Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in

Scienze Statistiche

Modelling associations in multivariate longitudinal data with INLA

Supervisor Prof. Livio Finos
Dipartimento di Scienze Statistiche

Co-supervisor Dr. Roula Tsonaka and Dr. Bart Mertens
Medical Statistics section, Leiden University Medical Center

Laureando Chiara Degan
Matricola N 2023355

Anno Accademico 2022/2023

Contents

Abstract 5

Introduction 7

1 Integrated Nested Laplace Approximation (INLA) 11

1.1 Latent Gaussian models and GMRF 12

1.2 Model structure 13

1.3 INLA approximations 16

1.3.1 Approximation of $\tilde{\boldsymbol{\pi}}(\boldsymbol{\theta} \mid \mathbf{y})$ 16

1.3.2 Approximation of $\tilde{\boldsymbol{\pi}}\left(x_i \mid \boldsymbol{\theta}, \mathbf{y}\right)$ 18

1.4 Methods of evaluation 20

1.4.1 Bayes Factor 20

1.4.2 Information-based criteria 21

1.4.3 Predictive measures 22

1.5 Predictive distribution 24

2 Mixed Effects Models 25

2.1 Linear Mixed Effects model 26

2.1.1 INLA model formulation 27

2.2 Generalized Linear Mixed Model 29

2.2.1 Beta Mixed Model 31

2.2.2 INLA model formulation 33

3 Joint Models 35

3.1 Time-varying covariates 35

3.1.1 Exogenous covariates . . . 36
3.1.2 Endogenous covariates . . . 38
3.2 Joint Mixed Model . . . 39
3.2.1 Association between outcome and endogenous covariate 40
3.2.2 INLA model formulation . . . 43
3.3 Joint Scaled Model . . . 46
3.3.1 Association between outcome and endogenous covariate 47
3.3.2 INLA model formulation . . . 49

4 Simulation 51
4.1 Data generation process . . . 52
4.1.1 Simulating data from a JMM . . . 54
4.1.2 Simulating data from a JSM . . . 55
4.2 Results . . . 56
4.2.1 Comparison between models with large variances . . . 56
4.2.2 Comparison between models with small variances . . . 59

5 Application 67
5.1 Presentation of the dataset . . . 67
5.2 Exploring the mean structure . . . 69
5.3 Modelling . . . 72
5.3.1 Joint Mixed Model . . . 73
5.3.2 Joint Scaled Model . . . 74
5.4 Results . . . 75
5.4.1 Analyzing *protein1* . . . 75
5.4.2 Analyzing *protein2* . . . 78
5.4.3 Comments . . . 80

Conclusion 85

Bibliography 89

A R Code 93

Abstract

Measuring the relationship between the longitudinal response variable and time-varying covariates is not always trivial in longitudinal studies, and the use of simple linear mixed models is no longer appropriate, especially when covariates depend on the prior values of the outcome (endogenous time-varying covariates). The failure to account for the dependence between the endogenous variable and the outcome history introduces significant bias. Moreover, the longitudinal variables can be measured at different points in time and may contain missing values. All of these motivations led us to use several multivariate models to assess the association between the response variable and the endogenous covariates. Some induce the association via correlated random effects, called Joint Mixed Models; others use a scaling factor to estimate the association, called Joint Scaled Models. Fitting either of these models, however, is not straightforward, and their computational intensity, due to the potentially high-dimensional integration over the random effects terms, limits their applicability. A flexible Bayesian estimation approach, known as INLA, will be used to fill this gap. We will evaluate its performance and applicability in the context of joint longitudinal models. In particular, it has been evaluated in a scenario with a low population dimension, which generally leads to low accuracy and precision of the estimates, small variances and covariaces of the random effects, resulting in a priors selection problem, and longitudinal variables of different types, one being normally distributed and the other following a Beta distribution. We will present analysis results from a simulation study and a clinical study conducted at the Leiden University Medical Center (LUMC).

Introduction

In longitudinal studies, data from various groups, such as subjects or clusters, are collected over time. Hence, repeated measurements are taken for each subject/cluster. As a result, the observations are no longer independent, necessitating the use of a specific statistical model that takes into account the correlation between them. The most common longitudinal models are Linear Mixed effects Models (LMM) and Generalized Linear Mixed effects Models (GLMM). These models' primary goal is to investigate changes or trends in the response variable over time, as well as the relationships between the response variable and one or more predictor variables. However, such an operation is not always straightforward. Everything becomes more challenging when time-varying independent variables, particularly endogenous variables, are considered.

Time-varying covariates can be categorized as exogenous or endogenous. Exogenous variables, such as age or daily climate change, have values that change over time based solely on their previous values. Endogenous variables, on the other hand, are influenced not only by their own history but also by previous values of the response variable. Examples of endogenous variables include treatment doses and biomarkers. Studying the association between endogenous variables and the outcome can be challenging. Mixed effect models may not be appropriate for analyzing the association for several reasons. For instance, if the outcome and covariate are measured at different time points or if there are missing values on both variables. Traditional GLMMs in fact require that the outcome and the time-varying biomarker are measured simultaneously. Furthermore, the covariate process must be specified and can not be ignored, otherwise, estimates of the parameters of interest are biased and inconsistent. Lastly, the functional form of the association is frequently unknown ahead of time and several options need to be explored. Multivariate models could be used to model the dependency structure between the endogenous and response variables. This type of model allows us to define the two processes separately, while also considering their relationship. We will examine at two multivariate models with different association structures. The joint mixed models (JMM) account for the association modelling the variance-covariance matrix of the random effects. The joint scaled models (JSM), instead, measure the association by introducing a scaling factor. However, because neither JMM nor JSM provide a known regression parameter, the interpretation of the estimated associations is not always clear; and, when the types of the two longitudinal variables are different, quantifying the association becomes especially difficult. Furthermore, their computational intensity, due to potentially high-dimensional integrations over the random effects terms, limits their applicability.

In this thesis, we will use a Bayesian method of estimation to try to overcome these problems: the Integrated Nested Laplace Approximation, well known as INLA. It is an approach that overcomes the issues coming from methods such as Markov Chain Monte Carlo (MCMC) and Laplace approximation, while still preserving the advantages. MCMC can be computationally intensive and can lead to different results depending on the starting points. The Laplace approximation, on the other hand, has limitations in handling complex priors and hierarchical models. The INLA approach, instead, is deterministic and more flexible.

Even though it has already been successfully and efficiently applied in biostatistics application, one goal of the thesis is to evaluate it under different conditions and assumptions. The dataset on which we will work has a very low population dimension and longitudinal variables that are not normally distributed. Indeed, we will present analysis results from a clinical study at Leiden University Medical Center (LUMC) to demonstrate the features of the proposed approach in the context of joint models. In particular, we will analyze data that comes from a study on children diagnosed with the rare disorder Duchenne Muscular Dystrophy (DMD). DMD is caused by mutations in the DMD gene and is characterized by delayed motor development, early loss of ambulation, progressive cardiac and respiratory failure, and premature death. All of these traits are summarized in functional scores, which are routinely recorded. In particular, we are interested in the *PUL 2.0* functional score, which records the upper limb motor performance as the disease progresses. The goal of the study is to detect the association between this outcome and some proteins, collected at different time points throughout the study period. Since the functional score has values limited from 0 to 42, joint models with gaussian and beta likelihoods will be implemented, which will lead to further complications that will be discussed throughout the thesis. Furthermore, the performance of INLA will be evaluated in different scenarios, to determine whether the results are reliable, accurate, and robust.

The thesis is then organized in the following manner. In chapter 1 we will go over the INLA estimation method in greater detail, presenting its hierarchical structure as well as the methods of evaluation that can be used in the context of Bayesian statistics. In chapter 2 a brief description of the Generalized Linear effects Models is provided, with a special focus on the Beta Mixed effects Model that will be used in the analysis of the data. We will present the mathematical form of the two joint models in chapter 3 , describing the process of derivation of the association quantities on both. Moreover, the INLA model formulation of them is described. Finally, some applications are shown in chapter 4 and chapter 5. First, a simulation study is reported to assess the performance of both INLA and the joint models. Secondly, we analyze the LUMC data.

Chapter 1

Integrated Nested Laplace Approximation (INLA)

Integrated Nested Laplace Approximation (INLA) is a deterministic Bayesian approach to statistical inference proposed by Rue, Martino, and Chopin 2009. As in every Bayesian problem, the main goal of this method is to provide an approximation to the posterior distribution, which often can not be defined in a closed-form. In particular, the focus is on the derivation of the univariate posterior marginal distributions and not on the joint posterior distribution of the parameters. Thus, it is not necessary to deal with multivariate distributions that are always more difficult to derive.

It is a deterministic method, which means it provides always the same results given a particular input, without the need to fix a seed in the software like in Markov Chain Monte Carlo (MCMC) or other methods used in Bayesian inference. In addition, it is faster and computationally less expensive than MCMC. Indeed, in order to speed the process some assumptions have been made by the authors: it should be possible to express the model as Latent Gaussian models with Gaussian Markov Random Field (GMRF) latent effects. So, not all the models could be fit in INLA.

In the following sections, everything will be described more in detail.

1.1 Latent Gaussian models and GMRF

The Integrated Nested Laplace Approximation works only for a specific class of model called Latent Gaussian Models. They are a subclass of the structured additive regression models, in which it is assumed that the response variable y_i distribution belongs to a distribution family (not necessarily the exponential family) of mean μ_i. In the general specification, the mean is linked through a function $g(\cdot)$ to a linear predictor η_i that has the following form:

$$\eta_i = \alpha + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_f} f^{(k)}\left(u_{ki}\right) + \varepsilon_i. \tag{1.1}$$

Here, α is the intercept, β_j's are the coefficients that quantify the linear effect of the covariates z_j, $j = 1, \ldots, n_\beta$, and $f^{(k)}$ are unknown functions, linear or non-linear transformations of the covariates u_k, $k = 1, \ldots, n_f$. Finally, the ε_i are the error terms, that can be absent depending on the likelihood. Based on this structure a lot of models can be defined, e.g., GLM, GLMM, GAM, time series, spatial models, mixed models, and many others.

As for every class of models, it is possible to specify the same structure in a Bayesian way just defining the prior distribution for each parameter. Based on this idea the specification of the Latent Gaussian Models is derived, in which, for each parameter of the structure additive predictor η_i, is assigned a Gaussian prior with zero mean and a specific precision matrix. Thus, the vector of the latent effects $x_i = (\eta_i, \alpha, \beta_j, f^{(k)})$ would be a multivariate normal distribution.

Furthermore, supplementary assumptions about the latent structure are made in the INLA framework. Considering a vector of n observations $\mathbf{y} = (y_1, \ldots, y_n)$, the latent effects vector could be written as follows

$$\mathbf{x} = (x_1, \ldots, x_n) = (\eta_1, \ldots, \eta_n, \alpha, \beta_1, \ldots, \beta_{n_\beta}, f^{(1)}, \ldots, f^{(n_f)}).$$

That vector x is assumed to be a Gaussian Markov Random Field (GMRF), means that it is a Gaussian random variable with Markov properties. Hence, some elements in the vector are conditionally independent; namely, given two components x_i and x_j, $i \neq j$, these are independent conditional on the remaining elements x_{-ij}. This property involves the precision matrix $\mathbf{Q}$ (inverse of the covariance matrix) that takes values equal to zero when conditional independence occurs. So, generally, $\mathbf{Q}$ is big $(10^2 - 10^5)$ and sparse. This allows for the use of a quicker and most efficient numerical method, based on the Cholesky-decomposition, to invert the precision matrix and obtain the covariance matrix Σ. In order to have a sparse matrix, one might also assume marginal independence between parameters, but it is a strong and generally unreasonable assumption. Conditional independence, instead, is a plausible assumption. For more information about GMRF and the properties used refer to Rue and Martino 2006.

To recap, the following assumptions are made in the INLA structure and they must be fulfilled to implement it:

- the model has to be expressed as a Latent Gaussian Model, this means that for every parameter is assumed a normal distribution with zero mean and specific precision matrix;

- the latent field has to be a Gaussian Markov Random Field, otherwise the conditional independence assumption fails and it is no longer possible to speed up the process of estimation due to the sparsity of the inverse matrix.

1.2 Model structure

As described above, the models in INLA are Bayesian hierarchical models, in particular, they are three-stages hierarchical models. In order to present in detail the structure is then necessary to define each stage, sorted as followed:

$$\begin{gathered}
Likelihood : \mathbf{y} \mid \ \mathbf{x}, \boldsymbol{\theta} \sim \boldsymbol{\pi}\left(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}\right), \\
Prior : \mathbf{x} \mid \boldsymbol{\theta} \sim \boldsymbol{\pi}\left(\mathbf{x} \mid \boldsymbol{\theta}\right), \\
Hyperprior : \boldsymbol{\theta} \sim \boldsymbol{\pi}(\boldsymbol{\theta}).
\end{gathered}$$

The vectors $\mathbf{x}$ and $\mathbf{y}$ have already been defined. Notes that from now on, the vector $\boldsymbol{\theta}$ will be used as a simplified notation to indicate the vector of all the hyperparameters. The latent effects distribution depends on a vector of some hyperparameters, said $\boldsymbol{\theta_1}$, which in turn depends on another vector of hyperparameters, said $\boldsymbol{\theta_2}$. Attaching these two results in the combined vector $\boldsymbol{\theta}$.

Likelihood

The first stage is the one regarding the vector of observations $\mathbf{y} = (y_1, \ldots, y_n)$. At this level it is necessary to define the likelihood, so the distribution of the observations given the latent parameters.

Assuming that the components are independent conditionally on the latent effects $\mathbf{x}$ and the hyperparameters $\boldsymbol{\theta}$ (See Section 1.1), the likelihood can be written as:

$$\boldsymbol{\pi}(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}) = \prod_{i \in I} \boldsymbol{\pi}\left(y_i \mid x_i, \boldsymbol{\theta}\right). \tag{1.2}$$

Here, the set I is a subset of $N = 1, \ldots, n$. It only includes the indexes of the observations that are detected, the indexes of the missing values of the response variable are not incorporated (see Section 1.5).

Prior

The second stage consists of delineating the prior distribution of the latent field $\mathbf{x}$. Based on the assumptions describe in Section1.1, the prior distribution is then a multivariate normal distribution

$$\boldsymbol{\pi}(\mathbf{x} \mid \boldsymbol{\theta}) = (2\pi)^{-n/2} |\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{Q}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu})\right\}, \tag{1.3}$$

where $\boldsymbol{\mu}$ is a zero vector and $\mathbf{Q}(\boldsymbol{\theta})$ is the sparse precision matrix with its own distribution that depends on some hyperparameters defines in the vector $\boldsymbol{\theta}$.

Hyperprior

Lastly, the third stage is the one regarding the unknown hyperparameters. Likely, most of the hyperparameters are independent, such as that the hyperprior $\boldsymbol{\pi}(\boldsymbol{\theta})$ could be defined as the product of the univariate hyperpriors.

The hyperprior selection is based on previous knowledge, as the name suggests. For instance, for the variance parameters, distributions with positive support are chosen, like *Beta* or *Gamma* distributions. Otherwise, when no information is available, non-informative priors are preferred. The possibility are many and in R-INLA several distributions are available but they will be described in more detail in Section 2.1.1. The aim of this work is not to present the Bayesian statistics theory, so to reach more information on how to select the best prior for a specific problem is recommended to refer to Nicenboim, Schad, and Vasishth 2022.

Specified the likelihood and the hyperpriors, the joint posterior distribution can be defined as:

$$\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) = \frac{\pi(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}, \boldsymbol{\theta})}{\pi(\mathbf{y})} \propto \pi(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}, \boldsymbol{\theta}). \tag{1.4}$$

The latter, considering equation (1.3) and the factorization of the joint posterior $\pi(\mathbf{x}, \boldsymbol{\theta}) = \pi(\mathbf{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})$, can be rewritten as follow:

$$\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) \propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{1/2} \exp\left\{-\frac{1}{2}\mathbf{x}^{\top}\mathbf{Q}(\boldsymbol{\theta})\mathbf{x}\right\} \prod_{i \in I} \pi\left(y_i \mid x_i, \boldsymbol{\theta}\right). \tag{1.5}$$

But the main goals of the INLA method is to estimate the univariate posterior marginal distributions of the latent field and the hyperparameters, respectively $\boldsymbol{\pi}\left(x_i \mid \mathbf{y}\right)$ and $\boldsymbol{\pi}\left(\theta_j \mid \mathbf{y}\right)$, defined as:

$$\boldsymbol{\pi}\left(x_i \mid \mathbf{y}\right) = \int \boldsymbol{\pi}\left(x_i \mid \boldsymbol{\theta}, \mathbf{y}\right) \boldsymbol{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) \mathrm{d}\boldsymbol{\theta}, \tag{1.6}$$

$$\boldsymbol{\pi}\left(\theta_j \mid \mathbf{y}\right) = \int \boldsymbol{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) \mathrm{d}\boldsymbol{\theta}_{-j}, \tag{1.7}$$

where $\boldsymbol{\theta}_{-j}$ is the vector of hyperparameters without the j-th element. In the next sections, it is described how INLA reaches this goal and which assumptions are made in order to do so.

1.3 INLA approximations

To compute the univariate marginal distributions (1.6) and (1.7) it is necessary to approximate the joint distribution of $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\pi}}(\boldsymbol{\theta} \mid \mathbf{y})$, and the posterior marginal $\tilde{\boldsymbol{\pi}}\left(x_i \mid \boldsymbol{\theta}, \mathbf{y}\right)$. The former is defined as

$$\tilde{\boldsymbol{\pi}}(\boldsymbol{\theta} \mid \mathbf{y}) \propto \left. \frac{\boldsymbol{\pi}(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\boldsymbol{\pi}}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}, \tag{1.8}$$

where $\tilde{\boldsymbol{\pi}}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation (see Section 1.3.2) to the full conditional of the latent field $\mathbf{x}$ and $\mathbf{x}^*(\boldsymbol{\theta})$ is the corresponding mode, give $\boldsymbol{\theta}$. For the latter, instead, Rue, Martino, and Chopin 2009 have proposed three types of approximations: Gaussian, Laplace and simplified Laplace approximation.

The INLA approach can then be subdivided into three stages. In the first one, the posterior marginal of $\boldsymbol{\theta}$ is estimated, in the second $\tilde{\boldsymbol{\pi}}\left(x_i \mid \boldsymbol{\theta}, \mathbf{y}\right)$ is defined and in the third one the previous two results are combined using numerical integration to obtain (1.6) and (1.7).

1.3.1 Approximation of $\tilde{\boldsymbol{\pi}}(\boldsymbol{\theta} \mid \mathbf{y})$

First of all, the mode is located optimizing the logarithm of $\tilde{\boldsymbol{\pi}}(\boldsymbol{\theta} \mid \mathbf{y})$ by respect to $\boldsymbol{\theta}$ with quasi-Newton's method. At the modal configuration, defined from now on as $\boldsymbol{\theta}^*$, it is then possible to compute the Hessian matrix

$\boldsymbol{H} = \boldsymbol{\Sigma}^{-1}$ using finite differences. Once this is done, to make the density more regular and to simplify the numerical integration, the vector $\boldsymbol{\theta}$ is reparameterized as follows:

$$\boldsymbol{\theta}(\mathbf{z}) = \boldsymbol{\theta}^* + \boldsymbol{\Sigma}^{1/2}\mathbf{z} = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{z},$$

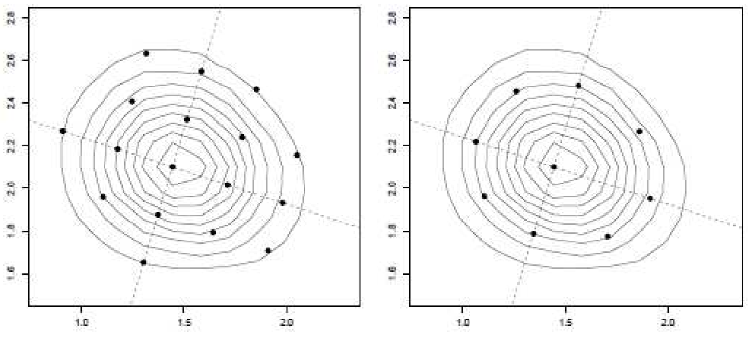
where $\mathbf{z}$ is the standardized vector of $\boldsymbol{\theta}$ and the matrix $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$ is defined via eigenvalue decomposition. In such a way the space of $\boldsymbol{\theta}$ is corrected for scale and rotation, thus it is easier to explore it.

From here it is possible to proceed with the approximation of $\log(\tilde{\boldsymbol{\pi}}(\boldsymbol{\theta} \mid \mathbf{y}))$. Based on the dimension of the hyperparameters vector there are two ways of proceeding: the grid strategy and the Central Composite Design (CCD) strategy. The first one explores the probability distribution moving in the $\mathbf{z}$ directions. It starts shifting from the mode, where $\mathbf{z} = 0$, to the positive direction of the first element $\mathbf{z}_1$ with steps of length δ_z until the distance between the two points is greater than a threshold δ. So, mathematically written, as long as

$$|\log(\tilde{\pi}(\boldsymbol{\theta}(\mathbf{0}) \mid \mathbf{y})) - \log(\tilde{\pi}(\boldsymbol{\theta}(\mathbf{z}) \mid \mathbf{y}))| < \delta.$$

The same process is repeated for each direction of $\mathbf{z}$ and, ultimately, also the combinations between all the points are considered. In such a manner, it is possible to detect the bulk of the probability mass.

The second strategy, instead, selects the points differently. The main idea of this approach is to consider the integration problem as a design problem, in particular as a two-level factorial experiment. From this, other points are included, called center and axial/star points. The former is the origin of the $\mathbf{z}$ axis, which in this case is the mode $\boldsymbol{\theta}^*$, the latter consists of two points for each direction at a distance $\pm\alpha$ from the central point. This approach requires less computational power than the other and leads to the same results. For this reason, Rue, Martino, and Chopin 2009 suggest using it, especially with a large number of hyperparameters. An example of both the strategy is shown in Figure1.1.



(a) (b)

Figure 1.1: Exemplification of the grid (a) and the CCD (b) strategy in a two-dimensional hyperparameters vector. Source: Rue, Thiago, et al. 2013

Once the marginal distribution $\tilde{\boldsymbol{\pi}}(\boldsymbol{\theta} \mid \mathbf{y})$ is approximated, it is possible to derive the univariate marginal distribution $\boldsymbol{\pi}\left(\theta_j \mid \mathbf{y}\right)$ by integrating $\boldsymbol{\theta_{-j}}$ out using numerical integration.

1.3.2 Approximation of $\tilde{\boldsymbol{\pi}}\left(x_i \mid \boldsymbol{\theta}, \mathbf{y}\right)$

As already mentioned above, for the marginals distributions of the latent effects three different approximations are available. The differences between them regard the computational expenses and the accuracy. The simplified Laplace approximation is the fastest computationally but it is less accurate than the Laplace approximation. This one, in turn, is more computationally expensive than the Gaussian approximation, but it is better in terms of accuracy.

Below they are described in more detail.

Gaussian approximation

The Gaussian approximation is generally called Laplace approximation in Bayesian statistics but it should not be confused with the one described below. It is based on the second-order Taylor series expansion around the mode of the logarithm of the distribution. Given a general distribution $g(x)$ and its mode $\hat{x}$, a possible approximation is then

$$\log g(x) \approx \log g(\hat{x}) + \frac{\partial \log g(\hat{x})}{\partial x}(x-\hat{x}) + \frac{1}{2}\frac{\partial^2 \log g(\hat{x})}{\partial x^2}(x-\hat{x})^2.$$

Since $\hat{x}$ is the mode, the first-order derivative term is equal to zero. Taking the exponent and the integral, the formula can be rewritten as

$$\int g(x)dx \approx \exp[\log g(\hat{x})] \cdot \int \exp\left[-\frac{1}{2}\frac{\partial^2 \log g(\hat{x})}{\partial x^2}(x-\hat{x})^2\right] dx,$$

where $\exp[\log g(\hat{x})]$ is a constant. Thus, the distribution $g(x)$ can be approximated with a Gaussian distribution with mean $\hat{x}$ and variance $\hat{\sigma}^2 = \left(\frac{\partial^2 \log g(\hat{x})}{\partial x^2}\right)^{-1}$.

In INLA inference, the distribution of interest is the full conditional of the latent fields. The idea is to start from the Gaussian distribution of $\tilde{\boldsymbol{\pi}}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and then to derive the univariate distributions $\tilde{\boldsymbol{\pi}}\,(x_i \mid \boldsymbol{\theta}, \mathbf{y})$ by marginalizing. In order to do that it is necessary to specify the marginal means $\boldsymbol{\mu_i}$ and variances $\boldsymbol{\sigma_i}$. The only extra cost derives from the fact that the variances need to be determined from the sparse precision matrix, but this is easily done by taking advantage of the GMRF properties. See Rue, Martino, and Chopin 2009 and Rue and Martino 2006 for the technicality.

Laplace approximation

The Laplace approximation has the same structure of the joint distribution of $\boldsymbol{\theta}$ (1.8). It is denoted as follows:

$$\pi_{LA}\left(x_i \mid \boldsymbol{\theta}, \mathbf{y}\right) \propto \left.\frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{GG}\left(\mathbf{x}_{-i} \mid x_i, \boldsymbol{\theta}, \mathbf{y}\right)}\right|_{\mathbf{x}_i = \mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta})}, \tag{1.9}$$

where $\tilde{\pi}_{GG}\left(\mathbf{x}_{-i} \mid x_i, \boldsymbol{\theta}, \mathbf{y}\right)$ is the Gaussian approximation to the vector $\mathbf{x}$ without the i-th element and $\mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta})$ is the corresponding mode. This approximation takes longer because needs to be computed for each value of x_i.

Simplified Laplace approximation

Lastly, the Simplified Laplace approximation is proposed to both correct the Gaussian approximation and to speed the process. It is defined as the product between a Gaussian approximation and a cubic splines:

$$\pi_{LA}\left(x_i \mid \boldsymbol{\theta}, \mathbf{y}\right) \propto \mathcal{N}\left(x_i \mid \mu_i(\boldsymbol{\theta}), \sigma_i^2(\boldsymbol{\theta})\right) \exp\left(\text{spline}\left(x_i\right)\right). \tag{1.10}$$

This is a compromise between the previous two approximations in terms of accuracy and speed time.

1.4 Methods of evaluation

In INLA there are a few ways used to evaluate the model assessment and to compare models and select the best one. In the frequentist framework, depending on whether the models are nested or not, different methods of evaluation are used. The method describe below, instead, can be utilized for every occasion.

1.4.1 Bayes Factor

The Bayes Factor compares the predictive performance of two models, i.e., $\mathcal{M}_1$ and $\mathcal{M}_2$, and it evaluates which model is more likely to have generated the data. It is defined as the ratio of Marginal Likelihood of the models fitted:

$$BF = \frac{\pi\left(\mathbf{y} \mid \mathcal{M}_1\right)}{\pi\left(\mathbf{y} \mid \mathcal{M}_2\right)} = \frac{\pi\left(\mathcal{M}_1 \mid \mathbf{y}\right)\pi\left(\mathcal{M}_2\right)}{\pi\left(\mathcal{M}_2 \mid \mathbf{y}\right)\pi\left(\mathcal{M}_1\right)}. \tag{1.11}$$

Here, $\pi\left(\mathbf{y} \mid \mathcal{M}_1\right)$ is the probability of the observed data under $\mathcal{M}_1$, and $\pi\left(\mathcal{M}_1\right)$ and $\pi\left(\mathcal{M}_1 \mid \mathbf{y}\right)$ are the corresponding prior and posterior distributions. The Marginal Likelihood is determined in INLA through a Gaussian approximation:

$$\tilde{\pi}(\mathbf{y}) = \int \left.\frac{\pi(\boldsymbol{\theta}, \mathbf{x}, \mathbf{y})}{\tilde{\pi}_{\mathrm{G}}(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})}\right|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} d\boldsymbol{\theta}, \tag{1.12}$$

where $\pi(\boldsymbol{\theta}, \mathbf{x}, \mathbf{y}) = \pi(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$. In the case of equal priors in the two models, the Bayes Factor is simply equal to the ratio of the posterior distribution. If it is greater than one, the model in the numerator is better. On the contrary, if it is smaller than one, there is evidence in favor of the denominator. A Bayes Factor close to one, instead, means that there is no significant difference between the predictive performance of the two models.

Generally, the logarithm of the Bayes Factor is considered:

$$\log(BF) = \log[\pi\left(\mathbf{y} \mid \mathcal{M}_1\right)] - \log[\pi\left(\mathbf{y} \mid \mathcal{M}_2\right)]. \tag{1.13}$$

This implies that when (1.13) is equal to zero, there is no evidence against or in favor of the models. When it is positive, $\mathcal{M}_1$ is the best one, when is negative, the opposite. However, a more specific interpretation has been proposed for both scales. See Table 1.1 and for more details refer to Nicenboim, Schad, and Vasishth 2022.

Table 1.1: Interpretation of the Bayes Factor

BF	log(BF)	Interpretation
¡ 0.10	¡ -2.30	Strong evidence for $\mathcal{M}_2$
$[0.10, 0.33)$	$[-2.30, -1.10)$	Moderate evidence for $\mathcal{M}_2$
$[0.33, 1)$	$[-1.10, 0)$	Weak evidence for $\mathcal{M}_2$
1	0	No evidence
$(1, 3]$	$(0, 1.10]$	Weak evidence for $\mathcal{M}_1$
$(3, 10]$	$(1.10, 2.30]$	Moderate evidence for $\mathcal{M}_1$
¿ 10	¿ 2.30	Strong evidence for $\mathcal{M}_1$

1.4.2 Information-based criteria

Similar to AIC and BIC in the frequentist framework, in Bayesian statistics the Deviance Information Criteria (DIC) and the Watanabe-Akaike Information Criteria (WAIC) are often utilized.

The idea of these criteria is to penalize the deviance in order to compare different models taking into consideration their complexity. The general formula is then

$$IC = D(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) + 2p_D.$$

Here $D(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}})$ is the deviance evaluated on the posterior mean of the latent effects and the posterior mode of the hyperparameters. For the second one the mode is considered because of the probable high skewness of the posterior marginals. The effective number of parameter $2p_D$ is different for each Information Criteria. For instance, for the DIC criterion, they are defined as the mean of the deviance minus the deviance of the mean. In particular, they are approximated as

$$p_D(\theta) \approx n - \operatorname{tr}\left\{Q(\theta)Q(\theta)^{-1}\right\},$$

where n is the number of observations and $Q(\theta)$ is the precision matrix of the latent field (defined in Section 1.2). These methods are used for the model choice, in which the model with the lowest value is the best one.

For more information see Spiegelhalter et al. 2002.

1.4.3 Predictive measures

The predictive measures are *leave-one-out* cross-validatory criteria used to both evaluate the good-of-fit of the models and select the best ones. Moreover, since they depend on single observations, they are useful in detecting outliers or the level of importance of the observations. However, it could still be dangerous to make final statements if one observation masks another.

In INLA two predictive measures are available: Conditional Predictive Ordinates (CPO) and Predictive Integral Transform (PIT). Both are evaluated for each observation.

The Conditional Predictive Ordinates are the posterior probabilities of observing the i-th measurements when the model is fitting on all the data except y_i. Thus, they are defined as

$$CPO_i = \pi\left(y_i \mid y_{-i}\right),$$

in which high values indicate the presence of outliers. The CPOs can be summarized across the data in the following way:

$$CPO = -\sum_{i=1}^{n} \log\left(CPO_i\right).$$

A smaller value of CPO means a better fitting of the model.

The Predictive Integral Transforms, instead, measure the probability for a new observation to be lower than the actually observed one when the model is fitted using all data except y_i. Hence, for continuous observations, it is defined as

$$PIT_i = \pi\left(y_i^{\text{new}} \leq y_i \mid y_{-i}\right),$$

and, for discrete observations, it is adjusted in the following way:

$$PIT_i^{\text{adjusted}} = PIT_i - 0.5 * CPO_i,$$

in which 0.5 is the probability to have y_i^{new} equal to y_i. In this case, unusually large or small values indicate possible outliers. Furthermore, there is not summary value. The only way to evaluate the model assessment over all the observations is by studying the PITs distribution. This one is expected to follow a Standard Uniform distribution.

To see how both are computed in INLA refers to Held, Schrödle, and Rue 2010, in which a comparison with MCMC is provided too.

Leave-one-out cross-validation is designed for models in which there is an assumption of independence between the observations. It generally leads to too optimistic results when this assumption does not apply. Thus, in this thesis in which we will study longitudinal data, CPO and PIT implemented as described above, are not really reliable since the information from the same subject is used for prediction. It would be better to consider a *subject-wise* cross-validation (Rue and Liu 2022) to estimate the posterior probabilities of observing the i-th subject when the model is fitting on all the data except the subject i (1.14). In this situation, all the measurements from a specific subject are used as a set to test the model trained on all the other observations. However, in R-INLA this kind of predictive measure is not available for everyone yet and it is still in progress. The computation itself is an approximation to nested integrals:

$$\pi\left(\mathbf{y}_i \mid \mathbf{y}_{-I_i}\right) = \int_{\boldsymbol{\theta}} \pi\left(\mathbf{y}_i \mid \boldsymbol{\theta}, \mathbf{y}_{-I_i}\right) \pi\left(\boldsymbol{\theta} \mid \mathbf{y}_{-I_i}\right) d\boldsymbol{\theta}, \tag{1.14}$$

$$\pi\left(\mathbf{y}_i \mid \boldsymbol{\theta}, \mathbf{y}_{-I_i}\right) = \int \pi\left(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\theta}\right) \pi\left(\mathbf{x}_i \mid \boldsymbol{\theta}, \mathbf{y}_{-I_i}\right) d\mathbf{x}_i. \tag{1.15}$$

The process of approximation of these distributions is similar to the one described in Section 1.3. The posterior probability (1.14) has to be computed by numerical integration and the integral (1.15) needs to be approximated by Gauss-Hermite quadratures. Hence, the calculation is not trivial and it will be not developed in this thesis since it is not the major focus of this work.

1.5 Predictive distribution

As we already mentioned in Section 1.2, in INLA the likelihood is based only on the observations that are detected. It is possible to estimate the missing values by using the Bayesian predictive distribution:

$$\pi\left(y_m \mid \mathbf{y}_{obs}\right) = \int \pi\left(y_m, \boldsymbol{\theta} \mid \mathbf{y}_{obs}\right) d\boldsymbol{\theta} = \int \pi\left(y_m \mid \mathbf{y}_{obs}, \boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta} \mid \mathbf{y}_{obs}\right) d\boldsymbol{\theta}. \tag{1.16}$$

The predictive distribution of the missing value y_m is then derived given all the observed values of the response variable. As with any other distribution, from this one as well we can simply get summary statistics like means, standard deviations and quantiles.

Chapter 2

Mixed Effects Models

Mixed models are regression techniques used with longitudinal, hierarchical and cluster data, i.e., medical data, social data, space data, and others. Such, the outcome of interest is repeatedly measured on the same individual or cluster, so the assumption of independence between the observations, which is usually made on models for cross-sectional data, does not apply anymore. In the mixed models, new unknown parameters, called random effects, are introduced to take into consideration the correlation between the repeated measurements. Thus, the unobservable features common to all observations related to the same unit are described as a realization of a random variable. Thanks to this, it is possible to make both predictions for a specific individual/cluster and marginal predictions, as well. Everything will be described more in detail in the following section.

There are different types of Mixed Models (see Wu 2019), but only the Linear Mixed Model (LMM) and Generalized Linear Mixed Model (GLMM) will be described below. Both are extensions of the corresponding models for cross-sectional data, Linear regression Models (LM) and Generalized Linear Models (GLM).

2.1 Linear Mixed Effects model

A general Linear regression Model for a subject i, where $i = 1, \dots, N$, can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{2.1}$$

where y_i is the response variable for the i-th subject and β_j's, $j = 1, \cdots, p$, are the regression coefficients that quantify the effect of the p independent variables x_j and are considered fixed over all the individuals. The ε_i's are the error terms assumed to be independent and to be normally distributed, with zero mean and variance σ^2. The model may be also be written in the following matrix-form:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbf{y} \sim \mathcal{N}(X\beta, \sigma^2 I_N).$$

For longitudinal or hierarchical data, these types of models are not appropriate. Indeed, the observations are no longer independent because they belong to the same e.g., subject or cluster. To take into consideration the correlation between them, new parameters are added to the Linear regression form (2.1). Random effects are introduced, assuming that each individual has his own profile which deviates from the population mean profile $X^\top \beta$:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad \mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D}), \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i). \tag{2.2}$$

Here, $\mathbf{y}_i = (y_{i1}, \dots, y_{n_i})$ is the vector of the n_i repeated observations of the i-th subject, $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})$ is the vector of the q random effects and Z_i is the corresponding design matrix. The random effects are assumed normally distributed, with a variance-covariance matrix $\mathbf{D}$ that gives information about the between-subjects variance. Thus, it quantifies the difference among the subject's trajectories. The bigger the variances (diagonal values of the matrix) are, the more considerable the difference between the individuals/clusters is. We can assume, for example, a model with random intercepts and random slopes. The variance of the former indicates how much difference there is between the subjects' response levels around the population mean.

The variance of the random slopes, on the other hand, provides information on the progression of the subjects' response over time. The variance increases if the progressions are significantly different, implying that the trajectories are not parallel and thus the slopes are disparate.

The vector of the error terms $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})$, instead, gives information about the within-subject variation. The bigger $\boldsymbol{\Sigma}_i = \sigma^2 I_{n_i}$ is, the higher the disparity among the repeated measurements of the subject i is. Random effects and error terms are assumed independent. In the above formulation, the Conditional Independence assumption is made, which means that the random effects capture all the correlations. So $\boldsymbol{\Sigma}_i$ is a scalar matrix with scale factor σ^2 positive. This suggests that the observations from the same subject are conditionally independent, given the random effects.

With LMM it is possible to provide both marginal and subject-specific information. The fixed-effect parameters $\boldsymbol{\beta}$ are assumed the same for all the individuals and measure the mean change of the response variable for a unit increase of the covariate, given that all the other independent variables are kept fixed. Thus, they have a population-averaged interpretation and describe the overall mean of the units' profile. The random-effect parameters b_i have instead a subject-specific interpretation. They measure the deviation of the profile of each individual/cluster i from the population mean.

2.1.1 INLA model formulation

Estimation of the LMM can be done using Maximum Likelihood approach (e.g., using the established R packages nlme or lme4) or Bayesian methods (e.g., using the R package MCMCglmm). In this thesis, we consider an alternative estimation approach, the INLA method (using the R package INLA), presented in chapter 1. Estimation with INLA requires that the model can be written as a Latent Gaussian Model. To show that this is the case for the GLMM in general, we consider a simple LMM with a single time fixed covariate x and two correlated random effects i.e., random intercept and random slope terms. The form of the model is then:

$$y_{ij} = \beta_0 + b_{i0} + \beta_1 \cdot x_{ij} + (\beta_t + b_{it}) \cdot t_{ij} + \varepsilon_{ij}, \tag{2.3}$$

where

$$\begin{bmatrix} b_{i0} \\ b_{it} \end{bmatrix} \sim \mathcal{N}_2 \left(0, \begin{bmatrix} \sigma_0^2 & \sigma_{(0,t)} \\ \sigma_{(t,0)} & \sigma_1^2 \end{bmatrix} \right), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2).$$

Thus, random intercepts b_{i0} and random slopes b_{i1} are considered and they followed a bivariate normal distribution. Up to here, the distributions for the random effects and the error terms are defined using a parametrization based on the variances. However, in INLA the internal representation of all the parameters of the model is based on the precisions.

In the Bayesian framework, the latent field and the hyperparameters, which in this case are the precisions, are also assigned a distribution. By default, a non-informative prior is assigned to all the parameters. However, the type of the priors and the respective parameters can be changed using a specific option in R-INLA. Based on the INLA structure, described in Section 1.2, only one distribution cannot be changed: for the regression coefficients, a Gaussian distribution is assumed (see equation 1.3). The intercept β_0 follows a Normal distribution with zero mean and zero precision. The x's coefficient β_1 and the t's parameter β_t are normally distributed with zero mean and precision $1/\sigma_1^2 = 0.001$. With more than two covariates the prior distribution for each coefficient is the same.

Let's now focus on the random effects. The hyperprior for the precision matrix $\boldsymbol{W}$ is a 2-dimensional Wishart with r degree of freedom and scale matrix $\mathbf{R}^{-1}$, defined as $\mathbf{R} = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$ and with $R_{12} = R_{21}$. By default, $r = 4$ and $\mathbf{R}$ is a diagonal matrix. In Bayesian statistics, the Wishart distribution is the conjugate prior of the precision matrix of a multivariate normal vector. It is a generalization to multiple dimensions of the gamma distribution. Given N subjects, the total number of random effects is $2N$, because, for each individual, one random intercept and one random slope are defined. In R-INLA they are internally represented into a single vector $(b_{0,1}, \ldots, b_{0,N}, b_{t,1}, \ldots, u_{t,N})$ of length $2N$.

Lastly, we define the vector of independent and identically distributed error terms $\boldsymbol{\varepsilon}_i$ with zero mean and general precision matrix $\tau\boldsymbol{\Sigma}$. The hyperparameter $\tau = 1/\sigma^2$ has a Gamma distribution, but in INLA the internal parameterization $\theta = \log(\tau)$ is used. A variable u is log-Gamma distributed if $u = \log(x)$ and x is Gamma(a, b) - distributed. Thus, θ is assigned by default a log-Gamma distribution with shape $a = 1$ and rate $b = 0.00005$. The matrix $\boldsymbol{\Sigma}$ is a diagonal matrix of scaling factors $\boldsymbol{s} = (s_1, \cdots, s_{n_i})$, which are all equal to one by default (Conditional Independence assumption). It is also possible to assume a different vector of scaled factors $\boldsymbol{s}$, for example, to consider a different precision for different groups of observations.

To recap, the default hierarchical structure (see Section 1.2) of the Linear Mixed Model (2.3) with random intercepts and random slopes, is as follows:

$$\mathbf{y}_i \mid \mathbf{b}_i \sim \mathcal{N}_{n_i}(X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \boldsymbol{\Sigma}_i),$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_t \end{bmatrix} \sim \mathcal{N}_3\left(0, \begin{bmatrix} 0 & & \\ & 0.001 & \\ & & 0.001 \end{bmatrix}^{-1}\right), \quad \begin{bmatrix} b_{i0} \\ b_{it} \end{bmatrix} \sim \mathcal{N}_2\left(0, \boldsymbol{W}^{-1}\right), \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}\left(0, \tau^{-1}I_{n_i}\right),$$

$$\boldsymbol{W} \sim Wishart_2\left(4, I_2\right), \quad \theta = \log(\tau) \sim LogGamma\left(1, 0.00005\right).$$

For more information about the structure of the models in INLA see Gómez-Rubio 2020. Code A.1 details coding of the Linear Mixed Model (2.3) in R-INLA.

2.2 Generalized Linear Mixed Model

In Generalized Linear Models the response variables can come from different distributions of the exponential family besides the Gaussian. So what is modeled here it is no longer the outcome but its expectation. In addition, they allow non-linear relationships between the expectation and the independent variables.

Generally, these types of models are defined by

$$g(\mu) = \eta, \quad \eta = X\beta = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$$

where the linear predictor η is linked to the mean $\mu = E\left(y\right)$ through the linear or non-linear function $g(\cdot)$. Depending on the selected likelihood, there are several options for the link functions $g(\cdot)$. For instance, with binary data, the response variable is assumed to follow a Bernoulli distribution and the link functions used more often are the *logit* or the *probit*.

The form described above is for cross-sectional data but can be easily extended to longitudinal or cluster data. The idea is the same as with the Linear Mixed Models. Introducing random effects we can incorporate the correlation between repeated measurements within each individual or cluster. Defining as $\mathbf{y}_i = (y_{i1}, \ldots, y_{n_i})$ the vector of observations of the subject i, the Generalized Linear Mixed Model has the following form:

$$g\left(\mu_i\right) = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \quad \mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D}), \tag{2.4}$$

where $\mu_i = E\left(\mathbf{y}_i \mid \boldsymbol{\beta}, \mathbf{b}_i\right)$ is the mean conditional on the random effects. Therefore, it is possible to consider the Linear Mixed Model as a special case of GLMMs with a normal likelihood and as a link function the *identity* function. Similar to the LMM, each subject has his own profile in time. As we can see from (2.4), the conditional model, i.e., the distribution of the outcome given the random effects $f\left(\mathbf{y}_i \mid \boldsymbol{\beta}, \mathbf{b}_i\right)$, is the distribution of the data with predictive mean μ_i (for instance, a Beta distribution 2.8). However, the marginal model, obtained by integrating out the random effects, is not as simple as in the LMM to derive. The marginal likelihood is defined as follows:

$$\begin{aligned} L(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{D}) &= \prod_{i=1}^{N} f\left(\mathbf{y}_i; \boldsymbol{\beta}, \mathbf{D}\right) \\ &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f\left(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\beta}\right) f\left(\mathbf{b}_i \mid \mathbf{D}\right) d\mathbf{b}_i. \end{aligned} \tag{2.5}$$

The data distribution $f\left(y_{ij} \mid \mathbf{b}_i, \boldsymbol{\beta}\right)$ is not conjugate with the normal distribution of the random effects anymore, so the integrals do not have a closed-form solution. In order to have the marginal model then, it is necessary to approximate the integrals using numerical integration techniques. But, the complexity of the integral and the computational cost increase as the number of random effects increases. And, base on the method used, we can obtain different estimations of the parameters.

Furthermore, the use of a non-linear link function $g(\cdot)$ implies some complications in the interpretation of the parameters too. The covariates are related to the mean response non-linearly, so the mean of the average subject is different from the average of the means of all the subjects. The fixed-effect parameters now describe the mean of the average individual, the one with random effects equal to zero. Thus, they do not have a marginal interpretation anymore. Their interpretation is more complicated and cannot be simply generalized. It depends on the chosen likelihood and link function. Instead, the random effects have the same role as in the LMM. Each random effect represents the influence of each subject/cluster on the repeated measurements, information not captured by the fixed effect. Given the random effect, the observations are assumed independent. Hence, in the GLMM we typically assume the Conditional Independence assumption as well.

2.2.1 Beta Mixed Model

Motivated by the data example which is briefly introduced in the Introduction and presented in detail inchapter 5, a special case of a GLMM will be used: the Beta Mixed Model. Beta regression models are a suitable choice for continuous response variables defined on a specific interval. Indeed, in such situations, the Gaussian likelihood with support in $\mathbb{R}$ is not appropriate. The beta distribution is defined in a unit range $(0, 1)$. However, it is also used in the presence of a dependent variable y limited in different intervals (a, b). In order to do that, the support of the y variable has to be changed. There are several possible transformations but the most widely used is the

Min-Max transformation $h(y) = \frac{y-min}{max-min}$.

The beta distribution with shape parameters α and β, has density

$$f(y \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad 0 < y < 1, \quad \alpha, \beta > 0, \tag{2.6}$$

where the normalization constant $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta-function and $\Gamma(x)$ is the Gamma-function. But, in Beta regression models, another parameterization is used:

$$f(y \mid \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-y)\phi-1}, \quad 0 < \mu < 1. \tag{2.7}$$

Here, $\mu = E\left(y\right) = \frac{\alpha}{\alpha+\beta}$ is the mean and $\phi = \alpha + \beta$ is the precision parameter of (2.6) such that $\phi > 0$ is positive. From this it follows that the variance of the Beta distribution (2.7) is $\text{Var}(y) = \frac{\mu(1-\mu)}{1+\phi}$. It is possible to derive one parameterization from the other by simply computing $\alpha = \mu\phi$ and $\beta = -\mu\phi + \phi$.

Following the GLMM form (2.4), the Beta Mixed Model is then specified by

$$\mathbf{y}_i \mid \mathbf{b}_i \sim Beta(\mu_i, \phi), \quad g\left(\mu_i\right) = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \quad \mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D}), \tag{2.8}$$

where ϕ is assuming constant and the link function $g(\cdot)$ can be chosen as either *logit, cauchit, probit, complementary loglog* and *log-log.*

For more information see Bonat, Ribeiro, and Zeviani 2015.

As mentioned in Section 2.2, parameters in the Beta Mixed Models have an interpretation in terms of $\mu_i = E\left(\mathbf{y}_i \mid \boldsymbol{\beta}, \mathbf{b}_i\right)$, so conditional on the random effects, which is not always desirable. Having parameters with population-averaged interpretation would be preferable. A way to proceed is by applying the marginalization idea of Hedeker et al. 2017. Let's consider the marginal mean μ_i^M for the i-th subject and the vector of the regression parameters with marginal interpretation $\boldsymbol{\beta}^M$. The marginal model of (2.8) is defined as $g\left(\mu_i^M\right) = X_i\boldsymbol{\beta}^M$, from which it follows the matrix form $g\left(\boldsymbol{\mu}^M\right) = X\boldsymbol{\beta}^M$, where $g\left(\boldsymbol{\mu}^M\right)$ is a vector of length equal to the number of observations N_{obs}. Multiplying both sides of this last equation by $\left(X^\top X\right)^{-1} X^\top$:

$$\left(X^\top X\right)^{-1} X^\top g\left(\boldsymbol{\mu}^M\right) = \left(X^\top X\right)^{-1} X^\top X^\top \boldsymbol{\beta}^M,$$

and solving for $\boldsymbol{\beta}^M$, the population-averaged regression coefficients can be express as:

$$\boldsymbol{\beta}^M = \left(X^\top X\right)^{-1} X^\top g\left(\boldsymbol{\mu}^M\right). \tag{2.9}$$

Following this logic, we only need to compute $\boldsymbol{\mu}^M$. The marginal mean is expressed as the integral over the random effects of the inverse link-function of the linear predictor:

$$\mu_{ij}^M = \int_b g^{-1}\left(x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \mathbf{b}_i\right) f(\mathbf{b}_i) d\mathbf{b}_i, \tag{2.10}$$

where $\boldsymbol{\beta}$ are the subject-specific regression coefficients. The integration can be approximated by a summation derived by the Gauss-Hermite quadrature. From here, it is also possible to obtain an approximation of the marginalized standard errors, simply using the Delta method.

2.2.2 INLA model formulation

Also in GLMM, parameter estimation can be performed using the Maximum Likelihood approach, implemented through the R packages lme4, MASS or GLMMadaptive. However, in this thesis we will consider the Bayesian estimation approach INLA, which helps us overcome problems related to the marginal integration. Each GLMM can be expressed as a Latent Gaussian Model, so as the Beta Mixed Model. The priors of the random and fixed effects are the same described in Section 2.1.1. However, in this case, the linear predictor η is associated with the mean μ using a link function. By default, the *logit*-link is adopted.

The likelihood associated with the observations changes from the LMM. The outcome does not have support in $\mathbb{R}$ anymore but it is limited in the interval $(0, 1)$. Thus, the Gaussian distribution is replaced by a Beta distribution written in the (2.7) reparameterization and with the precision parameter ϕ no longer constant. In some applications, observations close to 0 or 1 are censored and represented as exactly 0 and 1. For this, it is introduced a censor parameter $0 < \delta < \frac{1}{2}$ and all $y \leq \delta$ or $y \geq 1 - \delta$ are treated as censored observations. By default, no censoring is applied ($\delta = 0$).

The hyperparameters in the GLMM have the same priors as the LMM's hyperparameters. However, a new one is introduced in the context of Beta Mixed Models: the precision parameter ϕ. In R-INLA it is represented as $\phi = s_i \exp(\theta)$ where $s = (si) > 0$ is a fixed scaling parameter by default equal to 1. It is also possible to select a vector of scaled factors $\boldsymbol{s}$, for example, to consider a different precision for different groups of observations. The prior is defined on θ. It is assumed to follow by default a log-Gamma distribution with shape 1 and rate 0.1.

So, given only the time variable and one independent time-fixed variable x as in (2.3), the linear predictor would be:

$$\eta_{ij} = \beta_0 + b_{i0} + \beta_1 \cdot x_{ij} + (\beta_t + b_{it}) \cdot t_{ij} + \varepsilon_{ij}$$

and the default R-INLA hierarchical structure of a Beta Mixed Model with random intercepts and slopes can be written as:

$$\mathbf{y}_i \mid \mathbf{b}_i \sim Beta(\mu_i, \phi), \quad \mu_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}, \quad \phi = exp(\theta),$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_t \end{bmatrix} \sim \mathcal{N}_3 \left(0, \begin{bmatrix} 0 \\ 0.001 \\ 0.001 \end{bmatrix}^{-1} \right), \quad \begin{bmatrix} b_{i0} \\ b_{it} \end{bmatrix} \sim \mathcal{N}_2 \left(0, \boldsymbol{W}^{-1} \right),$$

$$\boldsymbol{W} \sim Wishart_2 \left(4, I_2 \right), \quad \theta \sim LogGamma \left(1, 0.1 \right).$$

From this, the structure can simply be generalized for a model with more than two independent variables. The R-code to define the Beta Mixed Models is shown in the Code A.2.

Chapter 3

Joint Models

In longitudinal studies, the covariates can be divided into two big groups: time-invariant and time-varying covariates. The former are constant over time, for example, sex or race; the latter instead are measured repeatedly during the trial and their values change over time, for instance, biomarkers or treatment doses. With this last type of covariates, measuring the association between the longitudinal response variable and the independent variables is not always trivial. Especially, in mixed models, the standard specification requires that time-varying covariates and outcomes be measured at the same time points. Although, this is not often the case (Wu 2019). Additionally, there are different types of time-varying covariates, called exogenous and endogenous covariates, and based on their nature further problems arise. In the following sections, we will then discuss more in detail the issues related to each type of time-dependent variable and we will introduce two types of joint models used to overcome these problems. In particular, the Joint Mixed Models (JMM) and the Joint Scaled Models (JSM) will be defined (Verbeke et al. 2014).

3.1 Time-varying covariates

As already mention, two types of time-dependent covariates exist: exogenous and endogenous covariates. It is important to be able to distinguish between them in order to understand which method is most appropriate for the data analysis.

To define both, the following notation will be used:

- $y_i(t)$ is the value of the response for the i-th subject at time t. It is an observation of the outcome's vector $\mathbf{y}_i$;
- $v_i(t)$ is the value of the time-varying covariate for the i-th subject at time t. It is an observation of the covariate's vector $\mathbf{v}_i$;
- $\mathcal{H}_i^Y(t)$ is the history of the response process until time t:
$$\mathcal{H}_i^Y(t) = \{y_i(t_{i1}), y_i(t_{i2}), \ldots, y_i(t_{ik}); t_{ik} \leq t\};$$
- $\mathcal{H}_i^V(t)$ is the history of the time-varying covariate process until time t:
$$\mathcal{H}_i^V(t) = \{v_i(t_{i1}), v_i(t_{i2}), \ldots, v_i(t_{ik}); t_{ik} \leq t\};$$
- $\mathbf{w}_i$ the vector of time-independent covariates;
- $\theta = (\theta_1, \theta_2)$ the vector of all the parameters, where θ_1 is the vector of parameters in the likelihood of the response variable and θ_2 the vector of parameters in the likelihood of the time-varying covariate.

3.1.1 Exogenous covariates

A time-varying covariate is exogenous if its current value at a specific time is only associated with its previous values, but is not further associated with previous values of the outcome. Mathematically speaking, $v_i(t)$ is exogenous if the exposure value at time t is conditionally independent of the history of the response variable, given the history of the exposure process:

$$f\left(v_i(t) \mid \mathcal{H}_i^Y(t), \mathcal{H}_i^V(t-1), \mathbf{w}_i\right) = f\left(v_i(t) \mid \mathcal{H}_i^V(t-1), \mathbf{w}_i\right).$$

Or, even simpler, $\mathbf{v}$ is exogenous with respect the outcome process $\mathbf{y}$ if $\mathbf{v}_i(t) \perp \bar{\mathbf{y}}_i(t) \mid \bar{\mathbf{v}}_i(t-1)$ (See Qian, Klasnja, and Murphy 2020). Examples of these kinds of variables are age and the time itself, or daily climate changes. For instance, we can consider a study in which the goal is to evaluate how much air pollution can affect the presence of asthma in patients of different generations. In this case the quantity of air pollution changes in time but it does not depend on the outcome.

For exogenous covariate holds that the joint likelihood $f\left(\mathbf{y}_i, \mathbf{v}_i \mid \mathbf{w}_i, \boldsymbol{\theta}\right)$ can be factorized:

$$\begin{aligned} f\left(\mathbf{y}_i, \mathbf{v}_i \mid \mathbf{w}_i, \boldsymbol{\theta}\right) = & \left[\prod_{t=1}^{T} f\left(y_i(t) \mid \mathcal{H}_i^Y(t-1), \mathcal{H}_i^V(t), \mathbf{w}_i, \boldsymbol{\theta}_1\right)\right] \\ & \cdot \left[\prod_{t=1}^{T} f\left(v_i(t) \mid \mathcal{H}_i^V(t-1), \mathbf{w}_i, \theta_2\right)\right] = \\ & = \mathcal{L}_Y\left(\boldsymbol{\theta}_1\right) \cdot \mathcal{L}_V\left(\boldsymbol{\theta}_2\right). \end{aligned} \tag{3.1}$$

Hence, the two processes can be modeled separately. It is not necessary to model the covariate process in order to make inference about the outcome.

In the context of LMM, exogeneity implies that $\mathbf{v}_i$ does not depend on the random effects $\mathbf{b}_i$, therefore it is possible to have both a conditional and marginal interpretation. In a linear mixed model defined as in equation 2.2, the relationship between the covariates at time t and the response variable at time $t+1$ is:

$$y_{it+1} = X_{it}\boldsymbol{\beta} + Z_{it}\mathbf{b}_i + \varepsilon_{it+1}, \quad \mathbf{b}_i \sim \mathcal{N}(0, \boldsymbol{D}), \quad \varepsilon_{it+1} \sim \mathcal{N}\left(0, \sigma\right),$$

in which all the covariates $X_{it} = (w_{it}, v_{it})$ are fixed or at least exogenous. We can simply obtain the conditional and the marginal distribution of y_{it+1} from this model. In particular, the conditional distribution of y_{it+1} given X_{it} and the random effects $\mathbf{b}_i$ is a Gaussian distribution with mean

$$E\left(y_{it+1} \mid X_{it}, \mathbf{b}_i\right) = X_{it}\boldsymbol{\beta} + Z_{it}\mathbf{b}_i.$$

Otherwise, the mean of the marginal distribution can be expressed as

$$E\left(y_{it+1} \mid X_{it}\right) = X_{it}\boldsymbol{\beta},$$

because $E\left(\boldsymbol{\beta} \mid X_{it}\right) = 0$ under the assumption of exogenous and fixed independent variables. Thus, the hierarchical model implies the marginal model. In the generalized linear mixed model, instead, this relation does not occur due to the nonlinear link function, even when all the covariates are exogenous.

3.1.2 Endogenous covariates

A time-varying covariate is endogenous if its value at a specific time is associated with its previous values and with values of the outcome at previous time points. Thus, $v_i(t)$ is endogenous if the exposure at time t is conditionally dependent on the history of the response variable, given the history of the exposure process:

$$f\left(v_i(t) \mid \mathcal{H}_i^Y(t), \mathcal{H}_i^V(t-1), \mathbf{w}_i\right) \neq f\left(v_i(t) \mid \mathcal{H}_i^V(t-1), \mathbf{w}_i\right).$$

Examples of these kinds of variables are the biomarkers or the treatment regimen. Let's for example consider again a study on patients with asthma. We want to define how much the use of inhalers affects the severity of symptoms. It may be then that the amount of inhaler use depends at the same time on the course of the disease during the observation period. So, the treatment regimen changes in time and it depends on the outcome.

Since the exposure and the response variable are not conditionally independent, the likelihood can not be factorized as in equation 3.1. The covariate's process can no longer be ignored and the two variables must be modeled jointly. Furthermore, the marginal distribution of y_{it+1} in a linear mixed model can no be defined easily:

$$E\left(y_{it+1} \mid X_{it}\right) = X_{it}\boldsymbol{\beta} + Z_{it}E\left(\mathbf{b}_i \mid X_{it}\right).$$

Dependence on the outcome implies dependence on random effects, so $E\left(\mathbf{b}_i \mid X_{it}\right)$ is usually not equal to zero. Hence, when the covariates are endogenous, the regression coefficients have only conditional interpretation, and the LMMs are not valid anymore. In order to obtain the marginal interpretation of the coefficient and to quantify correctly the association between the endogenous variable and the outcome, we will use two types of joint models. These kinds of models allow us to estimate the association even when the outcome and the time-varying variable are measured on different time points and there are missing values on both.

Notice that in the following sections, for the sake of simplicity, with v we will refer to exogenous covariates, and with x we will denote endogenous covariates.

3.2 Joint Mixed Model

The first type of joint model we examined is the Joint Mixed Model (see Weiss 2005, Fieuws and Verbeke 2006 and Fieuws and Verbeke 2004). Here the association between the endogenous time-varying variable x and the outcomes y is measured via the random effects variance-covariance matrix $\mathbf{D}$. The mathematical form of the joint model assuming only one time-varying covariate and one outcome is:

$$\begin{cases} x_i\left(s_{ij}\right) = \mathbf{w}_{xi}^{\top} \cdot \boldsymbol{\alpha}_x + \mathbf{v}_{xi}^{\top}\left(s_{ij}\right)\boldsymbol{\beta}_x + \mathbf{z}_{xi}^{\top}\left(s_{ij}\right)\mathbf{b}_{xi} + \varepsilon_{xi}\left(s_{ij}\right) \\ y_i\left(t_{ij}\right) = \mathbf{w}_{yi}^{\top} \cdot \boldsymbol{\alpha}_y + \mathbf{v}_{yi}^{\top}\left(t_{ij}\right)\boldsymbol{\beta}_y + \mathbf{z}_{yi}^{\top}\left(t_{ij}\right)\mathbf{b}_{yi} + \varepsilon_{yi}\left(t_{ij}\right) \end{cases} \tag{3.2}$$

where

$$\begin{bmatrix} \mathbf{b}_{xi} \\ \mathbf{b}_{yi} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{D}_{xx} & \boldsymbol{D}_{xy} \\ \boldsymbol{D}_{yx} & \boldsymbol{D}_{yy} \end{bmatrix}\right); \quad \begin{bmatrix} \varepsilon_{yi} \\ \varepsilon_{xi} \end{bmatrix} \sim \mathcal{N}_{2n_i}\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right).$$

Here, $\boldsymbol{\alpha}$'s are the fixed-effects vectors and $\boldsymbol{\beta}$'s are the vector of the regression coefficients of the exogenous variables. The indexes s_{ij} and t_{ij} are the time points of the covariate and the outcome respectively, and they can be different. $x_i\left(s_{ij}\right)$ is the value of the endogenous variable x for the i-th subject, $i = 1, \ldots, N$, at time s_{ij}, $j = 1, \ldots, n_i$. The value of the response variable y for the individual i at the time point t_{ij} is $y_i\left(t_{ij}\right)$. There are two sets of error terms, ε_{yi} and ε_{xi}, both vectors of length equal to n_i, whose variance-covariance matrix is generally assumed $\boldsymbol{\Sigma}_i = \sigma_{\varepsilon}^2 I$ implying Conditional Independence Assumption (see Chapter 2). They are independent of the random effects. We assume that the random effects follow a joint multivariate Gaussian distribution with zero mean and an unstructured variance-covariance matrix. If each process has both random intercepts and random slopes, **D** is a block matrix with as submatrices the variance-covariance matrix of the random effects of the covariate D_{xx}, the variance-covariance matrix of the random effects of the outcome D_{yy} and the covariance matrix $D_{yx} = D_{xy}$ between them. If D_{yx} is null, the random effects are not correlated and there is no association between the response and the endogenous variables.

It is important to stress the fact that if the two time vectors are different only the Conditional Independence Assumption is possible. The correlation between residual errors is allowed only for outcome and time-varying variable assessed at the same time. With non-simultaneously observations, the pair of residual errors are not measured at the same time, so it is not possible to evaluate the correlation between them. Thus, the association between x and y is only computed by the variance-covariance matrix of the random effects.

Here it is assumed a normal likelihood for both the outcome and the endogenous variable and the same model structure, but in general, they can have different distributions. We will see a particular case in Chapter 5 in which a beta mixed model and a linear mixed model will be joined.

3.2.1 Association between outcome and endogenous covariate

So far, the only way to evaluate the association between the endogenous and the response variables is by studying the covariance parameters in the variance-covariance matrix **D**. If these values are not significantly different from zero, the model can be reduced to a LMM without the time-varying covariate, and the two processes can be modelled separately. In the context of Bayesian inference, credible intervals are used to evaluate the significance of the parameters. However, the covariance parameters can not be easily interpreted. They can give information about the order of magnitude of the association (big covariance leads to strong association), but they can not provide any information about the direction. What we would like to have is an association coefficient with a similar interpretation of the other regression coefficients, i.e., a parameter that estimates the change in mean outcome for a unitary increase in the endogenous covariate, given that all other covariates are fixed (Gomon 2022). To obtain an estimation of this association parameter, we need to define the conditional distribution of the outcome given the endogenous variable. Let's assume for now that both the outcome and the time-varying covariate are continuous in $\mathbb{R}$ and they can be modeled with two LMMs. The joint distribution of the two processes is then a bivariate Gaussian:

$$f\left(x_i(t), y_i(t)\right) = \mathcal{N}_2\left(\begin{bmatrix} \mu_{x,i}(t) \\ \mu_{y,i}(t) \end{bmatrix}, \begin{bmatrix} \sigma^2_{x,i}(t) & \rho_i(t)\sigma_{x,i}(t)\sigma_{y,i}(t) \\ \rho_i(t)\sigma_{y,i}(t)\sigma_{x,i}(t) & \sigma^2_{y,i}(t) \end{bmatrix}\right)$$

where

$$\begin{aligned} \mu_{x,i}(t) &= \mathbf{w}_{xi}^\top \cdot \boldsymbol{\alpha}_y + \mathbf{v}_{xi}^\top(t)\boldsymbol{\beta}_x, \quad \sigma^2_{x,i}(t) = \mathbf{z}_{xi}^\top(t)\mathbf{D}_{xx}\mathbf{z}_{xi}(t) + \sigma^2_{\varepsilon,x} \\ \mu_{y,i}(t) &= \mathbf{w}_{yi}^\top \cdot \boldsymbol{\alpha}_y + \mathbf{v}_{yi}^\top(t)\boldsymbol{\beta}_y, \quad \sigma^2_{y,i}(t) = \mathbf{z}_{yi}^\top(t)\mathbf{D}_{yy}\mathbf{z}_{yi}(t) + \sigma^2_{\varepsilon,y}. \end{aligned}$$

Here, the variances $\sigma^2_{x,i}(t)$ and $\sigma^2_{y,i}(t)$ and the correlation $\rho_i(t)$ all depend on the random effects and error terms variance-covariance matrices $\mathbf{D}$ and $\boldsymbol{\Sigma}_i$. The specific form of the correlation parameter can not be specified because depends on how many random effects are assumed for each equation and on the structure of $\mathbf{D}$. Note that all the elements just described are time-dependent.

From the bivariate normal, the conditional distribution of the response variable y given the time-varying covariate x can be simply derived:

$$f\left(y_i(t) \mid x_i(t) = a\right) = \mathcal{N}\left(\mu_{y,i}(t) + \frac{\sigma_{y,i}(t)}{\sigma_{x,i}(t)}\rho_i(t)\left(a - \mu_{x,i}(t)\right), \left(1 - \rho_i^2(t)\right)\sigma^2_{y,i}(t)\right).$$

Using the conditional distribution we can answer different questions regarding the relationship between the two variables. We can measure the parameter that estimate the cross-sectional effect, so the change in mean outcome for a unitary increase in the endogenous covariate given the same time point:

$$
\begin{aligned}
&E\left[f\left(y_i(t) \mid x_i(t)=a+1\right)\right]-E\left[f\left(y_i(t) \mid x_i(t)=a\right)\right]= \\
&=\left(\mu_{y,i}(t)+\frac{\sigma_{y,i}(t)}{\sigma_{x,i}(t)}\rho_i(t)\left(a+1-\mu_{x,i}(t)\right)\right)-\left(\mu_{y,i}(t)+\frac{\sigma_{y,i}(t)}{\sigma_{x,i}(t)}\rho_i(t)\left(a-\mu_{x,i}(t)\right)\right)= \\
&=\frac{\sigma_{y,i}(t)}{\sigma_{x,i}(t)}\rho_i(t)=\frac{\operatorname{Cov}(y,x)(t)}{\sigma_{x,i}^2(t)}.
\end{aligned}
$$

But we can also estimate a coefficient for the lag-effect, which indicates how the covariate at the previous time points affects the response variable in the present. It is calculated following the same idea:

$$
\begin{aligned}
&E\left[f\left(y_i(t) \mid x_i(t-k)=a+1\right)\right]-E\left[f\left(y_i(t) \mid x_i(t-k)=a\right)\right]= \\
&=\frac{\sigma_{y,i}(t)}{\sigma_{x,i}(t-k)}\rho_i(t,t-k)=\frac{\operatorname{Cov}(y,x)(t,t-k)}{\sigma_{x,i}^2(t-k)}.
\end{aligned}
$$

In conclusion, the cross-sectional association coefficient and the lag-effect association coefficient are respectively:

$$
\beta_x^{jmm}(t)=\frac{\operatorname{Cov}(y,x)(t)}{\sigma_{x,i}^2(t)}, \tag{3.3}
$$

$$
\beta_x^{jmm}(t,t-k)=\frac{\operatorname{Cov}(y,x)(t,t-k)}{\sigma_{x,i}^2(t-k)}. \tag{3.4}
$$

Note that equations 3.3 and 3.4 are valid only when we have one outcome and one endogenous covariate and they do not apply with a higher number of variables. The conditional distribution of more than two variables is less easy to get. In the context of normal likelihoods with three variables already the derivation process is more complex. Moreover, it is not possible to obtain a closed-form estimation of the association coefficient when variables are of different types (binary, count, ...). An example is provided in chapter 5, with further discussions too, in which a joint mixed model with a Beta and a normal likelihood is implemented.

3.2.2 INLA model formulation

In the following section, we will show how to rewrite the JMM as Latent Gaussian Model. In order to provide a better explanation, for both the process only one fixed covariate w and random intercepts and slopes are assumed:

$$\begin{cases} x_i\left(s_{ij}\right) = \left(\beta_0^{(x)} + b_{0,i}^{(x)}\right) + \beta_w^{(x)} \cdot w_i + \left(\beta_s^{(x)} + b_{t,i}^{(x)}\right) \cdot s_{ij} + \varepsilon_i^{(x)}\left(s_{ij}\right) \\ y_i\left(t_{ij}\right) = \left(\beta_0^{(y)} + b_{0,i}^{(y)}\right) + \beta_w^{(y)} \cdot w_i + \left(\beta_t^{(y)} + b_{t,i}^{(y)}\right) \cdot t_{ij} + \varepsilon_i^{(y)}\left(t_{ij}\right) \end{cases} \tag{3.5}$$

where

$$\begin{bmatrix} b_0^{(x)} \\ b_0^{(y)} \\ b_t^{(x)} \\ b_t^{(y)} \end{bmatrix} \sim \mathcal{N}_4(\mathbf{0}, \mathbf{D}); \quad \begin{bmatrix} \varepsilon_i^{(x)} \\ \varepsilon_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2n_i}\left(\mathbf{0}, \begin{bmatrix} \sigma^2_{\varepsilon,x}\mathbf{I}_{n_i} & 0 \\ 0 & \sigma^2_{\varepsilon,y}\mathbf{I}_{n_i} \end{bmatrix}\right).$$

Note that the vector of random effects is defined differently from before. Here, the order is determined by the type of random effects and not by the variable to which they belong. The first two elements are the random intercepts of the endogenous covariate and the outcome, and the second ones are the corresponding random slopes. Their variance-covariance matrix $\mathbf{D}$ can have different structures. The two forms that we have used are the unstructured matrix and the pairwise-correlated matrix. The former assumes correlation between all the random effects:

$$\mathbf{D} = \begin{bmatrix} \sigma^2_{x,0} & \sigma_{(x,0),(y,0)} & \sigma_{(x,0),(x,t)} & \sigma_{(x,0),(y,t)} \\ \sigma_{(y,0),(x,0)} & \sigma^2_{y,0} & \sigma_{(y,0),(x,t)} & \sigma_{(y,0),(y,t)} \\ \sigma_{(x,t),(x,0)} & \sigma_{(x,t),(y,0)} & \sigma^2_{x,t} & \sigma_{(x,t),(y,t)} \\ \sigma_{(y,t),(x,0)} & \sigma_{(y,t),(y,0)} & \sigma_{(y,t),(x,t)} & \sigma^2_{y,t} \end{bmatrix}. \tag{3.6}$$

The latter, instead, assume correlation only between pair of random effects. This implies that the number of parameters to be estimated is lower, and could be a useful property in some cases. The matrix $\mathbf{D}$ is a block-diagonal matrix that has the following form:

$$\mathbf{D} = \begin{bmatrix} \sigma^2_{x,0} & \sigma_{(x,0),(y,0)} & 0 & 0 \\ \sigma_{(y,0),(x,0)} & \sigma^2_{y,0} & 0 & 0 \\ 0 & 0 & \sigma^2_{x,t} & \sigma_{(x,t),(y,t)} \\ 0 & 0 & \sigma_{(y,t),(x,t)} & \sigma^2_{y,t} \end{bmatrix}. \tag{3.7}$$

Other structures for the variance-covariance matrix are possible, for instance assuming auto-regressive or random-walk random effects. In this thesis though, we only use the two forms described above. In R-INLA, indeed, there are some limitations. In particular, it is not possible to estimate more than five correlated random effects (for more see Gómez-Rubio 2020).

The joint model presented is a two-likelihood model. The first likelihood for the endogenous covariate, while the second one will be of the types used to model the outcome. In order to fit multiple likelihoods in R-INLA, the data must be adapted to a specific structure. The response variable of the joint model is a matrix with a number of columns equal to the number of likelihoods considered, and with as many rows as total observations for each variable. In our case, it is a matrix of dimension $(2N_{obs} \times 2)$. The first N_{obs} rows are related to the endogenous variable, the remaining rows correspond to the outcome. Note that we are assuming the same number of observations N_{obs} per variable. It is also possible to have a different number of observations, hence to have missing values in the response and the covariate process. However, INLA automatically omits the missing values and estimates the models on the observed ones, so, for simplicity of notation, we can assume the same number of measurements. Even if the input is a matrix, the inner representation of R-INLA response variable output is a vector of length $2N_{obs}$:

$$(x_{1,1}, \dots, x_{N,n_N}, y_{1,1}, \dots, y_{N,n_N})$$

The same idea applies to store the joint linear predictor. It is a block-diagonal matrix in which each block corresponds to the model matrix of every variable. All blocks outside the main diagonal contain missing values. For the joint model described above the linear predictor is a matrix with six columns, one for each fixed effect, and $2N_{obs}$ rows. This concept is illustrated below based on the model (3.5).

$$\begin{array}{cc} \text{Response variables:} & \text{Linear predictor:} \\ (\mathbf{x}, \mathbf{y}) & (\text{Intercept}_x, w_x, t_x, \text{Intercept}_y, w_y, t_y) \\ \begin{bmatrix} x_{1,1} & NA \\ x_{1,2} & NA \\ \vdots & \vdots \\ x_{N,n_N} & NA \\ NA & y_{1,1} \\ NA & y_{1,2} \\ \vdots & \vdots \\ NA & y_{N,n_N} \end{bmatrix}, & \begin{bmatrix} 1 & w_{1,1} & t_{1,1} & NA & NA & NA \\ 1 & w_{1,2} & t_{1,2} & NA & NA & NA \\ \vdots & \vdots & & & & \\ 1 & w_{N,n_N} & t_{N,n_N} & NA & NA & NA \\ NA & NA & NA & 1 & w_{1,1} & t_{1,1} \\ NA & NA & NA & 1 & w_{1,2} & t_{1,2} \\ \vdots & \vdots & & & & \\ NA & NA & NA & 1 & w_{N,n_N} & t_{N,n_N} \end{bmatrix} \end{array}$$

For the random effects, the specification is based on the form of the variance-covariance matrix. If an unstructured variance-covariance matrix (3.6) is assumed, the random effects must be defined as a vector as in equation (3.5). If we consider a pairwise variance-covariance matrix, instead, the random effects need to be stored as a matrix with as many columns as random effects and $2N$ rows:

$$\begin{bmatrix} b^{(x)}_{0,1} & b^{(y)}_{0,1} & NA & NA \\ b^{(x)}_{0,2} & b^{(y)}_{0,2} & NA & NA \\ \vdots & \vdots & & \\ b^{(x)}_{0,N} & b^{(y)}_{0,N} & NA & NA \\ NA & NA & b^{(x)}_{1,1} & b^{(y)}_{1,1} \\ NA & NA & b^{(x)}_{1,2} & b^{(y)}_{1,2} \\ \vdots & \vdots & & \\ NA & NA & b^{(x)}_{1,N} & b^{(y)}_{1,N} \end{bmatrix}.$$

Note that with both the specification the random effects output vector is represented internally as one vector of length $4N$:

$$(b^{(x)}_{0,1}, \dots, b^{(x)}_{0,N}, b^{(y)}_{0,1}, \dots, b^{(y)}_{0,N}, b^{(x)}_{1,1}, \dots, b^{(x)}_{1,N}, b^{(y)}_{1,1}, \dots, b^{(y)}_{1,N}).$$

The R-code used to define the Joint Mixed Models with unstructured and pairwise variance-covariance matrices are respectively shown in Code A.3 and Code A.4. See also Niekerk et al. 2022.

Once the data are prepared, INLA implements the model assuming for each parameter a prior distribution. We have already described the default choices of R-INLA in Section 2.1.1 and they are valid for joint models as well.

3.3 Joint Scaled Model

The second type of joint model we present is the Joint Scaled Model. Usually, they are used to combine a longitudinal model and a survival model (D. Rizopoulos 2017). In these types of models the association between the endogenous time-varying variable x and the outcomes y is measured via a scaling factor γ. The mathematical notation is:

$$\begin{cases} x_i\left(s_{ij}\right) = m_i\left(s_{ij}\right) + \varepsilon_{xi}\left(s_{ij}\right) \\ y_i\left(t_{ij}\right) = \mathbf{w}_{yi}^{\top}\alpha_{\mathbf{y}} + \gamma m_i\left(t_{ij}\right) + \mathbf{v}_{yi}^{\top}\left(t_{ij}\right)\beta_y + \mathbf{z}_{yi}^{\top}\left(t_{ij}\right)\mathbf{b}_{yi} + \varepsilon_{yi}\left(t_{ij}\right) \end{cases} \tag{3.8}$$

where

$$m_i\left(s_{ij}\right) = \mathbf{w}_{xi}^{T}\alpha_{\mathbf{x}} + \mathbf{v}_{xi}^{T}\left(s_{ij}\right)\beta_{\mathbf{x}} + \mathbf{z}_{xi}^{T}\left(s_{ij}\right)\mathbf{b}_{xi}$$

and

$$\begin{aligned} \mathbf{b}_{xi} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{D}_x\right), \quad \varepsilon_{xi}\left(s_{ij}\right) \sim \mathcal{N}_{n_i}\left(\mathbf{0}, \sigma_{\varepsilon,x}^2\right) \\ \mathbf{b}_{yi} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{D}_y\right), \quad \varepsilon_{yi}\left(t_{ij}\right) \sim \mathcal{N}_{n_i}\left(\mathbf{0}, \sigma_{\varepsilon,y}^2\right) \end{aligned}$$

Here, the random effects and the error terms of x and y are independent. Thus, $\varepsilon_{xi}\left(s_{ij}\right) \perp \mathbf{b}_{xi}$ and $\varepsilon_{yi}\left(t_{ij}\right) \perp \mathbf{b}_{yi}$ and $\varepsilon_{xi}\left(s_{ij}\right) \perp \varepsilon_{yi}\left(t_{ij}\right)$ and $\mathbf{b}_{xi} \perp \mathbf{b}_{yi}$. The association between the response and the time-varying variables derives only from the linear predictor $m_i\left(s_{ij}\right)$ of the endogenous variable x at time s_{ij}, and the associated scaling factor γ. Note that the time points s_{ij} and t_{ij} can be different in this model as well.

Assuming identical covariates for the two processes, therefore also the same time vector $t_{ij} = s_{ij}\ \forall i, j$, the expression for the outcome can be rewritten in the following way:

$$\begin{aligned}
y_i\left(s_{ij}\right) = \gamma m_i\left(s_{ij}\right) + \mathbf{w}_{yi}^{\top}\alpha_{\mathbf{y}} + \mathbf{v}_{yi}^{\top}\left(s_{ij}\right)\beta_{\mathbf{y}} + \mathbf{z}_{yi}^{\top}\left(s_{ij}\right)\mathbf{b}_{yi} + \varepsilon_{yi}\left(s_{ij}\right) = \\
= \left(\gamma\alpha_{\mathbf{x}} + \alpha_{\mathbf{y}}\right)\mathbf{w}_{yi}^{\top} + \left(\gamma\beta_{\mathbf{x}} + \beta_{\mathbf{y}}\right)\mathbf{v}_{yi}^{\top}\left(s_{ij}\right) + \left(\gamma\mathbf{b}_{xi} + \mathbf{b}_{yi}\right)\mathbf{z}_{yi}^{\top}\left(s_{ij}\right) + \varepsilon_{yi}\left(s_{ij}\right) = \\
= \alpha_{\mathbf{y}}'\mathbf{w}_{yi}^{\top} + \beta_{\mathbf{y}}'\mathbf{v}_{yi}^{\top}\left(s_{ij}\right) + \mathbf{b}_{yi}'\mathbf{z}_{yi}^{\top}\left(s_{ij}\right) + \varepsilon_{yi}\left(s_{ij}\right).
\end{aligned} \tag{3.9}$$

Here, the random effects $\mathbf{b}_{yi}' = \gamma\mathbf{b}_{xi} + \mathbf{b}_{yi}$ follow a Gaussian distribution with zero mean and variance-covariance matrix $\gamma^2\mathbf{D}_x + \mathbf{D}_y$. The benefit of the combined coefficients is that they more closely resemble the coefficients obtained fitting the JMM. We will especially use this reparameterization in the simulation described in chapter 4.

3.3.1 Association between outcome and endogenous covariate

In the Joint Scaled Models the first way to evaluate and quantify the association between the endogenous covariate and the outcome it is by the scaling factor γ. Another possibility is to estimate the association coefficient following the same process as in Section 3.2.1. To measure the change in mean outcome y with a unitary increase in time-varying covariate x, we need to determine the expectation of the difference of the conditional distributions for a i-th subject:

$$\begin{aligned}
& E\left(f\left[y_i(t) \mid x_i(t) = a+1\right]\right) - E\left(f\left[y_i(t) \mid x_i(t) = a\right]\right) = \\
& = E\left(\mathbf{w}_{yi}^{\top}\alpha_{\mathbf{y}} + \gamma m_i(t) + \mathbf{v}_{yi}^{\top}(t)\beta_{\mathbf{y}} + \mathbf{z}_{yi}^{\top}(t)\mathbf{b}_{yi} + \varepsilon_{yi}(t) \mid x_i(t) = a+1\right) + \\
& - E\left(\mathbf{w}_{yi}^{\top}\alpha_{\mathbf{y}} + \gamma m_i(t) + \mathbf{v}_{yi}^{\top}(t)\beta_{\mathbf{y}} + \mathbf{z}_{yi}^{\top}(t)\mathbf{b}_{yi} + \varepsilon_{yi}(t) \mid x_i(t) = a\right) = \\
& = \gamma E\left(m_i(t) \mid x_i(t) = a+1\right) - \gamma E\left(m_i(t) \mid x_i(t) = a\right).
\end{aligned} \tag{3.10}$$

The only element in the linear predictor of y that depend on the values of the endogenous variable x, is the linear predictor $m_i(t)$. Therefore, all the other components can be taken out of the conditional expectation. Since we know the marginal distributions of all the members:

$$\begin{aligned} m_i(t) &= \mathcal{N}\left(\mathbf{w}_{xi}^{\top}\alpha_{\mathbf{x}} + \mathbf{v}_{xi}^{\top}(t)\beta_{\mathbf{x}}, \mathbf{z}_{xi}^{\top}(t)\mathbf{D}_x\mathbf{z}_{xi}(t)\right), \\ \varepsilon_i(t) &= \mathcal{N}\left(0, \sigma_x^2\right), \\ x_i(t) &= \mathcal{N}\left(\mathbf{w}_{xi}^{\top}\alpha_{\mathbf{x}} + \mathbf{v}_{xi}^{\top}(t)\beta_{\mathbf{x}}, \mathbf{z}_{xi}^{\top}(t)\mathbf{D}_x\mathbf{z}_{xi}(t) + \sigma_{\varepsilon,x}^2\right), \end{aligned}$$

the joint distribution of $m_i(t)$ and $x_i(t)$ is a bivariate normal, with as mean the vector of the two means and as variance the matrix

$$\left[\begin{array}{cc} \sigma_{x,i}^2(t) & \rho_i(t)\sigma_{x,i}(t)\sigma_{y,i}(t) \\ \rho_i(t)\sigma_{y,i}(t)\sigma_{x,i}(t) & \sigma_{y,i}^2(t) \end{array}\right].$$

Once again, the mean of the conditional distribution is simply derived and corresponds to:

$$E\left(m_i(t) \mid x_i(t) = a\right) = \mu_{m_i}(t) + \frac{\operatorname{Cov}\left(m_i(t), x_i(t)\right)}{\sigma_{x,i}^2(t)}\left(a - \mu_{x,i}(t)\right)$$

where

$$\begin{aligned} \mu_{x,i}(t) &= \mathbf{w}_i^{\top} \cdot \boldsymbol{\alpha}_y + \mathbf{v}_{xi}^{\top}(t)\boldsymbol{\beta}_x, \\ \mu_{y,i}(t) &= \mathbf{w}_i^{\top} \cdot \boldsymbol{\alpha}_x + \mathbf{v}_{yi}^{\top}(s)\boldsymbol{\beta}_y, \\ \operatorname{Cov}\left(m_i(t), x_i(t)\right) &= \operatorname{Var}\left(m_i(t)\right) = \mathbf{z}_{xi}^{\top}(t)\mathbf{D}_x\mathbf{z}_{xi}(t). \end{aligned}$$

In conclusion, replacing all the above elements in equation (3.10), the association coefficient is estimated:

$$\beta_x^{jsm}(t) = \gamma\left[\frac{\mathbf{z}_{xi}^{\top}(t)\mathbf{D}_x\mathbf{z}_{xi}(t)}{\mathbf{z}_{xi}^{\top}(t)\mathbf{D}_x\mathbf{z}_{xi}(t) + \sigma_x^2}\right] = \gamma\left[1 - \frac{\sigma_{\varepsilon,x}^2}{\mathbf{z}_{xi}^{\top}(t)\mathbf{D}_x\mathbf{z}_{xi}(t) + \sigma_{\varepsilon,x}^2}\right] \tag{3.11}$$

This parameter has the same interpretation of the association coefficient $\beta_x^{jmm}(t)$ in the JMM (3.3) and the fixed effect regression coefficients in any LMM. To evaluate the significance of these coefficients no tests are available. We shall construct credible intervals. Since we are working under a Bayesian structure, a marginal posterior distribution for each parameter and hyperparameter is automatically obtained. When the posterior marginal of a non-linear transformation of the hyperparameter or functions that depend on several hyperparameters are required, it is suggested to sample from their approximate joint posterior distribution (see Gómez-Rubio 2020). Once the posterior marginal distribution of expressions (3.3) and (3.11) is derived, the credible intervals are available too.

However, the closed-form (3.11) applies only with normal likelihoods. When we want to model jointly two variables of different types, the estimation is more complicated. In this last scenario, the mathematical convenience of multivariate normal is lost and the following integration needs to be evaluated numerically:

$$f\left(y_{ij}, x_{ij}\right) = \int f\left(y_{ij} \mid m_{ij}, \mathbf{b}_{yi}\right) f\left(x_{ij} \mid \mathbf{b}_{xi}\right) f\left(\mathbf{b}_{yi}\right) f\left(\mathbf{b}_{xi}\right) d\mathbf{b}_{yi} d\mathbf{b}_{xi}.$$

This integral can be solve by numerical integration or Monte Carlo sampling but we tried to find an approximation of the closed-form formula using as best we could the Bayesian output provided by R-INLA. For further discussion about this issue refers to Section 5.4.3, in which an example with Beta mixed model is shown.

3.3.2 INLA model formulation

Let us consider for simplicity the following model:

$$\begin{cases} m_i\left(s_{ij}\right) = \left(\beta_0^{(x)} + b_{0,i}^{(x)}\right) + \beta_w^{(x)} \cdot w_i + \left(\beta_t^{(x)} + b_{t,i}^{(x)}\right) \cdot s_{ij} \\ x_i\left(s_{ij}\right) = m_i\left(s_{ij}\right) + \varepsilon_i^{(x)}\left(s_{ij}\right) \\ y_i\left(t_{ij}\right) = \gamma \cdot m_i\left(t_{ij}\right) + \left(\beta_0^{(y)} + b_{0,i}^{(y)}\right) + \beta_w^{(y)} \cdot w_i + \left(\beta_t^{(y)} + b_{t,i}^{(y)}\right) \cdot t_{ij} + \varepsilon_i^{(y)}\left(t_{ij}\right) \end{cases}$$

where

$$\begin{bmatrix} b_0^{(x)} \\ b_t^{(x)} \end{bmatrix} \sim \mathcal{N}_2\left(0, \begin{bmatrix} \sigma_{x,0}^2 & \sigma_{x,(0,t)} \\ \sigma_{x,(t,0)} & \sigma_{x,t}^2 \end{bmatrix}\right),$$

$$\begin{bmatrix} b_0^{(y)} \\ b_t^{(y)} \end{bmatrix} \sim \mathcal{N}_2\left(0, \begin{bmatrix} \sigma_{y,0}^2 & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma_{y,t}^2 \end{bmatrix}\right),$$

$$\begin{bmatrix} \varepsilon_i^{(x)} \\ \varepsilon_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2n_i}\left(0, \begin{bmatrix} \sigma_{\varepsilon,x}^2 \mathbf{I}_{n_i} & 0 \\ 0 & \sigma_{\varepsilon,y}^2 \mathbf{I}_{n_i} \end{bmatrix}\right).$$

As we have already explained in Section 3.2.2, in order to implement joint models in R-INLA it is required to rewrite the dataset in a specific way. The idea is the same as before: everything has to be stored in matrices in which the first block of rows is related to the endogenous variable, and the second block to the outcome. What changes in the specification of the JSM is the classification between fixed and random effects. In R-INLA only the random effects are allowed to be scaled by a factor γ and be copied into a different likelihood. Thus, the exclusive way to copy and scaled the linear predictor $m(s_{ij})$ in the linear predictor of y is by treating the fixed effects as independent random effects. Their priors are then the same as the error terms described in Section 2.1.1. During the estimation process, they are copied with the same scaling factor plus some tiny noise in the linear predictor of y. More formally, let's assume a latent effect $\mathbf{u} = (u_1, \ldots, u_p)$. The copied effect $\mathbf{u}^*$ is defined as:

$$u_j^* = \gamma u_j + \varepsilon_j, \quad j = 1, \ldots, n$$

where n is the number of observations. The error ε_j has a Gaussian prior, with a very large precision equal to exp(14) by default (for computational reason). The scale factor γ can be estimated or set as fixed. It the context of JSM it is considered as a hyperparameter with as default prior a gaussian distribution with mean 1 and precision 10. So, the fixed effects are written as random effects in the endogenous variable linear predictor and then are copied in the linear predictor of the outcome rescaled by γ.

This way of proceeding is legitimate because within the Bayesian framework the difference between fixed and random effects is more subtle than in a frequentist approach. Both fixed and random effects, indeed, are random variables with a certain prior probability.

The R-code used to implement this trick is shown in Code A.5 and Niekerk et al. 2022.

Chapter 4

Simulation

The main goal of the thesis is to evaluate the performance of the joint models estimated by R-INLA. Thus, the problem questions are especially two: how accurate are the results obtained using INLA as a method of estimation, and if there are any differences between the performances of JMM, JSM and LMM. We are interested in understanding how good the JMM and the JSM are when we have an endogenous covariate, compared to the LMM, and if the INLA results are more or less accurate under different conditions. In particular, we shall investigate how is the performance when the variances of the random effects and the error terms are really small, thus near the boundary. In the context of the joint models, are parameters indeed very relevant for a good estimation because of their connection with the association coefficients (see equations 3.3 and 3.11). When they are not correctly estimated, the association coefficient estimates are in turn affected.

To accomplish this, we will simulate $2M = 2000$ samples, $M = 1000$ generated according to a JMM and the other $M = 1000$ simulated from a JSM. For each sample, the three models (LMM, JMM and JSM) are fitted and the M results are then summarized and compared. Once we have done this in datasets in which large variances and covariances of the random effects are assumed, we repeat the same simulation study in samples with smaller values.

In the following sections, we will describe in more detail the data generation process and how the simulation was performed.

4.1 Data generation process

In longitudinal studies, the data should be in a long format, which means that every row of the dataset refers to a specific subject and to only one of his repeated measurements. For each row then, we have a different value for the time variable, as well as a different value for the longitudinal outcome and the time-varying covariates. Generally, the number of repeated measurements for each subject is different and the data are not collected at the same time point for all the units, which means that the data are unbalanced. Furthermore, it is possible that also the time-varying covariate and the outcome are measured at different moments. In order to reproduce this design, two variables are introduced in the simulated dataset: y_obs and x_obs. They are vectors of zeros and ones, randomly sampled from Bernoulli distributions with probabilities $p_x = 0.72$ and $p_y = 0.55$. When an element of the vector y_obs is equal to zero, the respective value of the response variable is considered missing. The same idea applies to x_obs and the endogenous variable.

A continuous time variable t is simulated, in order to reproduce e.g., the variable age. For each patient, a sequence of 12 values, generated between 3.0 and 27.0, is randomly selected. Finally, the id variable is created to indicate the subjects. We considered only $N = 65$ subjects, so as to reproduce as accurately as possible the data used in Section 5.

In this simulation study, we will consider one endogenous variable x and one response variable y, which we will generate several times according to the JMM (Section 4.1.1) and the JSM (Section 4.1.2). So then the final datasets will have N_{obs} rows and six columns, with shape as in Table 4.1. To simulate the random effects and the error terms for x and y, it is necessary to define the variance-covariance matrices, respectively denoted as $\mathbf{D}_x$ and $\mathbf{\Sigma}_x$, $\mathbf{D}_y$ and $\mathbf{\Sigma}_y$. As said before, one of the goals is to study the performance of INLA

Table 4.1: Shape of the simulated datasets

id	t	y_obs	x_obs	y	x
1	4.4	1	1	$y_{1,1}$	$x_{1,1}$
⋮	⋮	⋮	⋮	⋮	⋮
1	25.8	1	0	$y_{1,12}$	NA
2	3.0	0	0	NA	NA
⋮	⋮	⋮	⋮	⋮	⋮
2	23.1	0	1	NA	$x_{2,12}$
⋮	⋮	⋮	⋮	⋮	⋮
N	4.8	1	1	$y_{N,1}$	$x_{N,1}$
⋮	⋮	⋮	⋮	⋮	⋮
N	25.3	0	1	NA	$x_{N,12}$

and joint models in the presence or absence of small values of variance and covariance. Thus, for each matrix, two different values are assumed and the results coming from them are compared in Section 4.2. Defining $\mathbf{\Sigma}_x = \sigma_x^2 I$ and $\mathbf{\Sigma}_y = \sigma_y^2 I$, the following matrices will be used for the two cases:

$$\begin{array}{ll}
\text{(a) Large values:} & \text{(b) Small values:} \\
\mathbf{D}_x = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix} & \mathbf{D}_x = \begin{bmatrix} 0.14 & -0.013 \\ -0.013 & 0.017 \end{bmatrix} \\
\mathbf{D}_y = \begin{bmatrix} 3 & 2.5 \\ 2.5 & 4 \end{bmatrix} & \mathbf{D}_y = \begin{bmatrix} 0.12 & -0.006 \\ -0.006 & 0.34 \end{bmatrix} \\
\sigma_x = 0.5 & \sigma_x = 0.32 \\
\sigma_y = 0.5 & \sigma_y = 3.58
\end{array} \tag{4.1}$$

Notice that in the following sections, for the sake of simplicity, with the expression "*Large values*" we will refer to the simulation study in which we assume great values for the variances and covariances. Otherwise, we will use the expression "*Small values*".

4.1.1 Simulating data from a JMM

We simulated data from a JMM having the simplest form:

$$\begin{cases} x_i\left(t_{ij}\right) = \left(\beta_0^{(x)} + b_{0,i}^{(x)}\right) + \left(\beta_t^{(x)} + b_{t,i}^{(x)}\right) \cdot t_{ij} + \varepsilon_i^{(x)}\left(t_{ij}\right) \\ y_i\left(t_{ij}\right) = \left(\beta_0^{(y)} + b_{0,i}^{(y)}\right) + \left(\beta_t^{(y)} + b_{t,i}^{(y)}\right) \cdot t_{ij} + \varepsilon_i^{(y)}\left(t_{ij}\right) \end{cases} \tag{4.2}$$

The independent error terms $\boldsymbol{\varepsilon}^{(x)}$ and $\boldsymbol{\varepsilon}^{(y)}$ are simulated separately from a normal distribution with zero mean and variance-covariance matrices as indicated in (4.1a) and (4.1b), depending on the case being considered. The correlated random effects are simulated from a multivariate normal distribution with zero mean and unstructured variance-covariance matrix $\mathbf{D} = \begin{bmatrix} D_x & D_{xy} \\ D_{yx} & D_y \end{bmatrix}$, where $D_{xy} = D_{yx}$ is another matrix of dimension 4 × 4 that has the following values based on the case of study under consideration:

$$\text{Large values: } D_{xy} = D_{yx} = \begin{bmatrix} 1.75 & 1.6 \\ 2 & 2.5 \end{bmatrix}; \tag{4.3}$$

$$\text{Small values: } D_{xy} = D_{yx} = \begin{bmatrix} 0.012 & -0.108 \\ -0.0005 & 0.021 \end{bmatrix}. \tag{4.4}$$

Even the coefficients of the fixed effects change depending on which case is being examined. Whit large values of variances and covariances, e.i., matrices (4.1a) and (4.3), the vector $\boldsymbol{\beta} = (\beta_0^{(x)}, \beta_t^{(x)}, \beta_0^{(y)}, \beta_t^{(y)})$ of the fixed parameters is assumed of the form (4.5a). In contrast, when small variances and covariances are assumed, as in (4.1b) and (4.4), the fixed effects coefficients are equal to (4.5b):

$$\begin{array}{cc} \text{(a) Large values:} & \text{(b) Small values:} \\ \boldsymbol{\beta} = \begin{bmatrix} 5 \\ 1 \\ 6 \\ 2.2 \end{bmatrix} & \boldsymbol{\beta} = \begin{bmatrix} 10.086 \\ -0.128 \\ 53.240 \\ -1.974 \end{bmatrix} \end{array} \tag{4.5}$$

Based on all this information, it is also possible to derive the actual value of the association coefficient over time (see Section 3.2.1), between the endogenous variable x and the outcome y. Considering the model (4.2), it is defined as:

$$\beta_x^{jmm}(t) = \frac{\sigma_{(x,0),(y,0)} + t\sigma_{(y,t),(x,0)} + t\sigma_{(y,0),(x,t)} + t^2\sigma_{(x,t),(y,t)}}{\sigma^2_{x,0} + t^2\sigma^2_{x,t} + 2t\sigma_{(x,t),(x,0)} + \sigma^2_{\varepsilon,x}}. \tag{4.6}$$

As is simple to notice, the value of $\beta_x^{jmm}(t)$ is closely related to the variance and covariance values. Thus, we expect that the worse the estimation of the elements of variance-covariance matrix $\mathbf{D}$ is, the less the estimated association coefficients will approach the actual values (4.6).

4.1.2 Simulating data from a JSM

Simulating data according to the JSM proceeds in a very similar manner. Here too, we consider the simplest structure for the model:

$$\begin{cases} m_i\left(t_{ij}\right) = \left(\beta_0^{(x)} + b_{0,i}^{(x)}\right) + \left(\beta_t^{(x)} + b_{t,i}^{(x)}\right) \cdot t_{ij} \\ x_i\left(t_{ij}\right) = m_i\left(t_{ij}\right) + \varepsilon_i^{(x)}\left(t_{ij}\right) \\ y_i\left(t_{ij}\right) = \gamma \cdot m_i\left(t_{ij}\right) + \left(\beta_0^{(y)} + b_{0,i}^{(y)}\right) + \left(\beta_t^{(y)} + b_{t,i}^{(y)}\right) \cdot t_{ij} + \varepsilon_i^{(y)}\left(t_{ij}\right) = \\ \qquad = \left(\beta_0^{(y')} + b_{0,i}^{(y')}\right) + \left(\beta_t^{(y')} + b_{t,i}^{(y')}\right) \cdot t_{ij} + \varepsilon_i^{(y)}\left(t_{ij}\right) \end{cases} \tag{4.7}$$

Notes that the vector time t is assumed to be the same in both the linear predictors, so the expression for the outcome is rewritten based on the reparametrization (3.9) described in Section 3.3. By doing this, the comparison with the data simulated according to the JMM is easier. The new fixed effects coefficients for the outcome are explicated as $\beta_0^{(y')} = \gamma\beta_0^{(x)} + \beta_0^{(y)}$ and $\beta_t^{(y')} = \gamma\beta_t^{(x)} + \beta_t^{(y)}$. The vector $\boldsymbol{\beta} = (\beta_0^{(x)}, \beta_t^{(x)}, \beta_0^{(y)}, \beta_t^{(y)})$ in the JSM is then defined by replacing the reparametrized parameter $\beta_0^{(y')}$ and $\beta_t^{(y')}$ with the value in (4.5). Moreover, the scaled parameter γ is assumed to be equal to 1.2 when the variances are supposed big, and equal to 1.30 otherwise. Thus, the fixed effects coefficients, in the scenario with small and big variances and covariances, are:

$$\text{(a) Large values:} \quad \boldsymbol{\beta} = \begin{bmatrix} 5 \\ 1 \\ 0 \\ 1 \end{bmatrix} \qquad \text{(b) Small values:} \quad \boldsymbol{\beta} = \begin{bmatrix} 10.086 \\ -0.128 \\ 40.1282 \\ -1.8076 \end{bmatrix} \tag{4.8}$$

The independent error terms $\boldsymbol{\varepsilon}^{(x)}$ and $\boldsymbol{\varepsilon}^{(y)}$ are simulated separately from a normal distribution with zero mean and variance-covariance matrices as indicated in (4.1a) and (4.1b), depending on the case being considered. The correlated random effects are simulated from two bivariate normal distributions, one for the random effects of x and the other one for the random effects of y, both with zero mean and variance-covariance matrix respectively equal to D_x and D_y.

Finally, the association coefficient is simply derived according to equation (3.11):

$$\beta_x^{jsm}(t) = \gamma \left[\frac{\sigma_{x,0}^2 + t^2\sigma_{x,t}^2 + 2t\sigma_{(x,t),(x,0)}}{\sigma_{x,0}^2 + t^2\sigma_{x,t}^2 + 2t\sigma_{(x,t),(x,0)} + \sigma_{\varepsilon,x}^2} \right]. \tag{4.9}$$

4.2 Results

We performed the same analysis twice, one with datasets having large variances of the random effects, the other one with the same amount of samples but having data with small variances and covariances. In the following sections, we will discuss the results of the two scenarios, presenting also possible solutions to overcome the issues encountered.

4.2.1 Comparison between models with large variances

For every simulated dataset, we compute the association coefficients and the methods of evaluation as the marginal likelihood, the information-based quantities WAIC and DIC (see Section 1.4.2) or the predictive measures CPO and PIT (see Section 1.4.3). We will use them to compare the results and evaluate the performance of the different models.

Evaluation of the association coefficient estimates

We examined the relationship between the endogenous covariate and the outcome, evaluating the estimation of the association coefficient over time. In Figure 4.1, are shown the time trajectories of the association coefficients calculated in either the LMM, the JMM and the JSM. For each time point, the mean of the M association coefficients is plotted, with the corresponding credible intervals.

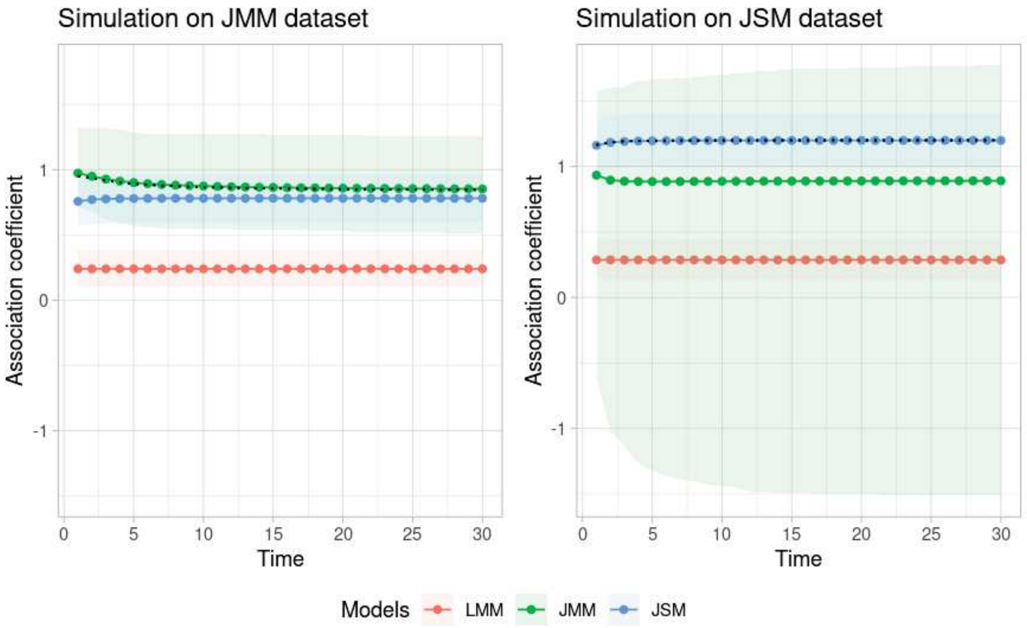


Figure 4.1: Trajectories of the estimated association coefficients and the actual values (black dotted line) when data are simulated according to either JMM and JSM.

On the left, are shown the results from the data simulated according to the JMM. We observed that the values of the JMM association coefficient are perfectly following the actual value of β_x^{jmm} (black dotted line), as we expect. The JSM seems to perform quite well too. Even though the mean values of the association coefficients over time are not exactly equal to the actual values, the credible intervals include them. The LMM, instead, fails to show any association.

The plot on the right in Figure 4.1, show the coefficient trajectories estimated from data simulated according to the JSM. The LMM, again, is not able to reach the actual value β_x^{jsm} (black dotted line): although the mean values are not far off from the actual values, the credible intervals are extremely narrow and do not include sample estimates of the association coefficient. Even here, both JSM and JMM fit quite well. However, in both plots, we notice that the JMM is less robust. The credible intervals, especially in the second plot, are indeed wider than the JSM credible intervals. This suggests that the M values estimated by the JMMs are pretty different from each other and are not actually always near the mean value. This may be due to the fact that in the JMM the number of parameters to be estimated is greater than in the JSM. In the JMM with an unstructured variance-covariance matrix $\mathbf{D}$ we have to estimate 12 parameters, besides the fixed effects coefficients. In the JSM, instead, only 8. This may not seem like such a big difference, but for INLA it is relevant, especially with a very low number of observations in which there is not enough information. We will explore this topic in more detail in the next Section 4.2.2.

Thus, INLA seems to be a good method of estimation and the joint models lead to reasonable conclusions. However, it seems that the estimations obtained by the JSM are quite accurate in every situation, unlike the JMM which is not robust and seems that its results are extremely sensitive to small departures from the model's assumption.

Notice that in the two joint models, the association coefficient tends to increase or decrease over time till it reaches a limit value:

$$\lim_{t\to\infty} \beta_x^{jmm}(t) = \frac{\sigma_{(y,t),(x,t)}}{\sigma_{x,t}^2} = 0.83 \tag{4.10}$$

$$\lim_{t\to\infty} \beta_x^{jsm}(t) = \gamma = 1.2 \tag{4.11}$$

In this simulation study, the association coefficients reach the limits soon.

Evaluation of the goodness of fit

The goodness of fit of the three models is evaluated too. In order to compare them, we compute quantities described in Section 1.4. The mean of the quantities CPO, WAIC, and DIC over the M datasets simulated using the JMM are shown in Table 4.2. In Table 4.3 are reported the same measures but with data generated from JSM. Notice that with measures of goodness, we arrive at the same conclusions as above. In both situations seem that the best model in terms of fitting and predictive performance is the joint scaled model. The worst, as we expected, is the linear mixed model. The marginal log-likelihood (ML) is also added to derive the Bayes Factor (1.4.1) and compare the two joint models. Because linear mixed models only have one likelihood, they cannot be compared, based on ML, to the other two models, which include outcome and covariate likelihoods. Indeed, the values of the marginal log-likelihood of the LMM are way smaller than the other. From values on Table 4.2, the logarithm of the Bayes Factor (see equation 1.13) is 14.62, indicating strong evidence for JSM even though the data are generated from the JMM. We arrive at the same conclusion in Table 4.3, where the logarithm of the Bayes Factor is 68.86. The results in this case are consistent with the simulation study.

	LMM	JMM	JSM
ML	-721.28	-1562.83	-1548.20
DIC	757.19	749.42	741.42
WAIC	759.27	752.25	743.05
CPO	394.11	389.01	386.29

Table 4.2: Goodness of fit measures with data from JMM

	LMM	JMM	JSM
ML	-770.45	-1648.02	-1579.16
DIC	769.41	778.98	742.48
WAIC	770.89	783.71	744.30
CPO	403.05	406.70	387.88

Table 4.3: Goodness of fit measures with data from JSM

4.2.2 Comparison between models with small variances

The simulation study described so far has also been applied to data in which the variances and covariances of the random effects are really small (data matrices are 4.1b and 4.4). Below the results are discussed.

Evaluation of the association coefficient estimates

In Figure 4.2 we show the trajectories of the association coefficients over time. Differently from before, here, the estimates are not so good. In the simulation where data are generated according to the JMM, not even the mean values of the JMM estimates reach the actual value β_x^{jmm} (black dotted line). The trajectory seems good in the early time points, but already after time 5 the values start to be way smaller than the real ones. The trend is similar but it seems that after 30 time points the coefficient estimations have not yet reached a limit and still tend to increase. The credible intervals contain the actual values throughout the timeline expect for the end, in which the estimated values coincide with the upper limit of the interval. The time trend of the JSM estimates is really different from the actual one. It rises slightly at first, but quickly reaches a limit, starting to include the actual values of the coefficient in the credible intervals from the middle of the timeline. Thus, for the first half of the timeline, the JMM provides the best estimates; for the second half, we rely more on the JSM values.

In the simulation where data are generated according to the JSM (plot on the right of Figure 4.2), the results are not better. The association coefficient estimates from the JMM do not reach the real values and also do not include them in the credible intervals. The JSM estimates are nearest to the actual values but are still biased.

In the scenario that we are considering, with small variances and covariances, the LMM estimate is nearest to the values of the other two models than before. This is probably because the small variances also indicate a less strong dependence between the endogenous variable and the outcome and a minor need to model the association through joint models. However, they all lead us to biased results. Finally, notice that, unlike in Section 4.2.1, the credible intervals of either LMM and JMM estimates are wide, suggesting that these models are not robust and their results are sensitive to modifica-

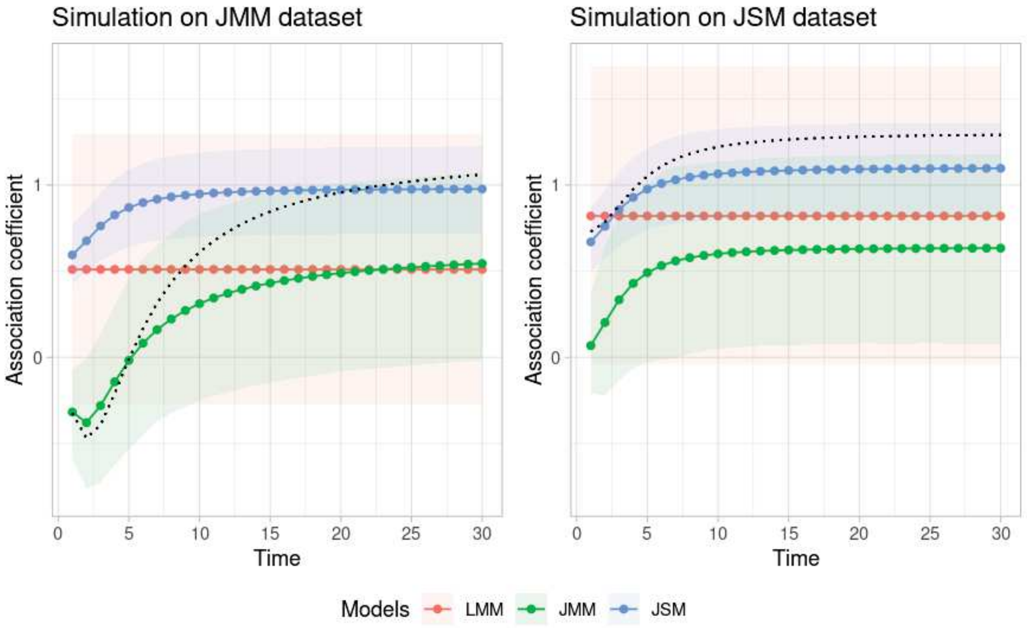


Figure 4.2: Trajectories of the estimated association coefficients and the actual values (black dotted line) when data are simulated according to either JMM and JSM.

tions of the data.

Evaluation of the variance and covariance estimates

Both the association coefficient estimates β_x^{jmm} (3.3) and β_x^{jsm} (3.11) are functions of the variances and covariances. Therefore, biased coefficient estimations are certainly due to poor estimation of variance-covariance matrices. Let's then study a little further the estimation provided by INLA. When the estimations of the variances are near the boundary, thus basically zero, it can happen that INLA fails to estimate correctly the posterior precisions. In Figure 4.3 we show two examples of poor estimation of the precision. The two posterior densities are not continuous, for values near zero, they go to infinity and the probability to observe all the others is practically zero. This situation does not allow us to calculate the distribution of the inverse function, i.e., of the variances. Even if we decide to work in terms of precisions, the mean value of these distributions does not give us proper information about them. Note that this happens in both the JSM and JMM, thus, the estimations of the association coefficients are not very good anyway (see Figure 4.2). We tried a few methods to overcome the inability of INLA to fit

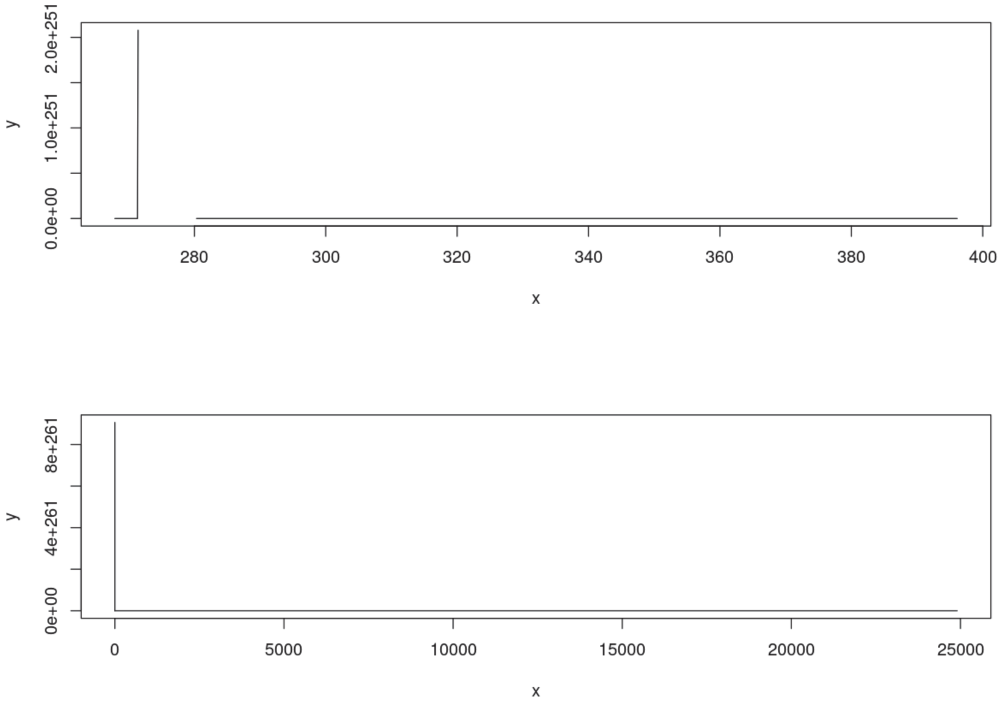


Figure 4.3: Posterior densities of two precisions estimated by INLA.

correctly the variance and covariance elements. Since we are working in a Bayesian framework, the issues may be caused by the prior selection. So far, we have simply used the default options, thus the uninformative priors. Although, in presence of small variances, may be better to choose a more informative prior for the hyperparameters. From the above sections, we saw that the worst performance is in the joint mixed model. Hence, in the following, we will describe the process of selection of a prior only for a JMM having an unstructured variance-covariance matrix of the random effects. Regard-less, the methods described here can also be applied in the case of different structures of the variance-covariance matrix or with JSM.

Let's remember that INLA works with the precision matrix, and not with the variance-covariance matrix. As usual in Bayesian statistics, the prior used for the precision matrix is a Wishart distribution (Zhang 2021). It is very popular because it is the conjugate to multivariate normal data, thus, we will not select another distribution but we will try to work on its parameter. The Wishart distribution of a precision matrix $\mathbf{W}$, with r degrees of freedom and scale matrix $\mathbf{R}$, is defined as:

$$f_{\mathbf{W}}(\mathbf{W}) = \frac{1}{2^{rp/2}|\mathbf{R}|^{r/2}\Gamma_p\left(\frac{r}{2}\right)}|\mathbf{W}|^{(r-p-1)/2}e^{-\frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\mathbf{W}\right)},$$

where p is the dimension of the multivariate normal. The mean and the variance of this distribution are:

$$E[\mathbf{W}] = r\mathbf{R}, \quad Var(w_{ij}) = r(r_{ij}^2 + r_{ii}r_{jj}), \tag{4.12}$$

where r_{ij} is the element of position (i, j) in the scaled matrix.

Following the Bayes theorem, the posterior distribution will be again a Wishart distribution having as a new number of degrees of freedom $r_1 = n + r$, where n is the sample size, and new scale matrix $\mathbf{R}_1 = (n\mathbf{S} + \mathbf{R}^{-1})^{-1}$, where $\mathbf{S}$ is the sample variance-covariance matrix. Using equations (4.12), the posterior mean and variance are simply derived. Notice that assuming a prior Wishart distribution for the precision matrix $\mathbf{W}$ is equivalent to assuming an Inverse-Wishart prior distribution for the variance-covariance matrix $\mathbf{D}$, with inverse scaling matrix: $\mathbf{D} \sim IW(r, \mathbf{R}^{-1})$. From this follows that the posterior means are:

$$\begin{aligned} E(\mathbf{W} \mid y) &= (n + r)\left(n\mathbf{S} + \mathbf{R}^{-1}\right)^{-1}, \\ E(\mathbf{D} \mid y) &= \frac{n\mathbf{S} + \mathbf{R}^{-1}}{n + r - p - 1} = \\ &= \frac{n}{n + r - p - 1}\mathbf{S} + \left(1 - \frac{n}{n + r - p - 1}\right)\frac{\mathbf{R}^{-1}}{r - p - 1}. \end{aligned} \tag{4.13}$$

This relation between the two posterior means can be useful to understand better how to select the best Wishart prior. Note in fact that the posterior mean of the variance-covariance matrix can be rewritten as the weighted average of the sample covariance matrix $\mathbf{D}$ and the prior mean $\mathbf{R}^{-1}/(r-p-1)$. Thus, when the sample size is large, the posterior mean tends to be closer to the sample mean given fixed r and p. In our case, the sample size is small, so the choice of prior has a greater influence on the posterior because there is not enough information from the data.

Considering 4 random effects, two random intercepts and two random slopes, all correlated with each other, the INLA default prior distribution for the precision matrix is $\mathbf{W} \sim Wishart_4(r = 11, \mathbf{R} = I_4)$. Based on this information, the parameters can be changed in the following way:

- changing the scale matrix while the degrees of freedom are assuming always equal to 11. With $\mathbf{R}$ as the identity matrix, the mean prior is equal to an identity scaled matrix $11I_4$ and the variance prior is an identity scaled matrix $22I_4$. The latest is too wide and often leads to biased estimation of the precision parameters (see Figure 4.3). An alternative idea could be to replace the scale matrix with a matrix that allows us to obtain a prior mean closer to the sample precision matrix $\mathbf{S}^{-1}$. However, the scale matrix choice affects the variance prior as well. Especially when we have small variances, i.e., big precisions, we risk imposing a prior with a mean closer to the true value but a variance even wider than the default option of INLA.

- changing the degree of freedom r while we are assuming the scale matrix is fixed as $\mathbf{R} = I_4$. Smaller are the degrees of freedom, smaller are the prior means and variances, as well as the posterior means, given all the other elements fixed. Indeed, both are defined as the product between the degrees of freedom and a quantity that depends on the scaled matrix. Furthermore, the smaller the degrees of freedom, the more information from the data is relevant and the posterior means are close to the sample precision matrix $1/\mathbf{S}$.

- changing both the degree of freedom r and the scaled matrix $\mathbf{R}$, maintaining the same prior mean, as suggested by Zhang 2021. Let's consider a generic choice for $\mathbf{R}$. To keep the prior mean equal to $\mathbf{R}$, we need to assume a Wishart distribution with scaled matrix $\mathbf{R}/r$ and r degree of freedom. This is the best option but also the most difficult to apply. Even if we choose to put the scaled matrix equal to the sample precision matrix $\mathbf{R} = \mathbf{S}^{-1}$, the selection of the degrees of freedom is substantial to keep control of the prior variance too.

In our simulations, it seems that none of these changes really improved the fitting of the models. However, only a few values have been tried, so the next research development could be to try new values. But the implementation is not straightforward and the only way to proceed is by doing a sensitivity analysis.

Notice that for all these tests is required to know the sample variance-covariance matrix $\mathbf{S}$. Although, with real data, we do not know the variance-covariance matrix of random variables as the random effects. Thus, we need to derive a priori the sample estimation of the matrix. An option may be to start with simpler models in order to obtain some information about the values of the variances and covariances. We can estimate a linear mixed model separately for the endogenous variable x and the outcome y, just to have an idea of the values of the variances of the random effects and their correlation. In the case of JSM, we can simply use the two outputs as the sample estimation of the variance-covariance matrices $\mathbf{D}_x$ and $\mathbf{D}_y$, and change the Wishart distribution based on them. What is missing for the JSM, instead, are the values of the covariances between the random effects of the two variables. We can start by assigning very small values, and then gradually increasing them. Finally, as the assigned values change, we change the parameters of the Wishart distribution.

However, this evaluation process takes time and is not the main goal of the thesis. Thus, another easier possible solution to derive a better estimation of the association coefficient could be to simply reduce the number of parameters in the JMM. Especially when the number of observations is small, this simple operation can reduce the bias and improve the goodness of fit.

For instance, instead of assuming an unstructured correlation matrix between the random effects, a pairwise variance-covariance matrix can be selected, or either considering only random intercepts, or only random intercepts for the endogenous variable and both random intercepts and random slopes for the outcome, or other combinations. For more discussion about these options sees chapter 5. Due to time limitations, these additional models were not applied to the simulated data.

Chapter 5

Application

In this chapter, we shall apply on real dataset all the models described above. The data comes from a study on children diagnosed with the rare disorder Duchenne Muscular Dystrophy, who have visited the Leiden University Medical Center (LUMC) in the last decades. The analysis was performed anonymously, i.e., without releasing crucial information about the exposures measured.

5.1 Presentation of the dataset

Duchenne Muscular Dystrophy (DMD) is the most frequent and best known of the childhood muscular dystrophies. It is a genetic disorder characterized by progressive muscle degeneration and weakness due to the alterations of a protein called dystrophin that helps keep muscle cells intact. The pathology is localized on the X chromosome, therefore, it affects males more often. The first symptoms start to appear at early ages, when the children are two or three years old. Then a degenerative progression begins. The shoulders and hip and thigh muscles begin to lose strength as early as five to seven years old, and this continues until the early teens, when the teenager starts to use full-time a wheelchair. Respiratory problems appear in their teens and twenties as well. More often ventilation support is required, till the increase of cardiac dysfunction leads to heart failure and death. So generally patients with DMD usually do not survive beyond their teen years.

There is not yet a cure or treatment to stop muscle degeneration, so monitoring disease progression is crucial for developing and assessing novel therapies. The disease progression, in every study of the DMD, is monitored via function scores, for instance, the number of meters the children can walk during a six minutes walk test. These scores are collected at each patient visit and provide the doctor with information about the patient's longitudinal progression. However, the goal of most recent research is to identify biomarkers that rely on blood and urine samples. They are easier and faster to collect and are non-invasive. To do so, it is necessary to determine which blood biomarkers are associated with the functional scores, so as to replace them.

In particular, in the dataset available, the aim is to study the association between seven thousand proteins and the *PUL 2.0* functional score. It is the total score of the performance of the upper limb test (see Mayhew et al. 2019). It collects information on several shoulder and hand movement tests, each of which is assigned a score. If a patient is able to properly move the upper body, the total score is equal to 42. *PUL 2.0* is thus a numerical variable, with a range $[0, 42]$. The proteins are in turn collected over time, not always at the same time points of the response variable, and depend on the history of the outcome. Thus, they are endogenous time-varying covariates (see Section 3.1).

The dataset contains information about a very low number of patients due to the fact that DMD is a rare disease. We will work on data from only 65 patients, each of them with a very different number of observations (see Figure 5.1). Some patients were visited only two times, others 13 times, with an average number of observations for each patient equals to 8. There are patients from 3 to 27 years old, with a mean age equal to 12.

The primary goal of this thesis is to use Joint Models to assess the relationship between endogenous time-varying proteins and functional score

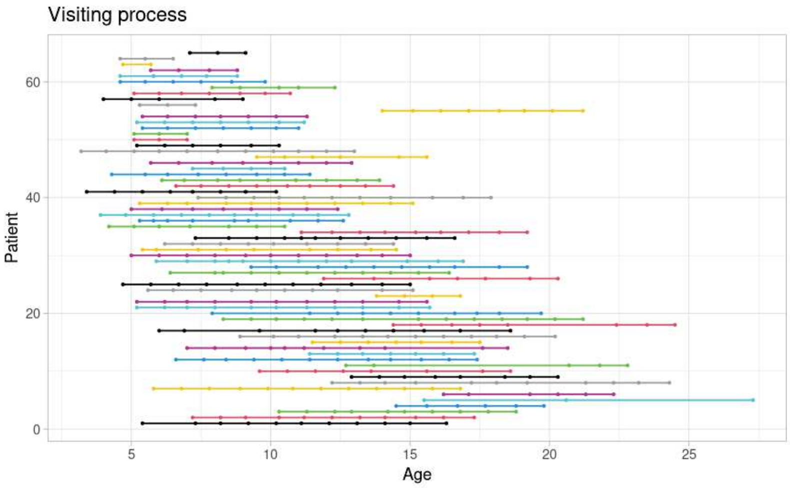


Figure 5.1: Visiting process for each patient

(described in Chapter 3). We will not estimate the association between the outcome and more than one protein at the time, as this leads to new issues that are beyond the scope of the thesis. In the following sections, we will not consider all seven thousand proteins, but rather only two of them to describe the estimation process. For issues regarding privacy, we will refer to them as *protein1* and *protein2* without mentioning their names.

5.2 Exploring the mean structure

Before proceeding with the model estimation, we examined the time progression of each variable (the outcome and the two proteins) in order to determine which form of the linear predictor was the most appropriate. First of all, spaghetti plots were used to examine each patient's longitudinal progression. The graphs for every variable are shown in Figure 5.2. The patients' trajectories appear to be very different from one another, especially for the outcome, and the between-subject variability is considerable. They differ not only in terms of variable level, but also in terms of line steepness. Based on that, for all the trends considered, a mixed model was implemented. In particular, at each step, a model with only random intercepts was compared to a model with both random intercepts and random slopes. In the end, the best of all the trends were selected and used on the Joint Mixed models.

The trend options used are:

- Linear trend over time. The general form of the linear predictor is:

$$\eta_{ij} = \beta_0 + \beta_1 \cdot x_{ij} + \beta_2 \cdot t_{ij} + \varepsilon_{ij},$$

  where other fixed and random effects and interactions can be added. It is the simplest mean structure, but it makes very strict assumptions about the progression in time;

- Quadratic trend over time, used when the form of the trajectories is more similar to a parabola. Mathematically speaking:

$$\eta_{ij} = \beta_0 + \beta_1 \cdot x_{ij} + \beta_2 \cdot t_{ij} + \beta_3 \cdot t_{ij}^2 + \varepsilon_{ij};$$

- Cubic trend over time. In this case, the trend of the mean is described by a third-degree polynomial:

$$\eta_{ij} = \beta_0 + \beta_1 \cdot x_{ij} + \beta_2 \cdot t_{ij} + \beta_3 \cdot t_{ij}^2 + \beta_4 \cdot t_{ij}^3 + \varepsilon_{ij}.$$

  These last two mean structures relax the assumption of the linear trend but increase the complexity of the model and its interpretation. Additionally, they are quite sensitive to outliers;

- Piecewise linear trend over time. The idea is to subdivide the time axis into intervals, on which a linear progression is assumed:

$$\eta_{ij} = \beta_0 + \beta_1 \cdot x_{ij} + \beta_2 \cdot t_{ij} + \beta_3 \cdot (t_{ij} - t^*)_+ + \varepsilon_{ij},$$

  where $(t_{ij} - t^*)_+ = t_{ij} - t^*$ when $t_{ij} > t^*$ and it is equal to 0 otherwise. The choice of the number and the location of the breakpoints t^*, generally called knots, can be done following different strategies. Here we used the information obtained by estimating a Multivariate

Adaptive Regression Splines (MARS), in which the piecewise linear basis functions are selected in a similar way as a forward stepwise linear regression (Hastie, Tibshirani, and Friedman 2009). Compared to polynomial trends, piecewise linear functions are more flexible and easier to interpret;

- Natural cubic splines trend over time. They are less sensitive to outliers and are smooth alternatives to the piecewise linear trend, so they estimate pretty well any kind of function. Here too, the time axis is divided into intervals. Beyond the boundary knots, a linear trend is assumed. In the internal intervals, instead, a cubic trend is presupposed. Although we obtain a more flexible model, we lose in interpretability. The model is in fact defined as a linear combination of K basis functions:

$$\eta_{ij} = \sum_{j=1}^{p} \sum_{m=1}^{M_j} \beta_{jm} h_{jm}\left(x_j\right),$$

where h_m are the basis functions defined as follows:

$$h_1(t) = 1, \quad h_2(t) = t, \quad h_{k+2}(t) = d_k(t) - d_{k-1}(t)$$
$$d_k(X) = \frac{(t-\xi_k)_+^3 - (t-\xi_K)_+^3}{\xi_K - \xi_k}.$$

However, in this study, the only parameters that we need to interpret are the association coefficients between the endogenous variable and the outcome, and they are derived analytically using variances and covariances. Hence, no fixed effects are really taken into consideration for the final comments.

Note that not all the trends were used for every variable, but only those deemed most appropriate based on the plots in Figure 5.2. The pink line represents the loess curve. Even if it is a local approximation, it may be useful to provide a starting point for the analysis. The further it deviates from a linear trend, the more likely it is that the best mean trend over time is not linear.

In conclusion, we chose a natural cubic splines trend over time for the *PUL*

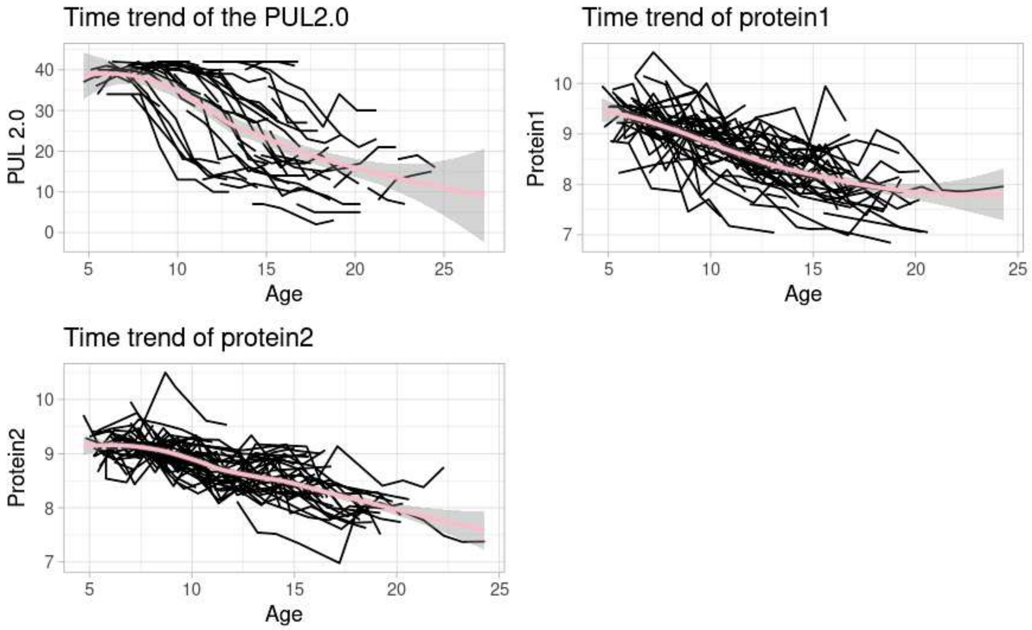


Figure 5.2: Trajectories per patient of the three variables over time.

2.0 variable and *protein1*. For the *PUL 2.0* outcome, we selected a cubic spline with two knots at age 8 and 18.5. For the *protein1* a cubic spline with only one knot at age 14. For *protein2*, instead, we chose a linear trend. Furthermore, for all of them, we preferred the models with both random intercepts and random slopes only on the linear effect of time. The selection was made based on the values of the information criteria WAIC and DIC (described in Section 1.4.2) and the marginal likelihoods. The predictive measures CPO and PIT (see Section 1.4.3) were considered too, but they provide a less accurate evaluation of the goodness of fit of the model due to the dependence between the observations.

5.3 Modelling

Once the trend over time was defined for each variable, we estimated the association between the proteins and the PUL functional score through the joint mixed model (JMM) and the joint scaled model (JSM). As has already been specified, hereafter the two proteins are studied separately. Hence, for each type of joint model, two models were estimated.

5.3.1 Joint Mixed Model

The joint mixed models measure the association between the endogenous variable selected and the functional score via the random effects variance-covariance matrix **D**. As we already have selected the best mean trend over time for each variable, the second step of the model building is to choose the appropriate structure of the matrix **D**. The strategy that we followed is:

- consider the most elaborate form for the matrix, the unstructured matrix (3.6), in which all the random effects are correlated;

- evaluate the significance of covariance values on the estimated matrix. The matrix form is simplified if the null hypothesis of values equal to zero is not rejected, which means that their credible intervals contain zero. For instance, a pairwise correlation matrix (3.7) is estimated or a new random effect structure is considered e.g., only random intercepts, one random slope instead of two, etc.; Furthermore, the model is evaluated on the basis of the measures described in Section 1.4;

- once again, in the newly selected matrix, the significance of the element outside the main diagonal is assessed. If they are still not significant, the protein's model and the outcome's model are estimated separately.

Note that if the variance and covariance values are still significant but really near zero, it is suggested to simplify the matrix form anyway in order to reduce the number of parameters to estimate.

In the end, for both the JMM, a pairwise correlation matrix was used.

The two models can be defined, with the same general structure, as follows:

$$\begin{cases} x_i\left(s_{ij}\right) = X_i\boldsymbol{\beta} + b_{0,i}^{(x)} + b_{t,i}^{(x)} \cdot s_{ij} + \varepsilon_i^{(x)}\left(s_{ij}\right) \\ y_i\left(t_{ij}\right) = \beta_0^{(y)} + \beta_1^{(y)} \cdot h_1(t_{ij}) + \beta_2^{(y)} \cdot h_2(t_{ij}) + \beta_3^{(y)} \cdot h_3(t_{ij}) + \\ \qquad\qquad\qquad\qquad + b_{0,i}^{(y)} + b_{t,i}^{(y)} \cdot t_{ij} + \varepsilon_i^{(y)}\left(t_{ij}\right) \end{cases} \tag{5.1}$$

where

$$\mathbf{b}_{0,i} = \begin{bmatrix} b_{0,i}^{(x)} \\ b_{0,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_2\left(0, \begin{bmatrix} \sigma_{x,0}^2 & \sigma_{(x,0),(y,0)} \\ \sigma_{(x,0),(y,0)} & \sigma_{y,0}^2 \end{bmatrix}\right),$$

$$\mathbf{b}_{t,i} = \begin{bmatrix} b_{t,i}^{(x)} \\ b_{t,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_2\left(0, \begin{bmatrix} \sigma_{x,t}^2 & \sigma_{(x,t),(y,t)} \\ \sigma_{(x,t),(y,t)} & \sigma_{y,t}^2 \end{bmatrix}\right),$$

$$\begin{bmatrix} \varepsilon_i^{(x)} \\ \varepsilon_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2n_i}\left(\mathbf{0}, \begin{bmatrix} \sigma_{\varepsilon,x}^2 \mathbf{I}_{n_i} & 0 \\ 0 & \sigma_{\varepsilon,y}^2 \mathbf{I}_{n_i} \end{bmatrix}\right).$$

The fixed effect part of the endogenous covariate $X_i\boldsymbol{\beta}$ is differently defined for *protein1* (5.2) and *protein2* (5.3):

$$X_i\boldsymbol{\beta}^{(p_1)} = \beta_0^{(x)} + \beta_1^{(x)} \cdot h_1(s_{ij}) + \beta_2^{(x)} \cdot h_2(s_{ij}) \tag{5.2}$$

$$X_i\boldsymbol{\beta}^{(p_2)} = \beta_0^{(x)} + \beta_1^{(x)} \cdot s_{ij} \tag{5.3}$$

Based on this model, the association coefficient can be then derived as follows:

$$\beta_x^{jmm}(t) = \frac{\mathrm{Cov}(y,x)(t)}{\mathrm{Var}(x)(t)} = \frac{\sigma_{(x,0),(y,0)} + t^2\sigma_{(x,t),(y,t)}}{\sigma_{x,0}^2 + t^2\sigma_{x,t}^2 + \sigma_{\varepsilon,x}^2}. \tag{5.4}$$

Note that this structure of β_x^{jmm} applies for both the proteins independently on the mean structure of the models. The association coefficient in the JMM depends only on the structure of the random effects and on the form of their variance-covariance matrix.

5.3.2 Joint Scaled Model

In the joint scaled model, the association between the endogenous variable selected and the functional score is measured via a scaled factor γ. No procedure is therefore required for the selection of the best variance-covariance matrices for the random effects, which are directly defined as:

$$\begin{cases} m_i\left(s_{ij}\right) = X_i\boldsymbol{\beta} + b_{0,i}^{(x)} + b_{t,i}^{(x)} \cdot s_{ij} + \varepsilon_i^{(x)}\left(s_{ij}\right) \\ x_i\left(s_{ij}\right) = m_i\left(s_{ij}\right) + \varepsilon_{xi}\left(s_{ij}\right) \\ y_i\left(t_{ij}\right) = \gamma m_i\left(t_{ij}\right) + \beta_0^{(y)} + \beta_1^{(y)} \cdot h_1(t_{ij}) + \beta_2^{(y)} \cdot h_2(t_{ij}) + \beta_3^{(y)} \cdot h_3(t_{ij}) + \\ \qquad + b_{0,i}^{(y)} + b_{t,i}^{(y)} \cdot t_{ij} + \varepsilon_i^{(y)}\left(t_{ij}\right) \end{cases} \tag{5.5}$$

whit

$$\begin{aligned} \mathbf{b}_{xi} &\sim \mathcal{N}\left(\mathbf{0}, \mathbf{D}_x\right), \quad \varepsilon_{xi}\left(s_{ij}\right) \sim \mathcal{N}_{n_i}\left(\mathbf{0}, \sigma^2_{\varepsilon,x}\right) \\ \mathbf{b}_{yi} &\sim \mathcal{N}\left(\mathbf{0}, \mathbf{D}_y\right), \quad \varepsilon_{yi}\left(t_{ij}\right) \sim \mathcal{N}_{n_i}\left(\mathbf{0}, \sigma^2_{\varepsilon,y}\right) \end{aligned}$$

where $X_i\boldsymbol{\beta}$ correspond to (5.2) for *protein1* and to (5.3) for *protein2*. Here, the association coefficient β_x^{jmm} is computed as in equation 4.6.

5.4 Results

5.4.1 Analyzing *protein1*

To understand if there is an association between the functional score *PUL 2.0* and the endogenous variable *protein1*, we have to evaluate the association coefficient. Additionally, in the JMM we need to inspect the covariances between the random effect and the corresponding credible intervals. $\sigma_{(x,0),(y,0)}$ and $\sigma_{(x,t),(y,t)}$ are particular of interest, since they are responsible for the association between the two variables. Instead, in the JSM, the significance of the scaling factor γ must be assessed. It is the element with which the joint scaled model considers the association, as well as the limit reached by the coefficient when time goes to infinity (see equation 5.4). In Figure 5.3 the results of the models (5.1) and (5.5) are shown. Also the credible intervals of the random effects' hyperparameters in JSM are reported, in order to understand if it is possible to consider a more parsimonious model. The association coefficients in both models increase over time following the same pattern. At low ages, the JMM provides an association fairly close to zero, but with increasing age, the coefficient tends to approach one. However, the credible intervals of the JMM estimates include zero, indicating a

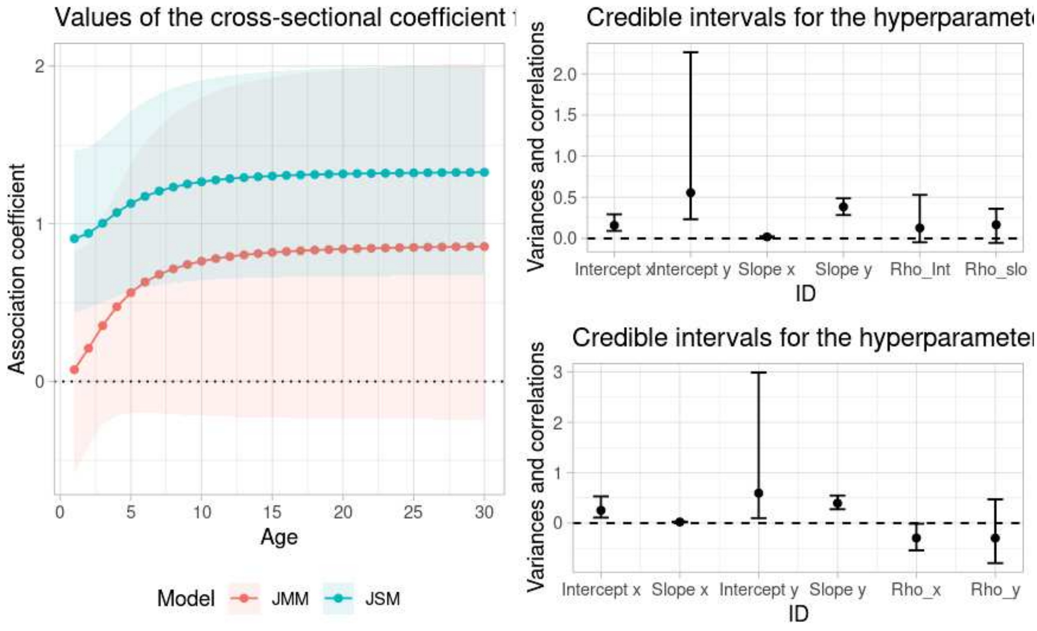


Figure 5.3: Trajectories of the estimated association coefficients and credible intervals of the hyperparameters of the random effects for both JMM and JSM models between the functional score and the endogenous variable *protein1*.

non-significant association, unlike the JSM estimates, which are significantly above zero. Remember that the JMM appeared to be not robust in the simulation study, hence the association coefficient estimate might be distorted. Furthermore, we are dealing with low variances and correlations. Thus, we can not fully trust the results, but the plot can still provide insight into the general behavior of the association.

On the plot in the right top corner of Figure 5.3, the credible intervals (CI) of variances and correlations of the random effects in JMM are shown. It's worth noting that the CIs for both correlations (correlation between random intercepts and correlation between random slopes) have zero inside. Having followed the strategy described in Section 5.3.1, the best way to proceed would be to consider a new model that does not include these terms. Indeed, as shown in Table Table 5.1, the best model is the one that assumes no correlation between random effects of different variables. This conclusion is consistent with the behavior of the association coefficient estimates. Only the marginal likelihood favors the model with an unstructured correlation matrix, which was excluded at the outset in order to simplify the model as much as possible. However, the differences between the measures are negligible. Thus, because our initial goal was to estimate the association between the endogenous covariate and the response variable, we still consider the model with a pairwise variance-covariance matrix to be reliable.

Table 5.1: Comparison between JMM model for *protein1* with an unstructured correlation matrix of the random effects, pairwise correlation matrix and no correlation.

	Unstructured	Pairwise	None
Marginal Likelihood	-1129.68	-1133.65	-1136.95
DIC	1509.97	1508.05	1506.92
WAIC	1506.57	1504.47	1503.62
CPO	760.90	759.94	759.72

Finally, on the plot in the right bottom corner of Figure 5.3 the credible intervals of variances and correlations of the random effects in JSM are shown. In this type of model, the correlations between the random effects of x and y are not used to evaluate the association between the *protein1* and the *PUL 2.0* outcome. Anyway, the association coefficient still depends on the covariances between the random effects of the endogenous variable, so the estimations are affected by the choice of different forms for the random effects of x. Moreover, removing non-significant elements can be relevant to obtain a better model. From our data, it appears that the covariance between the random intercepts and random slopes of the response variable is not significant. However, based on the measures of fitting in Table 5.2, the best model is the one that assumes a correlation between the random effects of the response variable.

Table 5.2: Comparison between a JSM with covariance between random intercepts and slopes of the response variable (covariances), and a JSM without (no-covariances).

	Covariances	No-covariances
Marginal Likelihood	-1152.51	-1188.18
DIC	1504.79	1556.21
WAIC	1502.05	1556.80
CPO	758.84	787.37

5.4.2 Analyzing *protein2*

The same analysis done for *protein1* now is reproduced for *protein2*. In Figure 5.4 the trajectories of the JMM and JSM estimated coefficients are shown with the corresponding credible intervals. This time, the estimated association coefficient of JMM is very close to zero across all ages. In contrast, the JSM values are all around one and the credible intervals do not include zero. Thus, JSM estimates are all significant, unlike the JMM estimates. Once again, in the JMM, the correlations between the random intercepts $\rho_{(x,0),(y,0)}$ and the one between the random slopes $\rho_{(x,t),(y,t)}$, are non-significant since their credible intervals have zero inside (plot on the right up corner of Figure 5.4). This leads us to the same conclusion of the association coefficient. Also the results in Table 5.3 confirm that the best LMM model is the one that assumes no association between *protein2* and *PUL 2.0*. However, the differences in goodness of fit measures are even smaller than before, allowing us to continue to use the pairwise model to estimate and provide information about the association over time. The correlations between the random effects of both the protein and the functional score in the JSM considered are pretty close to zero, and their credible intervals include it as well. Moreover, even the value of the random slopes of x is really near zero, and since we delete the correlation between the random effects of the endogenous variable, the random slopes can be removed too. The new model then has only random intercepts for the covariate and both the random effects for

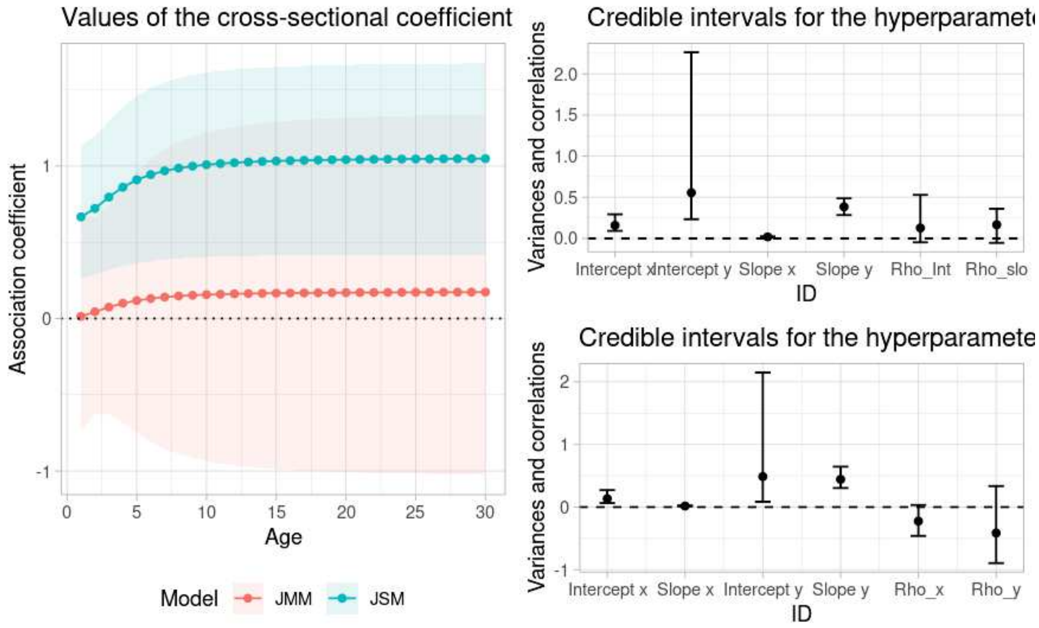


Figure 5.4: Trajectories of the estimated association coefficients and credible intervals of the hyperparameters of the random effects for both JMM and JSM models between the functional score and the endogenous variable *protein2*.

the functional score. Note that this implicates having an JSM association coefficient (3.11) that is no longer time dependent. The variance of the endogenous process is simply defined as the sum of the variance of the random intercept $\sigma^2_{x,0}$ and the variance of the error terms $\sigma^2_{\varepsilon,x}$. Instead, the variance of the linear predictor $m(t_{ij})$ coincides with only the variance of the random intercept $\sigma^2_{x,0}$. Thus, the association coefficient is:

$$\beta_x^{jsm} = \gamma \left[\frac{\text{Var}(m)}{\text{Var}(x)} \right] = \gamma \left[\frac{\sigma^2_{x,0}}{\sigma^2_{x,0} + \sigma^2_{\varepsilon,x}} \right] = 0.66$$

Once again we compare the models based on the measures of goodness described in Section 1.4 and the results are shown in Table 5.4. All of them suggest that the best model is the one that assumes random slopes for the endogenous variable and covariance between the random effects for both the

Table 5.3: Comparison between JMM model for *protein2* with an unstructured correlation matrix of the random effects, pairwise correlation matrix and none correlation.

	Ustructure	Pairwise	None
Marginal Likelihood	-1073.44	-1079.64	-1077.49
DIC	1510.28	1509.14	1506.65
WAIC	1506.70	1505.63	1503.00
CPO	761.20	760.67	759.44

variables, except for the Bayes Factor which is in favor of the model without.

Table 5.4: Comparison between a JSM with covariance between random effects for both the variables (covariance), and a JSM without covariances and random slopes for x (no-covariance).

	Covariances	No-covariances
Marginal Likelihood	-1086.20	-1029.05
DIC	1507.38	1563.62
WAIC	1504.21	1563.71
CPO	759.86	790.75

5.4.3 Comments

We presented results of models in which we assumed a normal likelihood for both the functional score and the endogenous variable. However, the outcome *PUL 2.0* is a score that can assume values only between 0 to 42. Assuming a Gaussian distribution for this variable is therefore formally wrong. The predictions obtained from such a model could, in fact, be projected out of the support $[0, 42]$. To overcome this problem we decide to use a Beta distribution, as described in Section 2.2.1. The beta distribution is generally defined in a range $(0, 1)$, but it is also used in the presence of a variable limited in different intervals. In order to do that, the support of the variable of interest has to be changed. There are several possible transformations but the one that we used is the *Min-Max* transformation $h(y) = \frac{y-min}{max-min}$.

On the transformed functional score then, we implemented a JMM and a JSM having a Beta likelihood for the dependent variable and a Gaussian likelihood for the endogenous variable. According to Table 5.5 and Table 5.6, the best models are those that assume a Beta distribution. The differences in goodness of fit measures are indeed substantial. The only quantities that strongly disagree are the CPOs, but as we have already stated, in the context of mixed models we cannot completely trust them (see Section 1.4.3).

Table 5.5: Comparison between JMM and JSM models for *protein1* with Gaussian and Beta likelihood.

	JMM		JSM	
	Gaussian	Beta	Gaussian	Beta
Marginal Likelihood	-1133.65	-27.07	-1152.51	-37.00
DIC	1508.05	-780.35	1504.79	-843.28
WAIC	1504.47	-779.06	1502.05	-846.05
CPO	759.94	1639.37	758.84	3664.29

Table 5.6: Comparison between JMM and JSM models for *protein2* with Gaussian and Beta likelihood.

	JMM		JSM	
	Gaussian	Beta	Gaussian	Beta
Marginal Likelihood	-1079.64	27.45	-1086.20	26.00
DIC	1509.14	-779.62	1507.38	-841.83
WAIC	1505.63	-778.26	1504.21	-848.63
CPO	760.67	1729.36	759.86	3702.65

However, there are some issues with these new models. First of all, the interpretation is not simple. The regression parameters can only be interpreted in terms of $E[y_{ij}]$ and not in terms of y_{ij}. This leads to a second problem concerning the association coefficient. As we have already discussed in Section 3.2.1 and 3.3.1, it is no longer possible to obtain a closed-form solution for the coefficient when one of the likelihoods is not normal. It is quite challenging to find the joint distribution and then the conditional on x as for the case with two Gaussian likelihoods. What we tried to do then, was to find an approximation of the closed-form formula using as best we could the Bayesian output provided by R-INLA. We need to derive the conditional posterior distribution of y given x and its mean $E(y_{ij} \mid x_{ij})$, defined as:

$$E(y_{ij} \mid x_{ij}) = \int_b \exp\left(x_{ij}^\top \boldsymbol{\beta} + z_{ij}^\top \mathbf{b}_i\right) f\left(\mathbf{b}_i\right) d\mathbf{b}_i,$$

that in the context of mixed models is the mean of the marginal distribution (see equation 2.10), in which the random effects are integrated out. Once we have it, the idea is to compare the marginal means for a unit change of x, thus $E(y_{ij} \mid x_{ij}+1) - E(y_{ij} \mid x_{ij})$, as we did when we derive the analytical form of the association coefficient in chapter 3. However, this software furnishes as output only univariate marginal posterior distributions of the latent field (1.6) and the hyperparameters (1.7). The only conditional distribution provides are the predictive distributions (1.16). For each record of the dataset, they furnish the preditive distribution of the response variable given all the parameters, hence, given the fixed and the random effects.

Even if we cannot obtain any information at the marginal level, using the predictive distribution (1.16), it is then possible to derive the subject-specific predictive mean $E(y_{ij} \mid x_{ij}, \mathbf{b}_i)$, so the mean of the conditional distribution on the random effects. By calculating the $E(y_{ij} \mid x_{ij}+1, \mathbf{b}_i) - E(y_{ij} \mid x_{ij}, \mathbf{b}_i)$ quantity, in fact, we obtain a subject-specific parameter that provides information on the mean change of the outcome for a unitary increase in the endogenous covariate for the i-th subject, given that all the other covariate are fixed. Some examples are shown in Table 5.7.

The values are obtained from the JSM, since is the model that has the best fitting. We select random rows of the dataset trying to understand if overall there is a pattern in the behavior of the association. First, we pick records of

Table 5.7: Values of the subject-specific parameters calculated on JSM for *protein1* and *protein2*.

Subject	Age	Protein1	Protein2
35	7.1	-0.000017	-0.000019
65	7.1	0.0017	0.0039
4	14.5	0.0031	-0.0066
15	14.5	-0.000029	0.000015
33	14.5	-0.00098	0.00044
6	17.1	-0.00025	-0.0050
34	17.1	0.00031	0.00014
14	7.0	0.0057	-0.0074
14	10.0	0.0066	-0.0176
14	14.1	0.0058	-0.0063
14	18.5	-0.000017	0.000013

subjects at the same ages to see if there is a possible relationship with time. Secondly, we select rows from the same subject but at different time points, to realize if the mean change of the outcome for a unitary increase in the *protein1* or *protein2* for the subject i depends on time. We reported only the values of patient 14 as an example because he/she is one of the few for which it is possible to select quite different age values. Seems that for both the variables, the quantities $E(y_{ij} \mid x_{ij}+1, \mathbf{b}_i) - E(y_{ij} \mid x_{ij}, \mathbf{b}_i)$ are all quite small, but this is not strange since we are studying the mean change at the subject-level. Between the results of the two proteins, there are no major differences. For each subject and age considered, the values show the same orders of magnitude, except for the row relating to subject 14 at age 10. For such a record, the unitary increase in *protein1* results in a 0.0066 increase in the mean of the outcome, whereas a unitary increase in *protein2* entails to a decrease in the mean equal to 0.0176. When comparing the values of the same subject 14 at different time points, it appears that the values are very similar for the first three ages, but there is no change in the mean outcome for the unitary increase of the two proteins for the last one. When we consider different subjects at the same time, we cannot draw firm conclusions. The results are all very small, but when we compare the order of magnitude, they are quite different from one another, so appears to be no pattern.

In conclusion, we do not have information about the marginal mean yet, what we can evaluate the association between the endogenous variable and the outcome for a specific subject at a specific time point. In our case, the subjects do not have very long trajectories because the patient's life expectancy is low. However, for longitudinal studies with large number of subjects and a longer period of observation, the association at the subject level can provide useful information as well. It may be especially useful in the medical field for identifying differences between subjects and applying appropriate treatments.

Conclusion

In this thesis, we focused on the use of joint models to model the association between longitudinal outcomes and endogenous time-varying covariates. We began by discussing endogeneity and how it affects the estimation of popular methods such as Linear Mixed Effects Models and Generalized Linear Mixed Models. We presented two types of joint models, Joint Mixed Models (JMM) and Joint Scaled Models (JSM), and their advantages over traditional methods. The joint models can produce more efficient and accurate estimates of the underlying relationships between the two correlated longitudinal variables by modeling their processes separately and combining them into a single model. We then introduced the Integrated Nested Laplace Approximation (INLA) methodology as a computationally efficient way of fitting joint models. We applied joint modeling in INLA to various examples, including analyzing the relationship between proteins and the performance of the upper limb test in patients having Duchenne Muscular Dystrophy. But first, we evaluated the performance of joint models in INLA through simulation studies.

The joint models that we used, provide information about the association between longitudinal outcomes and endogenous time-varying covariates only through the covariance among the random effects of the two processes (JMM) or a scaling factor (JSM). To more accurately measure the relationship, we also developed an association coefficient with the same interpretation as a generic regression parameter, i.e., a parameter that estimates the change in mean outcome for a unitary increase in the endogenous covariate, given that all other covariates are fixed. This makes the interpretation of the association easier and more practical to explain to doctors or scholars from outside the statistical field. However, a closed-form for it could only be defined when the outcome and the covariate are both normally distributed. To obtain an estimation of this parameter, in fact, the conditional distribution of the outcome given the endogenous covariate is required, which with both Gaussian variables, is itself a normal distribution. When they are of different types, though, we cannot exploit the property of normal distributions and we had to find another way to measure the association coefficient. The solution is not straightforward but using the output of INLA it has been possible to define an approximation. In particular, we provided an estimation of the change in mean outcome for a unitary increase in the endogenous covariate for a specific subject. Therefore, a parameter of association that has a subject-specific interpretation. The derivation of the coefficient of association with the marginal interpretation is left to future research because it necessitates a separate, more in-depth investigation.

We have introduced INLA in its Bayesian framework and show how, thanks to that, was possible to perform the joint models overcoming the problem of high-dimensional integration over the random effects. Overall, INLA is a powerful and flexible approach that can be used to address a wide range of research questions. It provided a computationally efficient way of fitting the joint models, allowing most of the time for quick inference. However, the implementation of the joint model in R-INLA is quite limited with the standard options, e.g., no more than five correlation random effects can be assumed. Furthermore, in order to implement joint models in R-INLA, the data required to be manually rewritten in a specific structure, which slows down the preparation process and leaves more room for possible careless mistakes and oversight. In the future, it would be interesting to develop a new R package that automates the generation of the dataset of interest and the estimation of the joint models.

Through simulation studies, we have shown INLA and the joint models in action and made recommendations on their use. First, we realized that whenever there are time-varying covariates, the JSM is the best choice to estimate the association. Unlike JMM, it is robust and leads to very accurate estimations. However, INLA does not always provide accurate estimations. In the presence of a small number of observations and small variances, INLA does not perfectly estimate the hyperparameters, such as the precision and the correlation parameters. In this thesis, this issue is especially relevant since the closed-form of the association coefficient is defined as a function of variances. Thus, poor precision estimations lead to poor variance estimations, which leads to poor estimation of the association coefficients. To solve this problem, we tried to change the prior distribution from the default uninformative one, but nothing seemed to improve the results. Future research could be conducted on the search for an optimal prior. Moreover, we would recommend simplifying the structure of the JMM random effects as much as possible, especially when the number of observations is low and not much information from the data is provided for the estimation.

In conclusion, we have demonstrated the importance of joint modeling of the outcome and endogenous time-varying covariates in longitudinal data analysis. Both are relevant, even though the JSM appeared to be more often accurate based on the simulation study results. They model the association in two different ways, which can lead to different conclusions, but it is crucial to use both to obtain more complete information about the association. We have implemented the joint mixed and joint scaled models within the Bayesian setting with INLA, and we have evaluated their performance. Although INLA occasionally results in biased estimations of the hyperparameter, overall it was proven to be an efficient and quick method of estimating the joint models.

Bibliography

Bates, Douglas et al. "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1 (2015), pp. 1–48. DOI: 10.18637/jss.v067.i01.

Bonat, W. H., P. J. Ribeiro, and W. M. Zeviani. "Likelihood analysis for a class of beta mixed models". In: *Journal of Applied Statistics* (2015), Vol. 42, No. 2, 252–266.

Fieuws, S. and G. Verbeke. "Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach". In: *Statistics in medicine* (2004), 23:3093–3104.

— "Pairwise Fitting of Mixed Models for the Joint Modeling of Multivariate Longitudinal Profiles". In: *Biometrics* (2006), pp. 62, 424–431.

Gómez-Rubio, V. *Bayesian Inference with INLA*. Boca Raton, FL: Chapman & Hall/CRC Press, 2020. URL: https://becarioprecario.bitbucket.io/inla-gitbook/index.html.

Gomon, G. *Joint Models: Implementation in INLA and Applications*. Leiden University, 2022.

Hadfield, Jarrod D. "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package". In: *Journal of Statistical Software* 33.2 (2010), pp. 1–22. URL: https://www.jstatsoft.org/v33/i02/.

Hastie, T., R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Heidelberg and New York: Springer-Verlag, 2009. URL: http://www-stat.stanford.edu/~tibs/ElemStatLearn/.

Hedeker, D. et al. “A note on marginalization of regression parameters from mixed models of binary outcomes”. In: *Biometrics* (2017), 74(1):354–361.

Held, L., B. Schrödle, and H. Rue. *Statistical Modelling and Regression Structures*. Berlin: Springer Verlag, 2010. Chap. Posterior and Cross-Validatory Predictive Checks: A Comparison of MCMC and INLA, pp. 91–110.

Mayhew, A.G. et al. “Performance of Upper Limb module for Duchenne muscular dystrophy”. In: *Developmental medicine & child neurology* (2019).

Muscular Dystrophy Association, About Duchenne Muscular Dystrophy. URL: https://www.mda.org/disease/duchenne-muscular-dystrophy.

Nicenboim, B., D. Schad, and S. Vasishth. *An Introduction to Bayesian Data Analysis for Cognitive Science*. 2022. URL: https://vasishth.github.io/bayescogsci/book/.

Niekerk, J. V. et al. “New frontiers in Bayesian modeling using the INLA package in R”. In: (2022).

Pinheiro, Jose et al. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-155. 2022. URL: https://CRAN.R-project.org/package=nlme.

Qian, T., P. Klasnja, and S. A. Murphy. “Linear Mixed Models with Endogenous Covariates: Modeling Sequential Treatment Effects with Application to a Mobile Health Study”. In: *Statistical science* (2020), Vol. 35, No. 3, 375–390.

R-INLA Manual. 2020. URL: http://www.r-inla.org/home.

Rizopoulos, D. *An Introduction to the Joint Modeling of Longitudinal and Survival Data, with Applications in R*. Department of Biostatistics, Erasmus University Medical Center, 2017.

Rizopoulos, Dimitris. *GLMMadaptive: Generalized Linear Mixed Models using Adaptive Gaussian Quadrature*. R package version 0.8-5. 2022. URL: https://CRAN.R-project.org/package=GLMMadaptive.

Rue, H. and Z. Liu. “Leave-group-out cross-validation for latent gaussian models”. In: (2022).

Rue, H. and S. Martino. *Approximate Bayesian Inference for Hierarchical Gaussian Markov Random Fields Models*. Department of Mathematical Sciences NTNU, Norway, 2006.

Rue, H., S. Martino, and N. Chopin. “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. In: *Journal of the Royal Statistical Society* (2009), pp. 319–392.

Rue, H., G. M. Thiago, et al. *Bayesian computing with INLA: new features*. Department of Mathematical Sciences Norwegian University of Science and Technology, 2013.

Spiegelhalter, D. et al. “Bayesian measures of model complexity and fit.” In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2002), 64(4):583–639.

Venables, W. N. and B. D. Ripley. *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer, 2002. URL: https://www.stats.ox.ac.uk/pub/MASS4/.

Verbeke, G. et al. “The analysis of multivariate longitudinal data: A review”. In: *National Institute of Health* (2014), 23(1): 42–59.

Weiss, R. E. *Modeling longitudinal data*. Springer, 2005.

Wu, L. *Mixed Effects Models for Complex Data*. University of British Columbia: Chapman & Hall/CRC Press, 2019.

Zhang, Z. “A Note on Wishart and Inverse Wishart Priors for Covariance Matrix”. In: *Journal of Behavioral Data Science* (2021), 1 (2), 119–126.

Appendix A

R Code

Code A.1: Linear Mixed Model in INLA

```
lmm ¡- function(data, N)–
  # random slopes position in the random effects vector
  data$time˙random ¡- data$id+N

  INLA˙formula ¡- y ˜ x + t +
    f(id, model = ”iid2d”, n=2*N) + # random intercepts
    f(time˙random, t, copy = ”id”) # random slopes

  LMM˙INLA ¡- inla(INLA˙formula, family = ”gaussian”, #
      likelihood
                                    data = data, control.compute = list
                                         (cpo = TRUE, dic = TRUE, waic =
                                         TRUE)) # evaluation methods

  return(list(lmm˙inla = summary(LMM˙INLA),
              CPO = -sum(log(LMM˙INLA$cpo$cpo), na.rm = TRUE),
         WAIC = LMM˙INLA$waic$waic,
         DIC = LMM˙INLA$dic$dic))
˝
```

Code A.2: Beta Mixed Model in INLA

```
bmm ¡- function(data, N)–
  # random slopes position in the random effects vector
  data$time˙random ¡- data$id+N

  INLA˙formula ¡- y ˜ x + t +
    f(id, model = ”iid2d”, n=2*N) + # random intercepts
    f(time˙random, t, copy = ”id”) # random slopes

  BMM˙INLA ¡- inla(INLA˙formula, family = ”beta”, #
      likelihood
                                  data = data, control.compute = list
                                       (cpo = TRUE, dic = TRUE, waic =
                                      TRUE), # evaluation methods
                                  control.predictor = list(compute =
                                      TRUE, link = 1), # link function
                                        for predictions
                                  control.family = list(beta.censor.
                                       value = 0)) # default censor
                                       value

  return(list(bmm˙inla = summary(BMM˙INLA),
                        CPO = -sum(log(BMM˙INLA$cpo$cpo), na.rm =
                             TRUE),
          WAIC = BMM˙INLA$waic$waic,
          DIC = BMM˙INLA$dic$dic))
˝
```

Code A.3: Joint Mixed Model in INLA with unstructure matrix

```
jmm ¡- function(data, N, time, i, failed)–
  N˙obs ¡- nrow(data)

  fixed.effects ¡- list(Intercept˙x = c(rep(1, N˙obs), rep(
     NA, N˙obs)),
         w˙x = c(data$w, rep(NA, N˙obs)),
                         t˙x = c(data$t, rep(NA, N˙obs)), #
                            fixed effects for x

                         Intercept˙y = c(rep(NA, N˙obs), rep
                            (1, N˙obs)),
                         w˙y = c(rep(NA, N˙obs), data$w),
                         t˙y = c(rep(NA, N˙obs), data$t), #
                            fixed effects for y

                         t = c(data$t, data$t))

  # Assuming unstructure variance-covariance matrix between
      random effects
  random.effects ¡- list(Random˙Intercept = c(data$id, data
     $id+N),
                          Random˙Slope = c(data$id+2*N, data
                             $id+3*N))

  INLA˙data ¡- c(fixed.effects, random.effects)
  INLA˙data$Y ¡- list(c(data$x, rep(NA, N˙obs)),
                      c(rep(NA, N˙obs), data$y))

  INLA˙formula ¡- Y ~ -1 +
        # fixed effects for x
        Intercept˙x + w˙x + t˙x +
    # fixed effects for y
        Intercept˙y + w˙y + t˙y +
  # random effects
  f(Random˙Intercept, model = ”iid4d”, n = 4*N) + f(
     Random˙Slope, t, copy = ”Random˙Intercept”)

print(paste(”Model number”, i))
tryCatch(JMM˙INLA ¡- inla(INLA˙formula, family = c(”
    gaussian”,”gaussian”),
                         data = INLA˙data, control.compute =
                              list(cpo = TRUE, dic = TRUE,
                            waic = TRUE, config=TRUE)),
               error = function(e) –print(”INLA failed: Model
                  JMM”); assign(’failed’,1,envir=globalenv());
                  failed ¡¡-1˝)
if (failed==1)–
  print(”Returning NA’s”)
  return(list(cbind(time, matrix(NA, nrow = length(time),
       ncol=3)),
                mlik = NA, overall˙dic = NA, y˙dic = NA,
                overall˙waic = NA, y˙waic = NA,
                overall˙cpo = NA, y˙cpo = NA))
˝



return(list(mlik = JMM˙INLA$mlik[1],
               overall˙dic = JMM˙INLA$dic$dic,
               y˙dic = sum(JMM˙INLA$dic$local.dic[(N˙obs+1)
                  :(2*N˙obs)], na.rm=TRUE),
               overall˙waic = JMM˙INLA$waic$waic,
               y˙waic = sum(JMM˙INLA$waic$local.waic[(N˙obs
                  +1):(2*N˙obs)], na.rm = TRUE),
               overall˙cpo = -sum(log(JMM˙INLA$cpo$cpo), na.
                  rm=TRUE),
               y˙cpo = -sum(log(JMM˙INLA$cpo$cpo[(N˙obs+1)
                  :(2*N˙obs)]), na.rm=TRUE)))
```

```
”
```

Code A.4: Joint Mixed Model in INLA with pairwise-correlation matrix

```
jmm ¡- function(data, N, time, i, failed)–
  N˙obs ¡- nrow(data)

  fixed.effects ¡- list(Intercept˙x = c(rep(1, N˙obs), rep(
      NA, N˙obs)),
          w˙x = c(data$w, rep(NA, N˙obs)),
                              t˙x = c(data$t, rep(NA, N˙obs)), #
                                  fixed effects for x

                              Intercept˙y = c(rep(NA, N˙obs), rep
                                  (1, N˙obs)),
                              w˙y = c(rep(NA, N˙obs), data$w),
                              t˙y = c(rep(NA, N˙obs), data$t), #
                                  fixed effects for y

                              t = c(data$t, data$t))

  # Assuming pairwise variance-covariance matrix between
      random effects
  random.effects ¡- list(Random˙Int˙x = c(data$id, rep(NA,
      N˙obs)),
                               Random˙slo˙x = c(data$id, rep(NA,
                                   N˙obs)),
                               Random˙Int˙y = c(rep(NA, N˙obs),
                                   data$id+N),
                               Random˙slo˙y = c(rep(NA, N˙obs),
                                   data$id+N))

  INLA˙data ¡- c(fixed.effects, random.effects)
  INLA˙data$Y ¡- list(c(data$x, rep(NA, N˙obs)),
                          c(rep(NA, N˙obs), data$y))
```

```
INLA˙formula ¡- Y ˜ -1 +
      # fixed effects for x
      Intercept˙x + w˙x + t˙x +
  # fixed effects for y
      Intercept˙y + w˙y + t˙y +
  # random effects
  f(Random˙Int˙x, model=”iid2d”, n=2*N) + f(Random˙Int˙y,
      copy = ”Random˙Int˙x”) +
              f(Random˙slo˙x, t, model=”iid2d”, n=2*N) +
                 f(Random˙slo˙y, t, copy = ”Random˙slo˙x”
                 )

print(paste(”Model number”, i))
tryCatch(JMM˙INLA ¡- inla(INLA˙formula, family = c(”
   gaussian”,”gaussian”),
                        data = INLA˙data, control.compute =
                            list(cpo = TRUE, dic = TRUE,
                           waic = TRUE, config=TRUE)),
         error = function(e) –print(”INLA failed: Model
            JMM”);assign(’failed’,1,envir=globalenv());
            failed¡¡-1˝)
if (failed==1)–
  print(”Returning NA’s”)
  return(list(cbind(time, matrix(NA, nrow = length(time),
      ncol=3)),
              mlik = NA, overall˙dic = NA, y˙dic = NA,
              overall˙waic = NA, y˙waic = NA,
              overall˙cpo = NA, y˙cpo = NA))
˝



return(list(mlik = JMM˙INLA$mlik[1],
            overall˙dic = JMM˙INLA$dic$dic,
                    y˙dic = sum(JMM˙INLA$dic$local.dic[(N˙obs+1)
                        :(2*N˙obs)], na.rm=TRUE),
                    overall˙waic = JMM˙INLA$waic$waic,
                    y˙waic = sum(JMM˙INLA$waic$local.waic[(N˙obs
                        +1):(2*N˙obs)], na.rm = TRUE),
                    overall˙cpo = -sum(log(JMM˙INLA$cpo$cpo), na.
                        rm=TRUE),
                    y˙cpo = -sum(log(JMM˙INLA$cpo$cpo[(N˙obs+1)
                        :(2*N˙obs)]), na.rm=TRUE)))
″
```

Code A.5: Joint Scaled Model in INLA

```
jsm ¡- function(data, N, time, i, failed)–
  N˙obs ¡- nrow(data)

  # Fixed effect part only for the y
  fixed.effects ¡- list(Intercept˙y = c(rep(NA, N˙obs), rep
      (1, N˙obs)),
                                                        w˙y=c(rep(
                                                           NA, N˙
                                                           obs),
                                                           data$w),
                              t˙y = c(rep(NA, N˙obs), data$t),

                              w =c(data$w, data$w),
                              t = c(data$t, data$t))

  # Fixed effects of the endogenous covariate x + random
      effects
  random.effects ¡- list(Intercept˙x = c(rep(1, N˙obs), rep
      (NA, N˙obs)),
        w˙x =c(rep(1, N˙obs), rep(NA, N˙obs)),
                                t˙x = c(rep(1, N˙obs), rep(NA, N˙
                                   obs)),

                                Intercept˙x˙scaled = c(rep(NA, N˙
                                   obs), rep(1, N˙obs)),
                                w˙x˙scaled=rep(NA, N˙obs), rep(1,
                                    N˙obs),
                                t˙x˙scaled = c(rep(NA, N˙obs),
                                   rep(1, N˙obs)),

                                Random˙intercept˙x = c(data$id,
                                   rep(NA, N˙obs)),
                                Random˙slope˙x = c(data$id+N, rep
                                    (NA, N˙obs)),
                                 Random˙intercept˙x˙scaled = c(rep
                                    (NA, N˙obs), data$id),
                                 Random˙slope˙x˙scaled = c(rep(NA,
                                     N˙obs), data$id+N),

                                 Random˙intercept˙y = c(rep(NA, N˙
                                    obs), data$id),
                                 Random˙slope˙y = c(rep(NA, N˙obs)
                                    , data$id+N))



INLA˙data ¡- c(fixed.effects, random.effects)
INLA˙data$Y ¡- list(c(data$x, rep(NA, N˙obs)),
                    c(rep(NA, N˙obs), data$y))

INLA˙formula = Y ˜ -1 +
  # fixed effects of x (define as iid random affects)
  f(Intercept˙x) + f(w˙x, w) + f(t˙x, t) +
  # scaled fixed effects of x
  f(Intercept˙x˙scaled, copy = ”Intercept˙x”, hyper =
     list(beta = list(fixed=FALSE))) +
  f(w˙x˙scaled, copy = ”w˙x”, same.as=’Intercept˙x˙scaled
     ’, hyper = list(beta = list(fixed=FALSE))) +
  f(t˙x˙scaled, t, copy = ”t˙x”, same.as = ’Intercept˙x˙
     scaled’, hyper = list(beta = list(fixed=FALSE))) +
  # fixed effects of y
  Intercept˙y + w˙y + t˙y +

  # random effects for x
  f(Random˙intercept˙x, model = ”iid2d”, n = 2*N) + f(
     Random˙slope˙x, t, copy = ”Random˙intercept˙x”) +
  # scaled random effects
  f(Random˙intercept˙x˙scaled, copy = ’Random˙intercept˙x
      ', same.as = 'Intercept˙x˙scaled', fixed = FALSE) +
  f(Random˙slope˙x˙scaled, t, copy = 'Random˙intercept˙x'
      , same.as = 'Intercept˙x˙scaled', fixed = FALSE) +
  # random effects of y
  f(Random˙intercept˙y, model = ”iid2d”, n = 2*N) + f(
      Random˙slope˙y, t, copy = ”Random˙intercept˙y”)

print(paste(”Model number”, i))
tryCatch(JSM˙INLA ¡- inla(INLA˙formula, family = c(”
    gaussian”,”gaussian”),
                              data = INLA˙data, control.
                                 compute = list(cpo =
                                 TRUE, dic = TRUE, waic =
                                 TRUE, config=TRUE)),
         error = function(e) –print(”INLA failed: Model 3
            ”);assign('failed',1,envir=globalenv());
            failed ¡¡-1˝)
if (failed==1)–
  print(”Returning NA's”)
  return(list(coeff˙jsm = cbind(time, matrix(NA, nrow =
     length(time), ncol=3)),
              mlik = NA, overall˙dic = NA, y˙dic = NA,
              overall˙waic = NA, y˙waic = NA,
              overall˙cpo = NA, y˙cpo = NA))
˝



return(list( mlik = JSM˙INLA$mlik[1],
            overall˙dic = JSM˙INLA$dic$dic,
            y˙dic = sum(JSM˙INLA$dic$local.dic[(N˙obs+1)
               :(2*N˙obs)], na.rm=TRUE),
            overall˙waic = JSM˙INLA$waic$waic,
            y˙waic = sum(JSM˙INLA$waic$local.waic[(N˙obs
               +1):(2*N˙obs)], na.rm = TRUE),
                         overall_cpo = -sum(log(JSM_INLA$cpo$cpo), na.
                              rm=TRUE),
                         y_cpo = -sum(log(JSM_INLA$cpo$cpo[(N_obs+1)
                              :(2*N_obs)]), na.rm=TRUE)))
"
```