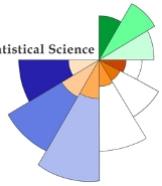
Joint Models: Implementation in INLA and Applications

Georgy Gomon (s1559370)

Daily Supervisor: Dr. S. Tsonaka
(LUMC)
Second supervisor: Dr. B. Mertens
(LUMC)

MASTER'S THESIS

Defended on ...



STATISTICS & DATA SCIENCE

Contents

Abstract

Longitudinal data is encountered when repeated measurements are performed on subjects over a period of time. Many models exist to fit longitudinal data, all sharing the feature that explanatory covariates are introduced into the model to explain the observed change over time. A special type of covariate found within the longitudinal framework is the time-varying covariate, such as the BMI or blood biomarker levels. These covariates, whose value changes over time, are a blessing in disguise. On one hand they allow the researcher to better model the change in outcome over time, and thus fit a better model. On the other hand bias can be introduced when the time-varying covariates depend on the outcome or its previous values.
Time-varying covariates that introduce such bias are called endogenous time-varying covariates: these are covariates whose current value, given their own history, depends on past values of the outcome. In the presence of such endogenous covariates, because of the cross-reliance of the endogenous covariate on the outcome, standard Mixed Models are no longer valid and one needs to resort to joint modelling of both the outcome and the endogenous covariate.
In this thesis several such joint longitudinal models will be discussed. Our focus will be on Joint Mixed Models and Joint Scaled Models. Both explicitly model the dependence between the outcome and the endogenous covariate, thereby removing the possible bias incurred by the time-varying covariate. We shall show how to fit these models using a novel Bayesian technique called INLA (Integrated Nested Laplace Integration), which is an elegant technique and a good alternative for the complex and long MCMC estimation procedure. Although INLA has seen rapid development over recent years, joint longitudinal models have so far received little attention. The goal of this thesis is therefore to implement several joint longitudinal models within the INLA framework and apply them on a simulation study as well as on a synthetic version of a clinical dataset.

Version of June 29, 2022– Created June 29, 2022 - 06:55

Chapter 1

Introduction

Longitudinal data analysis typically focuses on the effect covariates have on an outcome over time. As example we can imagine studying the effect of the covariates 'sex', 'age' and 'treatment regimen' on the outcome 'lung capacity' following a Covid-19 infection. Within the context of longitudinal data we can split the covariates into 3 groups. The first distinction is between time-invariant and time-varying covariates. In our hypothetical example 'sex' is a time-invariant covariate, as it does not change over time. The time-varying covariates can be further split into endogenous and exogenous time-varying covariates [2]. An exogenous time-varying covariate is a covariate whose current value, given its own history, does not depend on the value of the outcome at previous measurement times. In our example 'age' is such an exogenous time-varying covariate. 'Age' does change over time, but it is independent of 'lung capacity' (the outcome) at previous time points. Finally we have the endogenous time-varying covariates, which are covariates whose current value does depend on previous values of the outcome, given their own history. In our example 'treatment regimen' is such an endogenous time-varying covariate, since the lung capacity at previous measurements can influence the treatment regimen the patient is currently receiving, e.g: If the patient is recovering the treatment can be scaled down. Estimating the effect of such endogenous time-varying covariates (we shall call them endogenous covariates) can be challenging because the dependency structure between the outcome and the endogenous covariate needs to be modelled. For this reason a standard linear mixed model is no longer valid but instead the endogenous covariate and the outcome need to be properly modelled jointly, see [2]. This leads us into the framework of joint longitudinal models. Within the scope of this thesis the different approaches to joint modelling of the endogenous covariate and the outcome will be studied. The emphasis will be on 3 methods:

- First is a multivariate model in which the multiple outcomes are jointly Gaussian distributed. The association between the two outcomes is then modelled via correlated errors of the outcomes, see [1].
- Second is a joint mixed model in which the association between the multiple outcomes is given by multivariate normally distributed random effects, see [4],[18] & [19].
- Last a joint model is proposed in which the linear predictor of the endogenous covariate is inserted into the linear predictor of the outcome with an associated scaling factor, a joint model broadly used within the context of survival analysis, see [7], [10] & [11].

During the thesis these methods will be applied in R within the Bayesian framework. The emphasis will be to implement the methods using INLA (Integrated Nested Laplace Approximation) and its associated R package R-INLA. INLA is a new Bayesian framework based on Laplace Integration that removes the need for extensive MCMC estimation and is therefore much quicker than standard Bayesian methods. For more information on INLA we refer to [13]. For more information about the current joint models implementations of INLA we refer to [14], [9] & [15].

The thesis will be organized as follows: In the second chapter we shall mathematically define exogenous and endogenous time-varying covariates, followed by the definition of the Linear Mixed Model. The main part of the second chapter will concern the introduction of the different joint models and the manner in which the association between the outcome and the endogenous covariate is measured within these joint models. The third chapter will introduce INLA and several goodness of fit measures will be discussed. The fourth chapter will show how the different joint models can be fit within the R-INLA framework. Also, the implementation of joint models in other (non-Bayesian) R-packages will be discussed. In the fifth chapter a small scale simulation study is presented. The goal is two-fold: First to show that joint models with INLA offer a computationally viable solution. Secondly we obtain some indications about the settings in which joint models are a viable and useful alternative to standard mixed models. We end the thesis with a real-life application of joint models on a synthetic version of the LUMC Covid-19 dataset.

Chapter 2

Joint models

In this section we shall introduce endogenous and exogenous covariates. We shall continue by defining a Linear Mixed Model, before ending with the main part of this chapter: The definition and discussion of the different types of joint models.

2.1 Endogenous vs exogenous covariates

Within longitudinal studies we have both time-invariant and time-varying covariates. Examples of time-invariant covariates are sex and genetic profile. Examples of time-varying covariates are age, biomarkers, air-pollution exposure and treatment dose. The time-varying covariates can furthermore be divided into 2 groups: exogenous and endogenous covariates. To define exogenous and endogenous covariates (see [2]) the following notation is introduced:

- $y_i(t)$: Value of the response y for subject i at time t.
- $v_i(t)$: Value of the time-varying covariate v for subject i at time t.
- $\mathcal{H}_i^Y(t)$: History of the response process of subject i until time t:

$$\mathcal{H}_i^Y(t) = \{y_i(t_{i1}), y_i(t_{i2}), ..., y_i(t_{ik}); t_{ik} \leq t\}.$$

- $\mathcal{H}_i^V(t)$: History of the time-varying covariate process of subject i until time t:

$$\mathcal{H}_i^V(t) = \{v_i(t_{i1}), v_i(t_{i2}), ..., v_i(t_{ik}); t_{ik} \leq t\}.$$

- $\mathbf{w}_i$: Vector of time-independent covariates.

Definition 2.1.1 (Exogenous Covariate) *$v_i(t)$ is an exogenous covariate with respect to the outcome process if the covariate at time t is conditionally independent of the history of the out come process at time t, given the history of the covariate process at time t. Mathematically,*

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

Thus, for an exogenous covariate the exposure at time t does not depend on previous values of the response. Examples of exogenous covariates are age and air-pollution exposure.
For exogenous covariates the likelihood $f(\mathbf{y_i}, \mathbf{v_i}|\mathbf{w_i}, \boldsymbol{\theta})$ can be factorized:

$$\begin{aligned} f(\mathbf{y_i}, \mathbf{v_i}|\mathbf{w_i}, \boldsymbol{\theta}) &= \left[\prod_{t=1}^{T} f\left(y_i(t)|\mathcal{H}_i^Y(t-1), \mathcal{H}_i^V(t), \mathbf{w}_i, \boldsymbol{\theta_1}\right)\right] \cdot \left[\prod_{t=1}^{T} f\left(v_i(t)|\mathcal{H}_i^V(t-1), \mathbf{w}_i, \boldsymbol{\theta_2}\right)\right] = \\ &= \mathcal{L}_Y(\boldsymbol{\theta_1}) \cdot \mathcal{L}_V(\boldsymbol{\theta_2}). \end{aligned} \tag{2.1}$$

Here $\boldsymbol{\theta_1}$ & $\boldsymbol{\theta_2}$ are the parameters of interest in the likelihoods of the outcome $\mathbf{y}$ and the exogenous covariate $\mathbf{v}$ respectively, while $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$.
The factorization of the joint likelihood means that we do not need to model the covariate process of $\mathbf{v}$ or it's parameters $\boldsymbol{\theta_2}$ in order to make inference about θ_1 and the outcome $\mathbf{y}$.

Definition 2.1.2 (Endogenous Covariate) *$v_i(t)$ is an endogenous covariate with respect to the outcome process if the covariate at time t is conditionally dependent of the history of the outcome process at time t, given the history of the covariate process at time t. Mathematically,*

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

Thus, for an endogenous covariate the covariate at time t does depend on previous values of the response. An example might occur when investigating the effect of a certain treatment regimen on symptom severity. If no symptoms are present, the treatment regimen might be made less stringent and vice-versa, leading to dependence between the outcome (symptom severity) and the time-varying covariate (treatment regimen).
For endogenous covariates the factorization as shown in equation 2.1 does not hold, meaning that the endogenous covariate process cannot be ignored from the joint likelihood of the outcome and the covariate. To make the distinction between endogenous and exogenous covariates clear throughout the thesis we shall be using v to indicate exogenous time-varying covariates, while endogenous time-varying covariates will be indicated by x.

2.2 Linear mixed model

When assuming the time-varying covariate is exogenous, a standard linear mixed model (LMM) suffices to obtain valid inference. We shall therefore introduce the LMM as a benchmark to compare the joint models to. We shall limit ourselves to a brief introduction of mixed models, for more information we refer to [4], [5].

Definition 2.2.1 (Linear Mixed Model) *The LMM with exogenous time-varying covariates shall be of the following form:*

$$y_i(t_{ij}) = \mathbf{w}_i^\intercal \cdot \boldsymbol{\alpha} + \mathbf{v}_i^\intercal(t_{ij}) \cdot \boldsymbol{\beta} + \mathbf{z}_i^\intercal(t_{ij}) \cdot \mathbf{b_i} + \epsilon_i(t_{ij}), \tag{2.2}$$

with:

- $y_i(t_{ij})$*: Outcome for patient i at time* t_{ij}*. In total there are* $i = 1, ..., N$ *patients, with every patient having* $j = 1, ..., n_i$ *measurements.*
- $\mathbf{w}_i$*: Vector of time-invariant covariates*
- $\mathbf{v}_i(t_{ij})$*: Vector of time-varying covariates at time* t_{ij}*.*
- $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$*: Coefficient vectors of fixed time-invariant and time-varying covariates respectively.*
- $\mathbf{z}_i(t_{ij})$*: Vector of random effect covariates, possibly time-varying.*
- $\mathbf{b_i} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$*: Random effects vector. The exact structure of the random effects variance-covariance matrix* $\mathbf{D}$ *will depend on the question at hand.*
- $\epsilon_i(t_{ij}) \sim \mathcal{N}(0, \sigma^2)$*: Residual errors, with* $\epsilon_i(t_{ij}) \perp\!\!\!\perp \mathbf{b_i}$*. Note that hereby we assume independent errors at different measurement points. This is the Conditional Independence Assumption, which states that the random effects capture all correlation across time thus leading to independent errors.*

Throughout the thesis we shall focus on the association between the time-varying covariate and the outcome. In the LMM this association is easiest to quantify as it is given by the coefficient $\boldsymbol{\beta}$.

Definition 2.2.2 (LMM association coefficient) *It is well-known that in the LMM the coefficient $\boldsymbol{\beta}$ gives the association between the outcome $\mathbf{y}$ and the time-varying covariates $\mathbf{v}_i(t_{ij})$. Its interpretation is also well-known: A unitary increase in $\mathbf{v}_i(t_{ij})$ will yield a change of $\mathbf{v}_i^\intercal(t_{ij})\boldsymbol{\beta}$ in the outcome, given that all other covariates are fixed. We shall denote the coefficient $\boldsymbol{\beta}$ as $\boldsymbol{\beta}_v^{lmm}$, since this is the association between the time-varying covariate and the outcome in the LMM:*

$$\boldsymbol{\beta}_v^{lmm} := \boldsymbol{\beta}. \tag{2.3}$$

The subscript v indicates that the time-varying covariate is assumed to be exogenous in nature. The statistical significance of $\boldsymbol{\beta}_v^{lmm}$ can be tested by using any of the common statistical tests such as an F-test in the frequentist framework or by determining the highest posterior density interval (HPDI) in the bayesian framework.

Before continuing with joint models a few important notes have to be made concerning the LMM.

2.2.1 Lag in a linear mixed model

Within the context of the LMM it is straightforward to introduce lagged values of the time-varying covariate. These are values of the time-varying covariate at previous time-points used to predict the current outcome. A well known example of a lagged effect is the relation between smoking and lung cancer. It is widely known that smoking increases the risk of lung cancer. However, it is not the smoking on the day of the diagnosis that caused lung cancer. Instead, it is the lagged effect of smoking over multiple years that caused one to develop lung cancer.
There are many methods to introduce lagged values into a longitudinal model. In this thesis we shall discuss a simple additive model. This additive model is introduced within the framework of the LMM.

Definition 2.2.3 (LMM with lagged values) *Adding lagged values to a LMM is straightforward, one simply adds the measurements of the time-varying covariates at previous time-points. The resulting model is of the following form:*

$$y_i(t_{ij}) = \mathbf{w}_i^\intercal \cdot \boldsymbol{\alpha} + \mathbf{v}_i^\intercal(t_{ij}) \cdot \boldsymbol{\beta}_0 + \sum_{l=1}^{L} \mathbf{v}_i^\intercal(t_{ij} - l) \cdot \boldsymbol{\beta}_l + \mathbf{z}_i^\intercal(t_{ij}) \cdot \mathbf{b_i} + \epsilon_i(t_{ij}). \tag{2.4}$$

with:

- *L: The total degree of lag introduced into the model. All values of the time-varying covariate until time $t = t_{ij} - L$ are used as predictors in the model.*
- *$\boldsymbol{\beta}_l$ with $l = 1, ..., L$: The coefficient vectors of the lagged values.*
- *All other notation is similar to the notation introduced in the definition of the LMM, see section 2.2.1.*

Definition 2.2.3 represents a LMM with lag of degree L. All values of the time-varying covariate are included until time $t = t_{ij} - L$. Note that if for some lag l and patient i a measurement is missing at time $t = t_{ij} - l$, the lagged effect cannot be used to fit the model. Also note that the lagged effect is defined as $\mathbf{v}_i^\intercal(t_{ij} - l)$ instead of $\mathbf{v}_i^\intercal(t_{i,(j-l)})$. This is because the lagged effect of interest occurs at a specific time-point instead of the patient's previous measurement. The approach presented here is feasible if the study design is balanced and the amount of missing data is minimal. However, when the amount of missing data is substantial and/or the time is continuous, many of the lagged values of the time-varying covariate will not be available. In such cases a model needs to be postulated for the time-varying covariate as well, thus necessitating the use of joint (but independent) models.

2.2.2 Joint exogenous LMM

We saw that the inclusion of lagged values in case of an unbalanced design and/or missing data necessitates a model for the time-varying covariate.
A similar situation occurs when we want to use the LMM for prediction. Suppose we have fitted an LMM on available data and subsequently want to use the model to predict future observations of a new patient. Using the available repeated measurements we can determine the random effect level of this new patient. The time-independent covariates of the patient are known to us, as they are constant over time. A problem however arises of how to determine the values of the time-varying covariates at future time-points, as we ultimately need them to predict the outcome at future time-points.
Thus, the inclusion of lagged values as well as using an LMM for prediction necessitates a model to be postulated for the time-varying covariates. The factorization in equation 2.1 tells us that we can propose a LMM for the time-varying covariate independent of the LMM for the outcome.

Definition 2.2.4 (Joint Exogenous LMM) *The Mixed Model for predicting exogenous time-varying covariate v has the following form:*

$$v_i(t_{ij}) = \mathbf{w}_i^\intercal \cdot \boldsymbol{\alpha}_v + \mathbf{f}^\intercal(t_{ij}) \cdot \boldsymbol{\beta}_v + \mathbf{z}_i^\intercal(t_{ij}) \cdot \mathbf{b}_{vi} + \epsilon_{vi}(t_{ij})$$

Note that the notation is very similar to the LMM for the outcome, definition 2.2.1, with the only difference being that there are no time-varying covariates in the linear predictor. Instead, the only time-dependency allowed is via functions of time. Also note that because the time-varying covariate is assumed to be exogenous the notation v is used.
New notation introduced in the model is:

- $v_i(t_{ij})$*: time-varying covariate for patient i at time* t_{ij}*. In total there are* $i = 1, ..., N$ *patients, with every patient having* $j = 1, ..., n_i$ *measurements.*
- $\mathbf{f}^\intercal(t_{ij})$*: Vector of functions of time, the only time-dependency allowed for the prediction of exogenous covariates.*
- $\mathbf{z}_i(t_{ij})$*: Design vector of random effects. Note that here no time-varying covariates are allowed either. All time-dependent random effects should consist only of functions of time.*
- $\mathbf{b}_{vi} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$*: Random effects vector.*
- $\epsilon_{vi}(t_{ij}) \sim \mathcal{N}(0, \sigma^2)$*: Residual errors, with* $\epsilon_{vi}(t_{ij}) \perp\!\!\!\perp \mathbf{b_i}$.

Note that if multiple time-varying covariates are present in the LMM of the outcome, definition 2.2.1, multiple models need to be fitted.
By constructing an independent LMM for the exogenous covariate alongside a LMM for the outcome (equation 2.2) we have constructed a joint (albeit independent) model for both the outcome and the exogenous time-varying covariate. Note that the factorization of the joint likelihood in equation 2.1 allows for the independence of these models.

2.3 Joint (endogenous) models

Definition 2.1.2 showed that in case of endogenous covariates the likelihood $f(\mathbf{y_i}, \mathbf{x_i}|\mathbf{w_i}, \boldsymbol{\theta})$ of the outcome $\mathbf{y}$ and endogenous covariate $\mathbf{x}$ cannot be factorized into 2 independent likelihoods, and thus the covariate process cannot be ignored. Instead the outcome and time-varying covariate should be modelled jointly. Note that as we are talking about endogenous covariates we are using the notation $\mathbf{x}$ instead of $\mathbf{v}$ for the time-varying covariate.
In this thesis we shall be looking at 3 main methods which enable joint modelling of both the outcome and the endogenous covariate. For a more complete overview of joint models we refer to [16].

2.4 Multivariate joint model

The first joint model we shall be examining is a multivariate normal joint model, discussed in [1].

Definition 2.4.1 (Multivariate joint model) *In the multivariate joint model the association between the outcome y and the endogenous covariate x is realized via the residual errors covariance matrix $\mathbf{\Sigma}_i$. The model specification has the following form:*

$$\begin{cases} y_i(t_{ij}) = \mathbf{w}_i^\intercal \cdot \boldsymbol{\alpha}_y + \mathbf{v}_{yi}^\intercal(t_{ij})\boldsymbol{\beta}_y + \epsilon_{yi}(t_{ij}) \\ x_i(t_{ij}) = \mathbf{w}_i^\intercal \cdot \boldsymbol{\alpha}_x + \mathbf{v}_{xi}^\intercal(t_{ij})\boldsymbol{\beta}_x + \epsilon_{xi}(t_{ij}) \end{cases} \quad with \quad \begin{bmatrix} \epsilon_{yi} \\ \epsilon_{xi} \end{bmatrix} \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathbf{\Sigma}_i)$$

Most of the notation is similar to the notation introduced when presenting the LMM (definition 2.2.1). The only difference is the introduction of the endogenous time-varying covariate:

- *$x_i(t_{ij})$: Endogenous time-varying covariate for patient i at time t_{ij}. In total there are $i = 1, ..., N$ patients, with every patient having $j = 1, ..., n_i$ measurements.*
- *$\boldsymbol{\alpha}_x$ and $\boldsymbol{\beta}_x$: Coefficient vectors of fixed time-invariant and time-varying covariates respectively.*

We assume that the outcome $y_i(t_{ij})$ and the endogenous variable $x_i(t_{ij})$ need to be measured at the same time-points t_{ij}.

In this model the marginal association can be measured between any pair of time-points and missing data in the response and/or covariate can be handled simultaneously.
When implementing the multivariate joint model a choice must be made about the structure of the variance-covariance matrix $\mathbf{\Sigma}_i$. Possible choices are, among others, an Unstructured form, Compound symmetry, Auto-regressive and Toeplitz. The unstructured form is not feasible when the number of repeated measurements per patient increases beyond just a few, as for n repeated measurements per patient the unstructured variance-covariance matrix requires $\frac{2n(2n+1)}{2}$ parameters. For just 4 repeated measurements per patient this would result in 36 parameters to be estimated. Thus, to apply the multivariate joint model one will have to introduce some restrictions on the variance covariance matrix $\mathbf{\Sigma}_i$.
Note that in the definition it is assumed that the endogenous variable and the outcome need to be measured simultaneously. This is not necessarily the case, as options are available to use the multivariate joint models in case of unbalanced data, such as when using the continuous auto-regressive correlation variance covariance matrix $\mathbf{\Sigma}_i$. We shall however not dive deeper into these options in the course of this thesis.

2.5 Joint Mixed Model

Next we shall be examining the Joint Mixed Model (JMM).

Definition 2.5.1 (Joint Mixed Model) *The Joint Mixed Model (JMM) is a model where the association between the outcome y and the endogenous covariate x is measured via the random effects variance-covariance matrix $\mathbf{D}$ and potentially the residual errors covariance matrix $\mathbf{\Sigma}_i$. The mathematical notation of the model is given by:*

$$\begin{cases} y_i(t_{ij}) = \mathbf{w}_i \cdot \boldsymbol{\alpha}_y + \mathbf{v}_{yi}^\intercal(t_{ij})\boldsymbol{\beta}_y + \mathbf{z}_{yi}^\intercal(t_{ij})\mathbf{b}_{yi} + \epsilon_{yi}(t_{ij}) \\ x_i(s_{ij}) = \mathbf{w}_i^\intercal \cdot \boldsymbol{\alpha}_x + \mathbf{v}_{xi}^\intercal(s_{ij})\boldsymbol{\beta}_x + \mathbf{z}_{xi}^\intercal(s_{ij})\mathbf{b}_{xi} + \epsilon_{xi}(s_{ij}) \end{cases} \quad with$$
$$\begin{bmatrix} \mathbf{b}_{yi} \\ \mathbf{b}_{xi} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}); \quad \begin{bmatrix} \epsilon_{yi} \\ \epsilon_{xi} \end{bmatrix} \sim \mathcal{N}_{n_i}(\mathbf{0}, \mathbf{\Sigma}_i).$$

Here we have:

- $y_i(t_{ij})$: *Outcome for patient i at time* t_{ij}. *In total there are* $i = 1, ..., N$ *patients, with every patient having* $j = 1, ..., n_i$ *measurements.*
- $x_i(s_{ij})$: *Value of the endogenous covariate x for patient i at time* s_{ij}. *Note that the times* t_{ij} *and* s_{ij} *at which the outcome and the endogenous covariate are measured can be different.*
- *All other notation is similar to the notation introduced in the definition of the LMM (definition 2.2.1).*

Within the course of this thesis we shall be working under the Conditional Independence Assumption, the assumption that the random effects capture all correlation between repeated measurements. Hereby no correlation is left to be modelled by the error terms. Under the Conditional Independence Assumption the errors of repeated measurements are independent, thus giving $\mathbf{\Sigma}_i = \sigma_\epsilon^2 \mathbf{I}$.
An advantage of the JMM is that the outcome and endogenous covariate do not need to be measured at the same time, in contrast to the multivariate joint model. It is important to note that this holds only when working under the Conditional Independence Assumption.
Within the thesis the JMM shall be used as a means to jointly model the outcome and endogenous covariate. However, it can be also to used to model multiple associated outcomes, see [3].

2.5.1 Association between outcome and time-varying covariate

The association between the endogenous covariate and the outcome in the JMM is given by the covariances of the random effects in matrix **D**. These values are not easy to interpret. This stands in contrast to the LMM, where the association was given by $\boldsymbol{\beta}_v^{lmm}$ (definition 2.2.2) and can be interpreted as the change in mean outcome for a unitary increase in the exogenous covariate, given that all other covariates are fixed. We would like to have a similar value within the context of the JMM.
To obtain such an estimate the conditional distribution of the outcome given the endogenous covariate needs to be derived. The joint distribution of the outcome and the endogenous covariate at time t for subject i is bivariate normal and has the following form:

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

with

$$\mu_{y,i}(t) = \mathbf{w}_i^\intercal \cdot \boldsymbol{\alpha}_y + \mathbf{v}_{yi}^\intercal(t)\boldsymbol{\beta}_y,$$
$$\mu_{x,i}(t) = \mathbf{w}_i^\intercal \cdot \boldsymbol{\alpha}_x + \mathbf{v}_{xi}^\intercal(s)\boldsymbol{\beta}_x.$$

Note the explicit dependence on time of all elements involved in the formulation of the bivariate normal. It is clear that the means $\mu_{y,i}(t)$ and $\mu_{x,i}(t)$ are time-dependent. However, it should also be noted that the variance $\sigma^2_{y,i}(t)$ is time dependent if the term $\mathbf{z}_{yi}^\intercal(t)\mathbf{b}_y$ contains any functions of time. Similarly $\sigma^2_{x,i}(t)$ and $\rho_i(t)$ can also be functions of time. The expressions of $\sigma^2_{y,i}(t)$, $\sigma^2_{x,i}(t)$ and $\rho_i(t)$ can be derived from the model at hand and depend on the covariance matrix **D** of the random effects and the covariance matrix $\mathbf{\Sigma}$ of the residual errors. The derivation and time dependency will be illustrated when deriving the above mentioned quantities for a specific model in section 4.5.1.
Having obtained the bivariate normal distribution we can easily obtain the conditional distribution. The conditional distribution of the outcome given the endogenous covariate is given by:

$$f(y_i(t)|x_i(t) = a) = \mathcal{N}\left(\mu_{y,i}(t) + \frac{\sigma_{y,i}(t)}{\sigma_{x,i}(t)}\rho_i(t)(a - \mu_{x,i}(t)), (1 - \rho_i^2(t))\sigma^2_{y,i}(t)\right). \tag{2.5}$$

Using the conditional distribution we can answer questions relating to the association between the endogenous covariate and the outcome. We can simply calculate the effect on the outcome of increasing the endogenous covariate by a unitary amount by calculating the expectation of the following

quantity:

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

Thus, a unitary increase in the endogenous covariate $x_i(t)$ at time t will yield a change of $\frac{\sigma_{y,i}(t)}{\sigma_{x,i}(t)}\rho_i(t)$ in the outcome $y_i(t)$ at time t for subject i.

Definition 2.5.2 (JMM association coefficient) *The time-dependent coefficient which measures the change in the outcome $y(t)$ with a unitary increase in the endogenous covariate $x(t)$ at time t will be referred to as JMM association coefficient and will be denoted by:*

$$\beta_x^{jmm}(t) := \frac{\sigma_{y,i}(t)}{\sigma_{x,i}(t)}\rho_i(t)=\frac{Cov(y,x)_i(t)}{\sigma^2_{x,i}(t)}.$$

Note that $\beta_x^{jmm}(t)$ has the same interpretation as β_v^{lmm} in the LMM (see definition 2.2.2). A difference however is that the coefficient β_v^{lmm} in the LMM is time-independent, while the coefficient $\beta_x^{jmm}(t)$ in the JMM is time-dependent. As the random effects covariance matrix $\mathbf{D}$ (and thus also $\sigma_{y,i}(t)$, $\sigma_{x,i}(t)$ & $\rho_i(t)$) are equal for all subjects i there is no dependence of the JMM association coefficient $\beta_x^{jmm}(t)$ on subject i.

A problem arises of how to show whether $\beta_x^{jmm}(t)$ is significant. Unfortunately, no theoretical quantity has been found for the confidence interval of this quantity.
However, INLA offers a solution. It allows us to sample from the joint distribution of the hyperparameters $\sigma_y^2(t)$, $\sigma_x^2(t)$ and $\rho(t)$. This allows for the construction of empirical confidence interval of $\beta_x^{jmm}(t)$ via sampling. Note that this method is exclusively available in INLA. When analysing the COVID results we shall use sampling from the joint posterior to obtain credible intervals for the estimate $\beta_x^{jmm}(t)$, see section 6.3.2.
The significance of $\beta_x^{jmm}(t)$ and the covariance parameters in the variance-covariance matrix $\mathbf{D}$ indicates the significance of the association between the outcome and the endogenous covariate. If these terms are not significant our model reduces to a LMM without the time-varying covariate, see section 2.2.1. Note that the significance of $\beta_x^{jmm}(t)$ and the covariance parameters in the variance-covariance matrix $\mathbf{D}$ is not a test for whether the time-varying covariate is endogenous or exogenous in nature. In case of a significant time-varying exogenous covariate the JMM will show a significant association, a result that will be highlighted in the simulation study (section 5.5).

2.6 Joint Scaled Model

Lastly we present a joint model in which the linear predictor of the endogenous covariate is copied into the linear predictor of the outcome with an associated scaling factor γ. Models of this type are very common in survival analysis, where a longitudinal model and a survival model are combined via a scaling factor. More information about the use of such models within survival analysis can be found in [10] & [11].

Definition 2.6.1 (Joint Scaled Model (JSM)) *Joint models with scaled linear predictors shall be referred to as Joint Scaled Models (JSM). The mathematical notation of the JSM is given by:*

$$\begin{cases} x_i(t_{ij}) = m_i(t_{ij}) + \epsilon_{xi}(t_{ij}) \\ y_i(s_{ij}) = \mathbf{w}_{yi}^\intercal \alpha_{\mathbf{y}} + \gamma m_i(s_{ij}) + \mathbf{v}_{yi}^\intercal(s_{ij})\beta_{\mathbf{y}} + \mathbf{z}_{yi}^\intercal(s_{ij})\mathbf{b}_{yi} + \epsilon_{yi}(s_{ij}) \end{cases}$$

with

$$m_i(t_{ij}) = \mathbf{w}_{xi}^\intercal \alpha_{\mathbf{x}} + \mathbf{v}_{xi}^\intercal(t_{ij})\beta_{\mathbf{x}} + \mathbf{z}_{xi}^\intercal(t_{ij})\mathbf{b}_{xi}$$

and

$$\begin{aligned}
&\mathbf{b}_{xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_x), \quad \epsilon_{xi}(t_{ij}) \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_x^2) \\
&\mathbf{b}_{yi} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_y), \quad \epsilon_{yi}(s_{ij}) \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_y^2) \\
&\epsilon_{xi}(t_{ij}) \perp\!\!\!\perp \mathbf{b}_{xi}, \quad \epsilon_{yi}(t_{ij}) \perp\!\!\!\perp \mathbf{b}_{yi}, \quad \epsilon_{xi}(t_{ij}) \perp\!\!\!\perp \epsilon_{yi}(t_{ij}).
\end{aligned}$$

Hereby the following notation is used:

- $m_i(t_{ij})$*: The linear predictor of the endogenous variable* x *at time* t_{ij}*. In total there are* $i = 1, ..., N$ *patients, with every patient having* $j = 1, ..., n_i$ *measurements.*

Note that the association between the endogenous covariate and the outcome is only given via the linear predictor $m_i(t_{ij})$ *and its associated scaling factor* γ*. All random effects* $\mathbf{b}$ *and errors* ϵ *are independent between the outcome* y *and the endogenous covariate* x*.*

2.6.1 Association between outcome and time-varying covariate

For measuring the association between the outcome and the time-varying covariate the coefficients $\boldsymbol{\beta}_v^{lmm}$ in the LMM (definition 2.2.2) and $\beta_x^{jmm}(t)$ in the JMM (definition 2.5.2) have been obtained. Both measure the change in outcome y with a unitary increase in time-varying covariate x. To calculate such a coefficient for the JSM the expectation of the difference of the conditional distributions for a subject i should be investigated:

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

All fixed effects relating to the outcome y are independent of the value of the endogenous covariate x. This is also true for the random effects and errors in the linear predictor of the outcome. Therefore, all these values can be brought out of the conditional expectation.
This gives:

$$\begin{aligned}
E(f[y_i(t)|x_i(t) = a+1] - f[y_i(t)|x_i(t) = a]) = \\
E\left(\mathbf{w}_{yi}^\intercal \alpha_{\mathbf{y}} + \mathbf{v}_{yi}^\intercal(t)\beta_{\mathbf{y}} + \mathbf{z}_{yi}^\intercal(t)\mathbf{b}_{yi} + \epsilon_{yi}(t)\right) + E\left(\gamma m_i(t)|x_i(t) = a+1\right) + \\
- E\left(\mathbf{w}_{yi}^\intercal \alpha_{\mathbf{y}} + \mathbf{v}_{yi}^\intercal(t)\beta_{\mathbf{y}} + \mathbf{z}_{yi}^\intercal(t)\mathbf{b}_{yi} + \epsilon_{yi}(t)\right) - E\left(\gamma m_i(t)|x_i(t) = a+1\right) = \\
= \gamma E\left(m_i(t)|x_i(t) = a+1\right) - \gamma E\left(m_i(t)|x_i(t) = a\right). \qquad (2.6)
\end{aligned}$$

We are left with the expected value of $m_i(t)$ conditional on the value of $x_i(t)$. The distributions of $m_i(t), \epsilon_i(t)$ and $x_i(t)$ are:

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

Note the explicit dependence of both the mean and the variance of $m_i(t)$ and $x_i(t)$ on time.
The joint distribution $f(m_i(t), x_i(t))$ of $m_i(t)$ and $x_i(t)$ is multivariate normal, because $x_i(t)$ is a linear

combination of independent univariate normal distributions $m_i(t)$ and $\epsilon_i(t)$. We can therefore use equation 2.5 to obtain the conditional expectation of $m_i(t)$ given $x_i(t)$. This expectation will equal:

$$\begin{aligned} E(m_i(t)|x_i(t)=a) &= \mu_{m_i}(t) + \frac{Cov(m_i(t),x_i(t))}{\sigma^2_{x_i}(t)}(a-\mu_{x_i}(t)) = \\ &= \mathbf{w}_{xi}^\intercal\alpha_{\mathbf{x}} + \mathbf{v}_{xi}^\intercal(t)\beta_{\mathbf{x}} + \frac{Cov(m_i(t),x_i(t))}{\mathbf{z}_{xi}^\intercal(t)\mathbf{D}_x\mathbf{z}_{xi}(t)+\sigma^2_x}(a-\mathbf{w}_{xi}^\intercal\alpha_{\mathbf{x}} - \mathbf{v}_{xi}^\intercal(t)\beta_{\mathbf{x}}). \end{aligned}$$

The only unknown term within this equation is the covariance between $m_i(t)$ and $x_i(t)$. This covariance can easily be calculated as:

$$\begin{aligned} Cov(m_i(t),x_i(t)) &= Cov(m_i(t),m_i(t)+\epsilon_i) = Cov(m_i(t),m_i(t)) + Cov(m_i(t),\epsilon_i) = \\ &= Var(m_i(t)) = \mathbf{z}_{xi}^\intercal(t)\mathbf{D}_x\mathbf{z}_{xi}(t) \end{aligned}$$

The last step follows from the fact that $m_i(t)$ and ϵ_i are independent.
Having obtained the expected value of $m_i(t)$ given $x_i(t)$ we can continue with equation 2.6. We have:

$$\begin{aligned} &E(f[y_i(t)|x_i(t)=a+1] - f[y_i(t)|x_i(t)=a]) = \\ &\quad= \gamma\left[E\left(m_i(t)|x_i(t)=a+1\right) - E\left(m_i(t)|x_i(t)=a+1\right)\right] = \\ &\quad= \gamma\left[\frac{\mathbf{z}_{xi}^\intercal(t)\mathbf{D}_x\mathbf{z}_{xi}(t)}{\mathbf{z}_{xi}^\intercal(t)\mathbf{D}_x\mathbf{z}_{xi}(t)+\sigma^2_x}\right]. \end{aligned}$$

Definition 2.6.2 (JSM association coefficient) *The JSM time-dependent coefficient which measures the change of the outcome $y(t)$ with a unitary increase in the endogenous covariate $x(t)$ at time t will be referred to as JSM association coefficient and will be denoted by:*

$$\beta_x^{jsm}(t) := \gamma\left[\frac{\mathbf{z}_{xi}^\intercal(t)\mathbf{D}_x\mathbf{z}_{xi}(t)}{\mathbf{z}_{xi}^\intercal(t)\mathbf{D}_x\mathbf{z}_{xi}(t)+\sigma^2_x}\right] = \gamma\left[1-\frac{\sigma^2_x}{\mathbf{z}_{xi}^\intercal(t)\mathbf{D}_x\mathbf{z}_{xi}(t)+\sigma^2_x}\right].$$

Note that this coefficient has the same interpretation as the coefficient β_v^{lmm} in the LMM (see definition 2.2.2) and the coefficient $\beta_x^{jmm}(t)$ in the JMM (see definition 2.5.2). Also note that the association coefficients in both the JMM as well as the JSM are time dependent. As the random effects $\mathbf{z}_{xi}(t)$ are equal for all subjects i (as they are allowed to consist only of functions of time) there is no dependence of the JSM association coefficient $\beta_x^{jsm}(t)$ on the subject i.

As was the case with the coefficient $\beta_x^{jmm}(t)$, we have not been able to obtain a theoretical confidence interval for $\beta_x^{jsm}(t)$. Nevertheless, we can sample all necessary hyperparameters in INLA from their joint posterior distribution to construct an empirical confidence interval for $\beta_x^{jsm}(t)$.
When examining the LUMC covid dataset we shall explicitely give the coefficients involved in $\mathbf{z}_{xi}^\intercal(t)\mathbf{D}_x\mathbf{z}_{xi}(t)$ and we shall calculate the confidence intervals of $\beta_x^{jsm}(t)$, see section 6.9.

2.6.2 Reparametrization of the Joint Scaled Model

Joint scaled models can be re-parametrized to yield an expression for the combined coefficients of the outcome y. Assuming the covariates for the endogenous covariate x and the outcome y are identical, meaning $\mathbf{w}_{xi}^\intercal = \mathbf{w}_{yi}^\intercal, \mathbf{v}_{xi}^\intercal(s_{ij}) = \mathbf{v}_{yi}^\intercal(s_{ij})$ and $\mathbf{z}_{xi}^\intercal(s_{ij}) = \mathbf{z}_{yi}^\intercal(s_{ij})$, we can rewrite the expression for outcome y as:

$$\begin{aligned} y_i(s_{i,j}) &= \gamma m_i(s_{ij}) + \mathbf{w}_{yi}^\intercal\alpha_{\mathbf{y}} + \mathbf{v}_{yi}^\intercal(s_{ij})\beta_{\mathbf{y}} + \mathbf{z}_{yi}^\intercal(s_{ij})\mathbf{b}_{yi} + \epsilon_{yi}(s_{ij}) = \\ &= \left(\gamma\alpha_{\mathbf{x}}+\alpha_{\mathbf{y}}\right)\mathbf{w}_{yi}^\intercal + \left(\gamma\beta_{\mathbf{x}}+\beta_{\mathbf{y}}\right)\mathbf{v}_{yi}^\intercal(s_{ij}) + \left(\gamma\mathbf{b}_{xi}+\mathbf{b}_{yi}\right)\mathbf{z}_{yi}^\intercal(s_{ij}) + \epsilon_{yi}(s_{ij}) = \\ &= \alpha_{\mathbf{y}}'\mathbf{w}_{yi}^\intercal + \beta_{\mathbf{y}}'\mathbf{v}_{yi}^\intercal(s_{ij}) + \mathbf{b}_{yi}'\mathbf{z}_{yi}^\intercal(s_{ij}) + \epsilon_{yi}(s_{ij}), \end{aligned}$$

with $\alpha'_{\mathbf{y}} = \gamma\alpha_{\mathbf{x}} + \alpha_{\mathbf{y}}$, $\beta'_{\mathbf{y}} = \gamma\beta_{\mathbf{x}} + \beta_{\mathbf{y}}$ and $\mathbf{b}'_{yi} = \gamma\mathbf{b}_{xi} + \mathbf{b}_{yi}$.
For the random effects, they now have distribution:

$$\mathbf{b}_{yi} \sim \mathcal{N}(\mathbf{0}, \gamma^2\mathbf{D}_x + \mathbf{D}_y).$$

Inspecting the reparametrization we observe that the combined coefficients $\alpha'_{\mathbf{y}}$, $\beta'_{\mathbf{y}}$ and $\mathbf{b}'_{yi}$ are influenced by the coefficients of the endogenous covariate x with a scaling factor γ. The benefit of the combined coefficients is that they more closely resemble the coefficients obtained when fitting the LMM and JMM. This will be discussed in more detail when applying the joint models to a dataset, see section 6.4.2.

2.6.3 Lag within the Joint Scaled Model

From the JSM definition (definition 2.6.1) it is apparent that the endogenous covariate x and the outcome y do not have to be measured at the same time point. This provides for a convenient way of introducing lagged values into the model.

Definition 2.6.3 (Lagged values within the Joint Scaled Model) *Lagged values are introduced into the JSM by plugging in the linear predictors $m_{ij}(t)$ into the linear predictor of y at different time points. The model obtains the following form:*

$$\begin{cases} x_i(t_{ij}) = m_i(t_{ij}) + \epsilon_{xi}(t_{ij}) \\ y_i(s_{ij}) = \mathbf{w}_{yi}^\intercal\alpha_{\mathbf{y}} + \gamma_0 \cdot m_i(s_{ij}) + \sum_{l=1}^{L}\gamma_l \cdot m_i(s_{ij} - l) + \mathbf{v}_{yi}^\intercal(s_{ij})\beta_{\mathbf{y}} + \mathbf{z}_{yi}^\intercal(s_{ij})\mathbf{b}_{yi} + \epsilon_{yi}(s_{ij}) \end{cases}$$

with

$$m_i(t_{ij}) = \mathbf{w}_{xi}^\intercal\alpha_{\mathbf{x}} + \mathbf{v}_{xi}^\intercal(t_{ij})\beta_{\mathbf{x}} + \mathbf{z}_{xi}^\intercal(t_{ij})\mathbf{b}_{xi}.$$

Hereby we have:

- *L: The total degree of lag introduced into the model. All values of the linear predictor $m_i(s_{ij} - l)$ until time $s_{ij} - L$ are used as predictors in the model.*
- *γ_l: The scaling factors of the lagged valued $m_i(s_{ij} - l)$, with $l = 1, ..., L$.*

Note that within this framework lagged values can easily be added at any time point, without the need for the endogenous variable to be measured at that time.

2.6.4 Shared parameter models

A special case of the JSM described above is the shared parameter model. Shared parameter models allow for one or more elements of the linear predictor of the endogenous variable x to be scaled into the linear predictor of the outcome y, with different scaling factors γ.
An example of a Shared Parameter Model would be:

$$\begin{cases} x_i(t_{ij}) = \mathbf{w}_{xi}^\intercal\alpha_{\mathbf{x}} + \mathbf{v}_{xi}^\intercal(t_{ij})\beta_{\mathbf{x}} + b_{xi}^0 + b_{xi}^t t_{ij} + \epsilon_{xi}(t_{ij}) \\ y_i(s_{ij}) = \gamma_1 b_{xi}^0 + \gamma_2 b_{xi}^t s_{ij} + \mathbf{w}_{yi}^\intercal\alpha_{\mathbf{y}} + \mathbf{v}_{yi}^\intercal(s_{ij})\beta_{\mathbf{y}} + b_{yi}^0 + b_{yi}^t t_{ij} + \epsilon_{yi}(s_{ij}). \end{cases}$$

Here b_{xi}^0 and b_{xi}^t are the random intercept and random slope for the endogenous covariate, while b_{yi}^0 and b_{yi}^t fulfil a similar role for the outcome.
Note that within the shared parameter model it is not necessary to scale and copy the entire linear predictor of x. Furthermore, multiple scaling factors γ can be used for different elements of the linear predictor of x. Shared parameter models are discussed in more detail by [7]. Although in this thesis shared paramater models are not considered, they can easily be implemented in INLA in much the same way as the JSM, see section 4.6 for more details.

Chapter 3

Introduction to INLA

Many of the joint models discussed above can be implemented using standard statistical methods available within statistical programs such as R. The LMM, the multivariate model and the JMM are examples hereof. However, the JSM cannot be implemented.
When switching to the Bayesian framework joint scaled models can be implemented. Therefore, within this thesis, we shall be fitting the joint models using a Bayesian approach. For this goal we shall consider Integrated Nested Laplace Approximation (INLA) and its implementation in R with the R-INLA package [13], [6] & [17]. INLA combines the usage of Latent Gaussian Models (LGM's), Gaussian Markov Random Fields (GMRF's), Numerical methods for sparse matrices and Laplace approximation to derive approximate Bayesian inference. Overall INLA is much faster than Monte Carlo Markov Chain (MCMC) methods for Bayesian inference and does not necessitate the need for long sampling chains. Also, it has been shown that in terms of accuracy INLA is not inferior to MCMC methods [13].

3.1 Bayesian inference using INLA

3.1.1 Latent Gaussian model

INLA uses the fact that many statistical models, including the joint longitudinal models discussed in this thesis, can be rewritten as a Latent Gaussian Model (LGM). In fact, as we shall see, only models that can be rewritten as LGM's can be used within the INLA framework.
Imagine the most general form of a generalized linear mixed model:

$$y \sim \prod_{i}^{n} p(y_i|\mu_i) \quad \text{with} \quad g(\mu_i) \equiv \eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k \cdot z_{k_i} + \sum_{j=1}^{n_f} f^{(j)}(z_{ij}) + \epsilon_i.$$

Here $g()$ is the link function, α the intercept, β the regression parameters of covariates z and $f()$ function of covariates z, such as random effects. By n we denote the total number of observations.
A LGM can be constructed from this generalized linear mixed model consisting of the following elements:

- The Latent Field $\pi(\boldsymbol{\chi}|\boldsymbol{\theta_1}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}[\boldsymbol{\theta_1}])$. Note that the covariance matrix $\boldsymbol{\Sigma}$ is dependent upon hyperparameters $\boldsymbol{\theta_1}$. Here $\boldsymbol{\chi}$ contains all the element of the linear predictor: η, α, β and $f()$. Together these are assumed to be normally distributed. Thus, we can write:

  $$\pi(\boldsymbol{\chi}|\boldsymbol{\theta_1}) = \pi[(\eta, \alpha, \beta, f())\,|\boldsymbol{\theta_1}] \sim \mathcal{N}(0, \boldsymbol{\Sigma}(\boldsymbol{\theta}_1)).$$

  The dimension of $\boldsymbol{\chi}$ is usually very large (the number of observations n is large).

- Likelihood of the outcome $\pi(\mathbf{y}|\boldsymbol{\chi}, \boldsymbol{\theta}_2) \sim \prod_i p(y_i|\eta_i, \boldsymbol{\theta}_2)$. Here $\mathbf{y}$ is the observed data. We furthermore assume that the observations y_i are conditionally independent given $\boldsymbol{\chi}$ and $\boldsymbol{\theta_2}$.

- The Hyperpriors $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$. Note that although the dimension of the Latent Field $\boldsymbol{\chi}$ is usually very large, the dimension of the hyperpriors is commonly very low (e.g: The random effects are governed by just a few parameters, such as the covariances).

3.1.2 Gaussian Markov random field

The latent field $\boldsymbol{\chi} = [\eta, \alpha, \beta, f()]$ can now be thought of as a Gaussian Markov Random Field (GMRF). A GMRF is a normally distributed random vector $\boldsymbol{\chi} = (\chi_1, ..., \chi_n)$ with Markov properties, such as that for some $i \neq j$, $\chi_i \perp\!\!\!\perp \chi_j | \boldsymbol{\chi}_{-ij}$, which means that χ_i is independent of χ_j given all elements of $\boldsymbol{\chi}$ other than i and j ($\boldsymbol{\chi}_{-ij}$). The Markov properties are given in the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$, which is the inverse of the covariance matrix. Rue et all [13] showed that $\boldsymbol{Q}_{ij} = 0$ if and only if $\chi_i \perp\!\!\!\perp \chi_j | \boldsymbol{\chi}_{-ij}$. This result ensures that if in our vector $\boldsymbol{\chi} = [\eta, \alpha, \beta, f()] \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ the different elements are conditionally independent, the precision matrix $\boldsymbol{Q}$ will be sparse, allowing for easy and fast computations. For more information regarding Gaussian Markov Random Fields we refer to [12].

3.1.3 Laplace approximation

INLA uses the Laplace Approximation to estimate any distribution $g(x)$ with a normal distribution. The first 3 terms of the Taylor expansion around the mode ($\hat{x}$) are used to approximate $\log g(x)$ by:

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

The second term in this approximation, $\frac{\delta \log g(\hat{x})}{\delta x}(x - \hat{x})$, equals 0, since we are considering the derivative at the mode which is a maximum of the function.
We now estimate the variance as:

$$\hat{\sigma}^2 = -1 \Big/ \frac{\delta^2 \log g(\hat{x})}{\delta x^2}\Big|_{\hat{x}}$$

Using this we obtain:

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

With the last expression we can perform a normal approximation:

$$\int g(x)dx = \int \exp\left[\log g(x)\right] dx \approx \int \exp\left[\log g(\hat{x}) - \frac{1}{2\sigma^2}(x - \hat{x})^2\right] dx =$$
$$= \exp\left[\log g(\hat{x})\right] \cdot \int \exp\left[-\frac{1}{2\sigma^2}(x - \hat{x})^2\right] dx = \text{constant} \cdot \int \exp\left[-\frac{1}{2\sigma^2}(x - \hat{x})^2\right] dx.$$

Thus, the distribution of $g(x)$ is now approximated by a normal distribution with mean $\hat{x}$, which is found by solving $g'(x) = 0$ and with variance $\hat{\sigma}^2 = -1 \Big/ \frac{\delta^2 \log g(\hat{x})}{\delta x^2}\Big|_{\hat{x}}$, obtained at the mode $\hat{x}$.

3.1.4 Approximating the latent field

When conducting Bayesian inference we are interested in the marginals of the elements of the latent field (e.g: regression coefficients):

$$p(\chi_i|\mathbf{y}) = \int p(\chi_i, \theta|\mathbf{y})d\theta = \int p(\chi_i|\theta, \mathbf{y})p(\theta|\mathbf{y})d\theta,$$

and the elements of the hyperprior distribution (e.g: variances of random effects):

$$p(\theta_k|\mathbf{y}) = \int p(\theta|\mathbf{y})d\theta_{-k}.$$

To obtain these estimates we need to approximate $p(\chi_i|\theta, \mathbf{y})$ and $p(\theta|\mathbf{y})$.

3.1.5 Approximating $p(\theta|\mathbf{y})$

We can approximate the marginal distribution as:

$$p(\theta|\mathbf{y}) = \frac{p(\boldsymbol{\chi},\theta|\mathbf{y})}{p(\boldsymbol{\chi}|\theta,\mathbf{y})} \approx \left.\frac{p(\mathbf{y}|\boldsymbol{\chi},\theta)p(\boldsymbol{\chi}|\theta)p(\theta)}{\tilde{p}(\boldsymbol{\chi}|\theta,\mathbf{y})}\right|_{\boldsymbol{\chi}=\boldsymbol{\chi}^*(\theta)} = \tilde{p}(\theta|\mathbf{y}).$$

Here a Gaussian Laplace approximation is used for the denominator $p(\boldsymbol{\chi}|\theta,\mathbf{y})$ at the mode $x = x^*(\theta)$.

3.1.6 Approximating $p(\chi_i|\theta,\mathbf{y})$

To approximate $p(\chi_i|\theta,\mathbf{y})$ INLA has 3 options:

- Normal approximation, used in INLA when selecting the option 'Gaussian'. Here we approximate $p(\chi_i|\theta,\mathbf{y})$ using a standard Laplace approximation, and since we already computed $\tilde{p}(\boldsymbol{\chi}|\theta,\mathbf{y})$ during the exploration of $p(\theta|\mathbf{y})$ only the marginals are left to be computed. This method is by far the fastest of the three but often yields poor results.
- Laplace approximation, used in INLA when selecting the option 'Laplace'. Partitions the latent field $\boldsymbol{\chi} = [\chi_j, \boldsymbol{\chi}_{-j}]$ and uses Laplace approximation for each element χ_j in the latent field:

  $$p(\chi_j|\theta,\mathbf{y}) = \frac{p(\boldsymbol{\chi},\theta|\mathbf{y})}{p(\boldsymbol{\chi}_{-j}|\chi_j,\theta,\mathbf{y})} \propto \frac{p(\theta)p(\boldsymbol{\chi}|\theta)p(\mathbf{y}|\boldsymbol{\chi})}{\tilde{p}(\boldsymbol{\chi}_{-j}|\chi_j,\theta,\mathbf{y})}.$$

  Overall gives good results because the conditionals $p(\boldsymbol{\chi}_{-j}|\chi_j,\theta,\mathbf{y})$ are often close to normal, but is computationally expensive.
- Simplified Laplace approximation, default setting in INLA. Uses a compromise between the first 2 methods. Is computationally fast and almost always gives results very similar to the Laplace approximation.

For more information regarding Bayesian Inference with INLA we refer to [13].

3.2 Model assessment in INLA

Several methods are implemented in INLA to assess the goodness of fit of a model.

3.2.1 Marginal likelihood

The Marginal Likelihood, also called model evidence, is the probability that the data observed originates from a given model, independent of the parameters of that model (the parameters of the model are integrated out) but given a prior probability of the model. The marginal likelihood is a very convenient exclusively Bayesian model assessment tool which enables the comparison between models. In INLA the Marginal Likelihood is approximated as:

$$\tilde{\pi}(\mathbf{y}) = \int \left.\frac{\pi(\theta,\boldsymbol{\chi},\mathbf{y})}{\tilde{\pi}_G(\boldsymbol{\chi}|\theta,\mathbf{y})}\right|_{\boldsymbol{\chi}=\boldsymbol{\chi}^*(\theta)} d\theta.$$

Here $\tilde{\pi}_G(x|\theta,y)$ is the Gaussian approximation (see section 3.1.3) at the mode $\boldsymbol{\chi} = \boldsymbol{\chi}^*(\theta)$.
When considering a set of M models $\{\mathcal{M}_m\}_{m=1}^M$, the marginal likelihoods are written down as $\pi(y|\mathcal{M}_m)$. They form the basis of the Bayes factor K, which is given by:

$$K = \frac{\pi(y|\mathcal{M}_1)}{\pi(y|\mathcal{M}_2)}.$$

If supplying each model with a prior $\pi(\mathcal{M}_m)$, the Bayes factor can be written down as:

$$K = \frac{\pi(y|\mathcal{M}_1)}{\pi(y|\mathcal{M}_2)} = \frac{\pi(\mathcal{M}_1|y)\pi(\mathcal{M}_2)}{\pi(\mathcal{M}_2|y)\pi(\mathcal{M}_1)}.$$

In case of equal priors for the 2 models (the models are considered equally likely), the Bayes factor is simply the fraction of the Marginal Likelihoods of the models:

$$K = \frac{\pi(y|\mathcal{M}_1)}{\pi(y|\mathcal{M}_2)} = \frac{\pi(\mathcal{M}_1|y)}{\pi(\mathcal{M}_2|y)}.$$

The strength of evidence of a Bayes Factor $1 < K < 3.2$ is considered very weak, while only a Bayes Factor of $K > 10$ is considered to be indicative of strong evidence of model $\mathcal{M}_1$ versus model $\mathcal{M}_2$. Note that the Bayes factor is symmetrical, meaning that a Bayes Factor $1 > K > \frac{1}{3.6}$ indicates very weak evidence of model $\mathcal{M}_2$ versus $\mathcal{M}_1$, while a Bayes Factor of $K < \frac{1}{10}$ is indicative of strong evidence of model $\mathcal{M}_2$ against $\mathcal{M}_1$. INLA works with the natural logarithm of the Bayes Factor K, meaning that a difference in logarithms $1 < \ln(\mathcal{M}_1) - \ln(\mathcal{M}_2) < 1.16$ indicates very weak evidence while a difference of $\ln(\mathcal{M}_1) - \ln(\mathcal{M}_2) > 2.3$ indicates strong evidence of model $\mathcal{M}_1$ versus $\mathcal{M}_2$. Similarly, a logarithm of $1 > \ln(\mathcal{M}_1) - \ln(\mathcal{M}_2) > 0.86$ indicates weak evidence while a logarithm of $\ln(\mathcal{M}_1) - \ln(\mathcal{M}_2) < 0.43$ indicates strong evidence of model $\mathcal{M}_2$ against model $\mathcal{M}_1$.

3.2.2 Conditional predictive ordinates

The Conditional Predictive Ordinate (CPO) is computed for each observation i as:

$$CPO_i = \pi(y_i|y_{-i}).$$

It is the posterior probability of observing observation y_i when the model is fit using all data but y_i. A small value for an observation might indicate a possible outlier. INLA approximates this quantity for every observation without the need to re-analyse the model with the given observation removed. The CPO can be summarized over all the data by:

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

A smaller value indicates a better fit of the model over all observations.

3.2.3 Probability integral transform

The Probability Integral Transform (PIT) is very similar to the CPO and is computed for each continuous observation as:

$$PIT_i = \pi(y_i^{new} \leq y_i|y_{-i}).$$

The PIT measures the probability for a new observation y_i^{new} to be lower than the actual observation y_i when the model is fit using all data but y_i. Both the CPO and PIT thus apply techniques very similar to Leave-One-Out Cross-Validation (LOO CV). A very large or small PIT value for a given observation indicates a possibly surprising observation.
Over all the observations, in case of a good model, the PIT's should be approximately uniformly distributed on $[0, 1]$. The Kolmogorov Smirnov non-parametric test is used to test whether the PIT's are indeed uniformly distributed. In the remainder of this thesis we shall therefore summarize the PIT over all observations via the Kolmogorov-Smirnov test:

$$PIT = KS(PIT_i, U[0, 1]),$$

where KS is the Kolmogorov Smirnov test, PIT_i are the PIT values of all observations and $U[0, 1]$ is the standard uniform distribution.

3.2.4 DIC and WAIC

The DIC (Deviance Information Criteria) is a popular method for model selection, as it combines goodness of fit with penalization of the number of parameters used by the model. The DIC is given by:

$$DIC = D(\hat{\chi}, \hat{\theta}) + 2p_D.$$

Here $D(\hat{\chi}, \hat{\theta})$ is the model deviance, which is calculated using the posterior mean $\hat{\chi}$ and the posterior mode $\hat{\theta}$, as the distribution of θ can be severely skewed.
The effective number of parameters p_D is approximated as:

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

with n being the number of observations and Q being the precision of the Gaussian Markov Random Field, see section 3.1.2.
The Watanabe-Akaike Information Criterion is similar to the DIC, with the only difference being that the effective number of parameters p_D is calculated in a different way.

3.2.5 MSE

MSE (Mean Squared Error) is a well-known goodness of fit measure. Thus, even though MSE is not incorporated into the INLA package, we decided to use it nevertheless, by calculating it from the posterior means of fitted values. The MSE is given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

Here n is the total number of observations while y_i and $\hat{y}_i$ are the actual and fitted (posterior mean) outcomes respectively. In order to asses both the marginal and hierarchical model fit properties a total of 3 types of MSE were calculated:

- MSE_{train}: The MSE is calculated on the training set, the data-set supplied to fit the model. This metric serves to investigate which model can best fit the data at hand.
- $MSE_{subsequent}$: MSE determined on subsequent observations of subjects in the training set. This MSE was calculated in the following way:
  - For some subjects only half of the measurements are made available for fitting the model.
  - Using these measurements the model can determine the random effects of these subjects.
  - Finally, the random effects can be used to predict the omitted half of the measurements.

  This procedure supplies hierarchical results and thus allows us to inspect how well the model is able to fit the random effects of each individual separately.
  Note that this type of prediction is often needed in subject-specific questions. An example would occur when a physician wants to predict the status of a patient he is currently attending to. The physician can use the data he already acquired on the patient (and thus is subject-specific) to predict the status of the patient in the future.
- MSE_{test}: MSE calculated based on test subjects, whose data is not used to fit the model and whose random effects are therefore unknown. Here the interest is only on the marginal results, thus showing us how well the model can fit the population-averaged effects.

Chapter 4

Configuring joint models in R-INLA

An important goal of the thesis is to configure the joint models discussed in the previous chapters within the R-INLA package and test them on simulated data. In this chapter we shall introduce the models in the form that will be used for simulations in chapter 5. As the models are introduced we show how they can be rewritten into a Latent Gaussian Model (see section 3.1.1) and implemented within the R-INLA framework. During this testing phase the results obtained using R-INLA were compared, where possible, with results obtained using the R-packages nlme, lmer and MCMCglmm.

4.1 Linear mixed model

We shall start with the linear mixed model (LMM).

Definition 4.1.1 (Linear Mixed Model) *The LMM was introduced in section 2.2.1. As we are now within the Bayesian framework, we will introduce the model within the Bayesian setting. The Conditional Independence Assumption is presumed to be true thus leaving no correlation for the residual errors.*
Mathematically, the model is specified as:

$$y_{i,j} = (\beta_0 + u_{0,i}) + \beta_w \cdot w_i + \beta_v \cdot v_{i,j} + (\beta_t + u_{t,i}) \cdot t_{i,j} + \epsilon_{i,j} \qquad \text{with}$$
$$\begin{bmatrix} u_{0,i} \\ u_{t,i} \end{bmatrix} \sim \mathcal{N}_2 \left(\mathbf{0}, \begin{bmatrix} \sigma_0^2 & \sigma_{(0,t)} \\ \sigma_{(t,0)} & \sigma_t^2 \end{bmatrix} \right), \quad \epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2).$$

The following notation is used (which will be used for the remainder of this chapter):

- $y_{i,j}$*: Outcome for patient* $i = 1, ..., N$ *at time-points* $j = 1, ..., n_i$.
- β_0, β_w, β_v & β_t*: The fixed effect coefficients for the intercept, time-invariant covariate* w_i*, time-varying covariate* $v_{i,j}$ *and time* $t_{i,j}$ *(taken to be linear).*
  The priors for the fixed effect coefficients are $\beta_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$, $\beta_w \sim \mathcal{N}(\mu_w, \sigma_w^2)$, $\beta_v \sim \mathcal{N}(\mu_v, \sigma_v^2)$ & $\beta_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$ *with hyperparameters* $\mu_0, \mu_w, \mu_v, \mu_t, \sigma_0^2, \sigma_w^2, \sigma_v^2, \sigma_t^2$.
- $u_{0,i}^{(y)}$ & $u_{t,i}^{(y)}$*: Random intercept and random (time)-slope for patient i. The random effects* $u_0^{(y)}$ & $u_t^{(y)}$ *are joint normally distributed with mean* 0 *and covariance matrix* $\begin{pmatrix} \sigma_{y,0}^2 & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma_{y,t}^2 \end{pmatrix}$.
  The prior of the covariance matrix of the joint normal distribution is the Wishart distribution $\mathbf{W} \sim Wishart_2(n, \mathbf{R}^{-1})$*, where n and the elements of the matrix* $\mathbf{R}$ *are the hyperparameters.*
- $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2)$*: Errors for the outcome y.*
  The prior on the variance component σ^2 *is the* $\log(Gamma(a, b))$ *distribution, with a and b the hyperparameters.*

4.2 Priors in INLA

Having introduced the LMM within the Bayesian framework we continue with the main priors that shall be used within the thesis.

4.2.1 Fixed effect priors

The prior for fixed effects in INLA is a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, in which both the mean μ and the precision $\tau = 1/\sigma^2$ can be specified. The default values supplied by INLA are $\mu = 0$ & $\tau = 0.001$, yielding an uninformative prior.

4.2.2 Random effect priors

For the random effects we shall mainly be using the correlated random effect structure called 'iid**k**d' in INLA. Here **k** indicates the number of correlated random effects, with a maximum of 5. Suppose u and v are two correlated random effects, with a bivariate normal distribution:

$$\begin{bmatrix} u \\ v \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{W}^{-1}), \quad \text{with covariance matrix} \quad \mathbf{W}^{-1} = \begin{pmatrix} 1/\tau_u & \rho/\sqrt{\tau_u \tau_v} \\ \rho/\sqrt{\tau_u \tau_v} & 1/\tau_v \end{pmatrix}.$$

Here τ_u, τ_v (marginal precisions) and ρ (correlation coefficient) are hyperparameters.
The hyperparameters are represented internally in INLA as $\theta = (\log \tau_u, \log \tau_v, \phi)$, with $\rho = 2\frac{\exp(\phi)}{\exp(\phi)+1} - 1$. As we are more interested in the variances $\sigma_u^2 = 1/\tau_u$ & $\sigma_v^2 = 1/\tau_v$ rather than the precisions τ_u & τ_v we use the inverse of the posterior marginal distribution of the precisions to obtain the corresponding distributions of the variances.
The precision matrix $\mathbf{W}$ has a $p = 2$ dimensional Wishart distribution with support n:

$$\mathbf{W} \sim Wishart_2(n, \mathbf{R}^{-1}) \quad \text{with} \quad \mathbf{R} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \quad \text{and} \quad R_{12} = R_{21} \ \text{ due to symmetry.}$$

Some properties of the Wishart distribution are:

$$\mathbb{E}(\mathbf{W}) = n\mathbf{R}^{-1}, \quad \mathbb{E}(\mathbf{W}^{-1}) = \frac{\mathbf{R}}{n-(p+1)}.$$

The variance of the Wishart distribution has no direct overall form, but in general the variance is larger with increasing support n.
The 'iid**k**d' random effects thus have prior hyper-parameters n, R_{11}, R_{22} and $R_{21} = R_{12}$. The default values for the case $p = 2$ are $(4, 1, 1, 0)$.

4.2.3 Gaussian residuals prior

The prior for the Gaussian residuals is the log-Gamma distribution. This is also the prior for the scaling factor γ in the JSM, see definition 2.6.1. The $Gamma(a, b)$ distribution is given by:

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx),$$

with $a > 0$ the shape parameter and $b > 0$ the inverse scale parameter. The mean of the $Gamma(a, b)$ distribution equals a/b while the variance is a/b^2.
A variable u is log-Gamma distributed if $u = \log(x)$ and x is $Gamma(a, b)$-distributed. The default values for the hyperparameters of the log-Gamma prior in R-INLA are $a = 1$ & $b = 0.00005$.

The default values in INLA supply flat non-informative priors. As we will have no prior information in either the simulation study or the synthetic version of the Covid-19 dataset we shall be using these default non-informative priors throughout the thesis.

4.3 Fitting linear mixed model in R-INLA

Having defined the LMM in a Bayesian framework we show how to fit the LMM in R-INLA.
For implementing the LMM in definition 4.1.1 in INLA the following code is used:

```
LMM<-inla(y~f(v, model = "linear")+f(t, model = "linear")+
            f(w, model = "linear")+
            f(id, model = "iid2d", n=2*N)+f(time_id, t, copy="id"),
            data = data, family = "gaussian"),
```

Here `f(w, model = "linear")`, `f(v, model = "linear")` and `f(t, model = "linear")` are the fixed effects for the covariates w, v and time t respectively. Standard priors are used for these effects, see section 4.2.1.
Via the command `f(id, model = "iid2d", n=2*N)+f(time_id, t, copy="id")` we specify the multivariate normally distributed random slope and random intercept. As we have 2 jointly distributed random effects we shall be using the 'iid2d' random effect structure (described in section 4.2.2) with a total of $2N$ random effects (2 random effects for each individual: 1 random intercept + 1 random slope), with N being the total number of patients.
The copy feature tells INLA that the `time_id` term is the second of these correlated random effects. Note that the random effects vector is represented internally as one vector of length $2N$,

$$(u_{0,1}, ..., u_{0,N}, u_{t,1}, ..., u_{t,N}). \tag{4.1}$$

The variables `id` and `time_id` in the inla function call contain indexes that specify the data-rows that corresponds to the random effect in the internal representation vector 4.1. Thus, the total length of the indexes `id` and `time_id` is equal to the total length of the data at hand.

Note that to use the LMM for prediction we need values of the time-dependent covariate $v_{i,j}$. Within the simulations these values are obtained by constructing an additional LMM for the time-varying covariate $v_{i,j}$, as was described in section 2.2.2.

4.4 Multivariate joint models

Having completed the LMM as a baseline model we shall look at the first type of joint model, the multivariate joint model (section 2.4).

Definition 4.4.1 (Multivariate Joint Model) *In the multivariate joint model the association between the endogenous covariate x and the outcome y is modelled via residual errors. Mathematically, the model is given by:*

$$\begin{cases} x_i(t_{i,j}) = \beta_0^{(x)} + \beta_w^{(x)} \cdot w_i + \beta_t^{(x)} \cdot t_{i,j} + \epsilon_i^{(x)}(t_{i,j}) \\ y_i(t_{i,j}) = \beta_0^{(y)} + \beta_w^{(y)} \cdot w_i + \beta_t^{(y)} \cdot t_{i,j} + \epsilon_i^{(y)}(t_{i,j}) \end{cases} \tag{4.2}$$

with

$$\begin{bmatrix} \epsilon_{i,1}^{(x)} \\ \vdots \\ \epsilon_{i,n_i}^{(x)} \\ \epsilon_{i,1}^{(y)} \\ \vdots \\ \epsilon_{i,n_i}^{(y)} \end{bmatrix} \sim \mathcal{N}_{2n_i} \left(\mathbf{0}, \begin{bmatrix} \sigma_{x,1}^2 & \cdots & \sigma_{(x,1),(x,n_i)} & \sigma_{(x,1),(y,1)} & \cdots & \sigma_{(x,1),(y,n_i)} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \sigma_{(x,n_i),(x,1)} & \cdots & \sigma_{x,n_i}^2 & \sigma_{(x,n_i),(y,1)} & \cdots & \sigma_{(x,n_i),(y,n_i)} \\ \sigma_{(y,1),(x,1)} & \cdots & \sigma_{(y,1),(x,n_i)} & \sigma_{y,1}^2 & \cdots & \sigma_{(y,1),(y,n_i)} \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ \sigma_{(y,n_i),(x,1)} & \cdots & \sigma_{(y,n_i),(x,n_i)} & \sigma_{(y,n_i),(y,1)} & \cdots & \sigma_{y,n_i}^2 \end{bmatrix} \right) \tag{4.3}$$

The notation is similar to the LMM introduced in section 4.1.1. Note however that the time-dependent variable is denoted by x instead of v, as it now is assumed to be endogenous.

4.4.1 Multivariate joint model as LGM

Multivariate joint models cannot be put into the form of a Latent Gaussian Model as discussed in section 3.1.1, because the observations y_i are not conditionally independent given the latent field χ. This can be seen by inspecting the covariance of two measurements $y_{i,1}$ and $y_{i,2}$ on the same patient i at time points 1 and 2:

$$\begin{aligned} Cov(y_{i,1}, y_{i,2}|\chi) &= Cov\left(\beta_0^{(y)} + \beta_w^{(y)} \cdot w_i + \beta_t^{(y)} \cdot t_{i,1} + \epsilon_{i,1}^{(y)},\ \beta_0^{(y)} + \beta_w^{(y)} \cdot w_i + \beta_t^{(y)} \cdot t_{i,2} + \epsilon_{i,2}^{(y)}|\chi\right) = \\ &= Cov\left(\epsilon_{i,1}^{(y)}, \epsilon_{i,2}^{(y)}|\chi\right) = \sigma_{(i,1),(i,2)}. \end{aligned} \tag{4.4}$$

Note that the last step follows due to the definition of the latent field. The latent field $\chi = [\eta, \alpha, \beta, f()]$ contains all elements of the linear predictor except for the residual errors ϵ. Thus, conditional on the latent field the errors $\epsilon_{i,1}^{(y)}$ and $\epsilon_{i,2}^{(y)}$ are dependent and thus the outcomes $y_{i,1}$ and $y_{i,2}$ are not conditionally independent.

Thus, to fit the multivariate joint model using R-INLA a trick has to be used. The residual errors should be modelled in INLA as random effects. Simultaneously, the Gaussian errors are set fixed with a very high precision τ, in this way eliminating them.
The resulting model is thus given by:

$$\begin{cases} x_i(t_{i,j}) = \beta_0^{(x)} + \beta_w^{(x)} \cdot w_i + \beta_t^{(x)} \cdot t_{i,j} + u_{i,j}^{(x)} + \epsilon_i^{(x)}(t_{i,j}) \\ y_i(t_{i,j}) = \beta_0^{(y)} + \beta_w^{(y)} \cdot w_i + \beta_t^{(y)} \cdot t_{i,j} + u_{i,j}^{(y)} + \epsilon_i^{(y)}(t_{i,j}) \end{cases} \tag{4.5}$$

with

$$\begin{bmatrix} \mathbf{u}_i^{(x)} \\ \mathbf{u}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2j}\left(\mathbf{0}, \mathbf{\Sigma}_i\right) \text{ and } \quad \begin{bmatrix} \boldsymbol{\epsilon}_i^{(x)} \\ \boldsymbol{\epsilon}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2j}(\mathbf{0}, \tau \cdot \mathbf{I}_{2j}).$$

Here j is the number of subsequent measurements per subject, and since both the endogenous covariate $x_i(t_{i,j})$ as well as the outcome $y_i(t_{i,j})$ have j observations we have a multivariate normal with total dimension of $2j$. The residual variance-covariance matrix $\mathbf{\Sigma}_i$, which models the association between the endogenous covariate x and the outcome y, governs the random effects u instead of the errors ϵ. Note that since every measurements instance (i, j) is supplied with it's own random effect $u_{i,j}$, the model formulations in 4.2 and 4.5 are very comparable. The variance-covariance matrix $\mathbf{\Sigma}$ can be of any of the forms offered by the INLA package, such as the correlated random effect structure 'iid**k**d' mentioned in section 4.2.2. Furthermore, since the precision τ is set to be very high, the Gaussian noise is practically eliminated, thus allowing the random effect structure to fulfil the role of the residual errors in the original model parametrization.
Note that in the parametrization 4.5 the model is indeed an LGM. One can simply follow the proof given in equation 4.4 but note that the errors are now independent. The dependence between the outcomes $y_{i,1}$ and $y_{i,2}$ is now due to the random effects u, but since they are part of the latent field χ we do obtain conditional independence.

4.4.2 Fitting the multivariate joint model in INLA

The multivariate Joint Model is a multiple-likelihood model, as it incorporates the likelihoods for both the outcome y and the endogenous covariate x. In order to fit multiple-likelihood models in INLA a few tricks need to be performed. The code to run model 4.4.1 in R-INLA is given by:

```
INLA_data<-list(w_x=c(data$w, rep(NA, N_obs)),
                w_y=c(rep(NA, N_obs), data$w),
                t_x=c(data$t, rep(NA, N_obs)),
                t_y=c(rep(NA, N_obs), data$t))
```

```
INLA_data$Y<-list(c(data$x, rep(NA, N_obs)),
                  c(rep(NA, N_obs),data$y))
INLA_data$i<-c(1:(2*N_obs))
INLA_formula=Y~f(w_x, model='linear')+f(w_y, model='linear')+
              f(t_x, model='linear')+f(t_y, model='linear')+
              f(i, model="iid(2*j)d", n=2*N_obs)
set_residuals = list(list(initial=15, fixed=T), list(initial=15, fixed=T))
Multivariate<-inla(INLA_formula, family =c("gaussian","gaussian"),
                   data = INLA_data, control.family=set_residuals)
```

The first thing to note is that while N_{obs} is the total number of measurements, the data supplied to INLA has a dimension of $2N_{obs}$ instead. This is caused by the fact that the first N_{obs} entries of each variable are seen as belonging to the first likelihood (in this case the likelihood of the endogenous variable x), while the last N_{obs} entries belong to the second likelihood, the likelihood of the outcome y.
This concept is illustrated below. We see that for the vector of the endogenous covariate **x** the first N_{obs} elements are the actual values of **x**, while the last N_{obs} elements are NA's. This indicates that the first N_{obs} elements of each vector supplied to INLA apply only to the likelihood of the endogenous covariate **x**. This can be confirmed by looking at the covariate vectors in the R-INLA code 4.4.2, since the covariates relating to the model of **x** have entries only at the first N_{obs} indices. Similarly the outcome **y** contains values only in the last N_{obs} elements and so do the covariates corresponding to the outcome **y**.

$$\underset{\substack{\text{Response variables:}\\ \mathbf{x} \text{ and } \mathbf{y}}}{\begin{bmatrix} x_{1,1} & NA \\ x_{1,2} & NA \\ \vdots & \vdots \\ x_{N,n_N} & NA \\ NA & y_{1,1} \\ NA & y_{1,2} \\ \vdots & \vdots \\ NA & y_{N,n_N} \end{bmatrix}}, \quad \underset{\substack{\text{Covariate } \mathbf{w}:\\ w_x \text{ and } w_y}}{\begin{bmatrix} v_{1,1} & NA \\ v_{1,2} & NA \\ \vdots & \vdots \\ v_{N,n_N} & NA \\ NA & v_{1,1} \\ NA & v_{1,2} \\ \vdots & \vdots \\ NA & v_{N,n_N} \end{bmatrix}}, \quad \underset{\substack{\text{Time } \mathbf{t}:\\ t_x \text{ and } t_y}}{\begin{bmatrix} t_{1,1} & NA \\ t_{1,2} & NA \\ \vdots & \vdots \\ t_{N,n_N} & NA \\ NA & t_{1,1} \\ NA & t_{1,2} \\ \vdots & \vdots \\ NA & t_{N,n_N} \end{bmatrix}}$$

Within the INLA function call we set both likelihoods to be Gaussian by indicating `family =c("gaussian","gaussian")`. To ensure, as explained in section 4.4.1, that the residual errors have a very high precision, we set the precision of the residuals to be fixed at a high value: `set_residuals = list(list(initial=15, fixed=T), list(initial=15, fixed=T))`.
Lastly we set the random effects in a manner similar to what was explained in section 4.3, with a total dimension of $2j$.

It is important to note that the implementation of a multivariate marginal model within R-INLA is limited. Although INLA offers the unstructured form of the error variance-covariance matrix $\mathbf{\Sigma}$ (equation 4.3), this variance-covariance matrix is only feasible with few measurements per subject, as has been discussed in section 2.4.1. With just 4 repeated measurements per subject the unstructured form will results in 36 parameters to be estimated. Other commonly used variance-covariance structures for the errors, such as Compound symmetry & Toeplitz are not implemented within INLA. Because of these difficulties and since the interest of this thesis lies mainly with the joint mixed model and the joint scaled model we have not further looked into the implementation of the multivariate marginal model.

4.5 Joint Mixed Models

In joint mixed models (JMM) the association between the endogenous covariate and the outcome is modelled via dependence of the random effects.

Definition 4.5.1 (Joint Mixed Model) *The association at baseline and in time between the endogenous covariate and the outcome is modelled via random effects. We assume the Conditional Independence Assumption to hold, meaning that the residual errors are independent across measurements and play no role in the association. The model is given by:*

$$\begin{cases} x_i(t_{i,j}) = (\beta_0^{(x)} + u_{0,i}^{(x)}) + \beta_w^{(x)} \cdot w_i + \left(\beta_t^{(x)} + u_{t,i}^{(x)}\right) \cdot t_{i,j} + \epsilon_i^{(x)}(t_{i,j}) \\ y_i(s_{i,k}) = (\beta_0^{(y)} + u_{0,i}^{(y)}) + \beta_w^{(y)} \cdot w_i + \left(\beta_s^{(y)} + u_{s,i}^{(y)}\right) \cdot s_{i,k} + \epsilon_i^{(y)}(s_{i,k}) \end{cases} \tag{4.6}$$

with

$$\begin{bmatrix} u_0^{(x)} \\ u_0^{(y)} \\ u_t^{(x)} \\ u_t^{(y)} \end{bmatrix} \sim \mathcal{N}_4(\mathbf{0}, \mathbf{D}); \quad \begin{bmatrix} \boldsymbol{\epsilon}_i^{(x)} \\ \boldsymbol{\epsilon}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2n_i}\left(\mathbf{0}, \begin{bmatrix} \sigma_{\epsilon,x}^2 \mathbf{I}_{n_i} & 0 \\ 0 & \sigma_{\epsilon,y}^2 \mathbf{I}_{n_i} \end{bmatrix}\right).$$

We have:

- $x_i(t_{i,j})$*: Outcome for patient* $i = 1, ..., N$ *at time-points* $t_{i,j}$, $j = 1, ..., n_i^x$.
- $y_i(s_{i,k})$*: Outcome for patient* $i = 1, ..., N$ *at time-points* $s_{i,k}$, $k = 1, ..., n_i^y$.

Note that the times $t_{i,j}$ and $s_{i,k}$ at which the endogenous covariate x and the outcome y are measured can be different, as well as the total number of observations n_i^x & n_i^y.
The variance-covariance matrix $\mathbf{D}$ can be of many forms. Here we shall limit ourselves to an 'iid**k**d' variance-covariance matrix as discussed in section 4.2.2. In this case the matrix $\mathbf{D}$ has the following form:

$$\mathbf{D} = \begin{bmatrix} \sigma_{x,0}^2 & \sigma_{(x,0),(y,0)} & \sigma_{(x,0),(x,t)} & \sigma_{(x,0),(y,t)} \\ \sigma_{(y,0),(x,0)} & \sigma_{y,0}^2 & \sigma_{(y,0),(x,t)} & \sigma_{(y,0),(y,t)} \\ \sigma_{(x,t),(x,0)} & \sigma_{(x,t),(y,0)} & \sigma_{x,t}^2 & \sigma_{(x,t),(y,t)} \\ \sigma_{(y,t),(x,0)} & \sigma_{(y,t),(y,0)} & \sigma_{(y,t),(x,t)} & \sigma_{y,t}^2 \end{bmatrix}. \tag{4.7}$$

However, one can choose to set any of the covariances in the above matrix to 0, ensuring that less parameters need to be estimated. Also, one could choose a variance-covariance matrix of a completely different form, such as implementing auto-regressive random effects or random effects following a random walk. For the implementation of the JMM we have chosen for the Conditional Independence Assumption to hold, partly because of the difficulty implementing correlated errors within INLA, see section 4.4.1.

4.5.1 Measuring association

The association coefficient $\beta_x^{jmm}(t)$ for the joint mixed model as specified above is given by (section 2.5.2):

$$\beta_x^{jmm}(t) = \frac{Cov(y,x)(t)}{Var(x)(t)} = \frac{\sigma_{(y,0),(x,0)} + t\sigma_{(y,t),(x,0)} + t\sigma_{(y,0),(x,t)} + t^2\sigma_{(y,t),(x,t)}}{\sigma_{x,0}^2 + t^2\sigma_{x,t}^2 + \sigma_{\epsilon,x}^2 + 2t\sigma_{(x,t),(x,0)}}. \tag{4.8}$$

We can see that the coefficient $\beta_x^{jmm}(t)$ is indeed time-dependent. In the limit of $t \to \infty$ we have:

$$\lim_{t\to\infty} \beta_x^{jmm}(t) = \frac{\sigma_{(y,t),(x,t)}}{\sigma_{x,t}^2}. \tag{4.9}$$

4.5.2 Implementation in INLA

The code for fitting the Joint Mixed Model in R-INLA with the variance-covariance matrix structure as shown in equation 4.7 is given below and is borrowed from [14] and [15].

```
fixed.effects<-list(w_x=c(data$w, rep(NA, N_obs)),
                    w_y=c(rep(NA, N_obs), data$w),
                    t_x=c(data$t, rep(NA, N_obs)),
                    t_y=c(rep(NA, N_obs), data$t),
                    t=c(data$t, data$t))
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~f(w_x, model='linear')+ f(w_y, model='linear')+
               f(t_x, model='linear')+ f(t_y, model='linear')+
               f(Random_Intercept, model="iid4d", n=4*N)+
               f(Random_Slope, t,copy="Random_Intercept")
Joint_Mixed_Model<-inla(INLA_formula, family = c("gaussian","gaussian"),
                        data = INLA_data).
```

The JMM is a multiple likelihood model, thus necessitating the definition of vectors double the length of the number of observations N_{obs}, see section 4.4.1. The first N_{obs} elements of these vectors relate to the endogenous covariate **x**, while the second N_{obs} elements relate to the outcome **y**.
Note that N_{obs} is the total number of observations in our data while N is the total number of patients. When defining the random effects care is needed when supplying the indexes `Random_Intercept` and `Random_Slope`. We are indicating in `f(Random_Intercept, model="iid4d", n=4*N)` that we have an 'iid**4**d' correlated random effect structure with a total length of $4N$ elements. As was shown in section 4.3, the inner representation of the random effects in INLA is given by:

$$(u_{0,1}^{(x)}, ..., u_{0,N}^{(x)}, u_{0,1}^{(y)}, ..., u_{0,N}^{(y)}, u_{t,1}^{(x)}, ..., u_{t,N}^{(x)}, u_{t,1}^{(y)}, ..., u_{t,N}^{(y)}). \tag{4.10}$$

The indexes supplied tell INLA which row in the data matrix corresponds to which random effect in the internal representation (equation 4.10). This is exactly what we are doing when specifying `Random_Intercept=c(data$id, data$id+N)`. The indexes `data$id` tell INLA that the first N random effects are patient specific. Similarly, `data$id+N` ensures that the random effects with indexes ranging from $N+1$ to $2N$ are also subject specific.

4.6 Joint Scaled Models

Lastly we shall be consider the joint scaled model (JSM).

Definition 4.6.1 (Joint Scaled Model) *In the JSM the entire linear predictor of the endogenous covariate x is copied into the linear predictor of the outcome y with scaling factor γ. In both linear predictors we have dependent random intercept and random slope terms. The model is mathematically given by:*

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

with

$$\begin{bmatrix} u_{0,i}^{(x)} \\ u_{t,i}^{(x)} \end{bmatrix} \sim \mathcal{N}_2 \left(\mathbf{0}, \begin{bmatrix} \sigma_{x,0}^2 & \sigma_{x,(0,t)} \\ \sigma_{x,(t,0)} & \sigma_{x,t}^2 \end{bmatrix} \right); \quad \begin{bmatrix} u_{0,i}^{(y)} \\ u_{t,i}^{(y)} \end{bmatrix} \sim \mathcal{N}_2 \left(\mathbf{0}, \begin{bmatrix} \sigma_{y,0}^2 & \sigma_{y,(0,t)} \\ \sigma_{y,(t,0)} & \sigma_{y,t}^2 \end{bmatrix} \right);$$
$$\begin{bmatrix} \boldsymbol{\epsilon}_i^{(x)} \\ \boldsymbol{\epsilon}_i^{(y)} \end{bmatrix} \sim \mathcal{N}_{2j} \left(\mathbf{0}, \begin{bmatrix} \sigma_{\epsilon,x}^2 \mathbf{I}_{n_i^x} & 0 \\ 0 & \sigma_{\epsilon,y}^2 \mathbf{I}_{n_i^y} \end{bmatrix} \right)$$

4.6.1 Measuring association

The association coefficient $\beta_x^{jsm}(t)$ in this case is given by (section 2.6.2):

$$\beta_x^{jsm}(t) = \gamma \left[1 - \frac{Var(\epsilon^x)}{Var(x(t))} \right] = \gamma \left[1 - \frac{\sigma_{\epsilon,x}^2}{\sigma_{x,0}^2 + t^2 \sigma_{x,t}^2 + 2t\sigma_{x,(t,0)} + \sigma_{\epsilon,x}^2} \right]. \tag{4.11}$$

We can see that the coefficient $\beta_x^{jsm}(t)$ is indeed time-dependent. In the limit of $t \to \infty$ we have:

$$\lim_{t \to \infty} \beta_x^{jsm}(t) = \gamma. \tag{4.12}$$

4.6.2 Implementation in INLA

In order to implement the JSM in R-INLA a few tricks need to be used not previously discussed:

- INLA allows for random effects to be copied with a scaling factor γ into a different likelihood. For this the following syntax is used:

  ```
  f(Random_Intercept)+ f(Random_Intercept_copied, copy="Random_Intercept",
    hyper = list(beta = list(fixed=FALSE))).
  ```

  Hereby we indicate that we want to copy the element `Random_Intercept` ($u_0^{(x)}$) into a different likelihood with a non-fixed scaling factor γ, resulting in $\gamma u_0^{(x)}$.

- To ensure that all elements being copied use the same scaling factor γ, the following syntax is used:

  ```
  f(Random_Slope_copied, t, copy="Random_Slope",
    same.as = 'Random_Intercept', hyper = list(beta = list(fixed=FALSE))).
  ```

  Hereby we ensure that the `Random_Slope` ($u_t^{(x)}$) is copied with the same scaling factor γ that is used when copying and scaling the `Random_Intercept` ($u_0^{(x)}$).

- A problem within INLA is that only random effects can be copied and scaled in this way. Thus, the only way to copy and scale fixed effects is to implement them as random effects with 2 levels, one for the endogenous covariate **x** and one for the outcome **y**. As example we would have a random effect for the Intercept $\beta_0^k \sim \mathcal{N}(0, \sigma_0^2), \quad k = 1, 2$, with just 2 levels, one for the endogenous covariate x ($\beta_0^{(x)}$) and one for the outcome y ($\beta_0^{(y)}$), equal for all subjects. We are then not interested in the variance σ_0^2 of this random effect but instead in the random effect levels for both the endogenous covariate x and the outcome y. In this way all of the fixed effects are written down as random effects and copied with the same scaling parameter.
  As mentioned in [5], such an approach (with few random effect levels) gives results very comparable to simply implementing fixed coefficients per level. Particularly within the Bayesian framework the difference between fixed and random effects is more subtle than in a frequentist approach, as both fixed and random effects are random variables with a certain prior probability.

- All other aspects of the implementation of the JSM do not differ from the implementation of the JMM, see section 4.5.2.

4.6.3 Lag in a Joint Scaled Model

As discussed in section 2.6.3, implementing lag of the endogenous covariate into a JSM can be done very conveniently. Below the specification of such a lagged model is given:

$$\begin{cases} m_i(t_{i,j}) = & (\beta_0^{(x)} + u_{0,i}^{(x)}) + \beta_w^{(x)} \cdot w_i + \left(\beta_t^{(x)} + u_{t,i}^{(x)}\right) \cdot t_{i,j} \\ x_i(t_{i,j}) = & m_i(t_{i,j}) + \epsilon_i^{(x)}(t_{i,j}) \\ y_i(s_{i,k}) = & \sum_{l=0}^{L} \gamma_l \cdot m_i(s_{i,k} - l) + (\beta_0^{(y)} + u_{0,i}^{(y)}) + \beta_w^{(y)} \cdot w_i + \left(\beta_t^{(y)} + u_{s,i}^{(y)}\right) \cdot (s_{i,k}) + \epsilon_i^{(y)}(s_{i,k}) \end{cases}$$

We see that lag up to degree L is included. The implementation of lagged models in R-INLA uses much of the same tricks used when implementing the JSM.

4.7 Joint models and their implementation in other R packages

In this chapter we have focused on the implementation of joint models within the R-INLA framework. Within the thesis we have, however, also explored other R packages. These include nlme, lmer and MCMCglmm. We have examined the abilities and limitations of these R-packages in fitting joint models. For an implementation of the joint models in R-INLA, as well as nlme, lmer & MCMCglmm we refer to the R-code used within the framework of the thesis, which can be found on Github `https://github.com/georgygomon/Thesis_open`. Both nlme and lmer are well-known R packages, but MCMCglmm is less popular. For more information on the MCMCglmm package we refer to [8].
In Table 4.1 we present an overview of the different packages and their ability in fitting joint models.

Table 4.1: *The ability of select R-packages to fit joint models.*

	R-INLA	NLME	LMER	MCMCglmm
Linear Mixed Model	✓	✓	✓	✓
Multivariate Joint Model	✗[a]	✓	✗	✓
Joint Mixed Model[b]	✓	✓	✓	✓
Joint Scaled Model	✓	✗	✗	✗

[a] The implementation of Multivariate Joint Models within R-INLA is limited due to the unconventional implementation and limited number of measurements per subject that can be fitted.
[b] Under the Conditional Independence Assumption.

Chapter 5

Simulation study

Having introduced and implemented in R-INLA the linear mixed model (LMM), the joint mixed model (JMM) and the joint scaled model (JSM) we shall attempt to discover their features using a simulation study. Simulations will seek to answer several questions:

- First and foremost we shall compare the association coefficients derived for the different models. We shall thus be comparing β_v^{lmm} in the LMM, β_x^{jmm} in the JMM and β_x^{jsm} in the JSM.
  - We expect that in the case of an exogenous time-varying covariate all three association measures β_v^{lmm}, β_x^{jmm} and β_x^{jsm} will show similar results.
  - We are interested whether the coefficients show different results in the case of an endogenous covariate.
- Next we shall inspect whether the JMM and JSM perform better than the LMM, given that the data contains either an endogenous or an exogenous time-varying covariate. We shall be comparing the models by looking at several characteristics:
  - We shall be comparing the models using the model assessment tools provided in section 3.2.
  - Special attention will be given to prediction. We shall be using Mean Squared Error (MSE) to measure prediction accuracy.
- We shall be examining the goodness of fit of the models as we increase the number of subjects N in the data. We hypothesize that with a small number of subjects N the more complex JMM and JSM will not have enough information to properly fit the association between the time-varying covariate and the outcome. Therefore we expect that at low values of N the LMM might yield better results. However, we presume that as the number of subjects N increases the JMM and JSM will start to outperform the LMM and yield better results.
  - Thus, we shall be increasing the number of subjects N to determine at which level the JMM and JSM become superior as compared to the LMM. This will also serve as an indication on the data size one should have in order for endogenous models to prove useful.

5.1 Data for simulation study

To run the simulation study we use randomly generated data. The data consists out of the following elements and is shown in Table 5.1.

- A total of $N = N_1 + N_2$ patients is modelled, with the patients indexed by id. The first N_1 patients are part of the training set, while the last N_2 patients are part of the test set.

- A time-independent covariate w, randomly sampled from the $\mathcal{N}(0,1)$ distribution, unique to each patient.
- The variable time. Each patient has a total of $n = n_1 + n_2$ measurements. For patients in the training set, the first n_1 measurements are used to fit the model, while the subsequent n_2 measurements are used to calculate $\text{MSE}_{subsequent}$, see section 3.2.5. For patients in the test set, all $n = n_1 + n_2$ measurements are used to calculate the MSE on the test set.
- In order to make the data unbalanced, at each time point t there are probabilities p_1 & p_2 that the measurements of the time-dependent covariate and outcome are not observed respectively.
- The variable MSE_set. This variable indicates whether the particular data-point is part of the training set, test set or concerns subsequent measurements of subjects in the training set (train:same).

Table 5.1: *Data used for simulation study.*

id	t	w	y_observed	x_observed	MSE_set
1	1	0.3797	1	0	train
⋮	⋮	⋮	⋮	⋮	⋮
1	n_1	0.3797	1	1	train
1	n_1+1	0.3797	1	1	train: same
⋮	⋮	⋮	⋮	⋮	⋮
1	n_1+n_2	0.3797	1	0	train: same
2	1	0.9690	0	1	train
⋮	⋮	⋮	⋮	⋮	⋮
N_1	n_1+n_2	0.1632	1	1	train: same
N_1+1	1	-0.8350	1	1	test
⋮	⋮	⋮	⋮	⋮	⋮
N_1+N_2	n_1+n_2	-0.2765	1	1	test

The variables shown in table 5.1 form the basis to generating the eventual data needed for simulation. The outcome y and time-varying covariate x still need to be simulated. This happens according to either the LMM, JMM or JSM.

5.1.1 Simulating data according to the linear mixed model

To simulate data according to the LMM (see section 4.1.1) we proceed as follows:

- First we simulate $2 \times N$ (with N the number of subjects) correlated random effects $u_{0,i}^{(x)}$ and $u_{t,i}^{(x)}$ from a $\mathcal{N}_2(\mathbf{0}, \mathbf{D}_1)$-distribution and $2 \times N$ correlated random effects $u_{0,i}^{(y)}$ and $u_{s,i}^{(y)}$ from a $\mathcal{N}_2(\mathbf{0}, \mathbf{D}_2)$-distribution. The variance-covariance matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ used in the simulations are:

$$\mathbf{D}_1 = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 3 \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} 3 & 2.5 \\ 2.5 & 4 \end{pmatrix}. \tag{5.1}$$

- Next we simulate independent error terms from a $\mathcal{N}(0, 0.5)$ distribution.

- Finally we simulate the exogenous covariate v and the outcome y according to model 4.1.1. The coefficient for the exogenous covariate v is set to

$$\beta_v = 1.2,$$

while the other coefficients are set to the values shown in Table 5.2.

Table 5.2: *Fixed effects coefficients used for simulating data.*

For LMM	β_0^v	β_w^v	β_t^v	β_0^y	β_w^y	β_s^y
	5	3	1	6	4	2.2
For JMM	β_0^x	β_w^x	β_t^x	β_0^y	β_w^y	β_s^y

- The association coefficient between the time-varying covariate and the outcome is given by (see section 2.2.2):

$$\beta_v^{lmm} = 1.2.$$

5.1.2 Simulating data according to the joint mixed model

To simulate data according to the JMM (see section 4.5.1) we proceed as follows:

- First we simulate a $4 \times N$ matrix with random effects $u_{0,i}^{(x)}, u_{t,i}^{(x)}, u_{0,i}^{(y)}$ and $u_{s,i}^{(y)}$. We simulate them according to a joint normal distribution with mean 0 and variance-covariance matrix

$$D = \begin{pmatrix} 2 & 1.5 & 1.75 & 1.6 \\ 1.5 & 3 & 2 & 2.5 \\ 1.75 & 2 & 3 & 2.6 \\ 1.6 & 2.5 & 2.6 & 4 \end{pmatrix}.$$

- Next we simulate the independent error terms from a $\mathcal{N}(0, 0.5)$ distribution.
- Having simulated both the random effects and the errors we can continue by simulating the endogenous covariate x and the outcome y. We simply do so by using the linear regression formula shown in section 4.5.1 with fixed parameters set to the values shown in Table 5.2.
- The association coefficient is now given by (see section 4.5.1):

$$\beta_x^{jmm}(t) = \frac{\sigma_{(y,0),(x,0)} + t\sigma_{(y,t),(x,0)} + t\sigma_{(y,0),(x,t)} + t^2\sigma_{(y,t),(x,t)}}{\sigma_{x,0}^2 + t^2\sigma_{x,t}^2 + \sigma_{\epsilon,x}^2 + 2t\sigma_{(x,t),(x,0)}} = \frac{2 + 3.6t + 2.5t^2}{2.5 + 3t + 3t^2},$$

while the association in the limit of $t \to \infty$ equals:

$$\lim_{t\to\infty} \beta_x^{jmm}(t) = \frac{\sigma_{(y,t),(x,t)}}{\sigma_{x,t}^2} = \frac{2.5}{3} = 0.83.$$

5.1.3 Simulating data according to the joint scaled model

Simulating data according to the JSM proceeds in a very similar manner.

- First we simulate the correlated random effects $u_{0,i}^{(x)}$ and $u_{t,i}^{(x)}$ from a $\mathcal{N}_2(\mathbf{0}, \mathbf{D}_1)$-distribution and the correlated random effects $u_{0,i}^{(y)}$ and $u_{s,i}^{(y)}$ from a $\mathcal{N}_2(\mathbf{0}, \mathbf{D}_2)$-distribution. The variance-covariance matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are the same as those in the LMM, see equation 5.1.
- Next we simulate the independent error terms from a $\mathcal{N}(0, 0.5)$ distribution.

- Finally we simulate the endogenous covariate x and the outcome y according to model 4.6, using $\gamma = 1.2$ and fixed effect coefficients as given in Table 5.3. Note that using the coefficients in Table 5.3 results in the combined coefficients of the JMM (see section 2.6.2) to be exactly the same as the fixed coefficients shown in Table 5.2, e.g: $\beta_0^{y'} = \gamma\beta_0^x + \beta_0^y = 1.2 \cdot 5 + 0 = 6$.

Table 5.3: *Fixed effects coefficients used for simulating according to the JSM.*

β_0^x	β_w^x	β_t^x	β_0^y	β_w^y	β_s^y
5	3	1	0	0.4	1

- The association coefficient for this JSM is given by (see section 4.6.1):

$$\beta_x^{jsm}(t) = \gamma \left[1 - \frac{\sigma_{\epsilon,x}^2}{\sigma_{x,0}^2 + t^2\sigma_{x,t}^2 + 2t\sigma_{x,(t,0)} + \sigma_{\epsilon,x}^2}\right] = 1.2 \left[1 - \frac{0.5}{2.5 + 3t^2 + 1.5t}\right].$$

  In the limit of $t \to \infty$ we have:

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

5.2 Performing simulation study

The simulation study is performed as follows:

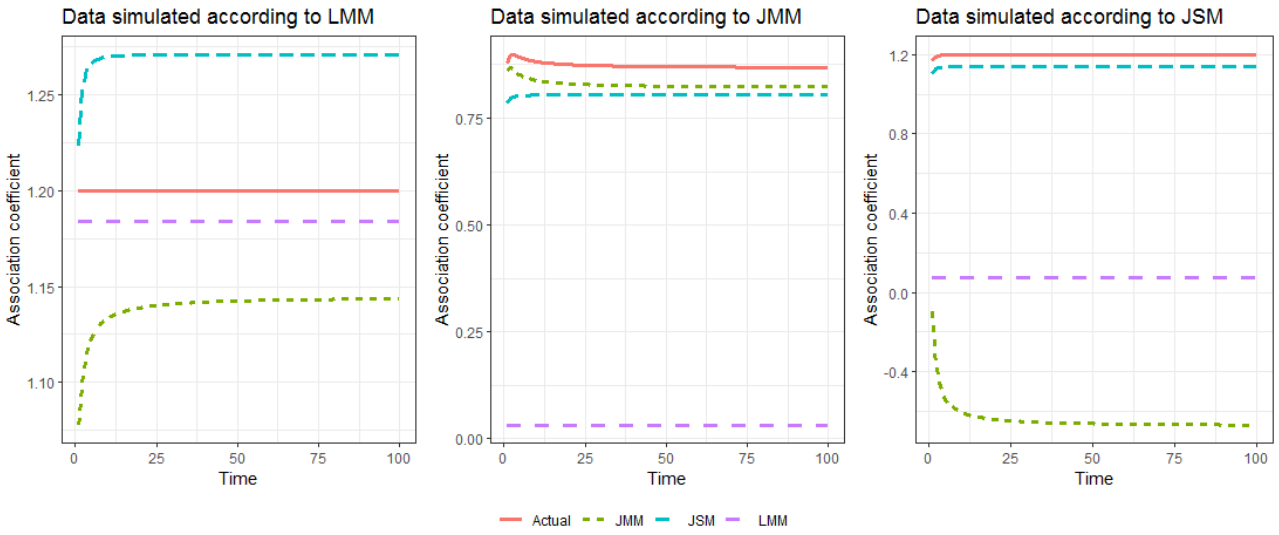
1. Set $n_1 = 15$, $n_2 = 10$, $p_1 = p_2 = 0.8$.
2. As we want to inspect the model performance with increasing number of patients N_1 in the training set, we run the simulation with values N_1 equal to 10, 25, 50, 75, 100, 150, 250, 500 and 750. We furthermore take N_2, the number of subjects in the test set, to equal $N_2 = 0.3 \cdot N_1$.
3. With a chosen number of patients N_1, simulate 250 datasets according to the LMM, JMM & JSM, see sections 5.1.1, 5.1.2 & 5.1.3 respectively.
4. Fit 3 models on every simulated dataset:
   - LMM presented in section 4.1.1.
   - JMM given in section 4.5.1.
   - JSM given in section 4.6.
5. For every simulated dataset, determine the following values:
   - Marginal Likelihood, DIC, WAIC, CPO, PIT, MSE
   - Posterior distributions of the fixed effects.
   - Significance of the association parameters.
6. Summarize all results over the 250 datasets.
7. Perform steps 3-6 with all values of N_1.

Note that all simulations have been run with the non-informative standard priors given in section 4.2.

5.3 Examining the association coefficients

First the association between the outcome and the time-varying covariate is examined. This association is given by the association coefficients $\boldsymbol{\beta}_v^{lmm}$ in the LMM, $\beta_x^{jmm}(t)$ for the JMM and $\beta_x^{jsm}(t)$ for the JSM. In Figure 5.1 are shown the values of the association coefficients for a single dataset with $N = 100$ subjects when the data is simulated according to either the LMM, JMM or JSM. Inspecting the coefficients when the data is simulated according to the LMM, we observe that the association coefficients of all models are quite close to the actual value of 1.2, with the LMM outperforming both the JMM and JSM. Also note that although the JMM and JSM coefficients are both time dependent, they quite quickly converge to the limits derived in sections 4.5.1 and 4.6.1.
When the data is simulated according to the JMM, we observe that the LMM fails to show any association. The JSM and JMM, however, are able to quite well model the association between the outcome and the time-varying covariate.
Lastly we inspect the association coefficients when the data is simulated according to the JSM. Once more we note that the LMM is not able to fit the association present in the model, while the JSM fits the actual association quite well. The JMM seem to fail in finding the association whatsoever, showing a negative association while the association is actually positive.
Note that the instances shown in Figure 5.1 are model fits for a single dataset. The reason to include them in the report is to show that the limiting behaviour is reached quite soon.

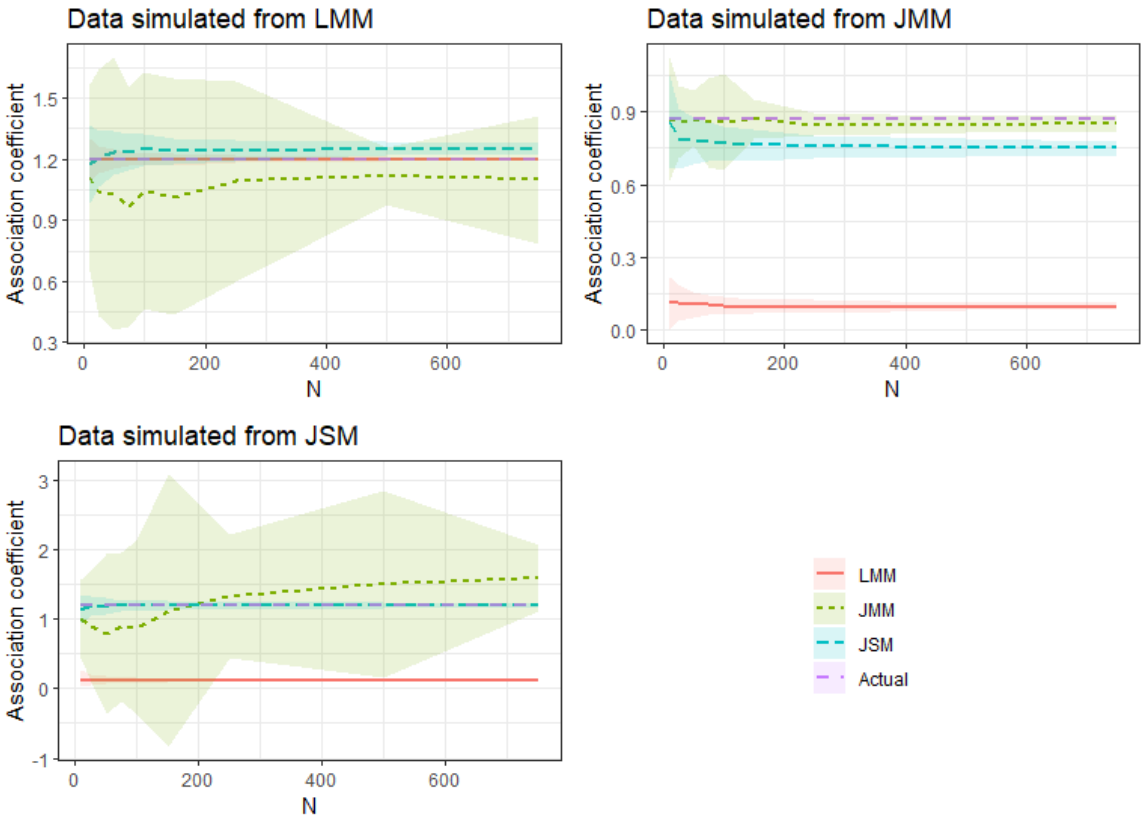
Figure 5.1: *Association coefficients β_v^{lmm}, $\beta_x^{jmm}(t)$ and $\beta_x^{jsm}(t)$ for a single dataset when the data is simulated according to either the Linear Mixed Model, Joint Mixed Model or Joint Scaled Model. Simulated data contains a total of $N = 100$ subjects.*



To better understand the behaviour of the association coefficients of the LMM, JMM & JSM we have inspected the behaviour of the association coefficients in their limit of $t \to \infty$ while increasing N (N being the number of patients), averaged over 250 instances of simulated data (see sections 5.1 & 5.2). The results are shown in Figure 5.2. Shown are the average values of the association coefficients over 250 datasets as well as the variances of these values, thus reflecting the bias-variance trade-off.
We observe that when data is simulated according to the LMM, both the LMM and the joint models can fit the actual association quite well. We hereby note that both the JMM as well as the JSM seem to be slightly biased, observing a small variance with the JSM estimate and a larger variance with the JMM estimate. In general we see that, especially with a high number of subjects N, both the JSM and to a lesser extent the JMM give the correct value for the association between the time-varying exogenous covariate and the outcome.
Next we turn to the results when data is simulated according to the JMM. We observe that in this case

the LMM does not detect the association whatsoever. Even more, the variance of the LMM estimand is very small, indicating that the association between the endogenous covariate and the outcome present in the model is never detected. Both the JMM as well as the JSM perform well in detecting the association coefficient, with both having small variances but the JMM coefficient being closer to the actual value.

Figure 5.2: *Association coefficients β_v^{lmm}, $\beta_x^{jmm}(t)$ and $\beta_x^{jsm}(t)$ in the limit $t \to \infty$ when the data is simulated according to either the Linear Mixed Model, Joint Mixed Model or Joint Scaled Model. Values are averaged over 250 datasets.*



Lastly we inspect the association coefficients when the data is simulated according to the JSM. We observe that the LMM does not show any association, and it does this with great certainty (small variance of the estimand). This was also the case with data simulated according to the JMM. The JSM shows the association present very well with a small variance. As for the JMM, even though its mean is close to the actual value, the variance of the estimand is very large, indicating that it might be of little use in finding the association between the endogenous covariate and the outcome.
The large variances observed when the data is fitted by the JMM can be explained by the fact that the covariance matrix of the JMM (see equation 4.7) is quite complex, while only 2 out of 10 elements of the covariance matrix ($\sigma_{x,t}^2$ and $\sigma_{(x,t),(y,t)}$) influence the association between the time-varying covariate and the outcome in the long run (see equation 4.9).
It should also be noted that INLA often fails to fit the elements of the covariance matrix. Table 5.4 shows the percentage of cases INLA fails to fit some elements of the covariance matrix 4.7 of the JMM. We observe that these percentages are very high when data is simulated according to either the LMM or JSM, especially with a low number of subjects N. With an increasing number of subjects N or when the data is simulated according to the JMM INLA is much better at fitting the elements of the covariance matrix 4.7.
Note that the parameters needed to calculate the association coefficients in the LMM (β_v) and JSM (γ) are always fitted, and thus the problem of an unknown association coefficient does not arise in these cases.

Table 5.4: *Table showing the percentage of cases in which an element of the covariance matrix of the JMM (see equation 4.7) was not fitted by INLA, with data simulated according to either the LMM, JMM or JSM.*

	N=10	N=25	N=50	N=75	N=100	N=150	N=250	N=500	N=750
LMM	59	58	52	38	42	36	27	22	16
JMM	13	2	1	1	1	1	0	0	0
JSM	64	65	55	54	56	55	40	35	37

In summary we can state the following. If the time-varying covariate is exogenous, the LMM is best at picking up the association between the exogenous covariate and the outcome with smallest variance of the estimand. However, both the JMM and JSM can also quite well model this association, especially in case of a large number of subjects N, when the variability of the JSM coefficient is comparable to that of the LMM. When the time-varying variable is endogenous, the LMM does not pick up the association present whatsoever. In contrast, both the JSM and the JMM do pick up this association between the outcome and the endogenous covariate. Here we note that the JMM has a very high variability of the estimand, probably caused by the complex covariance structure and by the fact that INLA fails to fit many of the covariance elements. Thus, one should take care when defining a random effect structure for a JMM model. One should also not forget that we are in the Bayesian setting, and the inability of the JMM to fit the covariance elements may be caused by the flat uninformative priors used within this simulation. When considering real data much more informative priors may be set, thereby improving the results of the JMM estimate.
It is also worth noting that one is not always interested in the association in the long run (as $t \to \infty$). In the simulation study we have inspected the behaviour as $t \to \infty$ as the results on a single dataset showed that for the proposed models this limit was reached very soon (see figure 5.1). Also, it enabled us to plot the association coefficient against the number of subjects N. We shall see when analysing the synthetic version of the Covid-19 dataset (section 6.4.2) that one is not always interested in the limiting behaviour.
In conclusion, when there is doubt over the nature of the time-varying covariate, joint models are always preferred over the LMM, especially if the number of subjects N is large. This is the case because even if the time-varying covariate is exogenous, the joint models will give a good estimate of the association. However, if the covariate is endogenous, the LMM will fail in estimating the coefficient. From our results it seems that the JSM outperforms the JMM in all situations, mainly because of the smaller variability of the estimates. However, several possible explanations for this behaviour have been given and should be examined more closely in the future.

5.4 Examining the GOF results

We shall continue by examining the goodness of fit results of the LMM, JMM and JSM. In Figure 5.3 the log marginal likelihood (see section 3.2.1) is shown for the different models when data is simulated according to the LMM, JMM or JSM.
First we note that the marginal likelihood is not comparable between the LMM and the JMM/JSM. While the LMM is a model for exclusively the outcome, the JMM and JSM are models for both the outcome and the time-varying covariate. However, we can still use the marginal likelihood to compare the goodness of fit between the JMM and the JSM. In Figure 5.3 one can see the Bayes Factor of the JMM versus the JSM plotted together with the log marginal likelihoods. In section 3.2.1 we discussed that a Bayes Factor > 2.3 is indicative of strong evidence of the JMM in favor of the JSM. We see that the marginal likelihood favours the joint model according to which the data has been generated. However, when the data is simulated according to the LMM, there is a strong preference for the JSM as opposed to the JMM. We also note that at low levels of N, the JSM is preferred even if the data is simulated according the the JMM.

Figure 5.3: *Log marginal likelihoods of the different models when the data is simulated according to either the LMM, JMM or JSM. Also shown is the Bayes Factor of the JMM versus the JSM, with a Bayes Factor>2.3 indicating a strong preference for the JMM as opposed to the JSM. Note that the marginal likelihood of the LMM is not comparable to the marginal likelihoods of the joint models.*

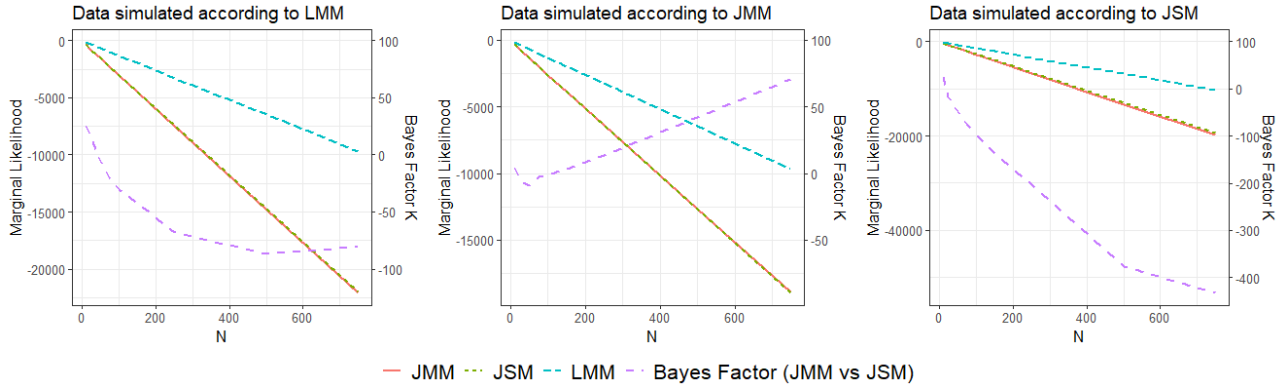
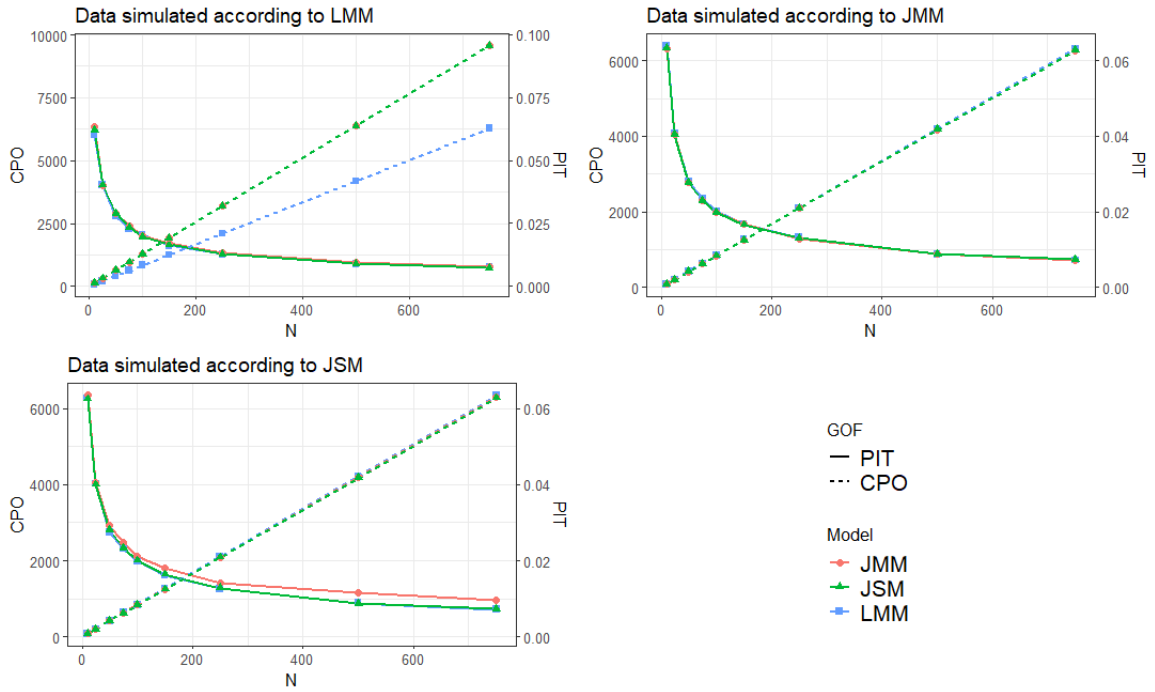


Figure 5.4 shows both the PIT and CPO for the LMM, JMM and JSM when data is simulated according to each of them. Note that both CPO and PIT are a type of Leave-one-out Cross validation (see sections 3.2.3 and 3.2.2) and thus provide a measure of goodness of fit without taking into consideration model complexity.

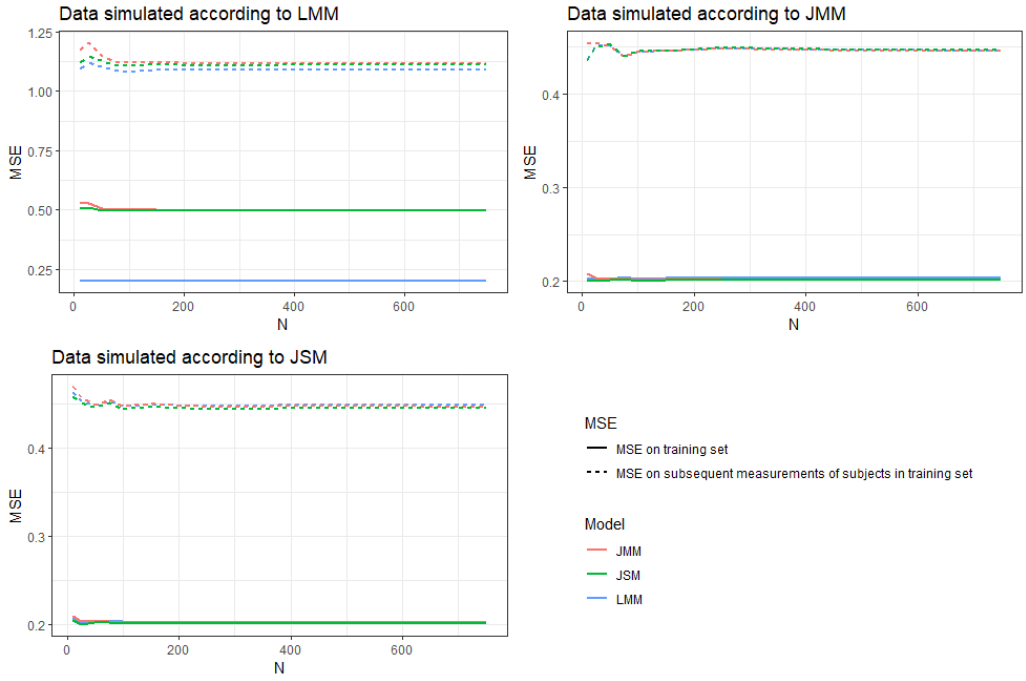
Figure 5.4: *CPO and PIT of the LMM, JMM and JSM when data is simulated according to each of them.*



We observe that according to the PIT all models perform very comparably, independent of the model according to which the data is simulated. This stands in contrast to the CPO, which shows that the LMM seems to outperform the joint models when data is simulated according to the LMM. While the PIT only measures the probability that the actual observation is larger than the fitted value, the CPO gives the posterior probability of the observed value. CPO is therefore particularly useful to detect outliers, while the PIT detects surprising fits. A possible explanation for the lower CPO value of the LMM when the data is simulated according to the LMM could be the following: When simulating according to the LMM, an outlier in the exogenous covariate yields an outlier in the outcome. The

association coefficient in the LMM can directly model this, while the association coefficients in the joint models have a more indirect approach, not giving a outlier in the outcome when the exogenous covariate is an outlier. This protects them against possible over-fitting, something we shall inspect closer when looking at the MSE.

Figure 5.5: *The Mean Squared Error on the training set and for subsequent measurements of subjects in the training set (see section 3.2.5) for the LMM, JMM and JSM. Results shown when data is generated according to either the LMM, JMM or JSM.*



In Figure 5.5 are shown the MSE on the training set as well as on subsequent measurements of subjects in the training set (see section 3.2.5). We observe that when the data is simulated according to the JMM and JSM, all models seem to perform equally well. When examining the LMM, however, we see a different picture. We note that the LMM seems to perform much better than both the joint models on the training data. This behaviour is no longer present when examining the MSE on subsequent measurements of subjects in the training set. This might indicate that the LMM is in fact overfitting the training data, thus yielding a very low MSE on the training set. This was also the postulated reason for the better CPO value of the LMM when data is simulated according to the LMM. When examining the MSE on the test set in Figure 5.6 we also see that the LMM does not perform better than any of the joint models. Thus, we are led to conclude that the better performance of the LMM on the training set when the data is simulated according to the LMM can mainly be attributed to overfitting.

Figure 5.6 shows the MSE on the test set for the different models. One thing to note is that for a small number of subjects N the JSM seems to be outperformed by the LMM and JMM, independent of what model the data is simulated according to. This effect disappears with increasing N, and with $N > 150$ the effect is negligible. The worse performance of the JSM at low levels of N was also noted in Figure 5.3. Here we saw that if the data is simulated according to the LMM , at first the JMM seems to be the preferred model, while as N increases the JSM becomes preferred.
Note further that because the values shown are obtained by averaging over just 250 datasets, the MSE results do not yield strictly decreasing lines, as would be expected with an increasing number of subjects N. However, performing a more extensive simulation was not feasible due to long running times. Nevertheless, it is remarkable that this simulation study could be done with INLA within a manageable time frame. If one would have resorted to MCMC this simulation study would

not have been possible, since every model fit (especially for larger N) would have taken hours. With INLA, a JMM fit for $N = 20$ took just a few seconds, while the fit for $N = 750$ took approximately 10 minutes. Thus, only because of INLA such an extensive simulation study could be performed. The entire simulation study took about 3 days to perform on a 8-core computer where the simulation was parallelized as efficiently as possible.

Figure 5.6: *MSE on the test set for the LMM, JMM and JSM when data is simulated according to each of them.*

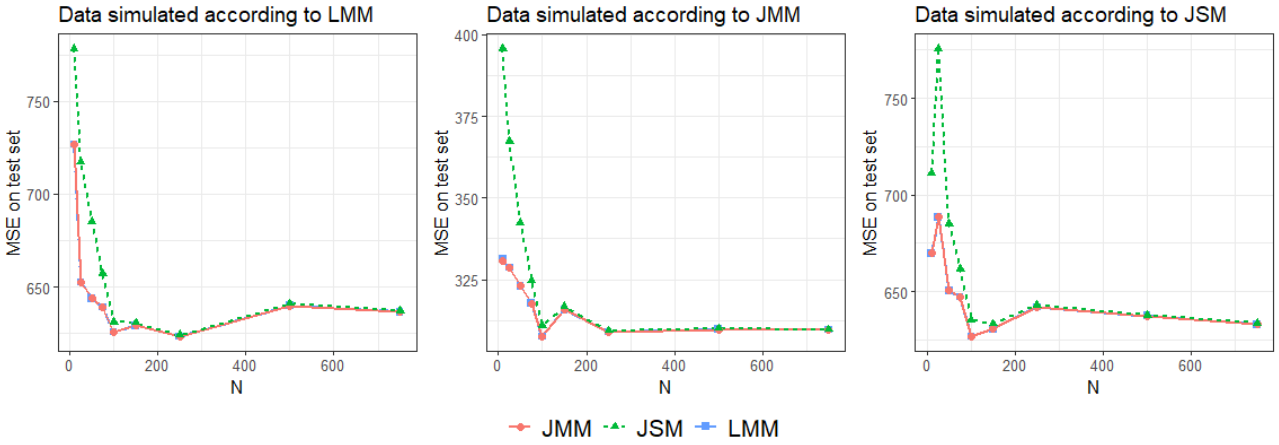
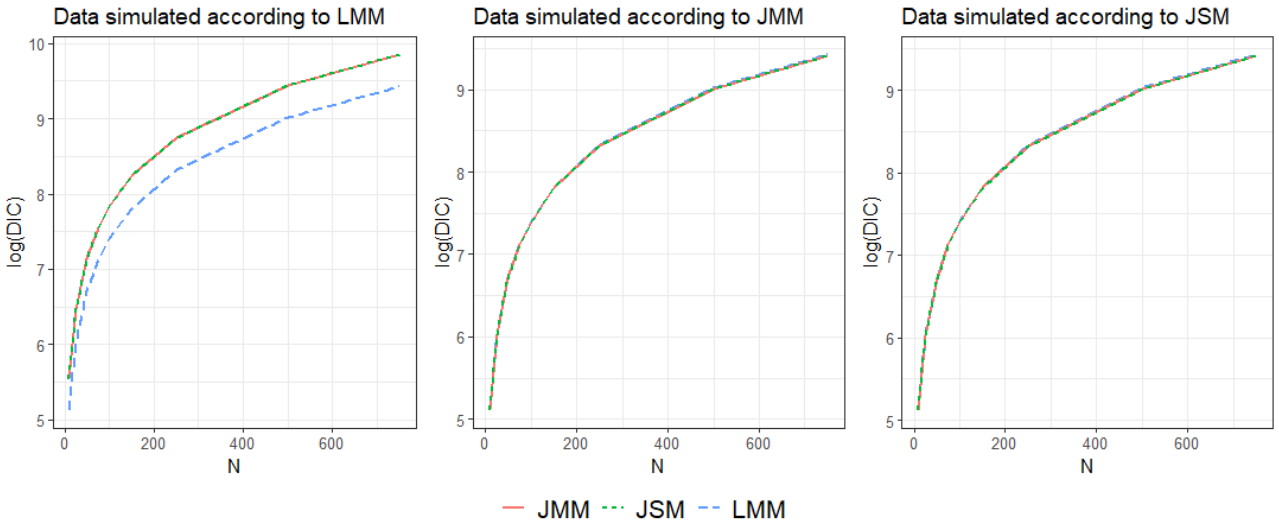


Figure 5.7 shows the logarithm of the Deviance Information Criterion (see section 3.2.4) for the different models. We observe that when the data is simulated according to the LMM, the LMM is the preferred method according to the DIC. This can be explained by the fact that DIC punishes elaborate models, and thus although the joint models fit equally well to the LMM, they are punished for their complexity. We note that when the data is simulated according to the JMM or JSM all models score equally well on the DIC, meaning that the complexness of the JMM and JSM is countered by a better fit as compare to the LMM. The WAIC is not shown but has very similar behaviour to the DIC .

Figure 5.7: *DIC of the LMM, JMM and JSM when data is simulated according to each of them.*



5.5 Recommendations based on simulation study

Based on the simulation study we can give the following recommendations for model fitting, depending on the nature of the time-varying covariate at hand:

- Time-varying covariate is exogenous: Use a LMM. The LMM will correctly give the association between the time-varying covariate and the outcome, while having a better fit overall as compared to the joint models. Also, a LMM is much easier to implement than the joint models.
- Time-varying covariate is endogenous: Use a joint model. The LMM should not be used as the association between the outcome and the endogenous covariate will not be fitted correctly. When choosing between the JMM and the JSM, attention should be given to the complexity of the random effects covariance matrix $\mathbf{D}$ of the JMM (see section 2.5.1). If the structure is complex, the JSM should be the preferred choice. Also, if the association parameter is the value of interest, the JSM should be the preferred choice, as the JSM seems to correctly fit the association coefficient in a wide variety of data. If the interest lies in lagged effect the JSM should also be preferred. In all other instances there is no strong preference, but it should be noted that the JMM is much easier to implement in INLA than the JSM.
- Nature of time-varying covariate is unknown: If the number of subjects is large enough ($N > 100$), use a joint model. If the time-varying covariate is endogenous, the joint models clearly outperform the LMM. Even if the time-varying covariate is exogenous, the joint models will give the correct association, as well as a good fit. If the number of subjects is small, use either a LMM or a joint model with a very simple structure.

Chapter 6

Application to synthetic version of LUMC Covid-19 Dataset

In chapter 4 the different techniques to model endogenous time-varying covariates have been configured within INLA. In chapter 5 it was shown within the context of a simulation study that joint models are preferable when dealing with endogenous time-varying covariates. In this chapter we shall apply the different techniques on a synthetic version of a real dataset. We shall be using a synthetic and anonymized version of the LUMC Covid-19 dataset to conduct this application.

6.1 Synthetic version of LUMC Covid-19 dataset

The portion of the LUMC Covid-19 dataset at our disposal consists out of 97 patients who were hospitalized at the LUMC because of Covid-19. The length of hospitalization differs per patient, with an average time of 13 days, a maximum of 47 and a mimimum of just 1 day. During the hospitalization the following parameters were noted:

- Day since infection: The day of infection was estimated at admission as well as possible by the physician.
- Cytokines: During hospitalization several blood cytokine levels were measured for each patient. Here we shall consider 3 cytokines, which we shall refer to as cytokines 1, 2 & 3. The cytokines are log transformed to achieve a linear relationship with the outcome. Furthermore, the cytokines chosen for this thesis have no limit of detection, thus not necessitating the use of any special treatment.
- Severity score: The severity score is a score based mainly upon clinical features to assess the state of the patient with regards to the Covid-19 infection. The severity score is available only on those days when the cytokines are measured. The maximum severity score equals 21, while the average score over all measurements in our sample is 8.

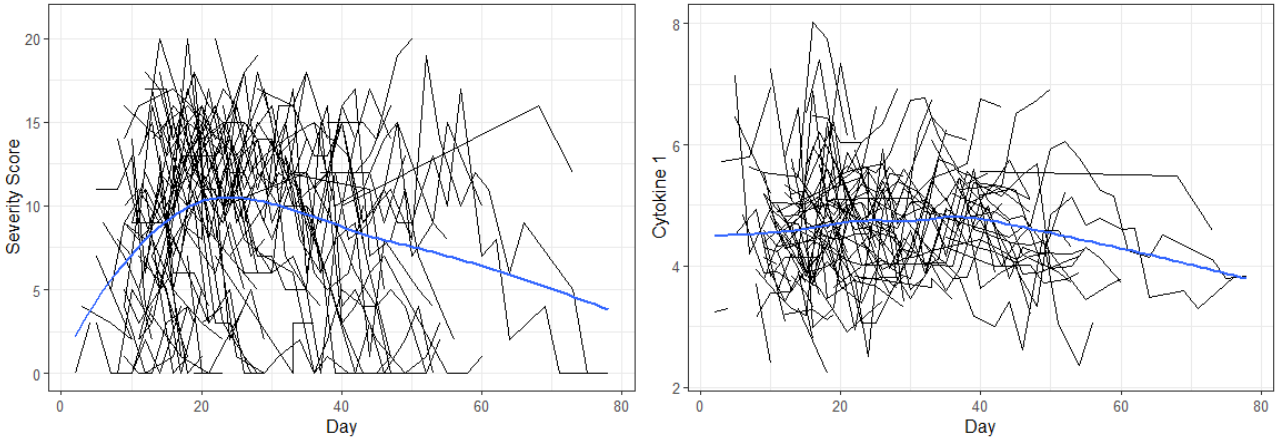
The data was synthesized by adding random noise to the cytokines as well as the severity score.
It is to be noted that a small percentage of patients in our sample died. We do not have this information at our disposal and it is not incorporated into the models.
The question of interest is the relation between the different cytokines and the severity score. Understanding this relation would greatly help physicians, as it would enable them to predict the condition of patients using cytokine levels. Thus, we would like to consider the severity score as outcome while the cytokines will have the role of time-varying explanatory variables.
It is very likely that the cytokines are in fact endogenous time-varying covariates, meaning that the cytokines at time t are, given their own history, dependent on the history of the outcome, the severity score. This is caused by the fact that the severity score of the patient probably has an effect on

cytokine levels, as within a recovering patient the cytokine levels are likely to be much lower than in a sick patient.
Thus, as the cytokines are most likely endogenous and as we have a total of about $N = 100$ patients, the recommendations from the simulation study (see section 5.5) point us towards using joint models while keeping the covariance structure **D** (see section 2.5.1) of the JMM as simple as possible.

6.2 Modelling time-progression

Before commencing the analysis of the relation between the cytokines and the severity score we will examine the data at hand more closely. In Figure 6.1 we see the spaghetti plots of the severity score and cytokine 1. In blue are fitted the corresponding loess lines, local polynomial regression lines. For the severity score one can observe an overall increasing trend up to approximately day 25, from where the severity score seems to decrease. For cytokine 1 a similar although less pronounced trend is seen: an increase until day 40 followed by a decrease. Note that the time-profiles of distinct patients differ tremendously. Both the total available time-span per patient as well as the moment of admittance differ considerably between patients. This is especially evident during the first and last days since infection, e.g: beyond day 60 there is only data for 3 patients.

Figure 6.1: *Spaghetti plots for both the severity score as well as cytokine 1 for all $N = 97$ patients in the synthetic version of the LUMC Covid-19 dataset. Plotted in blue are the corresponding Loess lines.*



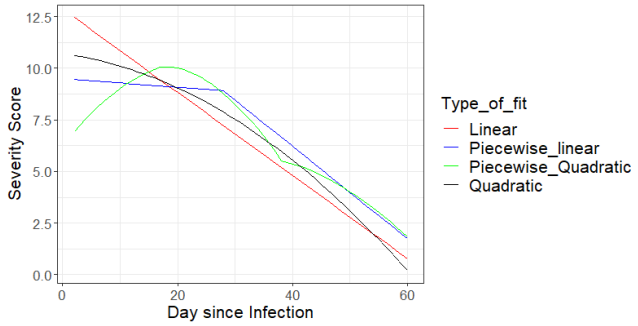
Because of the tremendous differences in time-profiles of the patients we have considered the use of other time-scales, such as the time since admittance and the normalized time since infection (normalized on the interval from 0 to 1). However, we have chosen to continue our analysis on the original time scale as it is most widely used within the clinical setting and is most useful within the setting of prediction.
As we have seen in Figure 6.1 the trend over time of the severity score is certainly not linear. Thus, a standard linear trend over time will not suffice to model the severity score. Instead we have looked at 4 options:

- Quadratic trend over time.
- Linear Splines: Piecewise linear trend over time with breakpoints. The breakpoints were obtained by applying MARS (Multivariate Adaptive Regression Spline) on the regression of the severity score on the day since infection. MARS yielded a single breakpoint at day 28. When looking at figure 6.1 we see that the loess fit also seems to indicate a change of direction at day 28.

- Quadratic Splines: Piecewise quadratic trend over time with breakpoints. As the location of breakpoints in the quadratic splines model is less important than in the linear splines model, the breakpoints were chosen at the commonly chosen locations of the 1st and 3rd quartiles of the time since infection (which correspond to days 17 and 38 respectively).
- Linear Model: Linear change over time. This model was added as a baseline model.

The four different models for time-progression are shown in Figure 6.2a. Here the fits are shown when applying linear mixed models with as outcome the severity score and as explanatory variable the day since infection. The corresponding model selection criteria can be found in Table 6.2b.

Figure 6.2: *Comparison of the Linear, Linear Splines, Quadratic and Quadratic Splines models when regressing the severity score on the day since infection.*



(a) *Time-profile of different time-progression models.*

	AIC	BIC	logLik
Linear	3134.90	3152.11	-1563.45
Piecewise Linear	3083.16	3113.28	-1534.58
Quadratic	3127.09	3148.61	-1558.55
Piecewise Quadratic	3074.33	3117.36	-1527.17

(b) *Model selection criteria for the different time-progression models.*

Table 6.2b shows that the piecewise fits outperform the non-piecewise fits. We also observe that the piecewise quadratic and piecewise linear fits perform almost equally well. For ease of interpretation and because the piecewise linear fit is a more parsimonious model we have chosen to continue with the piecewise linear fit, with as only breakpoint the 28^{th} day since infection.

Although there is no clear trend over time for the cytokines, we chose the same piecewise linear model with a breakpoint on the 28^{th} day since infection to fit all cytokines in the joint models. We have taken this approach for ease of interpretation and implementation, although we realise that a simpler model for the cytokines might have been preferable. It is also worth noting that the analysis on the synthetic version of the LUMC Covid-19 dataset is purely meant as a feasibility study on the implementation of INLA on a real dataset and no conclusions will be drawn from the results.

6.3 Modelling association between severity score and cytokines

The linear mixed model (LMM), the joint mixed model (JMM) and the joint scaled model (JSM) shall be applied on the Covid-19 dataset.

6.3.1 Linear Mixed Model

Definition 6.3.1 (Severity Score LMM) *We start with an exogenous model, in which the severity score is regressed on the time-varying cytokines without taking into account the bias invoked because of the possible endogenous nature of the cytokines. The mathematical notation of this model is:*

$$S_i(t_{i,j}) = (\beta_0^{(S)} + u_{0,i}^{(S)}) + \beta_c^{(S)} C_i(t_{i,j}) + (\beta_t^{(S)} + u_{t,i}^{(S)}) t_{i,j} + (\beta_{t28}^{(S)} + u_{t28,i}^{(S)})(t_{i,j} - 28)_+ + \epsilon_i(t_{i,j})$$

with

$$\mathbf{u}_i^{(S)} = \left(u_{0,i}^{(S)}, u_{t,i}^{(S)}, u_{t28,i}^{(S)}\right) \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma^2_{S,0} & \sigma_{S,(0,t)} & \sigma_{S,(0,t_{28})} \\ \sigma_{S,(t,0)} & \sigma^2_{S,t} & \sigma_{S,(t,t_{28})} \\ \sigma_{S,(t_{28},0)} & \sigma_{S,(t_{28},t)} & \sigma^2_{S,t_{28}} \end{bmatrix}\right),$$
$$\epsilon_i(t_{i,j}) \sim \mathcal{N}(0, \sigma^2)$$
$$\boldsymbol{\epsilon} \perp\!\!\!\perp \mathbf{u}.$$

Hereby the following notation is used:

- $S_i(t_{i,j})$*: The Severity Score of patient* $i = 1, ..., 97$*, with each patient having a total of* n_i *measusrements at times* $t_{i,j}$*, with* $j = 1, ..., n_i$*.*
- $C_i(t_{i,j})$*: The Cytokine level for patient i at time* $t_{i,j}$*. Note that* C_i *can represent any of the 3 Cytokines of interest. This will be apparent from the context.*
- $(t_{i,j} - 28)_+ = \max(0, t_{i,j} - 28)$*: The breakpoint introduced at day 28 since infection to fit a Linear Splines Model.*
- $\beta_0^{(S)}, \beta_C^{(S)}, \beta_t^{(S)}$ *and* $\beta_{t28}^{(S)}$*: The regression coefficients for the Intercept, Cytokine, Time and Time after breakpoint 28 respectively.*
  The priors for the fixed effect coefficients are $\beta_0^{(S)} \sim \mathcal{N}(\mu_0, \sigma_0^2)$*,* $\beta_C^{(S)} \sim \mathcal{N}(\mu_c, \sigma_c^2)$*,* $\beta_t^{(S)} \sim \mathcal{N}(\mu_t, \sigma_t^2)$ & $\beta_{t28}^{(S)} \sim \mathcal{N}(\mu_{t28}, \sigma_{t28}^2)$ *with hyperparameters* $\mu_0, \mu_c, \mu_t, \mu_{t28}, \sigma_0^2, \sigma_c^2, \sigma_t^2, \sigma_{t28}^2$*.*
- $u_{0,i}^{(S)}, u_{t,i}^{(S)}$ *and* $u_{t28,i}^{(S)}$*: The random effects for the Intercept, Time and Time after breakpoint 28 respectively. The prior of the joint normal distribution is the Wishart distribution* $\mathbf{W} \sim Wishart_3(n, \mathbf{R}^{-1})$*, where n and the elements of the matrix* $\mathbf{R}$ *are the hyperparameters.*
- $\epsilon_i(t_{i,j})$*: The error for patient i at time* $t_{i,j}$*.*
  The prior on the variance component σ^2 *is the* $\log(Gamma(a, b))$ *distribution, with a and b the hyperparameters.*

Note that we assume the Conditional Independence Assumption to be true, the assumption that the random effects capture all correlation and there is no correlation left for the error terms.

Standard uninformative priors shall be used when analysing the Covid-19 data, as no prior information is available regarding the cytokines. The same uninformative priors will also be used for the joint models and will therefore not be mentioned when defining the models. For the parameters of the uninformative priors we refer to section 4.2.
Lagged values can easily be included into the exogenous model by incorporating the values of covariates measured at previous time-points into the linear predictor. Note however that only the values that are actually measured can be used in this manner.

6.3.2 Joint Mixed Model

Definition 6.3.2 (Joint Mixed Model) *The JMM used to fit the Covid-19 data is given by:*

$$C_i(t_{i,j}) = (\beta_0^{(C)} + u_{0,i}^{(C)}) + (\beta_t^{(C)} + u_{t,i}^{(C)})t_{i,j} + (\beta_{t28}^{(C)} + u_{t28,i}^{(C)})(t_{i,j} - 28)_+ + \epsilon_i^{(C)}(t_{i,j}) \tag{6.1}$$
$$S_i(s_{i,j}) = (\beta_0^{(S)} + u_{0,i}^{(S)}) + (\beta_t^{(S)} + u_{s,i}^{(S)})s_{i,j} + (\beta_{t28}^{(S)} + u_{s28,i}^{(S)})(s_{i,j} - 28)_+ + \epsilon_i^{(S)}(s_{i,j}) \tag{6.2}$$

with

$$\mathbf{u}_{0,i} = \left(u_{0,i}^{(C)}, u_{0,i}^{(S)}\right) \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_{C,0}^2 & \sigma_{(C,S),0} \\ \sigma_{(C,S),0} & \sigma_{S,0}^2 \end{bmatrix}\right), \tag{6.3}$$

$$\mathbf{u}_{t,i} = \left(u_{t,i}^{(C)}, u_{t,i}^{(S)}\right) \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_{C,t}^2 & \sigma_{(C,S),t} \\ \sigma_{(C,S),t} & \sigma_{S,t}^2 \end{bmatrix}\right), \tag{6.4}$$

$$\mathbf{u}_{t28,i} = \left(u_{t28,i}^{(C)}, u_{t28,i}^{(S)}\right) \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_{C,t28}^2 & \sigma_{(C,S),t28} \\ \sigma_{(C,S),t28} & \sigma_{S,t28}^2 \end{bmatrix}\right) \tag{6.5}$$

$$\epsilon_i^{(C)}(t_{i,j}) \sim \mathcal{N}(0, \sigma_{(C)}^2), \quad \epsilon_i^{(S)}(t_{i,j}) \sim \mathcal{N}(0, \sigma_{(S)}^2) \tag{6.6}$$

$$\epsilon_i^{(C)}, \epsilon_i^{(S)} \perp\!\!\!\perp \mathbf{u}_{0,i}, \mathbf{u}_{t,i}, \mathbf{u}_{t28,i}, \quad \epsilon_i^{(C)} \perp\!\!\!\perp \epsilon_i^{(S)} \tag{6.7}$$

- *$C_i(s_{i,j})$: The Cytokine level for patient $i = 1, ..., 97$, with each patient having a total of n_i measusrements at times $s_{i,j}$, with $j = 1, ..., n_i$. Note that C_i can represent any of the 3 Cytokines of interest. This will be apparent from the context.*
- *All other notation and priors used are similar to the ones used for the LMM, see definition 6.3.1.*

Note that a simple structure was chosen for the random effects with few covariance components, as the simulation study indicated that the JMM estimate of the association coefficient might not be well fitted in case of a complex covariance structure of the random effects.

Estimating association in the Joint Mixed Model

In order to estimate the conditional association between the cytokine and the severity score (see section 2.5.1) the JMM should be written down as a bivariate normal distribution. For this we need to calculate the variances of the endogenous covariate and the outcome, as well as the covariance between them. As was mentioned in section 2.5.1, these quantities are dependent upon time. We shall give as example the calculation of the variance of the outcome, as well as the covariance between the outcome and the cytokine:

$$\begin{aligned}
Var\left[S_i(t_{i,j})\right] &= Var\left[u_{0,i}^{(S)} + u_{t,i}^{(S)} t_{i,j} + u_{t28,i}^{(S)}(t_{i,j} - 28)_+ + \epsilon_i^{(S)}(t_{i,j})\right] \overset{1}{=} \\
&= Var\left(u_{0,i}^{(S)}\right) + t_{i,j}^2 Var(u_{t,i}^{(S)}) + \left[(t_{i,j} - 28)_+\right]^2 Var\left(u_{t28,i}^{(S)}\right) + Var\left(\epsilon_i^{(S)}(t_{i,j})\right) = \\
&= \sigma_{S,0}^2 + t_{i,j}^2 \sigma_{S,t}^2 + \left[(t_{i,j} - 28)_+\right]^2 \sigma_{S,t28}^2 + \sigma_{(S)}^2 \\
Cov\left[S_i(s_{i,j}), C_i(t_{i,j})\right] &= Cov\left[u_{0,i}^{(C)} + u_{t,i}^{(C)} t_{i,j} + u_{t28,i}^{(C)}(t_{i,j} - 28)_+, u_{0,i}^{(S)} + u_{t,i}^{(S)} t_{i,j} + u_{t28,i}^{(S)}(t_{i,j} - 28)_+\right] \overset{1}{=} \\
&= Cov(u_{0,i}^{(C)}, u_{0,i}^{(S)}) + t_{i,j}^2 Cov(u_{t,i}^{(C)}, u_{t,i}^{(S)}) + \left[(t_{i,j} - 28)_+\right]^2 Cov(u_{t28,i}^{(C)}, u_{t28,i}^{(S)}) = \\
&= \sigma_{(C,S),0} + t_{i,j}^2 \sigma_{(C,S),t} + \left[(t_{i,j} - 28)_+\right]^2 \sigma_{(C,S),t28}.
\end{aligned}$$

In the above derivations the equalities denoted by 1 follow from the fact that all random effects and errors are independent. Now we can construct the bivariate normal distribution of both the cytokine and the severity score. Note that all elements in the bivariate normal distribution are time-dependent.

$$f(S(t_{i,j}), C(t_{i,j})) = \mathcal{N}_2\left(\begin{bmatrix} \beta_0^{(S)} + \beta_t^{(S)} t_{i,j} + \beta_{t28}^{(S)}(t_{i,j} - 28)_+ \\ \beta_0^{(C)} + \beta_t^{(C)} t_{i,j} + \beta_{t28}^{(C)}(t_{i,j} - 28)_+ \end{bmatrix}, \mathbf{\Sigma}\right)$$

with

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{S,0}^2 + t_{i,j}^2 \sigma_{S,t}^2 + \left[(t_{i,j} - 28)_+\right]^2 \sigma_{S,t28}^2 + \sigma_{(S)}^2 & \sigma_{(C,S),0} + t_{i,j}^2 \sigma_{(C,S),t} + \left[(t_{i,j} - 28)_+\right]^2 \sigma_{(C,S),t28} \\ \sigma_{(C,S),0} + t_{i,j}^2 \sigma_{(C,S),t} + \left[(t_{i,j} - 28)_+\right]^2 \sigma_{(C,S),t28} & \sigma_{C,0}^2 + t_{i,j}^2 \sigma_{C,t}^2 + \left[(t_{i,j} - 28)_+\right]^2 \sigma_{C,t28}^2 + \sigma_{(C)}^2 \end{bmatrix}.$$

It was derived in section 2.5.1 that a unitary increase in the cytokine at time t will give an increase of the severity at time t equal to:

$$\beta_C^{jmm}(t) = \frac{Cov(y,x)(t)}{\sigma_x^2(t)} = \frac{\sigma_{(C,S),0} + t_{i,j}^2\sigma_{(C,S),t} + \left[(t_{i,j}-28)_+\right]^2\sigma_{(C,S),t28}}{\sigma_{C,0}^2 + t_{i,j}^2\sigma_{C,t}^2 + \left[(t_{i,j}-28)_+\right]^2\sigma_{C,t28}^2 + \sigma_{(C)}^2}. \tag{6.8}$$

Within the context of the Covid-19 dataset our interest does not lie with the limiting behaviour of $\beta_C^{jmm}(t)$ as $t \to \infty$. Instead we are interested in the association between the severity score and the cytokine in the time-span of days 1 until 50 after infection. This is the time-span at which patients develop Covid-19 symptoms and are hospitalized.
We shall examine the association between the outcome and the endogenous covariate in the JMMin several ways:

- We shall inspect the credible intervals and therefore the significance of the covariance terms involved in equations 6.3-6.5: $\sigma_{(C,S),0}, \sigma_{(C,S),t}$ & $\sigma_{(C,S),t28}$. These are of particular interest since they are the terms responsible for the association between the outcome and the endogenous covariate. Also, credible intervals are available for these quantities.

- We shall be giving the expected value and credible intervals of the coefficient $\beta_C^{jmm}(t)$ as a function of time. The credible intervals are estimated by sampling one million values from the joint posterior density of the parameters involved and calculating the necessary expression.

6.3.3 Joint Scaled Model

Definition 6.3.3 (Joint scaled model) *Lastly, the joint scaled model (JSM) is given by:*

$$\begin{cases} m_i(t_{i,j}) = & (\beta_0^{(C)} + u_{0,i}^{(C)}) + (\beta_t^{(C)} + u_{t,i}^{(C)})t_{i,j} + (\beta_t^{(C),28} + u_{t,i}^{(C),28})(t_{i,j}-28)_+ \\ C_i(t_{i,j}) = & m_i(t_{i,j}) + \epsilon_i^{(C)}(t_{i,j}) \\ S_i(s_{i,j}) = & \gamma_0 m_i(s_{i,j}) + \sum_{l=1}^{L}\gamma_l \cdot m_i(s_{ij}-l) + \\ & +(\beta_0^{(S)} + u_{0,i}^{(S)}) + (\beta_s^{(S)} + u_{s,i}^{(S)})s_{i,j} + (\beta_s^{(S),28} + u_{s,i}^{(S),28})(s_{i,j}-28)_+ + \epsilon_i^{(S)}(s_{i,j}) \end{cases}$$

with

$$\mathbf{u}^C = \left(u_{0,i}^{(C)}, u_{t,i}^{(C)}, u_{t28,i}^{(C)}\right) \sim \mathcal{N}_3\left(\mathbf{0}, \begin{bmatrix} \sigma_{C,0}^2 & \sigma_{C,(0,t)} & \sigma_{C,(0,t_{28})} \\ \sigma_{C,(t,0)} & \sigma_{C,t}^2 & \sigma_{C,(t,t_{28})} \\ \sigma_{C,(t_{28},0)} & \sigma_{C,(t_{28},t)} & \sigma_{C,t_{28}}^2 \end{bmatrix}\right),$$

$$\mathbf{u}^S = \left(u_{0,i}^{(S)}, u_{t,i}^{(S)}, u_{t28,i}^{(S)}\right) \sim \mathcal{N}_3\left(\mathbf{0}, \begin{bmatrix} \sigma_{S,0}^2 & \sigma_{S,(0,t)} & \sigma_{S,(0,t_{28})} \\ \sigma_{S,(t,0)} & \sigma_{S,t}^2 & \sigma_{S,(t,t_{28})} \\ \sigma_{S,(t_{28},0)} & \sigma_{S,(t_{28},t)} & \sigma_{S,t_{28}}^2 \end{bmatrix}\right),$$

$$\epsilon_{i,j}^{(C)} \sim \mathcal{N}(0, \sigma_{(C)}^2), \quad \epsilon_{i,j}^{(S)} \sim \mathcal{N}(0, \sigma_{(S)}^2)$$

$$\epsilon_{i,j}^{(C)}, \epsilon_{i,j}^{(S)} \perp\!\!\!\perp \mathbf{u}^C, \mathbf{u}^S, \quad \epsilon_{i,j}^{(C)} \perp\!\!\!\perp \epsilon_{i,j}^{(S)}$$

with:

- $m_{i,j}$*: Linear predictor of the endogenous covariate C for patient* i, $i = 1, ..., N$ *at time* $t_{i,j}$.
- γ_0*: Scaling factor for the linear predictor* $m_i(s_{i,j})$ *at time point* $s_{i,j}$.
- γ_1*, ...,* γ_l*: Scaling factors for the lagged linear predictors* $m_i(s_{ij}-l)$*. Lagged linear predictors are included up to lag of degree* L.

The notation and priors used are similar to the ones used for the LMM, see definition 6.3.1.

Estimating association in the Joint Scaled Model

For the JSM, the association coefficient $\beta_x^{jsm}(t)$ introduced in section 2.6.1 is given by the following quantity:

$$\beta_x^{jsm}(t) = \gamma \frac{Var(m_i(t))}{Var(C_i(t_{i,j}))} = \gamma \left(1 - \frac{\sigma^2_{(C)}}{Var(C_i(t_{i,j}))}\right) \tag{6.9}$$

with

$$Var(C_i(t_{i,j})) = \sigma^2_{C,0} + t^2_{i,j}\sigma^2_{C,t} + \left[(t_{i,j} - 28)_+\right]^2 \sigma^2_{C,t_{28}} + \sigma^2_{(C)} + \tag{6.10}$$

$$+ 2t_{i,j}\sigma_{C,(t,0)} + 2(t_{i,j} - 28)_+ \sigma_{C,(t_{28},0)} + 2t_{i,j}\left[(t_{i,j} - 28)_+\right]\sigma_{C,(t,t_{28})}. \tag{6.11}$$

A clear time-dependence of the coefficient $\beta_x^{jsm}(t)$ is to be noted. The derivation of $Var(C_i(t_{i,j}))$ proceeds in much the same way as the derivation shown in section 6.3.2 and is thus omitted.
To examine the association between the outcome and the endogenous covariate in the COVID dataset we shall be examining 2 quantities:

- We shall be looking at the estimate of γ and its corresponding credible interval. This is of interest to us as $lim_{t\to\infty}\beta_x^{jsm}(t) = \gamma$.
- We shall be giving an estimate and corresponding credible intervals for the coefficient $\beta_x^{jsm}(t)$ as function of the time t. The credible intervals are estimated by sampling 1 million values from the joint posterior density of the parameters involved and calculating the necessary expression.

6.4 Results

6.4.1 Presence of a lagged effect

First we inspect whether a lagged association is present between the cytokines and the severity score. We inspect this by considering the joint scaled model (JSM) as well as the linear mixed model (LMM) with lag of order 5. For the LMM, only observed measurements are used as lagged values, thus not necessitating the construction of a LMM for the cytokine. The models are fit for each cytokine separately. The results can be seen in Table 6.1. Here $\gamma_1^{C1}, ..., \gamma_5^{C1}$ are the lagged effects of order 1 up to 5 for Cytokine 1. Similar notation is used for cytokines C2 and C3. Shown in the table are the estimates with corresponding 95% credible intervals. We observe that none of the lagged values for any of the cytokines are significant, thus meaning that a lagged association is most probably not present within the Covid-19 dataset.
In the dataset, only 0.5% of measurements had a lagged value of order 1: this is a measurement performed 1 day before. This in contrast to lagged values of order 2 and 5: a total of 46% and 44% of measurements had such lagged values respectively. The limited availability of lagged values a day beforehand can be seen in Table 6.1, as the LMM cannot calculate the coefficients of lag 1 for any of the cytokines. As was mentioned in section 2.2.1, the LMM can only fit lagged values that are actually measured. In order to fit lagged values at any time point an additional independent LMM has to be constructed for the cytokine. In contrast, the JSM can supply coefficients at any lagged time-point because of the scaled linear predictor (see section 2.6.3). Thus, in Table 6.1 the lagged values of order 1 are shown for the JSM but can not be calculated for the LMM.

Table 6.1: *Table showing the results of fitting a lagged model with lag of degree 5 on the Covid-19 data using the LMM and JMM. The models are fitted on each of the cytokines separately. The results are shown for all 3 cytokines, with $\gamma_1^{C1}, ..., \gamma_5^{C1}$ being the lagged values of degree 1 to 5 for cytokine C1. The notation is similar for cytokines C2 and C3.*

	Linear Mixed Model	Joint Scaled Model
γ_1^{C1}	NA	0.13 [-0.47 , 0.74]
γ_2^{C1}	0.12 [-0.04 , 0.27]	0.31 [-0.12 , 0.75]
γ_3^{C1}	0.07 [-0.12 , 0.25]	0.09 [-0.39 , 0.56]
γ_4^{C1}	0.06 [-0.13 , 0.26]	0.28 [-0.2 , 0.75]
γ_5^{C1}	0.12 [-0.04 , 0.28]	0.22 [-0.22 , 0.67]
γ_1^{C2}	NA	-0.21 [-0.79 , 0.38]
γ_2^{C2}	0.07 [-0.05 , 0.18]	-0.36 [-0.82 , 0.11]
γ_3^{C2}	0.05 [-0.09 , 0.19]	-0.13 [-0.62 , 0.37]
γ_4^{C2}	-0.02 [-0.16 , 0.13]	-0.16 [-0.67 , 0.35]
γ_5^{C2}	0.05 [-0.07 , 0.17]	-0.15 [-0.63 , 0.33]
γ_1^{C3}	NA	0.12 [-0.47 , 0.7]
γ_2^{C3}	0.07 [-0.06 , 0.21]	0.04 [-0.46 , 0.53]
γ_3^{C3}	0.05 [-0.11 , 0.21]	0.17 [-0.35 , 0.69]
γ_4^{C3}	0.01 [-0.16 , 0.18]	0.2 [-0.33 , 0.72]
γ_5^{C3}	0.07 [-0.07 , 0.21]	0.27 [-0.23 , 0.77]

6.4.2 Analysing cytokine 1

We have shown that there does not seem to be a lagged association between any of the cytokines and the severity score. We shall thus focus on models without lag. We shall first examine the association between the severity score and cytokine 1. In Table 6.2 are shown the goodness of fit measures for the LMM, JMM and JSM in fitting the severity score given cytokine 1. The models are implemented as shown in section 6.3, with as outcome the severity score and with $C_i(t_{i,j})$ being cytokine 1. For all GOF metrics (except for the marginal likelihood) a lower score indicates a better fit.
Note that the marginal likelihood cannot be used to compare the LMM to the joint models, as the joint models contain likelihoods for both the time-varying cytokine as well as the outcome. The joint models are comparable between themselves via the marginal likelihood, and Table 6.2 shows that the JSM performs significantly better than the JMM according to the marginal likelihood, as a difference in log marginal likelihood of greater than 2.1 indicates a strong preference for the JSM as compared to the JMM (see section 3.2.1).
However, overall we observe that the JMM appears to best fit the data at hand, as it scores best on all metrics except for the marginal likelihood, the PIT and the MSE on the test set. It is remarkable that both joint models score better on both the DIC and WAIC as compared to the LMM, as these metrics combine goodness of fit with a penalization of model complexity. It thus appears that although the joint models are more complex than the LMM, they compensate by a better fit.
The MSE tells us something about how well the models can fit the data at hand (MSE train), predict future instances of subjects in the data set (MSE subsequent) and fit similar but new data (MSE test). To calculate the MSE on the test set and on future observations of subjects in the training set we have performed 50-fold cross validation. In each of the cross-validation samples, 15 subjects were excluded from the training set to serve as test set, while another 15 subjects were used to to calculate MSE subsequent by excluding half their measurements from the training set to serve as unknown future observations.
We observe that the joint models outperform the LMM on all MSE measures. This was to be expected for the training set (where it could be attributed to overfitting), but a better fit is also observed for the test set and for subsequent observations. Thus, the better fit does not seem to be due to overfitting

but because the more complex joint models truly capture the underlying mechanisms of the processes involved.

Table 6.2: *Goodness of fit measures for fitting the LMM, JMM and JSM on the Covid-19 dataset when regressing the severity score on cytokine 1.*

	LMM	JMM	JSM
Marginal likelihood	-1579	-2376	-2356
DIC	2920	2880	2907
WAIC	2915	2876	2905
PIT	0.0240	0.0352	0.0360
CPO	1466	1452	1461
MSE (train)	8.12	7.13	7.71
MSE (subsequent)	24.87	22.80	28.43
MSE (test)	46.20	45.62	39.36

In Table 6.3 the coefficients for the severity score of the different models are shown (for the JSM the combined coefficients are shown to make the comparison easier, see section 2.6.2). We note that all models indicate a significant intercept as well as a significant decrease with time after $t = 28$. One thing to note is that although the intercept is significant in all models, the value of this parameter differs greatly between the joint models and the LMM. This is caused by the different model formulations. In the LMM, a large value of the cytokine at $t = 0$ greatly increases the severity score at $t = 0$, without the need for a large intercept. In the data the average value of cytokine 1 equals 4.7. This gives an average of value at $t = 0$ of $4.7\beta_{cy} + \beta_0 = 9.658$. This is very comparable to the intercept values observed in the mixed models.

Table 6.3: *Coefficients obtained by the LMM, JMM and JSM when regressing the severity score on cytokine 1.*

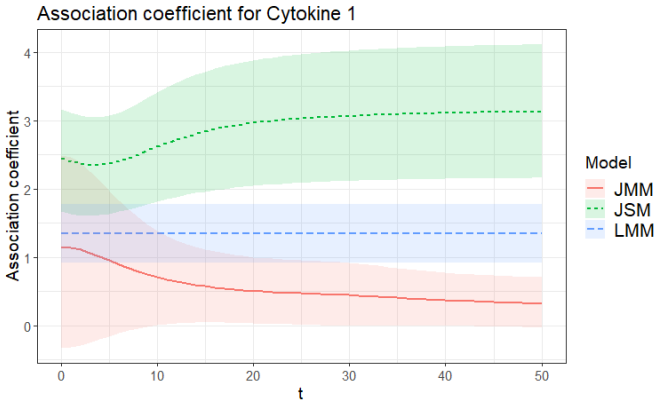
	LMM	JMM	JSM
β_0	**3.36 (0.75, 5.95)**	**10.67 (8.98, 12.4)**	**10.58 (6.08, 15.76)**
β_t	-0.06 (-0.16, 0.04)	**-0.12 (-0.21, -0.02)**	-0.06 (-0.19, 0.05)
β_{t28}	**-0.29 (-0.47, -0.09)**	**-0.23 (-0.4, -0.05)**	**-0.27 (-0.47, -0.05)**

In Figure 6.3a the trajectories of the association coefficients for the LMM (β_v^{lmm}), the JMM ($\beta_x^{jmm}(t)$) and the JSM ($\beta_x^{jsm}(t)$) with their corresponding credible intervals are shown. The credible intervals for both joint models were constructed by sampling from the joint posterior distribution of the elements involved. Note the time-dependency of the association coefficients in the joint models. In Figure 6.3a the association coefficients are shown from $t = 0$ until time $t = 50$, as this is the window of interest during which patients have symptoms relating to Covid-19. All models give a positive association between the cytokine and the outcome. For the JSM and the LMM the association is significant throughout the entire time window. For the JMM the association is only significant from day $t = 9$ to day $t = 29$.

To better examine the association within the JMM Table 6.3b shows the covariance elements of the JMM random effects, see section 6.3.2. These covariance elements are the parameters that determine the dependence of the outcome on the cytokine. These are also the parameters that constitute $\beta_x^{jmm}(t)$, the association coefficient in the JMM. We see that the covariance element of the random time effect is almost significant. This causes the significant association of the JMM from time $t = 9$ until $t = 29$, as before $t = 9$ the insignificant covariance term of the random intercept ($\sigma_{(C,S),0}$) governs the association, while after $t = 29$ the term $\sigma_{(C,S),t28}$ is responsible for an insignificant association. Note that the time span between days 9 and 29 is the time span we are most interested in, as during this time most patients are admitted to the hospital.

We see that the JSM and JMM show opposite behaviour in the association with time. The JSM seems to point at an increasing association with time, while the association for the JMM decreases with time. The disagreement between the models might be explained by a phenomena we observed in the simulation study. In the simulation study we saw that although the association coefficient of the JMM is unbiased, it is very variable. Thus, the association might have been wrongly fitted. Nevertheless, all 3 models indicate a positive association between the cytokine and the outcome.
Note that in our model specification the association in the LMM is independent of time. This does not necessarily need to be the case for a LMM, as an interaction between the cytokine and time can be added to the model.

Figure 6.3: *Association coefficients of the LMM, JMM and JSM obtained when regressing the severity score on cytokine 1. Also shown are the covariance elements that make up the JMM association coefficient* $\beta_x^{jmm}(t)$.



(a) *Association coefficients of the LMM, JMM and JSM over time with corresponding credible intervals.*

	JMM
$\sigma_{(C,S),0}$	0.35 (-0.11 , 0.71)
$\sigma_{(C,S),t}$	0.19 (-0.02 , 0.4)
$\sigma_{(C,S),t28}$	0.1 (-0.25 , 0.45)

(b) *Credible intervals for the covariance coefficients of the JMM, see section 6.3.2.*

6.4.3 Analysing cytokines 2 & 3

Table 6.4 shows the model fits when using either cytokine 2 or 3 to fit the severity score. Over all GOF measures, the JMM seems to perform best. The exceptions to this are the marginal likelihood (which favors the JSM) as well as the PIT (which favors the LMM). Similar results were noted when analysing cytokine 1 (see Table 6.2).

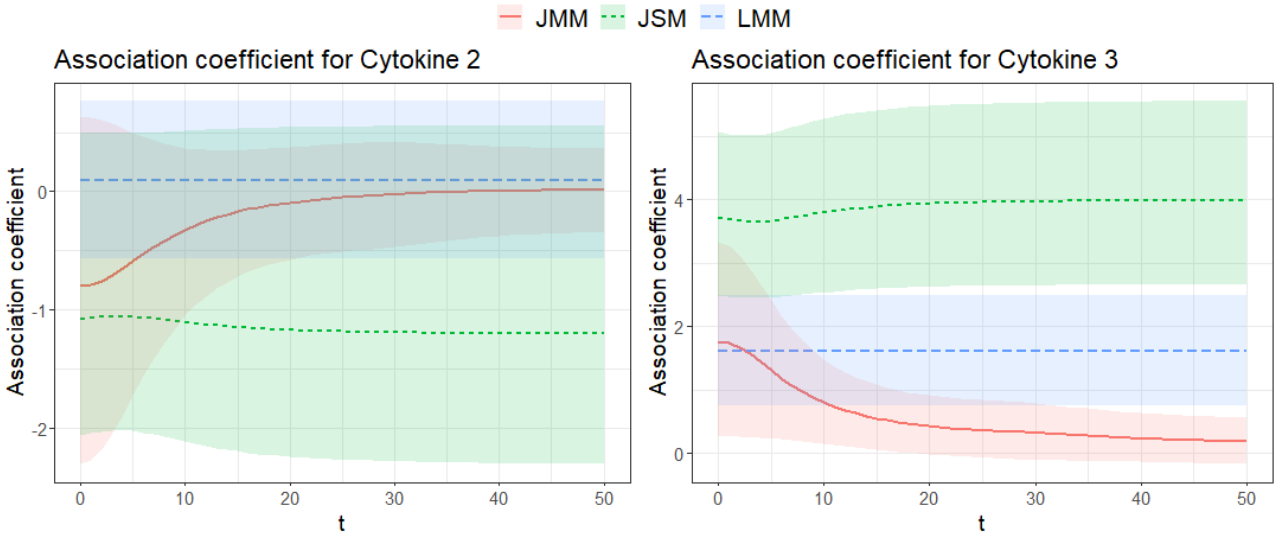
Table 6.4: *Goodness of fit values for fitting the LMM, JMM and JSM when regressing the severity score on either cytokine 2 or cytokine 3.*

	Cytokine 2			Cytokine 3		
	LMM	JMM	JSM	LMM	JMM	JSM
Marginal likelihood	-1597	-2188	-2184	-1591	-2057	-2036
DIC	2950	2885	2946	2955	2881	2927
WAIC	2948	2882	2945	2952	2877	2923
PIT	0.0318	0.0352	0.0349	0.0303	0.0371	0.0362
CPO	1484	1455	1483	1484	1452	1470
MSE (train)	8.70	7.19	8.59	8.89	7.15	8.19
MSE (subsequent)	24.75	23.95	24.96	24.90	24.43	28.48
MSE (test)	51.06	47.61	49.62	44.51	42.06	41.16

In Figure 6.4 the association coefficients can be seen for the LMM, JMM and JSM. The corresponding credible intervals are also shown. We see that for cytokine 2, none of the models shows a significant

association between the cytokine and the severity score. The JSM shows a non-significant negative association, which is also given by the JMM at low values of t. The LMM does not detect any association whatsoever.
For cytokine 3 a positive significant association is detected by both the LMM and the JSM. At low levels of t the JMM also gives a positive association, but with increasing t this association disappears.

Figure 6.4: *Association coefficients of the LMM, JMM and JSM when regressing the severity score on cytokines 2 & 3.*



To better understand the behaviour of the JMM association coefficients it is necessary to look at the covariance elements of the JMM. The covariance elements of the JMM for both cytokine 2 and 3 are given in Table 6.5. We observe that for cytokine 2 none of the covariance elements are significant, thus explaining the non-significant behaviour seen for the JMM association coefficient in Figure 6.4. For cytokine 3, on the other hand, we observe that the covariance element of the intercept is significant, while the covariance element of the random time slope is borderline significant. This explains the significant association of cytokine 3 with the outcome until time $t = 17$ observed in Figure 6.4.

Table 6.5: *Credible intervals for the covariance coefficients of the JMM (see section 6.3.2) when fitting cytokines 2 and 3.*

	Cytokine 2	Cytokine 3
$\sigma_{(C,S),0}$	-0.21 (-0.58 , 0.17)	0.39 (0.06 , 0.67)
$\sigma_{(C,S),t}$	0.02 (-0.19 , 0.23)	0.11 (-0.1 , 0.32)
$\sigma_{(C,S),t28}$	0 (-0.33 , 0.34)	0.02 (-0.32 , 0.36)

6.5 Conclusion

We can draw the following conclusions about the Covid-19 dataset:

- The joint models seem to be a better fit than the linear mixed model (LMM) for regressing the severity score on the cytokines, as the joint models outperform the LMM on all goodness of fit metrics except for the PIT.
- Cytokines 1 & 3 seem to have a significant positive association with the severity score, as this was shown by all models for at least some time periods. The association between cytokine 2 and the severity score is non-significant, but there seems to be a negative association. Overall, the exogenous and endogenous models seem to agree on the association for all cytokines.

- The question on whether the cytokines are in fact endogenous or exogenous time-varying covariates cannot be answered. Within the context of the simulation study we saw that the exogenous LMM could not detect the association in case of an endogenous time-varying covariate. However, we do not know whether this holds in general and thus applies to the Covid-19 dataset.
- There is some things we can conclude. The joint models are feasible to implement in INLA and apply to an existing dataset. The joint models perform better than the linear mixed model on almost all goodness of fit metrics. If the cytokines are endogenous, the results obtained with the joint models are less biased than the linear mixed model results. If the cytokines are exogenous, the fit of the joint models is better than the fit of the linear mixed model and thus joint models are still preferred.

Thus, the recommendations given at the end of the simulation study are still applicable (see section 5.5): If the time-varying covariate is endogenous or if the nature of the time-varying covariate is unknown (and enough data is available), joint models are preferred to a standard linear mixed model.

Chapter 7

Discussion

In this thesis we have looked at the problem of analysing time-varying covariates within the context of longitudinal data. We started by showing that endogenous time-varying covariates can lead to bias when not modelled jointly with the outcome. We pointed out that the most popular method for analysing longitudinal data, the linear mixed model (LMM), does not properly model endogenous covariates. We have therefore presented several joint models which do not have this bias, most notably the joint mixed model (JMM) and the joint scaled model (JSM). Within this thesis we have derived an association coefficient between the outcome and the endogenous time-varying covariate for the JMM and JSM, allowing for a straightforward comparison between these joint models and the LMM. Also, we have shown how to implement the joint models within the Bayesian framework using R-INLA. By way of a simulation study we have deduced that joint models are preferable over a LMM when the nature of the time-varying covariate is unknown/endogenous and the data at hand is large enough. Lastly, we have shown on a synthetic version of a real dataset that our models can be implemented on a real data-set.

An important application of joint models is to correctly model the association between the outcome and the time-varying covariate. We have therefore extensively focused on finding coefficients showing this association within the joint models. The obtained association coefficients are much more intuitive in revealing the relation between the outcome and the time-varying covariate than the distribution of the random effects (JMM) or the scaling factor (JSM). Also, they make a straightforward comparison of the joint models with the LMM. Therefore, especially in clinical practise, these association coefficients can be very useful in interpreting results obtained with joint models. We have been able to derive theoretical expected values for the association coefficients, however no theoretical confidence intervals could be obtained. This might be an avenue for further research. Also, we have mentioned several other uses of joint models, such as the inclusion of lagged values or shared parameter models. This could also be an area of future research.

We have introduced INLA in its Bayesian framework and shown how the different joint models can be implemented in R-INLA. We have discussed the multivariate joint model only briefly as its implementation in R-INLA is quite limited with the standard options. A more elaborate implementation of the multivariate joint model is a possible direction for future research, as R-INLA has multiple options for user-implemented random effects.

The simulation study allowed us to observe the joint models in practise and to obtain some recommendations on the use of joint models. We have shown that INLA is a feasible and fast (as compared to MCMC) method for implementing joint Bayesian models. In practise this means that INLA is a method that can be used to fit multiple joint models on large data-sets within a reasonable amount of time. The conclusions of the simulation study show that whenever the nature of the time-varying covariate is unknown or endogenous, a joint model yields a better fit than a standard linear mixed model, especially if the data-set is large ($>$100 subjects). This has great implications for practise, as often the nature of a time-varying covariate is unknown and thus a joint model would be the preferred option. The simulation study also shows some inconsistencies. Not all goodness of fit measures

monotonically decrease with an increasing number of subjects and the variance of the JMM association coefficient is quite high. Both could be caused by the fact that the simulation study results are based on just 250 cross-validation samples. Also, the random effects structure used for the JMM was quite complicated. For future simulation studies we would recommend using more cross-validation samples and a simple structure of the JMM random effects. Also, the simulation study does not give recommendations on when to use the JMM as opposed to the JSM and vice versa. Future research could be done on the comparison between the JMM and JSM.
Lastly, the joint models were implemented on a synthetic version of the LUMC Covid-19 dataset in which the time-varying covariates are of unknown nature. We have shown that joint models fit this data better than a LMM. This is in line with the conclusion of the simulation study, which stated that if the nature of the time-varying covariate is unknown a joint model would yield a better fit. Several limitations are observed when applying the joint models to the data. First, the association coefficient between the outcome and the time-varying covariate did not agree between the different joint models. This may be caused by the high variability of the JMM association coefficient, a property of the JMM that was observed during the simulation study. Also, the random effects structure of the time-varying covariate might have been to elaborate. This might partly explain the inconsistency of the association coefficients between the joint models and once again shows the need to carefully choose the random effects structure when fitting the joint models.

In conclusion, within this thesis we have made clear the necessity of joint modelling of the outcome and endogenous time-varying covariates in longitudinal data analysis. We have introduced and implemented the joint mixed and joint scaled models within the Bayesian setting with INLA. The simulation study showed that joint models are useful as soon as the nature of the time-varying covariate is unknown or endogenous, a claim that was observed to be true when analysing a synthetic version of the LUMC Covid-19 dataset.

Bibliography

[1] Cho, H. (2016). The analysis of multivariate longitudinal data using multivariate marginal models. *Journal of Multivariate Analysis*, 143:481–491.

[2] Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Zeger, S., et al. (2002). *Analysis of longitudinal data*. Oxford university press.

[3] Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62(2):424–431.

[4] Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008). *Longitudinal data analysis*. CRC press.

[5] Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

[6] Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. CRC Press.

[7] Guo, X. and Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The american statistician*, 58(1):16–24.

[8] Hadfield, J. D. (2010). Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package. *Journal of statistical software*, 33:1–22.

[9] Niekerk, J. v., Bakka, H., and Rue, H. (2021). Competing risks joint models using r-inla. *Statistical Modelling*, 21(1-2):56–71.

[10] Rizopoulos, D. (2017a). *An introduction to the joint modeling of longitudinal and survival data, with applications in R*.

[11] Rizopoulos, D. (2017b). An introduction to the joint modeling of longitudinal and survival data, with applications in r. *Department of Biostatistics, Erasmus University Medical Center*.

[12] Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman and Hall/CRC.

[13] Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392.

[14] Van Niekerk, J., Bakka, H., and Rue, H. (2019a). Joint models as latent gaussian models-not reinventing the wheel. *arXiv preprint arXiv:1901.09365*.

[15] Van Niekerk, J., Bakka, H., Rue, H., and Schenk, O. (2019b). New frontiers in bayesian modeling using the inla package in r. *arXiv preprint arXiv:1907.10426*.

[16] Verbeke, G., Fieuws, S., Molenberghs, G., and Davidian, M. (2014). The analysis of multivariate longitudinal data: a review. *Statistical methods in medical research*, 23(1):42–59.

[17] Wang, X., Yue, Y., and Faraway, J. J. (2018). *Bayesian regression modeling with INLA*. Chapman and Hall/CRC.

[18] Weiss, R. E. (2005). *Modeling longitudinal data*, volume 1. Springer.

[19] Wu, L. (2009). *Mixed effects models for complex data*. Chapman and Hall/CRC.

Appendices

Appendix A

R Code

The code used for this thesis can be found on Github: `https://github.com/georgygomon/Thesis_open`.