The EM algorithm for estimating parameters of normal mixed effects models

Ibirénoyé H. R. SODJAHIN*

August 17, 2024

Abstract

I lay out here the steps I went through when trying to understand the mechanic of the EM algorithm. I'm interested in situations where closed-form solution exists for the parameters of interest. Situations involving simulation of the conditional distribution are beyond the scope of this paper.

I start by deriving the explicit solutions for estimating linear mixed effects models by maximum likelihood through the EM algorithm, and provide formula for computing the information matrix (and thus, the standard errors) of obtained estimates. Finally, numerical examples using simulated data are also provided and estimated in each case.

Keywords: maximum likelihood, linear mixed effects models, random effects models, EM algorithm

Contents

*I'm highly indebted to Philippe KOUTCHADE for countless hours he spent to explain me the EM algorithm, its extensions and applications

1 Introduction

I outline the steps towards the derivation of the maximum likelihood estimator (MLE) of the normal mixed effects model via the Expectation-Maximization algorithm proposed by Dempster et al. (1977). This algorithm proves useful in computing the maximum likelihood estimator of models in which data missing issues are present. Its convergence properties are established in Boyles (1983) and Wu (1983). Furthermore, the computation of the covariance matrix of the random effects/random coefficients is straightforward in the EM framework, and in addition, we are not obliged to impose some restrictions on the covariance matrix to ease the estimation.

I start with the linear mixed effects model and derive the sufficient statistics needed to implement the EM algorithm. I provide formula to compute the variance of the estimates. I then derive the MLE of the random intercept model, and extend it to accommodate exogenous explanatory variables. Numerical example exploiting the data in Neath (2012) and simulated data are used to illustrate the derived formula.

2 Important preliminary results

A note of nomenclature: throughout the text, bold symbols in small capita are vectors, and those bold in capital letters are matrices. All what remain are scalars. The symbol $'$ is intended for matrix transposition.

The result below shows how to compute the first two moments of the conditional distribution of a random vector.

Let's consider 2 normal random vectors $\mathbf{X}$ and $\mathbf{Y}$, with $\mathbf{X} \sim \mathcal{N}(\mathbf{m}_X, \mathbf{V}_X)$ and $\mathbf{Y} \sim \mathcal{N}(\mathbf{m}_Y, \mathbf{V}_Y)$. Then, the random vector $(\mathbf{X}, \mathbf{Y})'$ is multivariate normal:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mathbf{m}_X \\ \mathbf{m}_Y \end{pmatrix}, \begin{pmatrix} \mathbf{V}_X & \mathbf{V}_{XY} \\ \mathbf{V}_{YX} & \mathbf{V}_Y \end{pmatrix}\right]$$

where $\mathbf{V}_{XY}$ is the covariance matrix between $\mathbf{X}$ and $\mathbf{Y}$.

Following Davidson (2018, p. 93) and adopting a vector/matrix notation, the random vector $\mathbf{Y} \mid \mathbf{X} \sim \mathcal{N}\left(\mathbf{m}_{\mathbf{Y}|\mathbf{X}}, \mathbf{V}_{\mathbf{Y}|\mathbf{X}}\right)$ with:

$$\mathbf{m}_{\mathbf{Y}|\mathbf{X}} = \mathbb{E}(\mathbf{Y} \mid \mathbf{X}) = \mathbf{m}_Y + \mathbf{V}_{YX}\mathbf{V}_X^{-1}(\mathbf{X} - \mathbf{m}_X)$$

$$\mathbf{V}_{\mathbf{Y}|\mathbf{X}} = \mathrm{Var}(\mathbf{Y} \mid \mathbf{X}) = \mathbf{V}_Y - \mathbf{V}_{YX}\mathbf{V}_X^{-1}\mathbf{V}_{XY}$$

In addition, as $\mathbf{V}_{\mathbf{Y}|\mathbf{X}} = \mathbb{E}(\mathbf{Y}\mathbf{Y}' \mid \mathbf{X}) - \mathbb{E}(\mathbf{Y} \mid \mathbf{X})\mathbb{E}(\mathbf{Y} \mid \mathbf{X})' = \mathbb{E}(\mathbf{Y}\mathbf{Y}' \mid \mathbf{X}) - \mathbf{m}_X\mathbf{m}_X'$, it turns out that

$$\mathbb{E}(\mathbf{Y}\mathbf{Y}' \mid \mathbf{X}) = \mathbf{V}_{\mathbf{Y}|\mathbf{X}} + \mathbf{m}_X\mathbf{m}_X'$$

3 The normal linear mixed effects model

Suppose that we have a sample $\mathscr{S}$ of $i = 1, \ldots, N$ independent observations (possibly observed over some time periods, $t \in \mathscr{T}_i = \{1, \ldots, T_i\}$ where T_i is the number of times individual i is observed) randomly drawn from a population[1], and we are interested in some of the population characteristics. The sample is described by the $(T_i \times 1)$ vectors $\mathbf{y}_i$, the $(T_i \times K)$ matrices $\mathbf{X}_i$ and $(T_i \times J)$ matrices $\mathbf{Z}_i$, where $K \geq J$. We note $n = \sum_i T_i$, the sample size, and recall that we sampled N individuals. Therefore, we write:

$$\mathbf{y}_i = \left(\mathrm{y}_{it}, t \in \mathscr{T}_i\right) = \begin{pmatrix} \mathrm{y}_{i1} \\ \vdots \\ \mathrm{y}_{iT_i} \end{pmatrix}, \quad \mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1}' \\ \vdots \\ \mathbf{x}_{iT_i}' \end{pmatrix} = \begin{pmatrix} \mathrm{x}_{1,it_1} & \mathrm{x}_{2,it_1} & \cdots & \mathrm{x}_{K,it_1} \\ \mathrm{x}_{1,it_2} & \mathrm{x}_{2,it_2} & \cdots & \mathrm{x}_{K,it_2} \\ \cdots & \cdots & \ddots & \vdots \\ \mathrm{x}_{1,iT_i} & \mathrm{x}_{2,iT_i} & \cdots & \mathrm{x}_{K,iT_i} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix},$$

$$\text{and } \mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i1}' \\ \vdots \\ \mathbf{z}_{iT_i}' \end{pmatrix} = \begin{pmatrix} \mathrm{z}_{1,it_1} & \mathrm{z}_{2,it_1} & \cdots & \mathrm{z}_{J,it_1} \\ \mathrm{z}_{1,it_2} & \mathrm{z}_{2,it_2} & \cdots & \mathrm{z}_{J,it_2} \\ \cdots & \cdots & \ddots & \vdots \\ \mathrm{z}_{1,iT_i} & \mathrm{z}_{2,iT_i} & \cdots & \mathrm{z}_{J,iT_i} \end{pmatrix}, \quad \boldsymbol{\mu}_i = \begin{pmatrix} \mu_{1,i} \\ \vdots \\ \mu_{J,i} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_i = (\varepsilon_{it}, t \in \mathscr{T}_i) = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT_i} \end{pmatrix}$$

We note $\mathbf{y} = \mathrm{vec}\left(\mathbf{y}_i,\ i = 1, \cdots, N\right)$ the vector of observations, obtained by stacking all the $\mathbf{y}_i$, and $\boldsymbol{\mu} = \mathrm{vec}\left(\boldsymbol{\mu}_i,\ i = 1, \cdots, N\right)$ the vector obtained by stacking all the $\boldsymbol{\mu}_i$.

We assume that the data generating mechanism producing $\mathbf{y}_i$ relates it to matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ (assumed non-random, while $\mathbf{Z}_i$ is a subset of $\mathbf{X}_i$) and random vector $\boldsymbol{\varepsilon}_i$ such

[1] We assume that this panel data is unbalanced, completely at random.

that $\forall t = 1, \cdots, T_i$:

(1) $y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\boldsymbol{\mu}_i + \varepsilon_{it}$ for $i, t \in \mathscr{S}$

where $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of fixed coefficients, $\boldsymbol{\mu}_i$ is a $(J \times 1)$ vector of random coefficients, and ε_{it} is an error term with no serial correlation $\left(i.e., \operatorname{cov}(\varepsilon_{it}, \varepsilon_{it'}) = 0 \text{ if } t \neq t'\right)$. We further assume that $\boldsymbol{\mu}_i$ and $\boldsymbol{\varepsilon}_i$ are independent and identically distributed, and that $\boldsymbol{\mu}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_0)$[2] and $\varepsilon_{it} \sim \mathcal{N}\left(0, \sigma_0^2\right)$.

Following the EM formalism, the observed data are $\mathbf{B}_i \equiv \left(\mathbf{y}_i, \mathbf{X}_i, \mathbf{Z}_i\right)$ and the complete data are $\mathbf{C}_i \equiv \left(\mathbf{y}_i, \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\mu}_i\right) \equiv \left(\mathbf{B}_i, \boldsymbol{\mu}_i\right)$. The set of parameters to estimate is $\boldsymbol{\gamma}_0 = \left(\boldsymbol{\beta}, \boldsymbol{\Omega}_0, \sigma_0^2\right)$. To alleviate the notational burden, we will write throughout this paper $\mathbf{y}_i \mid \boldsymbol{\mu}_i$ in lieu of $\mathbf{y}_i \mid \boldsymbol{\mu}_i, \mathbf{X}_i, \mathbf{Z}_i$, and $y_{it} \mid \boldsymbol{\mu}_i$ in lieu of $y_{it} \mid \boldsymbol{\mu}_i, \mathbf{x}_{it}, \mathbf{z}_{it}$.

From (1), $\forall t = 1, \cdots, T_i$, the random variable $y_{it} \mid \boldsymbol{\mu}_i$ is i.i.d. normal, *i.e.*,
$y_{it} \mid \boldsymbol{\mu}_i \sim \mathcal{N}\left(\mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{it}\boldsymbol{\mu}_i,\ \sigma_0^2\right)$. It follows that the joint distribution of
$\left(y_{i1} \mid \boldsymbol{\mu}_i,\ y_{i2} \mid \boldsymbol{\mu}_i, \cdots,\ y_{iT_i} \mid \boldsymbol{\mu}_i\right)' \equiv \mathbf{y}_i \mid \boldsymbol{\mu}_i$ is $\mathcal{N}\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\mu}_i,\ \sigma_0^2 \mathrm{Id}_{T_i}\right)$.

The density functions of the multivariate random vectors $\mathbf{y}_i \mid \boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_i$ write:

$$\mathbb{P}\left(\mathbf{y}_i \mid \boldsymbol{\mu}_i; \boldsymbol{\beta}, \sigma_0^2\right) = \frac{1}{\sqrt{(2\pi)^{T_i}} \mid \sigma_0^2 \mathrm{Id}_{T_i} \mid^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)' \left(\sigma_0^2 \mathrm{Id}_{T_i}\right)^{-1} \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)\right]$$

$$\mathbb{P}\left(\boldsymbol{\mu}_i; \boldsymbol{\Omega}_0\right) = \frac{1}{\sqrt{(2\pi)^{J}} \mid \boldsymbol{\Omega}_0 \mid^{1/2}} \exp\left[-\frac{1}{2}\boldsymbol{\mu}_i' \boldsymbol{\Omega}_0^{-1} \boldsymbol{\mu}_i\right]$$

Because σ_0^2 is a scalar:

- the properties of the determinant leads to: $\mid \sigma_0^2 \mathrm{Id}_{T_i} \mid^{1/2} = \left(\sigma_0^{2T_i} \mid \mathrm{Id}_{T_i} \mid\right)^{1/2} = \left(\sigma_0^2\right)^{T_i/2}$[3]
- $\left(\sigma_0^2 \mathrm{Id}_{T_i}\right)^{-1} = \frac{1}{\sigma_0^2} \mathrm{Id}_{T_i}$

It follows that:

$$\mathbb{P}\left(\mathbf{y}_i \mid \boldsymbol{\mu}_i; \boldsymbol{\beta}, \sigma_0^2\right) = \frac{1}{\sqrt{(2\pi)^{T_i}} \left(\sigma_0^2\right)^{T_i/2}} \exp\left[-\frac{1}{2\sigma_0^2}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)' \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)\right]$$

By the Bayes's rule, the joint probability of $\left(\mathbf{y}_i, \boldsymbol{\mu}_i\right)$, or the complete data likelihood

[2] $\boldsymbol{\Omega}_0$ is a $(J \times J)$ square matrix.

[3] $\mid cA \mid = c^m \mid A \mid$, and $\mid \mathrm{Id}_m \mid = 1$ for A a square matrix of order m, c a scalar and Id_m a identity matrix of order m.

writes:

$$\mathbb{P}\left(\mathbf{y}_i,\boldsymbol{\mu}_i;\boldsymbol{\gamma}_0\right)=\mathbb{P}\left(\mathbf{y}_i\mid\boldsymbol{\mu}_i;\boldsymbol{\beta},\sigma_0^2\right)\times\mathbb{P}\left(\boldsymbol{\mu}_i;\boldsymbol{\Omega}_0\right)=L_i\left(\mathbf{y}_i,\boldsymbol{\mu}_i;\boldsymbol{\gamma}_0\right)$$

The log-likelihood of the complete data for individual i writes:

$$\begin{aligned}\ln L_i\left(\mathbf{y}_i,\boldsymbol{\mu}_i;\boldsymbol{\gamma}_0\right)&=\ln\mathbb{P}\left(\mathbf{y}_i\mid\boldsymbol{\mu}_i;\boldsymbol{\beta},\sigma_0^2\right)+\ln\mathbb{P}\left(\boldsymbol{\mu}_i;\boldsymbol{\Omega}_0\right)\\&=-\frac{T_i}{2}\ln(2\pi)-\frac{T_i}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i\right)'\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i\right)\\&\quad-\frac{J}{2}\ln(2\pi)-\frac{1}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\boldsymbol{\mu}_i'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_i\end{aligned}$$

The sample log-likelihood of the complete data writes:

$$\begin{aligned}\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\boldsymbol{\gamma}_0\right)&=\sum_{i=1}^{N}\ln L_i\left(\mathbf{y}_i,\boldsymbol{\mu}_i;\boldsymbol{\gamma}_0\right)\\&=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i\right)'\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i\right)\\&\quad-\frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{\mu}_i'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_i\end{aligned}$$

$$\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\boldsymbol{\gamma}_0\right)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i-\frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{\mu}_i'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_i$$

and we recall that: $\boldsymbol{\varepsilon}_i=\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbf{Z}_i\boldsymbol{\mu}_i$.

In the EM framework, at iteration $(k+1)$, we aim to maximize the conditional expectation of $\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\boldsymbol{\gamma}_0\right)$, given $\mathbf{y}$ and given $\boldsymbol{\gamma}_0^{(k)}$, the value taken by the parameter vector $\boldsymbol{\gamma}_0$ at iteration (k). We denote this conditional expectation by $\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\boldsymbol{\gamma}_0\right)\mid\mathbf{y},\boldsymbol{\gamma}_0^{(k)}\right]$.

Note: Any conditining throughout the paper will be over $\mathbf{y}$ and the previous value of the parameter vector $\boldsymbol{\gamma}_0^{(k)}$. In order to alleviate the notational burden, we will omit the term $\boldsymbol{\gamma}_0^{(k)}$ when conditioning. For example, in lieu of writing $\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\boldsymbol{\gamma}_0\right)\mid\mathbf{y},\boldsymbol{\gamma}_0^{(k)}\right]$, we will write instead $\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\boldsymbol{\gamma}_0\right)\mid\mathbf{y}\right]$, and in lieu of writing $\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i,\boldsymbol{\gamma}_0^{(k)}\right)$, we will write $\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)$.

3.1 The E-step

See the details of the computation in Appendix A.1:

$$\mathbb{E}\left[\mathscr{L}(\mathbf{y},\boldsymbol{\mu};\gamma_0)\mid\mathbf{y}\right] = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma_0^2\right) - \frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\mathrm{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)\right]$$
$$-\frac{N\times J}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Omega}_0| - \frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}\left[\boldsymbol{\Omega}_0^{-1}\mathbb{E}\left(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\mid\mathbf{y}_i\right)\right]$$

We need to express $\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)$ and $\mathbb{E}\left(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\mid\mathbf{y}_i\right)$. Recall that $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i$. Thus, $\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right) = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)$. It turns out that the moments of the distribution of $\boldsymbol{\mu}_i\mid\mathbf{y}_i$ sufficiently determine $\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)$ and $\mathbb{E}\left(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\mid\mathbf{y}_i\right)$, and thus determines $\mathbb{E}\left[\mathscr{L}(\mathbf{y},\boldsymbol{\mu};\gamma_0)\mid\mathbf{y}\right]$.

But, we know that: $\mathbf{y}_i \sim \mathscr{N}\left(\mathbf{X}_i\boldsymbol{\beta},\ \mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)$ and $\boldsymbol{\mu}_i \sim \mathscr{N}(\mathbf{0},\ \boldsymbol{\Omega}_0)$.

In addition, the pairwise covariance between these random vectors write:

$$\mathrm{cov}\left(\mathbf{y}_i,\boldsymbol{\mu}_i\right) = \mathbb{E}\left(\mathbf{y}_i\boldsymbol{\mu}_i^{'}\right) - \mathbb{E}\left(\mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i\right)^{'} = \mathbb{E}\left(\mathbf{y}_i\boldsymbol{\mu}_i^{'}\right) = \mathbb{E}\left[\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i\right)\boldsymbol{\mu}_i^{'}\right] = \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\right) = \mathbf{Z}_i\boldsymbol{\Omega}_0$$
$$\mathrm{cov}\left(\boldsymbol{\mu}_i,\mathbf{y}_i\right) = \mathbb{E}\left(\boldsymbol{\mu}_i\mathbf{y}_i^{'}\right) = \mathbb{E}\left[\boldsymbol{\mu}_i\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i\right)^{'}\right] = \mathbb{E}\left(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\right)\mathbf{Z}_i^{'} = \boldsymbol{\Omega}_0\mathbf{Z}_i^{'} = \left(\mathbf{Z}_i\boldsymbol{\Omega}_0\right)^{'}$$

It follows that the distribution of the random vector $\left(\boldsymbol{\mu}_i,\mathbf{y}_i\right)^{'}$ writes:

$$\begin{pmatrix}\boldsymbol{\mu}_i\\ \mathbf{y}_i\end{pmatrix} \sim \mathscr{N}\left[\begin{pmatrix}\mathbf{0}\\ \mathbf{X}_i\boldsymbol{\beta}\end{pmatrix},\ \begin{pmatrix}\boldsymbol{\Omega}_0 & \left(\mathbf{Z}_i\boldsymbol{\Omega}_0\right)^{'}\\ \mathbf{Z}_i\boldsymbol{\Omega}_0 & \mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\end{pmatrix}\right]$$

We can refer to the results laid out at section 2. The conditional distribution of $\boldsymbol{\mu}_i\mid\mathbf{y}_i \sim \mathscr{N}(\mathbf{m}_i,\ \mathbf{V}_i)$, with:

$$\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right) = \left(\mathbf{Z}_i\boldsymbol{\Omega}_0\right)^{'}\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) = \mathbf{m}_i$$
$$\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right) = \boldsymbol{\Omega}_0 - \left(\mathbf{Z}_i\boldsymbol{\Omega}_0\right)^{'}\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\right) = \mathbf{V}_i$$
$$\mathbb{E}\left(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\mid\mathbf{y}_i\right) = \mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right) + \mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)^{'}$$

It follows that:

$$
\begin{aligned}
\mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right) &= \mathbb{E}\left[\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right) \mid \mathbf{y}_i\right] = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) \\
\mathbb{V}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right) &= \mathbb{V}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) = \mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\mathbf{Z}_i^{'} \\
\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'} \mid \mathbf{y}_i\right) &= \mathbb{V}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right) + \mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right)^{'} = \mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\mathbf{Z}_i^{'} + \mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right)^{'}
\end{aligned}
$$

In order to make the M-step more tractable, we need to rewrite the expressions of $\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$ and $\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$ before introducing them into the conditional expectation of the complete data likelihood. Following the suggestion in Pawitan (2001, p. 358), we write:

$$
\begin{aligned}
\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) &= \left(\mathbf{Z}_i\boldsymbol{\Omega}_0\right)^{'}\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \\
&= \underbrace{\left(\frac{1}{\sigma_0^2}\mathbf{Z}_i^{'}\mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1}\left(\frac{1}{\sigma_0^2}\mathbf{Z}_i^{'}\mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)}_{\mathrm{Id}_J}\left(\boldsymbol{\Omega}_0\mathbf{Z}_i^{'}\right)\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \\
&= \left(\frac{1}{\sigma_0^2}\mathbf{Z}_i^{'}\mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1}\left(\frac{1}{\sigma_0^2}\mathbf{Z}_i^{'}\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \mathbf{Z}_i^{'}\right)\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \\
&= \left(\frac{1}{\sigma_0^2}\mathbf{Z}_i^{'}\mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1}\left(\frac{1}{\sigma_0^2}\mathbf{Z}_i^{'}\right)\underbrace{\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}}_{\mathrm{Id}_{T_i}}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \\
\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) &= \frac{1}{\sigma_0^2}\left(\frac{1}{\sigma_0^2}\mathbf{Z}_i^{'}\mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1}\mathbf{Z}_i^{'}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)
\end{aligned}
$$

Remark: Comparing the initial and final expression of $\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$, we have demonstrated that $\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\right)^{'}\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i^{'} + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1} = \frac{1}{\sigma_0^2}\left(\frac{1}{\sigma_0^2}\mathbf{Z}_i^{'}\mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1}\mathbf{Z}_i^{'}$. We can then rewrite easily $\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$ as follow:

$$
\begin{aligned}
\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) &= \boldsymbol{\Omega}_0 - \left(\mathbf{Z}_i \boldsymbol{\Omega}_0\right)^{'} \left(\mathbf{Z}_i \boldsymbol{\Omega}_0 \mathbf{Z}_i^{'} + \sigma_0^2 \mathrm{Id}_{T_i}\right)^{-1} \left(\mathbf{Z}_i \boldsymbol{\Omega}_0\right) \\
&= \boldsymbol{\Omega}_0 - \frac{1}{\sigma_0^2} \left(\frac{1}{\sigma_0^2} \mathbf{Z}_i^{'} \mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1} \mathbf{Z}_i^{'} \left(\mathbf{Z}_i \boldsymbol{\Omega}_0\right) \\
&= \left[\mathrm{Id}_J - \frac{1}{\sigma_0^2} \left(\frac{1}{\sigma_0^2} \mathbf{Z}_i^{'} \mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1} \mathbf{Z}_i^{'} \mathbf{Z}_i\right] \boldsymbol{\Omega}_0 \\
&= \left(\frac{1}{\sigma_0^2} \mathbf{Z}_i^{'} \mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1} \left[\left(\frac{1}{\sigma_0^2} \mathbf{Z}_i^{'} \mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right) - \frac{1}{\sigma_0^2} \mathbf{Z}_i^{'} \mathbf{Z}_i\right] \boldsymbol{\Omega}_0 \\
&= \left(\frac{1}{\sigma_0^2} \mathbf{Z}_i^{'} \mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1} \left[\boldsymbol{\Omega}_0^{-1}\right] \boldsymbol{\Omega}_0 \\
&= \left(\frac{1}{\sigma_0^2} \mathbf{Z}_i^{'} \mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1}
\end{aligned}
$$

It is worth reminding that expressions of $\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$ and $\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$ above are evaluated at $\boldsymbol{\gamma}_0^{(k)} = \left(\boldsymbol{\beta}^{(k)}, \boldsymbol{\Omega}_0^{(k)}, \sigma_0^{2,(k)}\right)$. In the first order conditions of the M-step, terms of $\boldsymbol{\gamma}_0$ contained in $\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$ and $\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$ will be considered as known from the previous iteration.

To sum up:

$$
\begin{aligned}
\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\gamma_0\right)\mid\mathbf{y}\right] &= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)\right]\\
&-\frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\left[\boldsymbol{\Omega}_0^{-1}\mathbb{E}\left(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\mid\mathbf{y}_i\right)\right]\\
&= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\operatorname{tr}\left(\mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\mathbf{Z}_i^{'}+\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)^{'}\right)\\
&-\frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\left[\boldsymbol{\Omega}_0^{-1}\left(\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)+\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)^{'}\right)\right]\\
&= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\operatorname{tr}\left(\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)^{'}\right)\\
&-\frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\left(\boldsymbol{\Omega}_0^{-1}\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)^{'}\right)\\
&-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\operatorname{tr}\left(\mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\mathbf{Z}_i^{'}\right)-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\left(\boldsymbol{\Omega}_0^{-1}\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\right)\\
\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\gamma_0\right)\mid\mathbf{y}\right] &= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\operatorname{tr}\left(\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)^{'}\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)\right)\\
&-\frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)^{'}\boldsymbol{\Omega}_0^{-1}\mathbb{E}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\right]\\
&-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\operatorname{tr}\left(\mathbf{Z}_i^{'}\mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\right)-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\left(\boldsymbol{\Omega}_0^{-1}\mathbb{V}\left(\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\right)
\end{aligned}
$$

The last expression of $\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\gamma_0\right)\mid\mathbf{y}\right]$ is the so-called Q function, $Q\left(\gamma_0\mid\gamma_0^{(k)}\right)$.

3.2 The M-step

The separability of the log-likelihood function in the parameters γ_0 implies that optimal value of $\left(\boldsymbol{\beta},\sigma_0^2\right)$ can be obtained by maximizing only $\sum_{i=1}^{N}\mathbb{E}\left[\ln\mathbb{P}\left(\mathbf{y}_i\mid\boldsymbol{\mu}_i;\boldsymbol{\beta},\sigma_0^2\right)\mid\mathbf{y}_i,\gamma_0^{(k)}\right]$, and optimal value of $\boldsymbol{\Omega}_0$ can be obtained by maximizing only $\sum_{i=1}^{N}\mathbb{E}\left[\ln\mathbb{P}\left(\boldsymbol{\mu}_i;\boldsymbol{\Omega}_0\right)\mid\mathbf{y}_i,\gamma_0^{(k)}\right]$.

$$\left(\boldsymbol{\beta}, \sigma_0^2\right) = \underset{\boldsymbol{\beta}, \sigma_0^2}{\arg\max} \left\{ \begin{array}{l} -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma_0^2\right) - \frac{1}{2\sigma_0^2}\sum_{i=1}^{N} \operatorname{tr}\left(\mathbf{Z}_i'\mathbf{Z}_i \mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right) \\ -\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right]'\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right] \end{array} \right\}$$

$$\boldsymbol{\Omega}_0 = \underset{\boldsymbol{\Omega}_0}{\arg\max} \left\{ \begin{array}{l} -\frac{N \times J}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Omega}_0| - \frac{1}{2}\sum_{i=1}^{N} \operatorname{tr}\left(\boldsymbol{\Omega}_0^{-1}\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right) \\ -\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)'\boldsymbol{\Omega}_0^{-1}\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) \end{array} \right\}$$

In the matrix cookbook, we know that for any symmetric matrix **W**, and for any vectors x,s, and conformable matrices **A, B** and square matrix **D**:

$$\begin{cases} \frac{\partial}{\partial s}(\mathrm{x} - \mathrm{As})'\mathbf{W}(\mathrm{x} - \mathrm{As}) & = -2\mathrm{A}'\mathbf{W}(\mathrm{x} - \mathrm{As}) \\ \frac{\partial}{\partial \mathbf{D}}\ln\det(\mathbf{D}) & = \left(\mathbf{D}^{-1}\right)' = \left(\mathbf{D}'\right)^{-1} \\ \frac{\partial}{\partial \mathbf{D}}\mathrm{x}'\mathbf{D}^{-1}\mathrm{s} & = -\left(\mathbf{D}^{-1}\right)'\mathrm{xs}'\left(\mathbf{D}^{-1}\right)' \\ \frac{\partial}{\partial \mathbf{X}}\operatorname{tr}\left(\mathbf{A}\mathbf{X}^{-1}\mathbf{B}\right) & = -\left(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1}\right)' \end{cases}$$

The first order conditions for $\boldsymbol{\beta}$ (the matrix **W** is an identity matrix in our case):

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\gamma}_0 \mid \boldsymbol{\gamma}_0^{(k)}) = 0 &\Longleftrightarrow -\frac{1}{2\sigma_0^2}\sum_{i=1}^{N} -2\mathbf{X}_i'\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right] = 0 \\ &\Longleftrightarrow \sum_{i=1}^{N}\mathbf{X}_i'\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right] = 0 \\ &\Longleftrightarrow \sum_{i=1}^{N}\mathbf{X}_i'\left[\mathbf{y}_i - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right] = \sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\boldsymbol{\beta} \\ &\Longleftrightarrow \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\left[\mathbf{y}_i - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right] = \boldsymbol{\beta}^{(k+1)} \end{aligned}$$

where parameters $\boldsymbol{\gamma}_0$ present in $\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$ are evaluated at $\boldsymbol{\gamma}_0^{(k)}$.

The first-order conditions for σ_0^2:

$$\begin{aligned} \frac{\partial}{\partial \sigma_0^2} Q(\boldsymbol{\gamma}_0 \mid \boldsymbol{\gamma}_0^{(k)}) = 0 &\Longleftrightarrow -\frac{n}{2}\frac{1}{\sigma_0^2} - \frac{1}{2}\left(-\frac{1}{\sigma_0^4}\right)\sum_{i}^{N}\mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right)'\mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right) - \frac{1}{2}\left(-\frac{1}{\sigma_0^4}\right)\sum_{i=1}^{N}\operatorname{tr}\left(\mathbf{Z}_i'\mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right) = 0 \\ &\Longleftrightarrow -n + \frac{1}{\sigma_0^2}\sum_{i}^{N}\mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right)'\mathbb{E}\left(\boldsymbol{\varepsilon}_i \mid \mathbf{y}_i\right) + \frac{1}{\sigma_0^2}\sum_{i}^{N}\operatorname{tr}\left(\mathbf{Z}_i'\mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right) = 0 \end{aligned}$$

Finally, the update of parameter σ_0^2 writes:

$$\sigma_0^{2,(k+1)} = \frac{1}{n}\sum_i^N \left\{ \left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(k+1)} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right]' \left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(k+1)} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right] + \operatorname{tr}\left(\mathbf{Z}_i'\mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right)\right\}$$

The first-order conditions for Ω_0:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\Omega}_0} Q(\boldsymbol{\gamma}_0 \mid \boldsymbol{\gamma}_0^{(k)}) = 0 &\iff -\frac{N}{2}\boldsymbol{\Omega}_0^{-1} - \frac{1}{2}\sum_i^N \left[-\boldsymbol{\Omega}_0^{-1}\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)'\boldsymbol{\Omega}_0^{-1} - \left(\boldsymbol{\Omega}_0^{-1}\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\boldsymbol{\Omega}_0^{-1}\right)'\right] = 0 \\
&\iff \frac{1}{2}\boldsymbol{\Omega}_0^{-1}\left[-N + \left(\sum_i^N \mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)'\right)\boldsymbol{\Omega}_0^{-1} + \left(\sum_i^N \mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right)\boldsymbol{\Omega}_0^{-1}\right] = 0 \\
&\iff \left(\sum_i^N \mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)' + \mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right)\boldsymbol{\Omega}_0^{-1} = N \\
&\iff \frac{1}{N}\sum_i^N \left[\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)' + \mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right] = \boldsymbol{\Omega}_0^{(k+1)}
\end{aligned}$$

3.3 Summary of the EM algorithm for normal LME models

Parameter vector $\boldsymbol{\gamma}_0$ is evaluated at its value at iteration k in the conditional expectations. The EM algorithm involves inverting matrices: the smaller the dimension of a matrix is, the faster it will be inverted. Therefore, when the number of observations per individual, T_i, is larger than the number of random coefficients (*i.e.,* the number of columns of matrix $\mathbf{Z}_i$), the following formula will be preferred:

3.3.1 The E-step when $\forall i,\ T_i > J$

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) &= \frac{1}{\sigma_0^2}\left(\frac{1}{\sigma_0^2}\mathbf{Z}_i'\mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1}\mathbf{Z}_i'\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \\
\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) &= \left(\frac{1}{\sigma_0^2}\mathbf{Z}_i'\mathbf{Z}_i + \boldsymbol{\Omega}_0^{-1}\right)^{-1}
\end{aligned}$$

3.3.2 The E-step when $\forall i,\ T_i < J$

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) &= (\mathbf{Z}_i\boldsymbol{\Omega}_0)'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i' + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) \\
\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right) &= \boldsymbol{\Omega}_0 - (\mathbf{Z}_i\boldsymbol{\Omega}_0)'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i' + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}(\mathbf{Z}_i\boldsymbol{\Omega}_0)
\end{aligned}$$

3.3.3 The M-step

Next value of parameter vector $\boldsymbol{\gamma}_0^{(k+1)}$ is obtained by maximizing the expectation of the complete-data log-likelihood, conditionnally to the observed data, leading to the following expressions:

$$\boldsymbol{\beta}^{(k+1)} = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \left[\mathbf{y}_i - \mathbf{Z}_i \mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right]$$

$$\sigma_0^{2,(k+1)} = \frac{1}{n}\sum_{i}^{N} \left\{\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(k+1)} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right]' \left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(k+1)} - \mathbf{Z}_i\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right] + \operatorname{tr}\left(\mathbf{Z}_i'\mathbf{Z}_i\mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right)\right\}$$

$$\boldsymbol{\Omega}_0^{(k+1)} = \frac{1}{N}\sum_{i}^{N} \left[\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)' + \mathbb{V}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)\right]$$

3.4 The asymptotic variance of the estimator of the normal LME model

The EM algorithm is appealing by its simplicity to obtain MLE of LME models. However, its does not provide estimates of the variance of the MLE, required for inference purposes. The MLE is consistent and asymptotically normal under mild regularity conditions. Its asymptotic variance is given by the inverse of the information matrix. The information matrix can be derived as the opposite of the second-order derivative of the log-likelihood of the observed data. In the case of LME models, the second order derivative is complicated to obtain in close-form for the variance of the random effects. Then, we compute the score of each observation and approximate the information matrix as the cross-product of the scores.

3.4.1 The information matrix of $\boldsymbol{\beta}$

We write down the observed data log-likelihood (hereafter OD-LL). The likelihood of observing individual i writes:

$$\mathbb{P}\left(\mathbf{y}_i;\boldsymbol{\gamma}_0\right) = \frac{1}{\sqrt{(2\pi)^{T_i}} \mid \mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i' + \sigma_0^2\mathrm{Id}_{T_i} \mid^{1/2}} \exp\left[-\frac{1}{2}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i' + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right]$$

The sample log-likelihood writes:

$$\mathscr{L}\left(\mathbf{y};\boldsymbol{\gamma}_0\right) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln \mid \mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i' + \sigma_0^2\mathrm{Id}_{T_i} \mid - \frac{1}{2}\sum_{i}^{N}\left[\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i' + \sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right]$$

Let's state the Woodbury identity (found in the matrix cookbook):

$$\left(A+CBC'\right)^{-1}=A^{-1}-A^{-1}C\left(B^{-1}+C'A^{-1}C\right)^{-1}C'A^{-1}$$

It follows that:

$$\begin{aligned}\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i'+\sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}&=\frac{1}{\sigma_0^2}\mathrm{Id}_{T_i}-\frac{1}{\sigma_0^2}\mathrm{Id}_{T_i}\mathbf{Z}_i\left(\boldsymbol{\Omega}^{-1}+\mathbf{Z}_i'\frac{1}{\sigma_0^2}\mathrm{Id}_{T_i}\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i'\frac{1}{\sigma_0^2}\mathrm{Id}_{T_i}\\&=\frac{1}{\sigma_0^2}\left[\mathrm{Id}_{T_i}-\mathbf{Z}_i\left(\boldsymbol{\Omega}^{-1}+\frac{1}{\sigma_0^2}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i'\frac{1}{\sigma_0^2}\right]\\&=\frac{1}{\sigma_0^2}\left[\mathrm{Id}_{T_i}-\mathbf{Z}_i\left(\sigma_0^2\boldsymbol{\Omega}^{-1}+\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i'\right]\end{aligned}$$

The above expression of the inverse can be used if $T_i > J$, allowing to speed up the computation of the asymptotic variance to be derived below.

Using again the hints of the matrix cookbook stated above, we derive the second-order derivative of parameter vector $\boldsymbol{\beta}$:

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}}\mathcal{L}\left(\mathbf{y};\boldsymbol{\gamma}_0\right)&=-\frac{1}{2}\sum_i^N-2\mathbf{X}_i'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i'+\sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}\right)\\&=\sum_i^N\mathbf{X}_i'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i'+\sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\mathbf{y}_i-\left(\sum_i^N\mathbf{X}_i'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i'+\sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\mathbf{X}_i\right)\boldsymbol{\beta}\\\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\mathcal{L}\left(\mathbf{y};\boldsymbol{\gamma}_0\right)&=-\left(\sum_i^N\mathbf{X}_i'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i'+\sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\mathbf{X}_i\right)\\\mathbf{I}\left(\boldsymbol{\beta}\right)=-\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\mathcal{L}\left(\mathbf{y};\boldsymbol{\gamma}_0\right)&=\left(\sum_i^N\mathbf{X}_i'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i'+\sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\mathbf{X}_i\right)\end{aligned}$$

Therefore, the asymptotic variance of parameter vector $\boldsymbol{\beta}$ writes:

$$\mathbb{V}_{\mathrm{asym}}\left(\boldsymbol{\beta}\right)=\mathbf{I}\left(\boldsymbol{\beta}\right)^{-1}=\left(\sum_i^N\mathbf{X}_i'\left(\mathbf{Z}_i\boldsymbol{\Omega}_0\mathbf{Z}_i'+\sigma_0^2\mathrm{Id}_{T_i}\right)^{-1}\mathbf{X}_i\right)^{-1}$$

where the parameters $\boldsymbol{\Omega}_0$ and $\boldsymbol{\sigma}_0^2$ are replaced by their ML estimates.

The standard error of each of the K scalar parameters stacked in vector $\boldsymbol{\beta}$ is obtained as the square root of the diagonal of the asymptotic variance $\mathbb{V}_{\mathrm{asym}}\left(\boldsymbol{\beta}\right)$.

3.4.2 The information matrix of variance parameters $\boldsymbol{\Omega}_0$ and σ_0^2

It is difficult to obtain the second order derivative of the OD-LL in closed form. As a result, we approximate the information matrix by the cross-product of the score of each individual[4].

The log-likelihood of the complete data for individual i writes:

$$\ln L_i\left(\mathbf{y}_i,\boldsymbol{\mu}_i;\boldsymbol{\gamma}_0\right) = -\frac{T_i}{2}\ln(2\pi) - \frac{T_i}{2}\ln\left(\sigma_0^2\right) - \frac{1}{2\sigma_0^2}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)'\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right) - \frac{J}{2}\ln(2\pi) - \frac{1}{2}\ln|\,\boldsymbol{\Omega}_0\,| - \frac{1}{2}\boldsymbol{\mu}_i'\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_i$$

As a matter of verification, we will first write the information matrix of parameter $\boldsymbol{\beta}$. We note $\text{SC}_i\left(\boldsymbol{\beta}\right)$ the row vector (of length K) of scores of the parameter $\boldsymbol{\beta}$.

$$\text{SC}_i\left(\boldsymbol{\beta}\right) = \frac{\partial}{\partial\boldsymbol{\beta}}\ln L_i\left(\mathbf{y}_i,\boldsymbol{\mu}_i;\boldsymbol{\gamma}_0\right) = -\frac{1}{2\sigma_0^2}\left[-2\mathbf{X}_i'\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)\right] = \frac{1}{\sigma_0^2}\mathbf{X}_i'\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)$$

where $\boldsymbol{\mu}_i$ is replaced by $\mathbb{E}\left(\boldsymbol{\mu}_i\,|\,\mathbf{y}_i\right)$ and parameters $\boldsymbol{\beta}$ and σ_0^2 are replaced by their ML estimates. The N row vectors for the N individuals are stacked to form a matrix of dimensions $(N\times K)$, noted $\text{SC}\left(\boldsymbol{\beta}\right)$. Then, the approximate information matrix of $\boldsymbol{\beta}$ and its asymptotic variance, of dimension $(K\times K)$ writes:

$$\mathbf{I}\left(\boldsymbol{\beta}\right) = \text{SC}\left(\boldsymbol{\beta}\right)'\text{SC}\left(\boldsymbol{\beta}\right) \text{ and } \mathbb{V}_{\text{asym}}\left(\boldsymbol{\beta}\right) = \left[\text{SC}\left(\boldsymbol{\beta}\right)'\text{SC}\left(\boldsymbol{\beta}\right)\right]^{-1}$$

We do the same for the variance parameters:

$$\text{SC}_i\left(\sigma_0^2\right) = -\frac{T_i}{2\sigma_0^2} + \frac{1}{2(\sigma_0^2)^2}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)'\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\boldsymbol{\mu}_i\right)$$
$$\mathbf{I}\left(\sigma_0^2\right) = \text{SC}\left(\sigma_0^2\right)'\text{SC}\left(\sigma_0^2\right) \text{ and } \mathbb{V}_{\text{asym}}\left(\sigma_0^2\right) = \left[\text{SC}\left(\sigma_0^2\right)'\text{SC}\left(\sigma_0^2\right)\right]^{-1}$$

$\boldsymbol{\Omega}$ is a symmetric matrix. To avoid redundancy, we will compute the score of elements of lower triangular matrix, obtained with the ”vech” function, which outputs a vector

[4] We could also use this method to obtain the information matrix of parameter $\boldsymbol{\beta}$

of length $\frac{J(J+1)}{2}$.

$$\mathrm{SC}_i(\boldsymbol{\Omega}_0) = \mathrm{vech}\left\{-\frac{1}{2}\left(\boldsymbol{\Omega}_0^{-1}\right)'\left[\mathrm{Id}_J - \boldsymbol{\mu}_i\boldsymbol{\mu}_i'\left(\boldsymbol{\Omega}_0^{-1}\right)'\right]\right\}$$

$$\mathbf{I}(\Omega_0) = \mathrm{invvech}\left[\mathrm{diag}\left(\mathrm{SC}(\Omega_0)'\,\mathrm{SC}(\Omega_0)\right)\right] \text{ and } \mathbb{V}_{\mathrm{asym}}(\Omega_0) = [\mathbf{I}(\Omega_0)]^{-1}$$

Parameters $\boldsymbol{\beta}, \sigma_0^2$ and $\boldsymbol{\Omega}_0$ are replaced by their ML estimates, and $\boldsymbol{\mu}_i$ is replaced by $\mathbb{E}\left(\boldsymbol{\mu}_i \mid \mathbf{y}_i\right)$.

4 The random effects model without exogeneous regressors

Dealing with this case, which looks much simpler than the previous one (all the parameters of this model are scalar), was inspired to me by the paper of Neath (2012, p. 46). The case with exogeneous regressors can be easily guessed at the end of the derivation. I will follow all the steps of the previous section. The data generating mechanism producing y_{it} writes:

(2) $y_{it} = \mu_0 + \alpha_i + \epsilon_{it}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$

where μ_0 is a constant scalar, α_i and ϵ_{it} are random variables, with no serial correlation $\left(\textit{i.e.}, \mathrm{cov}(\epsilon_{it}, \epsilon_{it'}) = 0 \text{ if } t \neq t' \text{ and } \mathrm{cov}(\epsilon_{it}, \epsilon_{i't}) = 0 \text{ if } i \neq i'\right)$. We further assume that α_i and ϵ_{it} are independent and identically distributed normal random variables: $\alpha_i \sim \mathcal{N}\left(0, \sigma_\alpha^2\right)$ and $\epsilon_{it} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$. For an individual i, the data generating mechanism producing $\mathbf{y}_i$ writes

$$\mathbf{y}_i = \mu_0 \mathbb{1}_i + \alpha_i \mathbb{1}_i + \boldsymbol{\varepsilon}_i \text{ for } i = 1, \ldots, N$$

where $\mathbf{y}_i$ and $\boldsymbol{\varepsilon}_i$ are defined as in the previous section, and $\mathbb{1}_i$ is a vector of ones, of length T_i.

Following the EM formalism, the observed data is $\left(\mathbf{y}_i\right)$ and the complete data is $\left(\mathbf{y}_i, \alpha_i\right)$ $\forall i = 1, \ldots, N$. The set of parameters to be estimated is $\boldsymbol{\gamma}_0 = \left(\mu_0, \sigma_\alpha^2, \sigma_\epsilon^2\right)$.

By the Bayes' law on conditional probability, we write:

$$L_i\left(\mathbf{y}_i, \alpha_i; \boldsymbol{\gamma}_0\right) = \mathbb{P}\left(\mathbf{y}_i \mid \alpha_i; \mu_0, \sigma_\epsilon^2\right) \times \mathbb{P}\left(\alpha_i; \sigma_\alpha^2\right)$$

The density function of $\mathbf{y}_i \mid \alpha_i$ and α_i write:

$$\mathbb{P}\left(\mathbf{y}_i \mid \alpha_i; \mu_0, \sigma_\epsilon^2\right) = \frac{1}{\sqrt{(2\pi)^{T_i} \mid \sigma_\epsilon^2 \mathrm{Id}_{T_i} \mid^{1/2}}} \exp\left[-\frac{1}{2\sigma_\epsilon^2}\left(\mathbf{y}_i - \mu_0 \mathbb{1}_i - \alpha_i \mathbb{1}_i\right)' \left(\mathbf{y}_i - \mu_0 \mathbb{1}_i - \alpha_i \mathbb{1}_i\right)\right]$$

$$\mathbb{P}\left(\alpha_i; \sigma_\alpha^2\right) = \frac{1}{\sqrt{(2\pi)\,\sigma_\alpha^2}} \exp\left(-\frac{\alpha_i^2}{2\sigma_\alpha^2}\right)$$

The log-likelihood of the complete data for individual i writes (reminding by the way that $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - \mu_0 \mathbb{1}_i - \alpha_i \mathbb{1}_i$):

$$\begin{aligned}
\ln L_i\left(\mathbf{y}_i, \alpha_i; \boldsymbol{\gamma}_0\right) &= \ln \mathbb{P}\left(\mathbf{y}_i \mid \alpha_i; \mu_0, \sigma_\epsilon^2\right) + \ln \mathbb{P}\left(\alpha_i; \sigma_\alpha^2\right) \\
&= -\frac{T_i}{2} \ln(2\pi) - \frac{T_i}{2} \ln\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2} \boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln\left(\sigma_\alpha^2\right) - \frac{1}{2\sigma_\alpha^2} \alpha_i^2
\end{aligned}$$

We note $\boldsymbol{\alpha} = (\alpha_i, i = 1, \ldots, N)$, the vector obtained by stacking all the α_i. The sample log-likelihood of the complete data writes :

$$\begin{aligned}
\mathscr{L}\left(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\gamma}_0\right) &= \sum_{i=1}^{N} \ln L_i\left(\mathbf{y}_i, \alpha_i; \boldsymbol{\gamma}_0\right) \\
&= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^{N} \boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i - \frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln\left(\sigma_\alpha^2\right) - \frac{1}{2\sigma_\alpha^2} \sum_{i=1}^{N} \alpha_i^2
\end{aligned}$$

In the EM framework, at iteration $k+1$, we aim to maximize the conditional expectation of $\mathscr{L}\left(\mathbf{y}, \boldsymbol{\alpha}; \boldsymbol{\gamma}_0\right)$, given $\mathbf{y}$ and given $\boldsymbol{\gamma}_0^{(k)}$, the value taken by the parameter vector $\boldsymbol{\gamma}_0$ at iteration (k). We keep the same conventions as in the previous section.

4.1 The E-step

We will move faster here (possibly providing some computational details in the appendix), as we made detailled derivation in the previous section. The procedures are the same.

$$
\begin{aligned}
\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\alpha};\boldsymbol{\gamma}_0\right)\mid\mathbf{y}\right] &= \sum_{i=1}^{N}\mathbb{E}\left[\ln L_i\left(\mathbf{y}_i,\alpha_i;\boldsymbol{\gamma}_0\right)\mid\mathbf{y}_i\right] = -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_\epsilon^2\right)\\
&-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{N}\mathbb{E}\left(\boldsymbol{\varepsilon}_i^{'}\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)-\frac{N}{2}\ln(2\pi)-\frac{N}{2}\ln\left(\sigma_\alpha^2\right)-\frac{1}{2\sigma_\alpha^2}\sum_{i=1}^{N}\mathbb{E}\left(\alpha_i^2\mid\mathbf{y}_i\right)\\
&= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_\epsilon^2\right)-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{N}\operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)\right]\\
&-\frac{N}{2}\ln(2\pi)-\frac{N}{2}\ln\left(\sigma_\alpha^2\right)-\frac{1}{2\sigma_\alpha^2}\sum_{i=1}^{N}\mathbb{E}\left(\alpha_i^2\mid\mathbf{y}_i\right)
\end{aligned}
$$

We need to express $\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)$ and $\mathbb{E}\left(\alpha_i^2\mid\mathbf{y}_i\right)$. Recall that:
$\mathbf{y}_i\sim\mathcal{N}(\mu_0\mathbb{1}_i,\ \sigma_\alpha^2\mathbb{1}_i\mathbb{1}_i^{'}+\sigma_\epsilon^2\mathrm{Id}_{T_i})$; $\alpha_i\sim\mathcal{N}(0,\ \sigma_\alpha^2)$

$$
\begin{aligned}
\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right) &= \mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\\
\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right) &= \mathbb{V}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)+\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)^{'}\\
&= \mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\mathbb{1}_i^{'}+\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\\
\operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)\right] &= \operatorname{tr}\left[\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\mathbb{1}_i^{'}+\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\right]\\
&= T_i\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\operatorname{tr}\left\{\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\right\}\\
&= T_i\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\operatorname{tr}\left\{\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\right\}\\
&= T_i\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]
\end{aligned}
$$

The first two moments of the conditional distribution of $\alpha_i\mid\mathbf{y}_i$ completely define $\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)$ and $\mathbb{E}\left(\alpha_i^2\mid\mathbf{y}_i\right)$. As a result, $\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)$ and $\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)$ are sufficient statistics for deriving the conditional expectation of the complete data log-likelihood.

The pairwise covariance between $\mathbf{y}_i$ and α_i write:

$$
\begin{aligned}
&\operatorname{cov}\left(\mathbf{y}_i,\alpha_i\right)=\mathbb{E}\left(\mathbf{y}_i\alpha_i\right)-\mathbb{E}\left(\mathbf{y}_i\right)\mathbb{E}\left(\alpha_i\right)=\mathbb{E}\left(\mathbf{y}_i\alpha_i\right)=\mathbb{E}\left[\left(\mu_0\mathbb{1}_i+\alpha_i\mathbb{1}_i+\boldsymbol{\varepsilon}_i\right)\alpha_i\right]=\mathbb{E}\left(\alpha_i^2\right)\mathbb{1}_i=\sigma_\alpha^2\mathbb{1}_i\\
&\operatorname{cov}\left(\alpha_i,\mathbf{y}_i\right)=\mathbb{E}\left(\alpha_i,\mathbf{y}_i^{'}\right)=\mathbb{E}\left[\alpha_i\left(\mu_0\mathbb{1}_i+\alpha_i\mathbb{1}_i+\boldsymbol{\varepsilon}_i\right)^{'}\right]=\mathbb{E}\left(\alpha_i^2\right)\mathbb{1}_i^{'}=\sigma_\alpha^2\mathbb{1}_i^{'}
\end{aligned}
$$

It follows that the distribution of the random vector $\left(\alpha_i,\mathbf{y}_i\right)^{'}$ writes:

$$
\begin{pmatrix}\alpha_i\\ \mathbf{y}_i\end{pmatrix}\sim\mathcal{N}\left[\begin{pmatrix}0\\ \mu_0\mathbb{1}_i\end{pmatrix},\begin{pmatrix}\sigma_\alpha^2 & \sigma_\alpha^2\mathbb{1}_i^{'}\\ \sigma_\alpha^2\mathbb{1}_i & \sigma_\alpha^2\mathbb{1}_i\mathbb{1}_i^{'}+\sigma_\epsilon^2\mathrm{Id}_{T_i}\end{pmatrix}\right]
$$

We note: $\boldsymbol{\Psi}_i = \sigma_\alpha^2 \mathbb{1}_i \mathbb{1}_i^{'} + \sigma_\epsilon^2 \mathrm{Id}_{T_i}$. The conditional distribution $\alpha_i \mid \mathbf{y}_i \sim \mathcal{N}(\mathrm{m}_i, \mathrm{V}_i)$:

$$\mathrm{m}_i = \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right) = \sigma_\alpha^2 \mathbb{1}_i^{'} \boldsymbol{\Psi}_i^{-1} \left(\mathbf{y}_i - \mu_0 \mathbb{1}_i\right)$$
$$\mathrm{V}_i = \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) = \sigma_\alpha^2 - \sigma_\alpha^2 \mathbb{1}_i^{'} \boldsymbol{\Psi}_i^{-1} \sigma_\alpha^2 \mathbb{1}_i$$

We can express the inverse of $\boldsymbol{\Psi}_i$[5] as:

$$\boldsymbol{\Psi}_i^{-1} = \frac{1}{\sigma_\epsilon^2}\left(\mathrm{Id}_{T_i} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2} \mathbb{1}_i \mathbb{1}_i^{'}\right)$$

It follows that we can rewrite m_i and V_i as[6]

$$\mathrm{m}_i = \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right) = \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2}\left(\sum_{t=1}^{T_i} \mathrm{y}_{it} - \mu_0 T_i\right)$$
$$\mathrm{V}_i = \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) = \frac{\sigma_\epsilon^2 \sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2}$$
$$\mathbb{E}\left(\alpha_i^2 \mid \mathbf{y}_i\right) = \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]^2$$

4.2 The M-step

$$\left(\mu_0, \sigma_\epsilon^2\right) = \underset{\mu_0, \sigma_\epsilon^2}{\arg\max} \left\{ -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{N} \mathrm{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^{'} \mid \mathrm{y}_{it}\right)\right] \right\}$$
$$\sigma_\alpha^2 = \underset{\sigma_\alpha^2}{\arg\max} \left\{ -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left(\sigma_\alpha^2\right) - \frac{1}{2\sigma_\alpha^2}\sum_{i=1}^{N} \mathbb{E}\left(\alpha_i^2 \mid \mathbf{y}_i\right) \right\}$$

Recalling that

$$\mathrm{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^{'} \mid \mathbf{y}_i\right)\right] = T_i \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbf{y}_i - \mu_0 \mathbb{1}_i - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]^{'} \left[\mathbf{y}_i - \mu_0 \mathbb{1}_i - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]$$
$$\mathbb{E}\left(\alpha_i^2 \mid \mathbf{y}_i\right) = \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]^2$$

[5] See appendix B.1 for detailled computation of that inverse
[6] See appendix B.2 for computational details

Optimality condition for μ_0:

$$
\begin{aligned}
-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{N} -2\times \mathbb{1}_i^{'}\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]=0 \\
\iff \sum_{i=1}^{N}\left\{\mathbb{1}_i^{'}\left[\mathbf{y}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]-T_i\mu_0\right\}=0 \\
\iff \frac{1}{n}\sum_{i=1}^{N}\mathbb{1}_i^{'}\left[\mathbf{y}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]=\mu_0^{(k+1)}
\end{aligned}
$$

Optimality condition for σ_ϵ^2:

$$
\begin{aligned}
-\frac{n}{2\sigma_\epsilon^2}-\frac{1}{2}\left(-\frac{1}{\sigma_\epsilon^4}\right)\sum_{i=1}^{N}\Big[T_i\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\Big]=0 \\
\iff \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{N}\Big[T_i\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\Big]=n \\
\iff \frac{1}{n}\sum_{i=1}^{N}\Big[T_i\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i-\mu_0\mathbb{1}_i-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\Big]=\sigma_\epsilon^{2,(k+1)}
\end{aligned}
$$

Optimality condition for σ_α^2:

$$
\begin{aligned}
-\frac{N}{2\sigma_\alpha^2}-\frac{1}{2}\left(-\frac{1}{\sigma_\alpha^4}\right)\sum_{i=1}^{N}\left\{\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\left[\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\right]^2\right\}=0 \\
\iff \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\left[\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\right]^2\right\}=\sigma_\alpha^{2,(k+1)}
\end{aligned}
$$

4.3 Summary of the EM algorithm for normal random intercept models without exogeneous regressors

4.3.1 The E-Step

Computation of the sufficient statistics:

$$
\begin{aligned}
\mathrm{m}_i=\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)=\frac{\sigma_\alpha^2}{\sigma_\epsilon^2+T_i\sigma_\alpha^2}\left(\sum_{t=1}^{T_i}\mathrm{y}_{it}-\mu_0T_i\right) \\
\mathrm{V}_i=\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)=\frac{\sigma_\epsilon^2\sigma_\alpha^2}{\sigma_\epsilon^2+T_i\sigma_\alpha^2}
\end{aligned}
$$

4.3.2 The M-Step

Getting future optimal value for the parameters:

$$\mu_0^{(k+1)} = \frac{1}{n}\sum_{i=1}^{N} \mathbb{1}_i' \left[\mathbf{y}_i - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right] = \frac{1}{n}\sum_{i=1}^{N}\left[\sum_{t=1}^{T_i} \mathrm{y}_{it} - T_i\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]$$
$$\sigma_\epsilon^{2,(k+1)} = \frac{1}{n}\sum_{i=1}^{N}\left[T_i\mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbf{y}_i - \mu_0^{(k+1)}\mathbb{1}_i - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]'\left[\mathbf{y}_i - \mu_0^{(k+1)}\mathbb{1}_i - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]\right]$$
$$\sigma_\alpha^{2,(k+1)} = \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]^2\right\}$$

4.4 The asymptotic variance of the estimator of the random intercept model

The sample log-likelihood of the observed data writes:

$$\mathscr{L}\left(\mathbf{y};\gamma_0\right) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln|\mathbf{\Psi}_i| - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{y}_i - \mu_0\mathbb{1}_i - \alpha_i\mathbb{1}_i\right)'\mathbf{\Psi}_i^{-1}\left(\mathbf{y}_i - \mu_0\mathbb{1}_i - \alpha_i\mathbb{1}_i\right)$$

where $\mathbf{\Psi}_i$ and $\mathbf{\Psi}_i^{-1}$ are defined above:

$$\mathbf{\Psi}_i = \sigma_\alpha^2\mathbb{1}_i\mathbb{1}_i' + \sigma_\epsilon^2\mathrm{Id}_{T_i} \text{ and } \mathbf{\Psi}_i^{-1} = \frac{1}{\sigma_\epsilon^2}\left(\mathrm{Id}_{T_i} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i\sigma_\alpha^2}\mathbb{1}_i\mathbb{1}_i'\right)$$

Therefore, the information matrix of parameter μ_0 (and thus, the asymptotic variance of μ_0) can be derived from the second order derivative of the observed sample log-likelihood:

$$\frac{\partial}{\partial\mu_0}\mathscr{L}\left(\mathbf{y};\gamma_0\right) = -\frac{1}{2}\sum_{i=1}^{N} -2\times\mathbb{1}_i'\mathbf{\Psi}_i^{-1}\left(\mathbf{y}_i - \mu_0\mathbb{1}_i - \alpha_i\mathbb{1}_i\right) = \sum_{i=1}^{N}\mathbb{1}_i'\mathbf{\Psi}_i^{-1}\left(\mathbf{y}_i - \mu_0\mathbb{1}_i - \alpha_i\mathbb{1}_i\right)$$
$$\frac{\partial^2}{\partial\mu_0^2}\mathscr{L}\left(\mathbf{y};\gamma_0\right) = -\sum_{i=1}^{N}\mathbb{1}_i'\mathbf{\Psi}_i^{-1}\mathbb{1}_i$$
$$\mathbb{V}_{\text{asym}}\left(\mu_0\right) = \left(\sum_{i=1}^{N}\mathbb{1}_i'\mathbf{\Psi}_i^{-1}\mathbb{1}_i\right)^{-1}$$

For variance parameters, as before, we will rely on cross-product of scores. The complete data log-likelihood of individual i writes:

$$\ln L_i\left(\mathbf{y}_i,\alpha_i;\boldsymbol{\gamma}_0\right)=-\frac{T_i}{2}\ln(2\pi)-\frac{T_i}{2}\ln\left(\sigma_\epsilon^2\right)-\frac{1}{2\sigma_\epsilon^2}\left(\mathbf{y}_i-\mu_0\mathbb{1}_i-\alpha_i\mathbb{1}_i\right)'\left(\mathbf{y}_i-\mu_0\mathbb{1}_i-\alpha_i\mathbb{1}_i\right)$$
$$-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln\left(\sigma_\alpha^2\right)-\frac{\alpha_i^2}{2\sigma_\alpha^2}$$

Then the scores of the three parameters write:

$$\begin{aligned}
\mathrm{SC}_i\left(\mu_0\right)&=\frac{1}{\sigma_\epsilon^2}\mathbb{1}_i'\left(\mathbf{y}_i-\mu_0\mathbb{1}_i-\alpha_i\mathbb{1}_i\right)\\
\mathrm{SC}_i\left(\sigma_\epsilon^2\right)&=-\frac{T_i}{2\sigma_\epsilon^2}+\frac{1}{2(\sigma_\epsilon^2)^2}\left(\mathbf{y}_i-\mu_0\mathbb{1}_i-\alpha_i\mathbb{1}_i\right)'\left(\mathbf{y}_i-\mu_0\mathbb{1}_i-\alpha_i\mathbb{1}_i\right)\\
\mathrm{SC}_i\left(\sigma_\alpha^2\right)&=-\frac{1}{2\sigma_\alpha^2}+\frac{\alpha_i^2}{2(\sigma_\alpha^2)^2}
\end{aligned}$$

where α_i is replaced by $\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)$ and the other parameters are replaced with their ML estimates.

5 The random effects model with exogeneous regressors

The previous model (without exogeneous regressors) will be slightly modified as:

$\mathbf{y}_i=\mathbf{X}_i\boldsymbol{\beta}+\alpha_i\mathbb{1}_i+\boldsymbol{\varepsilon}_i$ for $i=1,\ldots,N$

where all the terms have already been defined. As before, the observed data is $\left(\mathbf{y}_i\right)$ and the complete data is $\left(\mathbf{y}_i,\alpha_i\right)$ $\forall i=1,\ldots,N$. The set of parameters to be estimated is $\boldsymbol{\gamma}_0=\left(\boldsymbol{\beta},\sigma_\alpha^2,\sigma_\epsilon^2\right)$.

The sample log-likelihood of the complete data writes :

$$\begin{aligned}
\mathscr{L}\left(\mathbf{y},\boldsymbol{\alpha};\boldsymbol{\gamma}_0\right)&=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_\epsilon^2\right)-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^N\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\alpha_i\mathbb{1}_i\right)'\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\alpha_i\mathbb{1}_i\right)\\
&-\frac{N}{2}\ln(2\pi)-\frac{N}{2}\ln\left(\sigma_\alpha^2\right)-\frac{1}{2\sigma_\alpha^2}\sum_{i=1}^N\alpha_i^2\\
&=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_\epsilon^2\right)-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^N\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}-\frac{N}{2}\ln(2\pi)-\frac{N}{2}\ln\left(\sigma_\alpha^2\right)-\frac{1}{2\sigma_\alpha^2}\sum_{i=1}^N\alpha_i^2
\end{aligned}$$

The conditional expectation of $\mathscr{L}\left(\mathbf{y},\boldsymbol{\alpha};\boldsymbol{\gamma}_0\right)$, given $\mathbf{y}$ and given $\boldsymbol{\gamma}_0^{(k)}$ writes:

$$\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\alpha};\boldsymbol{\gamma}_0\right)\mid\mathbf{y}\right]=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_\epsilon^2\right)-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^N\operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)\right]$$
$$-\frac{N}{2}\ln(2\pi)-\frac{N}{2}\ln\left(\sigma_\alpha^2\right)-\frac{1}{2\sigma_\alpha^2}\sum_{i=1}^N\mathbb{E}\left(\alpha_i^2\mid\mathbf{y}_i\right)$$

5.1 The E-step

$$\begin{aligned}
\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)&=\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\\
\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)&=\mathbb{V}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)+\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)\mathbb{E}\left(\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)^{'}\\
&=\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\mathbb{1}_i^{'}+\left[\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\left[\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\\
\operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)\right]&=\operatorname{tr}\left[\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\mathbb{1}_i^{'}+\left[\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\left[\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\right]\\
&=T_i\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\operatorname{tr}\left\{\left[\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]\left[\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\right\}\\
&=T_i\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\left[\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}-\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\mathbb{1}_i\right]
\end{aligned}$$

The distribution of the random vector $\left(\alpha_i,\mathbf{y}_i\right)^{'}$ writes:

$$\begin{pmatrix}\alpha_i\\ \mathbf{y}_i\end{pmatrix}\sim\mathscr{N}\left[\begin{pmatrix}0\\ \mathbf{X}_i\boldsymbol{\beta}\end{pmatrix},\begin{pmatrix}\sigma_\alpha^2 & \sigma_\alpha^2\mathbb{1}_i^{'}\\ \sigma_\alpha^2\mathbb{1}_i & \sigma_\alpha^2\mathbb{1}_i\mathbb{1}_i^{'}+\sigma_\epsilon^2\mathrm{Id}_{T_i}\end{pmatrix}\right]$$

We keep the same notation as before: $\boldsymbol{\Psi}_i=\sigma_\alpha^2\mathbb{1}_i\mathbb{1}_i^{'}+\sigma_\epsilon^2\mathrm{Id}_{T_i}$. The conditional distribution $\alpha_i\mid\mathbf{y}_i\sim\mathscr{N}\left(\mathrm{m}_i,\,\mathrm{V}_i\right)$:

$$\begin{aligned}
\mathrm{m}_i&=\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)=\sigma_\alpha^2\mathbb{1}_i^{'}\boldsymbol{\Psi}_i^{-1}\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}\right)\\
\mathrm{V}_i&=\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)=\sigma_\alpha^2-\sigma_\alpha^2\mathbb{1}_i^{'}\boldsymbol{\Psi}_i^{-1}\sigma_\alpha^2\mathbb{1}_i
\end{aligned}$$

The expression of $\boldsymbol{\Psi}_i$ hasn't changed; neither has its inverse. It follows that the corresponding expression of m_i and V_i can be deduced without difficulty:

$$\begin{aligned}
\mathrm{m}_i&=\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)=\frac{\sigma_\alpha^2}{\sigma_\epsilon^2+T_i\sigma_\alpha^2}\mathbb{1}_i^{'}\left(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta}\right)\\
\mathrm{V}_i&=\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)=\frac{\sigma_\epsilon^2\sigma_\alpha^2}{\sigma_\epsilon^2+T_i\sigma_\alpha^2}\\
\mathbb{E}\left(\alpha_i^2\mid\mathbf{y}_i\right)&=\mathbb{V}\left(\alpha_i\mid\mathbf{y}_i\right)+\left[\mathbb{E}\left(\alpha_i\mid\mathbf{y}_i\right)\right]^2
\end{aligned}$$

5.2 The M-step

$$
\begin{aligned}
\left(\boldsymbol{\beta}, \sigma_\epsilon^2\right) &= \underset{\boldsymbol{\beta}, \sigma_\epsilon^2}{\arg\max} \left\{ -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{N} \operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^{'} \mid \mathrm{y}_{it}\right)\right] \right\} \\
\sigma_\alpha^2 &= \underset{\sigma_\alpha^2}{\arg\max} \left\{ -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left(\sigma_\alpha^2\right) - \frac{1}{2\sigma_\alpha^2}\sum_{i=1}^{N} \mathbb{E}\left(\alpha_i^2 \mid \mathbf{y}_i\right) \right\}
\end{aligned}
$$

where:

$$
\begin{aligned}
\operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^{'} \mid \mathbf{y}_i\right)\right] &= T_i \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]^{'} \left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right] \\
\mathbb{E}\left(\alpha_i^2 \mid \mathbf{y}_i\right) &= \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]^2
\end{aligned}
$$

Optimality condition for $\boldsymbol{\beta}$:

$$
\begin{aligned}
-\frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{N} -2 \times \mathbf{X}_i^{'}\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right] &= 0 \\
\iff \sum_{i=1}^{N}\left\{\mathbf{X}_i^{'}\left[\mathbf{y}_i - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]\right\} &= \sum_{i=1}^{N}\mathbf{X}_i^{'}\mathbf{X}_i\boldsymbol{\beta} \\
\iff \left(\sum_{i=1}^{N}\mathbf{X}_i^{'}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i^{'}\left[\mathbf{y}_i - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right] &= \boldsymbol{\beta}^{(k+1)}
\end{aligned}
$$

Optimality condition for σ_ϵ^2:

$$
\begin{aligned}
-\frac{n}{2\sigma_\epsilon^2} - \frac{1}{2}\left(-\frac{1}{\sigma_\epsilon^4}\right)\sum_{i=1}^{N}\left[T_i \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]\right] &= 0 \\
\iff \frac{1}{\sigma_\epsilon^2}\sum_{i=1}^{N}\left[T_i \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]\right] &= n \\
\iff \frac{1}{n}\sum_{i=1}^{N}\left[T_i \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]^{'}\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\mathbb{1}_i\right]\right] &= \sigma_\epsilon^{2,(k+1)}
\end{aligned}
$$

Optimality condition for σ_α^2:

$$
\begin{aligned}
-\frac{N}{2\sigma_\alpha^2} - \frac{1}{2}\left(-\frac{1}{\sigma_\alpha^4}\right)\sum_{i=1}^{N}\left\{\mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]^2\right\} &= 0 \\
\iff \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]^2\right\} &= \sigma_\alpha^{2,(k+1)}
\end{aligned}
$$

5.3 Summary of the EM algorithm for normal random effect models with exogeneous regressors

5.3.1 The E-Step

$$\mathrm{m}_i = \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right) = \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2} \mathbb{1}_i^{'} \left(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}\right)$$

$$\mathrm{V}_i = \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) = \frac{\sigma_\epsilon^2 \sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2}$$

$$\mathbb{E}\left(\alpha_i^2 \mid \mathbf{y}_i\right) = \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]^2$$

5.3.2 The M-Step

$$\boldsymbol{\beta}^{(k+1)} = \left(\sum_{i=1}^{N} \mathbf{X}_i^{'} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^{'} \left[\mathbf{y}_i - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right) \mathbb{1}_i\right]$$

$$\sigma_\epsilon^{2,(k+1)} = \frac{1}{n} \sum_{i=1}^{N} \left[T_i \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(k+1)} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right) \mathbb{1}_i\right]^{'} \left[\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}^{(k+1)} - \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right) \mathbb{1}_i\right] \right]$$

$$\sigma_\alpha^{2,(k+1)} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) + \left[\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)\right]^2 \right\}$$

5.4 The asymptotic variance of the estimator of the random effects model with exogeneous regressors

The sample log-likelihood of the observed data writes:

$$\mathscr{L}\left(\mathbf{y}; \boldsymbol{\gamma}_0\right) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{N} \ln |\boldsymbol{\Psi}_i| - \frac{1}{2} \sum_{i=1}^{N} \left(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right)^{'} \boldsymbol{\Psi}_i^{-1} \left(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right)$$

where $\boldsymbol{\Psi}_i$ and $\boldsymbol{\Psi}_i^{-1}$ are defined above:

$$\boldsymbol{\Psi}_i = \sigma_\alpha^2 \mathbb{1}_i \mathbb{1}_i^{'} + \sigma_\epsilon^2 \mathrm{Id}_{T_i} \text{ and } \boldsymbol{\Psi}_i^{-1} = \frac{1}{\sigma_\epsilon^2} \left(\mathrm{Id}_{T_i} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2} \mathbb{1}_i \mathbb{1}_i^{'}\right)$$

Therefore, the information matrix of parameter μ_0 (and thus, the asymptotic variance of μ_0) can be derived from the second order derivative of the observed sample log-likelihood:

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} \mathscr{L}\left(\mathbf{y};\boldsymbol{\gamma}_0\right) &= -\frac{1}{2}\sum_{i=1}^{N} -2 \times \mathbf{X}_i^{'} \boldsymbol{\Psi}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right) = \sum_{i=1}^{N} \mathbf{X}_i^{'} \boldsymbol{\Psi}_i^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right) \\
\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^{'}} \mathscr{L}\left(\mathbf{y};\boldsymbol{\gamma}_0\right) &= -\sum_{i=1}^{N} \mathbf{X}_i^{'} \boldsymbol{\Psi}_i^{-1} \mathbf{X}_i \\
\mathbb{V}_{\text{asym}}\left(\mu_0\right) &= \left(\sum_{i=1}^{N} \mathbf{X}_i^{'} \boldsymbol{\Psi}_i^{-1} \mathbf{X}_i\right)^{-1}
\end{aligned}
$$

For variance parameters, as before, we will rely on cross-product of scores. The complete data log-likelihood of individual i writes:

$$
\begin{aligned}
\ln L_i\left(\mathbf{y}_i, \alpha_i; \boldsymbol{\gamma}_0\right) = &-\frac{T_i}{2}\ln(2\pi) - \frac{T_i}{2}\ln\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right)^{'}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right) \\
&- \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\sigma_\alpha^2\right) - \frac{\alpha_i^2}{2\sigma_\alpha^2}
\end{aligned}
$$

Then the scores of the three parameters write:

$$
\begin{aligned}
\mathrm{SC}_i\left(\boldsymbol{\beta}\right) &= \frac{1}{\sigma_\epsilon^2}\mathbf{X}_i^{'}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right) \\
\mathrm{SC}_i\left(\sigma_\epsilon^2\right) &= -\frac{T_i}{2\sigma_\epsilon^2} + \frac{1}{2(\sigma_\epsilon^2)^2}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right)^{'}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \alpha_i \mathbb{1}_i\right) \\
\mathrm{SC}_i\left(\sigma_\alpha^2\right) &= -\frac{1}{2\sigma_\alpha^2} + \frac{\alpha_i^2}{2(\sigma_\alpha^2)^2}
\end{aligned}
$$

where α_i is replaced by $\mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right)$ and the other parameters are replaced with their ML estimates.

References

Boyles, R. A. (1983), 'On the Convergence of the EM Algorithm', *Journal of the Royal Statistical Society: Series B (Methodological)* **45**(1), 47–50.

Davidson, J. (2018), *An Introduction to Econometric Theory*, Wiley.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum Likelihood from Incomplete Data Via the *EM* Algorithm', *Journal of the Royal Statistical Society* **39**(1), 1–22.

Neath, R. C. (2012), 'On convergence properties of the monte carlo em algorithm'.
URL: *https://arxiv.org/abs/1206.4768*

Pawitan, Y. (2001), *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford science publications, OUP Oxford.

Wu, C. F. J. (1983), 'On the Convergence Properties of the EM Algorithm', *The Annals of Statistics* **11**(1), 95–103.

A Linear mixed effects models

A.1 The conditional expectation of the complete-data LL

$$
\begin{aligned}
\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\gamma_0\right)\mid\mathbf{y}\right] &= \sum_{i=1}^{N}\mathbb{E}\left[\ln L_i\left(\mathbf{y}_i,\boldsymbol{\mu}_i;\gamma_0\right)\mid\mathbf{y}_i\right] \\
&= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\mathbb{E}\left(\boldsymbol{\varepsilon}_i^{'}\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right) \\
&- \frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}\left(\boldsymbol{\mu}_i^{'}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_i\mid\mathbf{y}_i\right) \\
&= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{\varepsilon}_i^{'}\boldsymbol{\varepsilon}_i\mid\mathbf{y}_i\right)\right] \\
&- \frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{\mu}_i^{'}\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_i\mid\mathbf{y}_i\right)\right] \\
&= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)\right] \\
&- \frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\mid\mathbf{y}_i\right)\right] \\
\mathbb{E}\left[\mathscr{L}\left(\mathbf{y},\boldsymbol{\mu};\gamma_0\right)\mid\mathbf{y}\right] &= -\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\left(\sigma_0^2\right)-\frac{1}{2\sigma_0^2}\sum_{i=1}^{N}\operatorname{tr}\left[\mathbb{E}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\mathbf{y}_i\right)\right] \\
&- \frac{N\times J}{2}\ln(2\pi)-\frac{N}{2}\ln\mid\boldsymbol{\Omega}_0\mid-\frac{1}{2}\sum_{i=1}^{N}\operatorname{tr}\left[\boldsymbol{\Omega}_0^{-1}\mathbb{E}\left(\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}\mid\mathbf{y}_i\right)\right]
\end{aligned}
$$

B Random effects models without exogeneous regressors

B.1 Computation of the inverse of Ψ_i matrix

Recall that $\boldsymbol{\Psi}_i = \sigma_\alpha^2 \mathbb{1}_i \mathbb{1}_i' + \sigma_\epsilon^2 \text{Id}_{T_i}$. Using the Sherman-Morrison formula found in the matrix cookbook, we have:

$$\begin{aligned}
\boldsymbol{\Psi}_i^{-1} &= \frac{1}{\sigma_\epsilon^2}\text{Id}_{T_i} - \frac{\frac{1}{\sigma_\epsilon^2}\text{Id}_{T_i}\sigma_\alpha^2 \mathbb{1}_i \mathbb{1}_i' \frac{1}{\sigma_\epsilon^2}\text{Id}_{T_i}}{1 + \mathbb{1}_i' \frac{1}{\sigma_\epsilon^2}\text{Id}_{T_i}\sigma_\alpha^2 \mathbb{1}_i} \\
&= \frac{1}{\sigma_\epsilon^2}\text{Id}_{T_i} - \frac{\frac{\sigma_\alpha^2}{\sigma_\epsilon^4}\text{Id}_{T_i}\mathbb{1}_i \mathbb{1}_i'\text{Id}_{T_i}}{1 + \frac{\sigma_\alpha^2}{\sigma_\epsilon^2}\mathbb{1}_i'\text{Id}_{T_i}\mathbb{1}_i} \quad \text{(Note that } \mathbb{1}_i'\text{Id}_{T_i}\mathbb{1}_i = T_i\text{)} \\
&= \frac{1}{\sigma_\epsilon^2}\text{Id}_{T_i} - \frac{\frac{\sigma_\alpha^2}{\sigma_\epsilon^4}\mathbb{1}_i \mathbb{1}_i'}{1 + T_i\frac{\sigma_\alpha^2}{\sigma_\epsilon^2}} \\
&= \frac{1}{\sigma_\epsilon^2}\text{Id}_{T_i} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^4}\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + T_i\sigma_\alpha^2}\mathbb{1}_i \mathbb{1}_i' \\
\boldsymbol{\Psi}_i^{-1} &= \frac{1}{\sigma_\epsilon^2}\left(\text{Id}_{T_i} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i\sigma_\alpha^2}\mathbb{1}_i \mathbb{1}_i'\right)
\end{aligned}$$

B.2 Rewriting the expressions of $\mathbf{m}_i$ and $\mathbf{V}_i$ taking into account the expression of Ψ_i^{-1}

We have shown that:

$$\begin{aligned}
\text{m}_i &= \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right) = \sigma_\alpha^2 \mathbb{1}_i' \boldsymbol{\Psi}_i^{-1}\left(\mathbf{y}_i - \mu_0 \mathbb{1}_i\right) \\
\text{V}_i &= \mathbb{V}\left(\alpha_i \mid \mathbf{y}_i\right) = \sigma_\alpha^2 - \sigma_\alpha^2 \mathbb{1}_i' \boldsymbol{\Psi}_i^{-1}\sigma_\alpha^2 \mathbb{1}_i
\end{aligned}$$

We have expressed the inverse of $\boldsymbol{\Psi}_i$ as:

$$\boldsymbol{\Psi}_i^{-1} = \frac{1}{\sigma_\epsilon^2}\left(\text{Id}_{T_i} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i\sigma_\alpha^2}\mathbb{1}_i \mathbb{1}_i'\right)$$

It follows that we can rewrite m_i and V_i as:

$$
\begin{aligned}
\mathrm{m}_i &= \mathbb{E}\left(\alpha_i \mid \mathbf{y}_i\right) = \sigma_\alpha^2 \mathbb{1}_i^{'} \mathbf{\Psi}_i^{-1} \left(\mathbf{y}_i - \mu_0 \mathbb{1}_i\right) \\
&= \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \mathbb{1}_i^{'} \left(\mathrm{Id}_{T_i} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2} \mathbb{1}_i \mathbb{1}_i^{'}\right) \left(\mathbf{y}_i - \mu_0 \mathbb{1}_i\right) \\
&= \left(\frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \mathbb{1}_i^{'} - \frac{\sigma_\alpha^4}{\sigma_\epsilon^2 \left(\sigma_\epsilon^2 + T_i \sigma_\alpha^2\right)} T_i \mathbb{1}_i^{'}\right) \left(\mathbf{y}_i - \mu_0 \mathbb{1}_i\right) \\
&= \left(\frac{\sigma_\alpha^2}{\sigma_\epsilon^2} - \frac{T_i \sigma_\alpha^4}{\sigma_\epsilon^2 \left(\sigma_\epsilon^2 + T_i \sigma_\alpha^2\right)}\right) \mathbb{1}_i^{'} \left(\mathbf{y}_i - \mu_0 \mathbb{1}_i\right) \\
&= \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \left(1 - \frac{T_i \sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2}\right) \mathbb{1}_i^{'} \left(\mathbf{y}_i - \mu_0 \mathbb{1}_i\right) \\
&= \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \times \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2} \mathbb{1}_i^{'} \left(\mathbf{y}_i - \mu_0 \mathbb{1}_i\right) \\
\mathrm{m}_i &= \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2} \left(\sum_{t=1}^{T_i} \mathrm{y}_{it} - T_i \mu_0\right) \\
\mathrm{V}_i &= \sigma_\alpha^2 \left(1 - \mathbb{1}_i^{'} \mathbf{\Psi}_i^{-1} \mathbb{1}_i \sigma_\alpha^2\right) \\
&= \sigma_\alpha^2 \left[1 - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \mathbb{1}_i^{'} \left(\mathrm{Id} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2} \mathbb{1}_i \mathbb{1}_i^{'}\right) \mathbb{1}_i\right] \\
&= \sigma_\alpha^2 \left[1 - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \left(\mathbb{1}_i^{'} - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2} T_i \mathbb{1}_i^{'}\right) \mathbb{1}_i\right] \\
&= \sigma_\alpha^2 \left[1 - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \left(T_i - \frac{\sigma_\alpha^2 T_i^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2}\right)\right] \\
&= \sigma_\alpha^2 \left[1 - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \left(\frac{T_i \sigma_\epsilon^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2}\right)\right] \\
&= \sigma_\alpha^2 \left(1 - \frac{T_i \sigma_\alpha^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2}\right) \\
\mathrm{V}_i &= \frac{\sigma_\alpha^2 \sigma_\epsilon^2}{\sigma_\epsilon^2 + T_i \sigma_\alpha^2}
\end{aligned}
$$