

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Helene Randi Behrens

Bayesian Mortality Modeling with Linearized Integrated Nested Laplace Approximations

Master's thesis in Applied Physics and Mathematics

Supervisor: Sara Martino

February 2022



Norwegian University of
Science and Technology

Helene Randi Behrens

Bayesian Mortality Modeling with Linearized Integrated Nested Laplace Approximations

Master's thesis in Applied Physics and Mathematics

Supervisor: Sara Martino

February 2022

Norwegian University of Science and Technology

Faculty of Information Technology and Electrical Engineering

Department of Mathematical Sciences



Norwegian University of
Science and Technology

ABSTRACT

Mortality forecasting is an important tool within, for example, demographic and actuarial sciences. Many popular mortality models include a multiplicative term, which means that they elude the group of models applicable to Bayesian inference with the popular INLA methodology (Rue et al. 2009). Our previous research (Behrens 2021) indicates that the method proposed by Lindgren and Bachl (2021), implemented in the R library `inlabru`, enables Bayesian inference with INLA also for these mortality models. In this research, we investigate how the approximate method of `inlabru` compares to asymptotically unbiased MCMC methods. Our research consists of three parts. The first part is a summary of the statistical field of mortality modeling. The second part is a comparison of `inlabru` and Stan (Stan Development Team 2019) when applied to a set of mortality models. We find that `inlabru` and Stan give similar, but not identical results and that the difference depends on the data. In the last part, we demonstrate the potential of using `inlabru` for Bayesian inference on mortality models through application to German cancer data.

SAMMENDRAG

Prediksjon av dødelighet er et viktig verktøy innen for eksempel aktuarvitenskap og demografi. Mange populære dødelighetsmodeller inneholder multiplikative ledd som gjør at de ikke inkluderes i gruppen av modeller der man kan bruke den populære metoden INLA (Rue et al. 2009) for å gjøre bayesiansk inferens. Våre tidligere undersøkelser har vist at metoden som ble foreslått av Lindgren and Bachl (2021), og som er implementert i R-biblioteket `inlabru`, gjør at bayesiansk inferens med INLA er mulig også for disse modellene. I denne studien sammenligner vi de tilnærmede resultatene til `inlabru` med resultatene til asymptotisk forventningsrette MCMC-metoder. Studien vår består av tre deler. Den første delen er en gjennomgang av dødelighetsmodellering som et statistisk felt. I den andre delen gjør vi en sammenligning av `inlabru` og Stan (Stan Development Team 2019) for et sett med dødelighetsmodeller. Vi finner at `inlabru` og Stan gir lignende, men ikke identiske resultater, og at forskjellen mellom de to avhenger av dataene. I den siste delen demonstrerer vi bruken av `inlabru` for å gjøre Bayesiansk inferens med dødelighetsmodeller for tyske kreftdata.

PREFACE

This work is my master's thesis in Industrial Mathematics, and it concludes my degree in Applied Physics and Mathematics at NTNU. This thesis is written in collaboration with my supervisor Sara Martino, with input from her colleague Andrea Riebler.

I first want to thank my supervisor Sara Martino. Regardless of how confused I have felt before our meetings, I have always returned feeling motivated, confident, and excited to continue the work. I am certain that a large part of the joy I have experienced working with this thesis is thanks to her.

I would also like to thank Andrea Riebler for valuable help and discussions. Finally, I want to thank my friends in Trondheim, for five unforgettable years.

CONTENTS

1	INTRODUCTION	1
2	MORTALITY MODELING AS A STATISTICAL FIELD	5
2.1	Age-Period-Cohort Models for Mortality Modeling	6
2.1.1	Age-Period Models	7
2.1.2	Cohort Effects in Mortality Modeling	8
2.1.3	Multivariate Mortality Models	10
2.2	Identifiability in Mortality Models	11
2.2.1	Identifiability of Age-Period Models	12
2.2.2	Identifiability of Age-Period-Cohort Models	13
2.2.3	Identifiability of Models with Different Time Intervals	16
3	GERMAN CANCER DATA	17
4	BAYESIAN MORTALITY MODELING	25
4.1	Markov Chain Monte Carlo	28
4.2	Integrated Nested Laplace Approximations	33
4.2.1	The method of INLA	35
4.2.2	Restrictions of the INLA framework	37
4.3	Linearization of Non-Linear Predictors within the INLA Framework	37
5	SIMULATION STUDY	41
5.1	Models with Gaussian Likelihoods	41
5.1.1	G1lin: Linear Gaussian Model with Fixed Precisions	45
5.1.2	G2lin: Linear Gaussian Model with Gamma Precisions	46
5.1.3	G1: Gaussian Model with Fixed Precisions	46
5.1.4	G2: Gaussian Model with Gamma Precisions	49
5.1.5	Conclusion on Comparison of Gaussian Models	52
5.2	Models with Poisson Likelihoods	52
5.2.1	P1lin: Linear Poisson Model with Fixed Precisions	53
5.2.2	P2lin: Linear Poisson Model with Gamma Precisions	54

5.2.3	P1: Poisson Model with Fixed Precisions	54
5.2.4	P2: Poisson Model with Gamma Precisions	57
5.2.5	Conclusion on Comparison of Poisson Models	61
5.3	Comparison of Run Times	62
6	APPLICATION TO GERMAN CANCER DATA	63
6.1	The Poisson Lee-Carter Cohort Model	64
6.2	Application to Full German Cancer Data	65
6.2.1	Quantifying Model Fit	70
6.2.2	Removing the Error Term	71
6.3	Forecasting German Cancer Data	74
6.4	Comparison of Model Choices for Period Effects	79
7	CLOSING REMARKS	85
8	REFERENCES	87
A	PROCEDURES IN THE SIMULATION STUDY	97
A.1	Comparison of Random Walk Models	97
A.2	Implementation of Constraints in Stan and inlabru	101
A.3	Generating Synthetic Data	103
B	COMPLEMENTARY RESULTS FROM THE SIMULATION STUDY	109
B.1	Models with Gaussian Likelihoods	109
B.2	Models with Poisson Likelihoods	118
B.3	Comparison of Run Times	138
C	COMPLEMENTARY FIGURES FROM THE APPLICATION TO GERMAN CANCER DATA	139
c.1	Application to Full German Cancer data	140
c.2	Forecasting German Cancer Data	148

INTRODUCTION

Good statistical methods for modeling trends in mortality rates are important tools in several scientific fields. In the public sector, projections of general and cause-specific mortality are of key interest in, among others, the development of health policies and the planning of social securities and pension funds. In actuarial applications, mortality forecasts play an important role in the pricing of life insurances and pension products (e.g., Lee and Carter (1992), Brouhns et al. (2002), Czado et al. (2005), Renshaw and Haberman (2006), Renshaw and Haberman (2009)).

A common way to model population development is to divide the population into age groups, for each period in time. This can be arranged in a so-called Lexis-diagram (Keiding (1990), Lexis (1875)), where age groups lie along the vertical axis and calendar periods lie along the horizontal axis. The age groups and periods are parted into intervals so that the diagram consists of squares of equal size, representing the part of the population that is in a given age group at a given period in time. Squares that lie on the same diagonal correspond to the same year of birth or the same "cohort". Constant mortality is usually assumed within these units (Czado et al. 2005), typically with some correlation or smoothing over the Lexis plane (Czado et al. (2005), Riebler (2010)). Mortality is then considered as a combination of age-, period- and cohort-specific effects.

Most mortality models are considered in a frequentist framework (see e.g., Lee and Carter (1992), Brouhns et al. (2002), Renshaw and Haberman

(2003), Blake et al. (2006), Renshaw and Haberman (2006), Currie (2016)). However, considering mortality models in a Bayesian framework provides several advantages. For example, parameter estimates and forecasts can be done in a single step, avoiding possible incoherence arising from the two-stage estimation procedures employed for most frequentist models (Czado et al. 2005). Including prior expert knowledge, with corresponding uncertainty, is straightforward in a Bayesian setting (Wiśniowski et al. 2015), and a Bayesian approach provides a natural framework for models of multiple populations (e.g., Berkum et al. (2017), Riebler et al. (2012a)). Finally, it is easier to interpret and include model and parameter uncertainty in Bayesian models (Czado et al. (2005), Wong et al. (2018)).

Bayesian mortality models, as Bayesian models in general, may demand heavy computations. Several methods have been developed to produce posterior inference. The traditional, and seemingly most widely used approach, is the Markov Chain Monte Carlo (MCMC) method for Bayesian inference. This method has successfully been used for mortality models by e.g., Czado et al. (2005), Wong et al. (2018) and Berkum et al. (2017). One of the MCMC algorithms that has gained popularity in recent years is the Hamiltonian Monte Carlo (HMC) algorithm (Duane et al. (1987), Neal (1993), Neal (2011)). This method is faster than traditional MCMC methods (Neal 1993), and thanks to the Stan software (Stan Development Team 2019) it is also easy to implement. Stan is available for the common user through for example, the `rstan` library of R.

MCMC methods give asymptotically unbiased estimates (e.g, Neal (1993)), but they do so at the price of computational complexity. An alternative to these methods is approximate Bayesian inference, such as the method of Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009). The popularity of INLA is also increasing within the field of Bayesian computationally intensive statistics, due to its computational powers (Rue et al. 2016). However, INLA is only applicable to a subset of mortality models. This is because INLA requires a linear relationship between the age, period, and cohort effects, that many mortality models do not fulfill.

One of the mortality models for which INLA is applicable is the APC (age-period-cohort) model (Clayton and Schifflers 1987). The APC model is commonly used in the field of medical and demographic sciences and is typically used to model cause-specific mortality. Indeed, there are several examples of the usage of INLA to obtain Bayesian inference with APC models that show convincing results (Riebler (2010), Riebler et al. (2012a), Riebler et al. (2012b)).

However, the scope of INLA excludes the class of mortality models based on the Lee-Carter model (Lee and Carter 1992), which are commonly used in the field of actuarial mortality modeling. A defining trait of these models is the application of an age-modulating factor to the period effect. This appears as a multiplicative term in the model, which conflicts with the requirement of a linear relationship between the time effects of INLA. This also applies to the Cairns-Blake-Dowd (CBD) model (Blake et al. 2006), which uses a simpler, but still non-linear, relationship between the age and period effects. For these models, MCMC-based methodologies are the most common option for doing Bayesian inference (Czado et al. (2005), Wiśniowski et al. (2015), Berkum et al. (2017), Wong et al. (2018), Hunt and Blake (2021a)). Since INLA is often significantly faster than traditional MCMC methods, producing results in seconds or minutes where an MCMC approach would take hours or days (Rue et al. 2009), it would be beneficial to find a way to use the INLA methodology on non-linear mortality models.

Recently, Bachl et al. (2019) proposed a method for applying the INLA methodology to models that do not fulfill its requirement of linearity. We refer to this method as `inlabru`, after the R library where it is implemented. `inlabru` was originally proposed as a solution to non-linearity problems arising in ecological data, and we wish to investigate if it is also applicable to non-linear mortality models. If this turns out to be the case, it would open up the possibility for more efficient analysis of e.g., Lee-Carter types of mortality models, by for instance making sensitivity analysis of model parameters faster and making the comparison of different model choices more feasible.

We already know that the approximate INLA method gives very similar results to asymptotically unbiased MCMC methods (Rue et al. 2009). However, `inlabru` introduces an additional approximation, namely a linearization of the non-linear parts of the model. We wish to investigate to what degree this estimation bias affects the resulting approximate inference when applied to non-linear mortality models.

This work is an extension of a previous investigation (Behrens 2021), where we demonstrated that `inlabru` was able to correctly identify the mortality rates and time effects of some simple synthetic data. We then used `inlabru` to apply both the APC model and a version of the Lee-Carter model to actual data of cancer mortality. This previous study showed promising results, indicating that it is indeed both possible and effective to use `inlabru` for Bayesian inference for non-linear mortality models. We now perform a more thorough investigation of the exact performance of the approximate `inlabru` method compared to the MCMC method. Specifically, we compare the inference results of `inlabru` to the results of inference with Stan, for a selection of mortality models both including and excluding a non-linear term.

We find that `inlabru` and Stan give very similar results in the cases where the models possess a linear structure. This is as expected since `inlabru` is equivalent to INLA in this case. For the cases with non-linear model structures, we do observe a difference in the estimation results of Stan and `inlabru`. We attribute some of this difference to unidentifiability in the data and some to the bias of the linearization process of `inlabru`. We discuss these differences in detail.

In the last part of our research, we demonstrate the abilities of `inlabru` as a tool for mortality modeling by using it to apply a Lee-Carter type of model to data of German cancer mortality. We apply the model both to the full set of data and to a set where the most recent years are left out, for which we use `inlabru` to produce forecasts for the omitted years.

MORTALITY MODELING AS A STATISTICAL FIELD

Good statistical methods for modeling trends in mortality rates are important tools in several scientific fields. In the public sector, projections of general and cause-specific mortality are of key interest in, among others, the development of health policies and planning of social securities and pension funds, and in actuarial applications, mortality forecasts play an important role in the pricing of life insurances and pension products (Lee and Carter (1992), Brouhns et al. (2002), Czado et al. (2005), Renshaw and Haberman (2006), Renshaw and Haberman (2009)).

The dominating approach to mortality modeling on a population level is to consider mortality, in the form of observed counts of mortality, mortality rates, or probability of mortality, as a combination of age- and period-specific effects. In some cases, a cohort effect is included, which describes lifelong effects for the population with the same birth year. This general structure is common for most mortality models in literature (Hunt and Blake 2021b). The differences between the numerous models that we see in the field of mortality modeling seem to lie in the structure of the relationship between the time effects, in what way variability is captured by the model, whether a separate cohort effect is included and whether the model is considered in a Bayesian setting (Clayton and Schifflers (1987) Lee and Carter (1992), Blake et al. (2006), Delwarde et al. (2007), Wong et al. (2018), Renshaw and Haberman (2006), Hunt and Villegas (2015), Riebler (2010), Czado et al. (2005), Wiśniowski et al. (2015)).

In the following chapter, we outline some of the discussions on mortality modeling that have been taking place in recent years. We describe some of the most commonly used models, as well as the estimation procedures that have been employed, and the challenges, properties, and discussions associated with these.

2.1 AGE-PERIOD-COHORT MODELS FOR MORTALITY MODELING

Mortality data typically consist of the observed deaths $Y_{x,t}$ for each age group x during the period t and the corresponding population at risk $E_{x,t}$. Most mortality models connect these observation to some age, period and cohort effects through a likelihood. This can for example be an assumed Gaussian mortality rate:

$$m_{x,t} = Y_{x,t}/E_{x,t}, \quad \log(m_{x,t}) \sim \mathcal{N}(\eta_{x,t}, 1/\tau),$$

as in e.g., Lee and Carter (1992). Here, $\mathcal{N}(\mu, 1/\tau)$ denotes a Gaussian distribution with mean μ and precision (inverse variance) τ . Alternatively, the observed deaths may be assigned a Poisson distribution as

$$Y_{x,t} \sim \mathcal{Poisson}(E_{x,t} \cdot e^{\eta_{x,t}})$$

(e.g., Brouhns et al. (2002)). Both models above describe the relation between mortality and the age, period, and cohort effects through a predictor $\eta_{x,t}$, and this predictor is usually given by some version of the structure

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^n \beta_x^{(i)} \kappa_t^{(i)} + \beta_x^{(0)} \gamma_c. \quad (2.1.1)$$

Here $x = 1, \dots, X$ denotes the age, $t = 1, \dots, T$ denotes the period, typically calendar years, and $c = 1, \dots, C$ denotes the birth year, or cohort, usually given by $c = t - x$. α_x denotes a static age effect, often modeled as the average mortality for age x over the observed period (e.g., Lee and Carter (1992)). The $\beta_x^{(i)}$ are modulating age effects, applied to period effects $\kappa_t^{(i)}$ or to the

cohort effect γ_c . We refer to these collectively as the random effects or the time effects.

This general way of expressing age-period-cohort mortality models was concretized by Hunt and Blake (2021b), as a way to summarize the increasing number of mortality models that have been proposed over the recent years. They emphasize, as do we, that this general structure is a result of combining all models proposed up until recently, and that the original mortality models were not proposed with the structure of Equation 2.1.1 in mind. Still, we consider it a useful tool in structuring the model comparison, and we will discuss the different proposed mortality models in relation to this general structure in the following section.

2.1.1 Age-Period Models

Several models use only age and period effects to model mortality. In the actuarial sciences, the model proposed by Lee and Carter (1992), hereby referred to as the Lee-Carter model, has been popular ever since its introduction. Variations of this model have largely dominated the field of actuarial mortality modeling (Renshaw and Haberman (2009), Booth and Tickle (2008)). The original Lee-Carter model assumes a Gaussian distribution of the log-mortality rates, where the mean of the distribution is linked to age and period effects through the predictor:

$$\eta_{x,t} = \alpha_x + \beta_x \kappa_t. \quad (2.1.2)$$

This corresponds to setting $n = 1$ and $\gamma_c = 0$ in Equation 2.1.1. Here $m_{x,t}$ is the mortality rate for an age group x at period t .

After its introduction, numerous extensions of the Lee-Carter model have been proposed. Brouhns et al. (2002) suggest exchanging the Gaussian log-mortality model structure with a Poisson structure, where they consider the observed number of deaths to follow a Poisson distribution:

$$\text{PLC: } Y_{x,t} \sim \mathcal{Poisson}(E_{x,t} \cdot e^{\eta_{x,t}}), \quad \eta_{x,t} = \alpha_x + \beta_x \kappa_t.$$

Here $Y_{x,t}$ is the observed number of deaths and $E_{x,t}$ is the exposed population of age x in year t . We refer to this model as the Poisson Lee-Carter (PLC) model. Czado et al. (2005) extended the PLC model further, by incorporating it into a Bayesian framework.

Delwarde et al. (2007) proposed extending the Bayesian PLC model with an error term to account for the variability that cannot be explained by either age, period or cohort effects, so-called overdispersion. The Poisson distribution of the basic PLC model limits the variability to be equal to the predictor $\eta_{x,t}$. Wong et al. (2018) demonstrated that the PLC model without the error term could not sufficiently capture the variability of their example data. Delwarde et al. (2007) proposed to assign the error term a Gamma prior, and Wong et al. (2018) suggested an alternative "Poisson log-normal Lee-Carter" (PLNLC) model that assigns the error term a Gaussian prior.

Another proposed extension of the original Lee-Carter model is the inclusion of multiple period terms to increase flexibility with respect to age (Renshaw and Haberman 2003). This corresponds to setting $n > 1$ and keeping $\gamma_c = 0$ in the structure of Equation 2.1.1. A variant of these types of models was introduced by Blake et al. (2006). Their model, known as the CBD model, is directed at modeling mortality at pensioner ages, and it is suited for modeling longevity risk. CBD types of models omit the static age effect α_x and instead include multiple period effects modulated by polynomial age functions. Additionally, the models are fitted to the probabilities of death transformed by a logit link function, as opposed to the mortality rates.

2.1.2 Cohort Effects in Mortality Modeling

A basic and commonly used model within the fields of medical and demographic sciences, is the APC (age-period-cohort) model (Clayton and Schifflers 1987).

The APC model defines a simple, additive relation between the time effects, which we arrive at by setting $\beta_x^{(i)} = 1$ and $n = 1$ in Equation 2.1.1:

$$\eta_{x,t} = \alpha_x + \kappa_t + \gamma_c.$$

Its simple structure makes the APC model easy to perform inference with, especially when it is considered in a Bayesian setting. However, due to the linear dependence between the time effects, arising from the relation

$$\text{cohort} = \text{period} - \text{age},$$

the additivity of this model makes it especially sensitive to identifiability issues between the three time effects (e.g., Riebler (2010), Hunt and Blake (2020b)). There is an entire field of literature dedicated to discussing the APC model and its performance (see e.g., Riebler and Held (2010)), and many studies discussing "age-period-cohort" models refer only to this model.

In the fields of medical mortality modeling, the inclusion of the cohort effect has long been considered a useful tool to correctly capture mortality, and the APC model has been used extensively in this field (see e.g., Riebler et al. (2012a), Riebler et al. (2012b)). In recent years, the need to include an explicit cohort effect also when describing tendencies in general mortality, has been pointed out increasingly often in the actuarial community as well (e.g., Hunt and Villegas (2015)). This is in contrast to more traditional models for actuarial mortality modeling, where cohort effects are often acknowledged, but treated as secondary to the age and period effects (Hunt and Villegas 2015). Although some argue that any observable cohort effects are simply misplaced age or period effects and that they should be modeled accordingly (for example Cairns et al. (2011)), there is increasing agreement that cohort effects are in any case useful in practical mortality modeling.

The inclusion of this cohort effect can be formulated in several ways. Renshaw and Haberman (2006) propose to simply add it as a term in the usual Lee-Carter model structure. They consider two versions of this model. In the

first, the cohort effect is modulated by an age effect, similarly to the period effect:

$$\text{M: } \log(\mu_{x,t}) = \alpha_x + \beta_x^{(1)} \kappa_t + \beta_x^{(0)} \gamma_c. \quad (2.1.3)$$

We refer to this model as model M, following the convention of Renshaw and Haberman (2006), and later Hunt and Villegas (2015). In the second version, the age-modulating effect on γ_c is removed, i.e., $\beta_x^{(0)} = 1$, which results in the simple model structure

$$\text{H1: } \log(\mu_{x,t}) = \alpha_x + \beta_x \kappa_t + \gamma_c, \quad (2.1.4)$$

which we denote by H1, also after Renshaw and Haberman (2006). Hunt and Villegas (2015) argue, after comparing the robustness to changes in data of the two models, that the H1 model should be preferred.

Similarly, different cohort extensions of the CBD model were proposed by Cairns et al. (2009), both with and without an age-modulation of the cohort effect. Cairns et al. (2009) showed that these models were effective for modeling mortality in older ages.

Multiple terms for the cohort effect have also been proposed, by e.g., Hatzopoulos and Haberman (2011). However, this has been found to over-complicate the models and does not seem to be needed for description of the cohort effects in most data, and so these models are rarely used (Hunt and Blake 2021b).

2.1.3 *Multivariate Mortality Models*

An additional field of mortality models are multivariate mortality models, aiming to capture differences in the age, period and cohort effects across subgroups of a population. This can for example be gender-specific differences in the time effects, or mortality patterns varying across geographical regions. Riebler and Held (2010) and Riebler et al. (2012a) discuss multivariate versions

of the APC model where some of the time effects are shared, while others are allowed to vary across different strata of the population.

As argued by Riebler et al. (2012a), multivariate modeling of subpopulations may be particularly useful when different quality of data is available across the different strata. Then the estimations for the strata with lower quality of data, e.g., more missing values, may borrow strength from the related populations. This benefit of multivariate models has also been emphasized and successfully employed in the field of actuarial mortality modeling by e.g., Berkum et al. (2017). We acknowledge the vast potential that multivariate mortality models present, while we limit the scope of our following research to univariate models.

2.2 IDENTIFIABILITY IN MORTALITY MODELS

The degree of identifiability of different mortality models has been a widely discussed topic within the field of mortality modeling. Unidentifiability of a model entails that there exist multiple different sets of time effects (age, period and cohort effects) that give the same fitted mortality rates. These sets of parameters are equivalent to each other, when considering nothing other than the available data. It is therefore common practice to apply some identifying constraints to the model. However, exactly which constraints are needed to ensure identifiability, and when they suffice to fully identify the model, has been thoroughly discussed (see e.g., Riebler (2010), Held and Riebler (2013), Hunt and Villegas (2015), Hunt and Blake (2020a), Hunt and Blake (2020b)).

The identifiability of the mortality models depends on the model structure, whether or not a cohort effect is included and the model choices for the time effects. We outline some of the discussion on identifiability in mortality models, as well as some recently proposed general conditions for the identifiability of age-period and age-period-cohort models.

Hunt and Blake (2020a) and Hunt and Blake (2020b) recently formalized the identifiability of models of the general form of Equation 2.1.1. We present some key points of their investigation here, as a way to summarize the discussions on identifiability in mortality models.

2.2.1 Identifiability of Age-Period Models

We begin with the simpler results of Hunt and Blake (2020a), which describe age-period models of the form

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_t^{(i)}. \quad (2.2.1)$$

Hunt and Blake (2020a) rewrite this to matrix notation as

$$\{\eta_{x,t}\} = \alpha \mathbf{1}^T + \beta \kappa.$$

Here, α is an $(X \times 1)$ matrix, $\mathbf{1}$ is a $(T \times 1)$ matrix of ones, β is an $(X \times N)$ matrix and κ is an $(N \times T)$ matrix. Hunt and Blake (2020a) then show that the only invariant transformations for this model structure are

$$\begin{aligned} \{\hat{\alpha}, \hat{\beta}, \hat{\kappa}\} &= \{\alpha, \beta A^{-1}, A \kappa\} \\ \{\hat{\alpha}, \hat{\beta}, \hat{\kappa}\} &= \{\alpha - \beta B, \beta, \kappa + B \mathbf{1}^T\}, \end{aligned} \quad (2.2.2)$$

where A is an invertible $(N \times N)$ matrix and B is an $(N \times 1)$ matrix. The parameter space of these transforms has dimension $N(N + 1)$, and thus $N(N + 1)$ constraints are needed for identifiability of the model (Hunt and Blake 2020a). For example, in the case where $N = 1$, such as for the Lee-Carter model, two constraints are needed to ensure identifiability.

We should point out, as emphasized by e.g., Hunt and Blake (2020a), that the identifiability constraints are arbitrary, and that they only serve to choose one out of many possible sets of parameters. This means that the constraints can be chosen freely, without altering the fit of the data. It also means that the constraints in themselves contain no information about the

data, which one should keep in mind for example when using the fitted model to forecast future rates. The identifiability constraints do, however, change the demographic significance of the time effects. Consequently, these constraints are often chosen in a way that makes the resulting model easily interpretable. Typically, for age-period models with the structure given in Equation 2.2.1, the identifying constraints are imposed on β_x and κ_t as

$$\sum_x \beta_x = 1, \quad \sum_t \kappa_t = 0. \quad (2.2.3)$$

These constraints have the advantage that they allow the interpretation of α_x as the average mortality rate for each age across the observed period (Hunt and Blake 2020a). Another common choice of constraint is for example $\kappa_{t_1} = 0$, as employed by e.g., Renshaw and Haberman (2003).

Hunt and Blake (2020a) specifically distinguish between models with parametric and non-parametric age effects $\beta_x^{(i)}$. They define parametric age effects as effects where some specific shape of the effect is assumed a priori, i.e., when $\beta_x^{(i)} = f^{(i)}(x, \theta^{(i)})$ for some parameters $\theta^{(i)}$. The results that we described above only generally apply to models with non-parametric age effects. This is because models with parametric $\beta_x^{(i)}$ may not be applicable to the invariant transformations of Equation 2.2.2, and the required number of identifying constraints may depend on the form of the age function. We mainly consider what Hunt and Blake (2020a) refer to as non-parametric functions for all of the time effects, so we do not elaborate further on this. However, we do note the importance of the distinction, since applying excessive identifying constraints may undesirably change the model fit (see e.g., Hunt and Villegas (2015)).

2.2.2 Identifiability of Age-Period-Cohort Models

When a cohort effect is added to the model structure, an additional identifiability issue arises, due to the linear relation cohort = year - age, which is not resolved simply by adding an identifying constraint to the cohort effect.

This has been widely discussed for APC models (see e.g., Riebler (2010)), as this unidentifiability of the time effects is especially apparent in the additive structure of these models. Solutions to this have been proposed, but none seem to be widely accepted (see e.g., Luo (2013) and Held and Riebler (2013) for an impression of the intricacies of the discussions on this topic).

Hunt and Blake (2020b) treat the topic of identifiability in age-period-cohort mortality models in a general way and consider models of the form

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^N \beta_x^{(i)} \kappa_t^{(i)} + \gamma_{c=t-x}. \quad (2.2.4)$$

Since the age-period model of Equation 2.2.1 is a submodel of this model, the identifiability issues discussed for the age-period model also applies here. With the introduction of the term γ_c , the parameter space increases, and so an additional constraint must be imposed on the cohort effect to ensure identifiability in the sense that it was discussed for the age-period model. This is typically chosen as

$$\sum_x \sum_t \gamma_{t-x} = 0,$$

as suggested by e.g., Cairns et al. (2009), or

$$\sum_{c=t_1-x_X}^{t_T-x_t} \gamma_c = 0, \quad (2.2.5)$$

as suggested by Hunt and Villegas (2015). The latter choice of constraint keeps the interpretation of α_x as the average mortality rate when it is combined with the constraints of Expression 2.2.3. However, simply adding this constraint does not necessarily ensure identifiability of the collinearity between the age, period and cohort effects, and this is examined by Hunt and Blake (2020b).

We note that Hunt and Blake (2020b) only consider models without an age-modulating factor in the term for the cohort effect, meaning that their findings do not necessarily apply to models with an age-modulated cohort effect, such as the model M of Renshaw and Haberman (2006). However, as Hunt and Blake (2021b) discuss, these models are quite rarely used, as they

have been shown to lack robustness (Hunt and Villegas 2015). Consequently, we do not consider this a great loss of generality.

Again, Hunt and Blake (2020b) make the distinction between models with parametric and non-parametric age functions $\beta_x^{(i)}$. They show that for non-parametric choices of $\beta_x^{(i)}$, models with the structure of Equation 2.2.4 do not possess any additional unidentifiability other than that discussed for the age-period model of Equation 2.2.1. This entails for example that the model H1 of Renshaw and Haberman (2006) with constraints such as those given in Expressions 2.2.3 and 2.2.5 is in theory fully identified, as long as the β_x effect is not given some functional shape $f(x)$ a priori. However, for parametric choices of β_x , models with the structure of Equation 2.2.4 are not necessarily identifiable. The APC model is considered a parametric model, since it is equivalent to the model in Equation 2.2.4 when we assign β_x the parametric choice $\beta_x = 1$. Hunt and Blake (2020b) attributes this lack of identifiability of the APC model to the invariant transformation

$$\{\hat{\alpha}_x, \hat{\kappa}_t, \hat{\gamma}_c\} = \{\alpha_x + a(x - \bar{x}), \kappa_t - a(t - \bar{t}), \gamma_c + a(c - \bar{c})\},$$

where a is a constant. They also discuss the unidentifiability for mortality models for other parametric model choices for β_x , for instance when β_x is modeled with a polynomial or exponential function, but we do not go into details of this discussion here.

Hunt and Blake (2020b) discuss different tactics for overcoming these identifiability issues in the APC and other parametric age-period-cohort mortality models. They emphasize the method of adding another constraint that removes linear drift from the cohort effect. Such a constraint can for example be

$$\sum_c n_c \gamma_c (c - \bar{c}) = 0, \quad (2.2.6)$$

where n_c is the number of available observations of cohort c . However, this practice is disputed, with some arguing that imposing demographic significance of the model in this way only serves to confuse, and that it is more correct to leave the model in its unidentified state (e.g., Held and Riebler (2013)).

Finally, we note that the theoretic identifiability of the non-parametric age-period-cohort models of Equation 2.2.4 does not guarantee identifiability in practice. For instance, Hunt and Villegas (2015) observe identifiability issues in their theoretically identifiable H1 model, and they attribute these issues to the near-parametric shape of κ_t that they find when fitting the H1 model to their data. Again, they suggest adding an "approximate" identifiability constraint, such as Equation 2.2.6, and they argue that this removes convergence issues that they see arising from the unidentifiability.

2.2.3 *Identifiability of Models with Different Time Intervals*

Finally, a third type of unidentifiability may arise in age-period-cohort models for which the age and period intervals are of unequal lengths. This issue has been discussed for APC models by e.g., Holford (2006) and Riebler (2010), who describe the unidentifiability appearing as cyclical patterns in the estimated cohort effects. Adding smoothing priors to the random effects, such as random walks of order two, is suggested as a way of removing this type of lack of identifiability.

GERMAN CANCER DATA

In our investigation, we will consider mortality modeling using data of mortality for lung and stomach cancer in Germany for the years 1999-2016, obtained from the German Centre for Cancer Registry (2021b). In these data sets, the observed mortality for the two cancer types is given by the number of cases for men and women, for each year, for five-year age intervals. The exception is for the ages above 85 years, which are collected in a single group. We denote the observed cases of deaths for one cancer type, for one sex, at age group x and cohort c during period t by $Y_{x,t}^{\text{type, sex}}$. We note that the cohort c of the observation is implicitly given by the age x and the year t .

We find data for the total German population, which serves as our at-risk population, for the corresponding years and ages at the German Centre for Cancer Registry (2021a). This dataset contains observations for age groups of one year, and we aggregate them to the same five-year age groups as for the cancer data. We denote the observed German population for a given sex, age group x and cohort c during period t by $E_{x,t}^{\text{sex}}$. We then find the corresponding mortality rates by

$$m_{x,t}^{\text{type, sex}} = \frac{Y_{x,t}^{\text{type, sex}}}{E_{x,t}^{\text{sex}}}, \quad (3.0.1)$$

or $m_{x,t}^{\text{type}} = Y_{x,t}^{\text{type}} / E_{x,t}$ in the case where we consider the two sexes together.

We index the age groups x , the periods t and the cohorts by

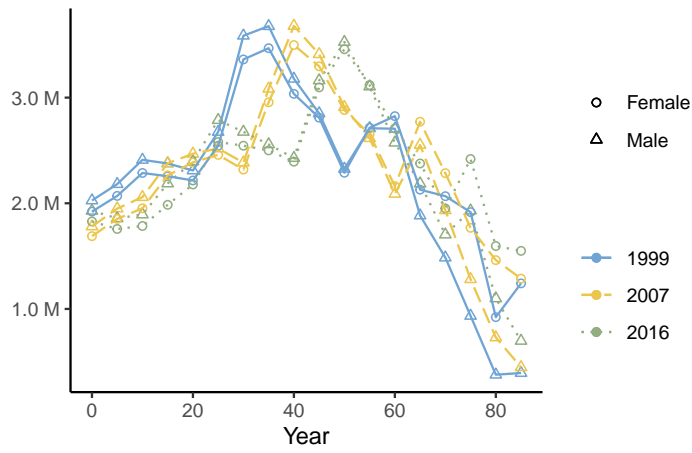
- The age indices x are $x = 0, \dots, X = 17$.
- The period indices t are $t = 0, \dots, T = 17$.
- The cohort indices are given by $c = 5(X - x) + t$, $C = 5X + T = 102$,

as proposed by Heuer (1997). We index the cohort effects this way, as opposed to the canonical $c = t - x$, to handle the fact that our data contain age- and period intervals of unequal lengths.

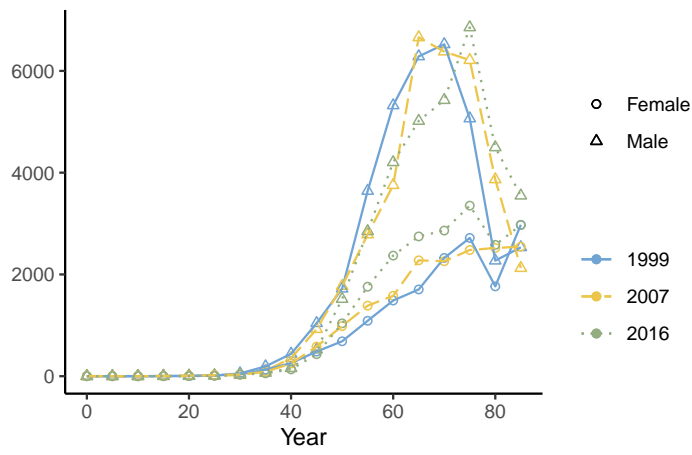
Figure 3.0.1 displays the observed population and cases of lung and stomach cancer deaths over different age groups. Figures 3.0.1b and 3.0.1c show that there are most observed cancer deaths at older ages, for both males and females. This is even though the general German population is largest at around 40 years old. From Figure 3.0.1a, we can observe aging of the German population, with the most common age shifting from around 40 years old in 1999 to closer to 50 years old in 2016.

Figure 3.0.2 displays the development of the observed population and cases of lung and stomach cancer deaths for the period 1999-2016. We see that the observed lung and stomach cancer cases seem to increase with time for women, but stay rather stable for men. There is a clear drop in the German population in the year 2011, which can be seen in Figure 3.0.2a. This is an effect of the population census held in Germany in 2011, which showed that the estimated population numbers were off by 1.5 million people compared to the previous years (Statistisches Bundesamt 2013).

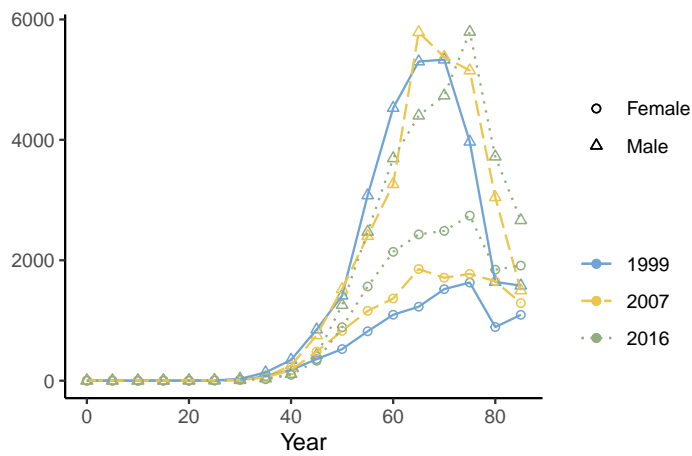
Figure 3.0.3 displays the observed male, female and combined cancer mortality rates, as found from the relation in Equation 3.0.1, averaged over all available years. We first note that the male and female mortality rates behave quite differently to each other and to the mortality rates for the full population. This is the case for both lung and stomach cancer mortality, and we interpret this as an indication that male and female mortality should be treated separately in modeling. Our previous work on mortality modeling for these data confirms this (Behrens 2021). The observed cases of male lung and stomach cancer follow each other very closely and we expect to get similar



(a) Observed German population

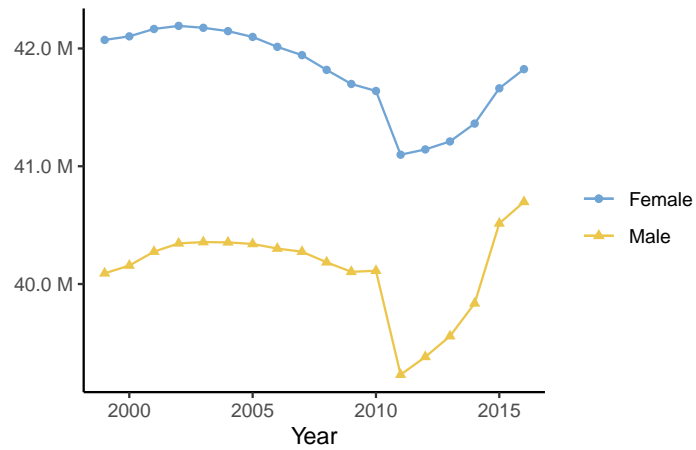


(b) Observed stomach cancer deaths

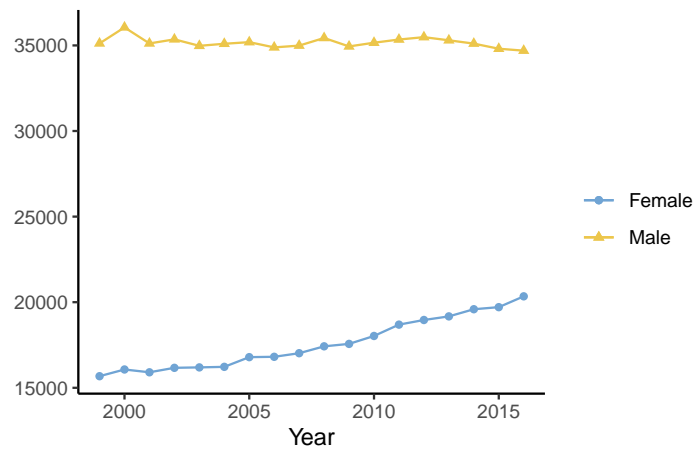


(c) Observed lung cancer deaths

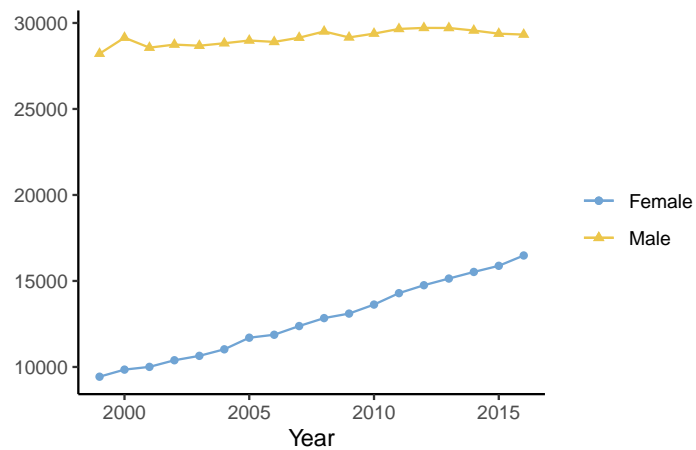
Figure 3.0.1: The total German population (upper) and the number of stomach and lung cancer deaths (middle and lower) are displayed by the different age groups, for the years 1999, 2007, and 2016.



(a) Observed German population



(b) Observed stomach cancer deaths



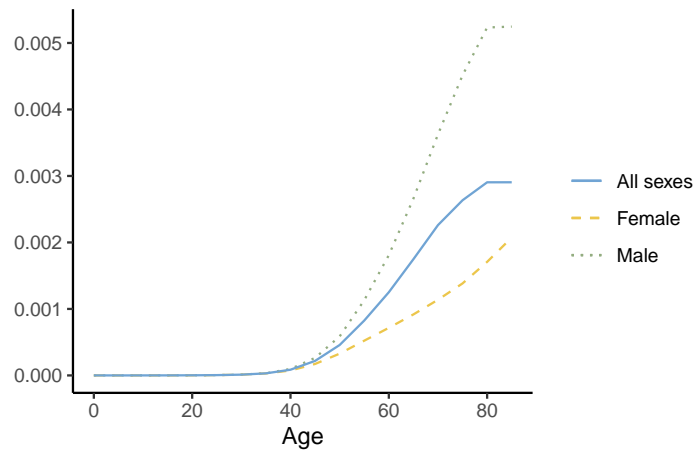
(c) Observed lung cancer deaths

Figure 3.0.2: The total number of people (upper) and stomach and lung cancer deaths (middle and lower), summed over all age groups and displayed for each year.

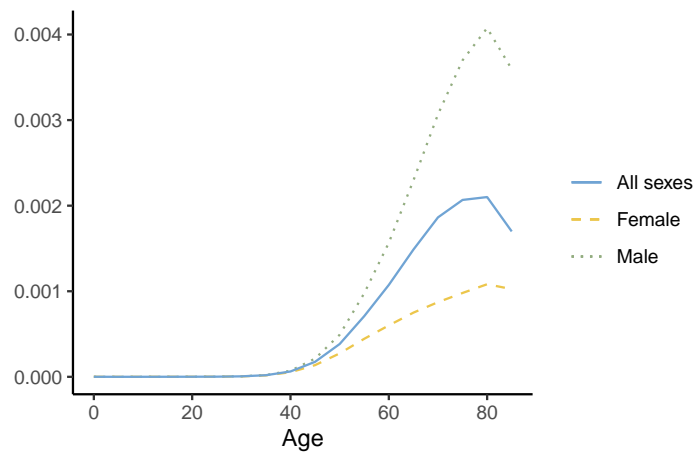
results when analyzing them. The same traits can be seen in Figure 3.0.4, which displays the observed mortality rates averaged over all age groups.

Figure 3.0.5 displays the observed mortality rates for each cohort for the years 1999, 2007, and 2016. Note that we do not have full observations for the older cohorts. For instance, for the year 2016, we only have observations of the cohorts down to $c = 17$. This results from the fact that all observations older than 85 are counted together, which is a simplification of reality. For example, there will be some small part of the population that was 85 years old or older in 1999, and thus part of the $c = 0$ cohort, that survived until 2016, where the $c = 0$ cohort is no longer observed. In 2016, these people would be assigned the cohort $c = 17$. Although we can technically consider this as an error in the cohort data, we assume that it will not affect our analysis significantly, since it applies to such a small part of the population. Still, we keep in mind that we have fewer observations for the older cohorts. Similarly, we also have fewer observations for the cohorts born after 1999. Additionally, Figure 3.0.5 shows that for many of the younger cohorts, mortality rates are close to zero during the entire observed period 1999-2016.

Mortality rates averaged over the period 1999-2016



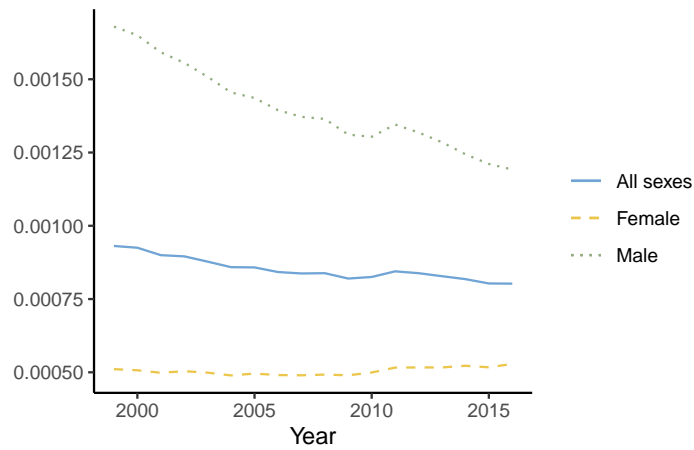
(a) Stomach: Average mortality rates for stomach cancer, male and female



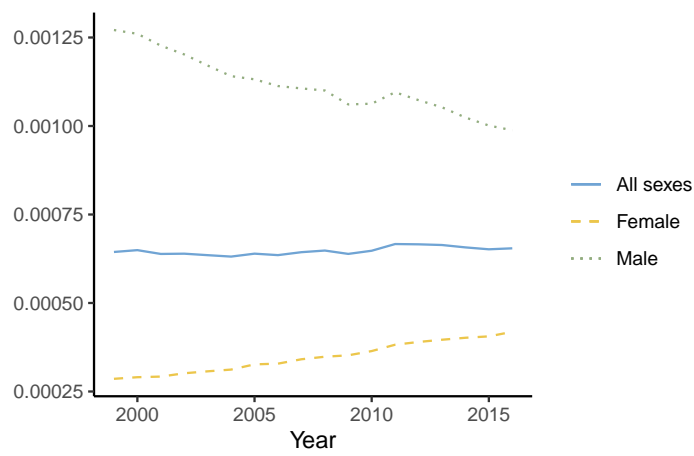
(b) Lung: Average mortality rates for lung cancer, male and female

Figure 3.0.3: Mortality rates averaged over all years, for males and females

Mortality rates averaged over ages 0-85+



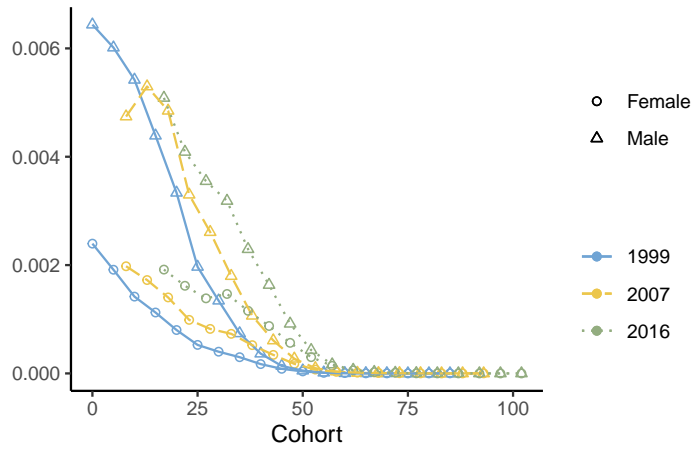
(a) Stomach: Average mortality rates for stomach cancer, male and female



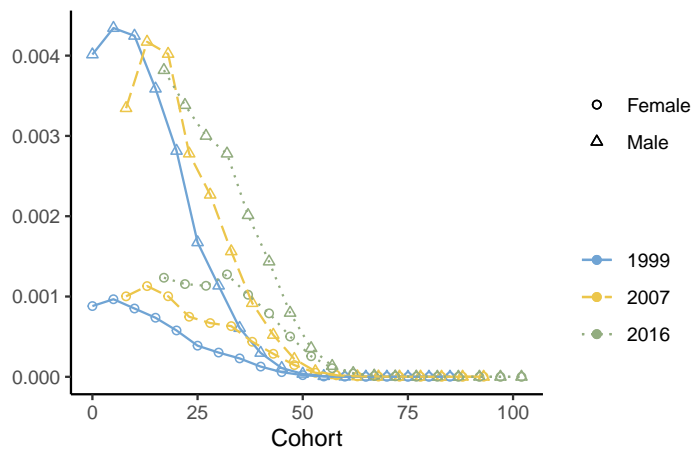
(b) Lung: Average mortality rates for lung cancer, male and female

Figure 3.0.4: Mortality rates averaged over all age groups, for male and females

Mortality rates averaged over all birth years



(a) Stomach: Average mortality rates for stomach cancer, male and female



(b) Lung: Average mortality rates for lung cancer, male and female

Figure 3.0.5: Mortality rates averaged over all cohorts, for male and females

BAYESIAN MORTALITY MODELING

In this research, we will consider mortality models in the framework of Bayesian statistics. Bayesian statistics differ from classical (frequentist) statistics in the way that all model parameters are considered as random variables (Gilks et al. 1995). Bayesian statistics can be described as a way to update one's current belief in light of new evidence. In practice, this entails that the prior, subjective, knowledge about some event is taken into account when considering the probability of that event, in addition to observed data.

As argued by e.g., Czado et al. (2005), Wiśniowski et al. (2015) and Wong et al. (2018), there are many advantages of considering mortality modeling and forecasting in a Bayesian framework. In Bayesian models, incorporating previous expert knowledge (or lack thereof) is intuitive and straightforward. Additionally, forecasts and estimates of model parameters may be found simultaneously through a single-step procedure, which eliminates the risk of incoherence and the reduction in the goodness of fit associated with alternative two-stage procedures (Czado et al. (2005), Hunt and Villegas (2015)).

Specifically, we consider a subgroup of Bayesian models, called hierarchical Bayesian models (see e.g., Allenby et al. (2005), Gamerman and Lopes (2006)). Hierarchical, or multilevel, models are typically used for modeling grouped observations (Albert and Hu 2019), and they can be thought of as a generalization of regression modeling where posterior probability distributions

are estimated also for the regression parameters (Gelman 2006). The models consist of three parts:

- \mathbf{y} : The observed data \mathbf{y} , generated by some unobserved process x
- x : The latent process x , conditional on some hyperparameters θ
- θ : The hyperparameter θ .

In a Bayesian framework, these three model components are all treated as random variables, with a corresponding probability distribution. \mathbf{y} is modeled with a likelihood model given x and θ , our previous knowledge about the latent process x is expressed through a prior distribution and θ is assigned a hyperprior distribution reflecting our knowledge of this hyperparameter. This results in the following structure:

$$\begin{aligned}\mathbf{y} &\sim \pi(\mathbf{y} | x) \\ x &\sim \pi(x | \theta) \\ \theta &\sim \pi(\theta)\end{aligned}$$

(see e.g., Gamerman and Lopes (2006)).

The mortality models that we have discussed in Chapter 2 can all be formulated as hierarchical models. The latent field x then consists of the time effects α_x , $\beta_x^{(i)}$, $\kappa_t^{(i)}$ and γ_c , which are linked to the observed mortality through a Gaussian or Poisson likelihood. Prior distributions $\pi(x | \theta)$ can then be assigned to the age, period and cohort effects. The hyperparameters of these processes are then the hyperparameters θ of the hierarchical model, which are in turn assigned some hyperprior $\pi(\theta)$.

Our interest lies in the posterior distribution $\pi(x, \theta | \mathbf{y})$. This is proportional to the likelihood models and the prior distributions of x and θ through the relation

$$\pi(x, \theta | \mathbf{y}) \propto \pi(\mathbf{y} | x, \theta) \cdot \pi(x | \theta) \cdot \pi(\theta),$$

following from Bayes' theorem (Bayes 1763). However, obtaining information from this posterior distribution often involves calculation of high-dimensional integrals, which are rarely analytically available (Beguin et al. 2012). There

are several integration methods that can be employed to solve this, with the Markov Chain Monte Carlo (MCMC) method being the traditionally dominant choice.

The MCMC method draws samples from the target distribution (Gilks et al. 1995), meaning that it produces asymptotically unbiased estimates (Neal 1993). The MCMC method is applicable to nearly every model formulation (Gilks et al. 1995), which makes it an extremely powerful tool for performing Bayesian inference. The downside of the classical MCMC algorithm is that it is computationally demanding, and even though it, in theory, converges to the correct distribution, it may do so only after a very long time (e.g., (Hoffman and Gelman 2014)). One possible solution is to use enhancements of the classical MCMC algorithm, such as the Hamiltonian Monte Carlo (HMC) algorithm (Duane et al. 1987). HMC utilizes principles of Hamiltonian dynamics to draw samples more efficiently and thus converges much quicker (see Section 4.1 for details).

An alternative to MCMC is the Integrated Nested Laplace Approximation (INLA) method (Rue et al. 2009). The INLA method, described further in Section 4.2, uses Laplace approximations of integrals to obtain approximate posterior marginals of the model parameters (Gómez Rubio 2020). It has been shown that for many problems, INLA can produce accurate results in much less time than the MCMC method would use to obtain the same accuracy (Rue et al. 2009), and it is therefore increasing in popularity (Rue et al. 2016).

However, INLA is only applicable to a limited set of models. Of the mortality models discussed in Chapter 2, it is only possible to use INLA with the APC model, since this is the only model with a linear relationship between the time effects. The theory behind this is explained further in Section 4.2. This means that for the wide class of mortality models with a multiplicative relationship between the age and period effects, it is not possible to use INLA as it is to perform inference.

In the following chapter, we give an introduction to these two methods for Bayesian inference. We first present the HMC method and the statistical

software Stan, as this is becoming a state-of-the-art approach for MCMC-type of inference. We then present the INLA method, together with the recently proposed `inlabru` method (Lindgren and Bachl 2021), which proposes a method for applying INLA also to non-linear models.

4.1 MARKOV CHAIN MONTE CARLO

The Markov Chain Monte Carlo (MCMC) method, usually attributed to Metropolis et al. (1953), is perhaps the most widely used method for performing Bayesian inference.

MCMC does inference by drawing samples from a target distribution, which, in the context of Bayesian statistics, is typically a posterior distribution that is not available in closed form. These samples can be used to calculate the desired quantities of the target distribution (the mean, mode, etc.). The driving principle behind the MCMC algorithm is that these samples are found through the construction of a Markov Chain with a limiting distribution equal to the target distribution of interest (see e.g., Robert and Casella (2004)). Commonly used MCMC algorithms include the Metropolis-Hastings algorithm of Hastings (1970) and the Gibbs-sampler (S. Geman and D. Geman 1984), which is a generalization of the former (Geyer 2011).

There are several strong benefits with MCMC. It is, for instance, applicable to most Bayesian models, it converges in distribution to the target distribution and it is asymptotically unbiased (see e.g., Neal (1993)). A downside to these classical MCMC methods, however, is its slowness in convergence, especially for models with a larger parameter space ((Hoffman and Gelman 2014), (Neal 1993)). As an alternative, computationally more efficient methods within the MCMC methodology have been proposed. Specifically, we consider the Hamiltonian Monte Carlo (HMC) method (Duane et al. 1987) with No-U-Turn Sampling (NUTS) (Hoffman and Gelman 2014), as it is implemented in the increasingly popular Stan software (Stan Development Team 2019). Stan

provides an interface with R through the library Rstan, and we will use this library in our analysis.

Hamiltonian Monte Carlo

The idea behind the Hamiltonian Monte Carlo method is to consider the target probability system as part of a Hamiltonian system of equations, from which we can find a Markov transition Betancourt (2016). The HMC then avoids the random walk behavior of most MCMC samplers, which scales poorly for complex models and leads to a slow exploration of the target distribution, and by that, slow convergence (see e.g., Betancourt (2018)). The HMC method was initially developed by Duane et al. (1987) in the field of quantum physics, but its usefulness in the statistical field was described by Neal (1993) and the method has since become increasingly popular.

The purpose of the method is similar to that of the MCMC method, namely to find some target density $\pi(\boldsymbol{\theta})$ of some parameters $\boldsymbol{\theta}$, usually given some observations \boldsymbol{y} . The HMC does this by introducing some auxiliary momentum variables $\boldsymbol{\rho}$, and considering the joint distribution between these momentum variables and the parameters $\boldsymbol{\rho}$:

$$\pi(\boldsymbol{\rho}, \boldsymbol{\theta}) = \pi(\boldsymbol{\rho} | \boldsymbol{\theta})\pi(\boldsymbol{\theta}). \quad (4.1.1)$$

Usually, the density of $\boldsymbol{\rho}$ is chosen so that it is independent of $\boldsymbol{\theta}$, and Equation 4.1.1 reduces to

$$\pi(\boldsymbol{\rho}, \boldsymbol{\theta}) = \pi(\boldsymbol{\rho})\pi(\boldsymbol{\theta}).$$

Specifically, in Stan, the density for $\boldsymbol{\rho}$ is chosen to be a multivariate normal distribution,

$$\boldsymbol{\rho} \sim \text{multiNormal}(0, M),$$

where M is the euclidean metric (Stan Development Team 2019). A Hamiltonian can be defined on this joint density as

$$\begin{aligned} H(\boldsymbol{\theta}, \boldsymbol{\rho},) &= -\log(\pi(\boldsymbol{\theta}, \boldsymbol{\rho})) \\ &= -\log(\pi(\boldsymbol{\rho} | \boldsymbol{\theta})) - \log(\pi(\boldsymbol{\rho})) \\ &= -\log(\pi(\boldsymbol{\rho})) - \log(\pi(\boldsymbol{\rho})) \\ &= T(\boldsymbol{\rho}) + V(\boldsymbol{\theta}), \end{aligned}$$

when the density of $\boldsymbol{\rho}$ is independent of $\boldsymbol{\theta}$. $H(\boldsymbol{\theta}, \boldsymbol{\rho})$ satisfies Hamilton's equations:

$$\begin{aligned} \frac{d\boldsymbol{\theta}}{dt} &= \frac{\partial H}{\partial \boldsymbol{\rho}} = \frac{\partial T}{\partial \boldsymbol{\rho}} \\ \frac{d\boldsymbol{\rho}}{dt} &= -\frac{\partial H}{\partial \boldsymbol{\theta}} = -\frac{\partial V}{\partial \boldsymbol{\theta}} \end{aligned} \tag{4.1.2}$$

As an analogy to the physical systems usually described by Hamiltonian equations, $T(\boldsymbol{\rho})$ is often referred to as the kinetic energy of the system, and the term $V(\boldsymbol{\theta})$ is often referred to as the potential energy of the system. These types of systems of equations are frequently used e.g., in the field of quantum physics (Betancourt 2018), and they can be solved by the leapfrog algorithm for numerical integration, which is adapted to be stable for Hamiltonian systems (Stan Development Team 2019).

The sampling in most HMC algorithms, including the HMC implementation of Stan, is done in two steps.

- First, a value for the momentum $\boldsymbol{\rho}$ is independently sampled.
- Then, the system in Equation 4.1.2 is developed using the leapfrog method for numerical integration, starting at the newly sampled value of $\boldsymbol{\rho}$ and the value of $\boldsymbol{\theta}$ from the previous sample.

One step of the leapfrog method, incrementing the trajectory of $(\boldsymbol{\theta}, \boldsymbol{\rho})$ for a time step ε , consists of the following three substeps:

$$\begin{aligned}\boldsymbol{\rho}_{t+\frac{1}{2}} &\leftarrow \boldsymbol{\rho}_t - \frac{\varepsilon}{2} \frac{\partial V}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_t) \\ \boldsymbol{\theta}_{t+1} &\leftarrow \boldsymbol{\theta}_t + \varepsilon \boldsymbol{\rho}_{t+\frac{1}{2}} \\ \boldsymbol{\rho}_{t+1} &\leftarrow \boldsymbol{\rho}_{t+\frac{1}{2}} - \frac{\varepsilon}{2} \frac{\partial V}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_{t+1}).\end{aligned}$$

These leapfrog steps are performed a total of L times, developing the trajectory for a time εL , and the resulting state $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\rho}})$ becomes the proposed sample.

Stan uses the NUTS algorithm, as described in the following Section 4.1, to perform the leapfrog development, and then uses a multinomial sampling procedure to find the final sample given the proposed sample $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\rho}})$. In the case where NUTS is not applied, some correcting Metropolis step is typically applied to find the final sample. The reason for this final step is that the leapfrog method is approximate, and so the proposal samples should not be used directly (Stan Development Team 2019).

The performance and sampling efficiency of HMC greatly depends on the choice of the parameters of the algorithm:

- The step size, or discretization in time, ε
- The metric M
- The number of leapfrog integration steps L

(Hoffman and Gelman (2014), Stan Development Team (2019)). To obtain efficient sampling, ε should be balanced so that a sufficiently large number of proposal samples are accepted, while still being large enough so that the integration procedure is not too computationally demanding. Similarly, the choice of L involves a trade-off between too large values of L resulting in making unnecessary steps in computation and too low values of L which give short leapfrog trajectories and sampling patterns that exhibit random walk behavior. Lastly, the inverse metric M^{-1} should be a sufficiently good estimate

of the posterior covariance. If it is not, ε must be kept small, and then larger values are needed for L as a compensation (Stan Development Team 2019).

The Stan methodology automatically optimizes all of these parameters. ε is adjusted to match a target acceptance rate, the warm-up samples are used to estimate M and the No-U-Turn Sampler is used to dynamically adapt L . We refer to the Stan reference manual (Stan Development Team 2019) for a detailed description of the adaption of ε and M and briefly outline the concept of the NUTS algorithm for adaption of L in the following section.

No-U-Turn Sampler

The No-U-Turn Sampler (NUTS) proposed by Hoffman and Gelman (2014) is a method for dynamical allocation of the number of leapfrog steps L in an HMC procedure for each sampling step. This removes the need for a potentially computationally demanding preliminary analysis to find an optimal value for L , while still keeping the computation time for the procedure low (Hoffman and Gelman 2014).

The driving idea behind NUTS is that the ideal proposal sample parameter $\tilde{\theta}$ of the HMC is as far as possible from the initially proposed parameter θ . This is because when successive samples are too close to one another, the chain displays the undesired random walk behavior and becomes slow in convergence. The leapfrog trajectories of the HMC act like loops, by moving away from the initial parameter values with the first integration steps, but at some point turning back towards the initial proposal (Hoffman and Gelman 2014). Consequently, the NUTS approach strives to find the number of integration steps L that corresponds to the point of the trajectory right before it turns back towards the initial values. Hoffman and Gelman (2014) finds this point by considering the quantity

$$\frac{d}{dt} \frac{(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)}{2} = (\tilde{\theta} - \theta) \frac{d}{dt} (\tilde{\theta} - \theta) = (\tilde{\theta} - \theta) \cdot \tilde{\rho}, \quad (4.1.3)$$

where θ is the initial parameter proposal, $\tilde{\theta}$ is the current parameter proposal and $\tilde{\rho}$ is the current momentum. This quantity is proportional to what can be interpreted as the amount of progress away from the initial parameter estimate θ one would make by running the algorithm for dt more time.

The above discussion would suggest an algorithm that runs leapfrog integration until the quantity in Equation 4.1.3 becomes less than zero, or conceptually speaking, when the trajectory begins to make a U-turn. However, this approach does not guarantee reversibility, and by that, it neither guarantees convergence to the correct distribution (Hoffman and Gelman 2014). To adjust for this, the NUTS algorithm develops the system in Expression 4.1.2 both forwards and backwards in time, creating a balanced binary tree rather than a single trajectory. The development is then halted when the binary tree begins to turn back on itself, or in other words, when the two trajectories from development of the system forwards and backwards in time begin to approach one another.

4.2 INTEGRATED NESTED LAPLACE APPROXIMATIONS

The Integrated Nested Laplace Approximation (INLA) is a method for approximate Bayesian inference on latent Gaussian models (LGM), first proposed by Rue et al. (2009). INLA is made readily available through the R package R-INLA and a thorough explanation of the method and the related framework can be found in Rue et al. (2009) and at <https://www.r-inla.org>. Bayesian inference with INLA is applicable to the class of Bayesian models named Latent Gaussian Models (LGMs) (Rue et al. 2009).

LGMs are hierarchical models, where the observations \mathbf{y} are conditionally independent given a latent Gaussian random field \mathbf{x} and some hyperparameters θ . LGMs are usually expressed through a predictor η_i related to the mean μ_i of observation y_i through some known link function $g^{-1}(\cdot)$:

$$\mu_i = g^{-1}(\eta_i). \quad (4.2.1)$$

The general form of the predictor is

$$\eta_i = \alpha + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_f} f^{(k)}(u_{ki}) + \varepsilon_i \quad (4.2.2)$$

(Rue et al. 2009). Here, α is the intercept, β are the linear effects on some covariates \mathbf{z} , the $f^{(k)}$'s are random effects on some covariates \mathbf{u} and ε_i is the error term. The model can be summarized as

$$\begin{aligned} \mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta} &\sim \prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta}) \\ \mathbf{x} \mid \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta})) \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}). \end{aligned} \quad (4.2.3)$$

The latent field \mathbf{x} is given by

$$\mathbf{x} = (\boldsymbol{\eta}, \alpha, \boldsymbol{\beta}, \mathbf{f}) \quad (4.2.4)$$

(Martino and Riebler 2019). The element that distinguishes LGMs from other Bayesian models is the assumed Gaussian prior on the latent field \mathbf{x} (Rue et al. 2009). Here $\mathbf{Q}(\boldsymbol{\theta})$ is the precision matrix (inverse of covariance matrix). $\boldsymbol{\theta}$ may have any suitable prior distribution.

The INLA method assumes that the Gaussian field is endowed with a Markov property, that is, \mathbf{x} is a Gaussian Markov Random Field (GMRF) (Rue and Held 2005). This entails that the matrix $\mathbf{Q}(\boldsymbol{\theta})$ is sparse. The methodology gives approximations to the marginal posterior distributions of the latent field \mathbf{x} and the associated hyperparameters $\boldsymbol{\theta}$. Rue et al. (2009) describe the method of INLA in its entirety, while Martino and Riebler (2019) give an updated and concise summary of the main components of the INLA method. We will briefly outline the estimation procedure of INLA here, and refer to these papers for a more detailed description.

4.2.1 The method of INLA

INLA produces deterministic approximations to the univariate marginal posterior distributions for the latent field \mathbf{x} and the hyperparameter vector $\boldsymbol{\theta}$ of LGMs. These quantities are given by the integrals

$$\pi(\theta_j | \mathbf{y}) = \int \int \pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{x}) d\mathbf{x} d\boldsymbol{\theta}_{-j} = \int \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_j \quad (4.2.5)$$

$$\pi(x_i | \mathbf{y}) = \int \int \pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{x}) dx_{-i} d\boldsymbol{\theta} = \int \pi(x_i | \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (4.2.6)$$

As described by Martino and Riebler (2019), the integral with respect to $\boldsymbol{\theta}$ can relatively easily be solved with numerical integration. To do this, approximations of $\pi(\boldsymbol{\theta} | \mathbf{y})$ and $\pi(x_i | \boldsymbol{\theta}, \mathbf{y})$ are needed. INLA approximates $\pi(\boldsymbol{\theta} | \mathbf{y})$, at a certain value $\boldsymbol{\theta}^k$ of the hyperparameter vector, as

$$\tilde{\pi}(\boldsymbol{\theta}^k | \mathbf{y}) \propto \frac{\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}^k) \pi(\mathbf{x} | \boldsymbol{\theta}^k) \pi(\boldsymbol{\theta}^k)}{\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}^k, \mathbf{y})}. \quad (4.2.7)$$

Here, $\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}^k, \mathbf{y})$ is a Gaussian approximation of $\mathbf{x} | \boldsymbol{\theta}^k, \mathbf{y}$, which is constructed from the value and the curvature at the mode. A Cholesky factorization of the matrix $Q(\boldsymbol{\theta})$ is needed to evaluate the denominator of Expression 4.2.7. Typically, this procedure is computationally expensive. However, the GMRF assumption of Rue et al. (2009) ensures sparsity of $Q(\boldsymbol{\theta})$, which makes the Cholesky factorization faster. In other words, as long as $\boldsymbol{\theta}$ is not too large, the above step is not very computationally demanding.

Next, an approximation for $\pi(x_i | \boldsymbol{\theta}, \mathbf{y})$ is needed. INLA offers three ways of doing this, which are all explained in Rue et al. (2009). The default method is a "Simplified Laplace approximation" (Rue and Held 2005), where the skewness and location of the marginal of the Gaussian approximation $\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})$ is corrected by a Taylor's series expansion around the mode of a Laplace approximation (Tierney and Kadane 1986). This method serves as a compromise between the accuracy of a full Laplace approximation and the speed of the Gaussian approximation (Martino and Riebler 2019).

A scheme describing the INLA algorithm is as follows:

1. $\boldsymbol{\theta}$ is explored through the approximation $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$, and K support points $\{\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^K\}$ are located in the high-density area of $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$
2. $\tilde{\pi}(\boldsymbol{\theta}^k | \mathbf{y})$ is calculated for each support point $\boldsymbol{\theta}^k$, $k = 1, \dots, K$ by Equation 4.2.7.
3. $\tilde{\pi}(x_i | \boldsymbol{\theta}^k, \mathbf{y})$ is calculated for each support point $\boldsymbol{\theta}^k$, $k = 1, \dots, K$ using a Simplified Laplace approximation, a full Laplace approximation or a Gaussian approximation
4. Numerical integration is used to solve Integral 4.2.6 by

$$\tilde{\pi}(x_i | \mathbf{y}) = \sum_{k=1}^K \tilde{\pi}(x_i | \boldsymbol{\theta}^k, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}^k | \mathbf{y}) \Delta_k,$$

where Δ_k are the appropriate weights for the support points. Integral 4.2.5 is solved in a similar manner

The main advantage of INLA compared to other methods for Bayesian inference is its computational power. Compared to the widely used Markov Chain Monte Carlo (MCMC) method, for instance, it has been shown that INLA can produce more accurate results in seconds or minutes than MCMC can produce in hours or days (Rue et al. 2009). This often makes INLA, when applicable, the preferred method of inference when working with model fitting, as it allows for efficient comparison of model fits and it opens up the possibility of sensitivity analysis, comparing the results when varying the prior distributions. Still, the usefulness of INLA is restricted by the requirements on the models it is applicable to.

4.2.2 Restrictions of the INLA framework

A limitation of the INLA approach is that it imposes some restrictions on the model used in the Bayesian inference. These restrictions can be formulated in three requirements:

1. It should be possible to write the model as an LGM
2. Each observation y_i should depend on the latent field \mathbf{x} only through the predictor η_i
3. The predictors η_i of the LGM should depend linearly on α , β and \mathbf{f} of Expression 4.2.3

(Martino and Riebler 2019).

Out of all the models discussed in Section 2, only the models where $\beta_x^{(i)}$ are constant fulfill these criteria. Among these is the APC model, for which INLA has indeed been widely and successfully used for performing inference (e.g., Riebler (2010)). However, for the large body of Lee-Carter types of models, this is not the case. This means that so far, it has not been possible to use the INLA framework to perform inference on Lee-Carter types of mortality models.

4.3 LINEARIZATION OF NON-LINEAR PREDICTORS WITHIN THE INLA FRAMEWORK

Recently, a new methodology was proposed that enables the usage of the INLA methodology on LGMs that include some non-linear terms in the predictor $\eta_{x,t}$. The method was first proposed by Bachl et al. (2019) in the setting of ecological surveys, but it is widely applicable to several other fields. We call this method `inlabru` after the associated R library that implements it. The `inlabru` method involves finding a linearization of the non-linear predictor

through a fixed-point iteration using INLA and then setting this linearization as the predictor in the final INLA run. Information about the installation and usage of the `inlabru` R library can be found at Lindgren and Bachl (2021). Here, we outline the approach of Bachl et al. (2019) for the linearization and following model approximations of LGMs with non-linear predictors. For a more thorough explanation of the method and the application to ecological data, we refer to their original paper (Bachl et al. 2019) and to Lindgren and Bachl (2021).

We assume that we have an LGM, with a structure as described in 4.2.3, but where the likelihood depends on a non-linear predictor $\tilde{\eta}(\boldsymbol{x})$:

$$\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta} \sim \pi(\boldsymbol{y} \mid \tilde{\eta}(\boldsymbol{x}), \boldsymbol{\theta}), \quad (4.3.1)$$

where \boldsymbol{x} is the latent field. `inlabru` then uses a linearization of $\tilde{\eta}(\boldsymbol{x})$ to be able to successfully run INLA. The linearization is found using a Taylor approximation around some point \boldsymbol{x}_0 :

$$\bar{\eta}(\boldsymbol{x}) = \tilde{\eta}(\boldsymbol{x}_0) + B(\boldsymbol{x} - \boldsymbol{x}_0), \quad (4.3.2)$$

where B is the Jacobian matrix of $\tilde{\eta}(\boldsymbol{x})$ evaluated at \boldsymbol{x}_0 . This linearized predictor is then used to run INLA, by substituting $\bar{\eta}(\boldsymbol{x})$ for $\tilde{\eta}(\boldsymbol{x})$ in the likelihood models to obtain an approximation for the likelihood:

$$\bar{\pi}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \pi(\boldsymbol{y} \mid \bar{\eta}(\boldsymbol{x}), \boldsymbol{\theta}) \approx \pi(\boldsymbol{y} \mid \tilde{\eta}(\boldsymbol{x}), \boldsymbol{\theta}) = \tilde{\pi}(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}). \quad (4.3.3)$$

By running INLA with this approximated likelihood, approximated posterior distributions are obtained for the latent effects, hyperparameters, and the predictor $\tilde{\eta}(\boldsymbol{x})$. `inlabru` finds the optimal linearization point \boldsymbol{x}_0 through a fixed-point iteration with INLA. For each step s in the iteration, the point \boldsymbol{x}' is found that maximizes the posterior distribution for the latent effects that resulted from the INLA approximation using the linearization from the last step \boldsymbol{x}_{s-1} . We denote this posterior approximated distribution as $\bar{\pi}_{\boldsymbol{x}_{s-1}}(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta})$. The fixed-point scheme to find the optimal linearization point \boldsymbol{x}_0 is given in Algorithm 1 (Lindgren and Bachl 2021). We note that the full scope of the `inlabru` method has not yet been fully investigated (Lindgren and Bachl 2021).

Since the method is based on the linearization of a non-linear expression, we might expect it to fail if the non-linearity is too extreme.

Algorithm 1 Fixed-point scheme for finding optimal linearization point x_0

```

 $s \leftarrow 0$ 
 $u_{s=0} \leftarrow$  initial linearization point
while not convergence do
   $B \leftarrow$  Jacobian matrix of  $\tilde{\eta}(x)$  evaluated at  $u_s$ 
   $\bar{\eta}(x)_{u_s} \leftarrow \tilde{\eta}(u_s) + B(x - u_s)$ 
  run INLA with  $\bar{\eta}(x)_{u_s}$  as the linear predictor
   $\bar{\pi}_{u_s}(\theta | y) \leftarrow$  posterior approximate distribution for  $\theta$  from INLA run
   $\bar{\pi}_{u_s}(x | y, \theta) \leftarrow$  posterior approximate distribution for  $x$  from INLA run
   $\hat{\theta} \leftarrow \operatorname{argmax}_{\theta} \bar{\pi}_{u_s}(\theta | y)$ 
   $u' \leftarrow \operatorname{argmax}_x \bar{\pi}_{u_s}(x | y, \hat{\theta})$ 
  if  $u_s \approx u'$  then ▷ Convergence
     $x_0 \leftarrow u'$ 
    return  $x_0$ 
  else
     $u_{s+1} \leftarrow (1 - \lambda)u_s + \lambda u'$  such that  $\lambda$  minimizes  $\|\tilde{\eta}(u_{s+1}) - \tilde{\eta}(u')\|$ 
     $s \leftarrow s + 1$ 
  end if
end while

```

SIMULATION STUDY

In this chapter, we perform a simulation study to compare the performance of `inlabru` and Stan on a series of mortality models with increasing complexity. To make the investigation as simple as possible, we only look at models with one multiplicative term $\beta_x \kappa_t$, and we do not consider cohort effects in this part of the investigation. This is because we are mainly interested in the results of `inlabru` when we apply it specifically to the multiplicative term, and these results should extend to models with multiple terms for the period effects or models with cohort effects.

We apply `inlabru` and Stan to two sets of mortality models, one set with Gaussian likelihoods and one set with Poisson likelihoods. The results are described in the following two sections.

The full code used to produce the results in this chapter can be found at <https://github.com/Helenerb/Masters-thesis>.

5.1 MODELS WITH GAUSSIAN LIKELIHOODS

We consider models of the form

$$m_{x,t} = Y_{x,t} / E_{x,t}, \quad \log(m_{x,t}) \sim \mathcal{N}(\eta_{x,t}, 1/\tau_\varepsilon). \quad (5.1.1)$$

Here, $m_{x,t}$ are the observed mortality rates for age groups $x = 1, \dots, X$ during periods $t = 1, \dots, T$. The values of $m_{x,t}$ are found from the observed number of deaths $Y_{x,t}$ and the exposed population $E_{x,t}$.

We include two versions of the predictor $\eta_{x,t}$. The first has a linear structure:

$$\eta_{x,t} = \alpha_x + \kappa_t.$$

For this predictor, `inlabru` should act exactly like INLA, which we know should produce very similar results to Stan (Rue et al. 2009). We include it mainly to check that our implementation is correct. Secondly, we consider the classical Lee-Carter predictor:

$$\eta_{x,t} = \alpha_x + \beta_x \cdot \kappa_t.$$

For both models, we assign the following priors to the time effects:

$$\begin{aligned} \alpha_x &\sim \text{RW1}(\tau_\alpha) \\ \beta_x &\sim \text{iid}(\tau_\beta) \\ \kappa_t &\sim \text{RW1}(\tau_\kappa). \end{aligned} \tag{5.1.2}$$

Here $\text{RW1}(\tau)$ denotes the first-order random walk and $\text{iid}(\tau)$ denotes the independent random noise model, as defined by Rue and Held (2005). τ denotes the precisions of the models. We impose the identifying constraints

$$\sum_x \beta_x = 1, \quad \sum_x \kappa_t = 0. \tag{5.1.3}$$

Finally, we let the two models have both fixed and varying hyperparameters τ_α , τ_β , τ_κ and τ_ε . In the cases where we let the hyperparameters vary, we assign them vague Gamma-priors:

$$\begin{aligned} \theta_\alpha &= \log(\tau_\alpha), & \theta_\alpha &\sim \log\mathcal{Gamma}(1, 0.00005) \\ \theta_\beta &= \log(\tau_\beta), & \theta_\beta &\sim \log\mathcal{Gamma}(1, 0.00005) \\ \theta_\kappa &= \log(\tau_\kappa), & \theta_\kappa &\sim \log\mathcal{Gamma}(1, 0.005) \\ \theta_\varepsilon &= \log(\tau_\varepsilon), & \theta_\varepsilon &\sim \log\mathcal{Gamma}(1, 0.00005) \end{aligned} \tag{5.1.4}$$

Model name	Predictor	Hyperparameters
G1lin	$\eta_{x,t} = \alpha_x + \kappa_t$	Fixed
G2lin	$\eta_{x,t} = \alpha_x + \kappa_t$	Gamma
G1	$\eta_{x,t} = \alpha_x + \beta_x \cdot \kappa_t$	Fixed
G2	$\eta_{x,t} = \alpha_x + \beta_x \cdot \kappa_t$	Gamma

Table 1: The models included in the analyses of Section 5.1. All models have the likelihood structure of Expression 5.1.1

where $\theta = \log(\tau) \sim \log\mathcal{Gamma}(a, b)$ if $\tau \sim \mathcal{Gamma}(a, b)$. We define the priors on the log-precision because that is the standard implementation in INLA (Rue and Held 2005). Variations of vague Gamma priors are commonly used in literature, see e.g., Czado et al. (2005), Wong et al. (2018).

A summary of the four models that we analyze in this section is presented in Table 1.

Preliminary Implementation Study

In the `inlabru` library, the random walk models and the constraints are implemented in a specific way, as defined in Rue and Held (2005). In Stan, there are several possible ways to define both the random walk models and the constraints (see e.g., Knorr-Held and Rainer (2001), Morris (2019) and discussions at e.g., Stan Discussion Forum (2019)). To be sure that the specific Stan implementations do not influence the results of the comparison, we perform a preliminary analysis to check that our implementation of these model components in Stan are equivalent to those of `inlabru`. The results of this study show that this is indeed the case. The full study and the results can be found in Appendices A.1 and A.2.

Synthetic Data

We produce four sets of synthetic data for the simulation study. These sets of data are all based on the German male lung cancer data, as described in

Chapter 3. Section A.3 gives a detailed description of how these sets of data are generated, and we present them briefly here.

We have one main set of data:

- Full.1 : Based on the full male lung cancer data. It has dimensions $x = 1, \dots, 18$, $t = 1, \dots, 18$ and it contains some observations where $Y_{x,t} = 0$.

Additionally, we have three supporting sets of data. We only discuss the results of these data when they provide some complementary insight to the results of the Full.1 data.

- Reduced.1 : Based on male lung cancer data for ages above 45 years old. This data has generally higher values for the counts $Y_{x,t}$ compared to the Full.1 data. It has dimensions $x = 1, \dots, 9$, $t = 1, \dots, 18$.
- Full.2 : Based on the full male lung cancer data. This data has the same dimensions as the Full.1 data, but no occurrences of $Y_{x,t} = 0$. κ_t is adjusted so that it is not too small compared to the age effect, and β_x is adjusted to be positive for all values of x .
- Reduced.2 : Based on male lung cancer data for ages above 10 years. This data has slightly different dimensions compared to the Full.1 data, $x = 1, \dots, 16$, $t = 1, \dots, 18$ and no occurrences of $Y_{x,t} = 0$.

Figure 5.1.1 displays the synthetic observed log-mortality of the Full.1 data. Similar plots of the Reduced.1, Full.2, and Reduced.2 are included in Appendix A.3 in Figures A.3.1, A.3.2 and A.3.3.

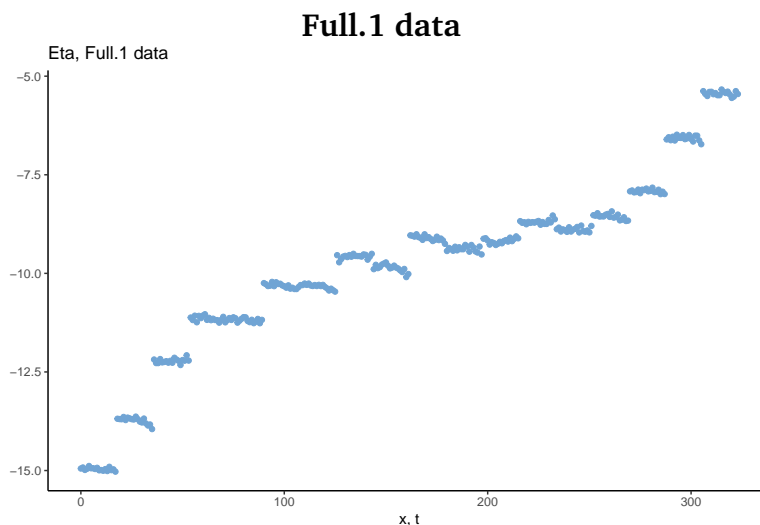


Figure 5.1.1: Observed log-mortality rates $\log(m_{x,t})$ of the Full.1 data for points $(x, t) = (1, 1), \dots, (1, 18), \dots, (18, 1), \dots, (18, 18)$

5.1.1 *G1lin*: Linear Gaussian Model with Fixed Precisions

We first evaluate the *G1lin* model of Table 1. This model is linear, with the hyperparameters fixed to the given values

$$\tau_{\alpha} = 1.96, \quad \tau_{\kappa} = 336, \quad \tau_{\varepsilon} = 420.$$

We apply this model to the Full.1 data using both `inlabru` and `Stan`. The Full.1 data is generated using a non-linear model, so we do not expect the estimation results to necessarily fit the data very well. However, we expect `inlabru` and `Stan` to give similar results, and hence we compare the results of `inlabru` and `Stan` to each other, but not to the Full.1 data. This applies to all linear models in this chapter.

The estimation results of `inlabru` and `Stan` are practically identical. This is as expected, since `inlabru` should produce the same results as INLA for models with linear predictors, and it has been thoroughly shown that the INLA approximations are very close to results produced by MCMC methods (Rue et al. 2009).

Plots of the results are included in the Appendix, where Figure B.1.1a displays some of the values of $\eta_{x,t}$ as estimated by Stan and inlabru, and Figure B.1.1b displays the corresponding estimated random effects.

5.1.2 G2lin: Linear Gaussian Model with Gamma Precisions

Model G2lin differs from the model in Section 5.1.1 by the fact that the hyperparameters τ_α , τ_κ , and τ_ϵ are not fixed, but assigned the Gamma-priors of Expression 5.1.4.

In this case, as well, results from the two inferential methods are practically identical. Detailed plots with a comparison of the results are displayed in Appendix B, in Figures B.1.2 and B.1.3.

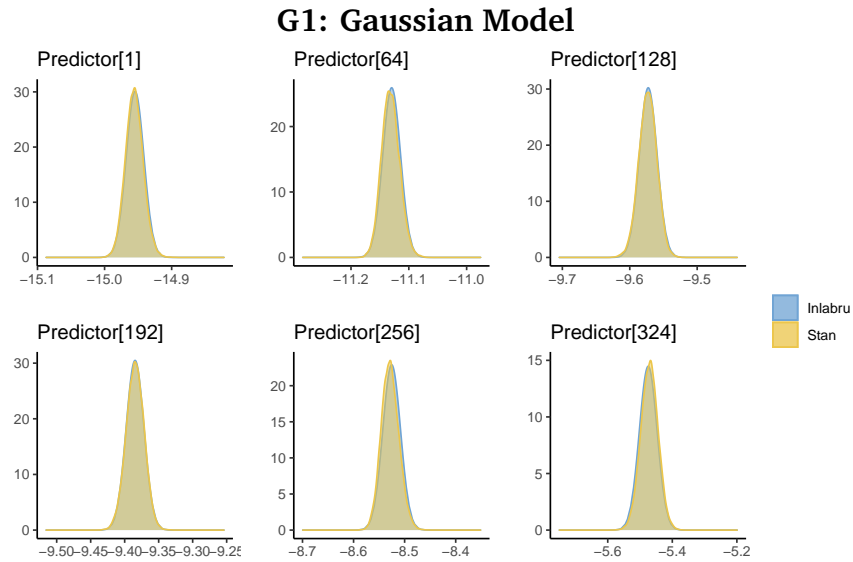
5.1.3 G1: Gaussian Model with Fixed Precisions

We now move on to the nonlinear models. First, we consider the G1 model where the hyperparameters are fixed to the following values:

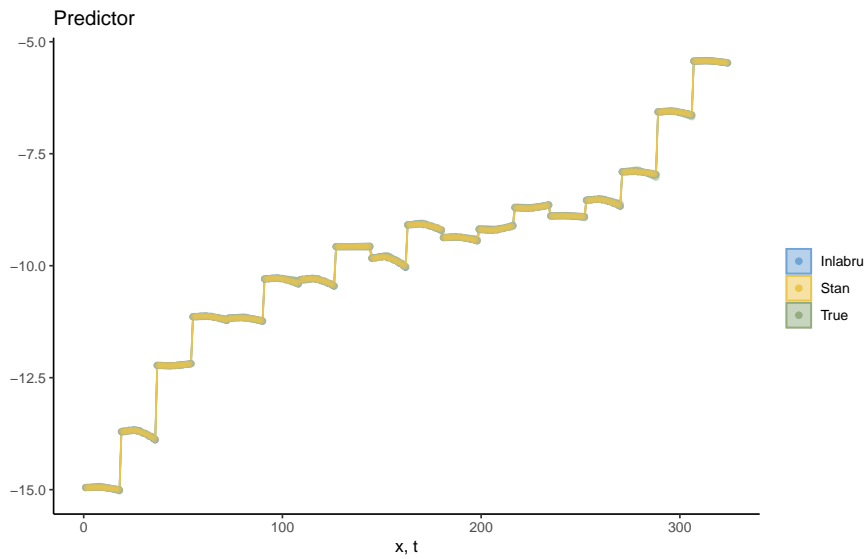
$$\tau_\alpha = 1.96, \quad \tau_\beta = 202, \quad \tau_\kappa = 30, \quad \tau_\epsilon = 420.$$

We apply this model to the Full.1 data and compare the estimate for the predictor $\eta_{x,t}$.

For most points (x, t) , the value of $\eta_{x,t}$ is estimated very similarly by the two methods, as can be seen from the posterior marginal distributions of $\eta_{x,t}$ that are displayed in Figure 5.1.2. We also see, from Figure 5.1.2b that both estimates follow the true values of $\eta_{x,t}$ quite well. However, for some points, we observe a slight shift between the estimations of inlabru and Stan.



(a) Marginal distributions for some arbitrary values of $\eta_{x,t}$



(b) Estimated means and 95% bounds together with observed values of $\eta_{x,t}$

Figure 5.1.2: Estimation results of linear predictor $\eta_{x,t}$ from applying the G1 model to the Full.1 data using Stan and inlabru

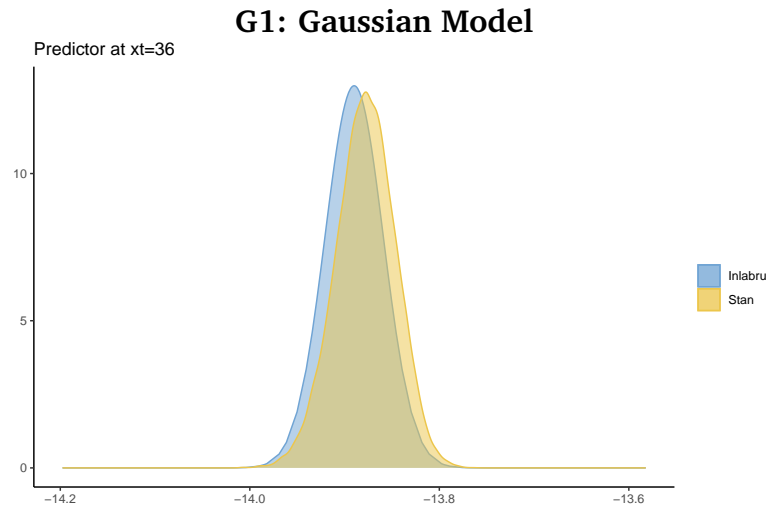


Figure 5.1.3: Estimated predictor $\xi_{x=2,t=18}$ from applying the G1 model to the Full.1 data using Stan and inlabru

Figure 5.1.3 displays the estimations of the value of $\eta_{x=2,t=18}$, where the difference is at its largest. We attribute this difference to the approximation that inlabru makes in the linearization of the predictor $\eta_{x,t}$. The difference that we observe is small, and we doubt that it would provide any practical problems.

Figure 5.1.4 displays the estimated random effects, and we observe that these are also estimated slightly differently by Stan and inlabru. More specifically, the two factors in the multiplicative term, β_x , and κ_t are estimated differently by the two methods, while the estimates for α_x seem to be identical. Plots of the posterior marginal distributions of β_x and κ_t can be found in the Appendix, in Figure B.1.4. In general, the estimations of the random effects seem accurate, as most of the true values are contained within the 95% estimation bounds of both inlabru and Stan.

We apply the G1 model to the Reduced.1 data set as well, to investigate how the results of inlabru and Stan compares for a different set of data, where the values of the predictor are generally higher. These results are included in the Appendix in Figures B.1.5 and B.1.6. For the Reduced.1 data, the difference in the estimation between inlabru and Stan is only barely observable.

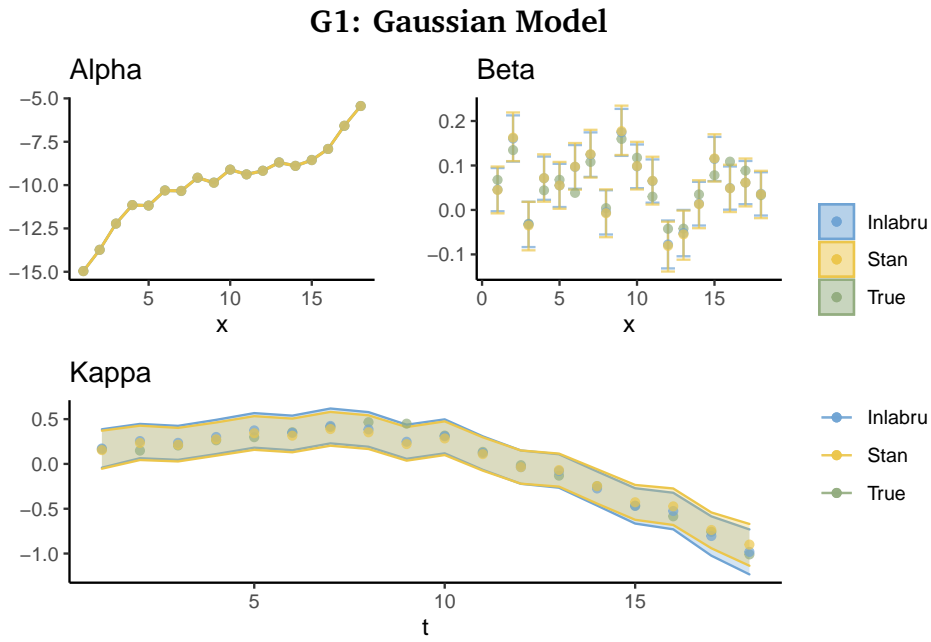


Figure 5.1.4: Estimated random effects α_x , β_x and κ_t from applying the G1 model to the Full.1 data using Stan and inlabru

One possible explanation of the different behavior for different sets of data, is that there are some unidentifiability between β_x and κ_t in the Full.1 data, that is not as apparent in the Reduced.1 data. Looking at the estimates of κ_t in Figures 5.1.4, we see that the change points of κ_t are similar in the Stan and inlabru estimation, and that the difference in the estimations of κ_t lies in the slope. The inlabru estimations are steeper than the Stan estimation, but they have the same shape. We note that theoretical identifiability of the basic Lee-Carter model which have been shown multiple times (see e.g Hunt and Blake (2020a)), does not guarantee identifiability in practice for all realized data.

5.1.4 G2: Gaussian Model with Gamma Precisions

We now let the hyperparameters vary, and we use the gamma priors of Expression 5.1.4 for the hyperparameters τ_α , τ_β , τ_κ and τ_ε .

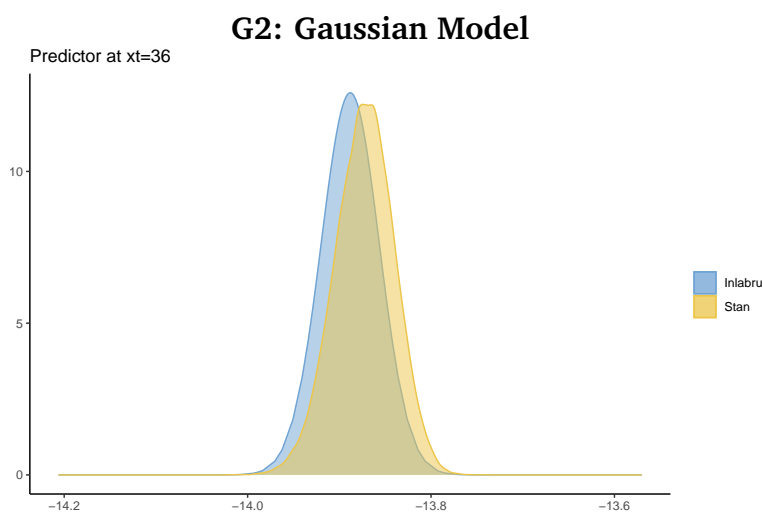
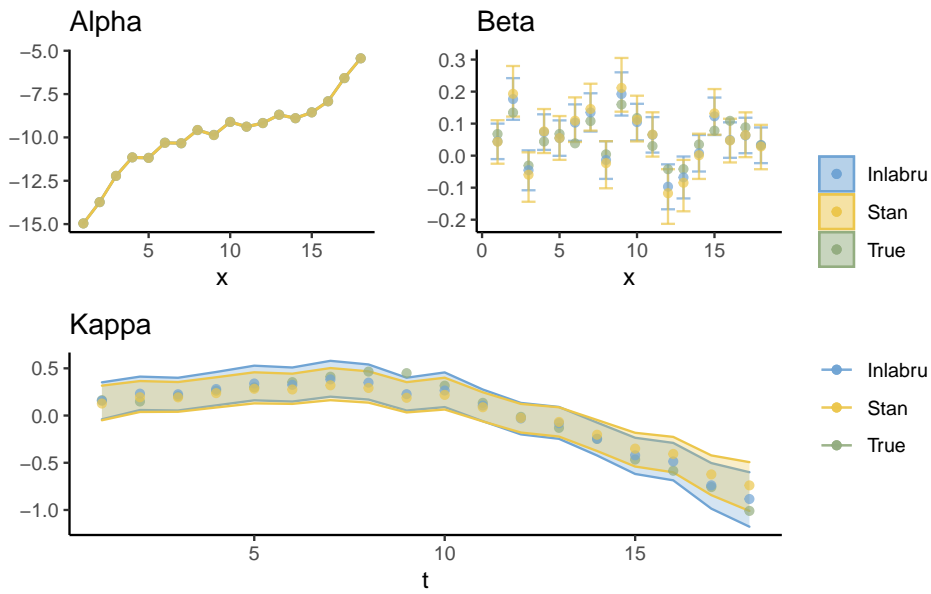


Figure 5.1.5: Estimation results of linear predictor $\eta_{x=2,t=18}$ from applying the G2 model to the Full.1 data using Stan and inlabru

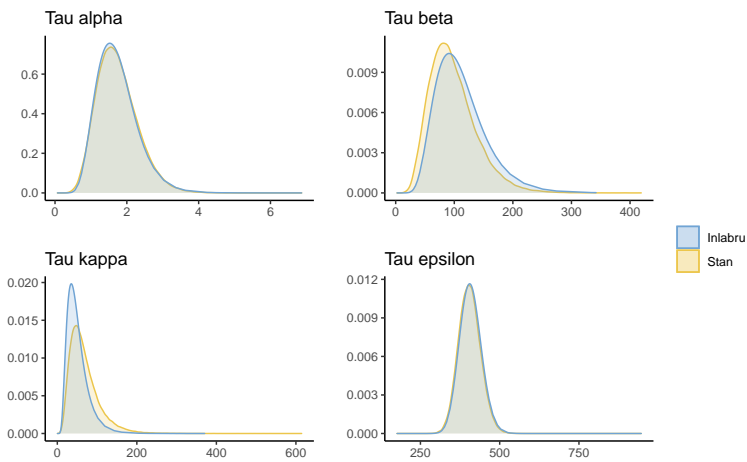
The results when we apply this model to the Full.1 data are displayed in Figures 5.1.5 and 5.1.6. Also in this case, the estimations from inlabru and Stan are similar, but not identical. Figure 5.1.5 displays the estimated value for $\eta_{x=2,t=18}$, which is one of the points where the difference between the two estimation methods seem to be largest. We observe that this difference is of the same size as the largest difference for the G1 model (Figure 5.1.3). The difference in the random effects however, is slightly more extreme compared to the estimated random effects of the G1 model. We observe that there are some differences in the estimation of the hyperparameters as well, as can be seen in Figure 5.1.6b. We note that the difference between inlabru and Stan for the predictor do not increase for the G2 model compared to the G1 model. This underlines the impression that there is some unidentifiability between the effects κ_t and β_x that inlabru and Stan interprets differently.

Complementary results for the G2 model can be found in Appendix B in Figures B.1.7 and B.1.8.

G2: Gaussian Model



(a) Estimated random effects α_x (upper, left), β_x (upper, right), κ_t (lower)



(b) Estimated hyperparameters τ_α (upper, left), τ_β (upper, right), τ_κ (lower, left), τ_ϵ (lower, right)

Figure 5.1.6: Estimation results from applying the G2 model to the Full.1 data using Stan and inlabru

5.1.5 Conclusion on Comparison of Gaussian Models

After considering the results of applying `inlabru` and Stan to four different models with Gaussian likelihoods, we make two remarks. The first is that `inlabru` seem to give the same results as Stan in the case where the predictor is linear. This is a clear indication that our implementations of the Stan model is equivalent to that of `inlabru`, and it confirms that `inlabru` performs in the same way as INLA for models with linear predictors.

Secondly, we have seen that `inlabru` produces similar, but not equal results to Stan, in the cases with non-linear models. Furthermore, we have observed that the difference between `inlabru` and Stan seem to depend on the underlying data. We attribute some of these difference to unidentifiability between the time effects in the data and some to biases in the linear approximation of `inlabru`. However, we are not able to quantify how much of the difference arises from each of these causes.

Complementary plots of the results in Section 5.1, that have not been explicitly discussed here, are included in Appendix B.1.

5.2 MODELS WITH POISSON LIKELIHOODS

We now consider models of the form

$$Y_{x,t} \sim \mathcal{P}oisson(E_{x,t} \cdot e^{\eta_{x,t}}). \quad (5.2.1)$$

Here, $Y_{x,t}$ are the observed numbers of deaths in the exposed populations $E_{x,t}$ of ages $x = 1, \dots, X$ during periods $t = 1, \dots, T$. We again consider one linear version of the predictor $\eta_{x,t}$:

$$\eta_{x,t} = \alpha_x + \kappa_t + \varepsilon_{x,t},$$

Model name	Predictor	Hyperparameters
P1lin	$\eta_{x,t} = \alpha_x + \kappa_t + \varepsilon_{x,t}$	Fixed
P2lin	$\eta_{x,t} = \alpha_x + \kappa_t + \varepsilon_{x,t}$	Gamma
P1	$\eta_{x,t} = \alpha_x + \beta_x \cdot \kappa_t + \varepsilon_{x,t}$	Fixed
P2	$\eta_{x,t} = \alpha_x + \beta_x \cdot \kappa_t + \varepsilon_{x,t}$	Gamma

Table 2: The models included in the analyses of Section 5.2. All models have the likelihood structure of Expression 5.2.1

and one non-linear version of the predictor:

$$\eta_{x,t} = \alpha_x + \beta_x \cdot \kappa_t + \varepsilon_{x,t}.$$

α_x , β_x and κ_t are modeled as in Section 5.1, Equations 5.1.2 and 5.1.3, and the error term $\varepsilon_{x,t}$ is modeled as an iid effect:

$$\varepsilon_{x,t} \sim \text{iid}(\tau_\varepsilon).$$

We include the error term to model overdispersion in the data. Again, we consider versions of the model where the hyperparameters τ_α , τ_β , τ_κ and τ_ε are fixed and when they are assigned flat priors. A summary of the models that are included in this section is presented in Table 2. We use the same data for the Poisson models as for the Gaussian models, described in Appendix A.3, with the exception that we fit the models to the numbers of deaths $Y_{x,t}$ and not to the log-mortality rates $m_{x,t}$.

5.2.1 P1lin: Linear Poisson Model with Fixed Precisions

The first Poisson model (P1lin) of Table 2 has a linear predictor and hyperparameters fixed to the following values:

$$\tau_\alpha = 1.96, \quad \tau_\kappa = 336, \quad \tau_\varepsilon = 420.$$

We apply this model to the Full.1 data.

As for the models with Gaussian likelihoods, the estimation results of Stan and `inlabru` are practically exactly equal for the P1 model, as can be seen from Figure B.2.1 in Appendix B.

5.2.2 *P2lin: Linear Poisson Model with Gamma Precisions*

The next model is the P2lin model. We assign this the same hyperpriors as for the Gaussian model, which are given in Expression 5.1.4. We apply this model to the Full.1 data using Stan and `inlabru`, and the results can be found in Figures B.2.2 and B.2.3 in Appendix B.

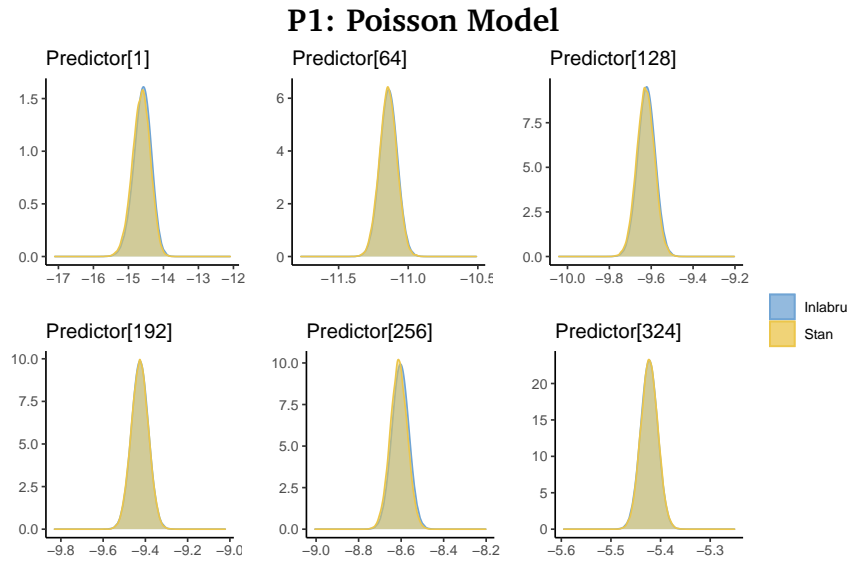
As for the other linear models and as expected, both the predictor $\eta_{x,t}$, the random effects and the precisions are estimated very similarly by `inlabru` and Stan.

5.2.3 *P1: Poisson Model with Fixed Precisions*

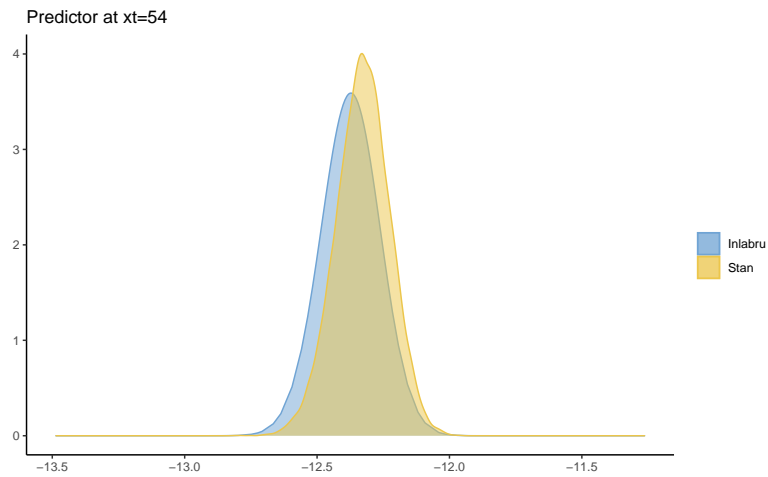
We now consider the non-linear P1 model, where the hyperparameters are fixed to the values

$$\tau_\alpha = 1.96, \quad \tau_\beta = 202, \quad \tau_\kappa = 30, \quad \tau_\varepsilon = 420.$$

As for the Gaussian G1 model, we see that while most of the values of $\eta_{x,t}$ are estimated similarly by Stan and `inlabru` (as can be seen from Figure 5.2.1a), there are some shifts between the estimations at some points. Figure 5.2.1b displays the estimation of $\eta_{x=3,t=18}$ where we see the largest difference. While this difference is still relatively small, we note that it is slightly larger than the differences observed for the G1 model, which was displayed in Section 5.1.3.



(a) Estimation results of non-linear predictor $\eta_{x,t}$



(b) Marginal distributions for $\eta_{x=3,t=18}$

Figure 5.2.1: Estimation results from applying the P1 model to the Full.1 data using Stan and inlabru

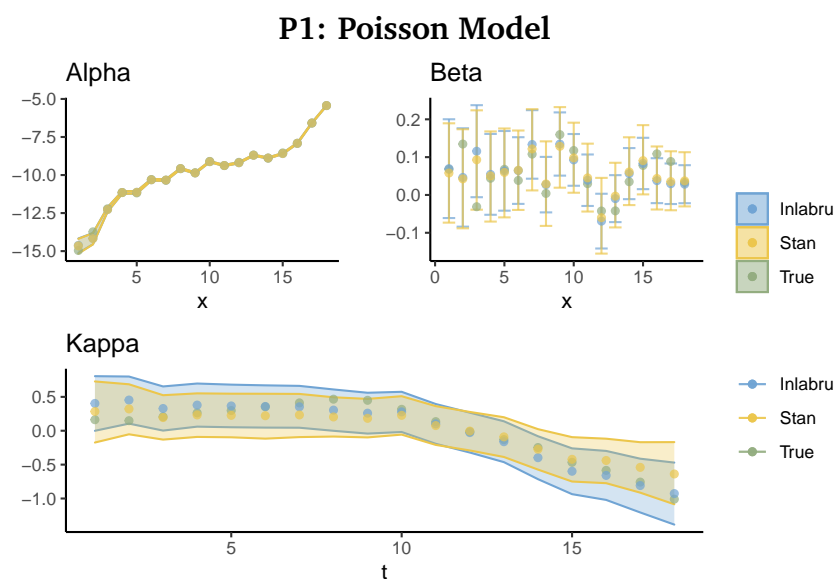


Figure 5.2.2: Estimation results of random effects α_x , β_x and κ_t from applying the P1 model to the Full.1 data using Stan and inlabru

For the random effects, the differences between the estimations by Stan and inlabru is somewhat larger for this P1 model compared to the G1 model, as seen in Figure 5.2.2. Again, we suspect that this difference originates from some unidentifiability in the data. It appears from the plot of the period effect κ_t that the estimates of inlabru and Stan are similar, except that they are tilted around the middle point (at around $t = 10$) when compared to each other. We recognize this "tilting" as a characteristic of unidentifiability, which have been described in other literature for similar mortality models (see e.g., Riebler (2010) and Hunt and Villegas (2015)). Furthermore, we note that the estimations of inlabru seem to lie closer to the true values of κ_t than the Stan estimates do.

We apply the P1 model also to the Reduced.1 data, to see how inlabru and Stan performs on a different set of data. As for the G1 model, the random effects is estimated much more similarly for the Reduced.1 data. This can be seen from Figures B.2.5 and B.2.6 in the Appendix.

5.2.4 P2: Poisson Model with Gamma Precisions

Finally, we consider the full P2 model, with the same hyperpriors as for the G2 model, given in Expression 5.1.4. We apply this model first to the Full.1 data.

Once again, we see that `inlabru` and Stan estimate the values of the predictor $\eta_{x,t}$ similarly for most points (x, t) , but gives different estimates for some values of $\eta_{x,t}$. This can be observed in Figure 5.2.3a, which displays the estimations of some arbitrary values of $\eta_{x,t}$ and from Figure 5.2.3b which displays the estimations of $\eta_{x=3,t=18}$, where the difference between `inlabru` and Stan is at its largest.

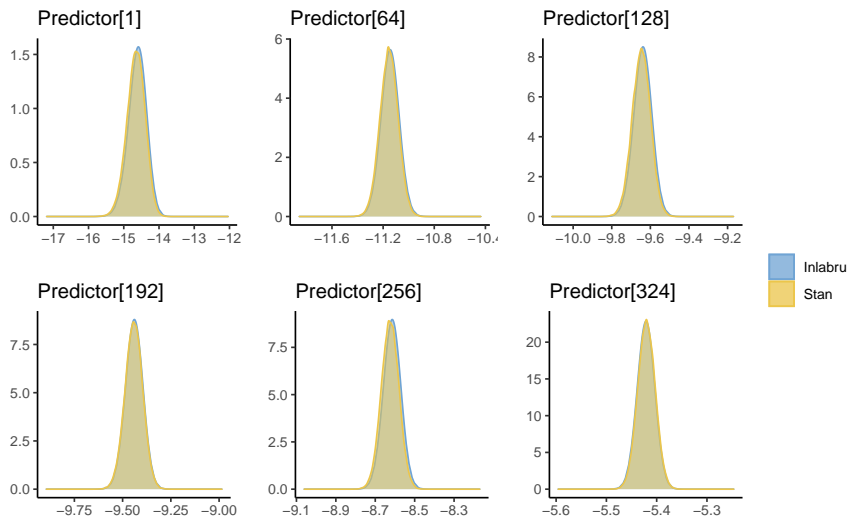
Unlike the models with Gaussian likelihoods, the largest difference between `inlabru` and Stan increased drastically for the P2 model compared to the P1 model. We see this by comparing Figure 5.2.3b for the P2 model with Figure 5.2.1b.

Figure 5.2.4 displays the random effects of the model as estimated by `inlabru` and Stan. From this figure, we see that the difference between the estimated random effects for `inlabru` and Stan have also increased for the P2 model compared to the P1 model.

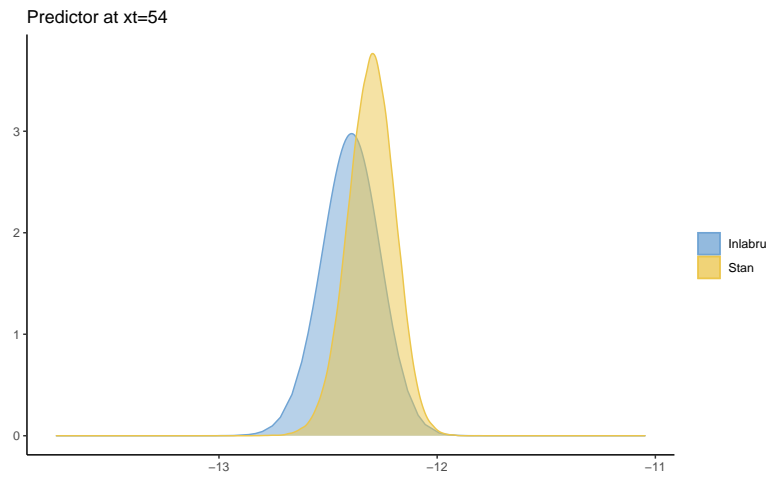
When comparing the estimates and the true values, it looks like `inlabru` produces more accurate estimations for both β_x and κ_t (Figure 5.2.4). For κ_t , in particular, it does not seem like the Stan estimation captures the curvature of the period effect, and it fails to include several of the true points in its 95% confidence bound ($t = 8, 9, 17, 18$). In comparison, the estimates of `inlabru` has wider confidence bounds with full coverage of the true values, and it displays some curvature similar to the true period effect. For the estimations of β_x , Stan gives much wider confidence bounds compared to `inlabru`.

To get a better impression of the performance of `inlabru` and Stan for the P2 model, we fit this model to three more sets of data: the Reduced.1 data, the Reduced.2 data and the Full.2 data.

P2: Poisson Model



(a) Estimation of arbitrary values of $\eta_{x,t}$



(b) Estimation results of linear predictor $\xi_{x=3,t=18}$

Figure 5.2.3: Estimation results from applying the P2 model to the Full.1 data using Stan and inlabru

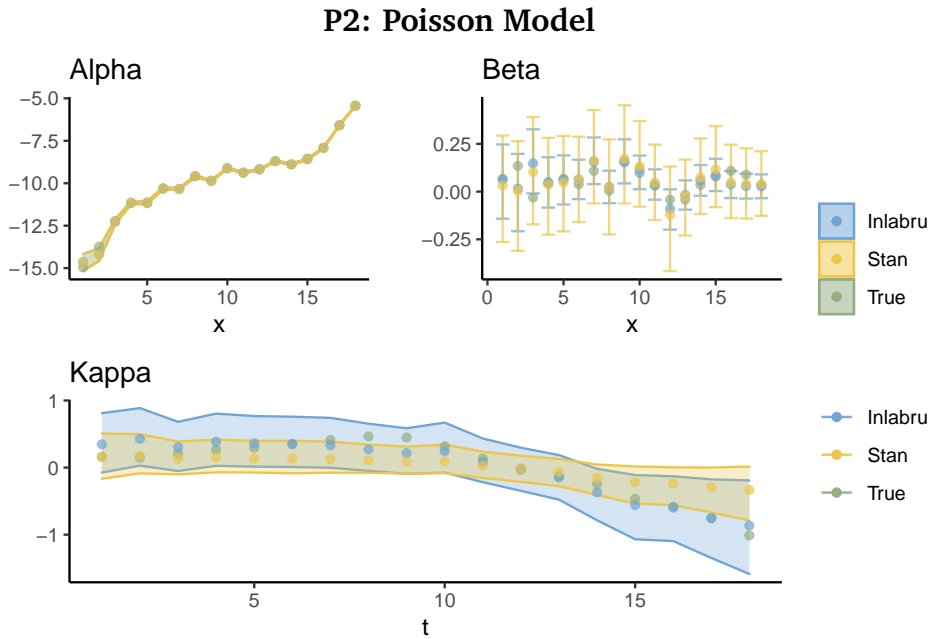


Figure 5.2.4: Estimated random effects α_x , β_x and κ_t from applying the P2 model to the Full.1 data using Stan and inlabru

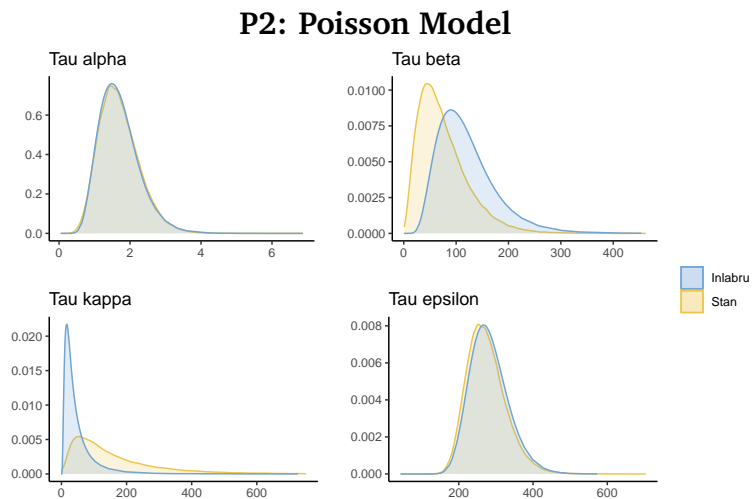


Figure 5.2.5: Estimated hyperparameters τ_α , τ_β , τ_κ , τ_ϵ from applying the P2 model to the Full.1 data using Stan and inlabru

The Reduced.1 data has generally higher values for $\eta_{x,t}$ and different dimensions from the Full.1 data. For the Reduced.1 data, `inlabru` and `Stan` give almost identical estimations results, as seen in Figure B.2.10a. The full estimation result for the Reduced.1 data are presented in Figures B.2.10, B.2.11 and B.2.12.

In the Full.2 data the values of $\eta_{x,t}$ are of roughly the same size as the Full.1 data, but the Full.2 data does not contain any points where $Y_{x,t}$ is zero. `inlabru` and `Stan` give very similar results when we apply the P2 model to the Full.2 data. The Full.2 data has the same dimensions of the Full.1 data, which indicates that the dimensions of the data are not the reason of the difference in the estimations. Plots of the estimations for the Full.2 data can be found in Figures B.2.13, B.2.14 and B.2.15.

The results from applying the P2 model to the Reduced.2 data show a situation where `inlabru` converges, while `Stan` does not. The Reduced.2 data does not contain any points where $Y_{x,t}$ is zero, and it has slightly reduced dimensions compared to the Full.1 data. Figure B.2.16 displays the trace plots from the `Stan` run with the Reduced.2 data, where we see that `Stan` does not converge for several model parameters. Figure B.2.18a displays the estimated random effects, and we see that `inlabru` gives estimations close to the true random effects. This is especially clear for the estimation of κ_t , where `inlabru` is able to correctly identify the period effect, while the diverging results of `Stan` give estimations with much too wide confidence bounds.

This is an interesting results, since it indicates that `inlabru` may be able to converge and give good results in cases where `Stan` does not converge. The full estimation results for the Reduced.2 data are included in Figures B.2.17, B.2.18 and B.2.19. Convergence issues and lack of robustness have been discussed for similar estimation procedures in other literature (by e.g., Hunt and Villegas (2015)), but this has usually been in relation to mortality models with a cohort effect. We note that the error term $\varepsilon_{x,t}$ might affect convergence in some of the same way as a cohort effect, since we let it vary with x and t .

We underline that we do not observe convergence issues for the Full.1 data similar to those of the Reduced.2 data. This can be seen from the trace plots of the Stan estimation for the Full.1 data, presented in Appendix B in Figure B.2.9. Consequently, we can not conclude that convergence issues can explain the differences we see for the P2 model applied to the Full.1 data.

5.2.5 Conclusion on Comparison of Poisson Models

The analysis of the performance of `inlabru` and Stan when applied to mortality models with Poisson likelihoods have presented several interesting results, which we summarize here.

Firstly, we have seen that for the models with linear predictors $\eta_{x,t}$, the results of Stan and `inlabru` are very similar, as expected. We see this as an indication that our model implementations are equivalent in `inlabru` and Stan.

Secondly, `inlabru` and Stan give overall similar, but not identical results also for the models with non-linear predictors. This aligns with the results from the study of mortality models with Gaussian likelihoods.

Unlike the Gaussian models discussed in Section 5.1, we have seen examples where the difference between `inlabru` and Stan is unexpectedly large for the Poisson models. We are not able to fully explain the reason for this difference. We observe that for the P2 model, Stan sometimes display difficulties with convergence that we do not see from the results of `inlabru`. Additionally, we note that the Full.1 data, for which the unexpected difference between `inlabru` and Stan occurs, several of the observed number of deaths are zero. However, we do not have sufficient evidence to conclude that any of these reasons fully explain the larges differences between Stan and `inlabru`.

When we apply the P2 model to the Full.1 data, our results indicate that `inlabru` estimates the random effects better than Stan. We emphasize that we

do not say that `inlabru` is better than Stan, as this would be a bold claim that we do not have evidence to support. Since Stan is asymptotically unbiased and `inlabru` is not, it also does not align well with theory. However, it could be the case that Stan experience convergence issues with some sets of data, that the INLA methodology more easily overcomes, and thus produces more accurate results. This hypothesis may provide an interesting starting point for further research on the performance of `inlabru` on mortality models.

Complementary plots of the results in Section 5.2, that have not been explicitly discussed here, are included in Appendix B.2.

5.3 COMPARISON OF RUN TIMES

A large part of the motivation behind using `inlabru` to do inference with mortality models is the computational power of the INLA methodology (Rue et al. 2009). Since `inlabru` does multiple runs of INLA as part of the linearization, we expect it to be somewhat slower compared to INLA. When comparing the runtimes for the eight models that we considered in this chapter, we see that in general, `inlabru` is faster than Stan. The computational advantage of `inlabru` seem to increase with the complexity of the model. `inlabru` provides the largest computational benefit for the P2 model, where we need many iterations in Stan to get a good result but `inlabru` still converge fairly quickly.

The exact runtimes, as well as the number of HMC iterations we used for Stan, for each model, are included in Appendix B.3 in Table 8.

APPLICATION TO GERMAN CANCER DATA

In this final part of our research, we use `inlabru` to perform Bayesian inference for mortality of German lung and stomach cancer in the years 1999-2016, described in Chapter 3. We define one mortality model and apply this to the full sets of data (Section 6.2) and to produce forecasts of future mortality (Section 6.3). We assess how our model compares to alternative, similar mortality models in Sections 6.2.2 and 6.4.

Throughout this chapter, we consider four subsets of the cancer data:

- Female lung cancer data
- Male lung cancer data
- Female stomach cancer data
- Male stomach cancer data

We treat these four data sets separately, as we expect this to give the best results. In our previous work (Behrens 2021), we investigated multivariate models for considering male and female mortality jointly, and we found that separate models performed better.

The full code for running the analyses in this chapter can be found at the GitHub repository <https://github.com/Helenerb/Masters-thesis>.

6.1 THE POISSON LEE-CARTER COHORT MODEL

To analyze German cancer mortality data we employ the model

$$\begin{aligned} \text{PLCC: } Y_{x,t} &\sim \mathcal{Poisson}(E_{x,t} \cdot e^{\eta_{x,t}}) \\ \eta_{x,t} &= \alpha_x + \beta_x \cdot \kappa_t + \gamma_c + \varepsilon_{x,t}, \end{aligned} \quad (6.1.1)$$

for age groups $x = 0, \dots, 17$, years $t = 0, \dots, 17$ and cohorts $c = 0, \dots, 102$. α_x is the main age effect, κ_t is the period effect, β_x is a modulating age effect and γ_c is the cohort effect. $\varepsilon_{x,t}$ is an error term, which included to account for overdispersion. We refer to the model as the Poisson Lee-Carter Cohort (PLCC) model.

To make the model identifiable, we apply the following constraints:

$$\sum_x \beta_x = 1, \quad \sum_t \kappa_t = 0, \quad \sum_c \gamma_c = 0. \quad (6.1.2)$$

According to Hunt and Blake (2020b), models similar to the PLCC model, but without the additional error term, should be fully identified, when the constraints in Expression 6.1.2 are applied. However, we have not seen equivalent results for models where extra error terms for overdispersion are added, and we keep this in mind when investigating the identifiability of our results.

Since we consider the model in a Bayesian setting, we assign prior distributions to the random effects and the error term as:

$$\begin{aligned} \alpha_x &\sim \text{RW1}(\tau_\alpha) \\ \beta_x &\sim \text{iid}(\tau_\beta) \\ \kappa_t &\sim \text{RW2}(\tau_\kappa) \\ \gamma_c &\sim \text{RW1}(\tau_\gamma) \\ \varepsilon_{x,t} &\sim \text{iid}(\tau_\varepsilon), \end{aligned} \quad (6.1.3)$$

where RW1, RW2, and iid denote the first and second-order random walks and the independent random noise models of Rue and Held (2005). We use RW1 models for the age effect α_x and the cohort effect γ_c since we expect some

smoothness in these effects. Previous literature (e.g., Czado et al. (2005)) has shown that β_x often takes a more erratic shape, and thus we assign the iid prior for this effect. We assign an RW2 model to the period effect κ_t , since RW2 models penalize deviation from a linear trend (Rue and Held 2005), and this linear trend can be used for projection into future periods. This kind of projection can be done with many different models, but since the second-order random walk is readily implemented in the `inlabru` framework, this is a natural choice.

To complete the specification of our Bayesian model, we need to specify prior distributions for the hyperparameters of the model as well. We apply penalized complexity priors (Simpson et al. 2017) for the hyperparameters τ_α , τ_κ , and τ_γ of the random walks, with parameters such that

$$\mathbb{P}(1/\sqrt{\tau} > 1) = 0.01, \quad (6.1.4)$$

as suggested by Rue (2022). For the priors of the iid effects, τ_β and τ_ε , we assign gamma priors:

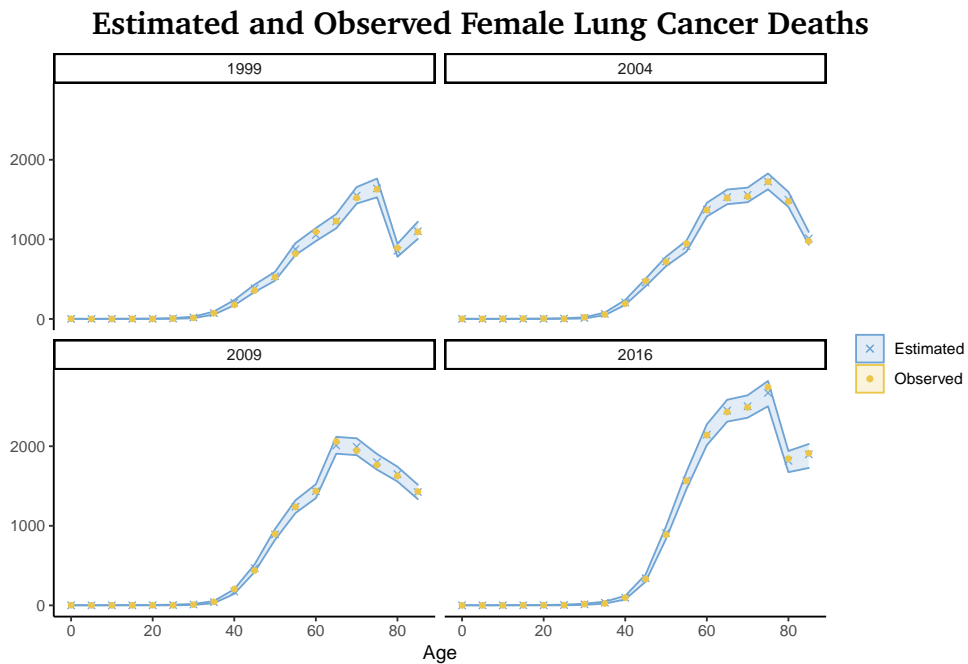
$$\tau \sim \mathcal{Gamma}(1, 0.00005).$$

6.2 APPLICATION TO FULL GERMAN CANCER DATA

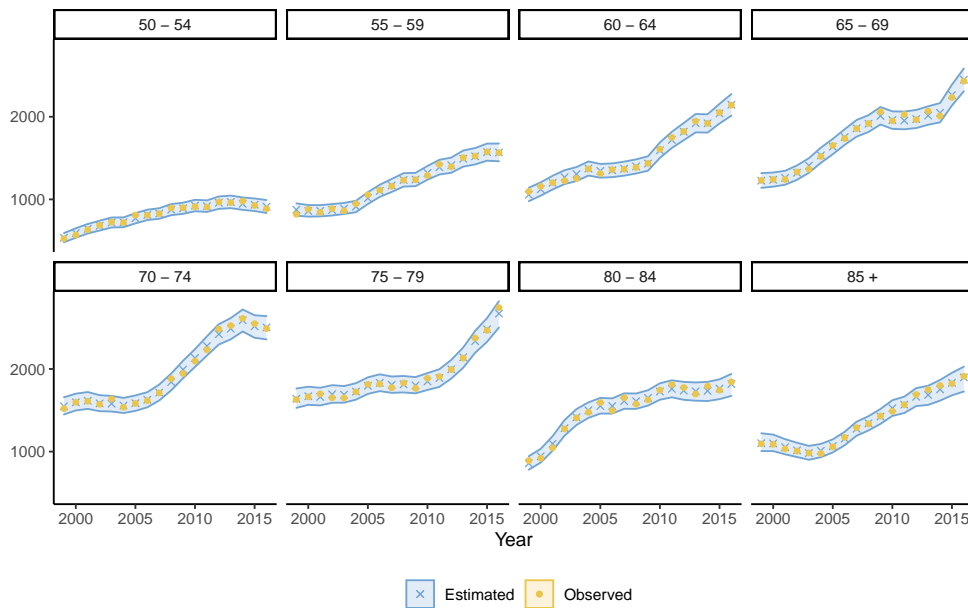
We fit the PLCC model presented in Section 6.1 to the female lung cancer data.

Figure 6.2.1 displays the estimated female lung cancer deaths, and we see that the estimated deaths follow the observed cancer deaths quite closely. For younger ages, both the observed and estimated numbers of deaths are very close to zero, and for this reason, we only display the results for ages where the cancer mortality rate is sufficiently non-zero in Figure 6.2.1b.

6.2.2 displays the estimated age, period, and cohort effects for the female lung cancer data. We observe that the estimated values for β_x have quite wide 95% confidence bounds for lower values of x . Since we are mainly



(a) Observed and estimated number of female lung cancer deaths, for years 1999, 2004, 2009 and 2016



(b) Observed and estimated number of female lung cancer deaths, for age groups 50-54 and older

Figure 6.2.1: Estimated (blue dots) and observed (yellow X-es) female lung cancer deaths, resulting from applying the PLCC to the full German female lung cancer data using inlabru

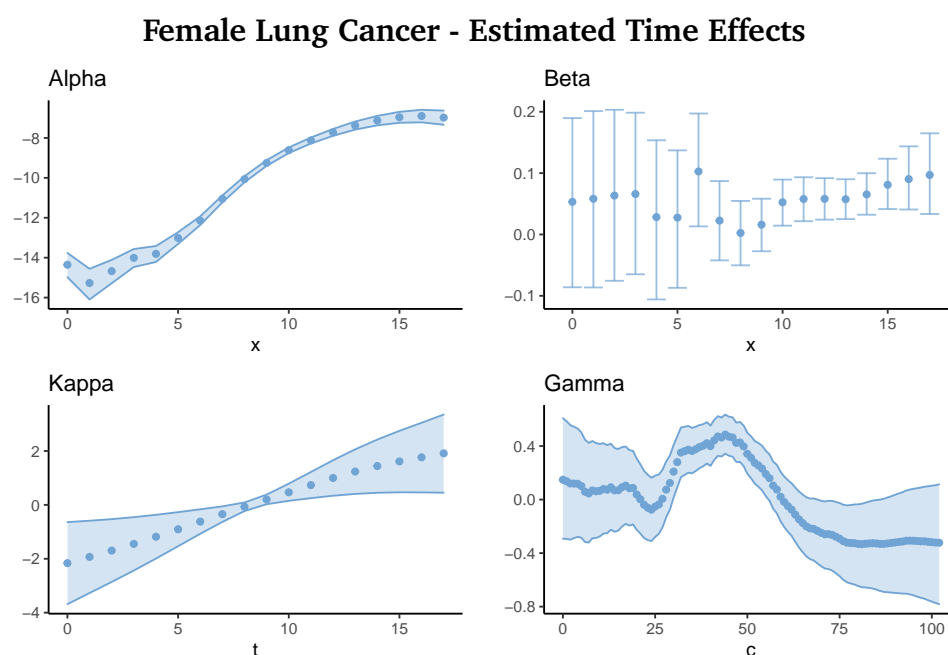


Figure 6.2.2: Estimated time effects α_x (upper left), β_x (upper right), κ_t (lower left) and γ_c (lower right), resulting from applying the PLCC model to the full German female lung cancer data using `inlabru`

interested in the estimates for older ages, we do not see this as a great concern. Furthermore, there seems to be some unidentifiability between the period effect κ_t and the cohort effect γ_c . Specifically, it seems like these effects may be tilted around their zero point. This kind of "tilting" unidentifiability has been described by Hunt and Villegas (2015) for a model similar to the PLCC model. They suggest adding an approximate identifiability constraint to the cohort effect, to remove linear drift. We hesitate from doing this, as it would slightly alter the model fit. Additionally, it would impose a different medical or demographic significance for the time effects, which we do not have the medical knowledge to support.

6.2.3 presents the estimated mean values of the error term $\varepsilon_{x,t}$, and we see that the magnitudes of the error terms $\varepsilon_{x,t}$ are relatively small compared to the magnitudes of the time effects. We interpret this as a sign that the main part of the variability in the data that may be attributed to some of the time effects, has correctly been attributed to these.

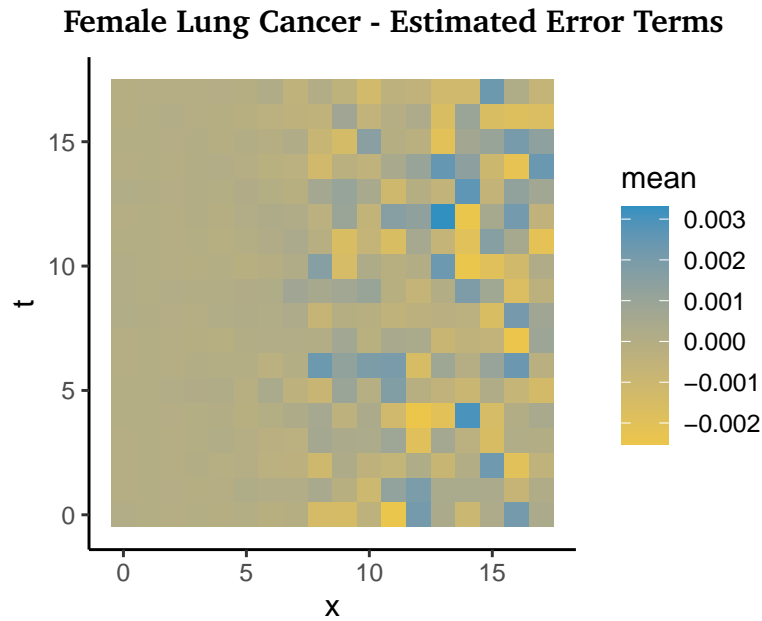


Figure 6.2.3: Estimated mean values for the error term $\varepsilon_{x,t}$, resulting from applying the PLCC model to the full German female lung cancer data using inlabru

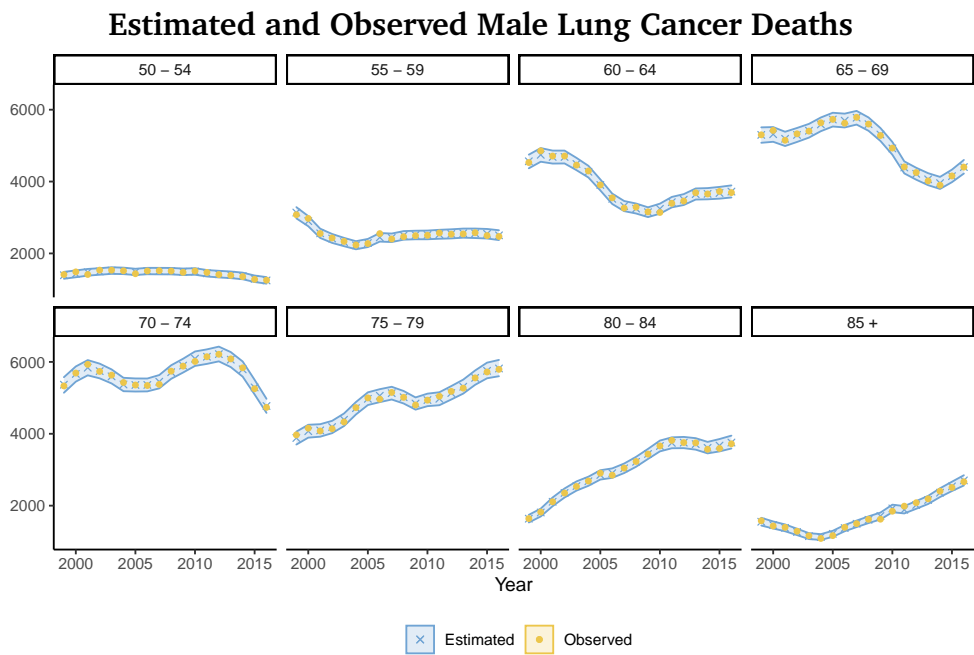


Figure 6.2.4: Observed and estimated number of male lung cancer deaths, for age groups 50-54 and older, resulting from applying the PLCC model to the full German male lung cancer using inlabru

Male Lung Cancer - Estimated Time Effects

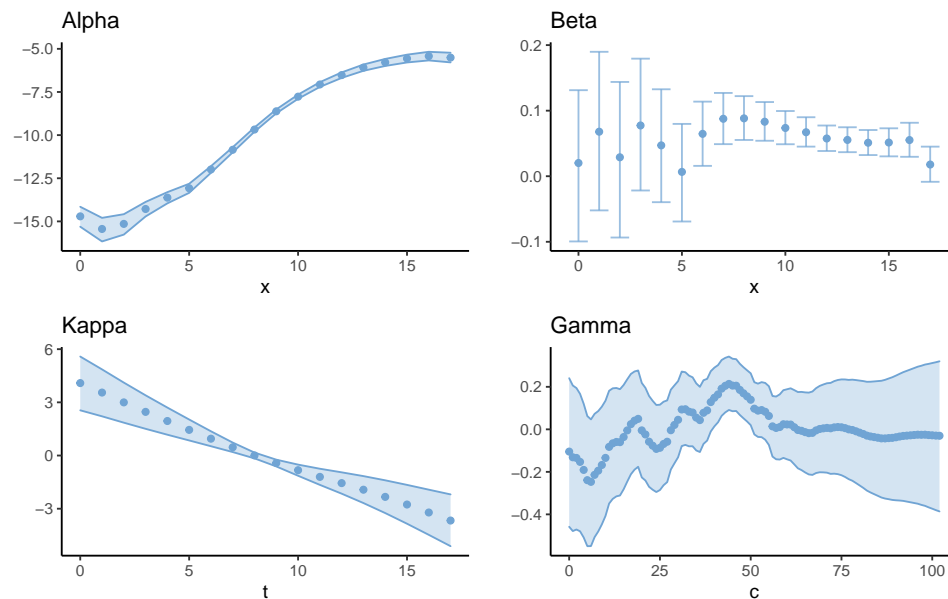


Figure 6.2.5: Estimated time effects α_x , β_x , κ_t and γ_c , resulting from applying the PLCC model to the German male cancer data using `inlabru`

We now fit the PLCC model to the male lung cancer data. Figure 6.2.4 displays the estimated number of deaths. Also for this data, the values of $Y_{x,t}$ as estimated by `inlabru` seem to align well with the observed values.

Figure 6.2.5 displays the estimated random effects for the male lung cancer data, and we see that many of the traits discussed for the female lung cancer data apply here as well. Additionally, we observe what might be some periodicity in the cohort effect. Other literature, e.g., Riebler (2010), have described similar periodicity in cohort effects in APC models, warning that it may be the result of unidentifiability arising from different lengths of the age and period intervals of the data. Since our data indeed has age and period intervals of different lengths, we treat this periodicity with skepticism and not as true cohort effects in the data.

Riebler (2010) suggests applying smoothing priors, such as second-order random walks, to the cohort effect to avoid periodicity. However, we note that assuming an RW2 on the cohort effect might lead to estimations with more lin-

ear drift, as the RW2 model penalizes deviation from a linear trend as opposed to the RW1 model that penalizes deviation from a constant (Knorr-Held and Rainer 2001). Linear drift in the cohort effect is described as problematic in some literature (e.g., Hunt and Blake (2021a)), which argues that the cohort effect should be driftless to avoid period effects being misclassified as cohort effects.

We apply the PLCC model to the female and male stomach cancer as well, and the resulting estimates display similar properties to what we have seen for the lung cancer data. Plots of these results are included in Appendix C.1 in Figures C.1.3, C.1.4 and C.1.5 for the female cancer data, and in Figures C.1.6, C.1.7 and C.1.8 for the male cancer data. We observe a similar periodicity in the cohort effect of the male stomach cancer data, as can be seen in Figure C.1.7a, but not for the female stomach cancer data.

Remaining plots of the full results, that have not explicitly been discussed here, are also included in Appendix C.1.

6.2.1 Quantifying Model Fit

To quantify the model fit, we calculate the percentage of the observed values for $Y_{x,t}$ that falls within the 95% confidence bounds of the `inlabru` estimations. This value is calculated by

$$\text{Coverage 95\% interval} = \frac{1}{(X+1) \times (T+1)} \sum_{x,t} \mathbb{1}(l_{x,t} \leq Y_{x,t} < u_{x,t}), \quad (6.2.1)$$

where $Y_{x,t}$ is the observed cancer deaths, $\mathbb{1}(\cdot)$ is the indicator function and $l_{x,t}$ and $u_{x,t}$ are the lower and upper 95% quantiles for the predicted cancer deaths for age group x and period t . We calculate the coverage for both the entire set of observations and for the set of observations where $x > 5$, omitting the values where the mortality rate is very close to zero. The results are presented in Table 3, together with the number of iterations before the `inlabru` linearization reaches convergence.

Data	Iterations	Coverage	Coverage, $x > 5$
Female, lung	34	0.97	0.98
Male, lung	12	0.98	0.98
Female, stomach	>100	0.98	0.98
Male, stomach	10	0.98	0.99

Table 3: The number of `inlabru` iterations before convergence and the coverage percentage of the 95% confidence bounds, when the PLCC model is applied to the different sets of data.

From Table 3, we see that the speed of convergence varies between the different sets of data. While the linearization procedure of `inlabru` reaches convergence after 12 and 10 iterations for male lung and stomach cancer, respectively, `inlabru` needs 34 iterations to converge for the female lung cancer data set. For the female stomach cancer data, `inlabru` does not reach convergence at all after running for a 100 steps. For all sets of data, the coverage percentage lies around 98%, which is slightly higher than the ideal 95%, but we still consider this quite good.

6.2.2 Removing the Error Term

As previously mentioned, the magnitude of the estimated error term $\varepsilon_{x,t}$ is in general quite small compared to the values of the time effects. This is especially the case for the male lung cancer data, as can be seen in Figure 6.2.6, which displays the estimated mean values of the error terms for the male lung cancer data. We wish to investigate whether the error term may be omitted.

Consequently, we slightly modify the model for male lung cancer data by removing the error term $\varepsilon_{x,t}$ as

$$\widetilde{\text{PLCC}}: Y_{x,t} \sim \mathcal{Poisson}(E_{x,t}e^{\eta_{x,t}}), \quad \eta_{x,t} = \alpha_x + \beta_x \kappa_t + \gamma_c, \quad (6.2.2)$$

with constraints and prior distributions as given by Equations 6.1.2 and 6.1.3.

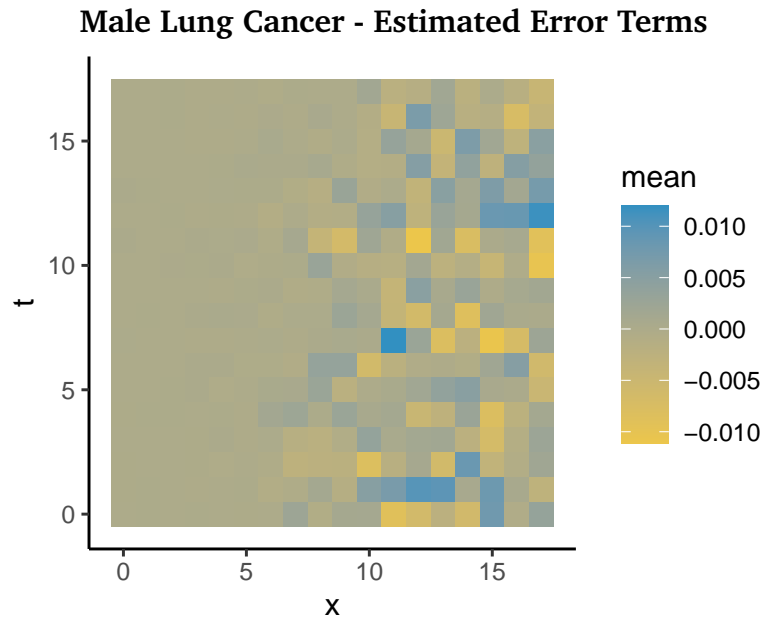
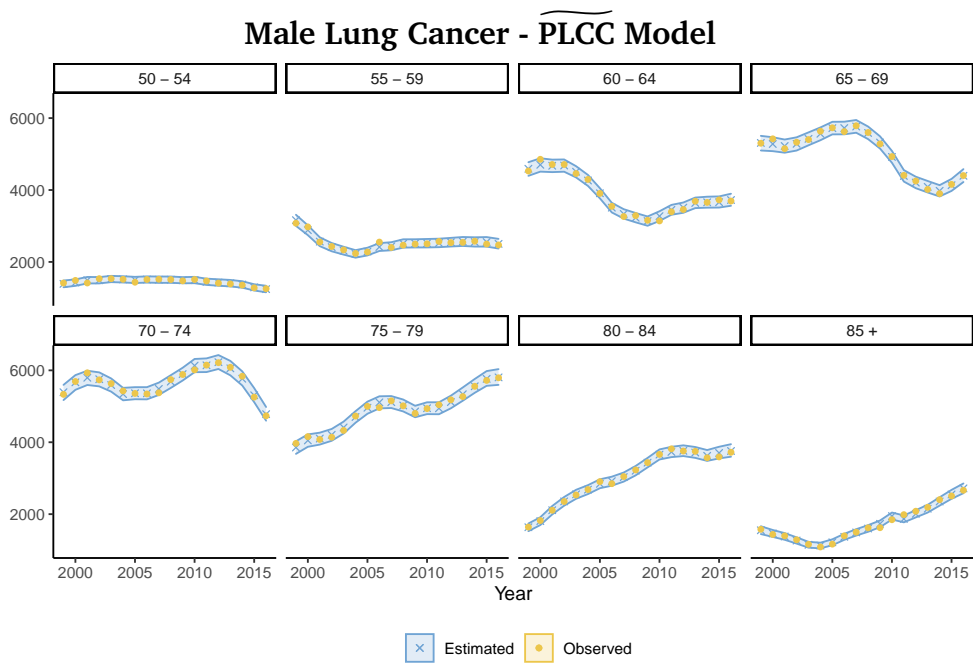


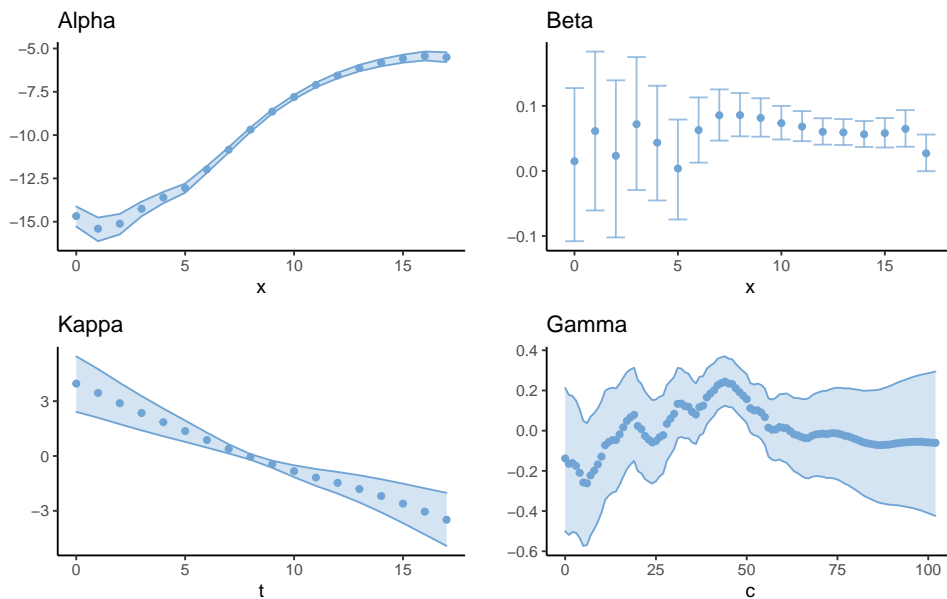
Figure 6.2.6: Estimated mean values for the error term $\varepsilon_{x,t}$.

Data	Iterations	Coverage	Coverage, $x > 5$
Male, lung, $\widehat{\text{PLCC}}$	10	0.96	0.95

Table 4: Score statistics for the $\widehat{\text{PLCC}}$ model, calculated as described for Table 3



(a) Observed and estimated number of male lung cancer deaths, for age groups 50-54 and older, resulting from applying the $\widetilde{\text{PLCC}}$ model to the full German male lung cancer using `inlabru`



(b) Estimated time effects α_x (upper left), β_x (upper right), κ_t (lower left) and γ_c (lower right).

Figure 6.2.7: Estimated number of deaths (upper) and time effects (lower), resulting from applying the $\widetilde{\text{PLCC}}$ model to the male lung cancer data

The resulting estimated deaths and time effects are displayed in Figure 6.2.7. We observe that these results look very similar to the results from the full PLCC model, displayed in Figures 6.2.4 and 6.2.5.

The number of `inlabru` iterations before convergence and the coverage percentage are displayed in Table 4, and they are relatively similar to the scores of Table 3 for the full model. The $\widetilde{\text{PLCC}}$ model converges faster and it has a coverage closer to 95%. The lower coverage of the $\widetilde{\text{PLCC}}$ model may indicate that the confidence bounds are more correct for this model. However, the differences are small, so we do not conclude that the $\widetilde{\text{PLCC}}$ model is much better than the PLCC model. In our further investigation, we keep the original PLCC model with the error term accounting for overdispersion.

6.3 FORECASTING GERMAN CANCER DATA

In this stage of our research, we investigate how the PLCC model performs when it is used for forecasting mortality. We hold back the data for years 2011-2016, and fit the model to the data for the years 1999-2011 while simultaneously generating forecasts for the period 2011-2016. We assume the population at risk, $E_{x,t}$ to be known for the full period.

In the following discussion, we find it necessary to treat the results for the estimations for the observed years 1999-2010 differently from the estimations for the forecast years 2011-2016. For simplicity, we refer to the data for the forecast years as the out-data and to the data for the observed years as the in-data.

We begin by finding forecasts for the female lung cancer data, and Figure 6.3.1 displays the estimations and forecasts of the cancer deaths $Y_{x,t}$. We observe that the forecasts match the observations quite well, with the majority of the observed deaths seemingly lying within the forecast 95% intervals.

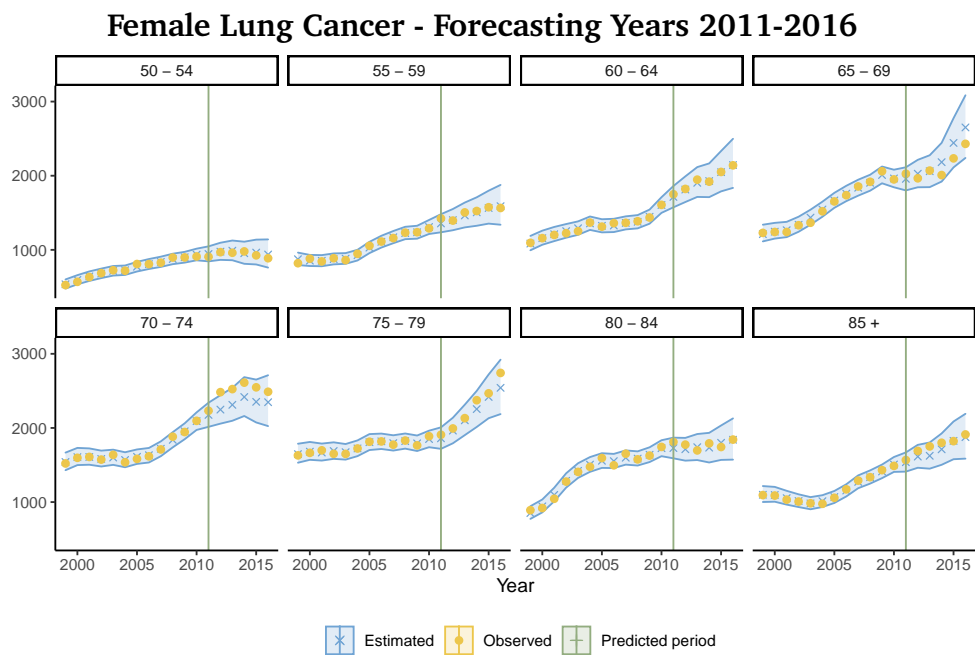


Figure 6.3.1: The estimated (blue dots) and the observed (yellow X-es) cancer deaths, for the ages 50-85+ and the period 2011-2016 for female lung cancer, when years 2011-2016 are forecast by the PLCC model. The green line marks the beginning of the forecast period

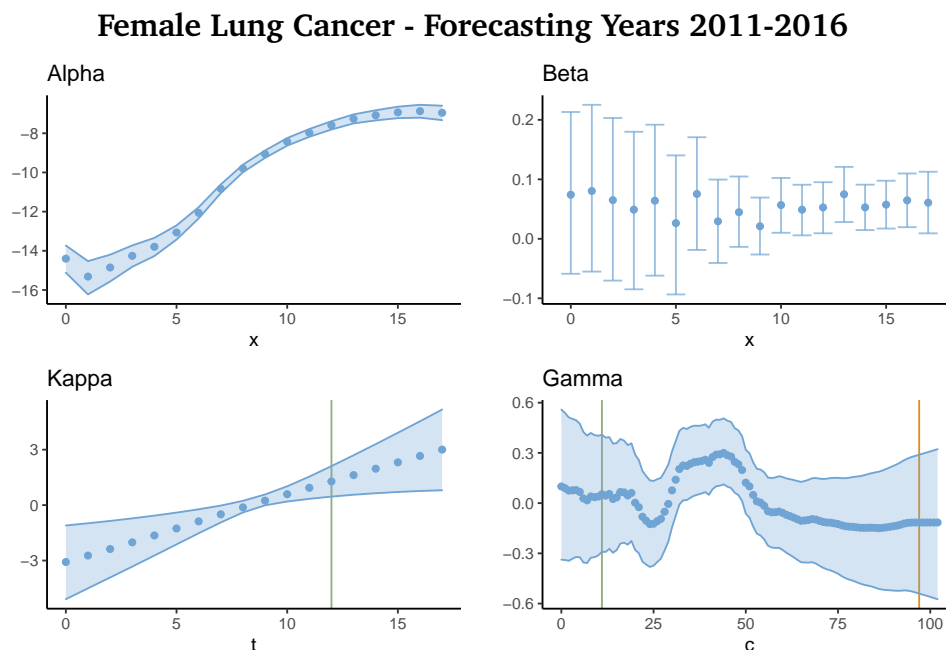


Figure 6.3.2: Estimated time effects α_x , β_x , κ_t and γ_c for female lung cancer, when years 2011-2016 are forecast by the PLCC model. The green line in lower left plot marks the beginning of the predicted period. The green and orange lines in the lower right plot marks the beginning of the partially and the fully unobserved cohorts

Although the PLCC model seem to produce good forecasts for most of the data, there are some observations that lie outside the forecast confidence bounds that we look further into. We observe that the age-group 70-74 is somewhat underestimated in the forecasts produced by `inlabru`, as seen in Figure 6.3.1. However, it is only the observation for the year 2012 that clearly lies outside of the forecast confidence interval, and we do not interpret this as a sign that the PLCC model is unsuited for forecasts with this type of data.

Figure 6.3.2 present the random effects as estimated for the forecasting model by `inlabru`. The time effects look quite similar to those resulting from fitting the model to the full data (Figure 6.2.2). Again, we observe unidentifiability between the period effect κ_t and the cohort effect γ_c .

We use the PLCC model to forecast the female stomach cancer, and the results are included in Appendix C.2 in Figure C.2.1. These results are similar

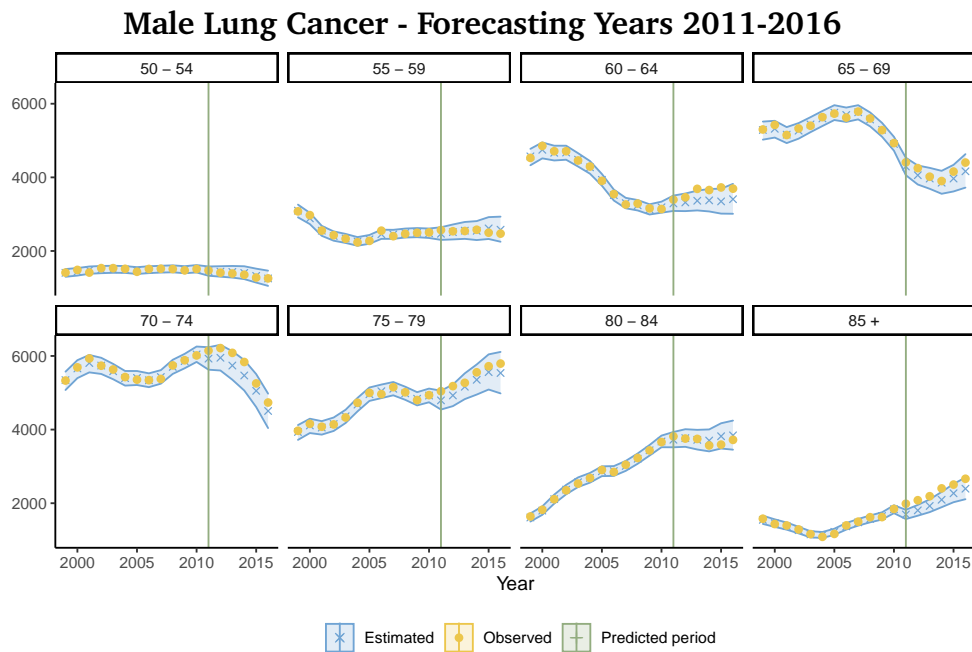


Figure 6.3.3: The estimated (blue dots) and the observed (yellow X-es) cancer deaths, for the ages 50-85+ and the period 2011-2016. for male lung cancer, when years 2011-2016 are forecast by the PLCC model. The green line marks the beginning of the forecast period.

to those of the female lung cancer, with similar tendencies in the estimated random effects and the number of deaths.

Next, we use the PLCC model to produce forecasts for the male lung cancer data. The resulting estimated and forecast number of deaths are displayed in Figure 6.3.3. We observe that the forecasts seem to underestimate the mortality for the oldest age group. We find this underestimation slightly worrisome, since the observed deaths for this age group consistently lie above the forecast 95% interval.

The random effects for the male lung cancer data, as estimated and forecast by `inlabru` are presented in Figure 6.3.4. For the PLCC model applied to the full male lung cancer data, the estimated modulating age effects β_x display a clear drop for the oldest age group compared to the younger ages (see Figure 6.2.5). We do not see a similar drop in the estimated β_x for the forecast model. This may be related to the poor forecasts for the male data for the 85+ age

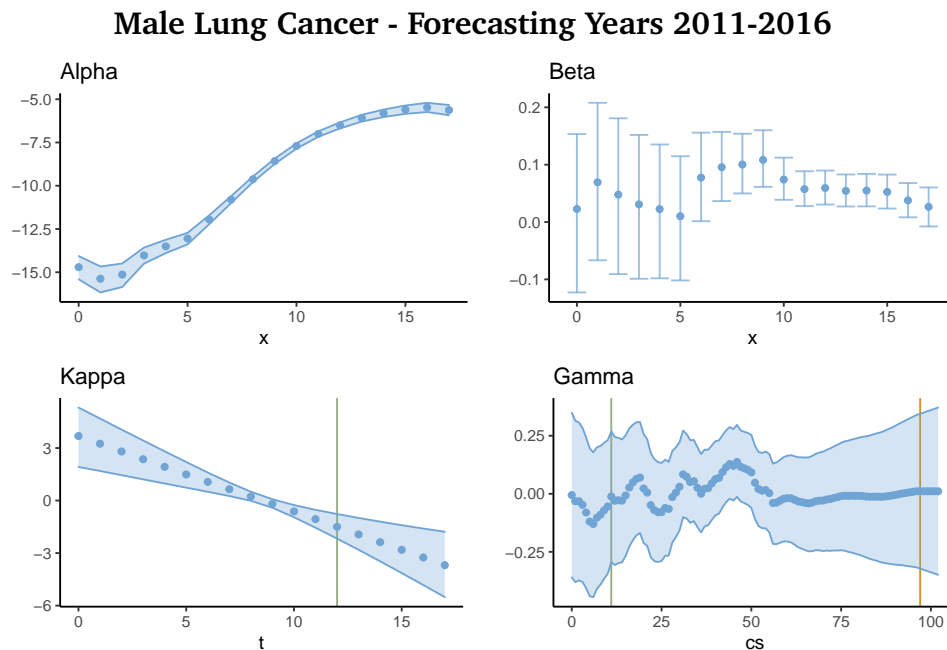


Figure 6.3.4: Estimated time effects α_x , β_x , κ_t and γ_c for male lung cancer, when years 2011-2016 are forecast by the PLCC model. The plot layout is as described for Figure 6.3.2

group. Since the period effect for the male cancer data is declining, a decrease in the value of the modulating β_x corresponds to an increase in the overall mortality, which supports this hypothesis. Additionally, we observe the same periodicity in the cohort effect as for the full data (Figure 6.2.5).

Finally, we use the PLCC model to estimate and forecast the male stomach cancer data. Plots of these results are included in Appendix C.2 in Figure C.2.2. The forecast results of the male stomach cancer are quite similar to those of male lung cancer, again displaying an underestimation in the forecast for the oldest age groups and similar patterns in the random effects.

We calculate the coverage percentage of the confidence bounds for the in- and out data observations combined, and we find that the coverage lies around 96% for all data sets. This could be an indication that the confidence bounds are slightly wider than they should be. However, since the calculation

Data	Iterations	Coverage
Female, lung	10	0.89
Male, lung	9	0.90
Female, stomach	81	0.88
Male, stomach	8	0.89

Table 5: The number of `inlabru` iterations before convergence and the coverage percentage of the out data, when the PLCC model is used for forecasting the years 2011-2016. The coverage are calculated for predictions where $x > 5$.

of the coverage is based on relatively few observations, we do not conclude that this is certainly the case.

Table 5 presents the number of linearization steps `inlabru` makes before convergence and the coverage percentage for the out-data. The coverage percentages are lower for the forecast data, at around 89% for all data sets, which indicate that the confidence bounds are slightly too narrow. We note that `inlabru` in general seem to converge more quickly when forecasting compared to the estimations given the full data. For example, `inlabru` now reaches convergence in under 100 iterations for the female stomach cancer model.

6.4 COMPARISON OF MODEL CHOICES FOR PERIOD EFFECTS

So far, we have used a second-order random walk model as the prior distribution for the period effect κ_t . This has also been done in previous literature, see e.g., Riebler (2010). However, the seemingly dominating model choice for the period effect in Lee-Carter types of models is a first-order random walk model with drift (see e.g., Lee and Carter (1992), Czado et al. (2005)). We wish to investigate how our choice of using a second-order random walk affects the quality of the forecasts, compared to models that use a random walk with drift.

A random walk model with drift is defined as

$$\kappa_t = \kappa_{t-1} + \phi + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \tau_\kappa),$$

where ϕ is a drift term. This model is not directly implemented as a random effect in `inlabru`. However, we can rewrite it as the sum of a linear term and a regular random walk of order one (RW1), as

$$\kappa_t = \kappa_t^* + \phi \cdot t, \quad \kappa_t^* = \kappa_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \tau_\kappa). \quad (6.4.1)$$

Separately, these two terms are implemented in `inlabru`. When we insert Expression 6.4.1 into the PLCC model, we get the new model

$$\text{PLCC}^* : \quad \eta_{x,t} = \alpha_x + \beta_x(\phi \cdot t + \kappa_t^*) + \gamma_c + \varepsilon_{x,t}.$$

We model α_x , β_x , γ_c and $\varepsilon_{x,t}$ in the same way as for the regular PLCC model (see Equations 6.1.2 and 6.1.3). The linear term $\phi \cdot t$ and the first-order random walk κ_t^* are modeled as described by Rue and Held (2005). We assign ϕ the prior

$$\phi \sim \mathcal{N}(0, 1),$$

and we use a PC-prior as described in Equation 6.1.4 for the precision of κ_t^* .

To ensure identifiability, we impose a sum-to-zero constraint on κ_t^* . The linear term is in itself constrained to $\phi \cdot t_0 = 0$ for period indices $t = 0, \dots, T$. This should not change the model fit, as we are free to choose the identifiability constraints as we wish, but it does change the demographic significance of the time effects compared to the regular PLCC model.

We fit the PLCC* model to the female lung cancer data, using `inlabru`, and we forecast the years 2011-2016 as described in Section 6.3. The resulting forecast cancer deaths are displayed in Figure 6.4.1. We compare these results to those of Figure 6.3.1 which displays the corresponding results for the regular PLCC model. The results from the PLCC and the PLCC* models overall look quite similar. The confidence bounds for the estimated and forecast deaths seem slightly wider for the PLCC* compared to the PLCC model, as we can

Female Lung Cancer - Forecasting Years 2011-2016 - PLCC* Model

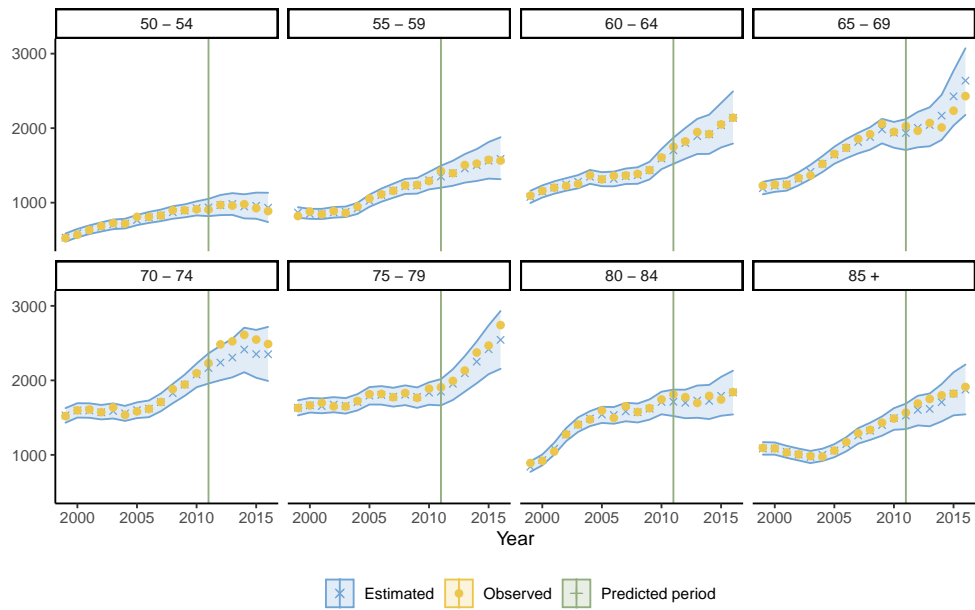


Figure 6.4.1: The estimated (blue dots) and the observed (yellow X-es) cancer deaths, for the ages 50-85+ and the period 2011-2016 for female lung cancer, when years 2011-2016 are forecast using the PLCC* model. The green line marks the beginning of the forecast period.

Female Lung Cancer - Forecasting Years 2011-2016 - PLCC* Model

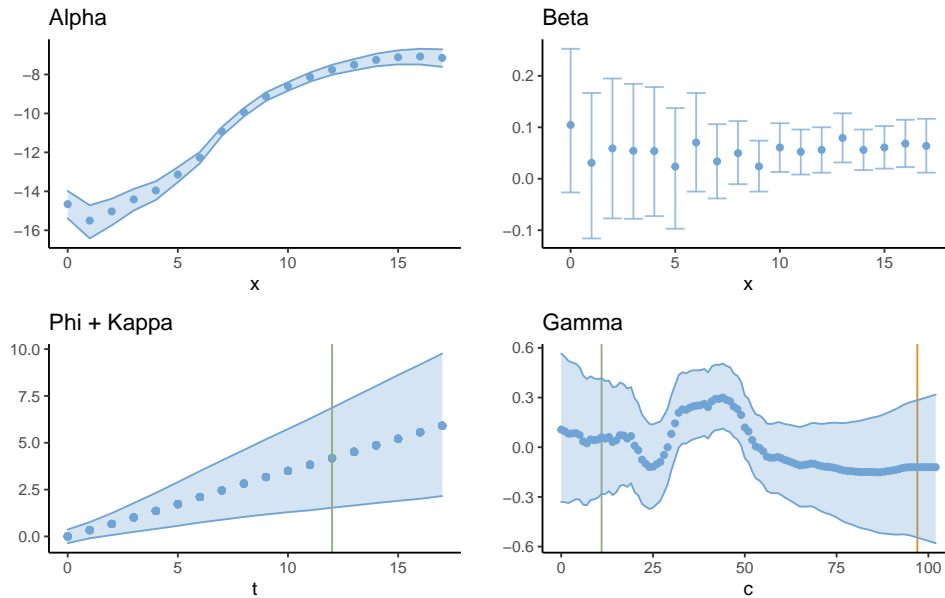


Figure 6.4.2: Estimated time effects α_x , β_x , $\kappa_t^* + \phi \cdot t$ and γ_c for female lung cancer, when years 2011-2016 are forecast using the PLCC* model. The plot layout is as described for Figure 6.3.2

see that the coverage of the PLCC* is higher than the PLCC model. Still, we emphasize that the difference is marginal.

Figure 6.4.2 displays the random effects for the PLCC* model, and we compare these as well to the random effects of the PLCC model (Figure 6.3.2). The estimated period effects are naturally differently defined, but the estimated age effect α_x and cohort effect γ_c are notably similar for the two models.

To further compare the performance of the two models, we use a statistic that consider both the accuracy and the sharpness of the predictions. Predictions with a higher accuracy are preferred to predictions with lower accuracy, and sharp predictions, with a narrow prediction interval, are preferred as long as the predictions are also fairly accurate. A score statistic that take all of this into consideration is the Dawid-Sebastiani Score (DSS), defined by

$$DSS = \left(\frac{Y_{x,t} - \mu_{x,t}}{\sigma_{x,t}} \right)^2 + 2 \cdot \log(\sigma_{x,t}). \quad (6.4.2)$$

Data	MDSS (in)	MDSS(out)	Iterations	Coverage
PLCC: female, lung	7.11	9.23	10	0.89
PLCC*: female, lung	7.33	9.57	10	0.92

Table 6: The MDSS for the in and out data, the number of iterations before `inlabru` converge and the coverage percentage for the out data, when the PLCC* and the PLCC models are applied to the female lung cancer data.

Here $Y_{x,t}$ is the observed number of deaths for age group x in year t , $\mu_{x,t}$ is the mean of the corresponding predicted number of deaths, and $\sigma_{x,t}$ is the standard deviation of the prediction (Gneiting and Raftery 2007). A lower DSS indicates a better prediction (see Keilman (2020) for details). Following e.g., Riebler et al. (2012a), we use the DSS to quantify goodness of fit for the models. Since the DSS is calculated for each predicted number of deaths at age x and year t , we take the average over all ages and periods, and use this measure, denoted MDSS, in our comparison.

The MDSS for the in and out data, the coverage percentage and the number of `inlabru` iterations before convergence for the PLCC* and the PLCC models are displayed in Table 6. These results confirm that the PLCC* model has a marginally larger coverage than the PLCC. When we compare the MDSS for the two models, we see that the PLCC model gives a somewhat lower MDSS score than the PLCC* model, which indicates a slightly better model fit. However, as for the difference in coverage, this difference is hardly large enough to draw any conclusions.

In any case, we do not see any signs that the PLCC model performs significantly worse than the more commonly used PLCC* model. Since the PLCC model is more easily implemented in `inlabru`, we argue that it is the preferred model in our case.

CLOSING REMARKS

We have shown that Bayesian inference using `inlabru` on mortality models with multiplicative terms in the predictor gives results close, but not identical, to those produced by MCMC methods. This was done by applying `inlabru` and Stan (a software that produces HMC estimates) to a set of mortality models for some synthetic data, and comparing the results. We have seen that the difference between Stan and `inlabru` depends on the model choice and of the data the model is applied to. For some sets of data, the results of `inlabru` and Stan were hardly distinguishable. For other sets of data, the results of `inlabru` and Stan were notably different for some of the models at some points in the data.

We are cautious with concluding on exactly why the results of `inlabru` and Stan are sometimes different. Instead, we make some remarks based on the results of our investigations:

- For some combinations of mortality models and data sets, `inlabru` reached convergence when Stan did not. For these cases, the results of `inlabru` seemed to be relatively correct when compared to the true model. This may indicate that `inlabru` can serve as a method to avoid convergence issues that are present for MCMC methods.

- The difference between `inlabru` and Stan was particularly large for the mortality models with a Poisson likelihood. The estimates of `inlabru` seemed to be closer to the true values. Furthermore, we observed that the data set where this difference was especially apparent contained multiple occurrences of zero observed deaths.
- We have observed signs of unidentifiability between the random effects for some of the sets of data. The results for these sets of data give rise to a suspicion that Stan and `inlabru` give different estimates for the random effects in the cases where they are unidentifiable.
- From the models where the results of Stan and `inlabru` were very different, `inlabru` seemed, to our judgment, to give the best estimates of the random effects. This may be the result of a lack of convergence in Stan, although the trace plots of Stan did not show any signs of this.
- In general, Stan seems to spend somewhat longer to produce results with the same accuracy as `inlabru`, according to our measurements. The computational advantage of `inlabru` seems to increase with the complexity of the model.

Further research on the topic may include an even more thorough investigation of the characteristics of the data that lead to similar and dissimilar results between `inlabru` and Stan. It may also be interesting to perform the same experiments with other MCMC software. And finally, running the Stan estimation for many more iterations may give further insight into what part of the difference is caused by lack of convergence.

Although we have not been able to show that `inlabru` produces exactly equivalent results to Stan, our results still suggest that `inlabru` is in practice a useful tool for Bayesian inference with mortality models. We have demonstrated this by using `inlabru` to produce mortality estimates and forecasts for German cancer data, and shown that these fit well to the observed German cancer mortality.

REFERENCES

- Albert, J. and J. Hu (2019). *Probability and Bayesian Modeling*. 1st ed. Chapman and Hall/CRC.
- Allenby, G., P. Rossi, and R. McCulloch (Jan. 2005). “Hierarchical Bayes Models: A Practitioners Guide”. In: *SSRN Electronic Journal*. DOI: [10.2139/ssrn.655541](https://doi.org/10.2139/ssrn.655541).
- Bachl, F. E. et al. (2019). “inlabru: an R package for Bayesian spatial modelling from ecological survey data”. In: *Methods in Ecology and Evolution* 10.6, pp. 760–766. DOI: <https://doi.org/10.1111/2041-210X.13168>. eprint: <https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13168>. URL: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13168>.
- Bayes, T. (1763). “LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S”. In: *Philosophical Transactions of the Royal Society of London* 53, pp. 370–418. DOI: [10.1098/rstl.1763.0053](https://doi.org/10.1098/rstl.1763.0053). eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rstl.1763.0053>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rstl.1763.0053>.

- Beguín, J. et al. (Oct. 2012). “Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation”. In: *Methods in ecology and evolution* 3. DOI: [10.1111/j.2041-210X.2012.00211.x](https://doi.org/10.1111/j.2041-210X.2012.00211.x).
- Behrens, H. R. (2021). *Bayesian Inference with the INLA Methodology on Extensions of the Lee-Carter Model for Mortality Projections*. URL: <https://github.com/Helenerb/Project-thesis>.
- Berkum, F. van, K. Antonio, and M. Vellekoop (July 2017). “A Bayesian Joint Model for Population and Portfolio-Specific Mortality”. In: *ASTIN Bulletin* 47, pp. 1–33. DOI: [10.1017/asb.2017.17](https://doi.org/10.1017/asb.2017.17).
- Betancourt, M. (2016). *Identifying the Optimal Integration Time in Hamiltonian Monte Carlo*. arXiv: [1601.00225 \[stat.ME\]](https://arxiv.org/abs/1601.00225).
- (2018). *A Conceptual Introduction to Hamiltonian Monte Carlo*. arXiv: [1701.02434 \[stat.ME\]](https://arxiv.org/abs/1701.02434).
- Blake, D., A. Cairns, and K. Dowd (Feb. 2006). “A Two-Factor Model for Stochastic Mortality With Parameter Uncertainty: Theory and Calibration”. In: *Journal of Risk Insurance* 73, pp. 687–718. DOI: [10.1111/j.1539-6975.2006.00195.x](https://doi.org/10.1111/j.1539-6975.2006.00195.x).
- Booth, H. and L. Tickle (2008). “Mortality Modelling and Forecasting: a Review of Methods”. In: *Annals of Actuarial Science* 3.1-2, pp. 3–43. DOI: [10.1017/S1748499500000440](https://doi.org/10.1017/S1748499500000440).
- Brooks, S., A. Gelman, and G. Jones (2011). *Handbook of Markov Chain Monte Carlo*. 1st ed. Chapman and Hall/CRC. DOI: <https://doi.org/10.1201/b10905>.
- Brouhns, N., M. Denuit, and J. K. Vermunt (2002). “A Poisson log-bilinear regression approach to the construction of projected lifetables”. In: *Insurance: Mathematics and Economics* 31.3, pp. 373–393. ISSN: 0167-6687. DOI: [https://doi.org/10.1016/S0167-6687\(02\)00185-3](https://doi.org/10.1016/S0167-6687(02)00185-3). URL: <https://www.sciencedirect.com/science/article/pii/S0167668702001853>.

- Cairns, A. J. G. et al. (2009). “A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States”. In: *North American Actuarial Journal* 13.1, pp. 1–35. DOI: [10.1080/10920277.2009.10597538](https://doi.org/10.1080/10920277.2009.10597538). eprint: <https://doi.org/10.1080/10920277.2009.10597538>. URL: <https://doi.org/10.1080/10920277.2009.10597538>.
- Cairns, A. J. G. et al. (2011). “Mortality density forecasts: An analysis of six stochastic mortality models”. In: *Insurance: Mathematics and Economics* 48.3, pp. 355–367. ISSN: 0167-6687. DOI: <https://doi.org/10.1016/j.insmatheco.2010.12.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0167668710001484>.
- Clayton, D. and E. Schifflers (1987). “Models for temporal variation in cancer rates. II: Age–period–cohort models”. In: *Statistics in Medicine* 6.4, pp. 469–481. DOI: <https://doi.org/10.1002/sim.4780060406>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.4780060406>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780060406>.
- Currie, I. (2016). “On fitting generalized linear and non-linear models of mortality”. In: *Scandinavian Actuarial Journal* 2016.4, pp. 356–383. DOI: [10.1080/03461238.2014.928230](https://doi.org/10.1080/03461238.2014.928230). eprint: <https://doi.org/10.1080/03461238.2014.928230>. URL: <https://doi.org/10.1080/03461238.2014.928230>.
- Czado, C., A. Delwarde, and M. Denuit (2005). “Bayesian Poisson log-bilinear mortality projections”. In: *Insurance: Mathematics and Economics* 36.3, pp. 260–284. ISSN: 0167-6687. DOI: <https://doi.org/10.1016/j.insmatheco.2005.01.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0167668705000302>.
- Delwarde, A., M. Denuit, and C. Partrat (Sept. 2007). “Negative binomial version of the Lee–Carter model for mortality forecasting”. In: *Applied Stochastic Models in Business and Industry* 23, pp. 385–401. DOI: [10.1002/asmb.679](https://doi.org/10.1002/asmb.679).

- Duane, S. et al. (1987). “Hybrid Monte Carlo”. In: *Physics Letters B* 195.2, pp. 216–222. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X). URL: <https://www.sciencedirect.com/science/article/pii/037026938791197X>.
- Gamerman, D. and H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2nd ed. Chapman and Hall/CRC. ISBN: 9781584885870.
- Gelman, A. (2006). “Multilevel (Hierarchical) Modeling: What It Can and Cannot Do”. In: *Technometrics* 48.3, pp. 432–435. DOI: [10.1198/0040170050000000661](https://doi.org/10.1198/0040170050000000661). eprint: <https://doi.org/10.1198/0040170050000000661>. URL: <https://doi.org/10.1198/0040170050000000661>.
- Geman, S. and D. Geman (1984). “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6, pp. 721–741. DOI: [10.1109/TPAMI.1984.4767596](https://doi.org/10.1109/TPAMI.1984.4767596).
- German Centre for Cancer Registry, The Robert Koch Institute (Jan. 2021a). Personal communication, January 2021. URL: <https://www.krebsdaten.de/>.
- (2021b). Database query with estimates of the incidence, prevalence and survival of cancer in Germany based on the epidemiological state cancer registry data. Mortality data provided by the German Federal Statistical Office (Statistisches Bundesamt). Last updated: 17.12.2019. DOI: [DOI: 10.18444/5.03.01.0005.0014.0001](https://doi.org/10.18444/5.03.01.0005.0014.0001). URL: www.krebsdaten.de/abfrage (visited on 03/10/2021).
- Geyer, C. J. (2011). “Introduction to Markov Chain Monte Carlo”. In: Chapter 5 of Brooks, Gelman, and Jones (2011).
- Gilks, W. R., S. Richardson, and D. Spiegelhalter (1995). “Markov Chain Monte Carlo in Practice”. In: *Taylor Francis*. DOI: [10.1201/b14835](https://doi.org/10.1201/b14835).
- Gneiting, v and A. E. Raftery (2007). “Strictly Proper Scoring Rules, Prediction, and Estimation”. In: *Journal of the American Statistical Association* 102.477,

- pp. 359–378. ISSN: 01621459. URL: <http://www.jstor.org/stable/27639845>.
- Gómez Rubio, V. (Feb. 2020). *Bayesian Inference with INLA*. ISBN: 9781138039-872. DOI: [10.1201/97811315175584](https://doi.org/10.1201/97811315175584).
- Hastings, W. K. (1970). “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. In: *Biometrika* 57.1, pp. 97–109. ISSN: 00063444. URL: <http://www.jstor.org/stable/2334940>.
- Hatzopoulos, P. and S. Haberman (Sept. 2011). “A dynamic parameterization modeling for the age–period–cohort mortality”. In: *Insurance: Mathematics and Economics* 49, pp. 155–174. DOI: [10.1016/j.insmatheco.2011.02.007](https://doi.org/10.1016/j.insmatheco.2011.02.007).
- Held, L. and A. Riebler (Oct. 2013). “Comment on "Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort (APC) Problem"”. In: *Demography* 50. DOI: [10.1007/s13524-013-0255-8](https://doi.org/10.1007/s13524-013-0255-8).
- Heuer, C. (1997). “Modeling of Time Trends and Interactions in Vital Rates Using Restricted Regression Splines”. In: *Biometrics* 53.1, pp. 161–177. ISSN: 0006341X, 15410420. URL: <http://www.jstor.org/stable/2533105>.
- Hoffman, M. D. and A. Gelman (2014). “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo”. In: *Journal of Machine Learning Research* 15.47, pp. 1593–1623. URL: <http://jmlr.org/papers/v15/hoffman14a.html>.
- Holford, T. (Mar. 2006). “Approaches to fitting age-period-cohort models with unequal intervals”. In: *Statistics in medicine* 25, pp. 977–93. DOI: [10.1002/sim.2253](https://doi.org/10.1002/sim.2253).
- Hunt, A. and D. Blake (2020a). “Identifiability in age/period mortality models”. In: *Annals of Actuarial Science* 14.2, pp. 461–499. DOI: [10.1017/S1748499520000111](https://doi.org/10.1017/S1748499520000111).

- Hunt, A. and D. Blake (2020b). “Identifiability in age/period/cohort mortality models”. In: *Annals of Actuarial Science* 14.2, pp. 500–536. DOI: [10.1017/S1748499520000123](https://doi.org/10.1017/S1748499520000123).
- (2021a). “A Bayesian Approach to Modeling and Projecting Cohort Effects”. In: *North American Actuarial Journal* 25.sup1, S235–S254. DOI: [10.1080/10920277.2019.1649157](https://doi.org/10.1080/10920277.2019.1649157). eprint: <https://doi.org/10.1080/10920277.2019.1649157>. URL: <https://doi.org/10.1080/10920277.2019.1649157>.
- (2021b). “On the Structure and Classification of Mortality Models”. In: *North American Actuarial Journal* 25.sup1, S215–S234. DOI: [10.1080/10920277.2019.1649156](https://doi.org/10.1080/10920277.2019.1649156). eprint: <https://doi.org/10.1080/10920277.2019.1649156>. URL: <https://doi.org/10.1080/10920277.2019.1649156>.
- Hunt, A. and A. M. Villegas (2015). “Robustness and convergence in the Lee–Carter model with cohort effects”. In: *Insurance: Mathematics and Economics* 64, pp. 186–202. ISSN: 0167-6687. DOI: <https://doi.org/10.1016/j.insmatheco.2015.05.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0167668715000852>.
- Keiding, N. (1990). “Statistical inference in the Lexis diagram”. In: *Philosophical Transactions: Physical Sciences and Engineering* (332), pp. 487–509.
- Keilman, N. (2020). “Evaluating Probabilistic Population Forecasts”. In: *Economie et Statistique / Economics and Statistics* 520-521, pp. 49–64. DOI: <https://doi.org/10.24187/ecostat.2020.520d.2033>. URL: <https://www.sciencedirect.com/science/article/pii/S0167947313001552>.
- Knorr-Held, L. and E. Rainer (2001). “Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction”. In: *Biostatistics* 2 (1), pp. 109–129. DOI: <https://doi.org/10.1093/biostatistics/2.1.109>.
- Lee, Ronald D. and Lawrence R. Carter (1992). “Modeling and Forecasting U. S. Mortality”. In: *Journal of the American Statistical Association* 87.419, pp. 659–671. ISSN: 01621459. URL: <http://www.jstor.org/stable/2290201>.

- Lexis, W. (1875). *Einleitung in die Theorie der Bevölkerungsstatistik*. Strasbourg: Karl Trübner.
- Lindgren, v and F. E. Bachl (2021). *Iterative INLA method*. URL: <https://inlabru-org.github.io/inlabru/articles/method.html> (visited on 04/03/2021).
- Luo, L. (Dec. 2013). “Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem”. In: *Demography* 50. DOI: [10.1007/s13524-013-0243-z](https://doi.org/10.1007/s13524-013-0243-z).
- Martino, S. and A. Riebler (2019). *Integrated Nested Laplace Approximations (INLA)*. arXiv: [1907.01248 \[stat.CO\]](https://arxiv.org/abs/1907.01248).
- Metropolis, N. et al. (1953). “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6, pp. 1087–1092. DOI: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114). eprint: <https://doi.org/10.1063/1.1699114>. URL: <https://doi.org/10.1063/1.1699114>.
- Morris, M. (2019). *Spatial Models in Stan: Intrinsic Auto-Regressive Models for Areal Data*. URL: https://mc-stan.org/users/documentation/case-studies/icar_stan.html.
- Neal, R. M. (1993). *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto Toronto, ON, Canada.
- (May 2011). “MCMC using Hamiltonian dynamics”. In: Chapter 5 of Brooks, Gelman, and Jones (2011). DOI: [10.1201/b10905](https://doi.org/10.1201/b10905). URL: <http://dx.doi.org/10.1201/b10905>.
- Renshaw, A. and S. Haberman (Oct. 2003). “Lee-Carter Mortality Forecasting With Age-Specific Enhancement”. In: *Insurance: Mathematics and Economics* 33, pp. 255–272. DOI: [10.1016/S0167-6687\(03\)00138-0](https://doi.org/10.1016/S0167-6687(03)00138-0).
- (2006). “A cohort-based extension to the Lee–Carter model for mortality reduction factors”. In: *Insurance: Mathematics and Economics* 38.3, pp. 556–570. ISSN: 0167-6687. DOI: <https://doi.org/10.1016/j.insmatheco>.

- 2005.12.001. URL: <https://www.sciencedirect.com/science/article/pii/S0167668705001678>.
- Renshaw, A. and S. Haberman (Oct. 2009). “On Age-Period-Cohort Parametric Mortality Rate Projections”. In: *Insurance: Mathematics and Economics* 45, pp. 255–270. DOI: [10.1016/j.insmatheco.2009.07.006](https://doi.org/10.1016/j.insmatheco.2009.07.006).
- Riebler, A. (2010). “Multivariate age-period-cohort models”. PhD thesis. University of Zurich. URL: <https://doi.org/10.5167/uzh-38448>.
- Riebler, A. and L. Held (2010). “The analysis of heterogeneous time trends in multivariate age-period-cohort models”. In: *Biostatistics* 11,1, pp. 57–69. DOI: [10.1093/biostatistics/kxp037](https://doi.org/10.1093/biostatistics/kxp037).
- Riebler, A., L. Held, and H. Rue (2012a). “Estimation and extrapolation of time trends in registry data—Borrowing strength from related populations”. In: *The Annals of Applied Statistics* 6.1, pp. 304–333. DOI: [10.1214/11-AOAS498](https://doi.org/10.1214/11-AOAS498). URL: <https://doi.org/10.1214/11-AOAS498>.
- Riebler, A. et al. (2012b). “Gender-specific differences and the impact of family integration on time trends in age-stratified Swiss suicide rates”. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 175.2, pp. 473–490. ISSN: 09641998, 1467985X. URL: <http://www.jstor.org/stable/23251246>.
- Robert, C. and G. Casella (2004). *Monte Carlo Statistical Methods*. Springer. DOI: <https://doi.org/10.1007/978-1-4757-4145-2>.
- Rue, H. (2022). *Random walk model of order 1 (RW1)*. URL: <https://inla.r-inla-download.org/r-inla.org/doc/latent/rw1.pdf> (visited on 01/18/2022).
- Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications*. 1st ed. Chapman and Hall/CRC. DOI: <https://doi.org/10.1201/9780203492024>.
- Rue, H., S. Martino, and N. Chopin (2009). “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approxi-

- mations”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 319–392. DOI: <https://doi.org/10.1111/j.1467-9868.2008.00700.x>. eprint: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2008.00700.x>. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00700.x>.
- Rue, H. et al. (2016). *Bayesian Computing with INLA: A Review*. arXiv: 1604.00860 [stat.ME].
- Simpson, D. et al. (2017). “Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors”. In: *Statistical Science* 32.1, pp. 1–28. ISSN: 08834237, 21688745. URL: <http://www.jstor.org/stable/26408114>.
- Stan Development Team (2019). *Stan Modeling Language Users Guide and Reference Manual*. Version 2.28. URL: <https://mc-stan.org>.
- Stan Discussion Forum (2019). *Time series random walk parameters*. URL: <https://discourse.mc-stan.org/t/time-series-random-walk-parameters/6983>.
- Statistisches Bundesamt (2013). *80.2 million inhabitants lived in Germany on 9 May 2011*. URL: https://www.zensus2011.de/SharedDocs/AktuellesEN/Press_release_of_the_federal_statistical_office.html?nn=3360426 (visited on 06/03/2021).
- Tierney, L. and J. B. Kadane (1986). “Accurate Approximations for Posterior Moments and Marginal Densities”. In: *Journal of the American Statistical Association* 81.393, pp. 82–86. DOI: 10.1080/01621459.1986.10478240. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1986.10478240>. URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478240>.
- Wiśniowski, A. et al. (May 2015). “Bayesian Population Forecasting: Extending the Lee-Carter Method”. In: *Demography* 52. DOI: 10.1007/s13524-015-0389-y.

Wong, J. S. T., J. J. Forster, and P. W. F. Smith (2018). “Bayesian mortality forecasting with overdispersion”. In: *Insurance: Mathematics and Economics* 83, pp. 206–221. ISSN: 0167-6687. DOI: <https://doi.org/10.1016/j.insmatheco.2017.09.023>. URL: <https://www.sciencedirect.com/science/article/pii/S0167668717302433>.



PROCEDURES IN THE SIMULATION STUDY

In this chapter, we describe some of the complimentary procedures and investigations that were performed in the simulation study of Chapter 5.

A.1 COMPARISON OF RANDOM WALK MODELS

In this section, we perform a simple analysis to ensure that our implementation of a random walk of order one in Stan is equivalent to the RW1 model that is implemented in `inlabru`.

In `inlabru`, a first-order random walk is defined by Gaussian distributed differences, as

$$\Delta x_i = x_i - x_{i-1} \sim \mathcal{N}(0, 1/\tau) \quad (\text{A.1.1})$$

(Rue and Held 2005).

We consider two different implementations of this in Stan. The first is a simple implementation, where we slightly rewrite Equation A.1.1 as

$$\begin{aligned} x_1 &\sim \mathcal{N}(0, 10^4) \\ x_{i+1} &= x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1/\tau). \end{aligned} \quad (\text{A.1.2})$$

This way of modeling a first-order random walk has previously been employed by e.g., Knorr-Held and Rainer (2001). The random walk on this form is easy

to implement in Stan, as we can simply give each step x_i a Gaussian prior with a mean of x_{i-1} . We refer to this implementation as the "Stepwise" random walk.

Secondly, we consider an implementation closer to Equation A.1.1. To implement this in Stan, we use a transformed parameter that is the differences of the random walk and define Gaussian priors on these. This implementation is slightly more complicated, as it involves two sets of parameters for each random walk in the model, one set containing the steps of the random walk and one set containing the differences. We refer to this implementation as the "Difference" implementation of a first-order random walk.

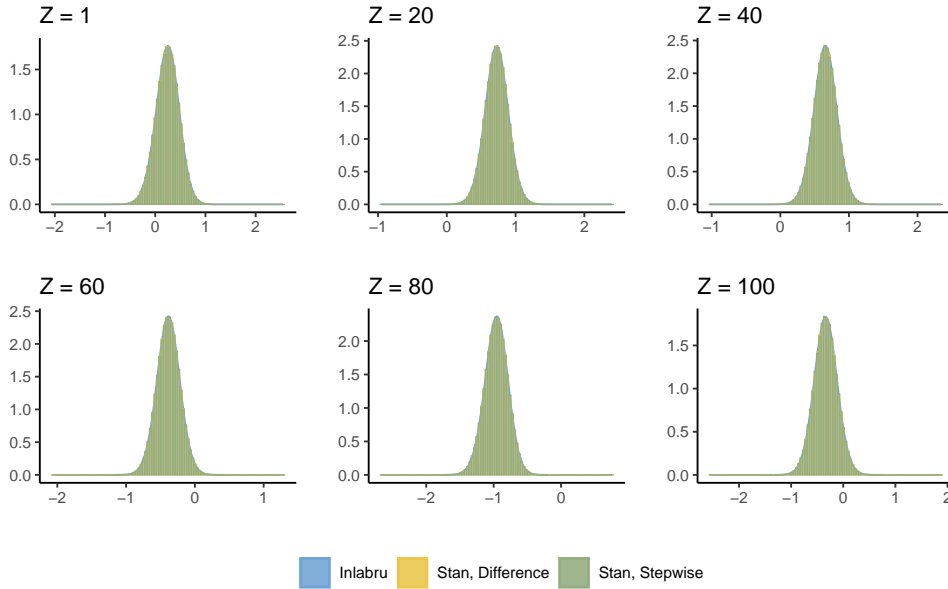
We use a simple example to compare the different implementations in Stan to the random walk model of `inlabru` and define the simple model

$$y_z \sim \eta_z + \varepsilon_z, \quad \eta_z \sim \text{RW1}(\tau_\eta), \quad \varepsilon_x \sim \mathcal{N}(0, 1/\tau_y), \quad (\text{A.1.3})$$

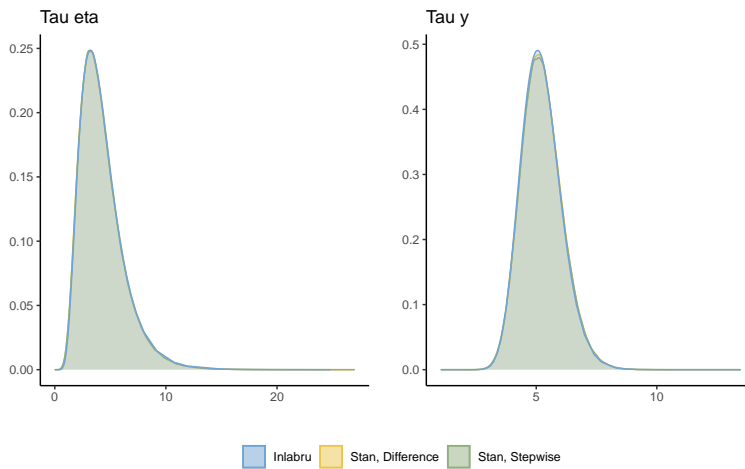
for $z = 0, \dots, 99$. We sample the data by $y_z = \sin(z/16.5) + \varepsilon_z$, where ε_z is independently sampled from a Gaussian distribution with zero mean and standard deviation $\sigma = 0.5$. We do inference on this data with the model of Expression A.1.3 with `inlabru` and the two implementations of Stan. The Stan programs, as well as the script for performing this comparison and produce the full results can be found at <https://github.com/Helenerb/Masters-thesis>.

Figure A.1.1a displays the estimated random walk, for some selected points, and Figure A.1.1b displays the estimated τ_η and τ_y for the three implementations. We observe that the results produced by `inlabru` and the two Stan implementations seem practically identical. The two Stan implementations require roughly the same amount of time to run for the same number of iterations, and we do not see any clear difference in the convergence between the two. Figure A.1.2 displays the trace plots of the precisions τ_η and τ_y , two steps of the random walk and the log-precisions θ_η and θ_y after running the "Stepwise" and the "Difference" Stan programs for four chains and 20000 iterations. The two processes seem to explore the parameter space equally well. Since both Stan implementations seem to be equivalent to the random

First-Order Random Walk Implementations in Stan and inlabru



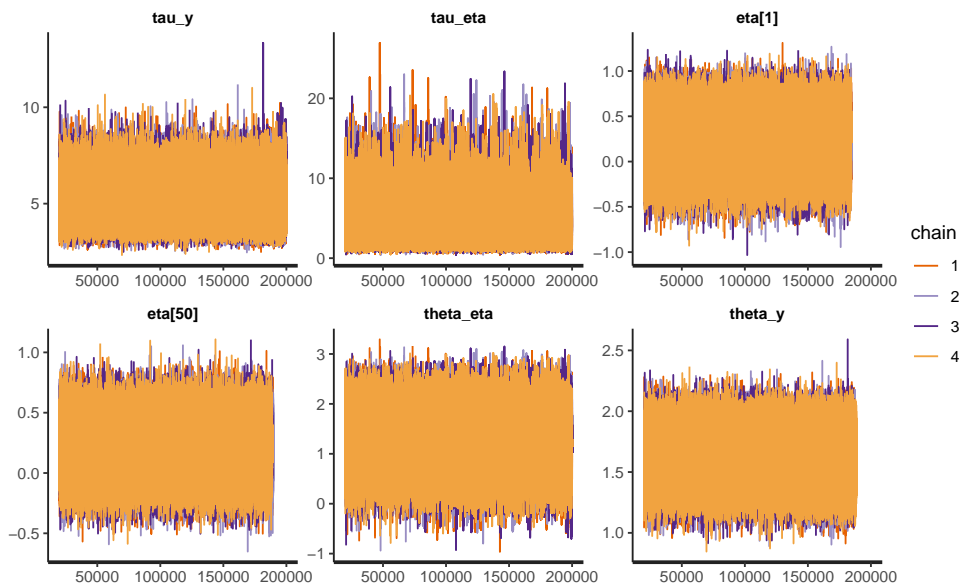
(a) Selected values of the random walk, as estimated by inlabru and two different implementations of Stan



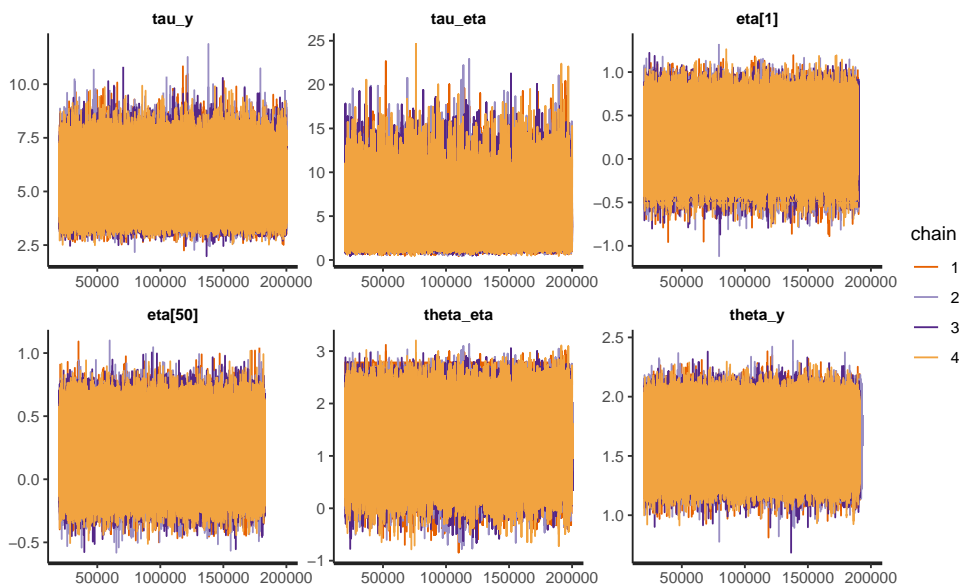
(b) The precisions of the random walk, as estimated by inlabru and two different implementations of Stan

Figure A.1.1: A first-order random walk, as estimated by inlabru , the "Stepwise" Stan implementation, and the "Difference" Stan implementation.

Trace Plots for the "Stepwise" and "Difference" Stan Implementations



(a) Trace plots for the "Stepwise" implementation



(b) Trace plots for the "Difference" implementation

Figure A.1.2: Trace plots of some selected parameters from running Stan for 20000 iterations and 4 chains for the "Stepwise" random walk implementation (upper) and the "Difference" random walk implementation (lower)

walk implementation and the two Stan programs seemingly perform equally well, we could use either in our further investigation. We use the "Stepwise" implementation in our investigation when considering random walk models, because of its simplicity.

A.2 IMPLEMENTATION OF CONSTRAINTS IN STAN AND inlabru

The sum-to-zero and sum-to-unit constraints that are used for the random effects throughout Chapter 5 (see Expression 5.1.3) are implemented differently in inlabru and Stan. The inlabru framework implements constraints as described by Rue and Held (2005). As their exact approach is complicated to implement in Stan, we use an alternative approach and model the constraints by imposing so-called soft constraints (Morris 2019). This involves imposing a sharp prior with mean at zero for on the sum of the effects:

$$\begin{aligned} \sum_x \beta_x &\sim \mathcal{N}(1, sd = 0.001 \cdot X), \\ \sum_t \kappa_t &\sim \mathcal{N}(0, sd = 0.001 \cdot T). \end{aligned}$$

To verify that this is equivalent to inlabru, we extend the simple example of Section A.1 to include a constraint:

$$y_z \sim \mathcal{N}(\mu + \eta_z, 1/\tau_y), \quad \eta_z \sim \text{RW1}(\tau_\eta), \quad \mu \sim \mathcal{N}(0, 1/0.001),$$

with

$$\sum_z \eta_z = 0. \tag{A.2.1}$$

We apply this model to the data described in Section A.1 using both inlabru and Stan, where the latter uses soft constraints for η_z .

Figure A.2.1 displays some values of η_z as estimated by inlabru and Stan, Figure A.2.2 displays the estimated intercept and Figure A.2.3 displays the hyperparameters. The results of inlabru and Stan are very similar for all

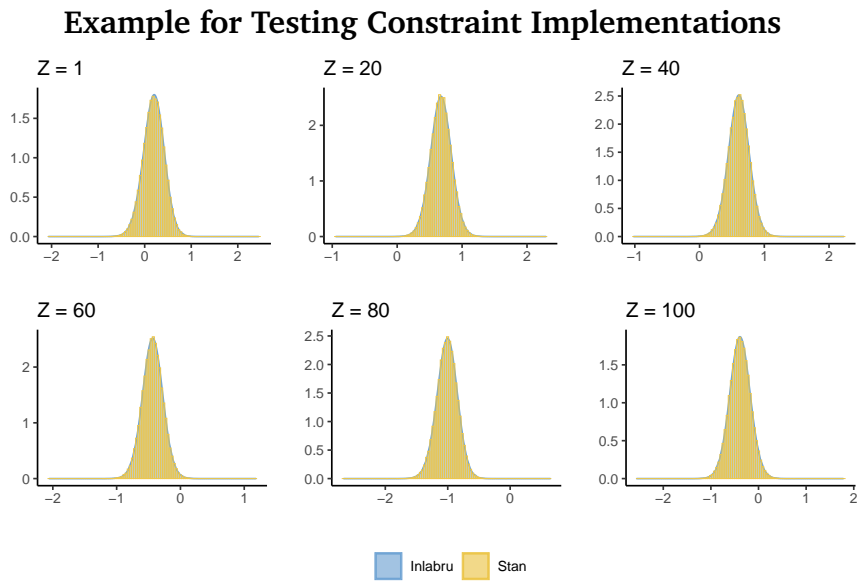


Figure A.2.1: Estimation results of some arbitrary values of η_z from the example discussed in Section A.2, comparing implementations of constraints

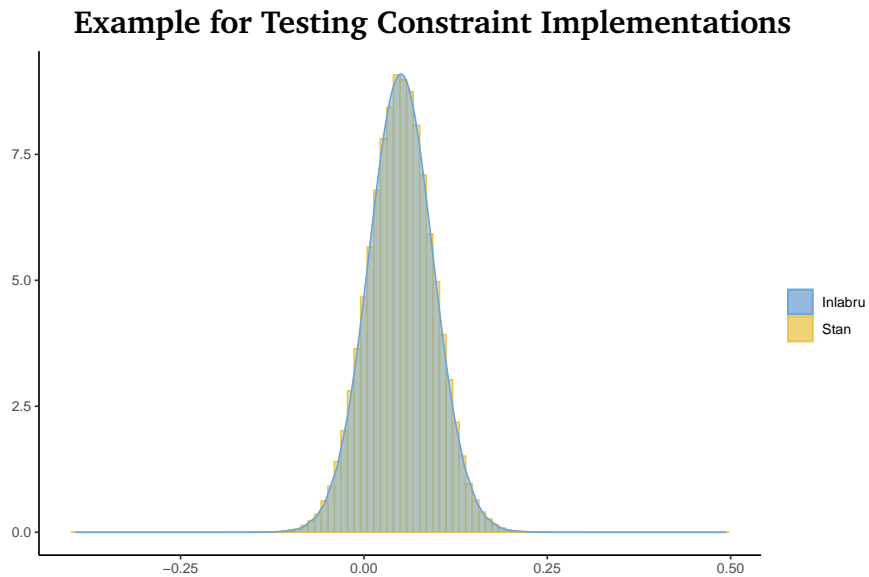


Figure A.2.2: Estimated intercept μ from the example discussed in Section A.2, comparing implementations of constraints

Example for Testing Constraint Implementations

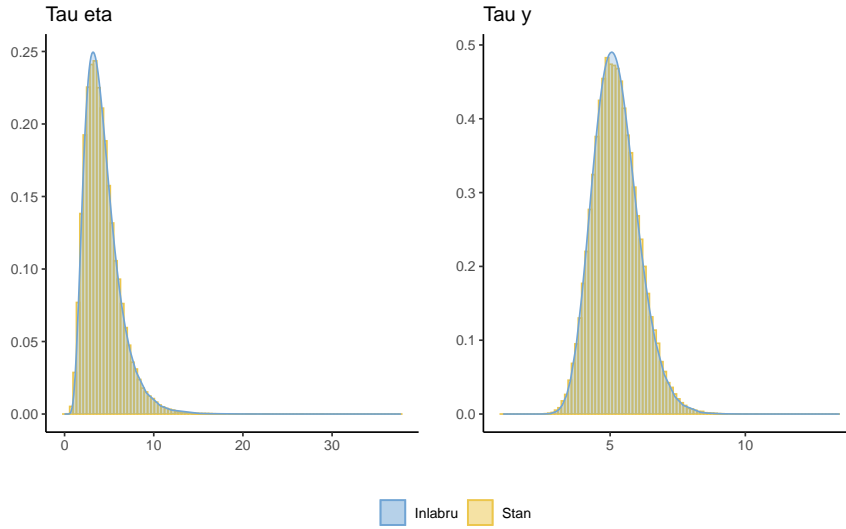


Figure A.2.3: Estimated precisions τ_η and τ_y from the example discussed in Section A.2, comparing implementations of constraints

model parameters, and we conclude that the soft-constraint implementation of Stan is equivalent to that of `inlabru`.

The full code for producing these results can be found at <https://github.com/Helenerb/Masters-thesis>.

A.3 GENERATING SYNTHETIC DATA

For the first part of our analysis, we use synthetic sets of mortality data, where we know the underlying values of the time effects.

In the simulation study of Chapter 5, we consider two sets of mortality models. One set of mortality models with Gaussian likelihoods,

$$\log(m_{x,t}) = \mathcal{N}(\eta_{x,t}, 1/\tau_\varepsilon),$$

for which the response variable is the mortality rate $m_{x,t}$, and one set with Poisson likelihoods

$$Y_{x,t} \sim \mathcal{Poisson}(E_{x,t} \cdot e^{\eta_{x,t}}), \quad (\text{A.3.1})$$

where the response variable is the observed cases of deaths $Y_{x,t}$. We generate the synthetic data from the following structure of the predictor $\eta_{x,t}$:

$$\eta_{x,t} = \mu + \alpha_x + \beta_x \cdot \kappa_t + \varepsilon_{x,t}, \quad (\text{A.3.2})$$

with the constraints

$$\sum_x \alpha_x = 0, \quad \sum_x \beta_x = 1, \quad \sum_t \kappa_t = 0. \quad (\text{A.3.3})$$

For the random effects, we make the following model choices

$$\begin{aligned} \mu &\sim \mathcal{N}(0, 1/0.001) \\ \alpha_x &\sim \text{RW1}(\tau_\alpha) \\ \beta_x &\sim \text{iid}(\tau_\beta) \\ \kappa_t &\sim \text{RW2}(\tau_\kappa) \\ \varepsilon_{x,t} &\sim \text{iid}(\tau_\varepsilon). \end{aligned} \quad (\text{A.3.4})$$

Here, RW1, RW2 and iid denote the first-order random walk, second-order random walk and independent random noise models of Rue and Held (2005). τ_α , τ_β , τ_κ and τ_ε denote the precisions. We use the following hyperpriors:

$$\begin{aligned} \tau_\alpha, \tau_\beta, \tau_\varepsilon &\sim \mathcal{Gamma}(1, 0.00005) \\ \tau_\kappa &\sim \mathcal{Gamma}(1, 0.005) \end{aligned} \quad (\text{A.3.5})$$

To generate the synthetic values for $Y_{x,t}$ and $m_{x,t}$, we employ the following procedure:

1. Fit the model in Expression A.3.1 to the male German lung cancer data described in Section 3:
 - (a) The observed deaths $Y_{x,t}$ are the observed cancer deaths
 - (b) The population at risk $E_{x,t}$ is the German male population

- (c) The predictor $\eta_{x,t}$ is defined as in Expressions A.3.2 through A.3.5
2. Extract the inlabru estimates of the precisions $\tau_\alpha, \tau_\beta, \tau_\kappa, \tau_\varepsilon$
 3. Extract the inlabru estimate for the intercept μ
 4. Use the values of $\tau_\alpha, \tau_\beta, \tau_\kappa, \tau_\varepsilon$ to sample values for the random effects $\alpha_x, \beta_x, \kappa_t$ and $\varepsilon_{x,t}$, by the models in Expression A.3.4
 5. Shift α_x, β_x and κ_t so they fulfill the constraints in Expression A.3.3
 6. Construct $\eta_{x,t}$ by Equation A.3.2, with the inlabru estimate for μ and and the sampled $\alpha_x, \beta_x, \kappa_t$ and $\varepsilon_{x,t}$
 7. Generate synthetic observations:
 - (a) Find synthetic mortality rates by $m_{x,t} = \exp(\eta_{x,t})$
 - (b) Find synthetic observed deaths by sampling

$$Y_{x,t} \sim \mathcal{Poisson}(E_{x,t} \cdot e^{\eta_{x,t}}).$$
 Here $E_{x,t}$ are the real values of the German male population.

We perform this procedure four times, producing four different sets of synthetic data:

- Full.1 : Generated using the full set of male cancer data for $x = 0, \dots, 17$ and $t = 0, \dots, 17$.
- Reduced.1 : Generated using a reduced set of male cancer data for $x = 9, \dots, 17$ and $t = 0, \dots, 17$.
- Full.2 : Generated using the full set of male cancer data for $x = 0, \dots, 17$ and $t = 0, \dots, 17$.¹
- Reduced.2 : Generated using a reduced set of male cancer data for $x = 2, \dots, 17$ and $t = 0, \dots, 17$.

¹ For the Full.2 data, we make a small adjustment to the data generating procedure. At step 4, we adjust the values of τ_β and τ_κ so that the range of κ_t is not too low compared to the range of α_x and so that all values of β_x are positive.

Data	μ	τ_α	τ_β	τ_κ	τ_ε
Full.1	-9.83	1.96	202	338	420
Reduced.1	-6.46	6.01	63.8	455	419
Full.2	-9.83	1.96	1262.5	202.8	420
Reduced.2	-9.83	1.97	174	364	419

Table 7: The exact values of μ , τ_α , τ_β , τ_κ and τ_ε that we use to generate the four sets of data

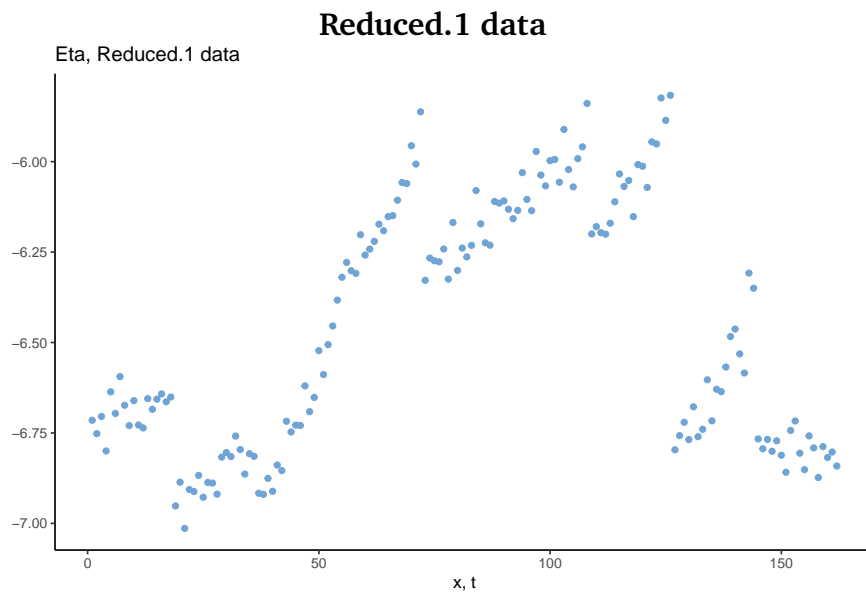


Figure A.3.1: Observed $\eta_{x,t}$ of the Reduced.1 data for the points $(x, t) = (1, 1), \dots, (1, 18), \dots, (9, 1), \dots, (9, 18)$

Throughout Chapter 5, we shift the indices of t and x , so that the age and period effects are one-indexed, since this simplifies the Stan implementation.

Table 7 presents the exact values of μ , τ_α , τ_β , τ_κ and τ_ε that we use to generate the four sets of data. The full code the generation of these sets of data can be found at <https://github.com/Helenerb/Masters-thesis>.

Figure 5.1.1 displays the quantity $\eta_{x,t}$ for the Full.1 data, while Figures A.3.1, A.3.2 and A.3.3 display $\eta_{x,t}$ for the Reduced.1, Full.2 and Reduced.2 data.



Figure A.3.2: Observed $\eta_{x,t}$ of the Full.2 data for the points $(x, t) = (1, 1), \dots, (1, 18), \dots, (16, 1), \dots, (16, 18)$

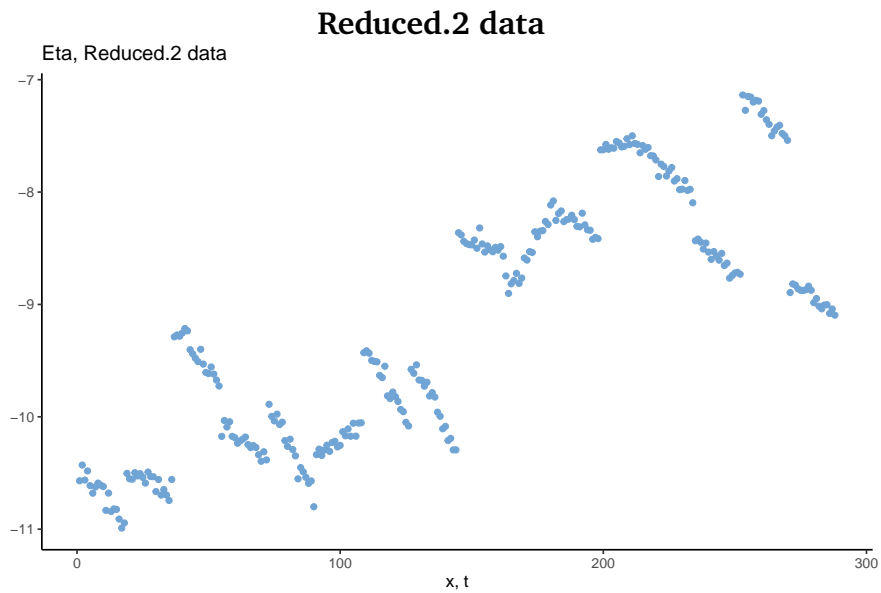


Figure A.3.3: Observed $\eta_{x,t}$ of the Reduced.2 data for the points $(x, t) = (1, 1), \dots, (1, 18), \dots, (18, 1), \dots, (18, 18)$

B

COMPLEMENTARY RESULTS FROM THE SIMULATION STUDY

In this chapter, complementary figures and supporting results of Chapter 5 are presented.

B.1 MODELS WITH GAUSSIAN LIKELIHOODS

In the following section, complementary figures for Section 5.1 are presented.

Figures B.1.1 and B.1.2 display the estimated predictor and random effects for the G1lin and the G2lin models respectively. Figure B.1.3 displays the estimated hyperparameters of model G2lin.

Figure B.1.4 shows the posterior marginal distributions for the G1 model.

Figure B.1.5 displays the estimated predictor and random effects for the G1 model applied to the Reduced.1 data. The corresponding posterior marginals are displayed in Figure B.1.6.

Figure B.1.7 displays the estimated predictor of model G2 at some arbitrary values. Figure B.1.8 shows posterior marginal distributions of β_x and κ_t for the same model.

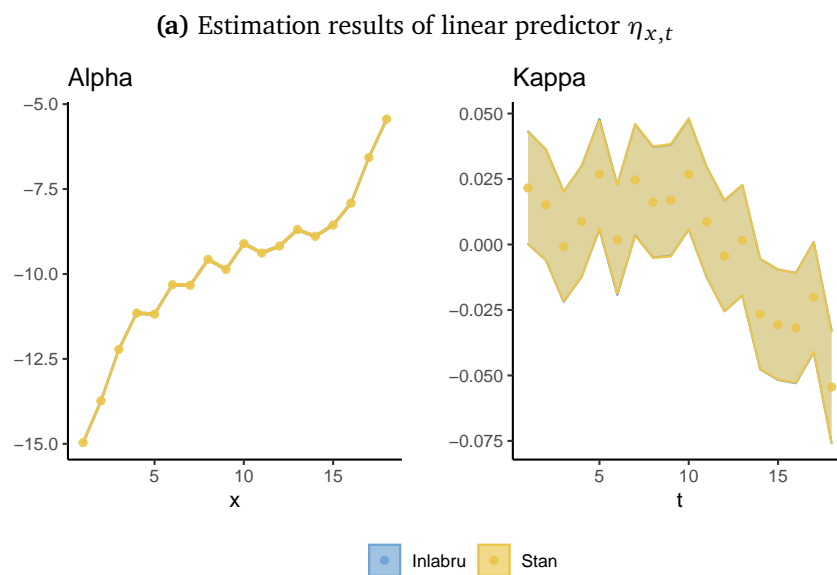
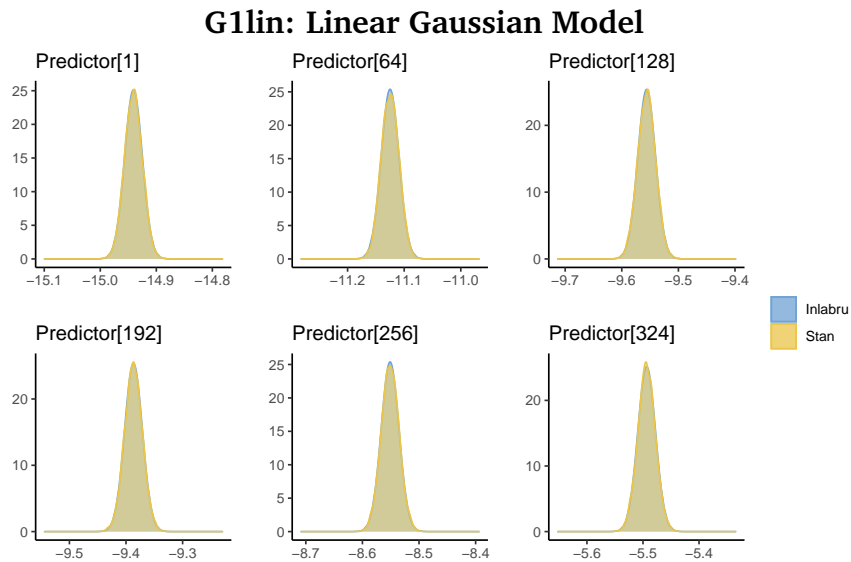
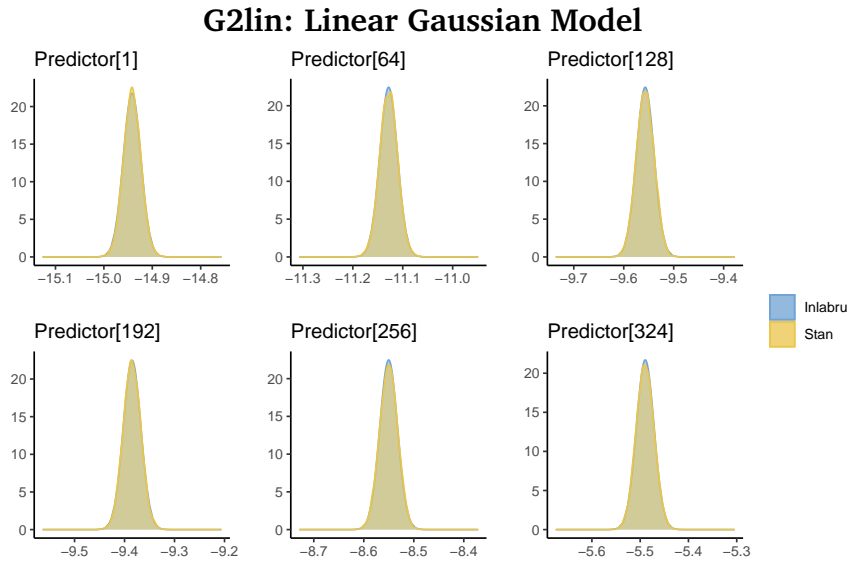
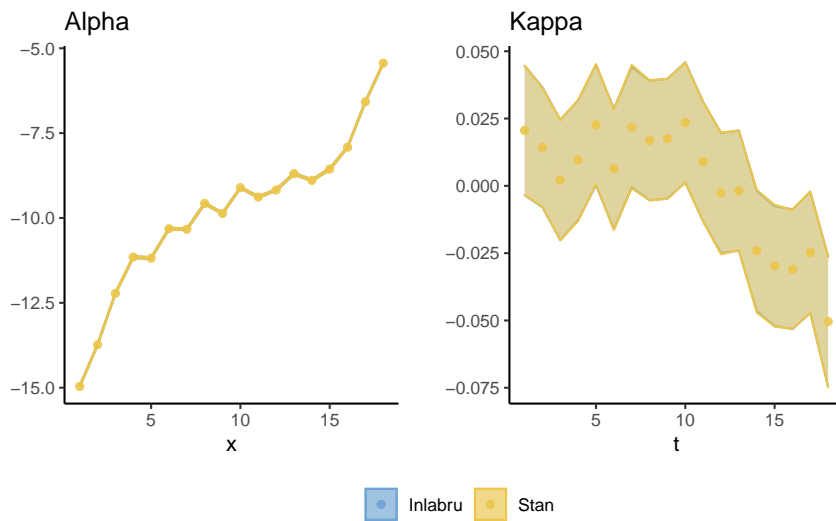


Figure B.1.1: Estimation results from applying the G1lin model to the Full.1 data using Stan and inlabru



(a) Estimation results of linear predictor $\eta_{x,t}$



(b) Estimated random effects.

Figure B.1.2: Estimation results from applying the G2lin model to the Full.1 data using Stan and inlabru

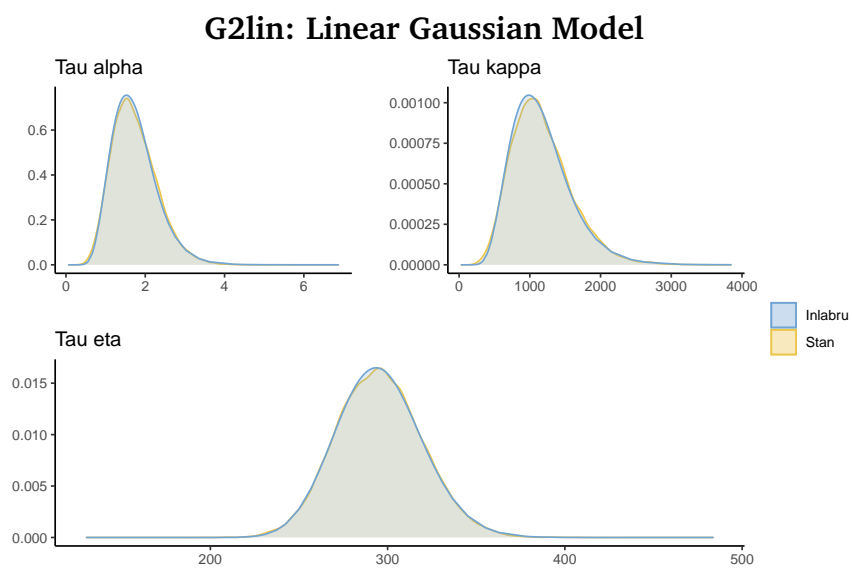
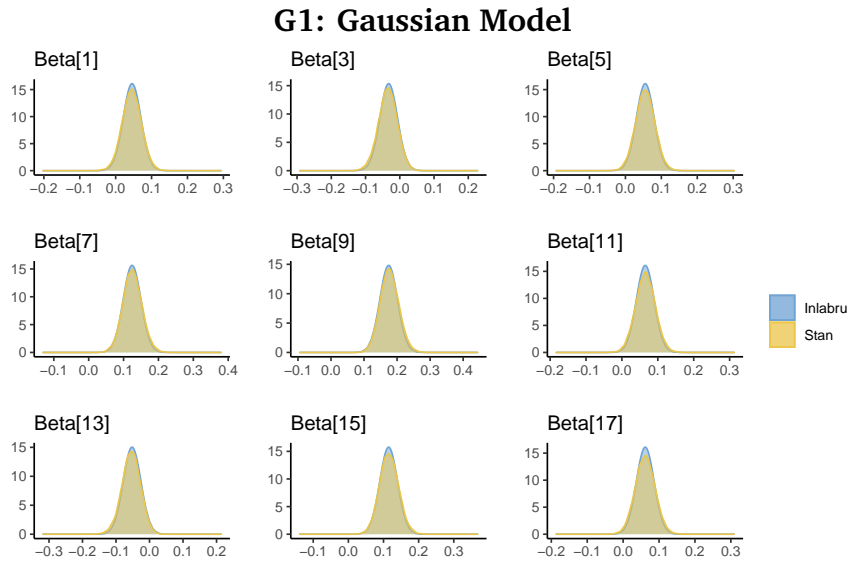
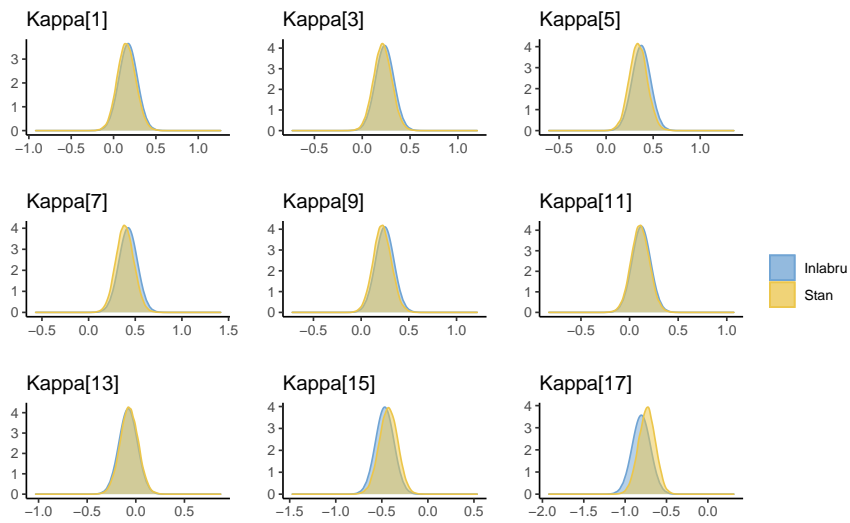


Figure B.1.3: Estimated hyperparameters τ_α , τ_κ and τ_η from applying the G2lin model to the Full.1 data using Stan and `inlabru`



(a) Estimation results of non-linear predictor $\eta_{x,t}$



(b) Estimated random effects

Figure B.1.4: Estimation results from applying the G1 model to the Full.1 data using Stan and inlabru

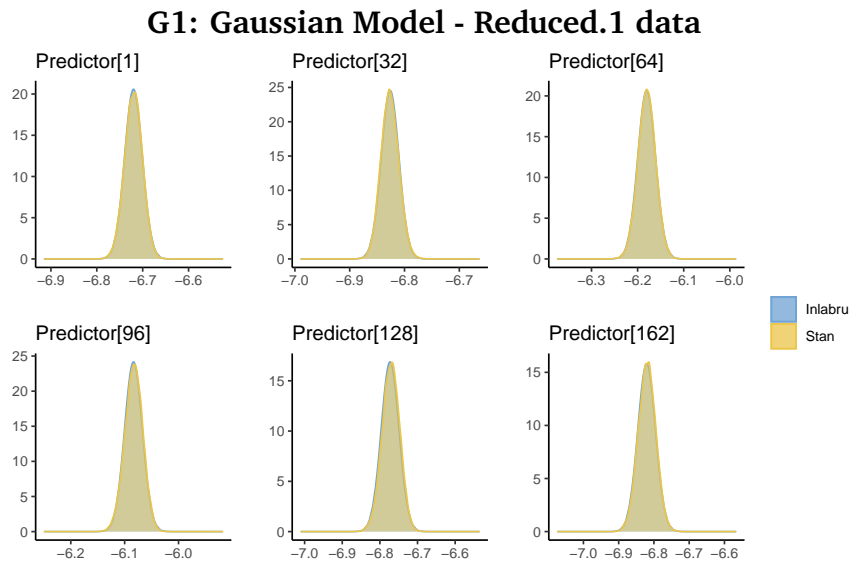
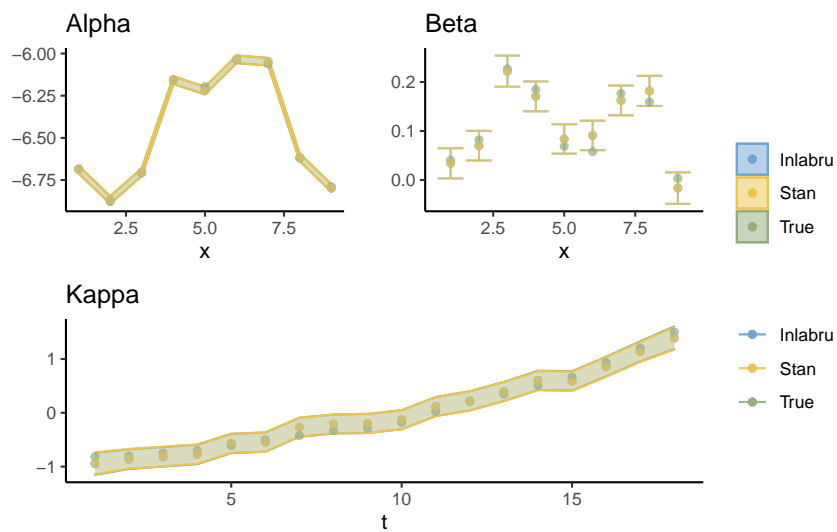
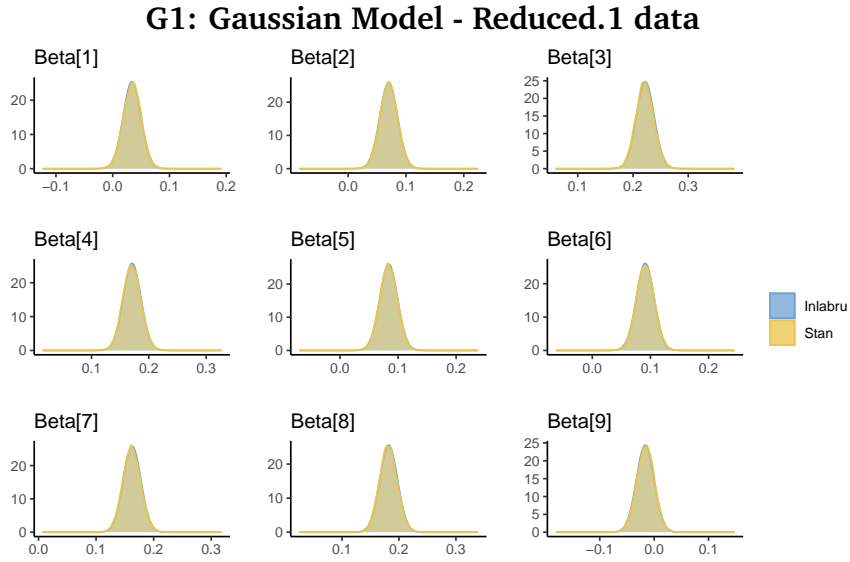
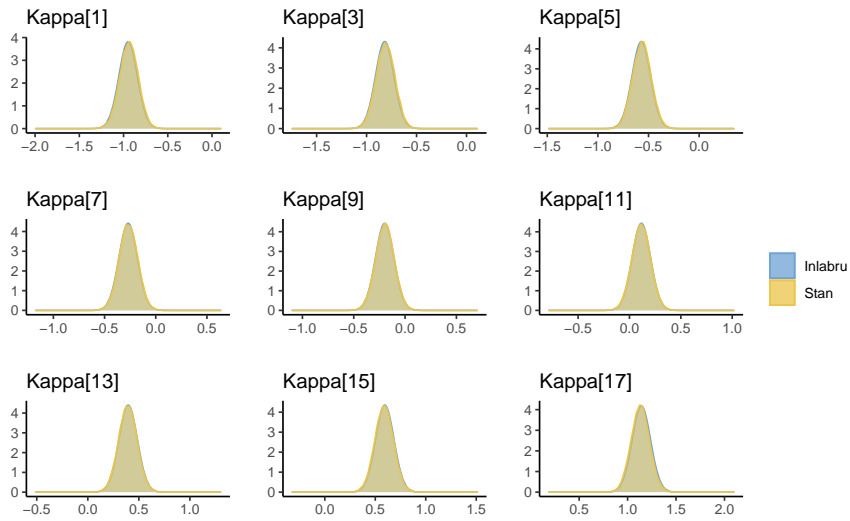
(a) Estimation results of linear predictor $\eta_{x,t}$ (b) Estimated random effects α_x , β_x and κ_t

Figure B.1.5: Estimation results from applying the G1 model to the Reduced.1 data using Stan and inlabru



(a) Estimated posterior marginals of β_x



(b) Estimated posterior marginals of κ_t

Figure B.1.6: Estimation results from applying the G1 model to the Reduced.1 data using Stan and inlabru

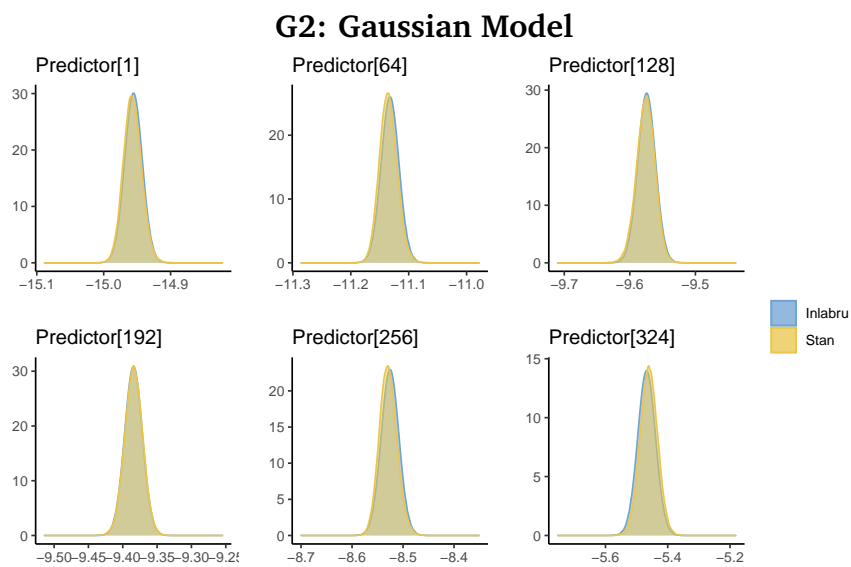
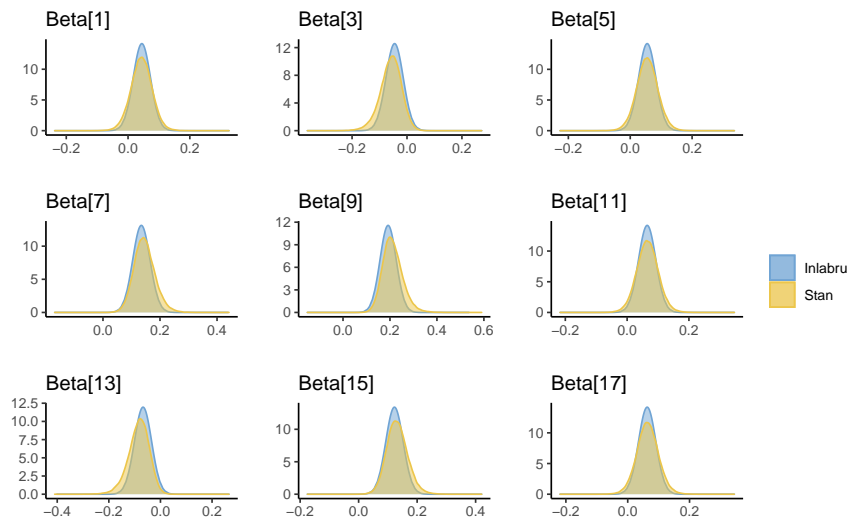
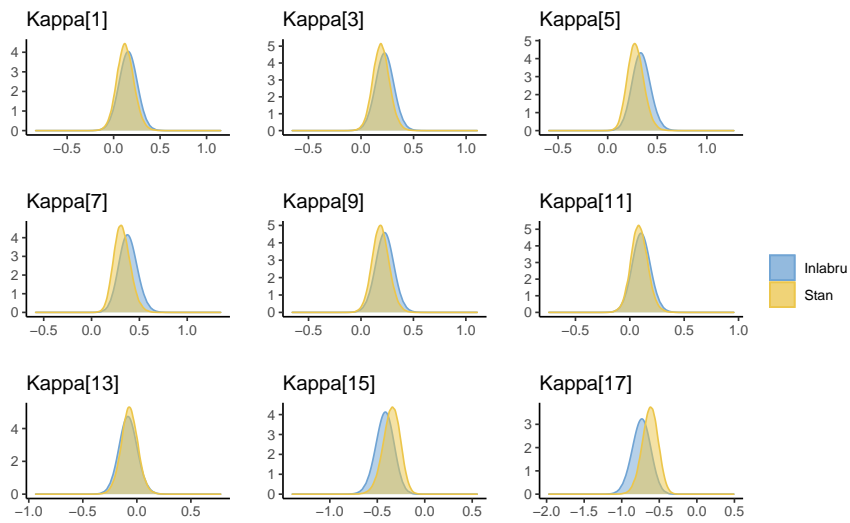


Figure B.1.7: Estimation results of the linear predictor $\eta_{x,t}$ from applying the G2 model to the Full.1 data using Stan and inlabru

G2: Gaussian Model



(a) Estimation of selected values of β_x



(b) Estimation of selected values of κ_t

Figure B.1.8: Estimation results from applying the G2 model to the Full.1 data using Stan and inlabru

B.2 MODELS WITH POISSON LIKELIHOODS

In the following section, complementary figures for Section 5.2 are presented.

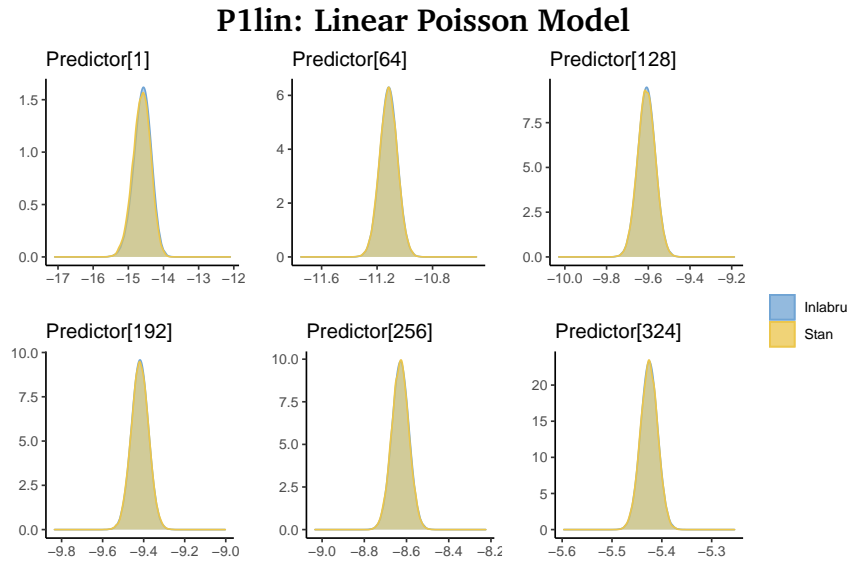
Figures B.2.1 and B.2.2 display the estimated predictors and random effects of the P1lin and P2lin models respectively. Figure B.2.3 display the estimated hyperparameters of the P2lin model.

Figure B.2.4 displays posterior marginal distributions of β_x and κ_t for the P1 model. The estimation results of applying the P1 model to the Reduced.1 data are displayed in Figures B.2.5 (predictor and random effects) and B.2.6 (posterior marginals of β_x and κ_t).

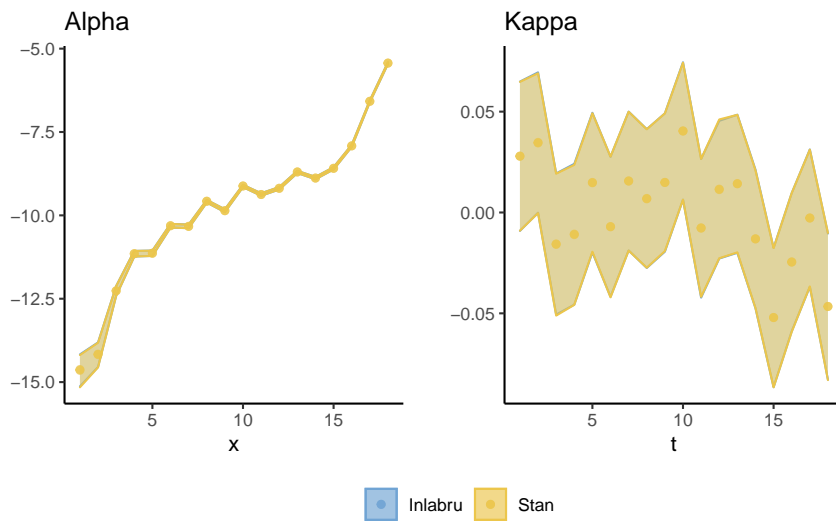
Figures B.2.7 and B.2.8 display posterior marginal distributions of β_x and κ_t when applying the P2 model to the Full.1 data. Figure B.2.9 displays the trace plots of the corresponding Stan run.

Figures B.2.10, B.2.11 and B.2.12 display the estimated predictor, random effects and hyperparameters of applying the P2 model to the Reduced.1 data. The corresponding plots for the Full.2 data are presented in Figures B.2.13, B.2.14 and B.2.15.

Figure B.2.16 displays the trace plots of the Stan run when applying the P2 model to the Reduced.2 data. The estimation results for the P2 model with the Reduced.2 data are included in Figure B.2.17 (the predictor), Figure B.2.18 (the random effects and hyperparameters), Figure B.2.19 (posterior marginals of β_x and κ_t) and Figure B.2.20 (posterior marginals of $\varepsilon_{x,t}$).



(a) Estimation results of linear predictor $\eta_{x,t}$.



(b) Estimation results of random effects α_x and κ_t

Figure B.2.1: Estimation results from applying the P1lin model to the Full.1 data using Stan and inlabru

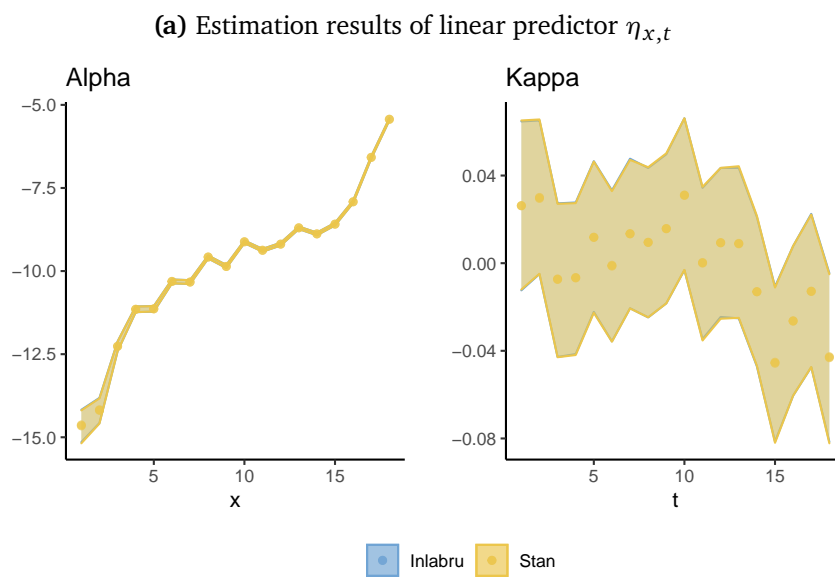
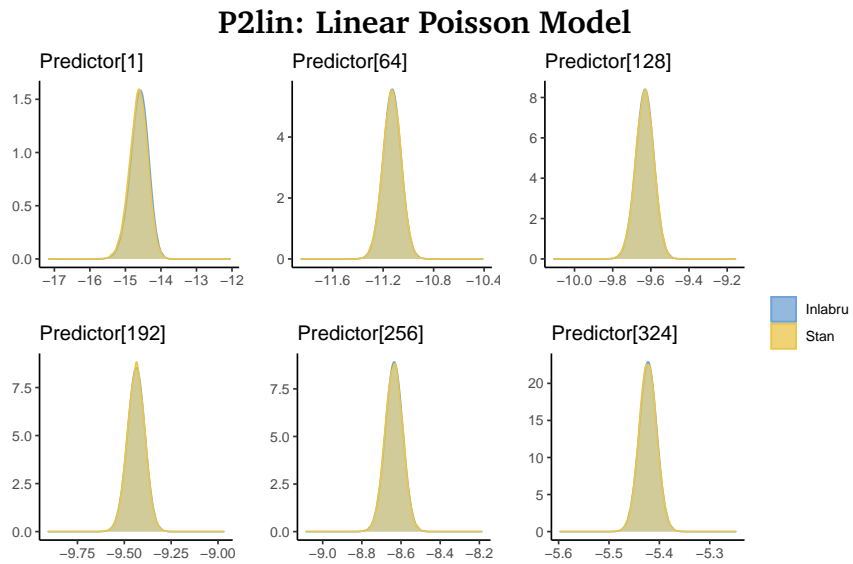


Figure B.2.2: Estimation results from applying the P2lin model to the Full.1 data using Stan and inlabru

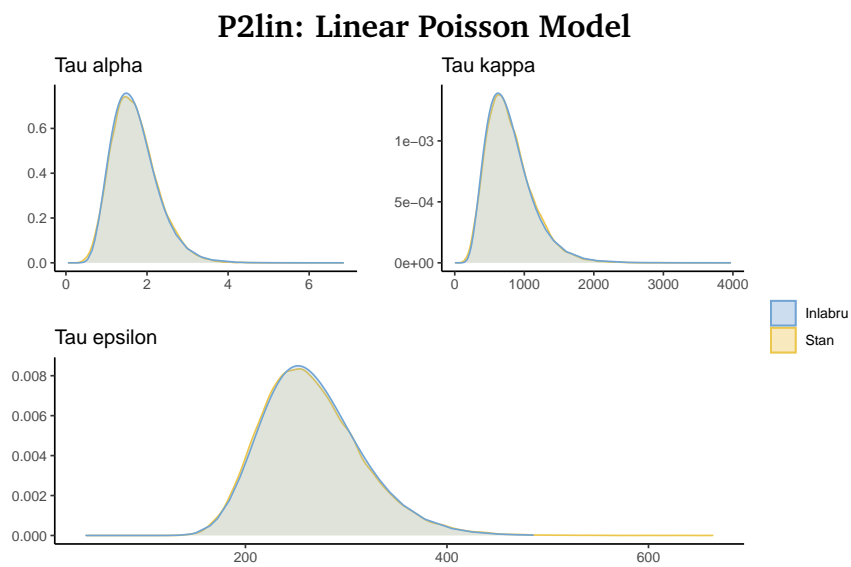


Figure B.2.3: Estimated hyperparameters τ_α , τ_κ and τ_ϵ from applying the P1lin model to the Full.1 data using Stan and inlabru

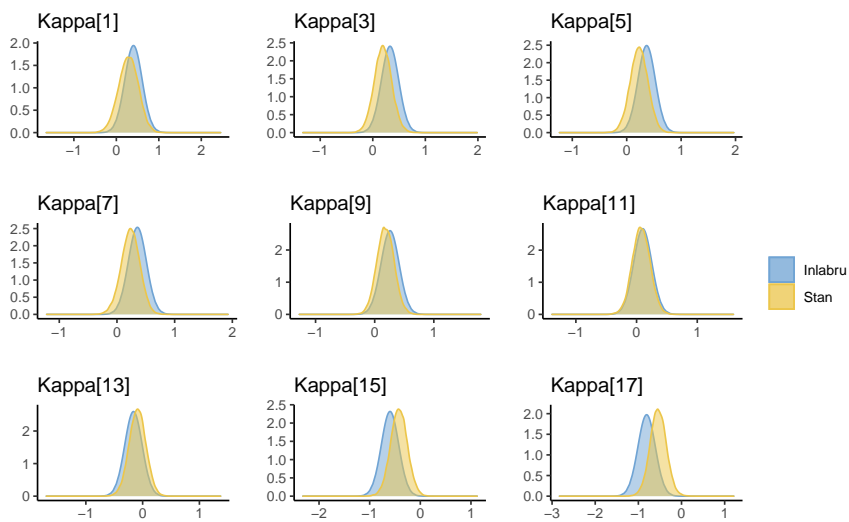
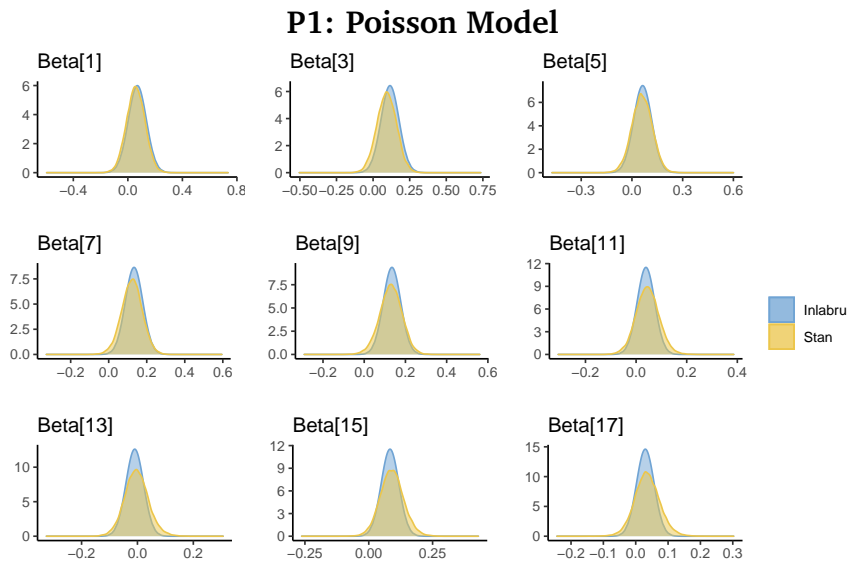
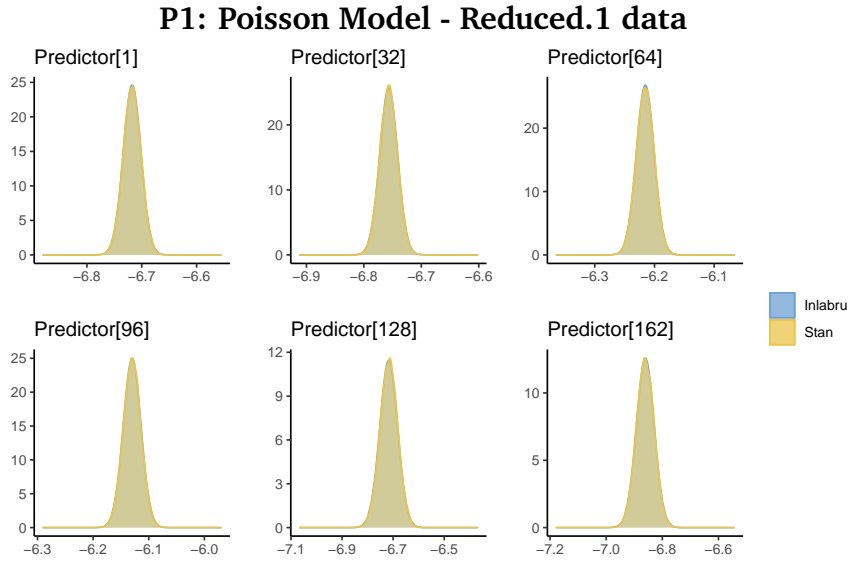
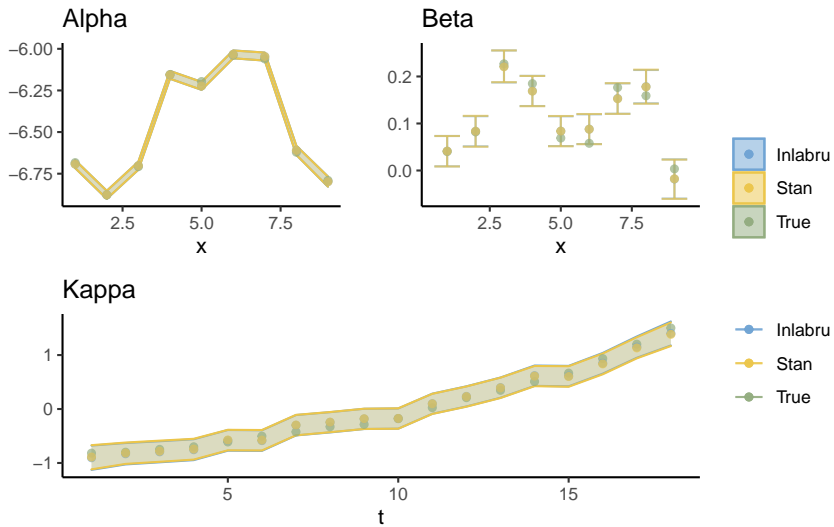


Figure B.2.4: Estimation results from applying the P1 model to the Full.1 data using Stan and inlabru



(a) Estimation results of non-linear predictor $\eta_{x,t}$.



(b) Estimation results of random effects α_x , β_x and κ_t .

Figure B.2.5: Estimation results from applying the P1 model to the Reduced.1 data using Stan and inlabru

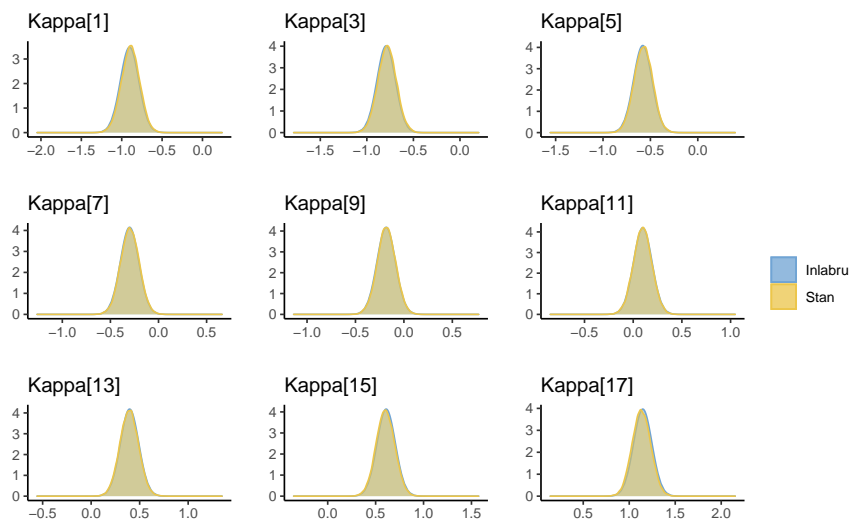
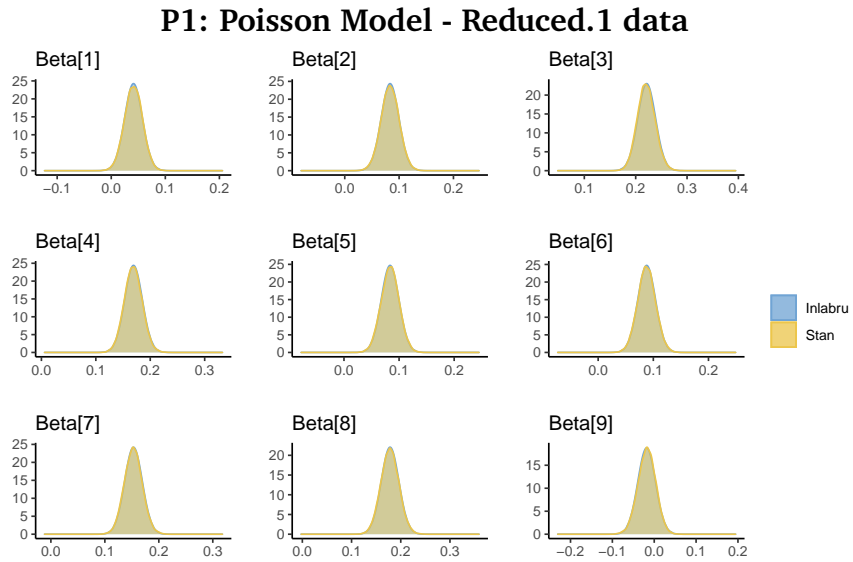
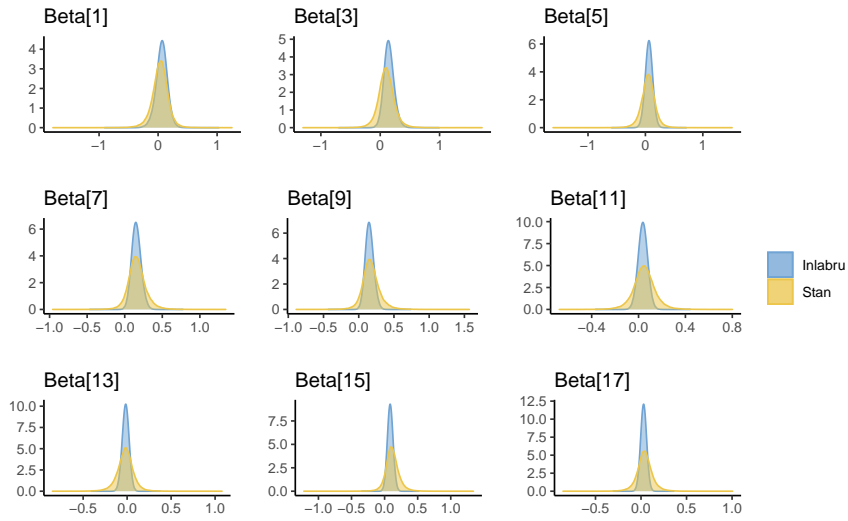
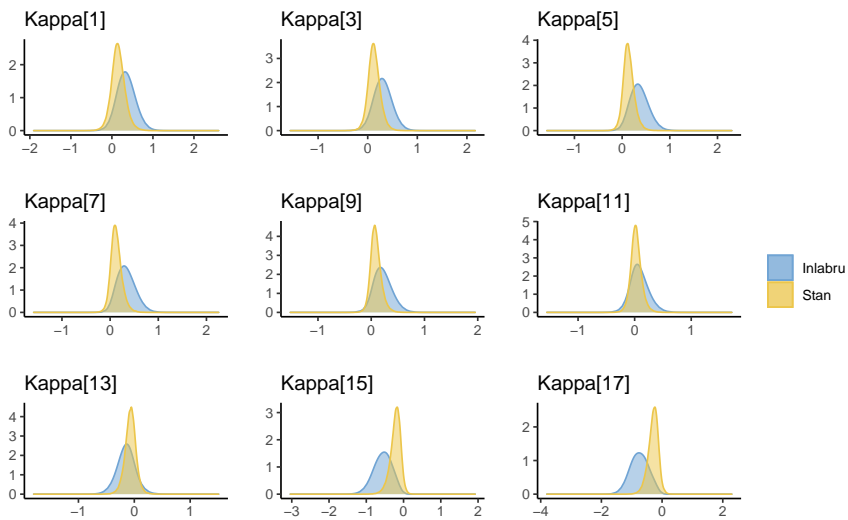


Figure B.2.6: Estimation results from applying the P1 model to the Reduced.1 data using Stan and inlabru

P2: Poisson Model

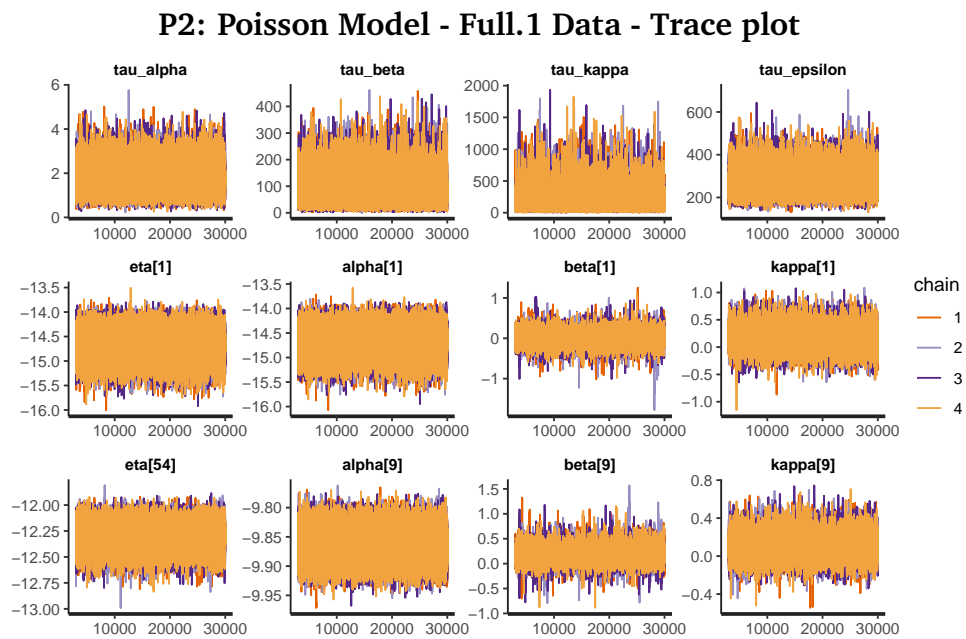
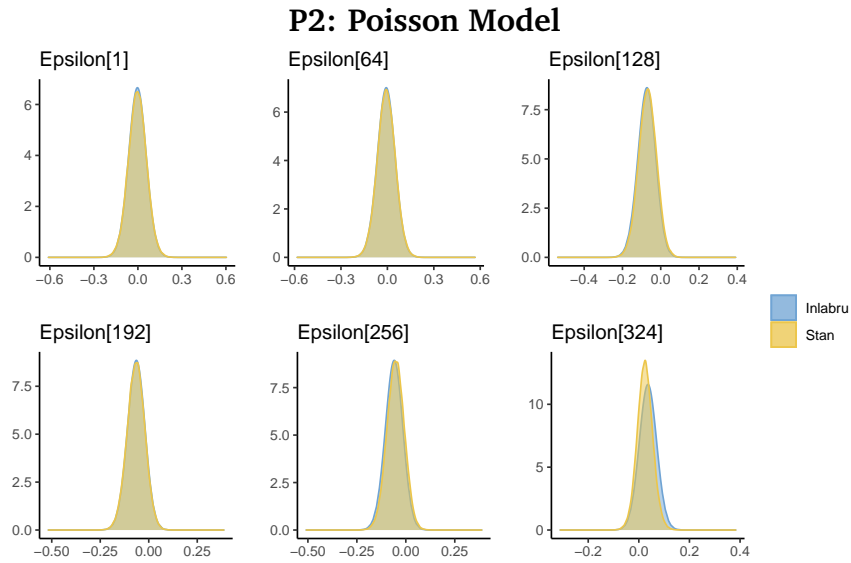


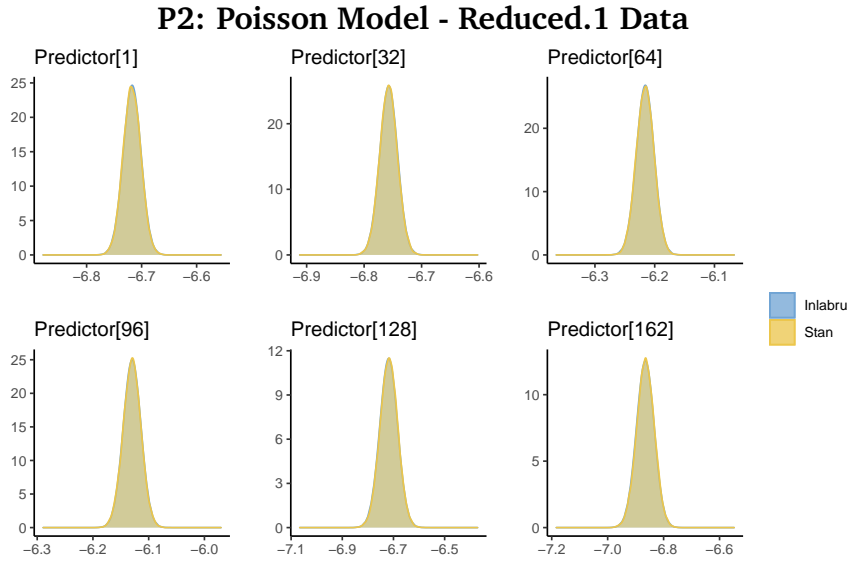
(a) Estimation of selected values of β_x



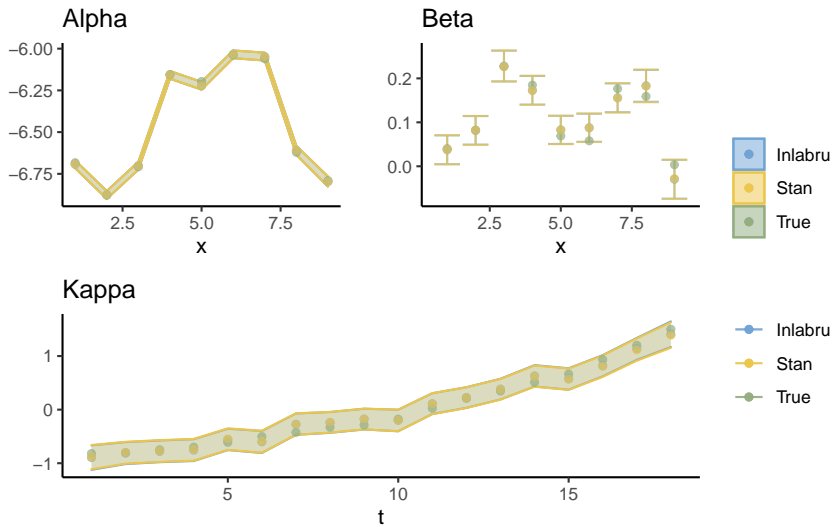
(b) Estimation of selected values of κ_t

Figure B.2.7: Estimation results from applying the P2 model to the Full.1 data using Stan and inlabru





(a) Estimation results of the predictor $\eta_{x,t}$.



(b) Estimation results of random effects α_x , β_x and κ_t

Figure B.2.10: Estimation results from applying the P2 model the the Reduced.1 data using Stan and inlabru

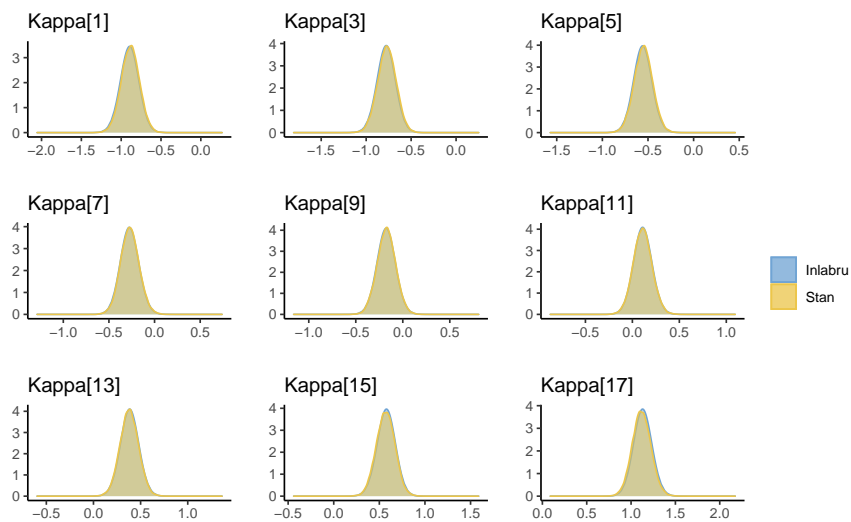
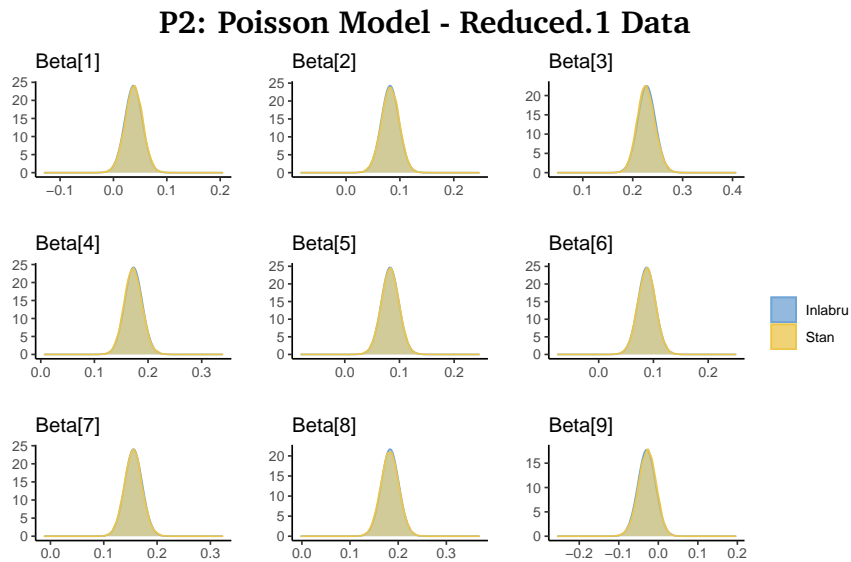
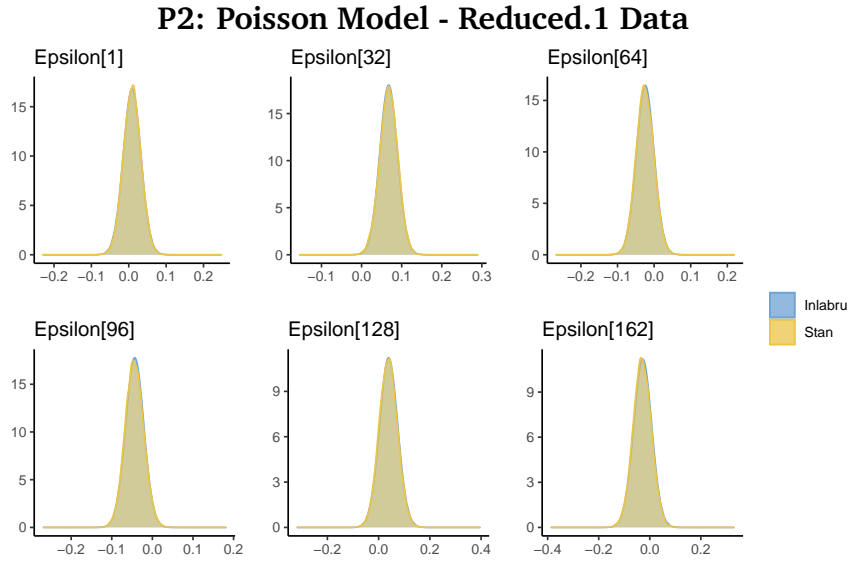
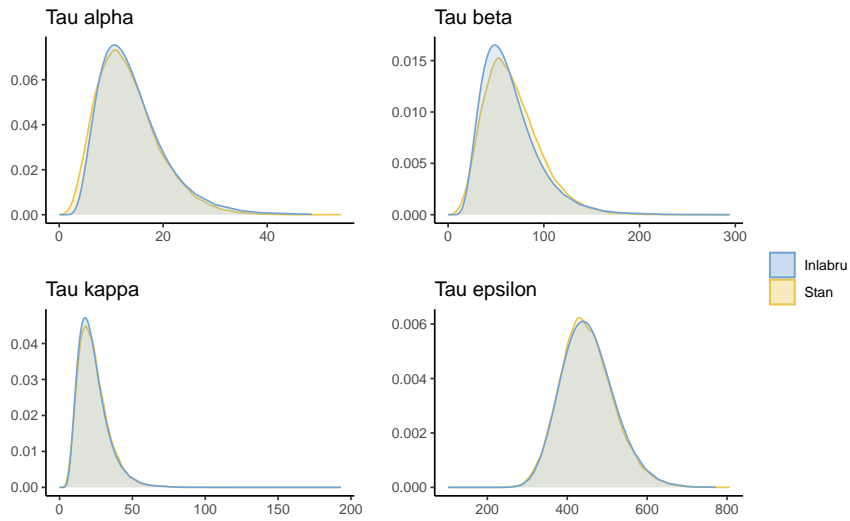


Figure B.2.11: Estimation results from applying the P2 model the the Reduced.1 data using Stan and inlabru



(a) Marginal posterior distributions for selected values of $\epsilon_{x,t}$



(b) Estimated hyperparameters τ_α , τ_β , τ_κ and τ_ϵ

Figure B.2.12: Estimation results from applying the P2 model the the Reduced.1 data using Stan and inlabru

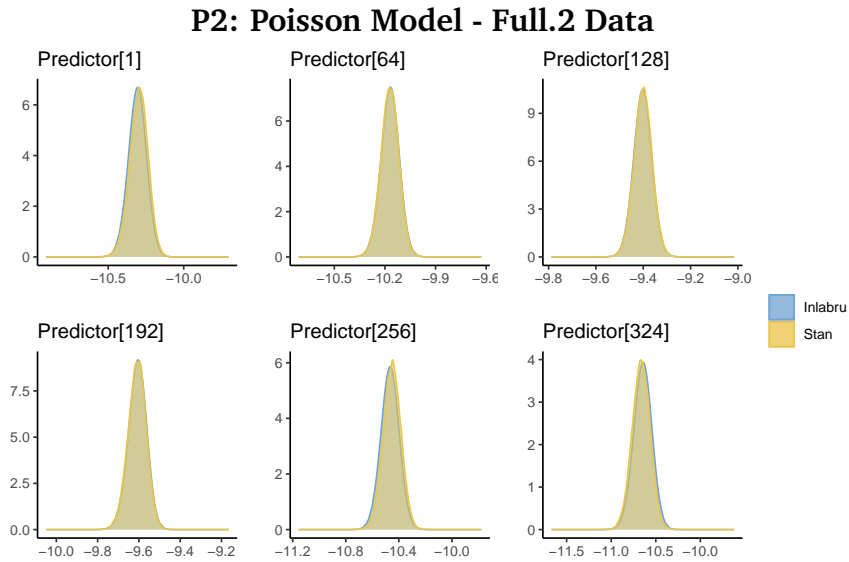
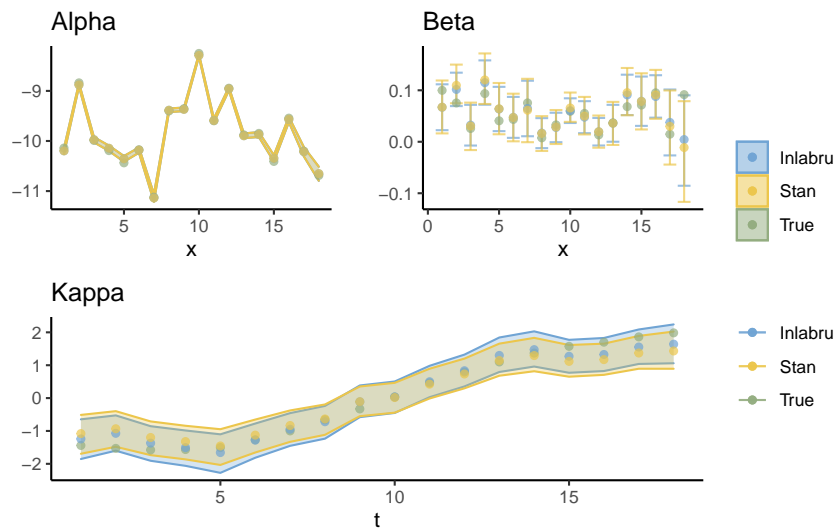
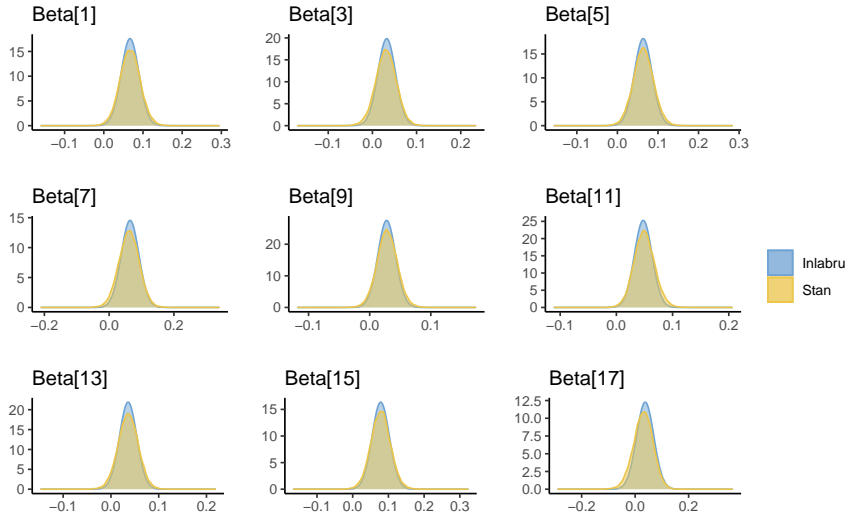
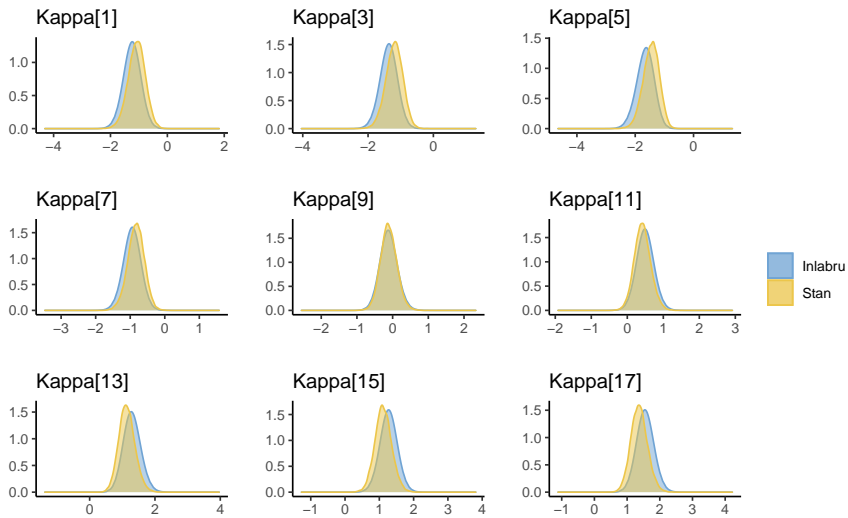
(a) Estimation results of the predictor $\eta_{x,t}$ (b) Estimation results of random effects α_x , β_x and κ_t

Figure B.2.13: Estimation results from applying the P2 model to the Full.2 data using Stan and inlabru

P2: Poisson Model - Full.2 Data

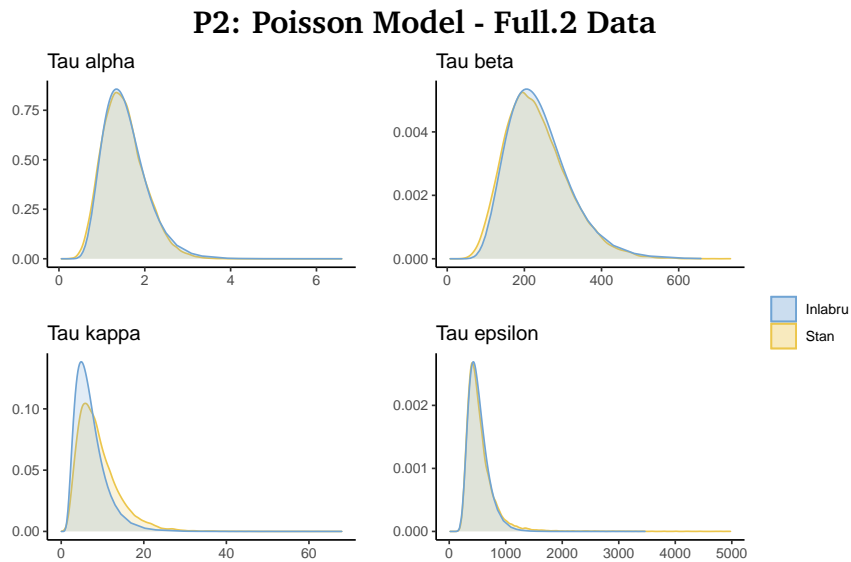


(a) Marginal posterior distributions for selected values of β_x

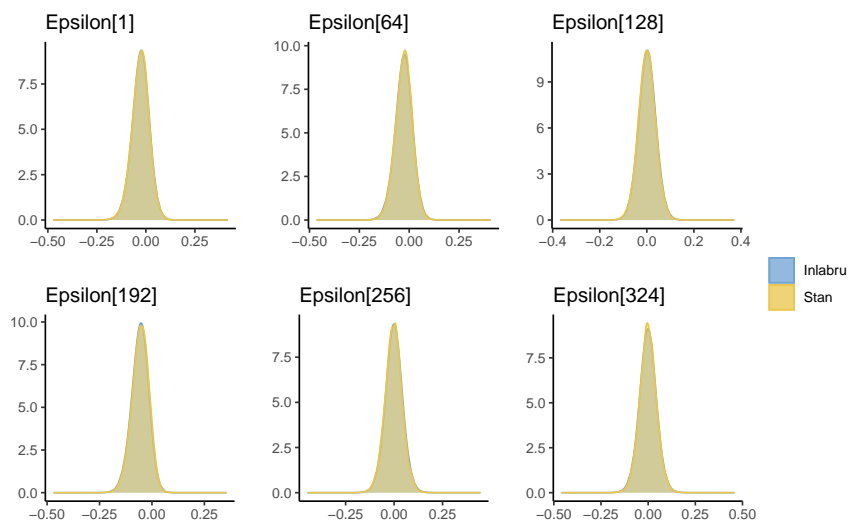


(b) Marginal posterior distributions for selected values of κ_t

Figure B.2.14: Estimation results from applying the P2 model to the Full.2 data using Stan and inlabru



(a) Estimated hyperparameters τ_α , τ_β , τ_κ and τ_ε



(b) Marginal posterior distributions for selected values of $\varepsilon_{x,t}$

Figure B.2.15: Estimation results from applying the P2 model to the Full.2 data using Stan and inlabru

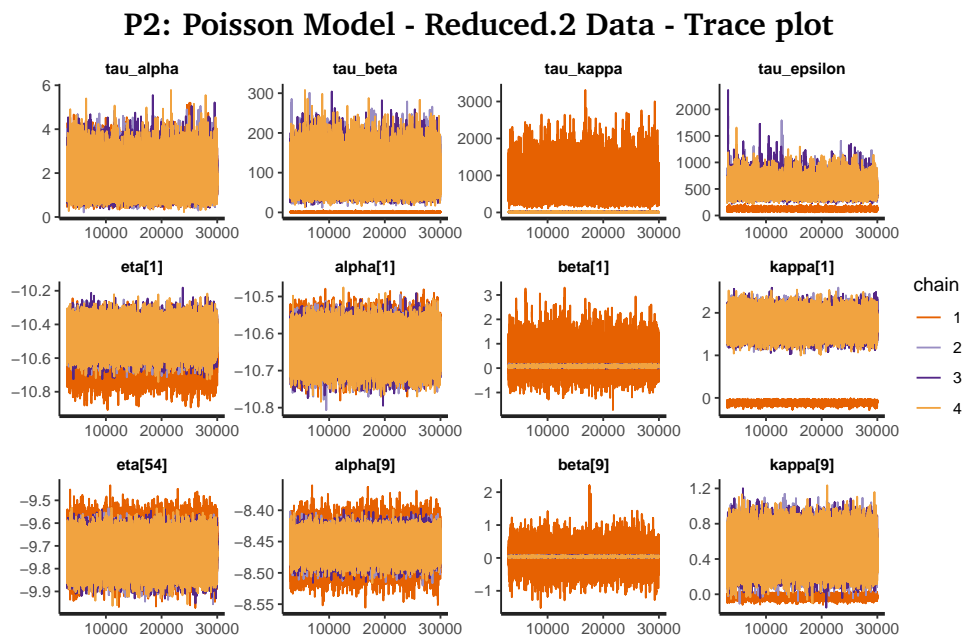
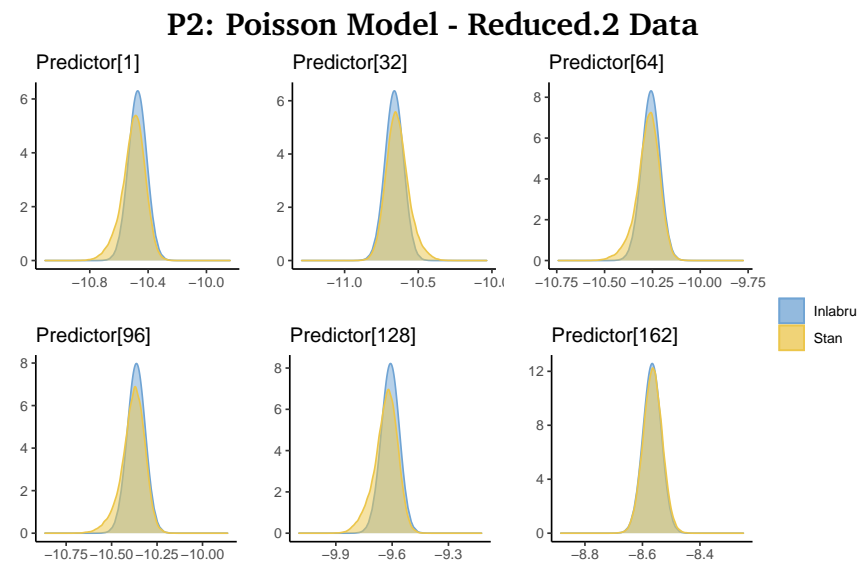
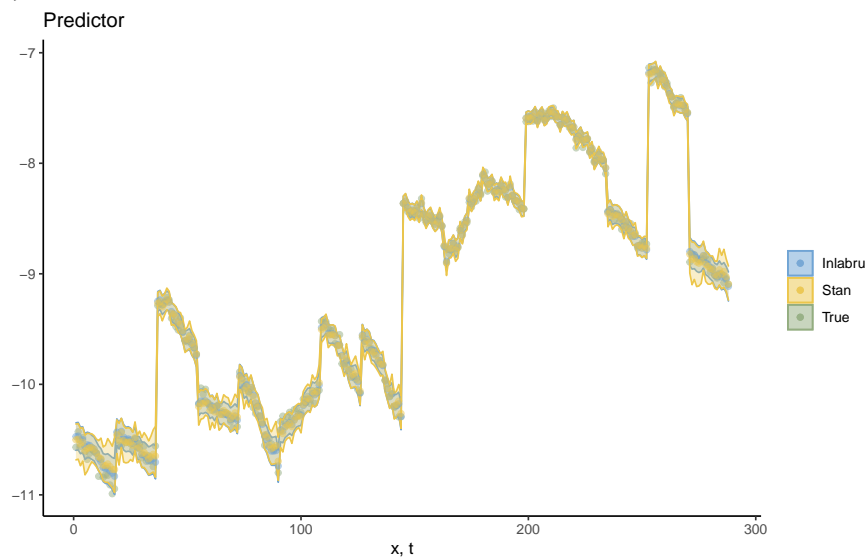


Figure B.2.16: Estimation results from applying Stan and inlabru to model P2

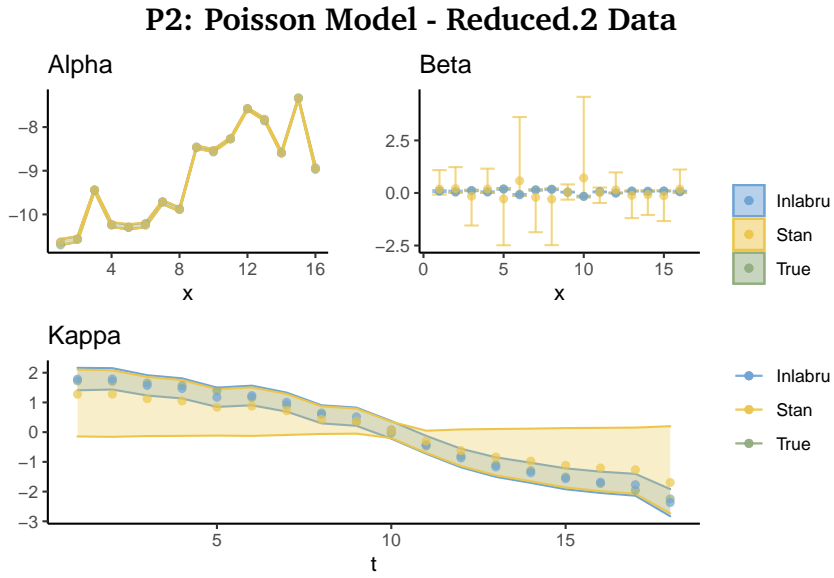


(a) Estimated posterior marginal distributions for some values of the predictor $\eta_{x,t}$.

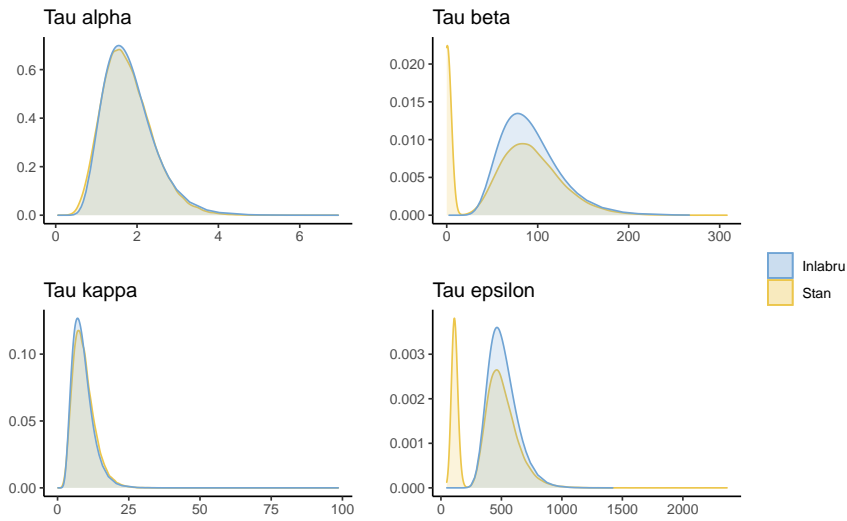


(b) Estimation results of the predictor $\eta_{x,t}$, together with the true value for the predictor

Figure B.2.17: Estimation results from applying the P2 model to the Reduced.2 data, using Stan and inlabru



(a) Estimation results of random effects α_x , β_x and κ_t



(b) Estimated hyperparameters τ_α , τ_β , τ_κ and τ_ϵ

Figure B.2.18: Estimation results from applying the P2 model to the Reduced.2 data using Stan and inlabru

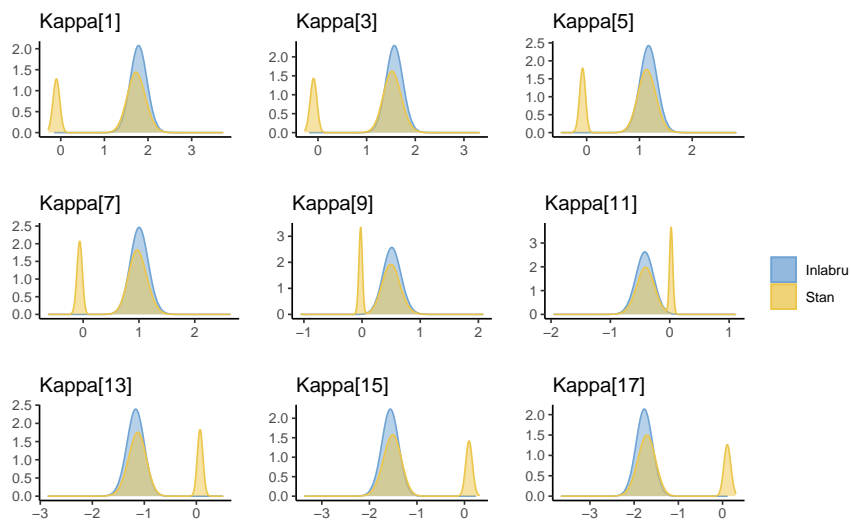
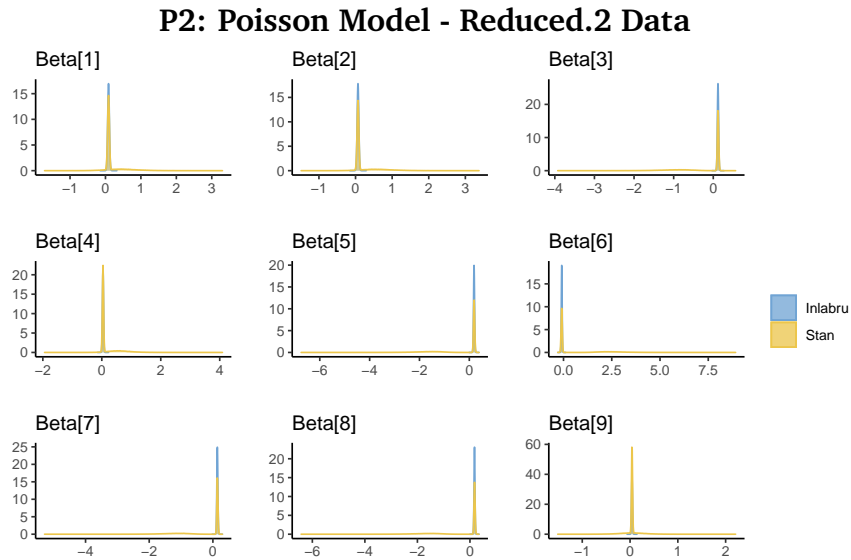


Figure B.2.19: Estimation results from applying the P2 model to the Reduced.2 data using Stan and inlabru

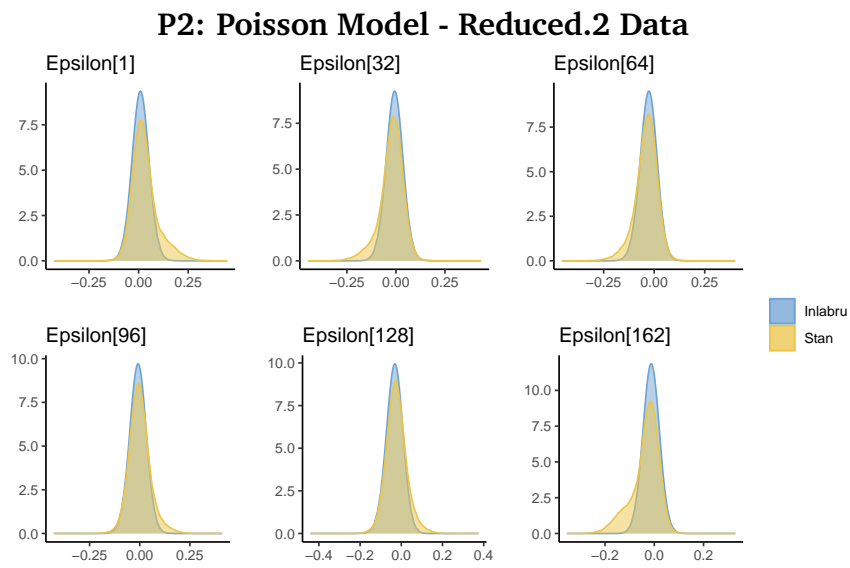


Figure B.2.20: Marginal posterior distributions for selected values of $\epsilon_{x,t}$ from applying the P2 model to the Reduced.2 data using Stan and inlabru

Model	inlabru	Stan	Stan iterations
G1lin	11	32	5000
G2lin	52	33	5000
G1	54	86	10000
G2	66	148	25000
P1lin	17	61	10000
P2lin	24	65	10000
P1	93	202	10000
P2	98	724	50000

Table 8: Run times in seconds for applying the models to the Full.1 data using `inlabru` and Stan, as well as the number of iterations per chain in Stan. All runs are timed using the author’s personal computer

B.3 COMPARISON OF RUN TIMES

Table 8 displays the computational time of running `inlabru` and Stan for each of the models discussed in Chapter 5 when applied to the Full.1 data. The number iterations per HMC chain that we use in the Stan estimations are also presented. Throughout the study, we have used four chains in all Stan programs. We set the number of iterations such that Stan does not report any convergence issues and that the samples seem to give sufficiently smooth densities. We report the full elapsed time for generating the results of `inlabru` and Stan on the author’s personal computer.

C

COMPLEMENTARY FIGURES FROM THE APPLICATION TO GERMAN CANCER DATA

In this chapter, complementary figures for Chapter 6 are presented.

Estimated and Observed Female Lung Cancer Deaths

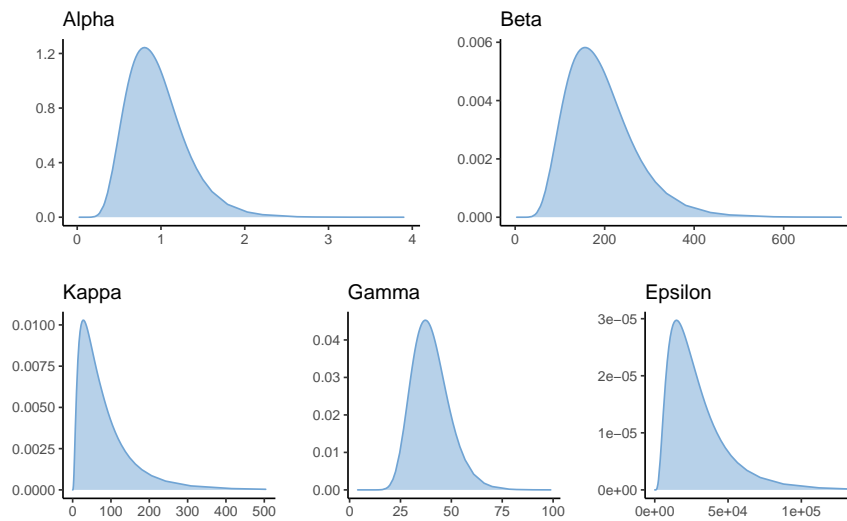


Figure C.1.1: Estimated values for the precisions τ_α (upper right), τ_β (upper right), τ_κ (lower left), τ_γ (lower middle) and τ_ε (lower right) from applying the PLCC model to the full German female lung cancer data using `inlabru`

C.1 APPLICATION TO FULL GERMAN CANCER DATA

In the following section, complementary figures for Section 6.2 are presented.

Figure C.1.1 displays the estimated hyperparameters resulting from applying the PLCC model to the female lung cancer data. Figure C.1.2 presents the estimated hyperparameters from applying the model to male lung cancer data.

Figure C.1.3 presents the number for female stomach cancer deaths, as observed and estimated by `inlabru`. Figures C.1.4 and C.1.5 display the corresponding estimated random effects, error terms and hyperparameters.

Similar results for the male stomach data are presented in Figure C.1.6 (number of deaths), Figure C.1.7 (random effects and error terms) and Figure C.1.8 (hyperparameters).

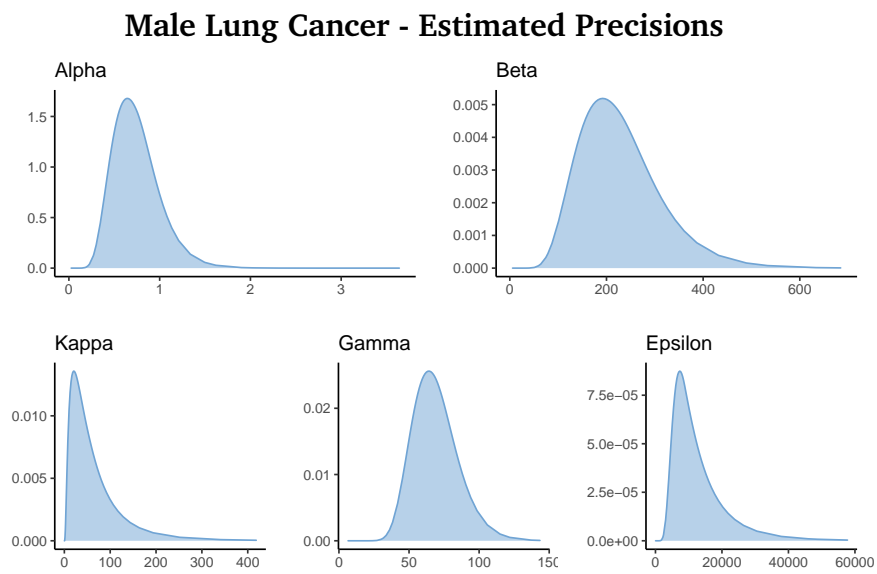
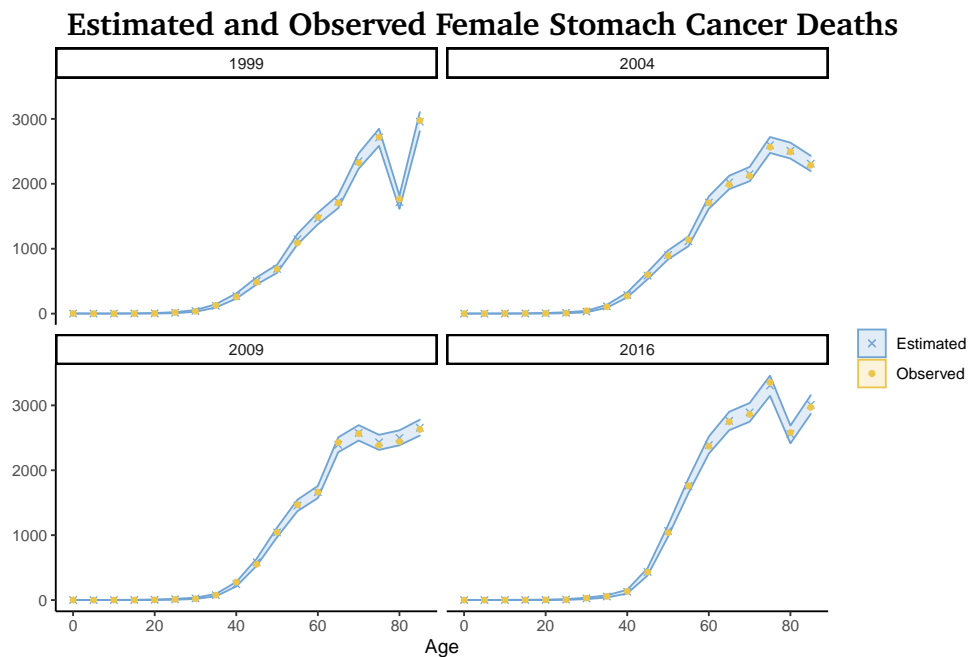
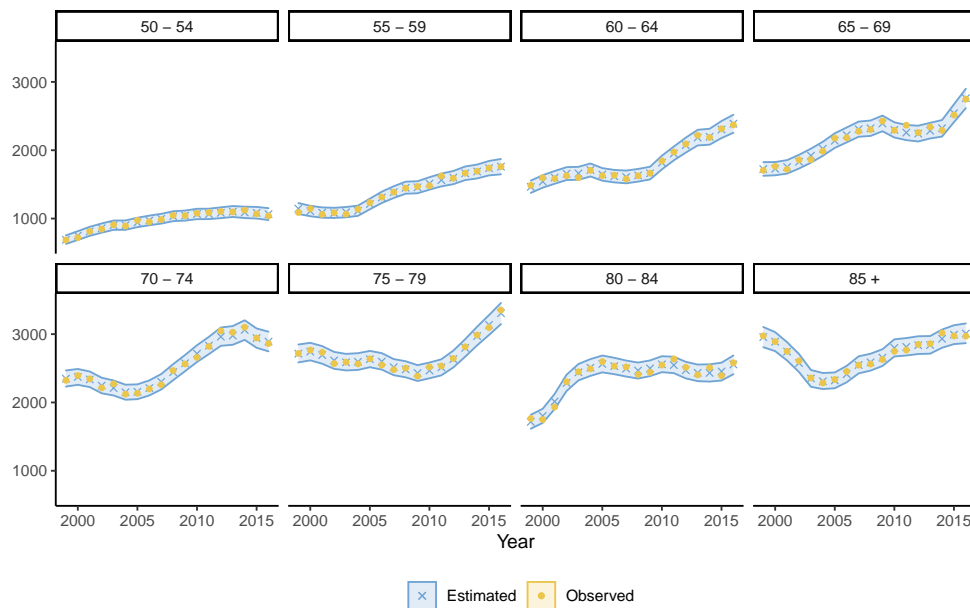


Figure C.1.2: Estimated values for the precisions τ_α , τ_β , τ_κ , τ_γ and τ_ϵ , resulting from applying the PLCC model to the German male cancer data using `inlabru`



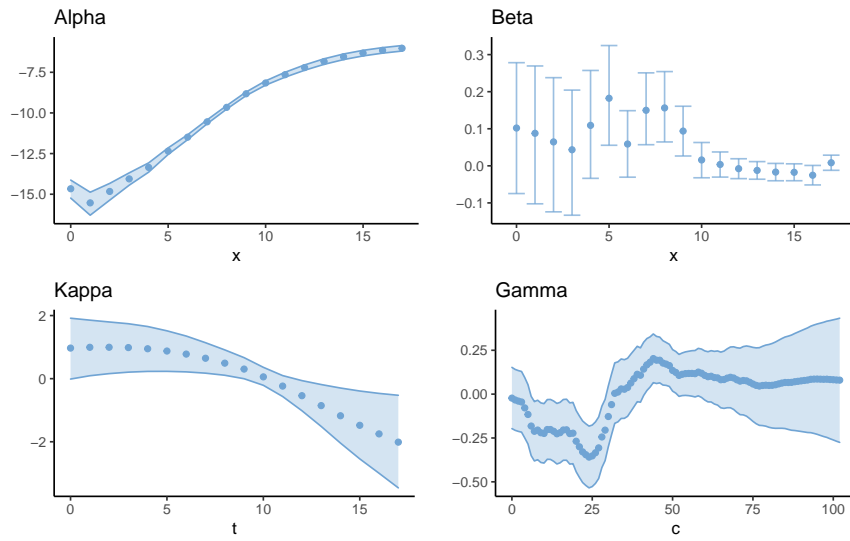
(a) Observed and estimated number of female stomach cancer deaths, for years 1999, 2004, 2009 and 2016



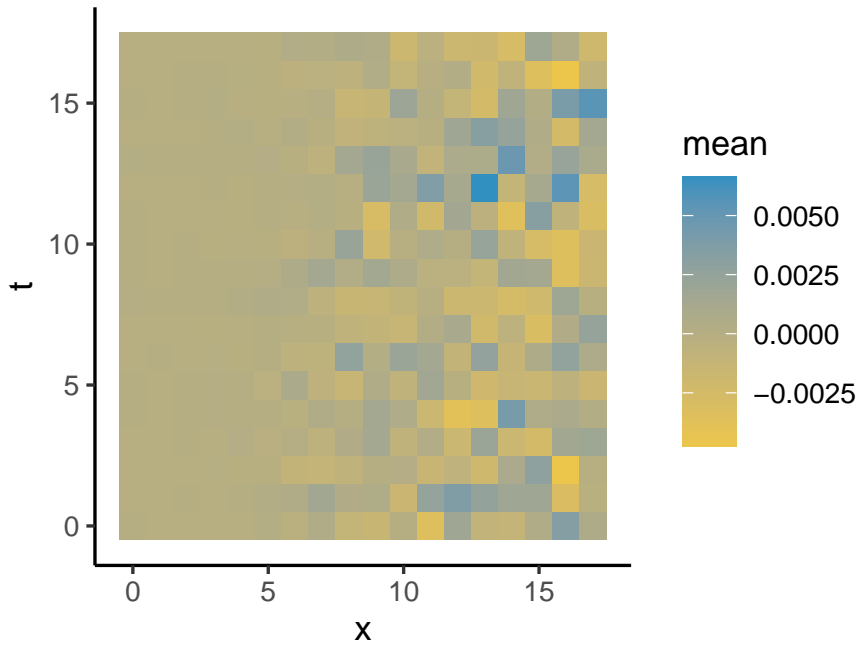
(b) Observed and estimated number of female stomach cancer deaths, for age groups 50-54 and older

Figure C.1.3: Estimated (blue dots) and observed (yellow X-es) female stomach cancer deaths, resulting from applying `inlabru` with the model defined in Equations 6.1.1 through 6.1.3 to the full German female stomach cancer data

Female Stomach Cancer - Estimated Time Effects



(a) Estimated time effects α_x (upper left), β_x (upper right), κ_t (lower left) and γ_c (lower right)



(b) Estimated mean values for the error term $\varepsilon_{x,t}$

Figure C.1.4: Estimation results from applying the PLCC model to the full German female stomach cancer data using `inlabru`

Female Stomach Cancer - Estimated Hyperparameters

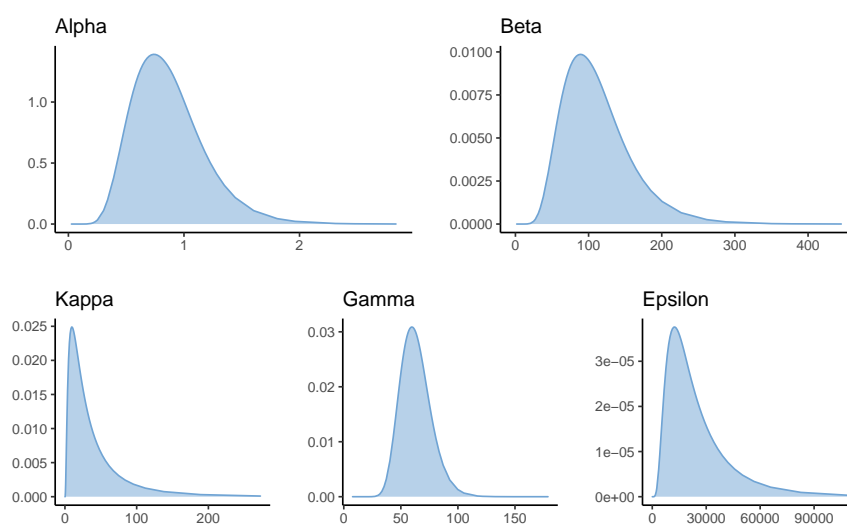
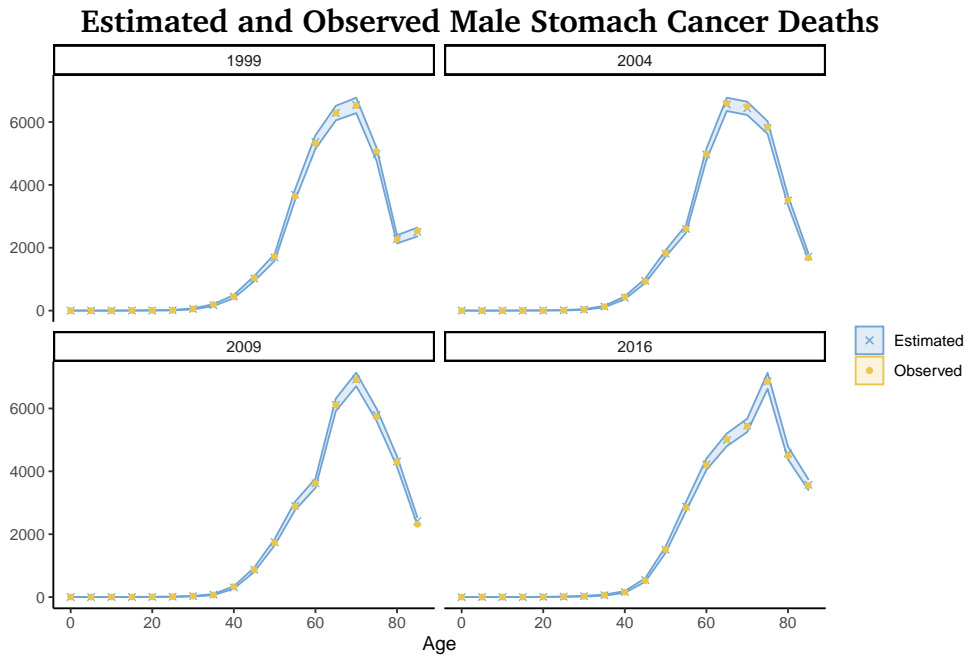
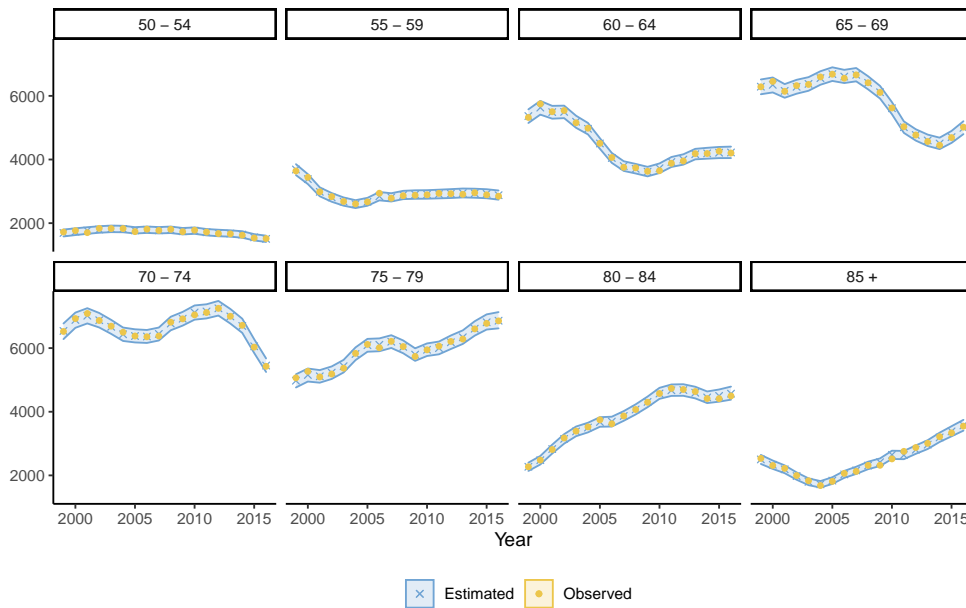


Figure C.1.5: Estimated values for the precisions τ_α , τ_β , τ_κ , τ_γ and τ_ε resulting from applying the PLCC model to the full German female stomach cancer data using `inlabru`

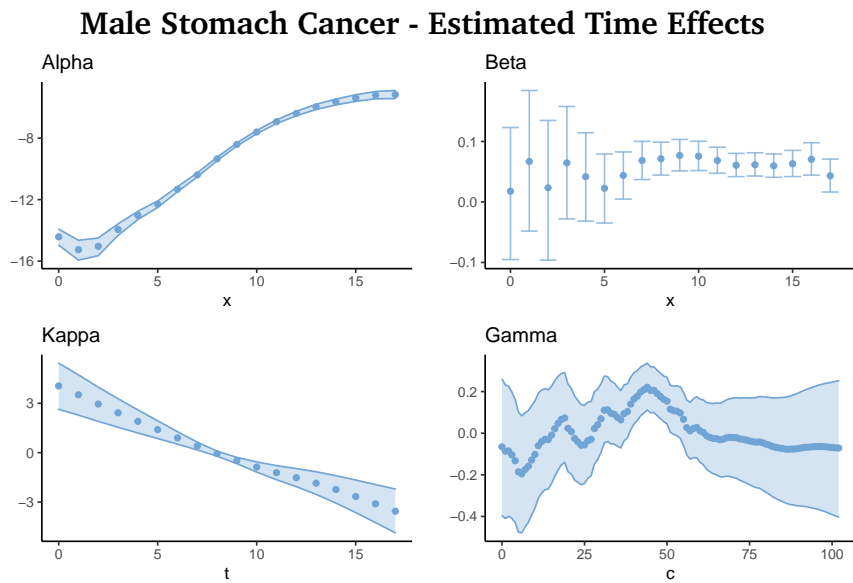


(a) Observed and estimated number of male stomach cancer deaths, for years 1999, 2004, 2009 and 2016

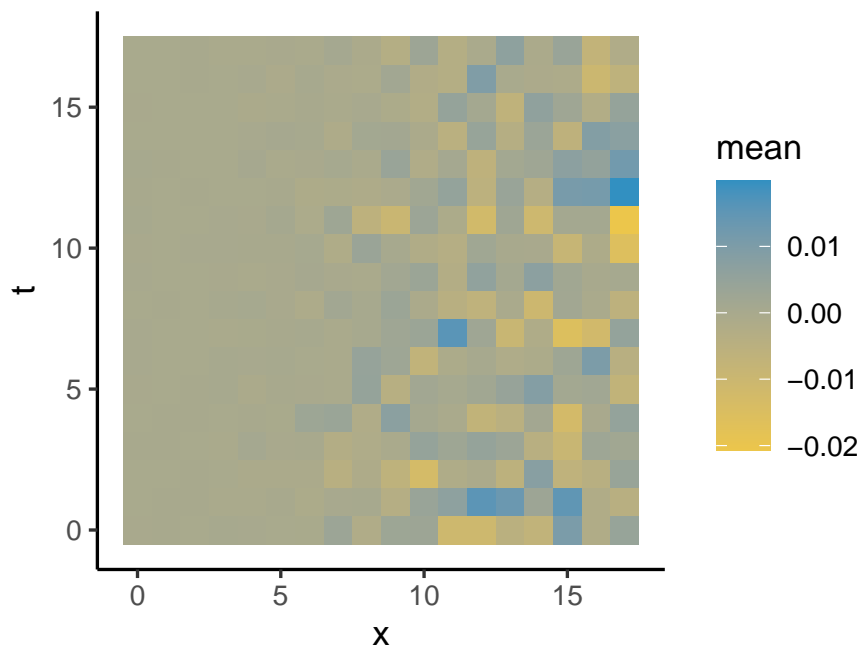


(b) Observed and estimated number of male stomach cancer deaths, for age groups 50-54 and older

Figure C.1.6: Estimated (blue dots) and observed (yellow X-es) male stomach cancer deaths, resulting from applying the PLCC model to the full German male stomach cancer data using inlabru



(a) Estimated time effects α_x , β_x , κ_t and γ_c



(b) Estimated values for the error term $\varepsilon_{x,t}$

Figure C.1.7: Estimation results from applying the PLCC model to the full German male stomach cancer data using `inlabru`

Male Stomach Cancer - Estimated Hyperparameters

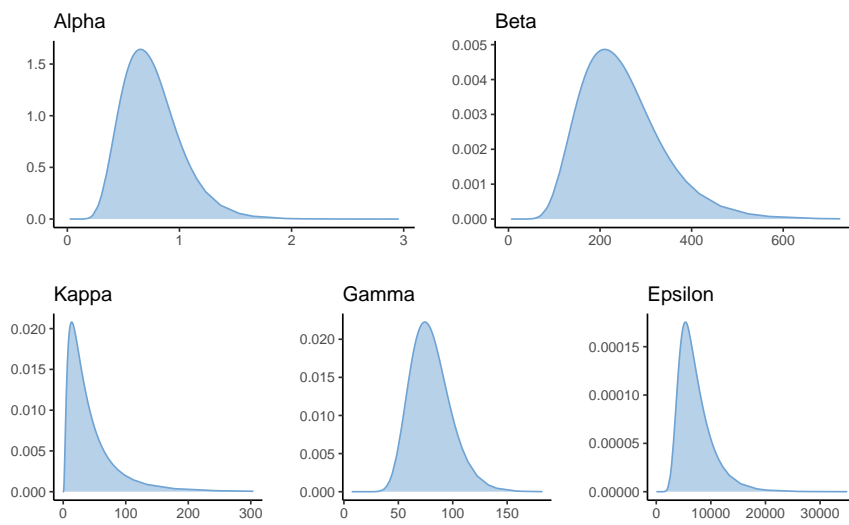
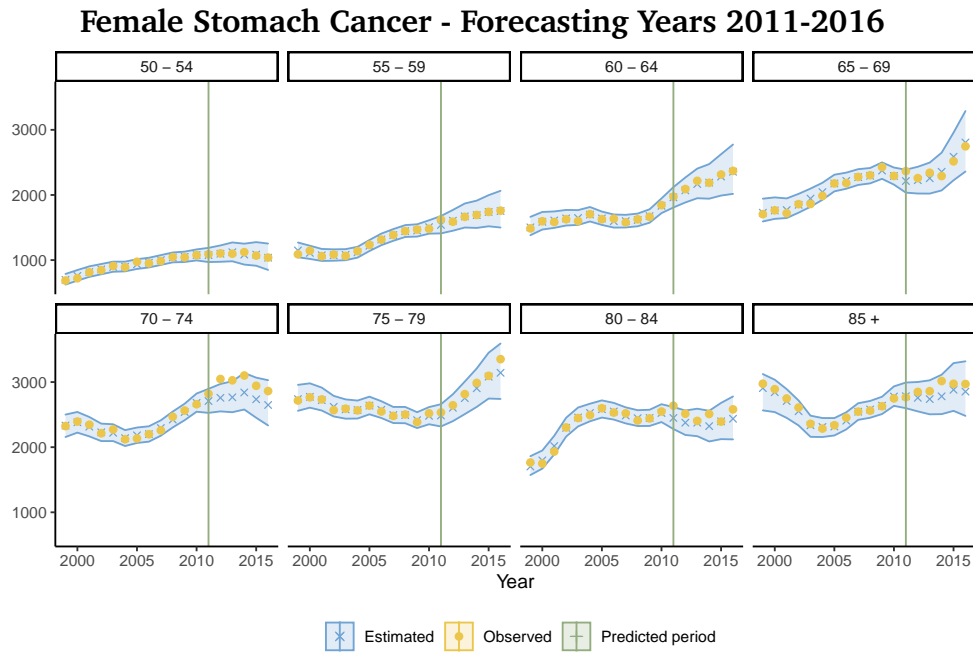


Figure C.1.8: Estimated values for the precisions τ_α , τ_β , τ_κ , τ_γ and τ_ε , resulting from applying the PLCC model to the full German male stomach cancer data using `inlabru`

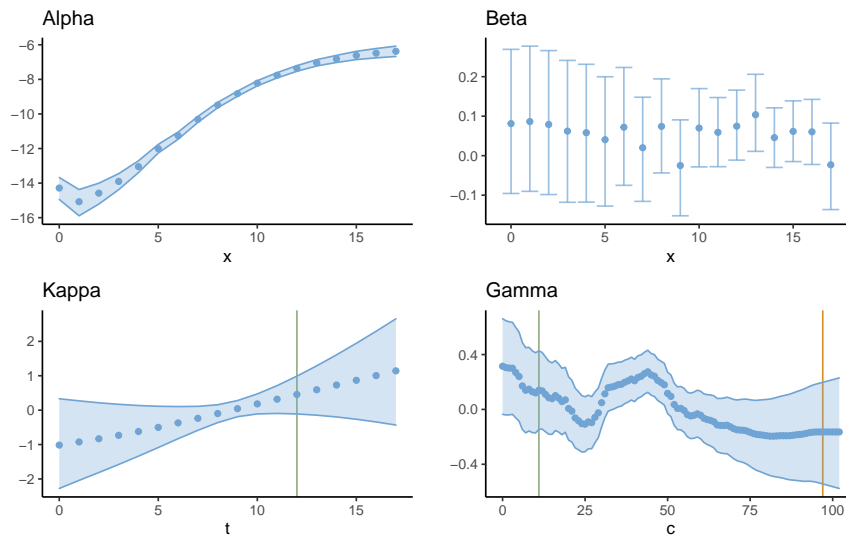
C.2 FORECASTING GERMAN CANCER DATA

In the following section, complementary figures for Section 6.3 are presented.

Figure C.2.1 presents the estimated and forecast number of deaths (Figure C.2.1a) and estimated random effects (Figure C.2.1b) when applying the PLCC model to the female stomach cancer. Figure C.2.2 displays similar plots for the results of forecasting male stomach cancer using the PLCC model.

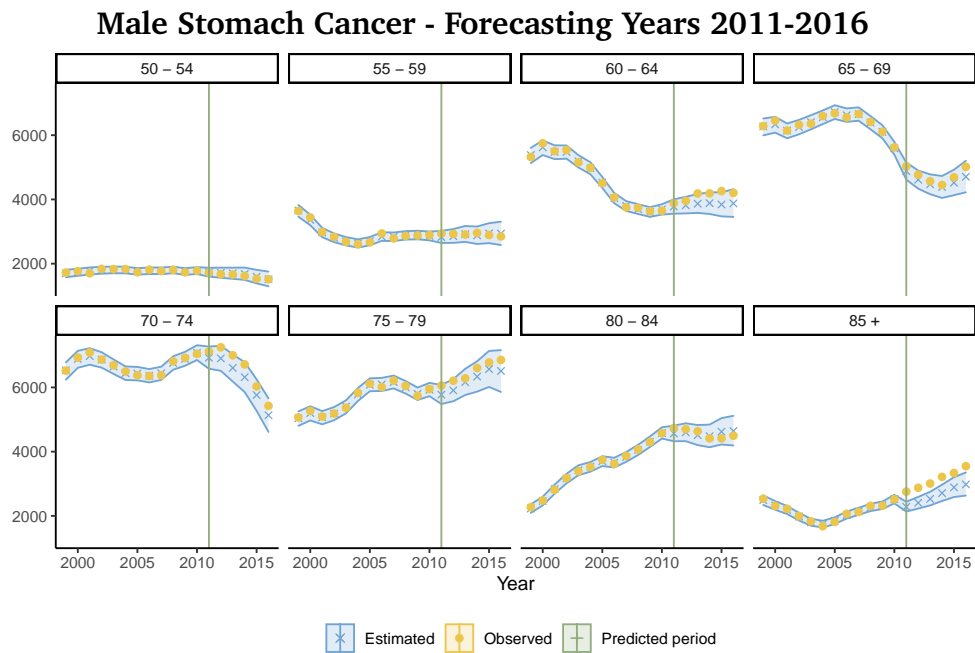


(a) The estimated (blue dots) and the observed (yellow X-es) cancer deaths, for the ages 50-85+ and the period 2011-2016. The green line marks the beginning of the forecast period.

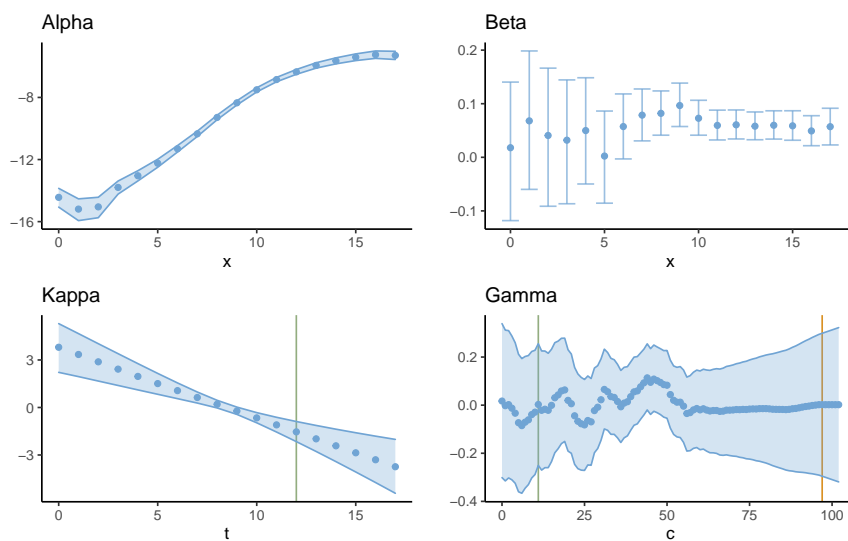


(b) Estimated time effects α_x (upper left), β_x (upper right), κ_t (lower left) and γ_c (lower right). The green line in the lower left plot marks the beginning of the predicted period. The green and orange lines in the lower right plot marks the beginning of the partially and the fully unobserved cohorts

Figure C.2.1: Estimated number of deaths (upper) and time effects (lower) for female stomach cancer, when years 2011-2016 are forecast by the PLCC model



(a) The estimated and the observed cancer deaths, for the ages 50-85+ and the period 2011-2016.



(b) Estimated time effects α_x , β_x , κ_t and γ_c .

Figure C.2.2: Estimated number of deaths and time effects for male stomach cancer, when years 2011-2016 are forecast by the PLCC model. The layout of the plots are as described for Figure C.2.1

