 1a: Restriction test results - FIM based on observed information matrix [Homoscedasticity case].

Notes: λ_W, λ_{LR} and λ_{LM} represent the Wald, likelihood and Lagrange multiplier statistics respectively. NA=not available. ^{a},^{b} and ^{c} represent statistical significant at 1%, 5% and 10% repectively.

One restriction($\beta_1 = 0$)			Two restrictions($\beta_1 = 0,\ \beta_2 = 0$)			Two restrictions($\beta_1 = 0,\ \beta_3 = 0$)			Two restrictions($\beta_1 = 0, \beta_4 = 0$)		
λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}
Swamy and Arora											
692.810^a	NA	$1.302e+11^a$	1709.655^a	NA	232103.990^a	861.639^a	575.066^a	867.050^a	1437.169^a	NA	235251.998^a
Wansbeek and Kappteyn											
242.551^a	236.413^a	4408906.542^a	336.928^a	NA	65958346.732^a	353.549^a	4991.859	$1.7904e+09^a$	272.816^a	2276.679^a	1141660.516^a
Wallace and Hussain											
370.771^a	383.875^a	23518.283^a	564.807^a	NA	609594.625^a	487.433^a	321.686^a	74512.198^a	452.761^a	NA	214563.054^a
Maximum likelihood											
377.302^a	NA	103473.807^a	570.571^a	NA	6709213.013^a	488.120^a	NA	158524.010^a	459.057^a	NA	173961.580^a

Three restrictions($\beta_1 = 0,\ \beta_2 = 0, \beta_3 = 0$)			Three restrictions($\beta_1 = 0,\ \beta_2 = 0, \beta_4 = 0$)			Three restrictions($\beta_1 = 0,\ \beta_3 = 0, \beta_4 = 0$)		
λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}
Swamy and Arora								
2138.526^a	NA	609110684.047^a	1109.311^a	NA	178319.656^a	857.074^a	NA	293757.012^a
Wansbeek and Kappteyn								
493.148^a	NA	839636025.188^a	364.119^a	NA	18431907.107^a	358.831^a	1433.875^a	1586347.584^a
Wallace and Hussain								
777.764^a	NA	5001783.708^a	624.210^a	NA	767226.002^a	581.321^a	NA	2843988.540^a
Maximum likelihood								
775.828^a	NA	$8.416138e+08^a$	629.157^a	NA	3688699.810^a	775.828^a	NA	$8.416138e+08^a$

Table 1b: Restriction test results - FIM based on exact information matrix [Homoscedasticity case].

Notes: λ_W, λ_{LR} and λ_{LM} represent the Wald, likelihood and Lagrange multiplier statistics respectively. NA=not available. ^{a},^{b} and ^{c} represent statistical significant at 1%, 5% and 10% repectively.

6.2 Example 2: Public capital productivity (Double Heteroscedasticity case)

6.2.1 Model specification

We still consider the model described previously, i.e.,

$$\ln(y_{it}) = \beta_0 + \beta_1 \ln(Pc_{it}) + \beta_2 \ln(PS_{it}) + \beta_3 \ln(L_{it}) + \beta_4 U_{\text{emp}_{it}} + \mu_i + \lambda_t + \varepsilon_{it}, \tag{6.2}$$

Next, we assume that the analysis is more complex and thereby proceed methodically. As will be seen later, the FIM here is based on heteroscedasticity of the two errors involved. The following steps are important to understand results reported in Table 2.

6.2.2 Total Number of Parameters

The total number of parameters is 16 since $\theta_\nu \in \mathbb{R}^4$ and $\theta_\mu \in \mathbb{R}^4$ (in fact they correspond to four independent variables).

6.2.3 Existence of Heteroscedasticity

We check the possibility of single or double heteroscedasticity on the individual term (μ_i) as well as on the rest of the perturbation (ν_{it}). Indeed, the double heteroscedasticity case based on the above data set is confirmed following [15], using the so called L_1, L_2 and L_3 tests e.g., see [15]. This step is crucial because having heteroscedasticity on the individual term (μ_i) or on the rest of the perturbation (ν_{it}) or both will obviously affect the structure of the variance covariance matrix of the error terms and thereby the FIM.

6.2.4 Expression of the Variances

For the variances, we use the expression given in [29] such that $\sigma^2_{\nu_{it}} = \sigma^2_\nu(1 + \theta'_\nu \overline{x}_{i.})^2$ and $\sigma^2_{\mu_i} = \sigma^2_\mu(1+\theta'_\mu \overline{x}_{i.})^2$, where $\overline{x}_{i.}$ is the vector of four values corresponding to the mean of each independent variable. [29] also proposed some alternative forms by replacing $(1+\delta'\overline{x}_{i.})^2$ by $\exp(\delta\overline{x}_{i.})$ for $\delta \in \{\theta_\nu, \theta_\mu\}$. As results, we present the Wald, LR and LM tests obtained by the approximations described in [3] and then obtain the same statistics with our approach, assuming the existence of heteroscedasticity and estimating the parameters θ_ν and θ_μ by maximum likelihood. Then, we re-calculate the previous tests in the case of heteroscedasticity.

6.2.5 Forms of Heteroscedasticity

In the absence of any discrimination test, we consider all potential forms of heteroscedasticity. In the literature, four potential cases exist: (i) $h_\nu(x) = (1+x)^2$ and $h_\mu(x) = (1+x)^2$, (ii) $h_\nu(x) = (1+x)^2$ and $h_\mu(x) = \exp(x)$, (iii) $h_\nu(x) = \exp(x)$ and $h_\mu(x) = (1+x)^2$ and (iv) $h_\nu(x) = \exp(x)$ and $h_\mu(x) = \exp(x)$. To conserve space and for the sake of conciseness, we present results related to case (i). Cases (ii), (iii) and (iv) results are available as supplementary material.

6.2.6 Results

In Table 2 the FIM involves exact information matrix based on double heteroscedasticity on the unobservable individual term as well as on the rest of the perturbation; since as previously indicated double heteroscedasticity case based on the above data set is confirmed using the so called L_1, L_2 and L_3 tests e.g., see [15]. It should be noticed that using any other information matrix (for example based on heteroscedasticity on the unobservable individual term or the rest of the perturbation) would have resulted in serious mis-specification and thereby mis-leading results.

Results reported are based on the FIM with exact information matrix and with double heteroscedastic errors; a case not addressed by existing testing procedures. Results clearly indicate that irrespective of the estimator used and no matter how many restrictions are used, public capital remains essential in private production. The results which are consistent with the study by [6] and [23]. are based on a new testing procedure.

6.2.7 Discussion

Some inconsistencies have been resolved. For example, the number of times some test results are not available remains relatively reasonable. A possible reason could be that the correct specification is used and the information is based on exact information. Note also in this case that the Hessian matrix is always invertible. The advantages of our new approach based on heteroscedasticity compared to the homoscedastic case are that (i) correct specification is used; (ii) correct information matrix is considered; (iii) correct estimated standard errors and the associated t-statistics are reported; (iv) correct F-statistics and their probabilities are reported; (v) testing procedure based on Wald, LR and LM tests is now using correct information and therefore gives fewer puzzling results.

6.2.8 Further comments

Some questions remain pending: (i) What if heteroscedasticity was not considered when there is one or double heteroscedasticity? A case of mis-specification and thereby misleading results since the appropriate variance-covariance matrix and thereby the appropriate FIM has not been taken into account. (ii) What if heteroscedasticity was only on the individual term, μ_i; when double heteroscedasticity has been assumed? Additional computations undertaken indicate some misleading and puzzling results as this is a serious case of mis-specification. (iii) What if heteroscedasticity was only on the rest of the perturbation, ν_{it}; when double heteroscedasticity or no heteroscedasticity has been assumed? Again, this case a serious case of mis-specification and would lead to serious inconsistencies. (iv) Other inappropriate cases not mentioned here will lead to mis-specifications as well.

One restriction($\beta_1=0$)			Two restrictions($\beta_1=0,\ \beta_2=0$)			Two restrictions($\beta_1=0,\ \beta_3=0$)			Two restrictions($\beta_1=0,\ \beta_4=0$)		
λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}
Swamy and Arora											
2571.678^a	862.849^a	354.361^a	87.522^a	193.918^a	79961.339^a	73.833^a	789.774^a	16209.709^a	176.438^a	126.214^a	83748.331^a
Wansbeek and Kapteyn											
3674.453^a	95.472^a	88338.056^a	109.841^a	158.616^a	79960.701^a	82.499^a	780.606^a	16181.654^a	175.484^a	101.969^a	83749.189^a
Wallace and Hussain											
2429.035^a	NA	384.378^a	83.029^a	41.617^b	385.879^a	70.582^a	742.483^a	389.649^a	169.421^a	11.781^c	384.498^a
Maximum likelihood											
2653.676^a	NA	384.422^a	89.529^a	79.667^a	385.866^a	73.317^a	707.826^a	392.209^a	172.957^a	7.466^c	384.592^a

Three restrictions($\beta_1=0,\ \beta_2=0,\beta_3=0$)			Three restrictions($\beta_1=0,\ \beta_2=0,\beta_4=0$)			Three restrictions($\beta_1=0,\ \beta_3=0,\beta_4=0$)		
λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}	λ_W	λ_{LR}	λ_{LM}
Swamy and Arora								
133.361^a	1234.732^a	5218.751^a	79.664^a	205.010^a	77275.198^a	277.756^a	910.836^a	11056.378^a
Wansbeek and Kappteyn								
134.893^a	1177.421^a	5219.120^a	84.288^a	161.857^a	77275.067^a	333.341^a	889.311	11059.297^a
Wallace and Hussain								
127.709^a	1078.167^a	411.086^a	76.811^a	72.769^b	385.722^a	256.284^a	859.921^a	389.792^a
Maximum likelihood								
131.810^a	1157.906^a	403.662^a	79.060^a	81.304^a	385.726^a	276.340^a	819.430^a	392.208^a

Table 2: Restriction test results - FIM based on exact information matrix with double heteroscedasticity. [Heteroscedasticity case].
Notes: λ_W, λ_{LR} and λ_{LM} represent the Wald, likelihood and Lagrange multiplier statistics respectively. NA=not available. ^{a},^{b} and ^{c} represent statistical significant at 1%, 5% and 10% repectively.

7 Final remarks

Correct model specification and selection have severe effects on modeling exercises. In this context, the Fisher Information Matrix (FIM) is critical. In this paper, we present a new approach to estimating the FIM in the specific case of the two-way random effects panel data model with and without heteroscedasticity. This is an attempt to possibly resolve earlier complexity in the use of the famous Cramer-Rao inequality statistic, an important aspect of which is the FIM. We derive the FIM of the two-way random effects panel data model in general as well as in specific cases of heteroscedasticity and homoscedasticity. Some examples based on real data are provided.

REFERENCES

[1] Baltagi, BH (2021). *Econometric Analysis of Panel Data*, Springer Texts in Business and Economics. 6th Edition.

[2] Baltagi, BH (1980). Simultaneous equations with error components *Journal of Econometrics*, **17**(2), 189–200.

[3] Baltagi, BH, Bresson, G and Pirotte, A (2006). Joint LM test for heteroscedasticity in a one-way error component model *Journal of Econometrics*, **134**(2), 401–417.

[4] Baltagi, BH and Griffin, JM (1988). A generalized error component model with heteroscedastic disturbances *International Economic Review* , **29**(4), 745–753.

[5] Baltagi, BH and Li, Q (1990). A Lagrange multiplier test for the error components model with incomplete panels *Econometrics Reviews*, **9**(1), 103–107.

[6] Baltagi, BH and Pinnoi, N (1995). Public capital stock and state productivity growth: Further evidence from an error components model *Empirical Economics*, **20**(2), 351–359.

[7] Baltagi, BH, Song, SH and Jung, BC (2002). A comparative study of alternative estimators for the unbalanced two-way error component regression model *Econometrics Journal*, **5**(2), 480–493.

[8] Bao Y and Hua (2014). On the Fisher information matrix of a vector ARMA process *Economics Letters*, **123**(1), 14–16.

[9] Bartlett, MS (1951). An inverse matrix adjustment arising in discriminant analysis *Annals of Mathematical Statistics* , **22**(11), 107–111.

[10] Baum, CF (2006). Testing for groupwise heteroscedasticity *The Stata Journal*, **6**(4), 590.

[11] Buse, A (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note *The American Statistician* , **36**(3a), 153–157.

[12] Cramer, H (1946). *Mathematical Methods of Statistics. Princeton Mathematical Series. Princeton: Princeton University Press.*

[13] Crowder, MJ (1976). Maximum likelihood estimation for dependent observations *Journal of The Royal Statistical Society, Series B (Methodological)*, **38**(1), 45–53.

[14] Feigin, PD (1976). Maximum likelihood estimation for continuous-time stochastic processes *Advances in Applied Probability* , **8**(4), 712-736.

[15] Feng S, Li G, Tong T and Luo S (2020). Testing for heteroskedasticity in two-way fixed effects panel data models *Journal of Applied Statistics*, **47**(1), 91-116.

[16] Hogg, RV, McKean J and Craig, AT (2005). *Introduction to Mathematical Statistics.* Pearson Education

[17] Holly, A and Gardiol, L (2000). A score test for individual heteroscedasticity in a one-way error component model *Chapter 10 in J. Krishnakumar and L. Ronchetti, eds. Panel Data Econometrics: Future Directions. North-Holland, Amsterdam* , 199-211.

[18] Kouassi, E, Sango, J, Brou, Bosson JM and Mougoue, M (2014). A joint score test for heteroscedasticity in the two-way error components model *Communications in Statistics Theory and Methods* , **43**(13), 2734-2751.

[19] Kouassi, E, Mougoue, M, Sango J, Brou, JM, Amba, CM, and Salisu, AA (2014) Testing for heteroskedasticity and spatial correlation in a two way random effects model *Computational Statistics and Data Analysis*, **70**(C), 153-171.

[20] Lejeune, B (2006). A full heteroscedastic one-way error component model pseudo maximum likelihood estimation and specification testing *Chapter 2 in B. Baltagi ed., Panel Data Econometrics: Theoretical Contributions and Empirical Applications. Elsevier: Amsterdam*, 31-66

[21] Mazodier, P, and Trognon, A (1978). Heteroscedasticity and stratification in error components models *Annales de lINSEE*, 451–482, Institut National de la Statistique et des Etudes Economiques.

[22] Montes-Rojas, G and Escudero, WS (2011). Robust tests for heteroscedasticity in the one-way error components model *Journal of Econometrics* , **160**(2), 300-310.

[23] Munnell, AH (1990). Why has productivity growth declined? *Productivity and Public Investment, New England Economic Review*, **30**(January), 3-22.

[24] Nadarajah, S (2006a). FIM for Arnold and Strauss's bivariate Gamma distribution *Computational Statistics and Data Analysis* , **51**(3), 1584-1590.

[25] Nadarajah, S (2006b). Fisher information for the elliptically symmetric Pearson distributions *Applied Mathematics and Computation* , **178**(2), 195-206.

[26] Note, AE (1962). The likelihood ratio, Wald and Lagrange multiplier tests: An expository note *The American Statistician* , **36**(3), 153-157.

[27] Petersen, MA (2009). Estimating standard errors in finance panel data sets: Comparing approaches *Review of Financial Studies* , **22**(1), 435-480.

[28] Randolph, WC (1988). A transformation for heteroscedastic error components regression models *Economics Letters* , **27**(4), 349-354.

[29] Roy, N (2002). Is adaptive estimation useful for panel models with heteroskedasticity in the individual specific error component? Some Monte Carlo evidence *Econometric Reviews*, **21**(2), 189-203.

[30] Su, LJ, Ullah, A and Wang, Y (2013). Nonparametric regression estimation with general parametric error covariance: A more efficient two-step estimator *Empirical Economics*, **45**(2), 1009-1024.

[31] Verbon, HAA (1980). Testing for heteroscedasticity in a model of seemingly unrelated Regression equations with variance component (SUREVC) *Economics Letters* , **5**(2), 149-153.

[32] Wansbeek, T and Kapteyn, A (1982). A simple way to obtain the spectral decomposition of variance components for balanced data *Communications in Statistics: Theory and Methods* , **A11**(18), 2105-2112.

[33] Wu, J (2016). Robust random effects tests for two-way error component models with panel data *Economic Modelling* , **59**(C), 1-8.

[34] Wu, JH and Zhu, LX (2011). Testing for serial correlation and random effects in a two-way error component regression model *Economic Modelling* , **28**(6), 2377-2386.

8 Appendices

8.1 Appendix A: Derivation of the coefficients $a_{i,j}, a_{i,j}(i), a_{i,j}(j), a_{i,j}(i,j)$.

The coefficients $a_{i,j}$ and $\mathbf{a}_{i,j}$ are given by

$$a_{1,1} = \frac{(T-1)}{2}\mathrm{tr}\big[C_1 D_\nu C_1 D_\nu\big] + \frac{1}{2}\mathrm{tr}\big[C_2 D_\nu C_2 D_\nu\big],$$

$$a_{2,2} = \frac{T^2}{2}\mathrm{tr}\big[C_2 D_\mu C_2 D_\mu\big],$$

$$a_{3,3} = \frac{(T-1)}{2}\mathrm{tr}\big[C_1 J_N C_1 J_N\big] + \frac{1}{2}\mathrm{tr}\big[C_2 J_N C_2 J_N\big],$$

$$a_{1,2} = a_{2,1} = \frac{T}{2}\mathrm{tr}\big[C_2 D_\nu C_2 D_\mu\big],$$

$$a_{1,3} = a_{3,1} = \frac{(T-1)}{2}\mathrm{tr}\big[C_1 D_\nu C_1 J_N\big] + \frac{1}{2}\mathrm{tr}\big[C_2 D_\nu C_2 J_N\big],$$

$$a_{2,3} = a_{3,2} = \frac{T}{2}\mathrm{tr}\big[C_2 D_\mu C_2 J_N\big],$$

$$\mathbf{a}_{1,4}(i) = \sigma_\nu^2\left(\frac{(T-1)}{2}\mathrm{tr}\big[C_1 D_\nu C_1 D^*_{\nu,i}\big] + \frac{1}{2}\mathrm{tr}\big[C_2 D_\nu C_2 D^*_{\nu,i}\big]\right), i = 1:p,$$

$$\mathbf{a}_{1,5}(j) = \frac{\sigma_\mu^2 T}{2}\mathrm{tr}\big[C_2 D_\nu C_2 D^*_{\mu,j}\big], j = 1:q,$$

$$\mathbf{a}_{2,4}(i) = \sigma_\nu^2\left(\frac{T}{2}\mathrm{tr}\big[C_2 D_\mu C_2 D^*_{\nu,i}\big]\right), i = 1:p,$$

$$\mathbf{a}_{2,5}(j) = \left(\frac{\sigma_\mu^2 T^2}{2}\mathrm{tr}\big[C_2 D_\mu C_2 D^*_{\mu,j}\big]\right), j = 1:q,$$

$$\mathbf{a}_{3,4}(i) = \mathbf{a}'_{4,3}(i) = \sigma_\nu^2\Big(\frac{(T-1)}{2}\mathrm{tr}\big[C_1J_NC_1D^*_{\nu,i}\big] + \frac{1}{2}\mathrm{tr}\big[C_2J_NC_2D^*_{\nu,i}\big]\Big), i = 1:p,$$

$$\mathbf{a}_{3,5}(i) = \frac{\sigma_\mu^2 T}{2}\mathrm{tr}\big[C_2J_NC_2D^*_{\mu,j}\big], j = 1:q,$$

$$\mathbf{a}_{4,4}(i,j) = \sigma_\nu^4\Big(\frac{(T-1)}{2}\mathrm{tr}\big[C_1D^*_{\nu,i}C_1D^*_{\nu,j}\big] + \frac{1}{2}\mathrm{tr}\big[C_2D^*_{\nu,i}C_2D^*_{\nu,j}\big]\Big), 1 \le i,j \le p,$$

$$\mathbf{a}_{4,5}(i,j) = \mathbf{a}'_{5,4}(i,j)\sigma_\mu^2\sigma_\nu^2\Big(\frac{T}{2}\mathrm{tr}\big[C_2D^*_{\nu,i}C_2D^*_{\mu,j}\big]\Big), 1 \le i, \le p; 1 \le j \le q,$$

$$\mathbf{a}_{5,5}(i,j) = \sigma_\mu^4\Big(\frac{T^2}{2}\mathrm{tr}\big[C_2D^*_{\mu,i}C_2D^*_{\mu,i}\big]\Big), 1 \le i,j \le q.$$

8.2 Appendix B: Proof of Proposition 5.4

The terms of $I_n(\overline{\theta})$ are written in terms Ψ_1, Ψ_2 and Ψ_3 by replacing $tr(C_1M_1C_2M_2)$ by the corresponding Ψ_i according to the following rule: $C_1 \leftrightarrow \vartheta$, $C_2 \leftrightarrow \phi$, $D_\nu \leftrightarrow \psi$ and $D_\mu \leftrightarrow \eta$. For example,

$$a_{1,1} = \frac{(T-1)}{2}\mathrm{tr}\big[C_1D_\nu C_1D_\nu\big] + \frac{1}{2}\mathrm{tr}\big[C_2D_\nu C_2D_\nu\big] = \frac{(T-1)}{2}\Psi_1(\vartheta,\psi,\vartheta,\psi) + \frac{1}{2}\Psi_1(\phi,\psi,\phi,\psi).$$

We obtain the other terms in the following way:

$$a_{2,2} = \frac{T}{2}\Psi_1(\phi,\eta,\phi,\psi),$$

$$a_{3,3} = \frac{(T-1)}{2}\Psi_3(\vartheta,\vartheta) + \frac{1}{2}\Psi_3(\phi,\phi),$$

$$a_{1,2} = a_{2,1} = \frac{T}{2}\Psi_1(\phi,\eta,\phi,\psi),$$

$$a_{1,3} = a_{3,1} = \frac{(T-1)}{2}\Psi_2(\vartheta,\eta,\vartheta) + \frac{1}{2}\Psi_2(\phi,\eta,\phi),$$

$$a_{2,3} = a_{3,2} = \frac{T}{2}\Psi_2(\phi,\psi,\phi),$$

$$\mathbf{a}_{1,4}(i) = \sigma_\nu^2\Big(\frac{T}{2}\Psi_1(\phi,\psi,\phi,\eta^i)\Big), i = 1:p,$$

$$\mathbf{a}_{1,5}(j) = \sigma_\mu^2\Big(\frac{T}{2}\Psi_1(\phi,\eta,\phi,\psi^j\Big), j = 1:q,$$

$$\mathbf{a}_{2,4}(i) = \sigma_\nu^2\Big(\frac{T}{2}\mathrm{tr}\big[C_2D_\mu C_2D^*_{\nu,i}\big]\Big), i = 1:p,$$

$$\mathbf{a}_{2,5}(j) = \sigma_\mu^2\Big(\frac{T^2}{2}\Psi_1(\phi,\psi,\phi,\psi^j)\Big), j = 1:q,$$

$$\mathbf{a}_{3,4}(i) = \mathbf{a}'_{4,3}(i) = \sigma_\nu^2\Big(\frac{(T-1)}{2}\Psi_2(\vartheta,\eta^i,\vartheta) + \frac{1}{2}\Psi_2(\phi,\eta^i,\phi)\Big), i = 1:p,$$

$$\mathbf{a}_{3,5}(i) = \sigma_\mu^2\bigg(\frac{T}{2}\Psi_2(\phi,\psi^j,\phi)\bigg), j = 1:q,$$
$$\mathbf{a}_{4,4}(i,j) = \sigma_\nu^4\bigg(\frac{(T-1)}{2}\Psi_1(\vartheta,\eta^i,\vartheta,\eta^j) + \frac{1}{2}\Psi_1(\phi,\eta^i,\phi,\eta^j)\bigg), 1 \le i,j \le p,$$
$$\mathbf{a}_{4,5}(i,j) = \mathbf{a}_{5,4}'(i,j)\sigma_\mu^2\sigma_\nu^2\bigg(\frac{T}{2}\mathrm{tr}\big[C_2D^*_{\nu,i}C_2D^*_{\mu,j}\big]\bigg), 1 \le i, \le p; 1 \le j \le q,$$
$$\mathbf{a}_{5,5}(i,j) = \sigma_\nu^4\bigg(\frac{T^2}{2}\Psi_1(\phi,\psi^i,\phi,\psi^j)\bigg), 1 \le i,j \le q.$$

8.3 Appendix C: Gradient of logarithm of likelihood function

By Lemma 3.1,

$$d\ell(\theta \mid u) = -\frac{1}{2}tr\left(\Omega^{-1}d\Omega\right) - u'\Omega^{-1}du + \frac{1}{2}u'\Omega^{-1}d\Omega\cdot\Omega^{-1}u.$$

Since from Proposition (4.1), $.\Omega^{-1}d\Omega = \sum_{j=1}^{r-k}\Omega_j d\bar{\theta}_j$, we have

$$d\ell(\theta \mid u) = -\frac{1}{2}\sum_{j=1}^{r-k}tr\left(\Omega_j\right)d\bar{\theta}_j + u'\Omega^{-1}Xd\beta + \frac{1}{2}\sum_{j=1}^{r-k}u'\Omega_j\cdot\Omega^{-1}ud\bar{\theta}_j,$$

where

$$\begin{aligned}
&\mathrm{tr}(\Omega_1) = (T-1)\mathrm{tr}(C_1D_\nu) + \mathrm{tr}(C_2D_\nu),\\
&\mathrm{tr}(\Omega_2) = T\mathrm{tr}(C_2D_\mu),\\
&\mathrm{tr}(\Omega_3) = (T-1)\mathrm{tr}(C_1J_N) + \mathrm{tr}(C_2J_N),\\
&\mathrm{tr}(\Omega_{3+j_1}) = \sigma_\nu^2\left(\mathrm{tr}(C_1D^*_{\nu,j_1})(T-1) + \mathrm{tr}(C_2D^*_{\nu,j_1})\right), 1 \le j_1 \le p,\\
&\mathrm{tr}(\Omega_{p+3+j_2}) = \sigma_\mu^2\mathrm{tr}(C_2D^*_{\mu,j_2})T, 1 \le j_2 \le q.
\end{aligned}$$

We then deduce that

$$\begin{aligned}
\frac{\partial\ell(\theta \mid u)}{\partial\sigma_\nu^2} &= -\frac{1}{2}\left[(T-1)\mathrm{tr}(C_1D_\nu) + \mathrm{tr}(C_2D_\nu)\right] + \frac{1}{2}u'\Omega_1\Omega^{-1}u,\\
\frac{\partial\ell(\theta \mid u)}{\partial\sigma_\mu^2} &= -\frac{1}{2}\left[T\mathrm{tr}(C_2D_\mu)\right] + \frac{1}{2}u'\Omega_2\Omega^{-1}u,\\
\frac{\partial\ell(\theta \mid u)}{\partial\sigma_\lambda^2} &= -\frac{1}{2}\left[(T-1)\mathrm{tr}(C_1J_N) + \mathrm{tr}(C_2J_N)\right] + \frac{1}{2}u'\Omega_3\Omega^{-1}u,\\
\frac{\partial\ell(\theta \mid u)}{\partial\theta^*_{\nu,j_1}} &= -\frac{1}{2}\left[\sigma_\nu^2\mathrm{tr}(C_1D^*_{\nu,j_1})(T-1) + \sigma_\nu^2\mathrm{tr}(C_2D^*_{\nu,j_1})\right] + \frac{1}{2}u'\Omega_{3+j_1}\Omega^{-1}u, 1 \le j_1 \le p,\\
\frac{\partial\ell(\theta \mid u)}{\partial\theta^*_{\mu,j_2}} &= -\frac{1}{2}\left[\sigma_\mu^2\mathrm{tr}(C_2D^*_{\mu,j_2})T\right] + \frac{1}{2}u'\Omega_{3+p+j_2}\Omega^{-1}u, 1 \le j_2 \le q,\\
\frac{\partial\ell(\theta \mid u)}{\partial\beta_k} &= u'\Omega^{-1}X[k], 1 \le k \le K.
\end{aligned}$$