General flexible modelling frameworks for multivariate and multi-state survival outcomes

Alessia Eletti

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Statistical Science

University College London

July 11, 2024

I, Alessia Eletti, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Health data often give rise to complex survival outcomes, which cannot be dealt with using traditional methods without incurring a loss of crucial information. We consider four such cases, motivated by different clinical settings, and present, for each, a general and flexible modelling framework, with the aim of achieving a better understanding of disease patterns and more accurate predictions.

We first focus on diseases which manifest through multiple organs, resulting in dependent time-to-events. We propose a copula-based framework for the joint modelling of bivariate survival outcomes, specified as flexible functions of time and the covariates of interest, with a mixed censoring scheme.

When interest lies in the progression of a disease, multi-state processes represent a powerful modelling approach. For the second case, we consider a continuously observed process, and propose a unified framework that exploits the simplification implied by the exact knowledge of the times-to-events. It combines the flexible specification of each transition, with a simulation-based approach to compute the transition probabilities, posing no limitations on the processes supported.

When constant monitoring of the process is not possible, existing models do not allow the information contained in the intermittently-observed data thus limiting the specifications supported to be fully exploited. The third framework proposed overcomes this challenge by exploiting a novel development, i.e. a closed-form expression for the local curvature information of the transition probability matrix, and supports flexible modelling for virtually any type of process.

Finally, we develop an approach to model two dependent multi-state processes. This is motivated by clinical applications which give rise to two (or more) associated diseases, making the modelling of their joint progression of interest.

The frameworks described are implemented in the `R` packages `GJRM` and `flexmsm` and are exemplified through case studies based on clinical data.

Impact Statement

The present doctoral thesis delivers actionable frameworks to flexibly model complex survival outcomes, with a focus on bivariate and multi-state time-to-events. This is motivated by the increasing interest there is in adequately modelling disease pathways, with the aim of improving the accuracy of predictions and of gaining a better understanding of the data patterns.

Existing literature lacks the generality warranted by the settings explored, both in terms of the methodology as well as of the supporting software. Often only standard settings are supported, thus failing to reflect the features of real-world data (e.g. mixed censoring schemes). General and openly accessible software is rarely provided, making the dissemination and applicability of proposed methods challenging.

To address these issues, each chapter proposes novel statistical methodology, tackling a specific setting. Due to the challenging nature of the problems considered, ample space is left for future developments, some of which are discussed. In addition to this, some of the results presented (e.g. the closed form expression of the second derivative of the transition probability matrix), can also be of use in other research areas, which can thus benefit from the advances described. Further, the model presented in the final chapter sets the foundation for the joint modelling of multi-state survival processes and thus leaves ample space for further extensions, both computational and methodological.

To support the straightforward use of these frameworks by applied users, we provide openly accessible general software for each. The tools are exemplified using real-world data to provide a starting point for the end-user to adapt them to their own setting. In fact, the rise of personalised medicine, with the aim to customise medical decisions and interventions to the individual person, has prompted the development of modelling tools which support the inclusion of information on the individual risk factors, to quantify their impact on the unfolding over time of the event of interest. These tools will aid subject-matter experts in being able to answer questions such as ”what is the expected amount of time a patient will spend in remission” or ”what is the probability that a patient will be treatment-free a given number of times?”.

The impact of the present thesis is, therefore, brought about via different channels. The theoretical developments discussed have been, and will continue to be, communicated through research papers submitted to top statistical journals and will be presented at national and international conferences. This will also provide the basis for collaborations with academics and non-academics. From a practical viewpoint, the tools proposed are implemented in the `R` packages `GJRM` and `flexmsm`, which are thoroughly documented, as per the standards upheld by CRAN. We, then, intend on expanding the exemplification of the packages through a tutorial-based paper that uses a running applied example to communicate how our methods can be used in practice.

Acknowledgements

I have been looking forward to writing the acknowledgments of my doctoral thesis far before I started writing it. My supervisors, my family and the people I was lucky to be surrounded by have made my Ph.D. a memorable experience, and I could not wait to express my gratitude towards them.

Foremost, I would like to express my deepest and most sincere gratitude towards my supervisors Professor Giampiero Marra and Professor Rosalba Radice. Their continuous technical and pastoral support, their patience and kindness, their generosity in sharing with me their time and their (academic) life lessons have been invaluable to my professional and personal journey over the past four years. I could not have asked for better supervisors for my Ph.D., and my hope is that I will be able to give back to them even just a small fraction of what they have given to me.

I would like to thank my parents and my sisters for their emotional support throughout my doctoral degree, for representing a solid harbour I could come back to during its challenging times and for celebrating me during my successes.

To my colleagues and dear friends from the Ph.D. office, thank you for the lunchtime (deep and lighthearted) conversations, for the times we brainstormed solutions on our (new) whiteboards, for reminding me that we all face similar obstacles and that, by sharing them, they become smaller. I am very grateful to have shared this journey with you.

I would like to thank the Department of Statistical Science at University College London for the Teaching Assistantship Studentship that supported me throughout these years and without which I could not have taken on this doctoral degree.

To the faculty and staff members, thank you for sharing advice and help when in need and for caring for and celebrating the Department's research community.

Chapter 3 is dedicated to the late statistician Brian D. Marx, of whom my supervisors attended, as PhD students, the short course "Splines, Knots, and Penalties: The Craft of Smoothing", in Galway. Although at that time they had just started learning about splines, the words used by Brian D. Marx to explain P-splines inspired them and always echoed and tormented their minds until they could appreciate their simplicity and versatility in the context of survival analysis.

The ELSA dataset used in Chapter 4, was made available through the UK Economic and Social Data Service.

Finally, I would like to thank my examiners Prof. Ardo van den Hout and Prof. Andrew Titman for their valuable comments and feedback.

Contents

List of Figures

List of Tables

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

1. **1. For a research manuscript that has already been published** (if not yet published, please skip to section 2)**:**

   (a) **What is the title of the manuscript?** Copula link-based additive models for bivariate time-to-event outcomes with general censoring scheme.

   (b) **Please include a link to or doi for the work:** https://doi.org/10.1016/j.csda.2022.107550

   (c) **Where was the work published?** Computational Statistics and Data Analysis.

   (d) **Who published the work?** Elsevier.

   (e) **When was the work published?** June 2022.

   (f) **List the manuscript's authors in the order they appear on the publication:** Danilo Petti, Alessia Eletti, Giampiero Marra, Rosalba Radice.

   (g) **Was the work peer reviewd?** Yes.

   (h) **Have you retained the copyright?** Yes.

   (i) **Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi** No.

   If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

   ☒ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)**:**

   (a) **What is the current title of the manuscript?** NA.

   (b) **Has the manuscript been uploaded to a preprint server 'e.g.**

medRxiv'?

If 'Yes', please please give a link or doi: Answer here: NA.

(c) **Where is the work intended to be published?** NA.

(d) **List the manuscript's authors in the intended authorship order:** NA.

(e) **Stage of publication:** NA.

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)**:** Danilo Petti and Alessia Eletti implemented the model, integrated it in the existing `R` package `GJRM`, contributed to the writing of the paper and carried out the preliminary data analysis. Giampiero Marra and Rosalba Radice supervised the work and contributed to the writing of the paper and of the data analysis.

4. **In which chapter(s) of your thesis can this material be found?** Chapter 2.

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)**:**

Candidate:

Date:

Supervisor/Senior Author signature (where appropriate)**:**

Date:

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

1. **1. For a research manuscript that has already been published** (if not yet published, please skip to section 2)**:**

   (a) **What is the title of the manuscript?** A spline-based framework for the flexible modelling of continuously observed multistate survival processes.

   (b) **Please include a link to or doi for the work:** https://doi.org/10.1177/1471082X231176120

   (c) **Where was the work published?** Statistical Modelling.

   (d) **Who published the work?** Sage Journals.

   (e) **When was the work published?** October 2023.

   (f) **List the manuscript's authors in the order they appear on the publication:** Alessia Eletti, Giampiero Marra, Rosalba Radice.

   (g) **Was the work peer reviewd?** Yes.

   (h) **Have you retained the copyright?** Yes.

   (i) **Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi** No.

   If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

   ☒ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)**:**

   (a) **What is the current title of the manuscript?** NA.

   (b) **Has the manuscript been uploaded to a preprint server 'e.g.**

medRxiv'?

If 'Yes', please please give a link or doi: NA.

(c) **Where is the work intended to be published?** NA.

(d) **List the manuscript's authors in the intended authorship order:** NA.

(e) **Stage of publication:** NA.

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)**:** Alessia Eletti wrote the code used for the case study presented in the work (this code is publicly available in a Github repository references in the work), wrote the paper and carried out the data analysis. Giampiero Marra and Rosalba Radice supervised the work and contributed to the writing of the paper.
4. **In which chapter(s) of your thesis can this material be found?** Chapter 3.

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)**:**

Candidate:
Date:

Supervisor/Senior Author signature (where appropriate)**:**
Date:

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

1. **1. For a research manuscript that has already been published** (if not yet published, please skip to section 2)**:**

   (a) **What is the title of the manuscript?** NA.

   (b) **Please include a link to or doi for the work:** NA.

   (c) **Where was the work published?** NA.

   (d) **Who published the work?** NA.

   (e) **When was the work published?** NA.

   (f) **List the manuscript's authors in the order they appear on the publication:** NA.

   (g) **Was the work peer reviewd?** NA.

   (h) **Have you retained the copyright?** NA.

   (i) **Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? If 'Yes', please give a link or doi**
   If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

   □ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)**:**

   (a) **What is the current title of the manuscript?** A General Estimation Framework for Multi-State Markov Processes with Flexible Specification of the Transition Intensities.

   (b) **Has the manuscript been uploaded to a preprint server 'e.g. medRxiv'?**

If 'Yes', please please give a link or doi:
https://doi.org/10.48550/arXiv.2312.05345

(c) **Where is the work intended to be published?** Journal of the Royal Statistical Society Series B: Statistical Methodology.

(d) **List the manuscript's authors in the intended authorship order:** Alessia Eletti, Giampiero Marra, Rosalba Radice

(e) **Stage of publication:** Under review.

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)**:** Alessia Eletti derived the novel result presented in the work, developed the `R` package `flexmsm`, contributed to the wroting of the paper and carried out the data analyses described in the two case studies. Giampiero Marra and Rosalba Radice supervised the work and contributed to the writing of the paper.

4. **In which chapter(s) of your thesis can this material be found?** Chapter 4.

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)**:**

Candidate:
Date:

Supervisor/Senior Author signature (where appropriate)**:**
Date:

Chapter 1

Introduction

Some key questions arising in fields as diverse as medicine, biology, public health, epidemiology, engineering, economics and demography can be conveniently formulated in terms of survival problems (Klein & Moeschberger, 2006). This has the twofold advantage of leading to an intuitive interpretation, centred around the time to occurrence of the event of interest, while allowing for a methodologically sound handling of the units for which the event itself is not observed, a common feature of real-world data, obtained through the notion of censoring.

While the statistical tools we will present are applicable to all these disciplines, our focus is on clinical applications. Here interest lies, for example, in assessing the time to death from a certain cause, the duration of response to treatment, the time to recurrence or development of a disease. The expansion of health registry data has implied an increasing interest towards combinations of two or more of the above, leading to complex survival outcomes, along with a heightened focus towards the proper handling of censored events and event times, often occurring in registry data.

The rise of personalised medicine, with the aim to tailor medical decisions and interventions to the individual person, has prompted the development of flexible models, which can capture multifaceted patterns within the data, by supporting the exploration of a variety of time and covariate effects. In this way, one can obtain more accurate predictions, which adequately take into account the diversity of patient features, hence better characterising disease evolution.

The aim of this thesis is to present general tools for the flexible modelling of complex survival outcomes, with a focus on bivariate and multi-state time-to-events, expanding the possibilities currently allowed in the literature, both in terms of methodology and of supporting software. This work has lead to the publication of two papers, with a third currently under review, the expansion of the `R` package `GJRM` (Marra & Radice, 2024) and the development of the `R` package `flexmsm` (Eletti et al., 2023a). The thesis is structured as a collection of articles and, as such, each chapter is self-contained in regard to its notation and in the layout of the setting. We are aware that this choice comes at the cost of some redundancies, however it benefits of better clarity. The chapters that use published material are duly noted in the *UCL Research Paper Declaration* forms included above.

Chapter 2 is based on Petti et al. (2022) and explores the setting where disease manifests through multiple organs, resulting in dependent time-to-events. Interest lies in accounting for their dependent nature, since not doing so would lead to biased estimates, while modelling the survival outcomes in a way that retains the interpretability on the individual organ level and that allows quantification of the strength of the dependence. To achieve this, we propose a copula-based framework for the joint modelling of bivariate survival outcomes, specified as flexible functions of time and covariates through a link-based additive model. The copula parameter is also specified as a flexible function of covariates, thus allowing investigation of the impact of patient characteristics on the strength of association between the disease manifestations in the two organs. Importantly, to reflect the nature of real health registry data, the framework supports mixed censoring, i.e. each survival outcome can take on a left-, right- and/or interval-censoring scheme. The method developed has been incorporated in the `R` package `GJRM` and is illustrated using data from the *Age-Related Eye Disease Study*. The analysis aims to quantify the effect of clinical risk factors on the joint risks of Age-related Macular Degeneration progression as well as to predict the progression profiles of patients with different characteristics. An extensive simulation study provides evidence on the empirical effectiveness of the proposed approach in recovering true covariate effects and baseline functions.

When interest lies in the progression of a disease rather than on a single outcome, multi-state processes provide a powerful modelling approach. Each manifestation of the disease is, in fact, represented by a ”state” and the unfolding over time of this chain of events is captured by a collection of time-dependent intensity functions, each associated with a ”transition” between a pair of states. If the time-to-events are known exactly, the estimation of the multi-state model can be broken down into that of a set of traditional survival models, one for each transition. Chapter 3, which is based on Eletti et al. (2023b), focuses on this case and provides a unified framework which combines a flexible link-based additive modelling approach for the specification of each transition intensity, supported in practice by the `R` package `GJRM`, with a general simulation-based method for the computation of the predicted transition probabilities, implemented in the `R` package `mstate`. The transition probabilities are key quantities for the interpretation of multi-state processes, since they provide an intuitive way to quantify the unfolding over time of the disease in terms of the probability of observing the process at a specific stage in a given time, conditional on a chosen starting point. Crucially, modelling takes place on the scale of the survival function, thus providing the quantities needed for the computation of the transition probabilities without the need for further intermediate steps, since the simulation-based approach requires transition-specific cumulative intensity estimates. Care was needed here to ensure the monotonicity of the transition-specific survival functions, which is elegantly embedded in the model design matrix. These choices ensure the seamless integration between the modelling and prediction as well as the overall computational efficiency of the framework. We exemplify its usage through a case study on breast cancer patients from the *Rotterdam Breast Cancer Study*, where we explored the effects of risk factors, such as progesterone level and the number of positive nodes, in a more general and flexible manner than previously possible in the literature.

Chapter 4, based on a submitted paper, addresses the case where constant monitoring of the multi-state process is infeasible, giving rise to various forms of censoring of the time-to-events and/or of the states occupied. This lack of knowledge makes the setting methodologically and practically challenging. Existing models do not allow full exploitation of the information contained in the intermittently-observed data, since they rely on the seminal paper by Kalbfleisch & Lawless (1985), which provides closed-form expressions for up to first order information only. In practice, this is insufficient to support the complexity of the setting, particularly given the degree of flexibility desired for the transition intensity models, which determine how the process unfolds and thus represent the core of the multi-state model. For this reason, existing literature is characterised either by basic parametric forms for the transition intensities or by works that propose flexible models, but only for simple process structures. We provide a closed-form expression for the local curvature information of the transition probability matrix. Such novel development allows one to model any type of process while supporting flexible time and covariate effects on the transition intensities, which are specified by means of spline-based additive predictors. The methodology is implemented in the `R` package `flexmsm` and is exemplified via two case studies. The first focuses on the postoperative recovery of heart transplant recipients, where the possible outcomes are remaining in a healthy state, onset of Cardiac Allograft Vasculopathy, i.e. a disease of the arterial walls, and death. The second study focuses on cognitive decline in a population of elder individuals who took part in the *English Longitudinal Study of Ageing*. Cognitive aptitude is measured using a memory-based test and each score is represented by a state in a five-state process, with forward and backward transitions, to reflect the patterns of cognitive decline and improvement observable in the data. In both case studies, our framework allows for model specifications and process structures which were not supported by the former state-of-the-art, thus leading to novel insight compared to existing analyses.

Chapter 5 builds on the previous developments and sets the foundation for the modelling of multiple dependent multi-state survival processes. This is motivated by cases in which a disease affects paired organ systems or where multiple disease manifestations stem from a common underlying condition. For example, in ophthalmology, damage caused by diabetic retinopathy - a progressive eye disease in diabetic patients - can occur in either or both eyes and the disease course in one eye is expected to be linked with that in the other. To gain a proper understanding of the disease mechanism, it is necessary to model simultaneously the temporal patterns of disease progression of both eyes and assess the influence of risk factors. In this chapter we propose to capture the dependence structure tying two multi-state processes through a copula-based model, which allows us to retain the interpretability of the marginal processes, while modelling each process by means of the framework proposed in Chapter 4. This work is at an early stage and further developments are currently under way, however what we propose is already more general compared to the current state-of-the-art. We exemplify our approach through a toy example based on simulated data.

In the above frameworks, estimation relies on an approach proposed in Marra & Radice (2020), carefully adapted here to each setting. This combines a computationally efficient and stable penalised maximum likelihood optimisation algorithm, with an integrated automatic multiple smoothing parameter selection algorithm. Tests carried out with alternative standard approaches have proven that these are insufficient to support the complexity of the models considered. In contrast, the method proposed makes an adequate use of the information contained in the data, thus ensuring that the modelling potential is fully exploited in practice.

Chapter 6 provides a general discussion and outlines some avenues for further research, some of which are currently under investigation.

Chapter 2

Copula Link-Based Additive Models for Bivariate Time-to-Event Outcomes with General Censoring Scheme

2.1 Introduction

Bivariate survival outcomes arise frequently in many research areas such as health and epidemiology. For example, bivariate survival data are often used in clinical trials studying diseases concerning paired organs, where the outcomes of interest are measured on the same individual. The main feature of survival data is censoring. For instance, bivariate interval censoring occurs when the events are not precisely observed due to intermittent assessment times and are indeed only known to belong to intervals. When individuals do not experience the two events at their last assessment times, the event statuses are undefined (bivariate right censoring). If some individuals have already experienced both events at the times they enter the study then the data are bivariate left-censored. Sometimes various types of censoring arise simultaneously. This would be the case when, e.g., a disease occurs in one of the paired organs between two consecutive visits and the condition does not occur in the other organ by the end of the study. The aim of this paper is to introduce a flexible regression modelling framework that can handle bivariate survival data under any censoring mechanism.

Several approaches for modelling bivariate censored data have been proposed. The literature is vast and here we mention a handful of works. Some of them are based on the frailty technique (e.g., Chen et al., 2009, 2014; Martins et al., 2019; Wen & Chen, 2013; Wang et al., 2015; Zhou et al., 2017; Zeng et al., 2017). Others, based on copulae and hence more relevant to this paper, are Barthel et al. (2018), Cook & Tolusso (2009), Hu et al. (2017), Kwon et al. (2021), Lo et al. (2020), Marra & Radice (2020), Romeo et al. (2018), Sujica & Van Keilegom (2018), Sun & Ding (2021a) and Wang et al. (2008). These works are not as general and versatile as our proposal. In fact, our modelling framework allows for: a) any bivariate combination of censoring types, whether left-, right-, interval-, or non-censored; b) the exploration of a wide array of dependence structures via copulae; c) all model parameters to be specified as functions of flexible covariate effects via the penalised regression spline methodology (e.g., Wood, 2017); d) the margins of the copula to be modelled via transformations of the survival functions, which give rise to link-based models with the proportional hazards and odds models being particular cases (e.g., Liu et al., 2018); e) the baseline survival functions to be modelled by means of monotonic P-splines which are theoretically advantageous and computationally tractable (e.g., Pya & Wood, 2015). Prior to this work, there were no such models (and related fitting procedures) available in the literature nor software implementations.

Despite the complexity of the proposed model, in that it allows for many layers of structure, there is no price to pay in terms of usability and interpretability. In fact, the model has been incorporated in the newly-revised software package `GJRM` (Marra & Radice, 2024), written for the programming language `R` (R Development Core Team, 2022), which significantly eases the use of the framework. An additional benefit is that post estimation functions have been extended and integrated within `GJRM` to allow any user to produce interpretable results. Parameter estimation relies on an extension of the stable and fast algorithm presented in Marra & Radice (2020) which is based on a simultaneous penalised maximum likelihood approach with integrated automatic multiple smoothing parameter selection. The proposed model together with fast and reliable software implementation represents a significant advance in modelling bivariate survival data. An interesting feature of the proposal is that it is very flexible and at the same time parametric. Sir David R. Cox, among others, has encouraged the broader use of parametric models for empirical modelling (e.g., Reid, 1994). In that spirit, our modelling framework enables a large amount of exploration via many and diverse functional structures which may help to uncover new patterns and trends in the data.

The potential of the approach is illustrated via a simulation study as well as using data from the Age-Related Eye Disease Study (AREDS), a multi-center randomised clinical trial exploring the development and progression of age-related macular degeneration (AMD), sponsored by the National Eye Institute (Group, 1999). The analysis aims to quantify the effect of clinical risk factors on the joint risks of AMD progression as well as to predict the progression profiles of AMD patients with different characteristics.

This chapter is organised as follows. Section 2.2 discusses various details of the proposed model. Section 2.3 introduces the model log-likelihood and explains how to perform parameter estimation, whereas Section 2.4 shows some inferential results. In Section 2.5, data from the AREDS are analysed and the main findings presented. Section 2.6 concludes the paper with a discussion. Supplementary Material A provides more details on the log-likelihood construction, reports the analytical expressions for the score and Hessian matrix, discusses the findings of a simulation study, and illustrates the use of `GJRM` on the AREDS data.

2.2 The Model

Let us consider the pair of survival times (T_{1i}, T_{2i}), a vector of covariates $\mathbf{x}_i$, for $i = 1, 2, \ldots, n$ where n represents the sample size, and a generic parameter vector $\boldsymbol{\delta} \in \mathbb{R}^W$ of dimension W. We assume that T_{1i} and T_{2i} have marginal survival functions written as $S_\nu(t_{\nu i}|\mathbf{x}_{\nu i}; \boldsymbol{\beta}_\nu) = P(T_{\nu i} > t_{\nu i}|\mathbf{x}_{\nu i}; \boldsymbol{\beta}_\nu) \in (0,1)$, for $\nu = 1, 2$, and a joint survival function expressed as $S(t_{1i}, t_{2i}|\mathbf{x}_i; \boldsymbol{\delta}) = P(T_{1i} > t_{1i}, T_{2i} > t_{2i}|\mathbf{x}_i; \boldsymbol{\delta})$. The survival functions are linked via a copula as follows

$$S(t_{1i},t_{2i}|\mathbf{x}_i;\boldsymbol{\delta}) = C\left(S_1(t_{1i}|\mathbf{x}_{1i};\boldsymbol{\beta}_1),S_2(t_{2i}|\mathbf{x}_{2i};\boldsymbol{\beta}_2);m\{\eta_{3i}(\mathbf{x}_{3i};\boldsymbol{\beta}_3)\}\right),$$

where $\boldsymbol{\delta}^\mathsf{T} = (\boldsymbol{\beta}_1^\mathsf{T},\boldsymbol{\beta}_2^\mathsf{T},\boldsymbol{\beta}_3^\mathsf{T})$, $\mathbf{x}_{1i}$, $\mathbf{x}_{2i}$ and $\mathbf{x}_{3i}$ are vectors of covariates, which can be sub-vectors of or equal to $\mathbf{x}_i$, with associated coefficient vectors $\boldsymbol{\beta}_1 \in \mathbb{R}^{W_1}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{W_2}$ and $\boldsymbol{\beta}_3 \in \mathbb{R}^{W_3}$, $W = W_1 + W_2 + W_3$, $C : (0,1)^2 \to (0,1)$ is a uniquely defined 2-dimensional copula function with coefficient $\theta_i = m\{\eta_{3i}(\mathbf{x}_{3i};\boldsymbol{\beta}_3)\}$ modelling the potentially varying dependence of (T_{1i},T_{2i}) across observations, $\eta_{3i}(\mathbf{x}_{3i};\boldsymbol{\beta}_3) \in \mathbb{R}$ is a predictor which includes generic additive covariate effects, and m is a monotonic and differentiable one-to-one transformation function ensuring that the restriction on the space of the parameter being considered is not violated. A similar specification has been previously adopted; see, e.g., Emura et al. (2021), Geerdens et al. (2018) and Marra & Radice (2020). The copulae implemented in `GJRM` are reported in Table 2.1, which also shows the relation between θ and the Kendall's $\tau \in [-1,1]$. If a copula can only account for positive dependence (e.g., Gumbel) then its counter-clockwise rotated versions can also be obtained (Brechmann & Schepsmeier, 2013).

The marginal survival functions can be written as

$$g_v\left[S_v(t_{vi}|\mathbf{x}_{vi};\boldsymbol{\beta})\right] = \eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v)), \tag{2.1}$$

where $g_v : (0,1) \to \mathbb{R}$ is a monotone and twice continuously differentiable link function with bounded derivatives, $\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v)) \in \mathbb{R}$ is an additive predictor which models the baseline hazard and several types of covariate effects, and $\mathbf{f}_v(\boldsymbol{\beta}_v)$ has the role of imposing a monotonicity constraint when evaluating the baseline function of time contained in the additive predictor (see the next section). Equation (2.1) can also be written as $S(t_{vi}|\mathbf{x}_{vi};\boldsymbol{\beta}_v) = G_v\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\}$, where G_v is an inverse link function. The cumulative hazard and hazard functions are defined as $H_v(t_{vi}|\mathbf{x}_{vi};\boldsymbol{\beta}_v) = -\log\left[G_v\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\}\right]$, and

$$h_v(t_{vi}|\mathbf{x}_{vi};\boldsymbol{\beta}_v) = -\frac{G_v'\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\}}{G_v\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}(\boldsymbol{\beta}_v))\}}\frac{\partial\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))}{\partial t_{vi}}, \tag{2.2}$$

Table 2.1: Definition of the copulae implemented in the `R` package `GJRM`, with corresponding parameter range of association parameter θ, one-to-one transformation function of θ, relation between Kendall's τ and θ, and range of τ. $\Phi_2(\cdot,\cdot;\theta)$ denotes the cumulative distribution function (cdf) of the standard bivariate normal distribution with correlation coefficient θ, and $\Phi(\cdot)$ the cdf of the univariate standard normal distribution. $t_{2,\zeta}(\cdot,\cdot;\zeta,\theta)$ indicates the cdf of the standard bivariate Student-t distribution with correlation θ and fixed $\zeta \in (2,\infty)$ degrees of freedom, and $t_\zeta(\cdot)$ denotes the cdf of the univariate Student-t distribution with ζ degrees of freedom. $A(t) = 1 - \left[t^{-\theta} + (1-t)^{-\theta}\right]^{-\frac{1}{\theta}}$ is the Pickands dependence function of the Galambos copula. $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{\exp(t)-1}dt$ is the Debye function and $D_2(\theta) = \int_0^1 t\log(t)(1-t)^{\frac{2(1-\theta)}{\theta}}dt$. Quantities Q and R are given by $1+(\theta-1)(u_1+u_2)$ and $Q^2 - 4\theta(\theta-1)u_1u_2$, respectively. The Kendall's τ for `"PL"` is computed numerically since no analytical expression is available. Argument `BivD` of `gjrm()` in `GJRM` allows the user to employ the desired copula and can be set to any of the values within brackets next to the copula names in the first column; for example, `BivD = "C0"`. For Clayton, Galambos, Gumbel and Joe, the number after the capital letter indicates the degree of rotation required: the possible values are `0`, `90`, `180` and `270`. The rotations are defined as $C_{90}(u_1,u_2;\theta) = u_2 - C(1-u_1,u_2)$, $C_{180}(u_1,u_2;\theta) = u_1+u_2-1+C(1-u_1,1-u_2)$ and $C_{270}(u_1,u_2;\theta) = u_1 - C(u_1,1-u_2)$.

Copula	$C(u_1,u_2;\theta)$	Range of θ	Transf.of θ	Kendall's τ	Range of τ
AMH (`"AMH"`)	$\frac{u_1u_2}{1-\theta(1-u_1)(1-u_2)}$	$[-1,1]$	$\tanh^{-1}(\theta)$	$-\frac{2}{3\theta^2}\left\{\theta+(1-\theta)^2\log(1-\theta)\right\}+1$	$[-0.1817,1/3]$
Clayton (`"C0"`)	$\left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1/\theta}$	$(0,\infty)$	$\log(\theta)$	$\frac{\theta}{\theta+2}$	$(0,1]$
FGM (`"FGM"`)	$u_1u_2\{1+\theta(1-u_1)(1-u_2)\}$	$[-1,1]$	$\tanh^{-1}(\theta)$	$\frac{2}{9}\theta$	$[-2/9,2/9]$
Frank (`"F"`)	$-\theta^{-1}\log\{1+(\exp\{-\theta u_1\}-1)$ $(\exp\{-\theta u_2\}-1)/(\exp\{-\theta\}-1)\}$	$\mathbb{R}\backslash\{0\}$	–	$1-\frac{4}{\theta}[1-D_1(\theta)]$	$(-1,1)\backslash\{0\}$
Galambos (`"GAL"`)	$u_1u_2\exp\left[\left\{(-\log u_1)^{-\theta}+(-\log u_2)^{-\theta}\right\}^{-1/\theta}\right]$	$(0,\infty)$	$\log(\theta)$	$\int_0^1\frac{t(1-t)}{A(t)}A''(t)dt$	$(0,1]$
Gaussian (`"N"`)	$\Phi_2\left(\Phi^{-1}(u_1),\Phi^{-1}(u_2);\theta\right)$	$[-1,1]$	$\tanh^{-1}(\theta)$	$\frac{2}{\pi}\arcsin(\theta)$	$[-1,1]$
Gumbel (`"G0"`)	$\exp\left[-\left\{(-\log u_1)^\theta+(-\log u_2)^\theta\right\}^{1/\theta}\right]$	$[1,\infty)$	$\log(\theta-1)$	$1-\frac{1}{\theta}$	$[0,1]$
Joe (`"J0"`)	$1-\left\{(1-u_1)^\theta+(1-u_2)^\theta-(1-u_1)^\theta(1-u_2)^\theta\right\}^{1/\theta}$	$(1,\infty)$	$\log(\theta-1)$	$1+\frac{4}{\theta^2}D_2(\theta)$	$(0,1]$
Plackett (`"PL"`)	$\left(Q-\sqrt{R}\right)/\{2(\theta-1)\}$	$(0,\infty)$	$\log(\theta)$	–	$(-1,1]$
Student's t (`"T"`)	$t_{2,\zeta}\left(t_\zeta^{-1}(u_1),t_\zeta^{-1}(u_2);\zeta,\theta\right)$	$[-1,1]$	$\tanh^{-1}(\theta)$	$\frac{2}{\pi}\arcsin(\theta)$	$[-1,1]$

respectively, where $G'_v\left\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\right\}=\partial G_v\left\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\right\}/\partial\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))$. Table 2.2 displays the functions g, G and G' implemented in `GJRM`.

Table 2.2: Link functions implemented in `GJRM`. Φ and ϕ are the cumulative distribution and density functions of a univariate standard normal distribution.

Model	Link $g(S)$	Inverse link $g^{-1}(\eta)=G(\eta)$	$G'(\eta)$
Prop. hazards (`"PH"`)	$\log\{-\log(S)\}$	$\exp\{-\exp(\eta)\}$	$-G(\eta)\exp(\eta)$
Prop. odds (`"PO"`)	$-\log\left(\frac{S}{1-S}\right)$	$\frac{\exp(-\eta)}{1+\exp(-\eta)}$	$-G^2(\eta)\exp(-\eta)$
probit (`"probit"`)	$-\Phi^{-1}(S)$	$\Phi(-\eta)$	$-\phi(-\eta)$

2.2.1 Predictor specification

The key difference between $\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))$, for $v=1,2$, and $\eta_{3i}(\mathbf{x}_{3i};\boldsymbol{\beta}_3)$, where in the latter $\mathbf{f}_3$ is the identity vector function, is that the two former predictors must include smooth functions of times t_{vi} which can be treated as regressors. In fact, the construction of the design matrices for the three additive predictors follows the same philosophy. We, therefore, consider a generic $\eta_{\nu i}$ $(\nu=1,2,3)$, where the dependence on the covariates and parameters is momentarily dropped, an overall covariate vector $\mathbf{z}_{\nu i}$ containing $\mathbf{x}_{\nu i}$ and $t_{\nu i}$ when $\nu=1,2$, and $\mathbf{z}_{3i}=\mathbf{x}_{3i}$. For simplicity, the dimensions of $\mathbf{z}_{1i}$ and $\mathbf{z}_{2i}$ are assumed to be W_1 and W_2.

An additive predictor can be defined as

$$\eta_{\nu i}=\beta_{\nu 0}+\sum_{k_\nu=1}^{K_\nu}s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i}),\ i=1,\ldots,n, \tag{2.3}$$

where $\beta_{\nu 0}\in\mathbb{R}$ is an overall intercept, $\mathbf{z}_{\nu k_\nu i}$ denotes the k_ν^{th} sub-vector of the complete vector $\mathbf{z}_{\nu i}$ and the K_ν functions $s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i})$ represent generic effects which are chosen according to the type of covariate(s) considered. Each $s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i})$ can be represented as a linear combination of $J_{\nu k_\nu}$ basis functions $b_{\nu k_\nu j_{\nu k_\nu}}(\mathbf{z}_{\nu k_\nu i})$ and regression coefficients $f_{\nu k_\nu j_{\nu k_\nu}}(\beta_{\nu k_\nu j_{\nu k_\nu}})\in\mathbb{R}$, that is (e.g., Wood, 2017)

$$\sum_{j_{\nu k_\nu}=1}^{J_{\nu k_\nu}}f_{\nu k_\nu j_{\nu k_\nu}}(\beta_{\nu k_\nu j_{\nu k_\nu}})b_{\nu k_\nu j_{\nu k_\nu}}(\mathbf{z}_{\nu k_\nu i}). \tag{2.4}$$

The above formulation implies that the vector of evaluations $\{s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu 1}),\ldots,s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu n})\}^{\mathsf{T}}$

can be written as $\mathbf{Z}_{vk_v}\mathbf{f}_{vk_v}(\boldsymbol{\beta}_{vk_v})$ with $\mathbf{f}_{vk_v}(\boldsymbol{\beta}_{vk_v}) = (f_{vk_v1}(\beta_{vk_v1}),\ldots,f_{vk_vJ_{vk_v}}(\beta_{vk_vJ_{vk_v}}))^{\mathsf{T}}$ and design matrix $\mathbf{Z}_{vk_v}[i,j_{vk_v}] = b_{vk_vj_{vk_v}}(\mathbf{z}_{vk_vi})$. Therefore, equation (2.3) can be written as

$$\boldsymbol{\eta}_v = \beta_{v0}\mathbf{1}_n + \mathbf{Z}_{v1}\mathbf{f}_{v1}(\boldsymbol{\beta}_{v1}) + \ldots + \mathbf{Z}_{vK_v}\mathbf{f}_{vK_v}(\boldsymbol{\beta}_{vK_v}),$$

where $\mathbf{1}_n$ is an n-dimensional vector made up of ones, or in a more compact way as $\boldsymbol{\eta}_v = \mathbf{Z}_v\mathbf{f}_v(\boldsymbol{\beta}_v)$, where $\mathbf{Z}_v = (\mathbf{1}_n,\mathbf{Z}_{v1},\ldots,\mathbf{Z}_{vK_v})$ and $\mathbf{f}_v(\boldsymbol{\beta}_v) = (\beta_{v0},\mathbf{f}_{v1}(\boldsymbol{\beta}_{v1})^{\mathsf{T}},\ldots,\mathbf{f}_{vK_v}(\boldsymbol{\beta}^{\mathsf{T}}_{vK_v}))^{\mathsf{T}}$. Note that smooth functions are subject to centering identifiability constraints (Wood, 2017). Each $\boldsymbol{\beta}_{vk}$ has an associated quadratic penalty $\lambda_{vk_v}\boldsymbol{\beta}^{\mathsf{T}}_{vk_v}\mathbf{D}_{vk_v}\boldsymbol{\beta}_{vk_v}$ which has to be used during model fitting to enforce specific properties on the k_v^{th} function, such as smoothness. Smoothing parameter $\lambda_{vk_v} \in [0,\infty)$ controls the trade-off between fit and smoothness, whereas $\mathbf{D}_{vk_v}$ depends on the choice of the basis functions. For example, for a cubic regression spline, $\mathbf{D}_{vk_v}$ is given by the integrated square second derivative of the basis functions, i.e. $\int \mathbf{d}_{vk_v}(z_{vk_v})\mathbf{d}_{vk_v}(z_{vk_v})^{\mathsf{T}} dz_{vk_v}$ with the $j^{th}_{vk_v}$ element of $\mathbf{d}_{vk_v}(z_{vk_v})$ defined as $\partial^2 b_{vk_vj_{vk_v}}(z_{vk_v})/\partial z^2_{vk_v}$. P-splines are, instead, characterised by a difference penalty applied directly to the parameters, to control function wiggliness. When second differences are assumed, $\mathbf{D}_{vk_v}$ is a tridiagonal matrix with -1 on the upper and lower diagonals, $\mathbf{D}_{vk_v}[\imath,\imath] = 1$ for $\imath = 1, J_{vk_v}$ and $\mathbf{D}_{vk_v}[\imath,\imath] = 2$ for $\imath = 2,\ldots,J_{vk_v} - 1$. The overall penalty can be defined as $\boldsymbol{\beta}^{\mathsf{T}}_v\mathbf{D}_v\boldsymbol{\beta}_v$, where $\mathbf{D}_v = \mathrm{diag}(0,\lambda_{v1}\mathbf{D}_{v1},\ldots,$ $\lambda_{vK_v}\mathbf{D}_{vKv})$. The above formulation allows for many types of flexible covariate effects (e.g., non-linear, random, spatial, interactions). In fact, several definitions of basis functions and penalty terms are supported in `GJRM` which are based on Wood (2017). The time effects are instead modelled using the monotonic P-spline approach which will guarantee that the estimated survival functions are monotonically decreasing or equivalently that the hazard functions are positive. To avoid redundancies, we refer the reader to Chapter 3, Section 3.3.2, for a detailed description of how this is done.

2.2.2 Remarks

When working with interval-censored observations, the implementation of the model set up needs to account for the information contained in the lower and upper bounds of the censoring intervals. Therefore, for each margin, two distinct design matrices (based on the two bounds) and hence additive predictors are required. The covariates and parameter vector $\boldsymbol{\beta}_v$ used in their construction will be the same.

In equation (2.2), $\partial\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))/\partial t_{vi}$ is required. Based on the results of the previous paragraph, $\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))$ can be written as $\mathbf{Z}_{vi}(t_{vi},\mathbf{x}_{vi})^{\mathsf{T}}\mathbf{f}_v(\boldsymbol{\beta}_v)$ which means that the quantity of interest can be calculated as

$$\lim_{\varepsilon\to 0}\left\{\frac{\mathbf{Z}_{vi}(t_{vi}+\varepsilon,\mathbf{x}_{vi})-\mathbf{Z}_{vi}(t_{vi}-\varepsilon,\mathbf{x}_{vi})}{2\varepsilon}\right\}^{\mathsf{T}}\mathbf{f}_v(\boldsymbol{\beta}_v)=\mathbf{Z}_{vi}^{\prime\mathsf{T}}\mathbf{f}_v(\boldsymbol{\beta}_v),$$

where $\mathbf{Z}'_{vi}$ can be conveniently obtained by finite differencing. In practice, ε is set to small value and the limit is approximated by the ratio. Through extensive experimentation we found this approach to work well and to not be sensitive to the exact choice of ε.

Formulation (2.4) requires a value for J_{vk_v}. This is especially relevant when modelling the effects of continuous covariates. As explained by Vatter & Chavez-Demoulin (2015), among others, all that is required is to set J_{vk_v} to an arbitrary value that allows for enough flexibility in estimating the related smooth term; penalisation during model fitting will then ensure that a good balance between fit and parsimony is achieved.

The general model formulation introduced in the previous two sections yields the proportional hazards and odds models as special cases; for details on this, we refer the reader to, e.g., Liu et al. (2018) whose developments are based on the same conceptual survival modelling framework adopted here. Other important benefits are that quantities such as $h_v(t_{vi}|\mathbf{x}_{vi};\boldsymbol{\beta}_v)$ can be directly obtained without the need for numerical integration, and that time-dependent effects can be easily incorporated in the model via terms like $s_{vk_v}(t_{vi})\mathbf{x}_{vk_vi}$.

2.3 Parameter Estimation

Let T_{vi} denote the true event time, for $v = 1,2$. In the case of censoring, T_{vi} is only known to lie within the interval (L_{vi}, R_{vi}), where L_{vi} and R_{vi} represent left and right censoring times. If $L_{vi} = 0$ then the i^{th} observation for the v margin is defined as left-censored. When $R_{vi} = \infty$, the observation is classified as right-censored. If L_{vi} and R_{vi} take on finite distinct non-zero values then the observation is interval-censored. Exact observations relate to the case $L_{vi} = R_{vi}$. Since we are dealing with a bivariate response, there will be sixteen possible censoring combinations to account for; these can be characterised through the indicator functions $\gamma_{I_{vi}}$ and $\gamma_{U_{vi}}$, where $\gamma_{I_{vi}}$ takes value 1 if the i^{th} observation is interval-, right- or left-censored and 0 otherwise. Similarly, $\gamma_{U_{vi}}$ is 1 if the i^{th} observation is uncensored and 0 otherwise.

$$
\begin{aligned}
\ell(\delta) &= \gamma_{U_{1i}}\gamma_{U_{2i}}\sum_{i=1}^{n}\log f(t_{1i},t_{2i}) + \gamma_{I_{1i}}\gamma_{I_{2i}}\sum_{i=1}^{n}\log P(T_{1i}\in(l_{1i},r_{1i}],T_{2i}\in(l_{2i},r_{2i}]) \\
&+ \gamma_{U_{1i}}\gamma_{I_{1i}}\sum_{i=1}^{n}\log\left[\int_{l_{21}}^{r_{2i}} f(t_{1i},y)dy\right] + \gamma_{I_{1i}}\gamma_{U_{1i}}\sum_{i=1}^{n}\log\left[\int_{l_{1i}}^{r_{1i}} f(y,t_{2i})dy\right] \\
&= \gamma_{U_{1i}}\gamma_{U_{2i}}\sum_{i=1}^{n}\log\left[\frac{\partial^2}{\partial t_{1i}\partial t_{2i}}C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}));\theta_i\}\right] \\
&+ \gamma_{I_{1i}}\gamma_{I_{2i}}\sum_{i=1}^{n}\log\Big[C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}));\theta_i\} - C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}));\theta_i\} \\
&- C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}));\theta_i\} + C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}));\theta_i\}\Big] \\
&+ \gamma_{U_{1i}}\gamma_{I_{1i}}\sum_{i=1}^{n}\log\Big[\frac{\partial}{\partial t_{1i}}\Big(C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}));\theta_i\} \\
&- C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}));\theta_i\}\Big)\Big] \\
&+ \gamma_{I_{1i}}\gamma_{U_{1i}}\sum_{i=1}^{n}\log\Big[\frac{\partial}{\partial t_{2i}}\Big(C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}));\theta_i\} \\
&- C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}));\theta_i\}\Big)\Big].
\end{aligned}
$$

The case of interval censoring incorporates both right and left censoring. So, if the i^{th} observation for the v margin is right-censored then $r_{vi} = \infty$. If it is left-censored then $l_{vi} = 0$. The terms of the above log-likelihood have been derived as follows:

- T_{1i} interval-censored and T_{2i} interval-censored:

$$
\begin{aligned}
&P(l_{1i} < T_{1i} < r_{1i}, l_{2i} < T_{2i} < r_{2i}) = P(T_{1i} < r_{1i}, T_{2i} < r_{2i}) - P(T_{1i} < l_{1i}, T_{2i} < r_{2i})\\
&- P(T_{1i} < r_{1i}, T_{2i} < l_{2i}) + P(T_{1i} < l_{1i}, T_{2i} < l_{2i})\\
&= F(r_{1i}, r_{2i}) - F(l_{1i}, r_{2i}) - F(r_{1i}, l_{2i}) + F(l_{1i}, l_{2i})\\
&= S(l_{1i}, l_{2i}) - S(l_{1i}, r_{2i}) - S(r_{1i}, l_{2i}) + S(r_{1i}, r_{2i})\\
&= C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\}\\
&\quad - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\} + C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\}.
\end{aligned}
$$

Recall that, using the above formulation, all scenarios deriving from any combination of right-, left- and interval-censored bivariate outcomes can be produced.

- T_{1i} uncensored and T_{2i} uncensored (in this case, $t_{1i} = r_{1i} = l_{1i}$ and $t_{2i} = r_{2i} = l_{2i}$):

$$
\begin{aligned}
f(t_{1i}, t_{2i}) &= \frac{\partial^2}{\partial t_{1i} \partial t_{2i}} F(t_{1i}, t_{2i}) = \frac{\partial^2}{\partial t_{1i} \partial t_{2i}} [1 - S(t_{1i}) - S(t_{2i}) + S(t_{1i}, t_{2i})]\\
&= \frac{\partial^2}{\partial t_{1i} \partial t_{2i}} C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i})); \theta_i\}.
\end{aligned}
$$

- T_{1i} uncensored and T_{2i} interval-censored (the “swapped” case can be trivially derived by switching the subscripts where required):

$$
\begin{aligned}
&\int_{l_{2i}}^{r_{2i}} f(t_{1i}, y)dy = \int_0^{r_{2i}} f(t_{1i}, y)dy - \int_0^{l_{2i}} f(t_{1i}, y)dy = \frac{\partial}{\partial t_{1i}} F(t_{1i}, r_{2i}) - \frac{\partial}{\partial t_{1i}} F(t_{1i}, l_{2i})\\
&= \frac{\partial}{\partial t_{1i}} [1 - S_1(t_{1i}) - S_2(r_{2i}) + S(t_{1i}, r_{2i})] - \frac{\partial}{\partial t_{1i}} [1 - S_1(t_{1i}) - S_2(l_{2i}) + S(t_{1i}, l_{2i})]\\
&= \frac{\partial}{\partial t_{1i}} \left[C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\} - C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\}\right].
\end{aligned}
$$

As above, the right- and left-censored cases can be easily worked out.

The reader is referred to Supplementary Material-Section A for the more explicit version of the log-likelihood. As explained in Section 2.2.1, quadratic penalties have to be employed during model fitting to calibrate the trade-off between fit and smoothness. Therefore, we maximise

$$\ell_p(\boldsymbol{\delta}) = \ell(\boldsymbol{\delta}) - \frac{1}{2}\boldsymbol{\delta}^{\mathsf{T}}\mathbf{S}\boldsymbol{\delta},$$

where ℓ_p is the penalised log-likelihood, $\mathbf{S} = \text{diag}(\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3)$, $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ are overall penalties that take the form specified in Section 2.2.1 and include $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$ and $\boldsymbol{\lambda}_3$. The smoothing parameters can be collected in the vector $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^{\mathsf{T}}, \boldsymbol{\lambda}_2^{\mathsf{T}}, \boldsymbol{\lambda}_3^{\mathsf{T}})^{\mathsf{T}}$.

Model fitting is challenging in this context because of the non-linear dependence of $\mathbf{f}_v(\boldsymbol{\beta}_v)$ on $\boldsymbol{\beta}_v$, the requirement of estimating $\boldsymbol{\lambda}$ automatically, and the need for providing a stable and fast implementation that is computationally solid and practically usable. To this end, we employ the stable and fast trust region algorithm presented in Marra & Radice (2020) which is based on a simultaneous penalised maximum likelihood approach with integrated automatic multiple smoothing parameter selection. A major challenge with the implementation of this algorithm is that the analytical score vector and Hessian matrix of $\ell(\boldsymbol{\delta})$ are required. Given the generality and complexity of the model, deriving such quantities has been a rather tedious and time-consuming task; these are given in Sections B and C of the Supplementary Material, and have been thoroughly checked and verified numerically. Starting values for the marginal survival models are obtained by combining the use of the shape constrained smoothing approach of Pya & Wood (2015) with the procedure detailed in Liu et al. (2018). An initial value for the copula parameter is worked out by using a transformation of the empirical τ between the responses. The simulation study in Supplementary Material-Section D supports the empirical effectiveness of the estimation framework. Briefly, several sample sizes ($n = 300, 1000, 1500$ and 2000) are considered as well as both mild (62.86% and 44.98%) and high (84.82% and 77.13%) censoring levels. Overall, the modelling framework performs consistently well, even for the lowest sample size. The parametric effects and smooth effects were properly recovered in all of the scenarios considered, exhibiting both low bias and RMSE. The estimation of the quantities related to the copula dependence parameter is more challenging and shows some bias when compared to the other model parameters, although performance is still deemed satisfactory. As expected, parameter estimation is more difficult in the presence of high censoring, due to the loss of information implied by censoring itself. Note, however, that both of these challenging settings improve markedly as the sample size increases.

The number of parameters in the model can be quantified using the notion of number of effective degrees of freedom (edf). The edf for a model containing only unpenalised terms would clearly be equal to W, whereas that for a penalised model can be written as $W-\zeta$, where $\zeta = \mathrm{tr}\left\{(-\boldsymbol{H}+\mathbf{S})^{-1}\mathbf{S}\right\}$ and $\boldsymbol{H}$ is the Hessian matrix. This shows the role that $\boldsymbol{\lambda}$ (contained in $\mathbf{S}$) plays in determining the model edf, which indeed is a value in the range $[W-\zeta, W]$. The definition of the edf of a single smooth or penalised term follows the same logic and has a value smaller than or equal to J_{vk_v}.

2.4 Inference

Inferential results can be borrowed from known theory for general penalised likelihood-based models. Specifically, at convergence, reliable confidence intervals for any linear or non-linear function of $\boldsymbol{\delta}$ are obtained by exploiting the Bayesian large sample approximation (e.g., Wahba, 1983; Wood et al., 2016) $\delta \dot{\sim} \mathcal{N}(\hat{\delta}, \mathbf{V}_{\boldsymbol{\delta}})$, where $\hat{\delta} = \arg\max_{\delta} \ell_p(\boldsymbol{\delta})$ and $\mathbf{V}_{\boldsymbol{\delta}} = (-\boldsymbol{H}(\hat{\boldsymbol{\delta}})+\mathbf{S})^{-1}$. One view of the smoothing process is that the penalty employed during fitting imposes the belief that the true function is more likely to be smooth than wiggly. This belief can be expressed in a Bayesian manner by defining a prior distribution on function wiggliness $f_{\delta} \propto \exp\left(-1/2\delta^{\mathsf{T}}\mathbf{S}\delta\right)$. The reason for adopting a Bayesian viewpoint when it comes to inference in penalised models is that the Bayesian covariance matrix gives close to across-the-function frequentist coverage probabilities, while the frequentist covariance matrix $-(\boldsymbol{H}(\hat{\boldsymbol{\delta}})-\mathbf{S})^{-1}\boldsymbol{H}(\hat{\boldsymbol{\delta}})(\boldsymbol{H}(\hat{\boldsymbol{\delta}})-\mathbf{S})^{-1}$ does not. The problem with constructing confidence intervals for the smooth terms in a model is the smoothing bias, which has to be corrected in order to obtain confidence intervals with good properties. It turns out that the Bayesian confidence intervals include a component accounting for bias, as elaborated by Wood (2017, Section 6.10, see also references therein), thus explaining the good coverage properties achieved by them and thus the reason we use them for inference.

Intervals for nonlinear functions of $\boldsymbol{\delta}$ can be conveniently obtained via posterior simulation (see, e.g., Marra & Radice (2020) for an example. P-values for the terms in the model can be reliably obtained by using the results summarised in Wood (2017, Section 6.12) which are based on $\mathbf{V}_{\boldsymbol{\delta}}$. Note that for the parametric (unpenalised) terms in the model, the corresponding entries in $\mathbf{S}$ (contained in $\mathbf{V}_{\boldsymbol{\delta}}$) are equal to zero. This would be equivalent to using the classical frequentist result, based on $-\boldsymbol{H}(\hat{\boldsymbol{\delta}})$, for such terms.

2.5 Application to AREDS data

The proposed approach is applied to a dataset from the AREDS available through the R package CopulaCenR (Sun & Ding, 2021b), which includes 629 Caucasian participants. The event of interest is the progression to late-AMD disease, which is the most common cause of blindness in developed countries (Swaroop et al., 2009). Due to intermittent assessment times (every 6 months up to the first 6 years and every 1 year thereafter), the exact time when each eye progressed to late-AMD is only known to lie in a certain interval. More specifically, less than half of the subjects developed late-AMD in both eyes (bivariate interval-censored); around 20% of the subjects developed late-AMD in one eye and did not develop late-AMD in the other eye before the end of the study (mixed interval- and right-censored); more than one third of the subjects did not develop late-AMD in both eyes (bivariate right-censored).

The dataset contains three covariates potentially related with AMD progression: SevScaleBL for baseline AMD severity score (a factor variable with values between 4 and 8 with a higher value indicating more severe AMD), ENROLLAGE for baseline age (a numeric variable), and rs2284665 for a genetic variant (a factor variable with levels 0, 1 and 2 which represent GG, GT and TT, respectively).

For the marginal equations, the smooth functions of `ENROLLAGE` and the time variables were represented using penalised thin plate regression splines with second order penalty (Wood, 2017) and monotonic penalised B-splines (see Section 2.2.1), respectively. The number of bases used for each smooth was 10; increasing this value did not lead to visible changes in the estimated curves. The remaining variables entered the predictors of the marginals linearly. All link functions shown in Table 2.2 were considered in the modelling. For both margins, `PO` was found to yield the smallest AIC and BIC. As for the copula, we started off with the Gaussian and then, based on the (negative or positive) sign of the dependence, we tried out alternative specifications that were consistent with this initial finding. Using a 2.60-GHz Intel(R) Core(TM) computer running Windows 10, the average computing time to fit a model was about 9 seconds and the length of the model parameter vector was 43. Using the AIC and BIC, where, in their construction, the model edf was used in place of the number of model parameters, the chosen model is based on the Plackett copula with `PO` margins. The `R` code used to fit the models, and to produce all the numerical and visual summaries commented below can be found in Supplementary Material-Section A. Using the second and third best copulae did not change the conclusions of the analysis.

	Left Eye		Right Eye	
Parametric Eff.	Estimate (Std.error)	$Pr(>\|z\|)$	Estimate (Std.error)	$Pr(>\|z\|)$
`(Intercept)`	-18.0368(4.39)	4.09e-05	-33.2811 (10.89)	0.002246
`ENROLLAGE`	-	-	0.0364 (0.01)	0.011592
`SevScale5`	0.6707 (0.24)	0.00556	0.8187 (0.25)	0.001365
`SevScale6`	1.0049 (0.22)	6.90e-06	1.2957 (0.23)	4.81e-07
`SevScale7`	1.9255 (0.23)	< 2e-16	2.4270 (0.25)	< 2e-16
`SevScale8`	2.8208 (0.31)	< 2e-16	3.2793 (0.32)	< 2e-16
`rs22846651`	0.3269 (0.16)	0.04966	0.4589 (0.16)	0.006467
`rs22846652`	0.6058 (0.23)	0.00927	0.7874 (0.22)	0.000481

Table 2.3: AREDS data. Parameters estimates, standard errors and p-values obtained from fitting the model using `gjrm()`. Note: the full output of the `R` model summary is reported. It is clear that the p-values relating to the categorical variables do not have a meaningful interpretation, these are only included for completeness.

The model parameters estimates are reported in Table 2.3. The estimated regression coefficients of `SevScaleBL`, which are 0.67, 1.00, 1.93, 2.82 in the equation for the left eye and 0.82, 1.21, 2.43, 3.28 in that for the right eye, imply, as expected, that the subjects with higher baseline AMD severity score have a higher risk than the subjects with lower baseline AMD severity score. As for the genetic variant, rs2284665, the estimated parameters are 0.33 and 0.61 for the left eye equation, and 0.46 and 0.79 for the right one. This is consistent with the interpretation that participants with TT genotype group have the highest risk of developing the disease, followed by participants with GT genotype group.

Figure 2.1 shows the estimated functional forms for the effect of ENROLLAGE and times of the selected model. Note that the smooth function for ENROLLAGE in the second equation has not been reported as the effect was linear ($edf = 1$), which indeed indicates that there is a constantly increasing risk associated with age. As for the first equation, the estimated smooth function confirms this increasing trend. Also, since there are few subjects who are younger than 60 and older than 80, the point-wise intervals are larger at lower and higher age values. The plots for the time variables exhibit increasing monotonic trends, suggesting again that the risk increases with time.

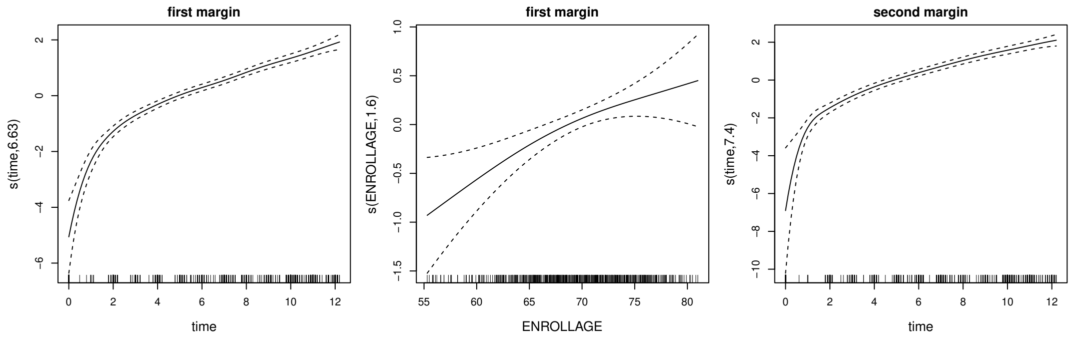

Figure 2.1: AREDS data. Baseline risks and smoothed effect of baseline age (ENROLLAGE), for the first equation only. 95% point-wise intervals are based on the result mentioned in Section 2.3. The rug plot, at the bottom of each graph, shows the values of the considered variable. The number in brackets in the y-axis caption of each plot represents the edf of the respective estimated smooth function.

The estimated Kendall's τ is 0.36, with 95% confidence interval $(0.304, 0.408)$, which implies moderate dependence in AMD progression between the two eyes. Given the capabilities of the proposed modelling framework, we also specified a model where the dependence parameter is expressed as a flexible function of the covariates. This feature can help understand how and which covariates modify the strength of the dependence across observations. In this case, however, the coefficients were found not to be significant (see Supplementary Material-Section E). It is worth noting that such specifications are likely to be more successful in finding covariate patterns when the number of observations is higher than that available for this study.

Using the chosen model, we produced joint survival functions under several scenarios. The left panel of Figure 2.2 displays the joint progression-free probability contours for subjects who are 69 years old, with AMD severity score equal to 6 for both eyes, but with different `rs2284665` genotypes. The middle panel of Figure 2.2 shows the joint progression-free probability contours for subjects who are 69 year old, with GT genotype, but with different severity scores (4, 6 and 8). Finally, the right panel of the figure plots the joint progression-free probability contours for GT genotype subjects, with AMD severity score equal to 6 in both eyes, but different ages (56, 69 and 81). In the left panel, it can be clearly seen that the three genotype groups are separated, with the GG group having the largest progression-free probabilities. In the middle panel, the difference between the three AMD severity groups is rather pronounced, with the highest AMD severity group having the smallest progression-free probabilities. Finally, the right panel shows how the progression-free probabilities are higher for younger subjects as compared to older subjects. The scenarios considered here illustrate how valuable the proposed modelling framework is in characterising and identifying AMD patients at a higher risk of developing late-AMD. Of course, several other scenarios can be considered and other quantities of interest worked out. For example, one could be interested in visualising conditional and marginal survival probabilities.

2.6 Discussion

We have introduced a copula link-based additive model for bivariate time-to-event outcomes under various types of censoring mechanisms. Model fitting is based on the simultaneous estimation of all model parameters and relies on a penalised

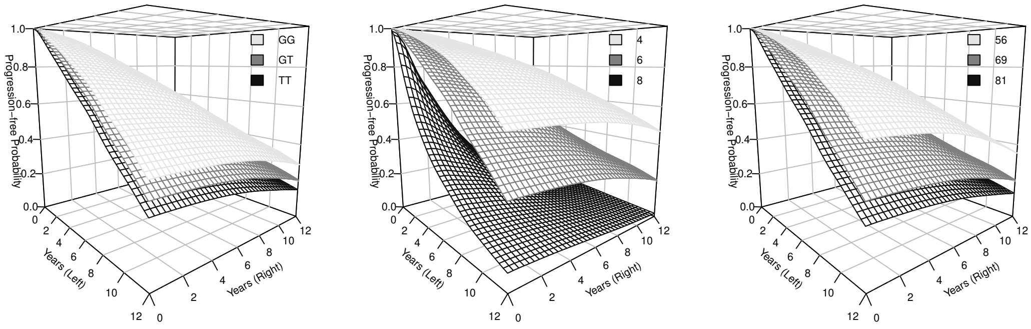

Figure 2.2: AREDS data. Joint progression-free probability contours for progression to late-AMD disease (in years) in the left and right eyes, under different scenarios. In left panel, age is set to 69, and AMD severity score to 6 for both eyes. In the middle panel, age is set to 69, and genotype to GT. In the right panel, genotype is set to GT, and AMD severity score to 6 in both eyes.

maximum likelihood approach with integrated stable and efficient automatic multiple smoothing parameter selection. Inferential results are also readily available. All developments have been integrated within the `R` package `GJRM` whose modularity allows for easy inclusion of potentially any parametric link marginal function and copula. The proposed approach makes a significant contribution in applied statistics as it is methodologically flexible, computationally sound and practically usable.

Although the literature in this area is reasonably ample, to the best of our knowledge, only Sun & Ding (2021a) provided a methodological framework together with software for modelling bivariate censored data. Unlike their copula approach, which allows the margins to be specified through semi-parametric transformation models, the baseline survival functions to be modelled using Bernstein polynomials and the dependence between events to be captured via one-parameter and two-parameter copulae, our proposal permits to specify all model parameters (including the dependence parameter) as flexible functions of covariate effects, model the baseline survival functions by means of monotonic P-splines which are theoretically and computationally advantageous, and conveniently characterise the marginals via links of the survival functions. Methodologically speaking, both approaches have been conceived to handle any combination of censoring mechanisms as well as have two different sets of regression coefficients for the marginal survival functions. However, from a computational point of view, the implementation provided by Sun & Ding (2021a) does not simultaneously support all possible bivariate combinations of censoring types and forces the two set of regression parameters to be the same.

Future research will focus on extending the approach to more than two event times (e.g., multi-morbidity) exploring, for instance, the use of multivariate Archimedean copulae, mixtures of powers, pair-copulae constructions, the multivariate Gaussian and Student's t distributions, and the composite likelihood approach (see, e.g., the supplementary material of Filippou et al., 2019, and references therein, which illustrates succinctly these ideas in a different context). Other potentially interesting extensions would be to account for informative and/or dependent censoring (e.g., Dettoni et al., 2020) as well as consider the case of excess hazard modelling (Eletti et al., 2022).

Chapter 3

A Spline-Based Framework for the Flexible Modelling of Continuously Observed Multistate Survival Processes

3.1 Introduction

When considering multistate processes for the modelling of life-history data, a particularly advantageous setting is that in which transition times are known exactly, i.e. the process is continuously observed. In this case, in fact, the overall model likelihood can be decomposed into the product of likelihoods referring to each specific transition only. Estimation then becomes equivalent to fitting one standard survival model for each transition, considering only the subset of the data relevant to that transition and including left-truncation times if the transition at hand can only happen once another has occurred. This is referred to as *separate estimation* (Putter et al., 2007; Putter, 2011; Crowther & Lambert, 2017). An important practical implication of this is that existing tools can be used to fit the transition-specific models. In particular, we propose to model each transition intensity through the general link-based additive modelling framework by Eletti et al. (2022), implemented in the `R` package `GJRM` (Marra & Radice, 2024). This modelling framework allows for the inclusion of virtually any type of covariate effects (including time-dependent effects) using any type of smoother (e.g., thin plate and cubic splines, and tensor products). Importantly, the use of shape constrained P-splines (SCOPs) to model time effects permits to approach the multiple univariate survival models directly on the survival scale, rather than on the hazards scale (which would require expensive numerical integration), while retaining a high degree of modelling flexibility. Specifically, SCOPs, developed by Pya & Wood (2015), extending the penalised B-splines discussed in the seminal work of Eilers & Marx (1996), elegantly embed the monotonicity required for the survival functions within the construction of the survival functions themselves, thus enabling very efficient parameter estimation. The exploration of different forms of dependence on past history also becomes considerably easier when the exact transition times are known. Indeed, assuming a semi-Markov process, the most common relaxation considered in the literature, rather than a Markov process, the most commonly made assumption, implies no further methodological difficulty.

When dealing with life-history data, one is often interested in assessing the effects of specific risk-factors on the probability of transitioning between states. When the process is assumed to be time-dependent and/or not-Markov, the computation of the transition probabilities is a nontrivial task. Two main approaches can be identified in the literature to address this problem and are detailed in Supplementary Material B.1. We adopt a simulation-based approach which allows one to compute the transition probabilities by simulating a number of paths through the assumed multistate process and counting the number of individuals experiencing each transition (Iacobelli & Carstensen, 2013; Touraine et al., 2016). This is appealing due its aptness at supporting any type of multistate process and was proposed in Fiocco et al. (2008) and implemented, amongst others, in the `R` package `mstate` (Putter et al., 2020), whose tools can be seamlessly integrated with the estimation approach implemented in the `R` package `GJRM`.

The remainder of the chapter is organised as follows. In Section 3.2, the mathematical setting of multistate survival processes is described, while Section 3.3 introduces the modelling framework. Sections 3.4, 3.5 and 3.6 discuss model estimation, the extraction of the transition probabilities and inference respectively. In Section 3.7, the *Rotterdam Breast Cancer Study* is introduced to exemplify the proposed framework. Finally, Section 3.8 provides some concluding remarks alongside directions of future work.

3.2 Mathematical setting of multistate survival processes

A continuous-time discrete-state stochastic process is a family of random variables $\{Z(t), t \in \mathcal{T}\}$ with some indexing set given by $\mathcal{T} = [0, \infty)$ in the survival setting. The set of all values that the process takes $\mathcal{S} := \{z : Z(t) = z, t \in \mathcal{T}\} \subseteq \{0, 1, 2, ...\}$ is called the state space, where $Z(t)$ denotes the state occupied at time t. A $p \times 1$ vector of left-continuous, time-dependent covariates is represented by $X(t)$. The history of the process, including the evolution of the covariates vector, is denoted by $\mathcal{F}_t = \{Z(u), X(u), 0 \leq u \leq t\}$. The transition intensities and the transition probabilities are then the two key quantities associated with the process. The former represent the rates of transition to a state s for an individual who is currently in another state r, formally

$$q^{(rs)}(t \mid \mathcal{F}_{t^-}) = \lim_{\Delta t \downarrow 0} \frac{P(Z(t + \Delta t^-) = s \mid Z(t^-) = r, \mathcal{F}_{t^-})}{\Delta t}, \quad r \neq s,$$

with $q^{(rs)}(t \mid \mathcal{F}_{t^-}) = 0$ if r is an absorbing state and $q^{(rr)}(t \mid \mathcal{F}_{t^-}) = -\sum\limits_{s \neq r} q^{(rs)}(t \mid \mathcal{F}_{t^-})$. The matrix with (r, s) element given by $q^{(rs)}(t \mid \mathcal{F}_{t^-})$ for every $r, s \in \mathcal{S}$ is called transition intensity matrix or generator matrix and we will denote it by $\mathbf{Q}(t \mid \mathcal{F}_{t^-})$. Similarly, we define the transition probability matrix associated with the time interval $[u, t]$ as the matrix with (r, s) element given by $P(Z(t) = s \mid Z(u) = r, \mathcal{F}_{u^-})$ and denote this by $\mathbf{P}(u, t \mid \mathcal{F}_{u^-})$. It is common to simplify the dependence on past history and time by assuming either a Markov or a semi-Markov process. The former implies that the probability of being in a given state at a given future time only depends on the current state occupied (Ross et al., 1996). The latter assumes that the future state only depends on the history of the process through the current state and through time since entry to the current state (Pyke, 1961). Note that we will consider only homogeneous semi-Markov processes, as defined in Yang & Nair (2011). The more general case of time-dependent semi-Markov process is out of the scope of this work. Exact knowledge of the transition times, as in our setting, allows for both assumptions to be modelled in an equally straightforward manner. The time for intermediate transitions will just need to be re-defined to be the time from entry to the current state.

3.3 Flexible transition-specific modelling

When a multistate process is continuously observed, each transition time can viewed as a standalone time-to-event and can thus be modelled through traditional survival analysis. It is well know that survival analysis can be undertaken on different scales. One such option is to model transformations of the survival function using generalised survival models, a class that was first introduced by Younes & Lachin (1997). Subsequent works further developed this approach (e.g., Royston & Parmar, 2002; Liu et al., 2018), each allowing for more modelling flexibility and ensuring the monotonicity of the survival function in different ways. More recently Marra & Radice (2020) proposed a generalised survival modelling framework which elegantly embeds the monotonicity of the survival function within the model design matrix by exploiting the properties of P-splines (see Section 3.3.2). We adopt this approach and thus describe it in the following in the context of transition-specific modelling. The model follows that described in Chapter 2 for a single marginal survival function. Here we adapt it to the context of transition-specific modelling.

Let $\mathcal{A} = \{(r,s) \mid r \neq s \in \mathcal{S}, q^{(rs)}(t_i) \neq 0\}$ be the set of transitions and N represent the sample size. For individual $i = 1,\dots,N$ and for $(r,s) \in \mathcal{A}$, let $H^{(rs)}(\cdot)$ be the cumulative hazard for the transition $r \rightarrow s$ defined in terms of the transition intensity $q^{(rs)}(\cdot)$ as $H^{(rs)}(t_i \mid \mathbf{x}_i; \beta^{(rs)}) = \int\limits_0^{t_i} q^{(rs)}(u \mid \mathbf{x}_i; \beta^{(rs)})du$. Then we will have a conditional transition-specific survival function denoted by $S^{(rs)}(t_i \mid \mathbf{x}_i; \beta^{(rs)}) = \exp\left\{-H^{(rs)}(t_i \mid \mathbf{x}_i; \beta^{(rs)})\right\} \in (0,1)$, where $\mathbf{x}_i$ represents a generic vector of patient characteristics that has an associated regression coefficient vector $\beta^{(rs)} \in \mathbb{R}^w$, where

w is the length of $\beta^{(rs)}$. A link-based additive transition-specific survival model can then be written as

$$g^{(rs)}\left\{S^{(rs)}(t_i \mid \mathbf{x}_i;\beta^{(rs)})\right\} = \eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)})), \tag{3.1}$$

where $g^{(rs)} : (0,1) \to \mathbb{R}$ is a monotone and twice continuously differentiable link function with bounded derivatives, hence invertible, which determines the scale of the analysis, $\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)})) \in \mathbb{R}$ is an additive predictor which includes a baseline function of time and several types of covariate effects and $\mathbf{f}(\beta^{(rs)})$ is a vector function of $\beta^{(rs)}$ through which the monotonicity required for the survival functions is imposed (see Section 3.3.2). Rearranging (3.1) yields $S^{(rs)}(t_i \mid \mathbf{x}_i;\beta^{(rs)}) = G^{(rs)}\left\{\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)}))\right\}$, where $G^{(rs)}$ is an inverse link function. Note that modelling directly on the survival scale implies a considerable advantage in this context (see Section 3.5). The cumulative transition-specific hazard is then $H^{(rs)}(t_i \mid \mathbf{x}_i;\beta^{(rs)}) = -\log\left[G^{(rs)}\left\{\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)}))\right\}\right]$ and the transition intensity function is defined as

$$q^{(rs)}(t_i \mid \mathbf{x}_i;\beta^{(rs)}) = -\frac{G^{(rs)\prime}\left\{\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)}))\right\}}{G^{(rs)}\left\{\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)}))\right\}}\frac{\partial\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)}))}{\partial t_i}, \tag{3.2}$$

where $G^{(rs)\prime}\left\{\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)}))\right\} = \partial G^{(rs)}\left\{\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)}))\right\}/\partial\eta_i^{(rs)}(t_i,\mathbf{x}_i;\mathbf{f}(\beta^{(rs)}))$. Table 2.2 in Chapter 2 lists the functions $g^{(rs)}$, $G^{(rs)}$ and $G^{(rs)\prime}$ available in the `R` package `GJRM`.

3.3.1 Additive predictor

Dropping the dependence on covariates and on parameters for the sake of simplicity, the additive predictor is defined as

$$\eta_i^{(rs)} = \beta_0^{(rs)} + \sum_{k=1}^{K^{(rs)}} s_k^{(rs)}(\mathbf{z}_{ki}), \quad i = 1,\ldots,n, \tag{3.3}$$

where $\beta_0^{(rs)} \in \mathbb{R}$ is an overall intercept, $\mathbf{z}_{ki}$ denotes the k^{th} sub-vector of the complete vector $\mathbf{z}_i$ and the $K^{(rs)}$ functions $s_k^{(rs)}(\mathbf{z}_{ki})$ denote effects which are chosen according to the type of covariate(s) considered. The observations made in Chapter 2 will then hold here as well. In particular, these functions can be expressed as a linear combination of basis functions $\mathbf{b}_k(\mathbf{z}_{ki}) = (b_{k1}^{(rs)}(\mathbf{z}_{ki}), \ldots, b_{kJ_k}^{(rs)}(\mathbf{z}_{ki}))^\mathsf{T}$ and regression coefficients $\mathbf{f}_k^{(rs)}(\beta_k^{(rs)}) = (f_{k1}^{(rs)}(\beta_{k1}^{(rs)}), \ldots, f_{kJ_k}^{(rs)}(\beta_{kJ_k}^{(rs)}))^\mathsf{T} \in \mathbb{R}^{J_k}$, that is $s_k^{(rs)}(\mathbf{z}_{ki}) = \mathbf{b}_k(\mathbf{z}_{ki})^\mathsf{T}\mathbf{f}_k^{(rs)}(\beta_k^{(rs)})$ (e.g., Wood, 2017). We can then write (3.3) compactly as $\eta_i^{(rs)} = \mathbf{Z}_i^{(rs)\mathsf{T}}\mathbf{f}^{(rs)}(\beta^{(rs)})$, where $\mathbf{Z}_i^{(rs)} = (1, \mathbf{b}_1(\mathbf{z}_{1i})^\mathsf{T}, \ldots, \mathbf{b}_{K^{(rs)}}(\mathbf{z}_{K^{(rs)}i})^\mathsf{T})^\mathsf{T}$ and $\mathbf{f}^{(rs)}(\beta^{(rs)}) = (\beta_0^{(rs)}, \mathbf{f}_1^{(rs)}(\beta_1^{(rs)})^\mathsf{T}, \ldots, \mathbf{f}_{K^{(rs)}}^{(rs)}(\beta_{K^{(rs)}}^{(rs)})^\mathsf{T})^\mathsf{T}$. The derivative with respect to time required in (3.2) can be expressed as $\partial\eta_i^{(rs)}(t_i, \mathbf{x}_i; \mathbf{f}^{(rs)}(\beta^{(rs)}))/\partial t_i = \mathbf{Z}_i^{(rs)}(t_i, \mathbf{x}_i)'^\mathsf{T}\mathbf{f}^{(rs)}(\beta^{(rs)})$ where, depending on the type of spline basis employed, $\mathbf{Z}_i(t_i, \mathbf{x}_i)' = \lim_{\varepsilon \to 0} \frac{\mathbf{Z}_i^{(rs)}(t_i+\varepsilon, \mathbf{x}_i) - \mathbf{Z}_i^{(rs)}(t_i-\varepsilon, \mathbf{x}_i)}{2\varepsilon}$ can be calculated either by a finite-difference method or analytically. Each $\beta_k^{(rs)}$ has an associated quadratic penalty $\lambda_k^{(rs)}\beta_k^{(rs)\mathsf{T}}\mathbf{D}_k^{(rs)}\beta_k^{(rs)}$, used in fitting, whose role is to enforce specific properties on the k^{th} function, such as smoothness, with matrix $\mathbf{D}_k^{(rs)}$ depending only on the choice of the basis functions. The smoothing parameter $\lambda_k^{(rs)} \in [0, \infty)$ controls the trade-off between fit and smoothness, and hence determines the shape of the estimated smooth function. The overall penalty can be defined as $\beta^{(rs)\mathsf{T}}\mathbf{S}_{\lambda^{(rs)}}^{(rs)}\beta^{(rs)}$, where $\mathbf{S}_{\lambda^{(rs)}}^{(rs)} = \mathrm{diag}(0, \lambda_1^{(rs)}\mathbf{D}_1^{(rs)}, \ldots, \lambda_{K^{(rs)}}^{(rs)}\mathbf{D}_{K^{(rs)}}^{(rs)})$ is a block diagonal matrix where each block is given by the k^{th} penalty, and where $\lambda^{(rs)} = (\lambda_1^{(rs)}, \ldots, \lambda_{K^{(rs)}}^{(rs)})^\mathsf{T}$ is the transition-specific overall smoothing parameter vector. Depending on the types of covariate effects one wishes to model, several definitions of basis functions are possible, e.g. thin plate, cubic and P- regression splines, tensor products, Markov random fields, random effects, Gaussian process smooths. These are handled automatically within the software proposed. We refer the reader to Section 3.7 for practical examples of the effects mentioned above and to Wood (2017) for the other available options.

3.3.2 Imposing monotonicity by means of SCOPs

When modelling life-history data through multistate processes, one is often interested in making statements in terms of the probabilities of transitioning from one state to another for specific combinations of risk-factors. In Section 3.5, it will be shown that we compute these by first extracting the transition-specific cumulative hazards at various time points. Direct modelling of the survival functions thus allows us to obtain the transition probabilities more cheaply, as we drop the intermediate step of having to first integrate the transition intensities. The only caveat is that one needs to ensure the survival functions are monotonically decreasing. Liu et al. (2018) propose to do this by means of a penalty applied to the hazard function such that the associated coefficient is iteratively doubled until the estimated hazard functions of all individuals are not negative. We employ a more theoretically founded approach. Indeed, in the proposed framework the properties of P-splines are exploited to elegantly embed the monotonicity within the construction of the survival functions themselves, while allowing for the flexible modelling of the time effect.

Let $s^{(rs)}(t_i) = \sum_{j=1}^{J^{(rs)}} f_j^{(rs)}(\beta_j^{(rs)}) b_j^{(rs)}(t_i)$, where the $b_j^{(rs)}(\cdot)$ are B-spline basis functions of at least second order built over the interval $[a,b]$, based on equally spaced knots, and the $f_j^{(rs)}(\beta_j)^{(rs)}$ are spline coefficients. Given the link functions listed in Table 2.2, we need $s^{(rs)'}(t_i) \geq 0$. Eilers & Marx (1996) combined B-spline basis functions with discrete penalties in the basis coefficients to produce the popular P-spline smoothers. Then Pya & Wood (2015) proposed shape constrained P-splines through a mildly nonlinear extension of these P-splines, with corresponding novel discrete penalties, thus allowing the development of efficient and stable model estimation frameworks, such as the one proposed. In particular, a sufficient condition for $s^{(rs)'}(t_i) \geq 0$ over $[a,b]$ is that $f_j^{(rs)}(\beta_j^{(rs)}) \geq f_{j-1}^{(rs)}(\beta_{j-1}^{(rs)}), \forall j$. Indeed, given a function $\eta(x) = a_0 + \sum_{j=1}^{m} a_j B_j(x,q)$, where $B_j(x,q)$ are the bases for a $(q+1)^{th}$ order B-spline, m is the number of basis functions, $\partial\eta(x)/\partial x = \frac{1}{h}\sum_{j=1}^{m-1}(a_{j+1} - a_j)B_j(x,q-1)$ with h the distance between equally spaced knots and so $a_{j+1} \geq a_j$ implies $\partial\eta(x)/\partial x \geq 0$ since $B_j(x,q-1) \geq 0$ (Leitenstorfer & Tutz, 2006). Such condition can be imposed by defining the vector function $\mathbf{f}^{(rs)}(\beta^{(rs)}) = \Sigma\left\{\beta_1^{(rs)}, \exp(\beta_2^{(rs)}), \ldots, \exp(\beta_{J^{(rs)}}^{(rs)})\right\}^{\mathsf{T}}$, where $\Sigma[\imath_1, \imath_2] = 0$ if $\imath_1 < \imath_2$ and $\Sigma[\imath_1, \imath_2] = 1$ if $\imath_1 \geq \imath_2$, with $\imath_1$ and $\imath_2$ denoting the row and column entries of Σ, and $\beta^{(rs)\mathsf{T}} = (\beta_1^{(rs)}, \beta_2^{(rs)}, \ldots, \beta_{J^{(rs)}}^{(rs)})$ is the parameter vector to estimate. Crucially, in practice Σ is absorbed into the design matrix containing the B-spline basis functions $\mathbf{Z}$, hence allowing the constraint to be elegantly embedded within the construction of the model design matrix itself. Finally, in a smoothing context, we are interested in having a penalty on the smooth function to control its ”wiggliness”. Eilers & Marx (1996) introduced the notion of directly penalising the difference in the basis coefficients of a B-splines basis, which is used with a relatively large number of basis functions to avoid underfitting. The adaptation to the shape-constrained case is straightforward as it implies penalising the squared differences between adjacent $\beta_j^{(rs)}$, starting from $\beta_2^{(rs)}$, using $\mathbf{D}^{(rs)} = \mathbf{D}^{(rs)*\mathsf{T}}\mathbf{D}^*$ where $\mathbf{D}^{(rs)*}$ is a $(J^{(rs)}-2) \times J^{(rs)}$ matrix made up of zeros except that $\mathbf{D}^{(rs)*}[\imath, \imath+1] = -\mathbf{D}^{(rs)*}[\imath, \imath+2] = 1$ for $\imath = 1, \dots, J^{(rs)}-2$. The penalty is zeroes when all the $\beta_j^{(rs)}$ after $\beta_1^{(rs)}$ are equal so that the $f_j^{(rs)}(\beta_j^{(rs)})$ form a uniformly increasing sequence and $s^{(rs)}(t_i)$ is an increasing straight line. As a result, the proposed penalty shares the basic feature of smoothing towards a straight line, but in a manner that is computationally convenient for constrained smoothing.

3.4 Estimation

Since each likelihood contribution refers to a specific transition only and every transition is exactly observed if and only if it occurs, it can be shown (see Supplementary Material B.2) that the overall model log-likelihood can be broken down into the sum of the log-likelihoods associated with each transition, which are functions only of the parameters relating to that transition, i.e. $\ell(\theta) = \sum_{(r,s)\in\mathcal{A}} \ell^{(rs)}(\beta^{(rs)})$, where $\theta = \{\beta^{(rs)} \mid (r,s) \in \mathcal{A}\}$ is an overall model parameter vector. Re-writing the log-likelihood in this way, rather than as a sum of contributions associated with each observation time, is more convenient as it breaks down the estimation task into a number of traditional survival problems, one for each transition. It is precisely to each of these transition-specific models that the framework developed in Eletti et al. (2022) is applied. Briefly, as the model allows for a high degree of flexibility, to prevent over-fitting, the log-likelihood is augmented with a penalty term $\ell_p^{(rs)}(\beta^{(rs)}) = \ell^{(rs)}(\beta^{(rs)}) - \frac{1}{2}\beta^{(rs)\mathsf{T}}\mathbf{S}_{\lambda^{(rs)}}^{(rs)}\beta^{(rs)}$ where $\mathbf{S}_{\lambda^{(rs)}}^{(rs)}$ is an overall penalty term defined in Section 3.3. The estimation framework then combines a carefully structured trust region algorithm which uses the analytical expressions of the gradient and Hessian of the log-likelihood and properly chosen starting values with a general automatic multiple smoothing parameter selection algorithm based on an approximate AIC measure.

3.5 Prediction on the transition probabilities scale

While estimation can be carried out entirely by-passing the computation of the transition probabilities, one is often interested in making statements in terms of the probability of transitioning from one state to another given a specific combination risk-factors. We choose the simulation-based approach proposed in Fiocco et al. (2008), which we briefly describe in the following. Let r be the starting state, entered at time $t_r = 0$, and $t_{\max}$ the maximum follow-up time. Then

- Let $\mathcal{B}$ be the set of states that can be reached from state r. If $\mathcal{B}$ is empty, stop. Otherwise, for $s \in \mathcal{B}$, let $H^{(rs)}(t)$ be the cumulative transition-specific hazard function for transition $r \rightarrow s$ and $H^{(r\cdot)}(t) = \sum\limits_{s \in \mathcal{B}} H^{(rs)}(t)$ refer to the event of leaving state r.

- Sample t^* from $H^{(r\cdot)}(t) - H^{(r\cdot)}(t_r)$. This refers to the conditional distribution of leaving state r given that the process is known to be in state r until time t_r thus ensuring that the sampled time $t^* > t_r$.

- If $t^* < t_{max}$, select the next state s with probability $dH^{(rs)}(t^*)/dH^{(r\cdot)}(t^*)$, which provides a weight for the specific transition $r \rightarrow s$ out of state r for each $s \in \mathcal{B}$ at the given time t^*, and set the new starting points for the next iteration, $r = s$ and $t_r = t^*$. Otherwise, stop: a full path through the process was obtained.

This is repeated to obtain M paths through the multistate model and to compute the transition probabilities by counting the number of paths for which each event occurred. The approach is implemented in the function `mssample()` of the `R` package `mstate` and is straightforward to use given the estimated transition-specific cumulative hazards, provided by the `hazsurv.plot()` function of the `R` package `GJRM`.

It should be noted that computing the transition probability matrix is a non-trivial problem, as it entails solving the so-called Kolmogorov forward differential equations, which in general do not have a closed-form solution. The advantage of the simulation-based method proposed here is that it allows us to circumvent this issue, while retaining a great degree of generality. In fact, it applies to both Markov and semi-Markov processes, with any number of states and transition types.

3.6 Inference

One view of the smoothing process is that the penalty employed during fitting imposes the belief that the true function is more likely to be smooth than wiggly. This belief can be expressed in a Bayesian manner through the form of a prior distribution on $\beta^{(rs)}$, i.e. $f_{\beta^{(rs)}} \propto \exp\left\{-\beta^{(rs)\mathsf{T}}\mathbf{S}^{(rs)}_{\lambda^{(rs)}}\beta^{(rs)}/2\right\}$. This leads to the Bayesian large sample approximation $\beta^{(rs)} \dot{\sim} \mathcal{N}(\widehat{\beta}^{(rs)}, \mathbf{V}_{\beta^{(rs)}})$, where $\mathbf{V}_{\beta^{(rs)}} = -\mathbf{H}_p(\widehat{\beta}^{(rs)})^{-1}$; using $\mathbf{V}_{\beta^{(rs)}}$ gives close to across-the-function frequentist coverage probabilities because it accounts for both sampling variability and smoothing bias, a feature that is particularly relevant at finite sample sizes (Wood et al., 2016). Following Pya & Wood (2015), we then consider the Taylor series expansion of $\mathbf{f}^{(rs)}(\beta^{(rs)})$ around $\mathbf{f}^{(rs)}(\tilde{\beta}^{(rs)})$. This gives $\mathbf{f}^{(rs)}(\beta^{(rs)}) - \mathbf{f}^{(rs)}(\tilde{\beta}^{(rs)}) \approx \text{diag}(\mathbf{E}^{(rs)})\big(\beta^{(rs)} - \tilde{\beta}^{(rs)}\big)$, where $\mathbf{E}^{(rs)}[k_{j_k}] = 1$ if $f^{(rs)}_{k_{j_k}}(\beta^{(rs)}_{k_{j_k}}) = \beta_{k_{j_k}}$ and $\exp(\beta^{(rs)}_{k_{j_k}})$ otherwise, showing that $\mathbf{f}^{(rs)}(\beta^{(rs)}) - \mathbf{f}^{(rs)}(\tilde{\beta}^{(rs)})$ is approximately a linear function of $\beta^{(rs)}$. Combining this with the result above we have that $\mathbf{f}^{(rs)}(\beta^{(rs)}) \dot{\sim} \mathcal{N}(\mathbf{f}^{(rs)}(\tilde{\beta}^{(rs)}), \mathbf{V}_{\mathbf{f}^{(rs)}(\beta^{(rs)})})$ where $\mathbf{V}_{\mathbf{f}^{(rs)}(\beta^{(rs)})} = \text{diag}(\mathbf{E}^{(rs)})\mathbf{V}_{\beta^{(rs)}}\text{diag}(\mathbf{E}^{(rs)})$, since linear functions of normally distributed random variables follow normal distributions. Confidence intervals for linear functions of the model coefficient can then be obtained using this result. P-values for the smooth components in the model are derived by adapting the result discussed in Wood (2017) and using $\mathbf{V}_{\mathbf{f}^{(rs)}(\beta^{(rs)})}$ as covariance matrix. For nonlinear functions of the model coefficients, e.g. the transition-specific cumulative hazard functions, instead, the intervals can be conveniently obtained by posterior simulations, hence avoiding computationally expensive parametric bootstrap or frequentist approximations, for instance.

3.7 Primary breast cancer modelling case study

To illustrate what the proposed approach adds compared to the existing literature, we consider the case study described in Crowther & Lambert (2017) which is based on data from 2892 patients with primary breast cancer for which the time to relapse and/or the time to death is known. See, e.g., Sauerbrei et al. (2007) for further details on the *Rotterdam Breast Cancer Study* from which the data originated. The code used to produce this analysis can be found in the public repository `https://github.com/AlessiaEletti/ContinObsMultistateProcesses`.

All patients begin in the initial post-surgery state, 1518 patients experience relapse, 195 die without relapse and 1075 die after experiencing relapse. A Markov illness-death model (IDM, see Figure B.1 in Supplementary Material B.3) will thus be used to model the data. As an aside, note that an attempt assuming a semi-Markov process was also made but this was not supported by the data according to the AIC values found for the fitted models.

As there are three transitions in the assumed IDM, three survival models will be fitted. For transitions which can occur only given that another transition has already taken place, i.e. the transition $2 \to 3$ in this case, one must account for the fact that the patient is at risk only after entering the new starting state, i.e. state 2. As long as this is done, each transition can be treated as a separate survival problem. The time at which the individual entered state 2 thus becomes the left-truncation time for the new transition $2 \to 3$.

To clarify how the separate estimations are carried out, recall that longitudinal survival data are characterised by multiple observations through time of at least one quantity of interest for the same individual. Typically the data are formatted in the so-called stacked (or long) form, i.e. each row represents a single time point per subject. In particular, each subject will have at least v rows, where v is the number of possible transitions exiting the initial state. Here, $v = 2$ as there are two ways of exiting state 1, i.e. going in state 2 or 3. A start and a stop time will then indicate, respectively, the first time after which the patient becomes at risk of the given transition and the time at which the transition itself occurred. The start time for transitions exiting the first state is 0, as is usually the case here. If the patient transitions to an intermediate state, u rows will be added, where u is the number of transitions exiting the intermediate transition state reached. Here, $u = 1$, as the only possible transition out of state 2 is $2 \rightarrow 3$, where 3 is an absorbing state. When estimating $q^{(12)}(\cdot)$, all of the rows relating to this transition are included in the estimation. Since every patient will at least have one row for each transition exiting the first state, this implies that the entire population is included. The same is true for $q^{(13)}(\cdot)$, for which the rows relating to the $1 \rightarrow 3$ transition will be used for estimation. The two resulting separate datasets can then be treated as traditional survival data with uncensored and right censored observations and with the event of interest given by the transition to the new state, i.e. state 2 for the former and state 3 for the latter. When estimating $q^{(23)}(\cdot)$, only individuals who have transitioned to state 2 at some point are included in the estimation. The data are then treated as traditional survival data with left-truncated uncensored and left-truncated right censored observations and where the event of interest is the transition to the absorbing state 3. We refer the reader to Supplementary Material B.4 for further details on the format of the data in this setting.

The dataset contains information on the age of the patient at primary surgery (in years), tumour size (divided into 3 classes: ≤ 20, $20-50$ and > 50 mm), number of positive nodes, progesterone levels (in fmol/L) and whether or not the patient was on hormonal therapy. These are all included as covariates. We then include a time-dependent effect for the progesterone level, as this has been found to be relevant in the reference paper, and include age, the progesterone level and the number of positive nodes nonlinearly, as supported by existing literature. Importantly, our framework allows for the exploration of these effects in a more general and flexible manner than previously possible in the literature thanks to the use of splines. In contrast, for instance, Sauerbrei & Royston (1999) modelled the number of positive nodes nonlinearly by using fractional polynomials with the degrees set heuristically. Similarly, in Crowther & Lambert (2017) the time-dependent effect is captured by a single interaction coefficient between time and the progesterone level . In particular, for $(r,s) \in \{(1,2),(1,3),(2,3)\}$, we specify the transition-specific models

$$\begin{aligned}\eta_i^{(rs)}(t_i, \mathbf{x}_i; \mathbf{f}(\beta^{(rs)})) = \beta_0^{(rs)} + s_0^{(rs)}(\log(\mathrm{t}_i)) + \beta_1^{(rs)} \mathrm{I}_{\mathrm{size}_i=20-50} + \beta_2^{(rs)} \mathrm{I}_{\mathrm{size}_i>50} + \beta_3^{(rs)} \mathrm{hormon}_i \\ + s_1^{(rs)}(\mathrm{age}_i) + s_2^{(rs)}(\mathrm{nodes}_i) + s_3^{(rs)}(\mathrm{pr}_i) + s_4^{(rs)}(\log(\mathrm{t}_i), \mathrm{pr}_i),\end{aligned}$$

where $s_0^{(rs)}(\log(\mathrm{t}_i))$ is a monotonic P-spline of the logarithm of time which ensures the monotonicity of the survival function associated with this transition, as explained in Section 3.3.2; $s_1^{(rs)}(\mathrm{age}_i)$, $s_2^{(rs)}(\mathrm{nodes}_i)$ and $s_3^{(rs)}(\mathrm{pr}_i)$ are thin-plate splines, while $s_4^{(rs)}(\log(\mathrm{t}_i), \mathrm{pr}_i)$ is a pure smooth interaction between time and the progesterone level, i.e. a time-dependent effect. In regard to the penalty associated with a nonlinear term, e.g., $s_1(\mathrm{age}_i)$, this takes the form of the quadratic penalty defined above with $\mathbf{D}_k$ given by the integrated square second derivative of the basis functions, i.e. $\int \mathbf{d}_k(z_k)\mathbf{d}_k(z_k)^\mathsf{T} dz_k$ with the j_k^{th} element of $\mathbf{d}_k(z_k)$ defined as $\partial^2 b_{kj_k}(z_k)/\partial z_k^2$. The penalty associated with the time-dependent effect is, instead, more complex as it entails combining two penalties (see Wood, 2017, Chapter 5). Finally, note that for parametric effects the spline representation simplifies to $s^{(rs)}(\mathrm{hormon}_i) = \beta_3^{(rs)}\mathrm{hormon}_i$. No penalty is typically assigned to parametric effects, hence the associated quadratic penalty is $D = 0$. Note that in cases such as those in which the categorical variable has many levels with some with few observations, it may be advisable to set the penalty as the identity matrix. In this way, a ridge penalty is imposed and it may help avoid that the parameters associated with the more sparse categories are weakly or nonidentified.

The estimated covariate effects for each transition are reported in Table 3.1. For the first transition, for instance, they are in line with our expectations: the larger the size of the tumor the higher the risk of experiencing relapse, while hormonal therapy has a beneficial effect. In Figure 3.1 we report the estimated transition intensities with their 95% confidence intervals as functions of time for a 54 year old patient with tumour size $\geq$ 50 mm, 10 positive nodes, progesterone level of 3

		Estimate	Std. Error	Pr(> $\lvert z \rvert$)
Transition 1 → 2	(Intercept)	-10.630	1.198	$< 1e-4$
	size20-50	0.284	0.059	$< 1e-4$
	size>50	0.477	0.089	$< 1e-4$
	hormon	-0.318	0.085	$2e-4$
Transition 1 → 3	(Intercept)	-12.543	2.585	$< 1e-4$
	size20-50	0.153	0.162	0.344
	size>50	0.390	0.236	0.098
	hormon	-0.135	0.236	0.567
Transition 2 → 3	(Intercept)	-2.915	1.023	0.004
	size20-50	0.139	0.072	0.053
	size>50	0.259	0.101	0.010
	hormon	-0.015	0.098	0.881

Table 3.1: Model estimates, standard errors and p-values for the three transitions. Note: the full output of the R model summary is reported. It is clear that the p-values relating to the categorical variables do not have a meaningful interpretation, these are only included for completeness.

and under hormonal therapy. We find, for instance, that the risk of experiencing relapse for this profile increases for approximately 2.5 years after surgery, then it decreases and plateaus over time. In Figure 3.2 we report the plots of the smooths and of the tensor interaction for the transition *health* → *relapse*. These show that the data particularly support nonlinear effects for the age and the number of positive nodes. For instance, the latter exhibits an increasing trend up to about 12 nodes, followed by a plateau. The time-dependence of the progesterone level effect is also clear from the surface representing the smooth interaction, with low levels of progesterone associated with a decreasing risk of experiencing relapse over time and, conversely, high levels of progesterone associated with an increasing trend for the risk of experiencing relapse over time. Any additional complexity not supported by the data is then suppressed automatically through the estimation of the smoothing parameter, rather than requiring the user to make restrictive and potentially arbitrary choices a priori. This can be seen in the plots of the smooths of the remaining two transitions, reported in Figures B.2 and B.3 of Supplementary Material B.3. The plot of the smooth of age for the *health* → *death* transition, for instance, shows that the data actually supported a linear effect for this term.

As mentioned above, interest usually lies in making statements in terms of the

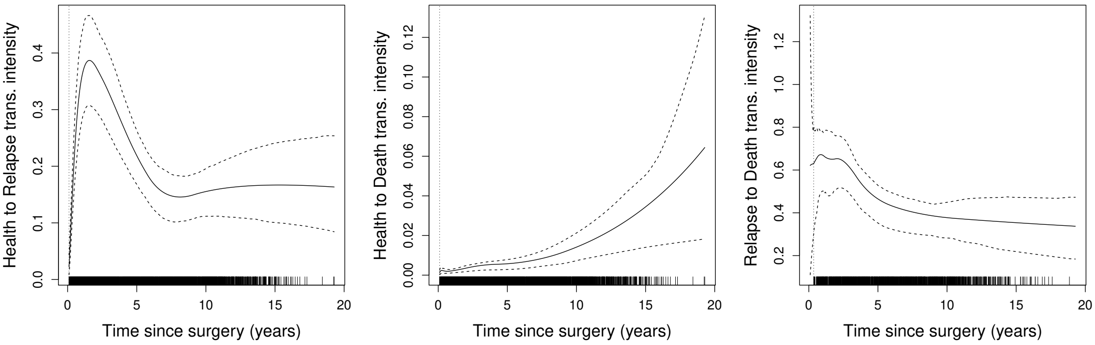

Figure 3.1: Fitted transition intensities and 95% confidence intervals (CIs) for a 54 year old patient under hormonal therapy with tumour size $\geq$ 50 mm, 10 nodes and progesterone level of 3, over 20 years. The vertical dashed line marks the smallest observed time: the transition intensities estimated at smaller times are extrapolations, thus explaining the wide CIs in the first section of the third plot. The width of the CIs in the final portion of the middle plot can be explained by the scarcity of observations in the final times, as shown by the rug plot. The width of the confidence intervals should also be related to the different range of values in each plot.

probabilities of transitioning between states thus, in Figure 3.3, we report stacked transition probability plots. Representing the probabilities in this stacked manner is a common way of quickly providing an overview of how risk evolves over time, however the uncertainty of the estimates cannot be easily portrayed. For this reason, in Figure 3.4, we report the predicted probabilities with their 95% confidence intervals for the individual corresponding to the top-left panel, i.e. a 54 year old patient under hormonal therapy, progesterone level of 3, 20 positive nodes and tumour size $\leq$ 20 mm. Note that the computation of the transition probabilities already entails a simulation, thus the process of obtaining confidence intervals for it will result in two nested simulations. The computational burden of this is not prohibitively high, however. Here, they are obtained by using 100 simulated cumulative hazards for each of the three transitions, over 100 distinct time points, and $M = 10000$ simulated paths through the process, which is a larger number of paths than typically needed. This required approximately 37 minutes using a laptop with Windows 10 (2.20 GHz processor, 16 GB RAM, 64-bit). Details on this, on how the model fitting is carried out and how the plots reported in this section were obtained can be found in

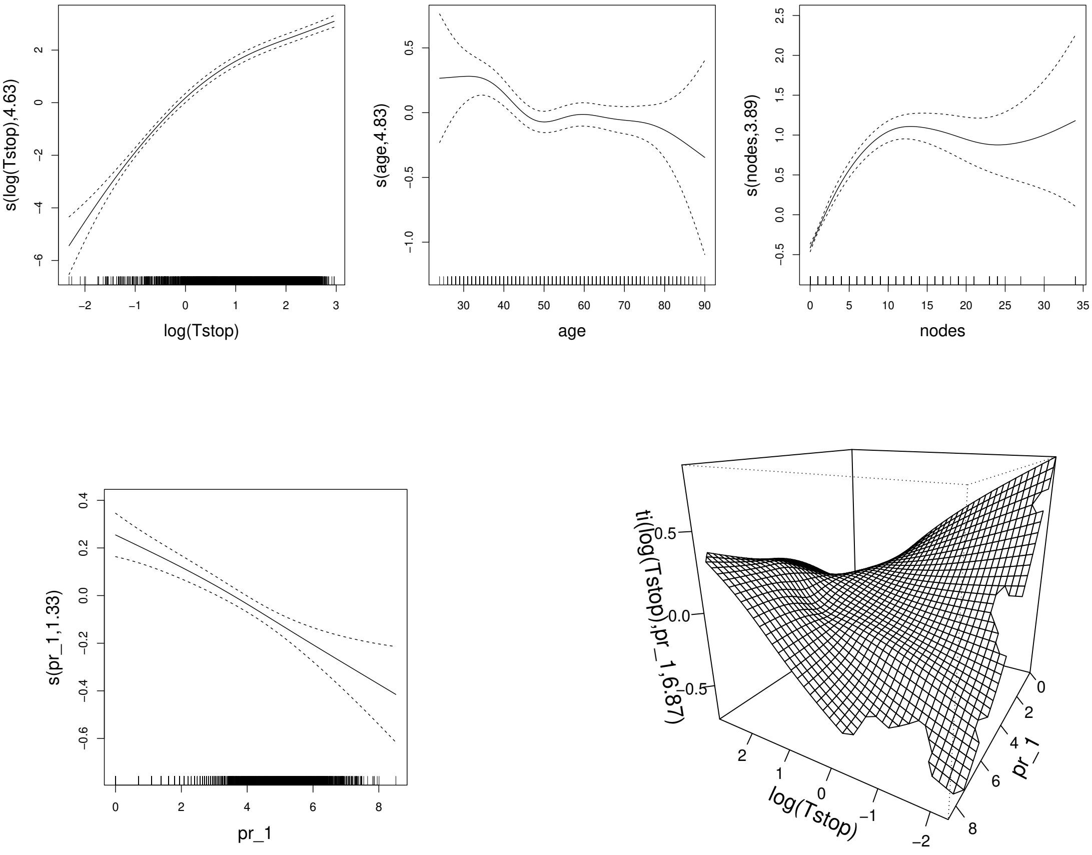

Figure 3.2: Smooth of log-time (top left), smooth of age (top middle), smooth of the number of positive nodes (top right), smooth of the progesterone level (bottom left) and smooth interaction between log-time and progesterone level (bottom right) for the transition $health \rightarrow relapse$.

Supplementary Material B.3.

3.8 Discussion

In this work we show how one can use existing tools to flexibly model multistate survival processes relating to continuously observed life-history data. In particular, we consider the survival estimation framework described in Eletti et al. (2022) and implemented in the `R` package `GJRM` which allows us to model virtually any type of covariate effect, including time-dependent ones. Direct modelling of the survival functions implies a considerable gain in efficiency when it comes to computing the transition probabilities of interest, which in turn are obtained through

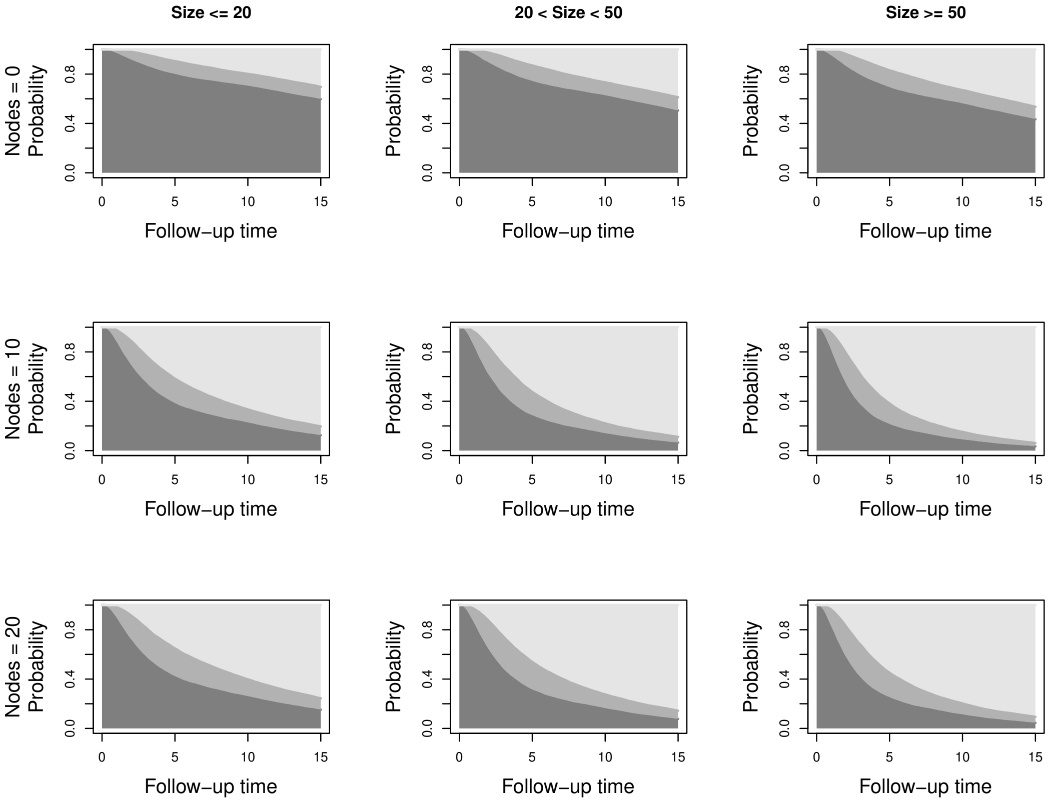

Figure 3.3: Stacked representation of estimated transition probabilities (dark grey: post-surgery; grey: relapse; light grey: death) for each combination of nodes (0, 10 and 20) and tumour sizes (≤ 20, $(20, 50)$ and ≥ 50) considered in a 54 year old patient under hormonal therapy with progesterone level of 3.

a simulation-based approach able to support any type of multistate process. Efficient modelling on the survival scale is achieved through shape constrained P-splines, developed by Pya & Wood (2015), building upon the work done in Eilers & Marx (1996). We exemplify our approach on data from the *Rotterdam Breast Cancer Study* and provide the code used for the analysis in the public repository `https://github.com/AlessiaEletti/ContinObsMultistateProcesses`.

With regard to directions of future work, we are interested in integrating the computation of the transition probabilities and the extraction of its confidence intervals directly within the `GJRM` package, so as to minimise the amount of user-written code needed and thus further simplify the use of these models by the practitioner. Similarly, for the visualisation tools available for the estimated transition probabilities. As the Markov assumption is quite common, we are also interested in implementing the method based on the numerical solution of the differential equations tying the transition probabilities to the intensities as well as to

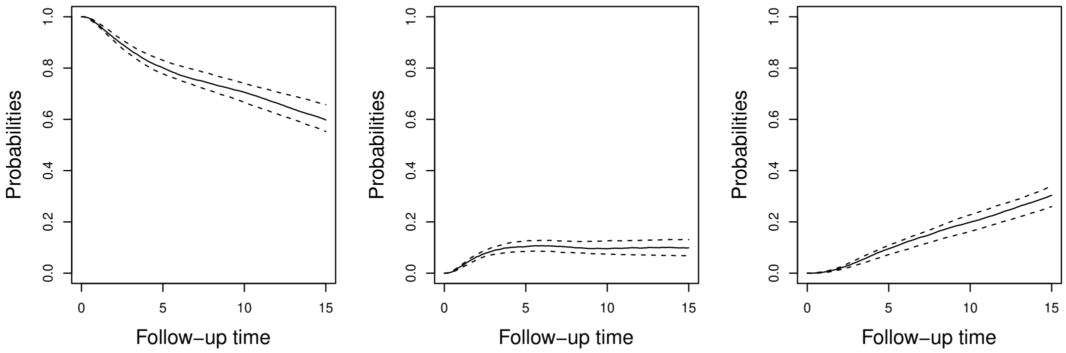

Figure 3.4: Estimated transition probabilities (left: post-surgery; middle: relapse; right: death) for the top-left pane in Figure 3.3.

implement our own simulation-based approach within the `GJRM` package, so that the user has all necessary instruments in the same place and the need for user-written code is reduced to the minimum.

Chapter 4

A General Estimation Framework for Multi-State Markov Processes with Flexible Specification of the Transition Intensities

4.1 Introduction

With the increase in the availability of longitudinal survival data, continuous-time multi-state Markov models have established themselves as powerful tools to model the progression of a phenomenon, while accounting for background information recorded for each individual throughout the follow-up period; see Yiu et al. (2017), Williams et al. (2020) and Gorfine et al. (2021) for some examples. In many applications, a non-homogeneous Markov process is assumed, i.e. the risks of moving across states depend on the current state and on time. This is typically addressed by employing parametric functional forms for the transition hazards, but some examples of more flexible (e.g., spline-based) specifications can be found as well (e.g., Cook & Lawless, 2018; Joly et al., 2002; Mariano Machado et al., 2021; Titman, 2011; Van Den Hout, 2016).

Constant monitoring of the progression of a phenomenon of interest is often not possible since it may be too expensive or altogether not feasible due to the nature of the event of interest. When this is the case, the process is only observed at a fixed set of times and is thus said to be intermittently observed or interval-censored. The lack of knowledge of the times in which the transitions occurred represents a methodological challenge. The literature on the subject is vast, however existing computational methods for fitting non-homogeneous multi-state Markov models in such setting have mainly been based on the estimation approach developed by Kalbfleisch & Lawless (1985), which relies on approximating the information matrix using the analytical score of the log-likelihood. The advantage of this method is that, in principle, it permits a great degree of generality by allowing for any number of states, forward and backward transitions, and any type of functional form for the transition intensities. However, only simpler models are supported in practice, with the most commonly used implementation provided via the R package msm (Jackson, 2019). Yet, convergence failures occur when the number of states and covariates increases; this can be attributed to the absence of the analytical information matrix which would provide valuable exact curvature information exploitable in model fitting. Based on Kalbfleisch & Lawless (1985), Mariano Machado et al. (2021) introduced an approach that allows to fit models that are more flexible than those considered in msm. The authors carried out comparisons with two currently available implementations and found their estimation approach to be superior in terms of empirical performance and modelling flexibility. However, Mariano Machado et al. (2021) only provided a bespoke code for the simple and well-known three-state Illness-Death Model (IDM). When applied to the CAV study, this approach was found to be too restrictive for the estimation of covariate effects. Another implementation is given via the R package nhm (Titman, 2023). Here, the transition probabilities are obtained as numerical solutions of the differential equations that ties them to the transition intensities (Titman, 2011). This package is as general as msm but it additionally supports the use of an unpenalized smooth function of time. When applied to the CAV data, convergence could only be achieved for a model with log-linear effects. For the cognitive study, no model could be fitted.

To widen significantly the scope of non-homogeneous multi-state Markov models, we propose an analytical expression for the local curvature information of the transition probability matrix. This allows us to introduce a modelling framework which is general and flexible, and that is applicable to far more complex empirical problems than those previously explorable in the literature. Specifically, the proposal allows for any type of multi-state process, with several states and various combinations of observation schemes (e.g., intermittent, exactly observed, censored), and for the transition intensities to be flexibly modelled through additive predictors. Parameter estimation is carried out by adapting to this context the stable and efficient estimation algorithm of Marra & Radice (2020) which can fully exploit the newly derived analytical observed information matrix. To allow for reproducible and transparent research, the framework is implemented in the `R` package `flexmsm` (Eletti et al., 2023a) which is very easy and intuitive to use; for instance, time and covariate effects of multi-state Markov models can be flexibly specified using the same syntax as that for generalised additive models in `R` (Wood, 2017).

In Section 4.2, we introduce the mathematical setting of multi-state Markov models and describe the regression spline-based approach employed for modelling the transition intensities. The penalised log-likelihood is presented in Section 4.3, while parameter estimation and how this is intertwined with the problem of computing the transition probabilities from the transition intensities are discussed in Section 4.4. This section also presents the closed-form expressions for the transition probability matrix and its first and second derivatives, which are needed to compute the analytical likelihood, gradient and Hessian exploited in estimation.

Section 4.5 describes how inference is carried out. Section 4.6 illustrates the potential of the proposal via a classical study, based on the IDM, that aims at modelling the onset of cardiac allograft vasculopathy, and a more complex one, about cognitive decline, which requires the use of a five-state process with both forward and backward transitions as well as an absorbing death state. Section 4.7 concludes the paper with some directions of future research. On-line Supplementary Materials C.1, C.2 and C.3 provide details on the log-likelihood contributions, the `R` package `flexmsm` and the algorithm employed for parameter estimation. Supplementary Material C.4 illustrates the empirical effectiveness of the proposal via two simulation studies. Supplementary Material C.5 contains a list of the mathematical symbols used and their meaning.

4.2 Multi-state processes with flexible transition intensities

We recover part of the notation introduced in Chapter 3 for multi-state processes and adapt it to our current setting, i.e. intermittently observed Markov processes. Let $\{Z(t), t > 0\}$ be a continuous-time Markov process, $\mathcal{S} = \{1, 2, \dots, C\}$ its discrete state space, where C is the total number of states, and $\mathcal{A} = \{(r, r') \mid r \neq r' \in \mathcal{S}, \exists\ r \to r'\}$ the set of transitions. The transition intensity function, i.e. the instantaneous rate of transition to a state r' for an individual who is currently in another state r, is defined as follows

$$q^{(rr')}(t) = \lim_{h \downarrow 0} \frac{P(Z(t+h) = r' \mid Z(t) = r)}{h}, \quad r \neq r',$$

with $q^{(rr')}(t) = 0$ if r is an absorbing state and $q^{(rr)}(t) = -\sum_{r \neq r'} q^{(rr')}(t)$. The matrix with (r, r') element given by $q^{(rr')}(t)$ for every $r, r' \in \mathcal{S}$ is called transition intensity matrix or generator matrix and can be denoted with $\mathbf{Q}(t)$. Similarly, the transition probability matrix associated with the time interval (t, t') is defined as the matrix with (r, r') element given by $p^{(rr')}(t, t') = P(Z(t') = r' \mid Z(t) = r)$ and can be denoted with $\mathbf{P}(t, t')$. Here, we assume a time-dependent process as opposed to the rather restrictive time-homogeneous process (i.e., $\mathbf{Q}(t) = \mathbf{Q}\ \forall t > 0$) often adopted in the literature for mathematical convenience.

The intensity for transition $r \rightarrow r'$, with $r \neq r'$, is generally represented using the proportional hazards specification, where the baseline intensity is typically specified using the exponential or Gompertz distribution (Van Den Hout, 2016). A more flexible representation for the transition intensity is

$$q^{(rr')}(t_\iota) = \exp\left[\eta_\iota^{(rr')}(t_\iota, \mathbf{x}_\iota; \beta^{(rr')})\right], \tag{4.1}$$

where t_ι and $\mathbf{x}_\iota$ are the time and the vector of characteristics for observation ι respectively, $\beta^{(rr')}$ is the associated regression coefficient vector and $\eta_\iota^{(rr')}(t_\iota, \mathbf{x}_\iota; \beta^{(rr')}) \in \mathbb{R}$ is an additive predictor, discussed in detail in the following section, which includes a baseline smooth function of time and several types of covariate effects. Note that, in contrast to the settings of Chapters 2 and 3, here the additive predictor is defined on the log-intensities scale, making unnecessary the monotonicity constraint introduced for the former. It follows that t_ι and $\mathbf{x}_\iota^\mathsf{T}$ can be treated in the same way.

4.2.1 Additive predictor

For simplicity, the dependence on covariates and parameters has been dropped when discussing the construction of $\eta_\iota^{(rr')}$. Also, since t_ι can be treated as a covariate, we define the overall vector $\tilde{\mathbf{x}}_\iota = (t_\iota, \mathbf{x}_\iota^\mathsf{T})^\mathsf{T}$.

An additive predictor allows for various types of covariate effects and is defined as

$$\eta_\iota^{(rr')} = \beta_0^{(rr')} + \sum_{k=1}^{K^{(rr')}} s_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota}), \quad \iota = 1, \ldots, \check{n}, \tag{4.2}$$

where $\check{n}$ is the sample size, $\beta_0^{(rr')} \in \mathbb{R}$ is an overall intercept, $\tilde{\mathbf{x}}_{k\iota}$ denotes the k^{th} sub-vector of the complete vector $\tilde{\mathbf{x}}_\iota$ and the $K^{(rr')}$ functions $s_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota})$ represent effects which are chosen according to the type of covariate(s) considered. For example, if we were interested in modelling a time-dependent effect of the covariate age_ι, then $\tilde{\mathbf{x}}_{\iota k}$ would be the vector $(age_\iota, t_\iota)^\mathsf{T}$ and $s_k^{(rr')}(age_\iota, t_\iota)$ the corresponding joint effect. Each $s_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota})$ can be represented as a linear combination of $J_k^{(rr')}$ known basis functions $\mathbf{b}_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota}) = \left(b_{k1}^{(rr')}(\tilde{\mathbf{x}}_{k\iota}),\dots,b_{kJ_k^{(rr')}}^{(rr')}(\tilde{\mathbf{x}}_{k\iota})\right)^\mathsf{T}$ and regression coefficients $\beta_k^{(rr')} = \left(\beta_{k1}^{(rr')},\dots,\beta_{kJ_k^{(rr')}}^{(rr')}\right)^\mathsf{T} \in \mathbb{R}^{J_k^{(rr')}}$, that is $s_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota}) = \mathbf{b}_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota})^\mathsf{T}\beta_k^{(rr')}$ (e.g., Wood, 2017). The above formulation implies that the vector of evaluations $\left\{s_k^{(rr')}(\tilde{\mathbf{x}}_{k1}),\dots,s_k^{(rr')}(\tilde{\mathbf{x}}_{k\check{n}})\right\}^\mathsf{T}$ can be written as $\tilde{\mathbf{X}}_k^{(rr')}\beta_k^{(rr')}$, where $\tilde{\mathbf{X}}_k^{(rr')}$ is the design matrix whose ι^{th} row is given by $\mathbf{b}_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota})^\mathsf{T}$ for $\iota = 1,\dots,\check{n}$. This allows the predictor in equation (4.2) to be written as $\eta^{(rr')} = \beta_0^{(rr')}\mathbf{1}_{\check{n}} + \tilde{\mathbf{X}}_1^{(rr')}\beta_1^{(rr')} + \dots + \tilde{\mathbf{X}}_{K^{(rr')}}^{(rr')}\beta_{K^{(rr')}}^{(rr')}$, where $\mathbf{1}_{\check{n}}$ is an $\check{n}$-dimensional vector made up of ones. This can also be represented in a more compact way as $\eta^{(rr')} = \tilde{\mathbf{X}}^{(rr')}\beta^{(rr')}$, where $\tilde{\mathbf{X}}^{(rr')} = (\mathbf{1}_n, \tilde{\mathbf{X}}_1^{(rr')},\dots,\tilde{\mathbf{X}}_{K^{(rr')}}^{(rr')})$ and $\beta^{(rr')} = \left(\beta_0^{(rr')\mathsf{T}},\beta_1^{(rr')\mathsf{T}},\dots,\beta_{K^{(rr')}}^{(rr')\mathsf{T}}\right)^\mathsf{T}$. Each $\beta_k^{(rr')}$ has an associated quadratic penalty $\lambda_k^{(rr')}\beta_k^{(rr')\mathsf{T}}\mathbf{D}_k^{(rr')}\beta_k^{(rr')}$, used in fitting, whose role is to enforce specific properties on the k^{th} function, such as smoothness. Matrix $\mathbf{D}_k^{(rr')}$ only depends on the choice of the basis functions. Smoothing parameter $\lambda_k^{(rr')} \in [0,\infty)$ has the crucial role of controlling the trade-off between fit and smoothness and hence it determines the shape of the corresponding estimated smooth function. The overall penalty can be defined as $\beta^{(rr')\mathsf{T}}\mathbf{S}_{\lambda^{(rr')}}^{(rr')}\beta^{(rr')}$, where $\mathbf{S}_{\lambda^{(rr')}}^{(rr')} = \mathrm{diag}(0,\lambda_1^{(rr')}\mathbf{D}_1^{(rr')},\dots,\lambda_{K^{(rr')}}^{(rr')}\mathbf{D}_{K^{(rr')}}^{(rr')})$ and $\lambda^{(rr')} = (\lambda_1^{(rr')},\dots,\lambda_{K^{(rr')}}^{(rr')})^\mathsf{T}$ is the transition-specific overall smoothing parameter vector. Note that smooth functions are subject to centering (identifiability) constraints which are imposed as described in Wood (2017). Several definitions of basis functions and penalty terms are supported by `flexmsm`; these include thin plate, cubic and P-regression splines, tensor products, Markov random fields, random effects, and Gaussian process smooths (see Wood (2017) for details).

An example of predictor specification is $\eta_\iota^{(rr')} = \beta_0^{(rr')} + s_1^{(rr')}(t_\iota) + \beta_2^{(rr')}sex_\iota$. Parametric effects usually, but not exclusively, relate to binary and categorical variables such as sex_ι. The spline representation introduced above thus simplifies to $s_2^{(rr')}(sex_\iota) = \beta_2^{(rr')}sex_\iota$. No penalty is typically assigned to parametric effects, hence the associated penalty is 0. However, there might be instances where some form of regularisation is required in which case a suitable penalisation scheme can be employed (e.g., Wood, 2017, Section 5.8). To explore a potentially nonlinear effect of t_ι, $s_1^{(rr')}(t_\iota)$ is specified as $\mathbf{b}_1^{(rr')}(t_\iota)^\mathsf{T}\beta_1^{(rr')}$, where

$\mathbf{b}_1^{(rr')}(t_\iota)$ are cubic regression spline bases, for example. In this case, the penalty is defined as

$$\beta_1^{(rr')\mathsf{T}}\mathbf{D}_1^{(rr')}\beta_1^{(rr')} = \int_{u_1}^{u_{J_1^{(rr')}}} \left(\frac{\partial^2}{\partial t^2}s_1^{(rr')}(u)\right)^2 du,$$

where u_1 and $u_{J_1^{(rr')}}$ are the locations of the first and last knots. For a smooth term in one dimension, such as $s_1^{(rr')}(t_\iota)$, the specific choice of spline definition (e.g., thin plate, cubic) will not have an impact on the estimated curve. As for $J_1^{(rr')}$, or more generally $J_k^{(rr')}$, this is typically set to 10 since such value offers enough flexibility in most applications. However, analyses using larger values can be attempted to assess the sensitivity of the results to $J_k^{(rr')}$. Regarding the selection of knots, these can be placed evenly throughout (or using the percentiles of) the values of the variable the smooth term refers to. For a thin-plate regression spline only $J_k^{(rr')}$ has to be chosen. See Wood (2017) for a thorough discussion.

As mentioned previously, our framework poses no limits on the types of splines that can be employed for specifying the transition intensities. For instance, as illustrated in Section 4.6.1, two-dimensional splines can be used to incorporate time-dependent effects. This would take the form of an interaction term involving, e.g., age_ι and the time variable through the smooth term $s_k^{(rr')}(age_\iota, t_\iota)$. Here we have two penalties, one for each of the arguments of the smooth function. These are summed after being weighted by smoothing parameters, which serve the purpose of controlling the trade-off between fit and smoothness in each of the two directions, thus allowing for a great degree of flexibility (Wood, 2017, Section 5.6).

4.3 Penalised log-likelihood

Let N be the number of statistical units, n_i the number of times the i^{th} unit is observed, $0 = t_{i0} < t_{i1} < \cdots < t_{in_i}$ the follow-up times, $z_{i0}, z_{i1}, \ldots, z_{in_i}$ the (possibly unobserved, i.e. censored) states occupied, and $\check{n} = \sum_{i=1}^{N}(n_i - 1)$ the sample size. If $L_{ij}(\theta)$ is the likelihood contribution for the j^{th} observation of the i^{th} unit and $\theta = \{\beta^{(rr')} \mid (r,r') \in \mathcal{A}\}$ the model parameter vector, then the log-likelihood is

$$\ell(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} \log\left(L_{ij}(\theta)\right), \tag{4.3}$$

where we have

$$L_{ij}(\theta) = \begin{cases} p^{(z_{ij-1}z_{ij})}(t_{ij-1}, t_{ij}), & \text{if } z_{ij} \text{ is an interval censored state} \\ \exp\left[\int\limits_{t_{ij-1}}^{t_{ij}} q^{(z_{ij-1}z_{ij-1})}(u)du\right] q^{(z_{ij-1}z_{ij})}(t_{ij}), & \text{if } z_{ij} \text{ is an exactly observed state} \\ \sum\limits_{c \in \tilde{\mathcal{S}} \subset \mathcal{S}} p^{(z_{ij-1}c)}(t_{ij-1}, t_{ij}), & \text{if } z_{ij} \text{ is a censored state} \\ \sum\limits_{\substack{c=1 \\ c \neq z_{ij}}}^{C} p^{(z_{ij-1}c)}(t_{ij-1}, t_{ij}) q^{(cz_{ij})}(t_{ij}), & \text{if } z_{ij} \text{ is an exactly observed death state} \end{cases} .$$

That is, the likelihood contribution for a given observation will depend on the nature of the states between which the transition occurred and the way in which it was observed. Note that, in the last contribution type, $q^{(cz_{ij})}(t_{ij})$ may depend on a time-dependent covariate whose value at the time of death t_{ij} may be unknown. One way to address this is to assume that the value of this covariate in t_{ij} is the same as the the one observed in the previous follow-up time t_{ij-1}. Supplementary Material C.1 provides details on each contribution type, whereas Supplementary Material C.2 describes the use of the `R` package `flexmsm` in such a general context.

To calibrate the trade-off between parsimony and complexity, we augment the objective function (4.3) with a quadratic penalty term. This results in the penalised log-likelihood

$$\ell_p(\theta) = \ell(\theta) - \frac{1}{2}\theta^\mathsf{T}\mathbf{S}_\lambda\theta, \tag{4.4}$$

where $\mathbf{S}_\lambda = \text{diag}\left(\{\mathbf{S}^{(rr')}_{\lambda^{(rr')}} \mid (r,r') \in \mathcal{A}\}\right)$ which is a block diagonal matrix where each block is given by the transition-specific penalty matrix $\mathbf{S}^{(rr')}_{\lambda^{(rr')}}$, and $\lambda = \{\lambda^{(rr')} \mid (r,r') \in \mathcal{A}\}$ is the overall multiple smoothing parameter vector. Both $\mathbf{S}^{(rr')}_{\lambda^{(rr')}}$ and $\lambda^{(rr')}$ are defined for a generic transition (r,r') in Section 4.2.1.

4.4 Stable estimation through exact local curvature information

Building a general and flexible multi-state Markov modelling framework hinges on the availability of the analytical information matrix of the transition probability matrix, for which we propose a version here. Parameter estimation is achieved by adapting to our setting the stable and efficient approach proposed in Marra & Radice (2020), which combines a trust region algorithm with automatic multiple smoothing parameter selection. The trust region method is known to perform better than its line search counterparts and has certain optimal convergence properties as long as the analytical observed information matrix is provided (Chapter 4, Nocedal & Wright, 2006). As for the smoothing parameters, we employ a general and fast estimation framework which removes the need for computationally expensive grid search-based methods and ad-hoc optimisers (see Supplementary Material C.3 for details). From (4.3), the w^{th} element of the gradient vector $\mathbf{g}(\theta)$ and the (w,w') element of the Hessian matrix $\mathbf{H}(\theta)$, for $w,w' = 1,\dots,W$ with $W = \sum_{(r,r')\in\mathcal{A}}\left(1+\sum_{k=1}^{K^{(rr')}} J_k^{(rr')}\right)$, are defined as

$$\frac{\partial}{\partial\theta_w}\ell(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i} L_{ij}(\theta)^{-1}\frac{\partial}{\partial\theta_w}L_{ij}(\theta),$$

$$\frac{\partial^2}{\partial\theta_w\partial\theta_{w'}}\ell(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\left(L_{ij}(\theta)^{-1}\frac{\partial^2}{\partial\theta_w\partial\theta_{w'}}L_{ij}(\theta) - L_{ij}(\theta)^{-2}\frac{\partial}{\partial\theta_w}L_{ij}(\theta)\frac{\partial}{\partial\theta_{w'}}L_{ij}(\theta)\right),$$

where $\dfrac{\partial L_{ij}(\theta)}{\partial\theta_w}$ is given by

$$\begin{cases}
\dfrac{\partial}{\partial\theta_w}p^{(z_{ij-1}z_{ij})}(t_{ij-1},t_{ij}), & \text{if } z_{ij} \text{ is an interval censored state}\\[2ex]
\exp\left(\int\limits_{t_{ij-1}}^{t_{ij}} q^{(z_{ij-1}z_{ij-1})}(u)du\right)\left[\dfrac{\partial}{\partial\theta_w}q^{(z_{ij-1}z_{ij})}(t_{ij})\right. & \text{if } z_{ij} \text{ is an exactly observed state}\\
\qquad\left. +q^{(z_{ij-1}z_{ij})}(t_{ij})\int\limits_{t_{ij-1}}^{t_{ij}}\dfrac{\partial}{\partial\theta_w}q^{(z_{ij-1}z_{ij-1})}(u)du\right], & \\[2ex]
\sum\limits_{c\in\tilde{\mathcal{S}}\subset\mathcal{S}}\dfrac{\partial}{\partial\theta_w}p^{(z_{ij-1}c)}(t_{ij-1},t_{ij}), & \text{if } z_{ij} \text{ is a censored state}\\[2ex]
\sum\limits_{\substack{c=1\\ c\neq z_{ij}}}^{C}\dfrac{\partial}{\partial\theta_w}p^{(z_{ij-1}c)}(t_{ij-1},t_{ij})q^{(cz_{ij})}(t_{ij}) & \text{if } z_{ij} \text{ is an exactly observed death state}\\
\qquad +p^{(z_{ij-1}c)}(t_{ij-1},t_{ij})\dfrac{\partial}{\partial\theta_w}q^{(cz_{ij})}(t_{ij}), &
\end{cases},$$

and $\dfrac{\partial^2 L_{ij}(\theta)}{\partial\theta_w\partial\theta_{w'}}$ is given by

$$\begin{cases} \dfrac{\partial^2}{\partial\theta_w\partial\theta_{w'}}p^{(z_{ij-1}z_{ij})}(t_{ij-1},t_{ij}), & \text{if } z_{ij} \text{ is an interval censored state} \\[2ex] \begin{aligned} &\frac{\partial}{\partial\theta_w}L_{ij}(\theta)\int\limits_{t_{ij-1}}^{t_{ij}}\frac{\partial}{\partial\theta_{w'}}q^{(z_{ij-1}z_{ij-1})}(u)du \\ &\quad+\exp\left(\int\limits_{t_{ij-1}}^{t_{ij}}q^{(z_{ij-1}z_{ij-1})}(u)du\right)\left[\frac{\partial^2 q^{(z_{ij-1}z_{ij})}(t_{ij})}{\partial\theta_w\partial\theta_{w'}}\right. \\ &\quad+\frac{\partial}{\partial\theta_{w'}}q^{(z_{ij-1}z_{ij})}(t_{ij})\int\limits_{t_{ij-1}}^{t_{ij}}\frac{\partial}{\partial\theta_w}q^{(z_{ij-1}z_{ij-1})}(u)du \\ &\quad\left.+q^{(z_{ij-1}z_{ij})}(t_{ij})\int\limits_{t_{ij-1}}^{t_{ij}}\frac{\partial^2 q^{(z_{ij-1}z_{ij-1})}(u)}{\partial\theta_w\partial\theta_{w'}}du\right], \end{aligned} & \text{if } z_{ij} \text{ is an exactly observed state} \\[2ex] \sum\limits_{c\in\tilde{\mathcal{S}}\subset\mathcal{S}}\dfrac{\partial^2}{\partial\theta_w\partial\theta_{w'}}p^{(z_{ij-1}c)}(t_{ij-1},t_{ij}), & \text{if } z_{ij} \text{ is a censored state} \\[2ex] \begin{aligned} &\sum_{\substack{c=1\\c\neq z_{ij}}}^{C}\frac{\partial^2}{\partial\theta_w\partial\theta_{w'}}p^{(z_{ij-1}c)}(t_{ij-1},t_{ij})q^{(cz_{ij})}(t_{ij}) \\ &\quad+\frac{\partial}{\partial\theta_w}p^{(z_{ij-1}c)}(t_{ij-1},t_{ij})\frac{\partial}{\partial\theta_{w'}}q^{(cz_{ij})}(t_{ij}) \\ &\quad+\frac{\partial}{\partial\theta_{w'}}p^{(z_{ij-1}c)}(t_{ij-1},t_{ij})\frac{\partial}{\partial\theta_w}q^{(cz_{ij})}(t_{ij}) \\ &\quad+p^{(z_{ij-1}c)}(t_{ij-1},t_{ij})\frac{\partial^2}{\partial\theta_w\partial\theta_{w'}}q^{(cz_{ij})}(t_{ij}), \end{aligned} & \text{if } z_{ij} \text{ is an exactly observed death state} \end{cases} .$$

The quantities needed for parameter estimation are the $C\times C$ dimensional matrices $\mathbf{P}(t_{ij-1},t_{ij})$, $\partial\mathbf{P}(t_{ij-1},t_{ij})/\partial\theta_w$ and $\partial^2\mathbf{P}(t_{ij-1},t_{ij})/\partial\theta_w\partial\theta_{w'}$ for $w,w'=1,\dots,W$.

4.4.1 Computation of the transition probability matrix and its first and second derivatives

Given the transition intensity matrix $\mathbf{Q}(t)$, the transition probability matrix is the solution of the Kolmogorov forward differential equations $\partial\mathbf{P}(t,t')/\partial t'=\mathbf{P}(t,t')\mathbf{Q}(t')$, which are not in general tractable. To tackle this, we employ the commonly adopted piecewise-constant approximation approach. As for the time grid over which such approximation is defined, we let it coincide with the observation times of the dataset at hand; this allows for satisfactory estimation of the model parameters at a contained computational cost. To investigate the performance of the piecewise-constant approximation, for example, Van Den Hout (2016) designed a simulation study where this approach is compared against a continuous-time-based estimation approach. The study showed that the former can lead to good results as long as the time between observations is not too long relative to the volatility of the multi-state process. Further, in the simulation study described in Supplementary Material C.4.1, the effect of the length of the gap occurring between two successive observations is explored. We found that the performance of our method decreased as the time gap increased, but the negative effect was only sensible for time grids with a four year or larger gap. The follow-up times found in applications tend to be denser than this, thus the piecewise-constant approximation can be safely assumed. In the case of particularly sparse follow-up times, it is possible to improve the piecewise-constant approximation by embedding the grid defined by the observation times in a finer grid (see e.g., Van den Hout & Matthews, 2008).

For each individual $i = 1,\ldots,N$, let the observed follow-up times $t_{i0} < t_{i1} < \cdots < t_{in_i}$ define the extremities of the intervals over which the transition intensities are assumed to be constant. The convention is to assume that the transition intensities remain constant on the value taken in the left extremity of each time interval. Then, for $t \in [t_{ij}, t_{ij+1})$, with $j = 0,1,\ldots,n_i - 1$, making explicit the dependence on the model parameters, we have $\mathbf{Q}(t;\theta) = \mathbf{Q}_j(\theta)$ and

$$\mathbf{P}(t_{ij},t_{ij+1}) = \mathbf{P}(t_{ij+1}-t_{ij}) = \exp[(t_{ij+1}-t_{ij})\mathbf{Q}_j(\theta)] = \sum_{\zeta=0}^{\infty} \frac{[(t_{ij+1}-t_{ij})\mathbf{Q}_j(\theta)]^{\zeta}}{\zeta!}. \quad (4.5)$$

It follows that computing the transition probability matrix and its derivatives entails calculating a number of matrix exponentials and their derivatives. The eigendecomposition approach popularised by Kalbfleisch & Lawless (1985) here is appealing because it provides a closed-form solution for these power series. The availability of a closed-form expression is crucial since solving the power series for the transition probability matrix and its derivatives is a rather involved process, due to the matrix-multiplication being non-commutative. In particular, the authors provide analytical expressions for $\mathbf{P}(t_{ij-1},t_{ij})$ and $\partial\mathbf{P}(t_{ij-1},t_{ij})/\partial\theta_w$, but not for $\partial^2\mathbf{P}(t_{ij-1},t_{ij})/\partial\theta_w\partial\theta_{w'}$, which is needed to derive the observed information matrix, and only when the eigenvalues of the transition intensity matrix are distinct. Much of the literature on interval-censored multi-state process followed this work and thus relies on up to first order information only.

A lesser known and so far unexploited result is that by Kosorok & Chao (1996), who provide a closed-form solution for $\partial^2\mathbf{P}(t_{ij},t_{ij+1})/\partial\theta_w\partial\theta_{w'}$. From this work it also emerged that the derived expressions do not require the eigenvalues of the transition intensity matrix to be distinct, which was not noted in Kalbfleisch & Lawless (1985) and the subsequent literature relying on this seminal paper.

In the following we report the full compact expressions of $\mathbf{P}(t_{ij}, t_{ij+1})$, $\partial \mathbf{P}(t_{ij}, t_{ij+1})/\partial \theta_w$ and $\partial^2 \mathbf{P}(t_{ij}, t_{ij+1})/\partial \theta_w \partial \theta_{w'}$. For simplicity, we will drop the dependence on i, j and θ and define $\delta t = t_{ij+1} - t_{ij}$.

Let $\mathbf{Q} = \mathbf{A}\Gamma\mathbf{A}^{-1}$ be the eigendecomposition of the transition intensity matrix, which is constant over the generic time interval of length δt, with $\mathbf{A}$ the matrix of eigenvectors and $\Gamma = \text{diag}[\gamma_1, \ldots, \gamma_C]$ the diagonal matrix containing the eigenvalues. Then

$$\mathbf{P}(\delta t) = \mathbf{A}\,\text{diag}(\exp\left[\gamma_1 \delta t\right], \ldots, \exp\left[\gamma_Y \delta t\right])\mathbf{A}^{-1}, \tag{4.6}$$

$$\frac{\partial}{\partial \theta_w}\mathbf{P}(\delta t) = \mathbf{A}\mathbf{U}_w\mathbf{A}^{-1}, \tag{4.7}$$

$$\frac{\partial^2}{\partial \theta_w \partial \theta_{w'}}\mathbf{P}(\delta t) = \mathbf{A}(\check{\mathbf{U}}_{ww'} + \dot{\mathbf{U}}_{ww'} + \dot{\mathbf{U}}_{w'w})\mathbf{A}^{-1}, \tag{4.8}$$

where $\mathbf{U}_w = \mathbf{G}^{(w)} \circ \mathbf{E}$ and $\check{\mathbf{U}}_{ww'} = \mathbf{G}^{(ww')} \circ \mathbf{E}$, with $\mathbf{E}[l,m] = \frac{\exp[\gamma_l \delta t] - \exp[\gamma_m \delta t]}{\gamma_l - \gamma_m}$ when $\gamma_l \neq \gamma_m$ and $\mathbf{E}[l,m] = \delta t e^{\gamma_l \delta t}$ when $\gamma_l = \gamma_m$, $\mathbf{G}^{(w)} = \mathbf{A}^{-1}\dfrac{\partial \mathbf{Q}}{\partial \theta_w}\mathbf{A}$, $\mathbf{G}^{(ww')} = \mathbf{A}^{-1}\dfrac{\partial^2 \mathbf{Q}}{\partial \theta_w \partial \theta_{w'}}\mathbf{A}$ and

$$\dot{\mathbf{U}}_{ww'}[l,m] = \sum_{y=1}^{Y} G_{ly}^{(w)} G_{ym}^{(w')} \begin{cases} \dfrac{e^{\gamma_l \delta t} - e^{\gamma_y \delta t}}{(\gamma_l - \gamma_y)(\gamma_y - \gamma_m)} - \dfrac{e^{\gamma_l \delta t} - e^{\gamma_m \delta t}}{(\gamma_l - \gamma_m)(\gamma_y - \gamma_m)}, & \gamma_l \neq \gamma_y \neq \gamma_m \\ \dfrac{t e^{\gamma_l \delta t}}{\gamma_l - \gamma_m} - \dfrac{e^{\gamma_l \delta t} - e^{\gamma_m \delta t}}{(\gamma_l - \gamma_m)^2}, & \gamma_l = \gamma_y \neq \gamma_m \\ \dfrac{e^{\gamma_l \delta t} - e^{\gamma_m \delta t}}{(\gamma_l - \gamma_m)^2} - \dfrac{t e^{\gamma_m \delta t}}{\gamma_l - \gamma_m}, & \gamma_m = \gamma_y \neq \gamma_l \\ \dfrac{t e^{\gamma_l \delta t}}{\gamma_l - \gamma_y} - \dfrac{e^{\gamma_l \delta t} - e^{\gamma_y \delta t}}{(\gamma_l - \gamma_y)^2}, & \gamma_l = \gamma_m \neq \gamma_y \\ \dfrac{1}{2}\delta t^2 e^{\gamma_l \delta t}, & \gamma_l = \gamma_m = \gamma_y \end{cases},$$

where $G_{lm}^{(w)}$ is the (l,m) element of matrix $\mathbf{G}^{(w)}$. $\dot{\mathbf{U}}_{w'w}$ is obtained in the same way as $\dot{\mathbf{U}}_{ww'}$ but with w and w' swapped wherever they appear. We refer the reader to Kalbfleisch & Lawless (1985) for the proofs of (4.6) and (4.7) and to Kosorok & Chao (1995) for the proof of (4.8).

Note that $\partial \mathbf{Q}/\partial \theta_w$ and $\partial^2 \mathbf{Q}/\partial \theta_w \partial \theta_{w'}$ are matrices whose (r, r') elements are given, respectively, by $\partial q^{(rr')}(t_{ij})/\partial \theta_w$ and $\partial^2 q^{(rr')}(t_{ij})/\partial \theta_w \partial \theta_{w'}$ for $w, w' = 1, \ldots, W$. Further, the first derivatives of the transition intensity matrix are already available from the computation of the first derivatives of the transition probabilities, hence only second derivatives have to be computed anew. Matrices $\mathbf{A}$, $\mathbf{A}^{-1}$ and Γ also need to be computed only once, when obtaining matrix $\mathbf{P}$.

Regarding the implementation of the quantities of interest, the number of operations grows quickly as n_i, N, C and W increase. Specifically, $\mathbf{Q}$ (and its eigendecomposition), $\partial\mathbf{Q}/\partial\theta_w$, $\partial^2\mathbf{Q}/\partial\theta_w\partial\theta_{w'}$, $\mathbf{P}$, $\partial\mathbf{P}/\partial\theta_w$ and $\partial^2\mathbf{P}/\partial\theta_w\partial\theta_{w'}$, for $w, w' = 1, \ldots, W$, have to be computed $\sum_{i=1}^{N} n_i - N$ times and then combined. To reduce computational cost, the proposed implementation exploited the upper-triangle form of the above mentioned matrices and the presence of structural zero-values in them. We also exploited parallel computing to obtain the log-likelihood, analytical score and information matrix more quickly; the overall run-time of the algorithm can be cut by a factor proportional to the number of cores in the user's computer.

4.5 Inference

To obtain confidence intervals, instead of using the classically derived frequentist covariance matrix $-\mathbf{H}_p^{-1}(\theta)\mathbf{H}(\theta)\mathbf{H}_p^{-1}(\theta)$, we follow Wood et al. (2016) and employ the Bayesian large sample approximation $\theta \dot{\sim} \mathcal{N}(\widehat{\theta}, \mathbf{V}_\theta)$, where $\mathbf{V}_\theta = -\mathbf{H}_p(\widehat{\theta})^{-1}$ with $\hat{\theta}$ the estimated model parameter and $\mathbf{H}_p(\theta) = \mathbf{H}(\theta) - \mathbf{S}_\lambda$ the penalised Hessian. Using $\mathbf{V}_\theta$ gives close to across-the-function frequentist coverage probabilities because it accounts for both sampling variability and smoothing bias, a feature that is particularly relevant at finite sample sizes and that is not shared by the frequentist covariance matrix. Note that applying the Bayesian approach to the modelling framework discussed in this paper follows the notion that penalisation in estimation implicitly assumes that wiggly models are less likely than smoother ones, which translates into the following prior specification for θ, $f_\theta \propto \exp\left\{-\theta^{\mathsf{T}}\mathbf{S}_\lambda\theta/2\right\}$.

Intervals for linear functions of the model coefficients, e.g. smooth components, can be obtained using the result just shown for θ. For nonlinear functions of the model coefficients, intervals can be conveniently obtained by posterior simulation. For example, to derive the $(1-\alpha)100\%$ intervals for the (r, r') transition intensity, the following procedure can be employed:

1. Draw n_{sim} random vectors $\beta^{(1,rr')}, \beta^{(2,rr')}, \ldots, \beta^{(n_{sim},rr')}$ from $\mathcal{N}(\widehat{\beta^{(rr')}}, \mathbf{V}_{\beta^{(rr')}})$, where $\widehat{\beta^{(rr')}}$ is the estimated model parameter.

2. Calculate n_{sim} simulated realisations of the quantity of interest, such as $q^{(rr')}(t)$. For fixed $\mathbf{x}$ and t, one would obtain $\mathbf{q}_{sim}^{(rr')} = (q^{(1,rr')}, q^{(2,rr')}, \ldots, q^{(n_{sim},rr')})^{\mathsf{T}}$ using $\beta^{(1,rr')}, \beta^{(2,rr')}, \ldots, \beta^{(n_{sim},rr')}$ respectively.

3. Using $\mathbf{q}_{sim}^{(rr')}$, calculate the lower, $\alpha/2$, and upper, $1-\alpha/2$, quantiles.

A small value of $n_{sim} = 100$ typically gives accurate results, whereas α is usually set to 0.05. Note that the distribution of nonlinear functions of the model parameters need not be symmetric. Intervals for the transition probabilities can be obtained by applying the above procedure to the **Q** matrices and then deriving the corresponding **P** matrices, as explained in Section 4.4.1.

P-values for the terms in the model can be reliably obtained by using the results summarised in Wood (2017, Section 6.12), which are based on $-\mathbf{H}_p(\theta)^{-1}$. Model building can be aided using tools such as the Akaike information criterion (AIC, Akaike, 1998), defined as $-2\ell(\theta) + 2edf$, where the log-likelihood is evaluated at the penalised parameter estimates and the effective degrees of freedom are given by $edf = \text{tr}(\mathbf{O})$, with $\text{tr}(\cdot)$ the trace function and $\mathbf{O} = \sqrt{-\mathbf{H}(\theta)}\,(-\mathbf{H}_p(\theta))^{-1}\sqrt{-\mathbf{H}(\theta)}$ (Marra & Radice, 2020). An alternative measure that can be used for model selection is the Bayesian information criterion (BIC, Schwarz, 1978), which penalises the log-likelihood by adding the number of effective degrees of freedom multiplied by the logarithm of the sample size. The BIC, however, requires the observations in the sample to be independent, an assumption that does not hold for longitudinal data. It follows that computing the BIC in this setting is not as straightforward as computing the AIC. The literature generally proposes to use the number of individuals, while recognising that this leads to a very conservative measure. Jones (2011) developed a method to compute an ”effective sample size”, whose minimum and maximum values are respectively the number of individuals and the number of observations, which can be used in the computation of the BIC. Adapting this approach to our setting is out of the scope of the work so we rely only on the AIC in the following, but it may represent an avenue of future work.

4.6 Case studies

The proposal is illustrated through two case studies. The first one uses flexible IDMs to model the onset of cardiac allograft vasculopathy (CAV), a deterioration of the arterial walls in heart transplant patients. The second one aims at modelling cognitive decline in the

English Longitudinal Study of Ageing (ELSA) population through a flexible five-state model with both forward and backward transitions as well as an absorbing death state. Note that more parsimonious models than those described below can be fitted as well. In the following analyses, however, we focus on transition intensity specifications with a high degree of flexibility since, thus far, these could not be explored.

4.6.1 CAV case study

The heart transplant monitoring data used here are openly accessible from the `R` package `msm`. The dataset contains 2846 observations, relating to 622 patients, and is about angiographic (approximately yearly) examinations of heart transplant recipients where the grade of CAV (not present, mild/moderate or severe) is recorded. The additional time event of death is also registered and known exactly (within one day). It follows that the likelihood contributions involved here are those relating to interval censored living states and to exactly observed absorbing states. Available baseline covariates include age of the donor (`dage`) and primary diagnosis of ischaemic heart disease (IHD, `pdiag`) which are known to be major risk factors for CAV onset. In line with Mariano Machado et al. (2021), we remove eight individuals for which the principal diagnosis is not known and exclude observations which occurred beyond 15 years from the transplant. The resulting dataset contains $\sum_{i=1}^{N} n_i = 2803$ observations of $N = 614$ patients. We consider flexible IDMs where the states are (1) health (2) CAV onset (mild/moderate or severe) and (3) death. A diagram representing the process is displayed in Figure 4.1 while Table 4.1 reports the number of observations available for each pair of states in the dataset. Note that the sum of these counts provides the sample size, $\breve{n} = 2189$.

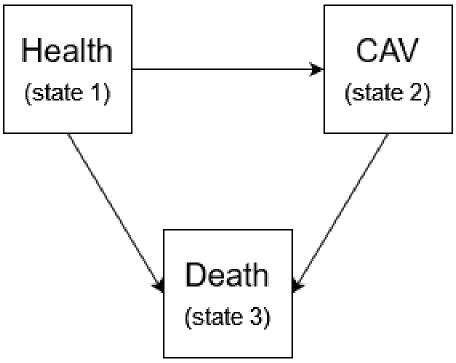

Figure 4.1: Diagram of the possible IDM disease trajectories.

	state 1	state 2	state 3
state 1	1314	223	136
state 2	0	411	105
state 3	0	0	0

Table 4.1: Number of observations for each pair of states in the CAV dataset.

The most flexible IDM considered in the literature for the CAV case study is based on

Mariano Machado et al. (2021)

$$q^{(rr')}(t_{ij}) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t_{ij}) + \beta_2 \texttt{dage}_i + \beta_3 \texttt{pdiag}_i\right], \tag{4.9}$$

for $(r,r') \in \{(1,2),(1,3),(2,3)\}$, where t is the time since transplant, the smooth term is represented by a cubic regression spline with 10 basis functions and second order penalty, and β_2 and β_3 are covariate effects which are constrained to be equal across the three transitions, hence the lack of superscript. Model fitting was conducted using the bespoke R code provided by Mariano Machado et al. (2021) which took 3.5 days to reach convergence, on a laptop with Windows 10, Intel 2.20 GHz core, 16 GB of RAM and eight cores. The resulting AIC was 2931.7. No justification was provided for setting $\beta_2^{(rr')} = \beta_2$ and $\beta_3^{(rr')} = \beta_3$ which may be too restrictive to estimate adequately the effects of dage and pdiag.

Using the proposed methodology, we considered the more general specification

$$q^{(rr')}(t_{ij}) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t_{ij}) + \beta_2^{(rr')} \texttt{dage}_i + \beta_3^{(rr')} \texttt{pdiag}_i\right], \tag{4.10}$$

which produced an AIC of 2915.2. The run-time of flexmsm was 59 minutes. Using different spline definitions and increasing $J_1^{(rr')}$ did not lead to tangible empirical differences. Note that employing an approximation of the Hessian, based on first order derivatives, led to convergence failures when fitting the above model as well as those considered at the end of this section. This finding is supported by the simulation study in Supplementary Material C.4.1.1 which shows that basing parameter estimation on an approximate information matrix leads to convergence issues. The study also demonstrates the empirical effectiveness of the proposed approach.

Table 4.2 reports the effects for dage and pdiag, and their standard errors, resulting from models (4.9) and (4.10). As the table shows, the constrained coefficients are, roughly speaking, the averages of the respective unconstrained ones. In this case, setting restrictions does not allow one to uncover the differing effects of the risk factors in the different trajectories. Specifically, the model (4.10) results indicate that dage and pdiag increase the risks of moving from state 1 to state 2 and from state 1 to state 3, and that these variables do not play a role in the transition 2 → 3. The curve estimates for the $s_1^{(rr')}(t_{ij})$ (not reported here) were similar across the two models.

Figure 4.2 shows the estimated transition intensities, and 95% intervals, when dage is

	`dage`	`pdiag`
$1 \to 2$	0.023 (0.006)	0.414 (0.132)
$1 \to 3$	0.040 (0.011)	0.341 (0.255)
$2 \to 3$	-0.016 (0.009)	0.002 (0.178)
$1 \to 2, 1 \to 3, 2 \to 3$	0.018 (0.004)	0.274 (0.096)

Table 4.2: Estimated covariate effects and related standard errors (between brackets) for donor age (`dage`) and principal diagnosis of IHD (`pdiag`) obtained using the proposed model fitted by `flexmsm` (first three lines) and the constrained model of Mariano Machado et al. (2021) fitted using the related bespoke `R` code.

equal to 26 years and `pdiag` is equal to 1 (i.e., the principal diagnosis is IHD). The risk of moving from state 1 to state 2 increases until about 7 years since transplant; after that the situation is uncertain. The risk for the transition $1 \to 3$ is fairly low and constant until about 10 years, after which it starts increasing steeply. For transition $2 \to 3$, the risk increases overall. As expected, the intervals are wide when the data are scarce. The same exercise can be repeated for different combinations of `dage` and `pdiag`. It should be noted that the CAV dataset provided in the `R` package `msm` does not indicate the amount of follow-up that occurred after the last angiogram for patients who survived. The stark upward trends exhibited by the estimated intensity functions for the transitions into the death state can be explained as an artifact of this. Titman (2008), in fact, noted that if censored subjects are taken out at their censoring time but patients who die are left as under observation until the final follow-up time, then the observed prevalence in the death state will be systematically overestimated.

Estimated transition intensities provide valuable information about the risks of moving across states. However, interpretation is more intuitive and easier when transition probabilities are considered. Setting `dage = 26` and `pdiag = 1` and assuming yearly piecewise constant transition intensities, the estimated five-year transition probabilities can be obtained by exploiting the Chapman-Kolmogorov equations (Cox & Miller, 1977). These allow us to write $\hat{\mathbf{P}}(0,5) = \hat{\mathbf{P}}(0,1) \times \hat{\mathbf{P}}(1,2) \times \cdots \times \hat{\mathbf{P}}(4,5)$, where the probabilities over each sub-interval are obtained using the corresponding transition intensity matrix, i.e. $\hat{\mathbf{Q}}(t)$, for $t = 0, 1, \ldots, 4$ respectively. The resulting estimated transition probability matrix and 95%

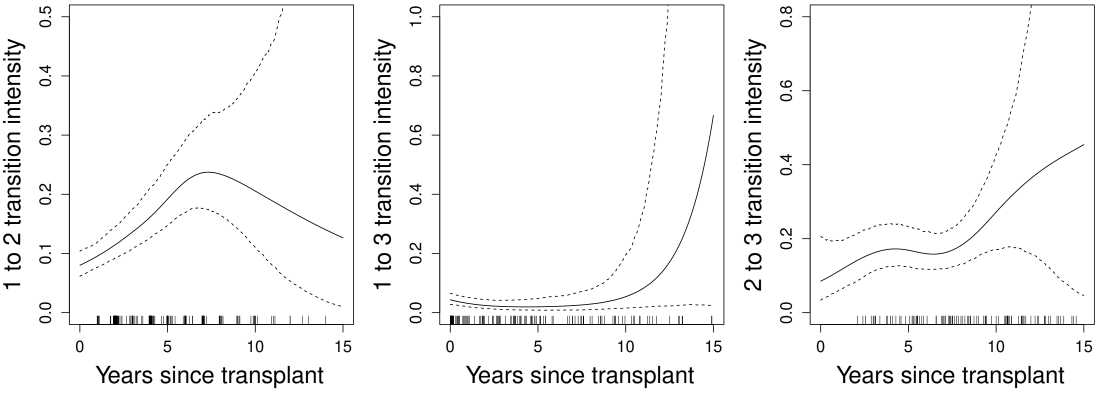

Figure 4.2: Estimated transition intensities obtained with `flexmsm` for $q^{(12)}(\cdot)$, $q^{(13)}(\cdot)$ and $q^{(23)}(\cdot)$ (from left to right) when `dage = 26` and `pdiag = 1`, with 95% intervals derived as detailed in Section 4.5. The ‘rug plot’, at the bottom of each graph, shows the empirical distribution of the transition times. Because we are dealing with an intermittent observation scheme, the time intervals have been represented by plotting the right extremity of each observed interval (the left extremity or mid-point could have been equivalently chosen). Recall that the aim of the rug plot is to highlight regions where the occurrence of a specific transition is rare, hence explaining the width of the intervals across sections.

intervals (obtained through the method detailed in Section 4.5) are

$$\hat{\mathbf{P}}(0,5) = \begin{bmatrix} 0.48 & (0.42, 0.53) & 0.29 & (0.24, 0.34) & 0.23 & (0.19, 0.29) \\ 0 & & 0.51 & (0.35, 0.63) & 0.49 & (0.37, 0.64) \\ 0 & & 0 & & 1 & \end{bmatrix}.$$

For instance, given a healthy starting point, there is a 29% chance of developing CAV five years after the transplant procedure occurred. Similarly, there is a 23% chance of dying within the same time frame, given the same starting point.

We also assessed the possible presence of nonlinear effects of `dage`. This was achieved by replacing $\beta_2^{(rr')}\texttt{dage}_i$ with $s_2^{(rr')}(\texttt{dage}_i)$ in model (4.10), where the smooth terms were represented as for $s_1^{(rr')}(t_{ij})$; the effects were found to be linear. Finally, to illustrate the generality of the proposal, we considered the specification

$$q^{(rr')}(t_{ij}) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t_{ij}) + s_2^{(rr')}(\texttt{dage}_i) + s_3^{(rr')}(t_{ij}, \texttt{dage}_i) + \beta_4^{(rr')}\texttt{pdiag}_i\right],$$

where $s_3^{(rr')}(t_{ij}, \texttt{dage}_i)$ is a tensor product interaction between `dage` and time whose marginals are cubic regression splines. Here, the main effects and their interaction are modelled separately, thus leading to more flexibility in determining the complexity of the effects (Wood, 2017, Section 5.6.3). Figure 4.3 shows the results for transition 1 → 2. In the left panel, we report the estimated transition intensity surface, which is a bivariate function of time and dage. This plot can be read by sectioning the surface, with respect to either of the two arguments, and assessing how the resulting curve varies with respect to the other covariate. In the right panel, we report two sections of the surface obtained by fixing dage at 26 and 56 years, along with their 95% confidence intervals. The scarcity of data for the two sections helps to explain the wide confidence intervals, particularly past a certain point. For this reason, we will focus the interpretation on the first few years since the transplant took place. One can see that the risk of developing CAV is almost three times higher with a 56 year old donor than it is with a 26 year old donor right after the transplant, and remains higher overall in the following few years. This is in line with expectations that older donors are associated with higher chances of disease occurrence.

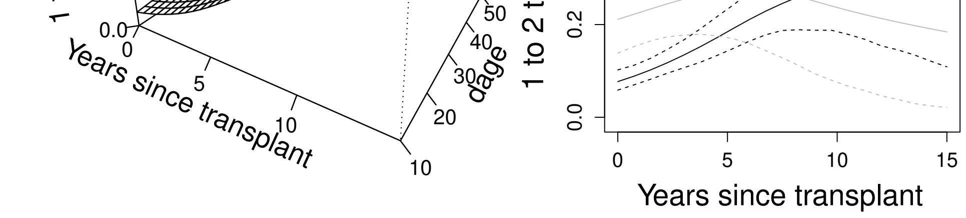

Figure 4.3: Left panel: estimated transition intensity surface, obtained with flexmsm when including a time-dependent effect of the donor age. Right panel: sections of the estimated transition intensity surface at dage = 26 (black) and dage = 56 (grey), along with their respective 95% confidence intervals (black and grey dashed lines, respectively).

Supplementary Material C.4.1 discusses a simulation study based on the IDM. The results support the empirical effectiveness of the proposed modelling framework and the related implementation in `flexmsm`.

4.6.2 ELSA case study

The ELSA collects data from people aged over 50 to understand all aspects of ageing in England. More than 18000 people have taken part in the study since it started in 2002, with the same people re-interviewed every two years, hence giving rise to an intermittently observed scheme. ELSA collects information on physical and mental health, wellbeing, finances and attitudes around ageing, and tracks how these change over time. The data can be downloaded from the UK Data Service by registering and accepting an End User Licence.

For this study, interest lies in assessing cognitive function in the older population. This is measured through the score obtained on a test in which participants are asked to remember words in a delayed recall from a list of ten, with the score given by the number of words remembered. In line with Mariano Machado et al. (2021), we use a random sample of $N = 1000$ individuals from the full population, leading to 4597 observations, and create four score groups to obtain a five-state process with the fifth state given by the occurrence of death (which is an exactly observed absorbing state). The intermediate states are given by $\{10,9,8,7\}$, $\{6,5\}$, $\{4,3,2\}$ and $\{1,0\}$ words remembered, respectively. Both forward and backward transitions are allowed between the intermediate states to account for possible improvements or fluctuations through the years in the cognitive function of the participants. In fact, although interest lies mostly in cognitive decline, the opposite trend is also observed as shown in Table 4.3. A diagram representing the assumed process is reported in Figure 4.4. Further, 221 participants die during the observation period. The time scale is defined by subtracting 49 years to the age of the individuals. A potential drawback of this analysis is that the quantity defining the states, i.e. the number of words recalled, is a noisy measure, which has the potential to lead to classification error. Future work will focus on extending the current approach to support hidden Markov models (Jackson et al., 2003), which provide a way to handle misclassification such as the one which may arise in this setting.

The most flexible five-state model considered in the literature for the ELSA data is based on Mariano Machado et al. (2021)

$$q^{(rr')}(t_{ij}) = \begin{cases} \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t_{ij})\right] & \text{for } (r,r') \in \{(1,2),(2,3),(2,5),(3,4),(3,5),(4,5)\} \\ \exp\left[\beta_0^{(rr')}\right] & \text{for } (r,r') \in \{(1,5),(2,1),(3,2),(4,3)\} \end{cases},$$

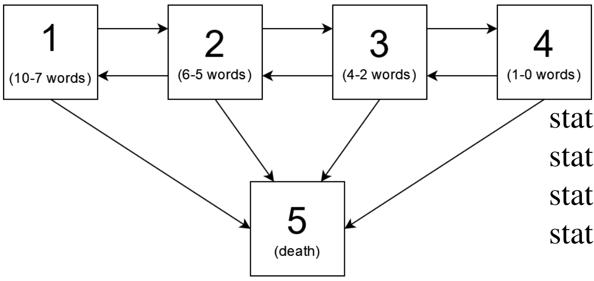

Figure 4.4: Diagram of the possible five-state process disease trajectories.

	state 1	state 2	state 3	state 4	state 5
state 1	225	194	58	5	11
state 2	209	600	384	54	46
state 3	59	383	732	152	94
state 4	8	42	117	154	70

Table 4.3: Number of observations for each pair of states in the ELSA dataset.

where each smooth term is represented by a cubic regression spline with $J_1^{(rr')} = 5$ and second order penalty, and upper bounds for the smoothing parameters were set at $\exp(20)$. The authors justify the specifications for the $q^{(rr')}(t_{ij})$ and the other settings by arguing that the limited information across the age range is probably what causes algorithmic convergence failures in more general models.

Using the proposed methodology, we considered the general specification

$$q^{(rr')}(t_{ij}) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t_{ij})\right] \text{ for } (r,r') \in \mathcal{A}, \tag{4.11}$$

with $J_1^{(rr')} = 10$ cubic regression spline bases instead. Figure 4.5 shows the estimated transition intensities, and related 95% intervals, obtained with `flexmsm`. As expected, the instantaneous risks of dying are overall smaller than the risks of experiencing further cognitive impairment. As the starting stage reflects more advanced decline, the risk of transitioning to a worse stage becomes a progressively flatter function of time. This shows that once the individuals in the population reach a stage of cognitive impairment, they will typically stay there for the rest of the observation period. Note that there is added value from having modelled the backward transitions through smooth functions of time. For example, we find that the instantaneous chance of improving back to state 3 from a state of cognitive impairment of level 4 decreases considerably faster through time than that of returning to state 1 from state 2. This is in line with expectations as the intermediate stages of cognitive health, i.e. stages 2 and 3, are by far the most frequently observed, with 72% of the population that is still alive at the end of the observation period found in these categories. The wide 95% intervals for transitions 1 → 5 and 2 → 5 can be explained by observing, from Table 4.3, that these transitions are characterised by the lowest number of observations.

Model (4.11) is general in that no prior assumptions are made with regard to the way each transition depends on time. Instead, they are all defined through flexible functional forms by means of splines. The proposed estimation approach then suppresses any complexity not supported by the data, resulting in final estimated shapes which may be either flat, linear or non-linear. This avoids the need for setting manual constraints or enforcing ad-hoc fixes.

We also quantified the effects of two commonly investigated risk factors: `sex` (0 for male and 1 for female) and `higherEdu` (0 if the individual has had less than 10 years of education and 1 otherwise) as extracted from the ELSA datasets. This was achieved by simply including $\beta_2^{(rr')}\texttt{sex}_{ij}$ and $\beta_3^{(rr')}\texttt{higherEdu}_{ij}$ in (4.11). We found, for example, that older people with a higher level of education have better memory function, although this does not protect them from cognitive decline as they age (e.g., Cadar et al., 2017). Overall, the effect of `sex` was found not to be significant.

In Figure 4.6, we present transition probability plots over 10 years for a 60 year old male with less than 10 years of education. We observe, e.g., that for such individual with stage 2 cognitive health and $\texttt{higherEdu} = 0$, the probability of reaching stage 3 by the age of 65 is approximately 40.3%, with 95% interval $(33.3\%, 44.7\%)$.

Finally, supplementary Material C.4.2 discusses a simulation study based on a five-state process. The results show our framework's ability to recover the true underlying transition intensities in a context which strays from the traditionally explored IDM.

4.7 Discussion

We propose a general framework for multi-state Markov modelling that allows for different types of process, with several states and various observation schemes, and that supports time-dependent flexible transition intensities with any type of covariate effects. This is motivated by the interest in modelling the evolution through time of diseases, with the aim of making statements on their course given specific scenarios or risk factors. The degree of flexibility allowed for the specification of the transition intensities determines the extent to which we can explore and describe the different factors influencing the evolution of a disease. Previous methodological developments have mainly focused on simple parametric forms and time-constant transition intensities, which can be attributed to the lack of an estimation framework capable of supporting more realistic specifications. Attempts addressing this have not been backed by adequate estimation procedures and software implementations.

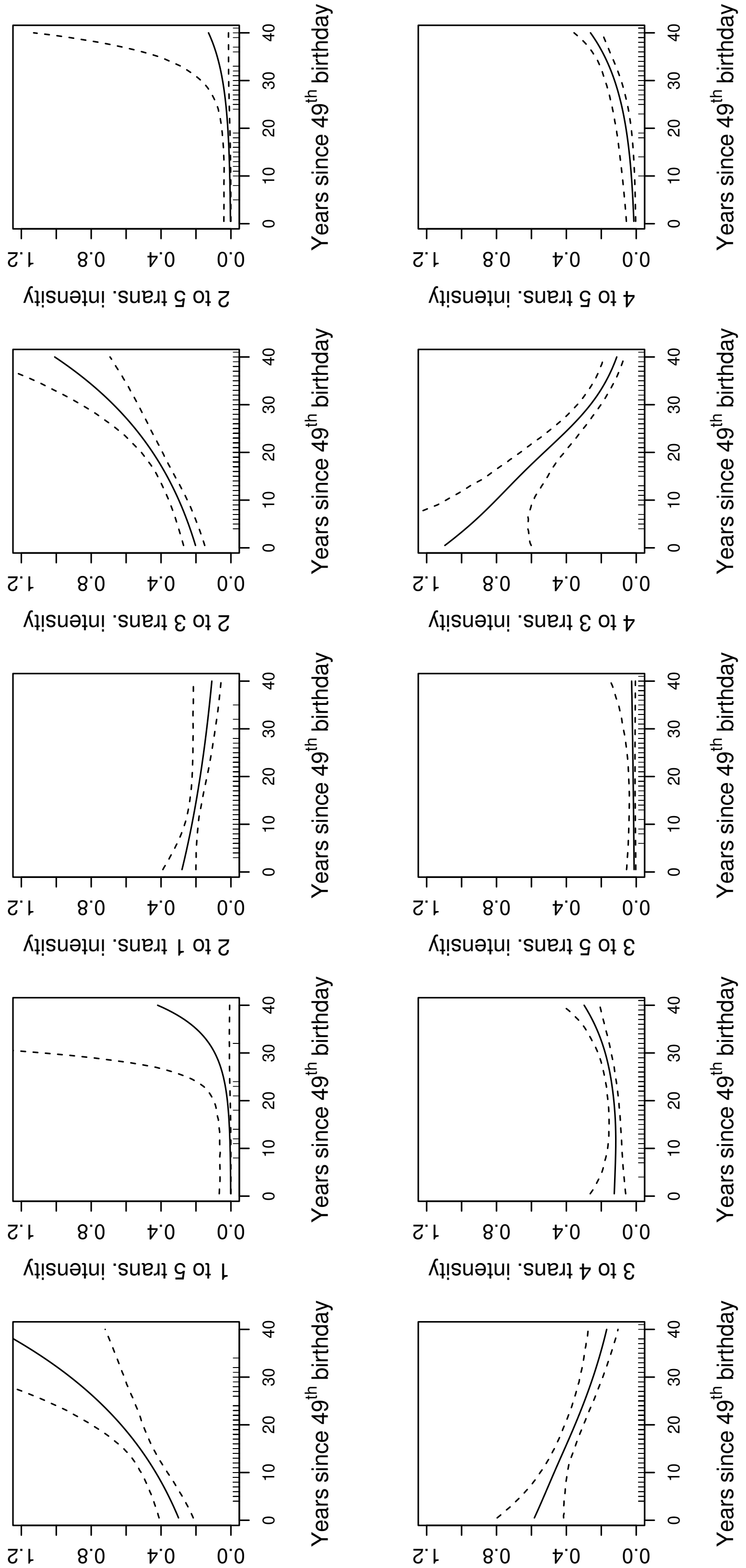

Figure 4.5: Estimated transition intensities obtained with `flexmsm` with the 95% confidence intervals derived as detailed in Section 4.5.

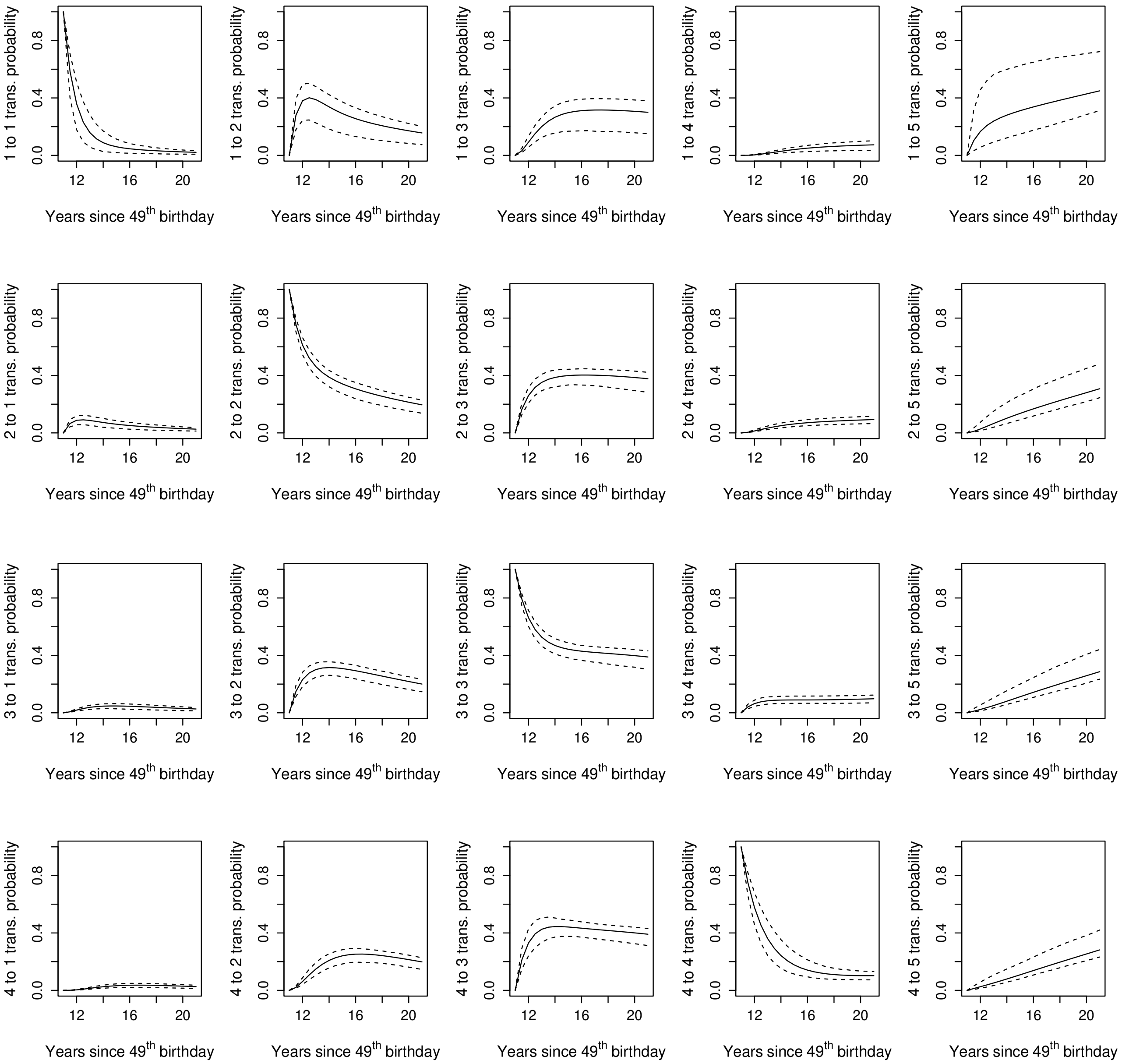

Figure 4.6: Transition probabilities for a male individual with less than 10 years of education estimated between 11 and 21 years from their 49^{th} birthday, i.e. $\hat{\mathbf{P}}(11,\ t)$ and $t \in (11,21)$. The dashed lines represent the corresponding 95% intervals.

The key contribution of the paper is the development of an approach that implements and exploits the knowledge of the exact local curvature information. Access to this source of information has allowed us to introduce a modelling framework that has unlocked a host of processes and specifications which were not previously attainable, as demonstrated via the two case studies on cardiac allograft vasculopathy and cognitive decline. To support applicability and reproducibility, we also introduced the `R` package `flexmsm`, which is easy and intuitive to use.

Future work will look into further improving the run-time required for model fitting. We are also interested in exploring transformations alternative to the exponential, to enhance the flexibility allowed by the framework. Note that we have assumed a Markov process throughout. Checking whether this property was appropriate for the data considered in this paper was outside of the scope of this work. Future efforts will look into goodness-of-fit tests (e.g., Titman, 2009) as well as the possibility of extending the current model to relax the Markov assumption. There is, however, theoretical and empirical evidence that assuming the Markov property when the true underlying process is non-Markov will still lead to a model that performs well and that has desirable properties. Using right censored data from a general multi-state model which is not Markov, for example, Datta & Satten (2001) show that the Nelson–Aalen estimator for the integrated transition hazard of a Markov process consistently estimates a population quantity even when the underlying process is not Markov. They also show that the Aalen–Johansen estimators of the stage occupation probabilities constructed from these integrated hazards via product integration are consistent for a general multi-state model that is not Markov. Using landmarking, consistency was proven for transition probabilities too (Putter & Spitoni, 2018). More recently, Nießl et al. (2023) extended these results to include multi-state data subject to left-truncation as well and provided a rigorous proof of consistency of the Aalen-Johansen estimator for state occupation probabilities, on which also correctness of the landmarking approach hinges, correcting and simplifying the earlier results.

Finally, there are circumstances which give rise to multiple dependent multi-state processes, such as the analysis of the evolution of a disease in paired organ systems. In these cases, interest lies in jointly modelling the evolution through time of these events, as the course of one is expected to affect the course of the others. Existing approaches rely on very simple specifications for the marginal processes and restrictive dependence structures among them. The framework proposed in this article will serve as the foundation for the flexible modelling of joint multi-state processes.

Chapter 5

Copula-Based Modelling of Multiple Dependent Multi-State Processes

5.1 Introduction

In the life sciences, interest lies in describing how a phenomenon will unfold over time, while accounting for factors with the potential to influence its course. Diseases, for example, exhibit multiple phases, depending on the degree of severity or the appearance or disappearance of specific symptoms. Modelling the sequence of events observed in the individuals of a population, then, allows us to predict the disease's evolution for a given set of features. This includes the worsening of their condition under different treatment strategies or, in health economics, the long-term costs connected to a specific health policy.

There are circumstances that give rise to interconnected phenomena, e.g. when a disease is expressed in multiple organ systems or locations. When this is the case, there is interest in modelling their evolution jointly, since the progression of one is expected to affect the progression of the others. Multi-state processes provide a powerful way of handling these complex settings: each phenomenon can be represented by a single process, whose states are given by the stages of the phenomenon; their joint evolution can, then, be described by linking the processes together, and modelling this system of linked processes. The aim is to make statements on the future course of a single phenomenon, while accounting for the effect that the evolution of the others has on the former.

This is the case for the progression of damage in the left and right sacroiliac joints in patients with Psoriatic Arthritis (Cook & Lawless, 2018). The extent of damage is assessed based on the analysis of radiographic images and by assigning a discrete score, from 0 to

3, to the degrees of damage found. Each sacroiliac joint can be represented by a four-state progressive process and one may be interested in answering questions such as: ”how does the fact that the left sacroiliac joint transitioned from stage 2 to stage 3 affect the probability of the right joint experiencing a similar worsening in the degree of damage?”.

In patients with lupus nephritis, the estimated glomerular filtration rate (eGFR) and the urine protein content (PU) are of interest in the treatment and management the disease (O’Keeffe et al., 2018). Movements by patients among the eGFR and PU levels can be represented by two three-state processes, where the states are based on clinically defined thresholds, with both forward and backward transitions, since improvement may occur. Evidence has been found that the two processes are associated, with a higher rate of transitioning from state 1 to state 2 for the glomerular filtration rate process when the proteinuria process is in state 2 versus when it is in state 1.

Another example stems from paired organ systems, such as retinopathy in diabetic patients or age-related macular degeneration, both of which are diseases affecting the eyes. Cook & Lawless (2014), for example, use the so-called Early Treatment Diabetic Retinopathy Study (ETDRS) scores to obtain a joint measure of retinopathy severity in the eyes. In particular, they group the scores into five levels and model disease progression using a five-state process. A more granular analysis may consider separate scores for each eye and model them as two associated processes. Nephropathy, a disease involving the kidneys, is also often found in diabetic patients, and is of vascular nature, as is retinopathy. This common origin motivates the interest in modelling the progression of the two diseases jointly. In particular, kidney disease progression can be represented through a three-state process in which each state reflects a level from the so-called KDIGO score, measuring the degree of severity of the diseases (Lintu et al., 2022). The resulting five- and three-state processes can then be modelled jointly to gain novel insight on the progression of the two related conditions.

Recent studies have also attempted to answer questions on the relationship between mental health and disease, such as disentangling the role of marital quality on average glycemic levels among adults 50 years and older in the UK (Ford & Robitaille, 2023). The quality of marital life is assessed by defining scores based on the answers given to survey questions on spouse support and strain. The average glycemic levels are summarised by testing for hemoglobin HbA1c levels in the blood and the states defining this process can be obtained by discretising these levels into three states: healthy, prediabetic, diabetic. The relation between the two can thus be modelled dynamically by means of two dependant multi-state processes, hence better characterising, e.g., the effect of spousal support on long-term prediction of type 2 diabetes, which is connected to glycemic health.

Examples that stray from the life sciences include, for example, the systems reliability literature, where jointly modelling the lifetimes of components belonging to the same system ensures better reliability design and analysis (Eryilmaz, 2014). Components that are used in the same environment and/or share the same load are, in fact, expected to affect their respective lifetimes, making an independence assumption inappropriate.

In the literature, the three main approaches used to account for dependence between multi-state processes are: (a) random effects; (b) intensity based models; (c) copulae. In (a), between-process dependence is modelled through specifying transition intensities conditionally on shared or correlated random effects. For example, Cook et al. (2004) considered the analysis of dependent progressive multi-state processes with a discrete multivariate random-effects distribution used to account for correlation. O'Keeffe et al. (2011) extend this to allow the use of gamma-distributed random effects. An important drawback of this approach is that the interpretation of time and/or covariate effects may be limited, as these are typically assumed to be conditional on random effects. In fact, although the conditional processes are assumed to be Markov, the marginal processes obtained by integrating out the random effects lose the Markov property. In approach (b), the state occupied at a given time by one process is included as a (time-dependent) covariate in the transition intensity equations of the other process, and vice-versa (Cook & Lawless, 2014), offering more freedom when specifying dependence structures. However, interpretation of effects can be problematic and it is not possible to quantify the strength of the dependence among processes through well defined quantities (e.g., correlation or association parameters). In (c), multi-state processes are linked through a copula function. This permits substantial modelling flexibility for both the usual interpretation of time and covariate effects and for several types of associations amongst the transitions of the processes.

In this chapter, we extend the work by Diao & Cook (2014), laying the foundation for more general modelling of dependent multi-state survival processes. In particular, each process is modelled by means of the flexible framework described in Chapter 4, and a number of copulae (see Chapter 2, Table 2.1 for the available options) can be chosen to specify the association between the two processes. To contain the computational cost of this complex setting, estimation is based on a composite likelihood, built using the ”construction method” (Varin, 2008), as described in Section 5.2. In Section 5.3 we exemplify the approach using simulated data which recreates the setting explored in the reference work, i.e. two dependent progressive three-state processes. What we present here is only preliminary work, further developments are needed to achieve the full intent of the proposal. These next steps are detailed in Section 5.4.

5.2 Composite likelihood based estimation

Let $\{Z^{(1)}(t), t > 0\}$ and $\{Z^{(2)}(t), t > 0\}$ be two intermittently observed continuous-time multi-state Markov processes, each defined as in Chapter 4, with $\mathcal{S}^{(v)} = \{1, 2, \dots, C^{(v)}\}$ the discrete state space of process $v = 1, 2$, where $C^{(v)}$ is an absorbing state. Let N be the number of statistical units, $n_i^{(v)}$ the number of times process v is observed for the i^{th} unit, $0 = t_{i0}^{(v)} < t_{i1}^{(v)} < \dots < t_{in_i}^{(v)}$ its follow-up times, and $z_{i0}^{(v)}, z_{i1}^{(v)}, \dots, z_{in_i}^{(v)}$ its observed states. Let $\ell_v(\theta_v)$ be the log-likelihood associated with process v, where θ_v is the model parameter vector, $\mathcal{C} : (0,1)^2 \to (0,1)$ a uniquely defined 2-dimensional copula function with coefficient ϕ, and $\psi = (\theta_1{}^\top, \theta_2{}^\top, \phi)^\top$ the overall parameter vector for the joint model. Following Diao & Cook (2014), we define the following composite likelihood

$$\mathrm{CL}(\psi) = \exp\left(\ell_1(\theta_1) + \ell_2(\theta_2) + \sum_{i=1}^{n} \mathrm{DM}_i(\theta_1, \theta_2; \phi)\right), \tag{5.1}$$

where $\mathrm{DM}_i(\theta_1, \theta_2; \phi)$ represents the contribution of the i^{th} unit to the dependence model (DM) capturing the association between the two processes. In particular $\mathrm{DM}_i(\theta_1, \theta_2; \phi) = P\left(T_{C^{(v)}}^{(v)} \in (t_{in_i-1}, t_{in_i}], v = 1, 2; \psi\right)$, i.e. it is the joint probability that the times of the transitions into the absorbing states, i.e. $T_{C^{(v)}}^{(v)}$, are in the observed censoring intervals for both processes. Note that we are assuming that the dependence involves only the transition times into the absorbing states. This is motivated by the fact that, in the case study presented in the reference work, interest lies in the transitions into the last states for both processes, due to the clinical meaning that these states have. Alternative approaches would be to model the association in the first transition times, as done in Jiang & Cook (2020), or to model dependence in the sojourn time in a particular state. The assumption considered here provides a meaningful and tractable starting point for the development of our general framework as, to date, there are no approaches to flexibly model multiple dependant multi-state processes. Future work will focus on generalising the specification of the dependence structure. Care will be needed in this case, as it will imply a significant increase in the complexity of the model. As we are using copulae to express the dependence between the processes, the i^{th} contribution of the dependence model $\mathrm{DM}_i(\theta_1, \theta_2; \phi)$ is given by

$$\begin{cases} \mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i-1}), \mathcal{F}_2(t^{(2)}_{in_i-1}); \phi) + \mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i}), \mathcal{F}_2(t^{(2)}_{in_i}); \phi) & \text{if } (z^{(1)}_{in_i}, z^{(2)}_{in_i}) \text{ are absorbing states} \\ -\mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i-1}), \mathcal{F}_2(t^{(2)}_{in_i}); \phi) - \mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i}), \mathcal{F}_2(t^{(2)}_{in_i-1}); \phi), & \\ & \\ \mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i}), \mathcal{F}_2(t^{(2)}_{in_i-1}); \phi) - \mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i}), \mathcal{F}_2(t^{(2)}_{in_i}); \phi), & \text{if only } z^{(2)}_{in_i} \text{ is an absorbing state} \\ & \\ \mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i-1}), \mathcal{F}_2(t^{(2)}_{in_i}); \phi) - \mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i}), \mathcal{F}_2(t^{(2)}_{in_i}); \phi), & \text{if only } z^{(1)}_{in_i} \text{ is an absorbing state} \\ & \\ \mathcal{C}(\mathcal{F}_1(t^{(1)}_{in_i}), \mathcal{F}_2(t^{(2)}_{in_i}); \phi), & \text{if } (z^{(1)}_{in_i}, z^{(2)}_{in_i}) \text{ are not absorbing states} \end{cases}$$

where $\mathcal{F}_v(t) = 1 - p^{(v,1C)}(0,t)$ is the marginal survivor function of the entry time to the absorbing state $C^{(v)}$ and $p^{(v,1C)}(0,t)$ is the $(1,C)$ element of the transition probability matrix $\mathbf{P}^{(v)}(0,t)$ of process v, in the time interval $(0,t)$. As discussed in Chapter 4, computing the transition probabilities for a general time-dependent multi-state process is a non-trivial task. Here, we exploit the Chapman-Kolmogorov equation to obtain $\mathcal{F}_v(t^{(v)}_{in_i-1})$ and $\mathcal{F}_v(t^{(v)}_{in_i})$ respectively by

$$\begin{aligned} \mathbf{P}^{(v)}(0, t_{in_i-1}) &= \prod_{j=1}^{n_i-1} \mathbf{P}^{(v)}(t_{ij-1}, t_{ij}), \\ \mathbf{P}^{(v)}(0, t_{in_i}) &= \mathbf{P}^{(v)}(0, t_{in_i-1}) \times \mathbf{P}^{(v)}(t_{in_i-1}, t_{in_i}). \end{aligned} \tag{5.2}$$

Note that the terms $\mathbf{P}^{(v)}(t_{ij-1}, t_{ij})$ need be computed for the contributions to the log-likelihood $\ell_v(\theta_v)$ as well, thus, at the cost of storing these matrices, they are already available for the dependence model. The analytical gradient of the composite log-likelihood then follows

$$\frac{\partial}{\partial\psi}\mathrm{CL}(\psi)=\begin{bmatrix}\frac{\partial}{\partial\theta_1}\ell_1(\theta_1)+\frac{\partial}{\partial\theta_1}\sum_{i=1}^{n}\log\left(\mathrm{DM}_i(\theta_1,\theta_2;\phi)\right)\\ \frac{\partial}{\partial\theta_2}\ell_2(\theta_2)+\frac{\partial}{\partial\theta_2}\sum_{i=1}^{n}\log\left(\mathrm{DM}_i(\theta_1,\theta_2;\phi)\right)\\ \frac{\partial}{\partial\phi}\sum_{i=1}^{n}\log\left(\mathrm{DM}_i(\theta_1,\theta_2;\phi)\right)\end{bmatrix}, \tag{5.3}$$

where $\partial/\partial\theta_v\ell_v(\theta_v)$ is the gradient associated with the marginal process v, obtained using the framework described in Chapter 4. When $(z_{in_i}^{(1)},z_{in_i}^{(2)})$ are both absorbing states, the i^{th} contribution to the first derivative of the DM term is defined as follows

$$\begin{aligned}&\frac{\partial}{\partial\theta_1}\log\left(\mathrm{DM}_i(\theta_1,\theta_2;\phi)\right)=\mathrm{DM}_i(\theta_1,\theta_2;\phi)^{-1}\\ &\cdot\left[\frac{\partial\mathcal{C}(\mathcal{F}_1(t_{in_i-1}^{(1)}),\mathcal{F}_2(t_{in_i-1}^{(2)});\phi)}{\partial\mathcal{F}_1(t_{in_i-1}^{(1)})}\frac{\partial\mathcal{F}_1(t_{in_i-1}^{(1)})}{\partial\theta_1}+\frac{\partial\mathcal{C}(\mathcal{F}_1(t_{in_i}^{(1)}),\mathcal{F}_2(t_{in_i}^{(2)});\phi)}{\partial\mathcal{F}_1(t_{in_i}^{(1)})}\frac{\partial\mathcal{F}_1(t_{in_i}^{(1)})}{\partial\theta_1}\right.\\ &\left.-\frac{\partial\mathcal{C}(\mathcal{F}_1(t_{in_i-1}^{(1)}),\mathcal{F}_2(t_{in_i}^{(1)});\phi)}{\partial\mathcal{F}_1(t_{in_i-1}^{(1)})}\frac{\partial\mathcal{F}_1(t_{in_i-1}^{(1)})}{\partial\theta_1}-\frac{\partial\mathcal{C}(\mathcal{F}_1(t_{in_i}^{(1)}),\mathcal{F}_2(t_{in_i-1}^{(2)});\phi)}{\partial\mathcal{F}_1(t_{in_i}^{(1)})}\frac{\partial\mathcal{F}_1(t_{in_i}^{(1)})}{\partial\theta_1}\right],\end{aligned}$$

similar expressions will hold for $\partial/\partial\theta_2\log\left(\mathrm{DM}_i(\theta;\phi)\right)$. For the cases where $z_{in_i}^{(1)}$ or $z_{in_i}^{(2)}$ are not absorbing states, the expressions of the derivatives are special cases of the above. The terms $\partial\mathcal{C}(\mathcal{F}_1(t),\mathcal{F}_2(t');\phi)/\partial\mathcal{F}_1(t)$ are copula densities, whose form will depend on the copula function chosen. The derivatives of the marginal survivor functions $\partial\mathcal{F}_v(t)/\partial\theta_v=-\partial p^{(v,1C)}(0,t)/\partial\theta_v$, for $t=t_{in_i-1}$ and $t=t_{in_i}$ respectively, follow from (5.2) and the application of the chain rule

$$\begin{aligned}\frac{\partial}{\partial\theta_v}\mathbf{P}^{(v)}(0,t_{in_i-1})&=\sum_{j=1}^{n_i-1}\mathbf{P}^{(v)}(0,t_{ij-1})\left(\frac{\partial}{\partial\theta_v}\mathbf{P}^{(v)}(t_{ij-1},t_{ij})\right)\mathbf{P}^{(v)}(t_{ij},t_{in_i-1}),\\ \frac{\partial}{\partial\theta_v}\mathbf{P}^{(v)}(0,t_{in_i})&=\left(\frac{\partial}{\partial\theta_v}\mathbf{P}^{(v)}(0,t_{in_i-1})\right)\mathbf{P}^{(v)}(t_{in_i-1},t_{in_i})+\mathbf{P}^{(v)}(0,t_{in_i})\left(\frac{\partial}{\partial\theta_v}\mathbf{P}^{(v)}(t_{in_i-1},t_{in_i})\right),\end{aligned}$$

where the terms that span over more than a single time-interval are, in turn, computed by repeatedly applying the Chapman-Kolmogorov equation to the stored set of matrices $\mathbf{P}^{(v)}(t_{ij-1},t_{ij})$. The derivatives $\partial/\partial\theta_v\mathbf{P}^{(v)}(t_{ij-1},t_{ij})$ need be computed for the contributions to the gradient $\partial/\partial\theta_v\ell_v(\theta_v)$, thus, at the cost of storing these matrices, they are already available for the gradient of the DM.

Note that, following Diao & Cook (2014), the full likelihood arising from intermittent inspection of a joint multi-state process fixed inspection times is given by

$$L(\psi) = P(T_c^{(v)} \in (l_c^{(v)}, r_c^{(v)}]; c = 1, \ldots, C^{(v)} - 1; v = 1, 2; \psi), \tag{5.4}$$

where, for process $v = 1,2$, $T_c^{(v)}$ represents the $c \to c+1$ transition time and $(l_c^{(v)}, r_c^{(v)}]$ represent the left and right end points of the censoring interval for this transition. The likelihood in (5.4) is obtained by computing $2 \times (C^{(v)} - 1)-$dimensional integrals over the joint density. In the setting considered here, $C^{(v)} = 3$ and the joint density, expressed in terms of one of the possible copula-based decompositions, is given by

$$\begin{aligned} f(t;\psi) =& f(t_1^{(1)}, t_2^{(1)}; \theta_1) \cdot c(\mathcal{F}_1(t_2^{(1)}), \mathcal{F}_2(t_2^{(2)})) \cdot f(t_1^{(2)}, t_2^{(2)}; \theta_2) \\ & \cdot c(\mathcal{F}(t_1^{(1)} \mid t_2^{(1)}), \mathcal{F}(t_2^{(2)} \mid t_2^{(1)})) \cdot c(\mathcal{F}(t_2^{(1)} \mid t_2^{(2)}), \mathcal{F}(t_1^{(2)} \mid t_2^{(2)})) \\ & \cdot c(\mathcal{F}(t_1^{(1)} \mid t_2^{(1)}, t_2^{(2)}), \mathcal{F}(t_1^{(2)} \mid t_2^{(1)}, t_2^{(2)})), \end{aligned}$$

where $t_c^{(v)}$ are the (unobserved) realisations of $T_c^{(v)}$, $c(\cdot,\cdot)$ indicates the copula density function and where we are omitting the copula parameters for simplicity. Integrals of this function over the four-dimensional cube defined by the (observed) censoring intervals $\{(l_c^{(v)}, r_c^{(v)}]\}_{v=1,2}^{c=1,2}$ are required to compute the full joint likelihood. It follows that the likelihood involves computationally demanding high-dimensional integration, particularly when the number of processes or the number of states increases. For this reason, we adopt a composite likelihood-based approach, where it is not necessary to specify the full joint process for estimation. In fact, the use of composite likelihood enables some simplification in the model specification and increases the robustness to model misspecification.

We exemplify the proposed method on a toy example based on simulated data. Estimation is carried out through the algorithm described in the previous chapters, which combines a trust region algorithm with an automatic multiple smoothing parameter selection algorithm, adapted to this novel setting. The objective function is the composite log-likelihood defined in (5.1) and the analytical gradient defined in (5.3) is provided. The Hessian is computed numerically as the Jacobian of the analytical gradient. This has been found to be more accurate, in practice, than the numerical Hessian obtained from the log-likelihood. Ideally, the analytical Hessian should be provided to the estimation algorithm, as done in the previous chapters. Due to its complex structure, more work is needed to derive and implement it. In the toy example presented in Section 5.3, the numerical Jacobian can be obtained efficiently and can thus be used. As discussed in Chapter 4, this will in general be insufficiently accurate and too computationally expensive to support real-world data applications and/or non-trivial process structures. However, this is a first step that allows us to test simple versions of joint multi-state models, thus setting the foundations for the generalisation of the flexible framework proposed here. Work for this is currently underway and relies upon these preliminary results.

5.3 Toy example

We assume two intermittently observed progressive time-homogeneous three-state Markov processes, where state 1 represents a “normal” condition, state 2 represents an “abnormal” condition, and state 3 represents the absorbing state of “organ damage”. Figure 5.1 represent our setting graphically, highlighting the association between the absorbing states. The observations are simulated as described in Diao & Cook (2014). $N = 1000$ individuals are observed at ten common inspection times, evenly spaced over the time interval $(0,1]$. For process v, the time-constant intensity of the transition $r \to r'$ is defined as

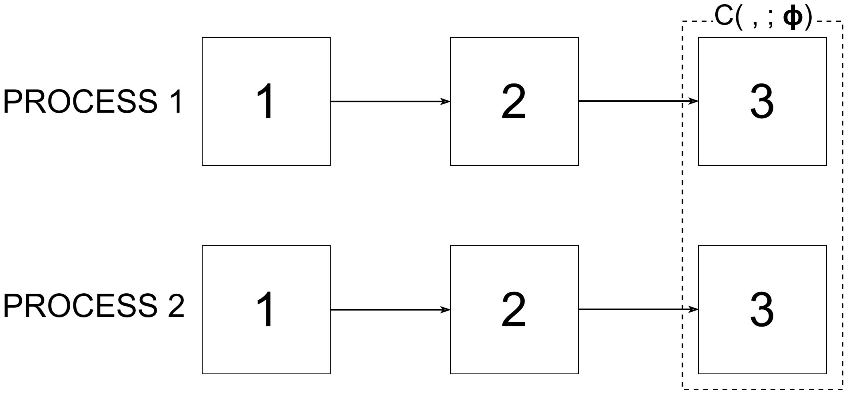

Figure 5.1: Two dependent progressive three-state processes. The association is present in the absorbing states and is captured via a copula-based approach.

$$q^{(v,rr')} = \exp\left[\beta_0^{(v,rr')} + x\,\beta_1^{(v,rr')}\right],$$

where x is a binary covariate observed at the baseline time for each individual. For simplicity, the parameters are set to be equal across the two processes which is equivalent to assuming that the processes are clustered, and they are chosen under the following constraints: (i) $\exp\left(\beta_0^{(v,23)}\right) = 1.5\exp\left(\beta_0^{(v,12)}\right)$, i.e. the baseline transition rate out of state 2 is 1.5 times of that out of state 1; (ii) the joint probability that the processes are in the respective absorbing states, when the binary covariate is null, is $P\left(Z^{(1)}(1) = 3, Z^{(2)}(1) = 3; x = 0\right) = 0.4$; (iii) the binary covariate x has the effect of mildly increasing the risk of transition from state 1 to state 2; (iv) the binary covariate x has the effect of moderately increasing the risk of transition from state 2 to state 3. These constraints give $\beta_0^{(v,12)} = \log(1.8148)$, $\beta_0^{(v,23)} = \log(2.7221)$, $\beta_1^{(v,12)} = \log(1.25)$ and $\beta_1^{(v,23)} = \log(1.4)$. For the association model, we assume a Clayton copula with a strong dependence, in particular Kendall's $\tau = 0.8$, giving the copula parameter $\phi = 8$.

Under this setting and using the method described in Section 5.2, we are able to retrieve the true underlying parameters, as shown in Table 5.1. The small discrepancies can be explained by the loss of efficiency implied by the use of the composite log-likelihood in place of the full log-likelihood of the joint model. We expect these to improve with the use of the exact Hessian, as this provides more accurate second order information of the objective function. Alternative definitions of the composite likelihood can also be explored and may lead to an improved performance, as shown in the reference work.

Finally, note that at the maximum likelihood estimates $\hat{\psi}$, the maximum in the absolute value of gradient vector is of the order of 10^{-6}, while the smallest eigenvalue of the Jacobian is strictly positive and far from zero ($\approx 10^2$), thus ensuring that a true maximum of the objective function has been found.

	Process 1		Process 2	
	True	Estim.	True	Estim.
$\beta_0^{(v,12)}$	0.596	0.685	0.596	0.619
$\beta_1^{(v,12)}$	0.223	0.164	0.223	0.258
$\beta_0^{(v,23)}$	1.001	0.983	1.001	1.086
$\beta_1^{(v,23)}$	0.336	0.403	0.336	0.241

	True	Estim.
$\log(\phi)$	2.079	2.121

Table 5.1: True and estimated parameters for the data generating process of the two dependent three-state processes.

When handling two associated multi-state processes, interest lies, for example, in making statements on the future course of a single phenomenon, while accounting for the effect that the evolution of the other has on the former. This can be quantified through a conditional probability such as $P(Z^{(1)}(t) = 3 \mid \mathbf{Z}(t_0))$, i.e. the probability that the first process has a high degree of severity given the states occupied by both processes at the beginning of the observation period. Here $\mathbf{Z}(t) = (Z^{(1)}(t), Z^{(2)}(t))$ and t_0 is the starting time. Similarly, one may be interested in the joint probability of observing ”organ damage”, i.e. $P\left(Z^{(1)}\left(t^{(1)}\right) = 3, Z^{(2)}\left(t^{(2)}\right) = 3; x\right)$ where $t^{(1)}$ and $t^{(2)}$ are the times in the corresponding process, given the patient characteristics. For example, this is meaningful in individuals with Psoriatic Arthritis (PsA), since one of the New York criteria (Moll & Wright, 1973) for diagnosis of ankylosing spondylitis, a complication of PsA, is satisfied if $\left(Z^{(1)}(t), Z^{(2)}(t)\right) = (3,3)$, where each process represents one sacroiliac joint and state 3 here represents the most severe degree of damage. In Figure 5.2, using the running toy example considered in this section, we plot the joint probability of observing ”organ damage” in the two processes as a bivariate function of time, while setting the binary covariate x to 0 and 1 (left and right pane, respectively). The two surfaces obtained in this way, confirm the effect of x as being that of increasing the risk of transitioning towards later stages, as per the true data generating process. Sections of the joint probability surface can, in turn, be interpreted as conditional probabilities of one process being in state 3, given that the other process was observed in this last state at a given time. This viewpoint provides insight on how disease will develop in one process, when the history of the other process is known and accounted for.

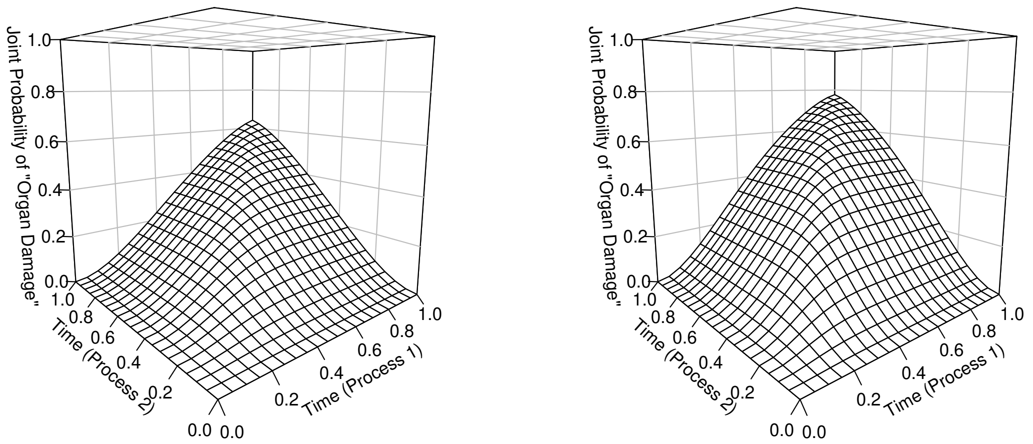

Figure 5.2: Joint probability of observing organ damage in both processes, i.e. $P\left(Z^{(1)}\left(t^{(1)}\right) = 3, Z^{(2)}\left(t^{(2)}\right) = 3; x\right)$. In the left pane $x = 0$, in the right pane $x = 1$.

5.4 Discussion

This chapter lays the foundation for a more general and flexible approach to the modelling of dependent multi-state processes. In contrast to the reference work from Diao & Cook (2014), in fact, the transition intensities can be specified through a flexible additive predictor-based model, which allows for virtually any type of time and covariate effects. The code, available upon request, is based on the general modelling approach discussed in Chapter 4, which supports time-dependent processes and poses no limitations on the number of states or on the types of transitions allowed. The dependence model, in turn, can be obtained via any of the copulae listed in Chapter 2, Table 2.1. In the reference work, estimation is based on a two-stage approach and a general purpose optimiser which relies on the analytical expression of the model log-likelihood alone is used. A two-stage estimation approach may make the fitting problem easier to deal with, in exchange for some loss in efficiency. However, the use of the composite likelihood, in place of the full joint likelihood, already implies a loss of efficiency, thus making the two-stage approach inapt at supporting more complex model structures. The estimation algorithm proposed in this chapter, thus, relies on a simultaneous approach which uses the analytical expression of the gradient, as well as second order information of the objective function. Through extensive experimentation we have found that the latter is needed to achieve model identification, and more accurate inference, in such challenging settings.

Through a toy example, based on simulated data, we have shown that the proposed framework adequately recovers the parameters of the true data generating process. This represents a preliminary experiment and relies on second order information provided by the Jacobian of the analytical gradient, which we have found to be more accurate than the numerical Hessian obtained from the objective function. This approximation is sufficient to identify the basic setting explored here and in the reference work. In general, the analytical expression of the Hessian is warranted, as this allows us to achieve the optimal convergence rate known to hold for the trust region algorithm when the exact Hessian is employed and provides crucial information on the objective function.

Future work will thus entail deriving the analytical expression of the Hessian and implementing it. The developed code will be integrated in the `R` package `flexmsm` to provide the end-user with a host of tools for (multiple) multi-state survival modelling. Further, the investigation of a case study where interest lies in the joint modelling of retinopathy and nephropathy in diabetic patients is currently under way. These diseases are both vascular in nature and are thus expected to be associated. The use of the modelling framework developed in Chapter 4 for the individual processes, combined with the approach described here, will allow us to investigate the impact of time and of potential risk factors in a more flexible way than currently allowed in the literature, while modelling the dependence between the two conditions. The aim is to achieve a deeper understanding of disease course. For example, the level of HbA1c level, i.e. the amount of blood sugar attached to the hemoglobin, has been identified as a risk factor from the related literature, but has only ever been included log-linearly in models.

From a methodological perspective, future efforts will focus on generalising the dependence structure, both in terms of the number of processes supported as well as the in way the processes are associated to one another. For the former aim, we will investigate the methods discussed in Chapter 2 in the context of multi-state event times, including multivariate Archimedean copulae, pair-copulae constructions, the multivariate Gaussian and Student's t distributions. In regard to the latter aim, one may be interested in modelling associations between transition times which are different from the absorbing times. The choice of where the links between the processes should be placed is far from trivial. It can be motivated by the subject matter expert or it can be inferred by the data. Future efforts will go towards developing an approach for the latter, where a general dependence structure is defined as a starting point and links which are not supported by the data are dropped at a later step.

The additional generality sought will imply an increase in the computational burden of the framework, which is already intrinsically high. Interest thus lies in investigating alternative ways to handle this methodological and computational challenge, thus providing ample space for future developments.

Overall, there are numerous avenues for future research in the area of joint multi-state modelling and they are motivated by the high clinical relevance of improving our understanding of associated disease pathways.

Chapter 6

General Conclusions and Future Research

In the present thesis, we propose four general frameworks, tied together by a number of common themes, for the modelling of complex survival outcomes.

At the core of each, there is the flexible modelling of the time-to-events by means of splines-based additive predictors. This allows us to specify a variety of time and covariate effects, thus enabling us to uncover the complex dynamics of the unfolding over time of the events of interest.

Through extensive practical experimentation, we recognised the importance of supporting the flexible modelling of the complex survival outcomes through a carefully designed estimation algorithm, which makes an optimal use of the information provided in the data. Time-to-event data is, in fact, inherently characterised by a systematic lack of information, due to censoring. This, combined with the complexity of the model structure, implies an increase in the methodological and practical difficulty of the optimisation problem. We thus propose a stable and efficient penalised likelihood based estimation approach, which relies on the use of the analytical expressions of the gradient and of the Hessian.

We then recognised the need for general software supporting each of the methodological developments proposed, and the lack thereof in the literature. This provides the end-user with the tools necessary to define, fit and visualise the output of the flexible models for each of the composite survival outcomes discussed.

On a high level, there is a common thread connecting the four frameworks. Chapter 2 proposes as copula-based approach to modelling two dependent time-to-events, while retaining the interpretability and flexibility of the marginal model specifications. Chapters 3 and 4 move the focus from traditional survival outcomes, where a single event is of interest, to multi-state survival outcomes, addressing the cases in which the transition times are known exactly (or censored) and in which they are only known to lie within a certain interval (or censored), respectively. Finally, Chapter 5, building on the previous chapters, proposes a copula-based framework to model to two dependent multi-state survival processes, each of which are flexibly defined.

The bivariate survival events discussed in Chapter 2 often arise in clinical trials studying diseases concerning paired organs, where the outcomes of interest are measured on the same individual and are therefore associated. Ignoring this dependence leads to biased estimates and thus an improper understanding of the disease mechanism. Interest also lies in quantifying the strength of the dependence, an aim which is supported by the straightforward interpretation of the copula dependence parameter. In this work we propose to model the (possibly mixed-censored) time-to-events and the copula parameter by means of additive predictors. In this way, smooth effects of time and of the risk factors of interest can be modelled for both the individual disease manifestations as well as the strength of their association, which can thus vary as a function of patient characteristics. Future research will focus on extending the approach to more than two event times (e.g., multi-morbidity), to accounting for informative and/or dependent censoring, as well as to considering the case of excess hazard modelling, all of which are of practical interest to subject matter experts.

Chapters 3 and 4 are motivated by longitudinal health data where each individual may experience multiple disease manifestations, and potentially death. Here, interest lies in predicting the disease trajectory given the patient characteristics at baseline as well as dynamically over time. Multi-state models represent a versatile and powerful tool to do so, as they allow each stage to be specified as a separate state and then capture the path through the collection of states by means of the transition intensities, which are specified as flexible functions of time and of the covariates of interest. Crucially, the cases where the time-to-events are known exactly and where the process is observed only intermittently require vastly different methodological treatments. This motivated the development of two separate frameworks and related software implementations. In this way, the end-user can benefit from the strengths of each setting without paying the costs of the other. For each observation scheme, we have provided a different approach to compute the transition probabilities, a key quantity for the interpretation of the multi-state model due its more intuitive scale compared to the transition intensities, whose computation is not trivial. The exact information available in the continuously observed setting implies that both the Markov and the semi-Markov assumption can be made, with no changes in the model specification except for the time-scale, which needs to be reset at each transition in the latter case. To reflect this generality also in the computation of the transition probabilities, we proposed a simulation based approach, which can be used in the same manner regardless of the time-scale chosen. In contrast to this, the loss of information, in which we incur when the process is only intermittently observed, implies a considerable methodological difficulty which is usually handled in the literature by assuming the process is Markov. Further, in this case, the transition probabilities are needed for estimation as well, and thus need to be computed in a far more efficient way. This motivated the tedious derivation of the closed form expression of the second derivatives of the transition probability matrix, which was not available in the literature, thus far, but which was a necessary element to adequately support estimation in this complex case. Future work in the continuously observed setting will focus on improving the integration between the software supporting the flexible modelling of the transition intensities and the computation of the predicted transition probabilities. We are also interested in implementing, for this setting, an alternative general approach to compute the transition probabilities which is based on numerically solving the Kolmogorov differential equations. In the context of intermittently observed processes, the most relevant when it comes to health registry data, we are interested in improving the scalability of the framework, through computational improvements (e.g. the use of C++) and/or theoretical results. Other potentially interesting directions include experimenting with relaxations of the Markov property; investigating alternative ways to compute the transition probability matrix, such as through Padé or Taylor expansion based approximation (this has partially been explored, but has not been included in the present work due its inferior performance compared to the closed form expressions); adding diagnostic tools to better understand convergence failures (the previous point provides one such tool, as it would allows the user to carry out sensitivity analyses on the computation of the transition probability matrix).

Chapter 5 recovers the motivation at the base of Chapter 2 and extends it the case of multi-state survival outcomes, which are modelled using the framework developed in Chapter 4. This reflects the more complex case where interest lies in modelling multiple degrees of severity of two interdependent diseases, rather than two dependent outcomes. For example, the score-based classification of the stages of retinopathy and nephropathy lends itself naturally to being represented through a multi-state process. These are two diseases which arise as a consequence of diabetes and are expected to be related due to both being vascular complications. Interest then lies in modelling the unfolding of each, while accounting for how the progression of retinopathy may affect the progression of nephropathy, and vice versa. This is a complex setting, with multiple methodological and practical challenges. In this work, we have begun to explore it in a simple but representative setting, with two three-state processes linked together through the transition times into their respective absorbing states. In this way, we provide a foundation for an approach that is more general than that currently available in the literature. Future work will focus on exploring alternative dependence structures, beyond that of tying the absorbing times together, and in extending the work to accommodate more than two processes, to reflect diseases affecting multi-organ systems such as Psoriatic Arthritis, which affects the joints. Investigating solutions to contain the computational burden of this complex setting is also of interest, particularly when extending it to support more general structures.

Overall, the future work in the field of complex survival outcomes is motivated by the needs sparking from health registry data and clinical applications, particularly those of improving the accuracy of predictions and available tools for the analysis of disease patterns. Several computational and methodological challenges arise due to the multiple levels of difficulty characterising this problem, which thus warrants and provides a fertile ground for statistical innovation and the development of novel computational solutions.

Appendix A

Supplementary Material A

A.1 Log-likelihood

The more explicit version of the log-likelihood is

$$
\begin{aligned}
\ell(\delta) = \sum_{i=1}^{n} & \gamma_{U_{1i}}\gamma_{U_{2i}} \log\left[\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot G_2'(\eta_{2i}(t_{2i}))\right. \\
& \qquad\qquad\qquad \left. \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}}\right] \\
& + \gamma_{R_{1i}}\gamma_{R_{2i}} \log\Big[C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}\Big] \\
& + \gamma_{L_{1i}}\gamma_{L_{2i}} \log\Big[1 - G_1(\eta_{1i}(l_{1i})) - G_2(\eta_{2i}(l_{2i})) + C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}\Big] \\
& + \gamma_{I_{1i}}\gamma_{I_{2i}} \log\Big[C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\} \\
& \qquad\qquad - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} + C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}\Big] \\
& + \gamma_{U_{1i}}\gamma_{R_{2i}} \log\left[-\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}\right] \\
& + \gamma_{R_{1i}}\gamma_{U_{2i}} \log\left[-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))} \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}}\right] \\
& + \gamma_{U_{1i}}\gamma_{L_{2i}} \log\left[\left(\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} - 1\right) \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}\right] \\
& + \gamma_{L_{1i}}\gamma_{U_{2i}} \log\left[\left(\frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))} - 1\right) \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}}\right] \\
& + \gamma_{U_{1i}}\gamma_{I_{2i}} \log\left[\left(\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} - \frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\right)\right. \\
& \qquad\qquad\qquad \left. \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\gamma_{I_{1i}}\gamma_{U_{2i}}\log\Bigg[\bigg(\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))}-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))}\bigg)\\
&\qquad\qquad\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial t_{2i}}\Bigg]\\
&+\gamma_{R_{1i}}\gamma_{L_{2i}}\log\Big[G_1(\eta_{1i}(r_{1i}))-C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}\Big]\\
&+\gamma_{L_{1i}}\gamma_{R_{2i}}\log\Big[G_2(\eta_{2i}(r_{2i}))-C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}\Big]\\
&+\gamma_{R_{1i}}\gamma_{I_{2i}}\log\Big[C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}-C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}\Big]\\
&+\gamma_{I_{1i}}\gamma_{R_{2i}}\log\Big[C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}-C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}\Big]\\
&+\gamma_{L_{1i}}\gamma_{I_{2i}}\log\Big[G_2(\eta_{2i}(l_{2i}))-G_2(\eta_{2i}(r_{2i}))+C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}\\
&\qquad\qquad-C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}\Big]\\
&+\gamma_{I_{1i}}\gamma_{L_{2i}}\log\Big[G_1(\eta_{1i}(l_{1i}))-G_1(\eta_{1i}(r_{1i}))+C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}\\
&\qquad\qquad-C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}\Big].
\end{aligned}
$$

Derivation of each term

For $v = 1,2$, $i = 1,...,n$, we define the following set of dummy variables:

$$
\gamma_{U_{vi}}=\begin{cases}1, & \text{if the i-th obs. is uncensored}\\ 0, & \text{otherwise}\end{cases}
\qquad
\gamma_{L_{vi}}=\begin{cases}1, & \text{if the i-th obs. is left-censored}\\ 0, & \text{otherwise}\end{cases}
$$

$$
\gamma_{R_{vi}}=\begin{cases}1, & \text{if the i-th obs. is right-censored}\\ 0, & \text{otherwise}\end{cases}
\qquad
\gamma_{I_{vi}}=\begin{cases}1, & \text{if the i-th obs. is interval-censored}\\ 0, & \text{otherwise}\end{cases}
$$

In the bivariate case the log-likelihood function is made up of sixteen terms corresponding to the following combinations of the indicator terms:

The derivation of each of these terms follows:

	Uncens	Left-cens	Right-cens	Interval-cens
Uncens	$\gamma_{U_{1i}}\ \gamma_{U_{2i}}$	$\gamma_{U_{1i}}\ \gamma_{L_{2i}}$	$\gamma_{U_{1i}}\ \gamma_{R_{2i}}$	$\gamma_{U_{1i}}\ \gamma_{I_{2i}}$
Left-cens	$\gamma_{L_{1i}}\ \gamma_{U_{2i}}$	$\gamma_{L_{1i}}\ \gamma_{L_{2i}}$	$\gamma_{L_{1i}}\ \gamma_{R_{2i}}$	$\gamma_{L_{1i}}\ \gamma_{I_{2i}}$
Right-cens	$\gamma_{R_{1i}}\ \gamma_{U_{2i}}$	$\gamma_{R_{1i}}\ \gamma_{L_{2i}}$	$\gamma_{R_{1i}}\ \gamma_{R_{2i}}$	$\gamma_{R_{1i}}\ \gamma_{I_{2i}}$
Interval-cens	$\gamma_{I_{1i}}\ \gamma_{U_{2i}}$	$\gamma_{I_{1i}}\ \gamma_{L_{2i}}$	$\gamma_{I_{1i}}\ \gamma_{R_{2i}}$	$\gamma_{I_{1i}}\ \gamma_{I_{2i}}$

- T_{1i} uncensored and T_{2i} uncensored (in this case $t_{1i} = r_{1i} = l_{1i}$ and $t_{2i} = r_{2i} = l_{2i}$):

$$
\begin{aligned}
f(t_{1i},t_{2i}) &= \frac{\partial^2}{\partial t_{1i}\partial t_{2i}} F(t_{1i},t_{2i}) \\
&= \frac{\partial^2}{\partial t_{1i}\partial t_{2i}} [1 - S(t_{1i}) - S(t_{2i}) + S(t_{1i},t_{2i})] = \\
&= \frac{\partial^2}{\partial t_{1i}\partial t_{2i}} C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\} = \\
&= \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}}.
\end{aligned}
$$

- T_{1i} right-censored and T_{2i} right-censored:

$$
P(T_{1i} > r_{1i}, T_{2i} > r_{2i}) = S(r_{1i}, r_{2i}) = C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}.
$$

- T_{1i} left-censored and T_{2i} left-censored:

$$
\begin{aligned}
P(T_{1i} < l_{1i}, T_{2i} < l_{2i}) &= F(l_{1i}, l_{2i}) = P(T_{1i} < l_{1i}) - [P(T_{2i} > l_{2i}) - S(l_{1i}, l_{2i})] = \\
&= 1 - S_1(l_{1i}) - S_2(l_{2i}) + S(l_{1i}, l_{2i}) = \\
&= 1 - G_1(\eta_{1i}(l_{1i})) - G_2(\eta_{2i}(l_{2i})) + C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}.
\end{aligned}
$$

- T_{1i} uncensored and T_{2i} right-censored (the swapped case can be trivially derived by switching the subscripts where required):

$$\begin{aligned}
&\int_{r_{2i}}^{+\infty} f(t_{1i},y)dy = \int_{0}^{+\infty} f(t_{1i},y)dy - \int_{0}^{r_{2i}} f(t_{1i},y)dy = \\
&= f_1(t_{1i}) - \frac{\partial}{\partial t_{1i}} F(t_{1i},r_{2i}) = f_1(t_{1i}) - \frac{\partial}{\partial t_{1i}}[1 - S_1(t_{1i}) - S_2(t_{2i}) + S(t_{1i},r_{2i})] = \\
&= f_1(t_{1i}) - f_1(t_{1i}) - \frac{\partial}{\partial t_{1i}} S(t_{1i},r_{2i}) = -\frac{\partial}{\partial t_{1i}} C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\} = \\
&= -\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}.
\end{aligned}$$

- T_{1i} interval-censored and T_{2i} interval-censored:

$$\begin{aligned}
&P(l_{1i} < T_{1i} < r_{1i}, l_{2i} < T_{2i} < r_{2i}) = \\
&= P(T_{1i} < r_{1i}, T_{2i} < r_{2i}) - P(T_{1i} < l_{1i}, T_{2i} < r_{2i}) - \\
&- P(T_{1i} < r_{1i}, T_{2i} < l_{2i}) + P(T_{1i} < l_{1i}, T_{2i} < l_{2i}) = \\
&= F(r_{1i},r_{2i}) - F(l_{1i},r_{2i}) - F(r_{1i},l_{2i}) + F(l_{1i},l_{2i}) = \\
&= [1 - S_1(r_{1i}) - S_2(r_{2i}) + S(r_{1i},r_{2i})] - [1 - S_1(l_{1i}) - S_2(r_{2i}) + S(l_{1i},r_{2i})] + \\
&\quad - [1 - S_1(r_{1i}) - S_2(l_{2i}) + S(r_{1i},l_{2i})] + [1 - S_1(l_{1i}) - S_2(l_{2i}) + S(l_{1i},l_{2i})] = \\
&= S(l_{1i},l_{2i}) - S(l_{1i},r_{2i}) - S(r_{1i},l_{2i}) + S(r_{1i},r_{2i}) = \\
&= C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\} + \\
&\quad - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} + C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}.
\end{aligned}$$

- T_{1i} right-censored and T_{2i} left-censored (the swapped case can be trivially derived by switching the subscripts where required):

$$\begin{aligned}
&P(T_{1i} > r_{1i}, T_{2i} < l_{2i}) = P(T_{2i} < l_{2i}) - P(T_{1i} < r_{1i}, T_{2i} < l_{2i}) = F_2(l_{2i}) - F(r_{1i},l_{2i}) = \\
&= 1 - S_2(l_{2i}) - [1 - S_1(r_{1i}) - S_2(l_{2i}) + S(r_{1i},l_{2i})] = \\
&= G_1(\eta_{1i}(r_{1i})) - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}.
\end{aligned}$$

- T_{1i} uncensored and T_{2i} left-censored (the swapped case can be trivially derived by switching the subscripts where required):

$$\int_0^{l_{2i}} f(t_{1i},y)dy = \frac{\partial}{\partial t_{1i}} F(t_{1i},l_{2i}) =$$
$$= \frac{\partial}{\partial t_{1i}}[1 - S_1(t_{1i}) - S_2(l_{2i}) + S(t_{1i},l_{2i})] =$$
$$= -\frac{\partial}{\partial t_{1i}} G_1(\eta_{1i}(t_{1i})) + \frac{\partial}{\partial t_{1i}} C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\} =$$
$$= -G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} + \frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} =$$
$$= \left[\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} - 1\right] \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}.$$

- T_{1i} uncensored and T_{2i} interval-censored (the swapped case can be trivially derived by switching the subscripts where required):

$$\int_{l_{2i}}^{r_{2i}} f(t_{1i},y)dy = \int_0^{r_{2i}} f(t_{1i},y)dy - \int_0^{l_{2i}} f(t_{1i},y)dy =$$
$$= \frac{\partial}{\partial t_{1i}} F(t_{1i},r_{2i}) - \frac{\partial}{\partial t_{1i}} F(t_{1i},l_{2i}) =$$
$$= \frac{\partial}{\partial t_{1i}}[1 - S_1(t_{1i}) - S_2(r_{2i}) + S(t_{1i},r_{2i})] - \frac{\partial}{\partial t_{1i}}[1 - S_1(t_{1i}) - S_2(l_{2i}) + S(t_{1i},l_{2i})] =$$
$$= \frac{\partial}{\partial t_{1i}} C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\} - \frac{\partial}{\partial t_{1i}} C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\} =$$
$$= \left[\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} - \frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\right] \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}.$$

- T_{1i} right-censored and T_{2i} interval-censored (the swapped case can be trivially derived by switching the subscripts where required):

$$P(T_{1i} > r_{1i}, l_{2i} < T_{2i} < r_{2i}) =$$
$$= P(T_{2i} < r_{2i}) - P(T_{2i} < l_{2i}) - P(T_{1i} < r_{1i}, T_{2i} < r_{2i}) + P(T_{1i} < r_{1i}, T_{2i} < l_{2i}) =$$
$$= F_2(r_{2i}) - F_2(l_{2i}) - F(r_{1i},r_{2i}) + F(r_{1i},l_{2i}) =$$
$$= 1 - S_2(r_{2i}) - 1 + S_2(l_{2i}) - [1 - S_1(r_{1i}) - S_2(r_{2i}) + S(r_{1i},r_{2i})] +$$
$$+ [1 - S_1(r_{1i}) - S_2(l_{2i}) + S(r_{1i},l_{2i})] =$$
$$= S(r_{1i},l_{2i}) - S(r_{1i},r_{2i}) =$$
$$= C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}.$$

- T_{1i} left-censored and T_{2i} interval-censored (the swapped case can be trivially derived by switching the subscripts where required):

$$\begin{aligned}
&P(T_{1i} < l_{1i}, l_{2i} < T_{2i} < r_{2i}) = F(l_{1i}, r_{2i}) - F(l_{1i}, l_{2i}) = \\
&= [1 - S_1(l_{1i}) - S_2(r_{2i}) + S(l_{1i}, r_{2i})] - [1 - S_1(l_{1i}) - S_2(l_{2i}) + S(l_{1i}, l_{2i})] = \\
&= S_2(l_{2i}) - S_2(r_{2i}) + S(l_{1i}, r_{2i}) - S(l_{1i}, l_{2i}) = \\
&= G_2(\eta_{2i}(l_{2i})) - G_2(\eta_{2i}(r_{2i}))\} \\
&+ C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}.
\end{aligned}$$

Notation

In order to provide more concise and readable expressions of the derivatives of the log-likelihood, the following quantities have been defined.

$$
\begin{aligned}
D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}} &= \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))} \\
D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}} &= G_1'(\eta_{1i}(t_{1i})) \\
D_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}} &= G_2'(\eta_{2i}(t_{2i})) \\
D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}} &= \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \\
D_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}} &= \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}} \\
D_{\gamma_{R_{1i}}\gamma_{R_{2i}}} &= C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\} \\
D_{\gamma_{L_{1i}}\gamma_{L_{2i}}} &= \Big[1-G_1(\eta_{1i}(l_{1i}))-G_2(\eta_{2i}(l_{2i}))+C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}\Big] \\
D_{\gamma_{I_{1i}}\gamma_{I_{2i}}} &= \Big[C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}-C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}+ \\
&\qquad -C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}+C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}\Big] \\
D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}} &= \frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} \\
D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}} &= -G_1'(\eta_{1i}(t_{1i})) \\
D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}} &= \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \\
D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}} &= \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))} \\
D_{2;\gamma_{R_{1i}}\gamma_{u_{2i}}} &= -G_2'(\eta_{2i}(t_{2i})) \\
D_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}} &= \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}} \\
D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}} &= \left(\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}-1\right) \\
D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}} &= G_1'(\eta_{1i}(t_{1i})) \\
D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}} &= \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \\
D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}} &= \left(\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))}-1\right)
\end{aligned}
$$

$$\begin{aligned}
D_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}} &= G_2'(\eta_{2i}(t_{2i}))\\
D_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}} &= \frac{\partial\eta_{2i}(t_{2i})}{\partial t_{2i}}\\
D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}} &= \left(\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} - \frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\right)\\
D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}} &= G_1'(\eta_{1i}(t_{1i}))\\
D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}} &= \frac{\partial\eta_{1i}(t_{1i})}{\partial t_{1i}}\\
D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}} &= \left(\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))} - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))}\right)\\
D_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}} &= G_2'(\eta_{2i}(t_{2i}))\\
D_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}} &= \frac{\partial\eta_{2i}(t_{2i})}{\partial t_{2i}}\\
D_{\gamma_{R_{1i}}\gamma_{L_{2i}}} &= G_1(\eta_{1i}(r_{1i})) - C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}\\
D_{\gamma_{L_{1i}}\gamma_{R_{2i}}} &= G_2(\eta_{2i}(r_{2i})) - C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}\\
D_{\gamma_{R_{1i}}\gamma_{I_{2i}}} &= C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\} - C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}\\
D_{\gamma_{I_{1i}}\gamma_{R_{2i}}} &= C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\} - C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}\\
D_{\gamma_{L_{1i}}\gamma_{I_{2i}}} &= \Big[G_2(\eta_{2i}(l_{2i})) - G_2(\eta_{2i}(r_{2i})) + C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}+\\
&\qquad - C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}\Big]\\
D_{\gamma_{I_{1i}}\gamma_{L_{2i}}} &= \Big[G_1(\eta_{1i}(l_{1i})) - G_1(\eta_{1i}(r_{1i})) + C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}+\\
&\qquad - C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}\Big]
\end{aligned}$$

First derivatives with respect to β_1

$$\begin{aligned}
\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1} &= \left[\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right]\\
\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1} &= \left[G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right]\\
\frac{\partial D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1} &= \left[\frac{\partial^2\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1}\right]
\end{aligned}$$

$$\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial \beta_1} = \left\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \beta_1} \right\}$$

$$\begin{aligned} \frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial \beta_1} = &\left\{ -G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial \eta_{1i}(l_{1i})}{\partial \beta_1} + \right. \\ &\left. + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial \eta_{1i}(l_{1i})}{\partial \beta_1} \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1} = &\left\{ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial \eta_{1i}(l_{1i})}{\partial \beta_1} + \right. \\ &- \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial \eta_{1i}(l_{1i})}{\partial \beta_1} + \\ &- \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \beta_1} + \\ &\left. + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \beta_1} \right\} \end{aligned}$$

$$\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial \beta_1} = \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2} \cdot G_1'(\eta_{1i}(t_{1i}) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \beta_1}$$

$$\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial \beta_1} = \left[-G_1''(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \beta_1} \right]$$

$$\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial \beta_1} = \frac{\partial^2 \eta_{1i}(t_{1i})}{\partial t_{1i} \partial \beta_1}$$

$$\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial \beta_1} = \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i})) \partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \beta_1}$$

$$\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial \beta_1} = \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \beta_1}$$

$$\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial \beta_1} = G_1''(\eta_{1i}(t_{1i})) \frac{\partial \eta_{1i}(t_{1i})}{\partial \beta_1}$$

$$\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial \beta_1} = \frac{\partial^2 \eta_{1i}(t_{1i})}{\partial t_{1i} \partial \beta_1}$$

$$\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial \beta_1} = \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i})) \partial G_1(\eta_{1i}(l_{1i}))} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial \eta_{1i}(l_{1i})}{\partial \beta_1}$$

$$\begin{aligned} \frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1} = &\left[\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \beta_1} + \right. \\ &\left. - \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \beta_1} \right] \end{aligned}$$

$$\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1} = \left[G_1''(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \beta_1} \right]$$

$$\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1} = \left[\frac{\partial^2 \eta_{1i}(t_{1i})}{\partial t_{1i} \partial \beta_1} \right]$$

$$\begin{aligned} \frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial \beta_1} = &\left\{ \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i})) \partial G_2(\eta_{2i}(t_{2i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \beta_1} + \right. \\ &\left. - \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}) \partial G_2(\eta_{2i}(t_{2i}))} \cdot G_1'(\eta_{1i}(l_{1i}) \cdot \frac{\partial \eta_{1i}(l_{1i})}{\partial \beta_1} \right\} \end{aligned}$$

$$
\begin{aligned}
\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1} &= \Bigg\{G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\Bigg\}\\
\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1} &= \Bigg\{-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\Bigg\}\\
\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1} &= \Bigg\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\Bigg\}\\
\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1} &= \Bigg\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\Bigg\}\\
\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1} &= \Bigg\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\Bigg\}\\
\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1} &= \Bigg\{G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}-G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\qquad +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\Bigg\}
\end{aligned}
$$

First derivatives with respect to β_2

$$
\begin{aligned}
\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} &= \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\\
\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} &= G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\\
\frac{\partial D_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} &= \frac{\partial^2\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2}\\
\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2} &= \Bigg\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\Bigg\}\\
\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2} &= \Bigg\{-G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}+\\
&\qquad +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\Bigg\}
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2} = & \left\{ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right.\\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_1'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}+\\
& \left.+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_1'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right\}
\end{aligned}$$

$$\frac{\partial D_{\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2} = \left\{\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right\}$$

$$\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} = \left[\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right]$$

$$\frac{\partial D_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} = \left[-G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right]$$

$$\frac{\partial D_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} = \frac{\partial^2\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2}$$

$$\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2} = \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}$$

$$\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} = \left[\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right]$$

$$\frac{\partial D_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} = \left[G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right]$$

$$\frac{\partial D_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1} = \left[\frac{\partial^2\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2}\right]$$

$$\begin{aligned}
\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2} = & \left\{\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\right.\\
& \left.-\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} = & \left[\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}+\right.\\
& \left.-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right]+
\end{aligned}$$

$$\frac{\partial D_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} = G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}$$

$$\frac{\partial D_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2} = \frac{\partial^2\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2}$$

$$\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2} = \left\{-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right\}$$

$$\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2} = \left\{G_2'(\eta_{2i}(r_{2i}))\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right\}$$

$$
\begin{aligned}
\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial \beta_2} &= \left\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})} \cdot G_2'(\eta_{2i}(l_{2i}) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2} + \right. \\
&\qquad \left. - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2} \right\} \\
\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial \beta_2} &= \left\{ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2} + \right. \\
&\qquad \left. - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2} \right\} \\
\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial \beta_2} &= \left\{ G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2} - G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2} + \right. \\
&\qquad + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2} + \\
&\qquad \left. - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2} \right\} \\
\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial \beta_2} &= \left\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2} + \right. \\
&\qquad \left. - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i}) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2} \right\}
\end{aligned}
$$

First derivatives with respect to β_3

$$
\begin{aligned}
\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial \beta_3} &= \left\{ \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(x_{3i},\beta_3)}{\partial \beta_3} \right\} \\
\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial \beta_3} &= \left\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(x_{3i},\beta_3)}{\partial \beta_3} \right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial \beta_3} &= \left\{ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(x_{3i},\beta_3)}{\partial \beta_3} \right\} + \\
\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial \beta_3} &= \left\{ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(x_{3i},\beta_3)}{\partial \beta_3} + \right. \\
&\qquad - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(x_{3i},\beta_3)}{\partial \beta_3} + \\
&\qquad - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(x_{3i},\beta_3)}{\partial \beta_3} + \\
&\qquad \left. + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(x_{3i},\beta_3)}{\partial \beta_3} \right\}
\end{aligned}
$$

$$\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3} = \left\{ \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \right\}$$

$$\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3} = \left\{ \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \right\}$$

$$\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3} = \left\{ \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \right\}$$

$$\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3} = \left\{ \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \right\}$$

$$\begin{aligned}\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} + \\ & - \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \Bigg\}\end{aligned}$$

$$\begin{aligned}\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} + \\ & - \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \Bigg\}\end{aligned}$$

$$\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3} = \left\{ - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \right\}$$

$$\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3} = \left\{ - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \right\}$$

$$\begin{aligned}\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3} = \Bigg\{ & \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} + \\ & - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \Bigg\}\end{aligned}$$

$$\begin{aligned}\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3} = \Bigg\{ & \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} + \\ & - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \Bigg\}\end{aligned}$$

$$\begin{aligned}\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3} = \Bigg\{ & \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} + \\ & - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \Bigg\}\end{aligned}$$

$$\begin{aligned}\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3} = \Bigg\{ & \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} + \\ & - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \Bigg\}\end{aligned}$$

Second derivatives with respect to β_1

$$
\begin{aligned}
\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_1{}^T} = \Bigg\{ & \frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^3\partial G_2(\eta_{2i}(t_{2i}))}\left(G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+ \\
& + \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1''(\eta_{1i}(t_{1i}))\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+ \\
& + \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial\beta_1{}^T}
\end{aligned}
$$

$$
\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_1{}^T} = \left\{G_1'''(\eta_{1i}(t_{1i}))\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2 + G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial\beta_1{}^T}\right\}
$$

$$
\frac{\partial^2 D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_1{}^T} = \frac{\partial^3\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1\partial\beta_1{}^T}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_1{}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2}\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial\beta_1{}^T}\Bigg\}+
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_1{}^T} = \Bigg\{ & -G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2 - G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial\beta_1{}^T}+ \\
& + \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial\beta_1{}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_1{}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2 \\
& + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial\beta_1{}^T}+ \\
& - \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+ \\
& - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+ \\
& - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial\beta_1{}^T}+ \\
& - \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T}+\\
&+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+\\
&\left.+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_1}^T}&=\left\{+\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^3}\cdot\left(G_1'(\eta_{1i}(t_{1i})\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+\right.\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1''(\eta_{1i}(t_{1i})\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+\\
&\left.+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1'(\eta_{1i}(t_{1i})\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial{\beta_1}^T}\right\}\\
\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_1}^T}&=\left\{-G_1'''(\eta_{1i}(t_{1i}))\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2-G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial\beta_1}\right\}\\
\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_1}^T}&=\frac{\partial^3\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1\partial{\beta_1}^T}\\
\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_1}^T}&=\left\{\frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(r_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+\right.\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2\\
&\left.+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial\beta_1}\right\}\\
\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_1}^T}&=\left\{+\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^3}\cdot\left(G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2\right.\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1''(\eta_{1i}(t_{1i}))\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+\\
&\left.+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial{\beta_1}^T}\right\}\\
\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_1}^T}&=\left\{G_1'''(\eta_{1i}(t_{1i}))\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+G_1''(\eta_{1i}(t_{1i}))\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial{\beta_1}^T}\right\}\\
\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_1}^T}&=\frac{\partial^3\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1\partial{\beta_1}^T}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ &+ \frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}\\
\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ &\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^3}\cdot\left(G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1''(\eta_{1i}(t_{1i}))\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial{\beta_1}^T}+\\
&-\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^3}\cdot\left(G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1''(\eta_{1i}(t_{1i}))\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}\\
\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ &G_1'''(\eta_{1i}(t_{1i}))\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right)^2+G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}\\
\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = {} &\frac{\partial^3\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1\partial{\beta_1}^T}\\
\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ &\frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2\partial G_2(\eta_{2i}(t_{2i}))}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T}+\\
&-\frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i})^2\partial G_2(\eta_{2i}(t_{2i}))}\cdot\left(G_1'(\eta_{1i}(l_{1i})\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i})\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1''(\eta_{1i}(l_{1i})\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i})\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1'(\eta_{1i}(l_{1i})\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ & G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 + G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T} + \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 + \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ & -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2 + \\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2 + \\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 + \\
& +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 + \\
& +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T} + \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2 + \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2 + \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial{\beta_1}^T} + \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 + \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2 + \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial{\beta_1}^T}+\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}\\
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_1}^T} = \Bigg\{ & G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial{\beta_1}^T}+\\
&-G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2-G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T}+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1''(\eta_{1i}(r_{1i}))\cdot\left(\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial^2\eta_{1i}(r_{1i})}{\partial\beta_1\partial{\beta_1}^T}+\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))^2}\cdot\left(G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\beta_1\partial{\beta_1}^T}\Bigg\}
\end{aligned}
$$

Second derivatives with respect to β_1 and β_2

$$
\begin{aligned}
\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i})\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}
\end{aligned}
$$

$$\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} = \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}$$

$$\begin{aligned}\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\ &-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\ &-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\ &+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\end{aligned}$$

$$\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(r_{2i})}\cdot G_2'(\eta_{2i}(r_{2i})\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(t_{1i})\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}$$

$$\begin{aligned}\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}+\\ &-\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_1}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\ &-\frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i})\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i})\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\end{aligned}$$

$$\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} = -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} = -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}$$

$$\begin{aligned}\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\ &-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\ &-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\
&\quad-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_2}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\quad-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial G_2(\eta_{2i}(l_{2i})}\cdot G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}
\end{aligned}$$

Second derivatives with respect to β_1 and β_3

$$\begin{aligned}
\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_3}^T} &= \frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_3}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_3}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial{\beta_3}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\
&\quad-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\
&\quad-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\quad+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial{\beta_3}^T} &= \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(t_{1i})\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_3}^T} &= \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial{\beta_3}^T} &= \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\\
\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial{\beta_3}^T} &= \frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}
\end{aligned}$$

$$\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3{}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1} +$$
$$- \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_3{}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1} +$$
$$- \frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i})\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(l_{1i}) \cdot \frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_3{}^T} = -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_3{}^T} = -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3{}^T} = \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1} +$$
$$- \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_3{}^T} = \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1} +$$
$$- \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(r_{1i}))\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3{}^T} = \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1} +$$
$$- \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}$$

$$\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_3{}^T} = \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1} +$$
$$- \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}$$

Second derivatives with respect to β_2 and β_3

$$\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3{}^T} = \frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^2\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}$$

$$\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_3{}^T} = \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}$$

$$\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} = \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}$$

$$\begin{aligned}\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}+\\ &-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\\ &-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}+\\ &+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\end{aligned}$$

$$\frac{\partial^2 D_{\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}$$

$$\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}$$

$$\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i}})\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}$$

$$\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} = \frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}$$

$$\begin{aligned}\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} &= \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\\ &-\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i})} m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} &= \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}+\\ &-\frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\end{aligned}$$

$$\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} = -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}$$

$$\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} = -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}$$

$$\begin{aligned}\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}+\\ &-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\\ &-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\\
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial{\beta_3}^T} &= \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}+\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\cdot G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}
\end{aligned}$$

Second derivatives with respect to β_3

$$\begin{aligned}
\frac{\partial^2 D_{\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} &= \Bigg\{\frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&\quad+\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&\quad+\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}\\
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} &= \Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&\quad+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&\quad+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}\\
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} &= \Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&\quad+\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&\quad+\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}\\
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} &= \Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&\quad+\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&\quad+\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+
\end{aligned}$$

$$
\begin{aligned}
&-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3\partial{\beta_3}^T}=\Bigg\{&\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3\partial{\beta_3}^T}=\Bigg\{&\frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3\partial{\beta_3}^T}=\Bigg\{&\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3\partial{\beta_3}^T}=\Bigg\{&\frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2\\
&+\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} = \Bigg\{ & \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\left(\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+\\
& -\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} = \Bigg\{ & \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+\\
& -\frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} = \Bigg\{ & -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} = \Bigg\{ & -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+\\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+ \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+ \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3\partial{\beta_3}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+ \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}
\end{aligned}
$$

$$\begin{aligned}\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial^2\beta_3\partial{\beta_3}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\ & +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\ & +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}+ \\ & -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\ & -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3}\right)^2+ \\ & -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(x_{3i},\beta_3)}{\partial\beta_3\partial{\beta_3}^T}\Bigg\}\end{aligned}$$

Second derivatives with respect to β_2

$$\begin{aligned}\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \Bigg\{ & \frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^3}\cdot\left(G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right)^2+ \\ & +\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2''(\eta_{2i}(t_{2i}))\cdot\left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right)^2+ \\ & +\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^T}\Bigg\} \\ \frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = & \, G_2'''(\eta_{2i}(t_{2i}))\cdot\left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right)^2+G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^T} \\ \frac{\partial^2 D_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = & \, \frac{\partial^3\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2\partial{\beta_2}^T} \\ \frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\ & +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\ & +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}\Bigg\}\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \biggl\{ & -G_2''(\eta_{2i}(l_{2i}))\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2 - G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}+\\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2''(\eta_{2i}(l_{2i}))\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}\biggr\}\\
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \biggl\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2''(\eta_{2i}(l_{2i}))\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}+\\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+\\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+\\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}+\\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2''(\eta_{2i}(l_{2i}))\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}+\\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+\\
& +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+\\
& +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}\biggr\}\\
\frac{\partial^2 D_{\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \biggl\{ & \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+\\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+\\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}\biggr\}
\end{aligned}
$$

$$\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \left\{ \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^3} \cdot \left(G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2} \right)^2 + \right.$$

$$+ \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2''(\eta_{2i}(t_{2i})) \cdot \left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2} \right)^2 +$$

$$\left. + \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^T} \right\}$$

$$\frac{\partial^2 D_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = -G_2'''(\eta_{2i}(t_{2i})) \cdot \left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2} \right)^2 - G_2''(\eta_{2i}(t_{2i})) \cdot \frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^T}$$

$$\frac{\partial^2 D_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \frac{\partial^3\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2\partial{\beta_2}^T}$$

$$\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \left\{ \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))^2} \cdot \left(G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2} \right)^2 + \right.$$

$$+ \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2''(\eta_{2i}(l_{2i})) \cdot \left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2} \right)^2 +$$

$$\left. + \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T} \right\}$$

$$\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \left\{ \frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^3} \cdot \left(G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2} \right)^2 + \right.$$

$$+ \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2''(\eta_{2i}(t_{2i})) \cdot \left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2} \right)^2 +$$

$$\left. + \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^T} \right\}$$

$$\frac{\partial^2 D_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = G_2'''(\eta_{2i}(t_{2i})) \cdot \left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2} \right)^2 + G_2''(\eta_{2i}(t_{2i})) \cdot \frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^T}$$

$$\frac{\partial^2 D_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \frac{\partial^3\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2\partial{\beta_2}^T}$$

$$\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \left\{ \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))^2} \cdot \left(G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2} \right)^2 + \right.$$

$$+ \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2''(\eta_{2i}(r_{2i})) \cdot \left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2} \right)^2 +$$

$$+ \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T} +$$

$$- \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))^2} \cdot \left(G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2} \right)^2 +$$

$$- \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2''(\eta_{2i}(l_{2i})) \cdot \left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2} \right)^2 +$$

$$\left. - \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T} \right\}$$

$$
\begin{aligned}
\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} &= \Bigg\{ \frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^3}\cdot\left(G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right)^2+\\
&\quad+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2''(\eta_{2i}(t_{2i}))\cdot\left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right)^2+\\
&\quad+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^T}+\\
&\quad-\frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^3}\cdot\left(G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right)^2+\\
&\quad-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2''(\eta_{2i}(t_{2i}))\cdot\left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right)^2+\\
&\quad-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^2}\Bigg\}\\
\frac{\partial^2 D_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} &= G_2'''(\eta_{2i}(t_{2i}))\cdot\left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right)^2+G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial^2\eta_{2i}(t_{2i})}{\partial\beta_2\partial{\beta_2}^T}\\
\frac{\partial^2 D_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} &= \frac{\partial^3\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2\partial{\beta_2}^T}\\
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} &= \Bigg\{-\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
&\quad-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2''(\eta_{2i}(l_{2i}))\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
&\quad-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}\Bigg\}\\
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} &= \Bigg\{G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}+\\
&\quad-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+\\
&\quad-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+\\
&\quad-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})^2}\cdot\left(G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})}\cdot G_2''(\eta_{2i}(l_{2i})\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})}\cdot G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}+ \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}\Bigg\} \\
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}+ \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}\Bigg\} \\
\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \Bigg\{ & G_2''(\eta_{2i}(l_{2i}))\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}+ \\
& -G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2-G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}+ \\
& +\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2''(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right)^2+ \\
& +\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\beta_2\partial{\beta_2}^T}+ \\
& -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot\left(G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2''(\eta_{2i}(l_{2i}))\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+ \\
& -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial{\beta_2}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2''(\eta_{2i}(l_{2i}))\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
&+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^2}+\\
&-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))^2}\cdot\left(G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2''(\eta_{2i}(l_{2i})\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right)^2+\\
&-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\beta_2\partial{\beta_2}^T}\Bigg\}
\end{aligned}
$$

A.2 Gradient

First derivative of log-likelihood with respect to β_1

$$
\begin{aligned}
\frac{\partial \ell(\delta)}{\partial \beta_1} =& \gamma_{U_{1i}}\gamma_{U_{2i}}\left\{D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\left[\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right]+\right.\\
&\qquad\qquad + D^{-1}_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}\left[G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right]+\\
&\qquad\qquad\left. + D^{-1}_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}\left[\frac{\partial^2\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1}\right]\right\}\\
&+\gamma_{R_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\left\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\left\{-G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\right.\\
&\qquad\qquad\left.+\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right\}+\\
&+\gamma_{I_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\left\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\right.\\
&\qquad\qquad-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\
&\qquad\qquad-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\qquad\qquad\left.+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right\}+\\
&+\gamma_{U_{1i}}\gamma_{R_{2i}}\left\{D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial^2 G_1(\eta_{1i}(t_{1i}))}\cdot G_1'(\eta_{1i}(t_{1i})\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}+\right.\\
&\qquad\qquad+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}\left[-G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\right]+\\
&\qquad\qquad\left.+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}\frac{\partial^2\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1}\right\}\\
&+\gamma_{R_{1i}}\gamma_{U_{2i}}D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\left\{\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\right\}+\\
&+\gamma_{U_{1i}}\gamma_{L_{2i}}\left\{D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}+\right.\\
&\qquad\qquad+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}G_1''(\eta_{1i}(t_{1i}))\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}+\\
&\qquad\qquad\left.+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}\frac{\partial^2\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{U_{2i}}\left\{D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\right\}+
\end{aligned}
$$

$$
\begin{aligned}
&+\gamma_{U_{1i}}\gamma_{I_{2i}}\Biggl\{D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\Biggl[\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\Biggr]+\\
&\qquad +D^{-1}_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}\Biggl[G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\beta_1}\Biggr]+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}\Biggl[\frac{\partial^2\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\beta_1}\Biggr]\Biggr\}+\\
&+\gamma_{I_{1i}}\gamma_{U_{2i}}D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\Biggl\{\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i})\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1'(\eta_{1i}(l_{1i})\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\Biggr\}+\\
&+\gamma_{R_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\Biggl\{G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\\
&\qquad \cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\Biggr\}+\\
&+\gamma_{L_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\Biggl\{-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\Biggr\}+\\
&+\gamma_{R_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\Biggl\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\Biggr\}+\\
&+\gamma_{I_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\Biggl\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}\Biggr\}+\\
&+\gamma_{L_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\Biggl\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\Biggr\}\\
&\gamma_{I_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\Biggl\{G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}-G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\qquad +\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\beta_1}+\\
&\qquad -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\beta_1}\Biggr\}
\end{aligned}
$$

First derivative of log-likelihood with respect to β_2

$$
\begin{aligned}
\frac{\partial \ell(\delta)}{\partial \beta_2} = & \gamma_{U_{1i}}\gamma_{U_{2i}}\Bigg\{D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial \eta_{2i}(t_{2i})}{\partial \beta_2}+ \\
& \qquad + D^{-1}_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}} G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial \eta_{2i}(t_{2i})}{\partial \beta_2}+ \\
& \qquad + D^{-1}_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}\frac{\partial^2 \eta_{2i}(t_{2i})}{\partial t_{2i}\partial \beta_2}\Bigg\}+ \\
& + \gamma_{R_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\Bigg\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2}\Bigg\}+ \\
& + \gamma_{L_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\Bigg\{-G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2}+ \\
& \qquad + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2}\Bigg\}+ \\
& + \gamma_{I_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\Bigg\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2}+ \\
& \qquad - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2}+ \\
& \qquad - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2}+ \\
& \qquad + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2}\Bigg\}+ \\
& + \gamma_{U_{1i}}\gamma_{R_{2i}}D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2}\Bigg\}+ \\
& + \gamma_{R_{1i}}\gamma_{U_{2i}}\Bigg\{D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\left[\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial^2 G_2(\eta_{2i}(t_{2i}))}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial \eta_{2i}(t_{2i})}{\partial \beta_2}\right]+ \\
& \qquad + D^{-1}_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}\left[-G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial \eta_{2i}(t_{2i})}{\partial \beta_2}\right]+ \\
& \qquad + D^{-1}_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}\frac{\partial^2 \eta_{2i}(t_{2i})}{\partial t_{2i}\partial \beta_2}\Bigg\}+ \\
& + \gamma_{U_{1i}}\gamma_{L_{2i}}D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2}\Bigg\}+ \\
& + \gamma_{L_{1i}}\gamma_{U_{2i}}\Bigg\{D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\left[\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial \eta_{2i}(t_{2i})}{\partial \beta_2}\right]+ \\
& \qquad + D^{-1}_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}\left[G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial \eta_{2i}(t_{2i})}{\partial \beta_2}\right] + D^{-1}_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}\left[\frac{\partial^2 \eta_{2i}(t_{2i})}{\partial t_{2i}\partial \beta_2}\right]\Bigg\}+ \\
& + \gamma_{U_{1i}}\gamma_{I_{2i}}D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial \eta_{2i}(r_{2i})}{\partial \beta_2}+ \\
& \qquad - \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial \eta_{2i}(l_{2i})}{\partial \beta_2}\Bigg\}+
\end{aligned}
$$

$$
\begin{aligned}
&+\gamma_{I_{1i}}\gamma_{U_{2i}}\left\{D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\left[\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}+\right.\right.\\
&\qquad\left.-\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2}\cdot G_2'(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\right]+\\
&\qquad+D^{-1}_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}G_2''(\eta_{2i}(t_{2i}))\cdot\frac{\partial\eta_{2i}(t_{2i})}{\partial\beta_2}\\
&\qquad\left.+D^{-1}_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}\frac{\partial^2\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\beta_2}\right\}+\\
&+\gamma_{R_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\left\{-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\left\{G_2'(\eta_{2i}(r_{2i}))\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\right.\\
&\qquad\left.\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right\}+\\
&+\gamma_{R_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\left\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})}\cdot G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}+\right.\\
&\qquad\left.-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right\}+\\
&\gamma_{I_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\left\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\right.\\
&\qquad\left.-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\left\{G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}-G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\right.\\
&\qquad+\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G_2'(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\beta_2}+\\
&\qquad\left.-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right\}+\\
&\gamma_{I_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\left\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))}\cdot G_2'(\eta_{2i}(l_{2i}))\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}+\right.\\
&\qquad\left.-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})}\cdot G_2'(\eta_{2i}(l_{2i})\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\beta_2}\right\}+
\end{aligned}
$$

First derivative of log-likelihood with respect to β_3

$$
\begin{aligned}
\frac{\partial\ell(\delta)}{\partial\beta_3} = & \gamma_{U_{1i}}\gamma_{U_{2i}}D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\Bigg\{\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{R_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\Bigg\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+
\end{aligned}
$$

$$
\begin{aligned}
& +\gamma_{L_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\Bigg\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{I_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\Bigg\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\\
& \qquad -\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\\
& \qquad -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\\
& \qquad -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{U_{1i}}\gamma_{R_{2i}}D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{R_{1i}}\gamma_{U_{2i}}D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{U_{1i}}\gamma_{L_{2i}}D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{L_{1i}}\gamma_{U_{2i}}D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{U_{1i}}\gamma_{I_{2i}}D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\\
& \qquad -\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{I_{1i}}\gamma_{U_{2i}}D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\Bigg\{\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\\
& \qquad -\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{R_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\Bigg\{-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{L_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\Bigg\{-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+\\
& +\gamma_{R_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\Bigg\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\\
& \qquad -\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\Bigg\}+
\end{aligned}
$$

$$
\begin{aligned}
&+\gamma_{I_{1i}}\gamma_{R_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\left\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\right.\\
&\qquad\left.-\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{I_{2i}}D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\left\{\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\right.\\
&\qquad\left.-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\right\}+\\
&+\gamma_{I_{1i}}\gamma_{L_{2i}}D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\left\{\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}+\right.\\
&\qquad\left.-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\mathbf{x}_{3i},\beta_3)}{\partial\beta_3}\right\}
\end{aligned}
$$

A.3 Hessian

Second derivative of log-likelihood with respect to β_1

$$
\begin{aligned}
\frac{\partial^2\ell(\delta)}{\partial\beta_1\partial\beta_1^T} =&\gamma_{U_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_1^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_1^T}+\\
&(-1)D^{-2}_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_1^T}\Bigg\}+\\
&+\gamma_{R_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_1^T}\Bigg\}+\\
&+\gamma_{L_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_1^T}\Bigg\}+\\
&+\gamma_{I_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_1^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_1^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_1^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_1^T}\Bigg\}+\\
&+\gamma_{R_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_1^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_1^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_1^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_1^T}\Bigg\}+\\
&+\gamma_{L_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_1^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_1^T}+
\end{aligned}
$$

$$
\begin{aligned}
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}} \cdot \left(\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}} \cdot \frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1 \partial \beta_1^T} + \\
&\left.(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}} \cdot \left(\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}} \cdot \frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1 \partial \beta_1^T}\right\} + \\
+ \gamma_{I_{1i}}\gamma_{U_{2i}} &\left\{(-1)D^{-2}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial \beta_1 \partial \beta_1^T}\right\} + \\
+ \gamma_{R_{1i}}\gamma_{L_{2i}} &\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{L_{2i}}} \cdot \left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}} \cdot \frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial \beta_1 \partial \beta_1^T}\right\} + \\
+ \gamma_{L_{1i}}\gamma_{R_{2i}} &\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{R_{2i}}} \cdot \left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}} \cdot \frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial \beta_1 \partial \beta_1^T}\right\} + \\
+ \gamma_{R_{1i}}\gamma_{I_{2i}} &\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{I_{2i}}} \cdot \left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}} \cdot \frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1 \partial \beta_1^T}\right\} + \\
+ \gamma_{I_{1i}}\gamma_{R_{2i}} &\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{R_{2i}}} \cdot \left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}} \cdot \frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial \beta_1 \partial \beta_1^T}\right\} + \\
+ \gamma_{L_{1i}}\gamma_{I_{2i}} &\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{I_{2i}}} \cdot \left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}} \cdot \frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial \beta_1 \partial \beta_1^T}\right\} + \\
+ \gamma_{I_{1i}}\gamma_{L_{2i}} &\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{L_{2i}}} \cdot \left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial \beta_1}\right)^2 + D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}} \cdot \frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial \beta_1 \partial \beta_1^T}\right\}
\end{aligned}
$$

Second derivative of log-likelihood with respect to β_1 and β_2

$$
\begin{aligned}
\frac{\partial^2 \ell(\delta)}{\partial\beta_1\partial\beta_2^T} =& \gamma_{U_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_2^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_2^T}+\\
&(-1)D^{-2}_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{R_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{L_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{I_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{U_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_2^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_2^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{R_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{1;\delta R_{1i}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\delta R_{1i}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{U_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_2^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_2^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{L_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{U_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_2^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_2^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{I_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+\\
&\gamma_{R_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_2^T}\Bigg\}+
\end{aligned}
$$

$$
\begin{aligned}
&\gamma_{L_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_2^T}\right\}+\\
&\gamma_{R_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_2^T}\right\}+\\
&\gamma_{I_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_2^T}\right\}+\\
&\gamma_{L_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_2^T}\right\}+\\
&\gamma_{I_{1i}}\gamma_{L_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_2^T}\right\}
\end{aligned}
$$

Second derivative of log-likelihood with respect to β_1 and β_3

$$
\begin{aligned}
\frac{\partial^2 \ell(\delta)}{\partial\beta_1\partial\beta_3^T} =& \gamma_{U_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_3^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_3^T}+\\
&(-1)D^{-2}_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{4;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{R_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{L_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{I_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{U_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_3^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_3^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{4;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{4;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{R_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\delta R_{1i}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\delta R_{1i}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{U_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_3^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_3^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{L_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{U_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3^T}+\\
&(-1)D^{-2}_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{I_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1}+D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+\\
&\gamma_{R_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_3^T}\Bigg\}+
\end{aligned}
$$

$$\gamma_{L_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_3^T}\right\}+$$

$$\gamma_{R_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3^T}\right\}+$$

$$\gamma_{I_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_1\partial\beta_3^T}\right\}+$$

$$\gamma_{L_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_1\partial\beta_3^T}\right\}+$$

$$\gamma_{I_{1i}}\gamma_{L_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1}+D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_1\partial\beta_3^T}\right\}$$

Second derivative of log-likelihood with respect to β_2

$$
\begin{aligned}
\frac{\partial^2\ell(\delta)}{\partial\beta_2\partial\beta_2^T} =& \gamma_{U_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}+\\
&(-1)D^{-2}_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{R_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{L_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{I_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{R_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}+\\
&(-1)D^{-2}_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}+\\
&(-1)D^{-2}_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{L_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}+\\
&(-1)D^{-2}_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}+\\
&(-1)D^{-2}_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{I_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}+\\
&(-1)D^{-2}_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}+\\
&(-1)D^{-2}_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{R_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+
\end{aligned}
$$

$$+\gamma_{L_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_2^T}\right\}+$$

$$+\gamma_{R_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial\beta_2^T}\right\}+$$

$$+\gamma_{I_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_2^T}\right\}+$$

$$+\gamma_{L_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial\beta_2^T}\right\}+$$

$$+\gamma_{I_{1i}}\gamma_{L_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}\right)^2+D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial\beta_2^T}\right\}$$

Second derivative of log-likelihood with respect to β_2 and β_3

$$
\begin{aligned}
\frac{\partial^2\ell(\delta)}{\partial\beta_2\partial\beta_3^T} =&\gamma_{U_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}+\\
&(-1)D^{-2}_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{2;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}+\\
&(-1)D^{-2}_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{5;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{R_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{L_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{I_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{R_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}+\\
&(-1)D^{-2}_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}+\\
&(-1)D^{-2}_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{\gamma_{3;R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_2^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{1;\partial D_{\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{L_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}+\\
&(-1)D^{-2}_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}+\\
&(-1)D^{-2}_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{U_{1i}}\gamma_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{I_{1i}}\gamma_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}+\\
&(-1)D^{-2}_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}+\\
&(-1)D^{-2}_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial D_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2}+D^{-1}_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{3;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+\\
&+\gamma_{R_{1i}}\gamma_{L_{2i}}\Bigg\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial\beta_3^T}\Bigg\}+
\end{aligned}
$$

$$
\begin{aligned}
&+\gamma_{L_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_3^T}\right\}+\\
&+\gamma_{R_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial\beta_3^T}\right\}+\\
&+\gamma_{I_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_2\partial\beta_3^T}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_2\partial\beta_3^T}\right\}+\\
&+\gamma_{I_{1i}}\gamma_{L_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\cdot\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2}+D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_2\partial\beta_3^T}\right\}
\end{aligned}
$$

Second derivative of log-likelihood with respect to β_3

$$
\begin{aligned}
\frac{\partial^2 \ell(\delta)}{\partial\beta_3\partial\beta_3^T} =&\gamma_{U_{1i}}\gamma_{U_{2i}}\left\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{R_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{R_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{L_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{L_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{I_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{I_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{U_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{R_{1i}}\gamma_{U_{2i}}\left\{(-1)D^{-2}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{R_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{U_{1i}}\gamma_{L_{2i}}\left\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{U_{2i}}\left\{(-1)D^{-2}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{L_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{U_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{U_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{I_{1i}}\gamma_{U_{2i}}\left\{(-1)D^{-2}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\left(\frac{\partial D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\gamma_{I_{1i}}\gamma_{U_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{R_{1i}}\gamma_{L_{2i}}\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{R_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{L_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{R_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{R_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{R_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{I_{1i}}\gamma_{R_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{I_{1i}}\gamma_{R_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{R_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{L_{1i}}\gamma_{I_{2i}}\left\{(-1)D^{-2}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{L_{1i}}\gamma_{I_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{L_{1i}}\gamma_{I_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+\\
&+\gamma_{I_{1i}}\gamma_{L_{2i}}\left\{(-1)D^{-2}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\left(\frac{\partial D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3}\right)^2+D^{-1}_{\gamma_{I_{1i}}\gamma_{L_{2i}}}\cdot\frac{\partial^2 D_{\gamma_{I_{1i}}\gamma_{L_{2i}}}}{\partial\beta_3\partial\beta_3^T}\right\}+
\end{aligned}
$$

A.4 Simulation study

This section provides evidence on the empirical effectiveness of the proposed approach in recovering true covariate effects and baseline functions.

Survival time T_{1i} was generated from a proportional hazards (PH) model, in particular it was obtained as the solution of $\log(-\log(U_{1i})) = \log[-\log S_{10}(t_{1i})] + \beta_{11}x_{1i} + s_{11}(x_{2i})$, where U_{1i} is uniform in [0,1] and $S_{10}(t_{1i}) = 0.9\exp(-0.4t_{1i}^{2.5}) + 0.1\exp(-0.1t_{1i})$. Time T_{2i} was generated from a proportional odds (PO) model, and it was obtained as the solution of $\log\left[\frac{1-U_{2i}}{U_{2i}}\right] = \log\left[\frac{1-S_{20}(t_{2i})}{S_{20}(t_{2i})}\right] + \beta_{21}x_{1i} + \beta_{22}x_{3i}$, where U_{2i} is uniform in [0,1] and $S_{20}(t_{2i}) = S_{10}(t_{2i}) = 0.9\exp(-0.4t_{1i}^{2.5}) + 0.1\exp(-0.1t_{1i})$. Observations were generated using the Brent's univariate root finding algorithm. The two survival times were joined using a Clayton copula where the predictor for the dependence parameter was specified as $\eta_{3i} = \beta_{31}x_{1i} + s_{31}(x_{2i})$. In practice this was achieved using the conditional sampling approach. The specification of η_3 allowed dependence to vary across observations, with Kendall's τ values ranging approximately from 0.10 to 0.90. The smooth functions were $s_{11}(x_i) = \sin(2\pi x_i)$, $s_{31}(x_i) = 3\sin(\pi x_i)$ and the parameters were defined as $\beta_{11} = -1.5$, $\beta_{21} = 1.2$, $\beta_{22} = 1.2$, $\beta_{31} = -1.5$. Correlated covariates were generated using a multivariate standard normal distribution with a correlation parameter $\rho = 0.5$, and then transformed using the distribution function of a standard normal distribution. Covariate x_{1i} was dichotomised by simply rounding it. The random censoring times were generated using the lower and upper bounds from two uniform random variables. Specifically, this was achieved by comparing such bounds with the simulated times. Uncensored observations were obtained from a subset of the interval- and left-censored observations, using a binomial random variable. Table A.1 shows the censoring rates for two scenarios: mild and high censoring. In the former case, the overall percentage of censoring for the two outcomes is 62.86% and 44.98%, and in the latter we have 84.82% and 77.13%.

Sample sizes were set to $1000, 1500, 2000$ while the number of replicates to 1000. The models were fitted using `gjrm()` in `GJRM` with the Clayton copula. The smooth components of the covariates were represented using penalized low rank thin plane splines with second order penalty and 10 basis functions, and the smooths of times using monotonic penalised B-splines with penalty defined in Section 2.1 of the main paper, and 10 bases. For each replicate, curve estimates were constructed using 200 equally spaced fixed values in the (0,8) range for the monotonic functions and (0,1) otherwise.

	Mild	High
II	2.29	9.01
IL	2.81	10.14
IR	1.15	2.09
IU	7.95	6.04
LI	1.60	5.47
LL	2.38	8.18
LR	0.48	0.86
LU	5.70	4.53
RI	6.46	11.98
RL	7.54	13.70
RR	4.15	4.15
RU	20.35	8.67
UI	6.06	4.58
UL	7.62	5.85
UR	2.44	1.12
UU	21.02	3.63

		I	L	R	U
Mild	cens1	14.20	10.16	38.50	37.14
	cens2	16.41	20.35	8.22	55.02
High	cens1	27.28	19.04	38.50	15.18
	cens2	31.04	37.87	8.22	22.87

Table A.1: Proportions of censoring rates by type, for two scenarios: mild and high censoring. These have been obtained by averaging the censoring rates obtained over 1000 simulated datasets.

The main findings of the simulation study are summarised below:

Parametric effects:

Figures A.1 and A.2 show that overall the mean estimates are very close to the respective true values and improve as the sample size increases, and that the variability of the estimates decreases as the sample size grows large. The estimates for β_{31} (the effect of x_{1i} contained in the additive predictor of the copula parameter) are more variable and exhibit some bias as compared to those of the other parameters, although the bias is somewhat negligible. However, the situation improves as more observations are available for model fitting. This result was completely in line with expectations and has also been documented by Romeo et al. (2018) and Marra & Radice (2020). The latter authors investigated this issue and found that the profile log-likelihood of the copula coefficient tends to be less sharp around the optimum which is to be expected.

Smooth effects:

Figures A.3 and A.4 with Tables A.2 and A.3 show that overall the true smooth functions are recovered well by the proposed estimation method and that the results improve in terms of bias and efficiency as the sample size increases. As expected, estimation of $s_{31}(x_{1i})$ is more challenging, for the reasons given earlier on. However, the performance improves dramatically as the sample size grows large. This suggests that complex model specifications should be adopted if the information content in the data is deemed sufficient to estimate reliably such effects.

Impact of censoring rates:

Comparing the plots of Figure A.1 (mild censoring rates) with those of Figure A.2 (high censoring), and the bias and root mean squared error (RMSE) of Table A.2 (mild censoring rates) with those of Table A.3 (high censoring), we see that the presence of high censoring deteriorates the estimation performance. Moreover, the most affected parameters are those belonging to the copula's additive predictor. These results do not come as a surprise given the loss of information caused by a higher level of censoring. As expected, as the sample size increases the estimates improve considerably which is reassuring. Finally, high censoring caused the algorithm to fail to converge for a few simulation replicates which were discarded from the results.

Coverage probabilities for the model terms were also checked. The empirical coverages were overall close to the respective nominal levels; for instance, for a 95% nominal level the coverages were in the range $[0.93, 0.96]$. The exceptions were observed for the difficult scenario of high censoring rate and small sample size where, for the effects related to the dependence parameter, the coverages were in the range $[0.88, 0.93]$ for a nominal level of 95%. This was expected since, as mentioned in the previous paragraphs, estimation of such effects is more difficult. The situation improved significantly for larger sample sizes.

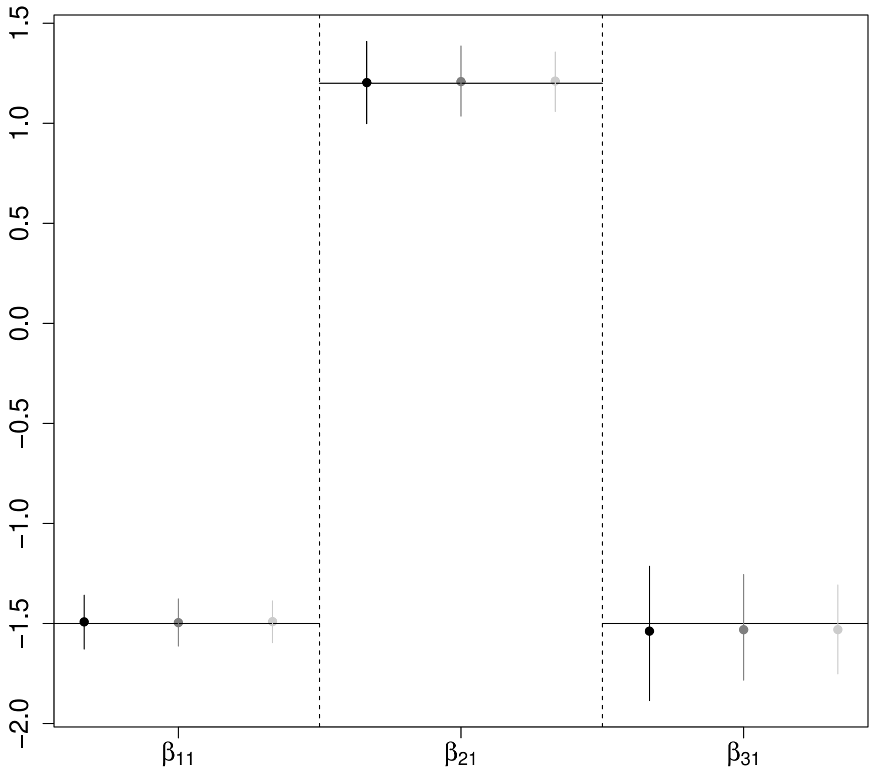

Figure A.1: Linear coefficient estimates obtained by applying `gjrm` to bivariate survival data with mild censoring rates. Circles indicate mean estimates, whereas vertical bars represent estimate ranges (5%-95% quantiles). True values are denoted with black solid lines. Black circles refer to the results obtained for $n = 1000$, whereas those in dark grey and light grey are for $n = 1500$ and $n = 2000$.

Using sample sizes smaller than the ones considered here would clearly affect the estimation performance. In fact, a small number of observations necessarily implies a limit to the amount of modeling complexity allowed by the data (e.g. Marra & Radice, 2020). In such a case, one would have to employ simpler model specifications as the use of splines clearly requires the availability of more information. For example, one could assume linear covariate effects instead of smooth (non-linear) ones. Nevertheless, following a reviewer suggestion, we checked how far we can push the model. When setting, e.g., n at 300, we had to discard around 30% of the replicates (those related to the non-converged models). However, surprisingly, results were still reasonable; see Figures A.5 and A.6.

	Bias			RMSE		
	n = 1000	n = 1500	n = 2000	n = 1000	n = 1500	n = 2000
β_{11}	0.008	0.004	0.009	0.082	0.072	0.063
β_{21}	0.003	0.008	0.011	0.124	0.104	0.088
β_{31}	-0.038	-0.031	-0.031	0.209	0.161	0.139
h_{10}	0.040	0.034	0.028	0.154	0.115	0.110
h_{20}	0.026	0.018	0.015	0.144	0.115	0.104
s_{11}	0.021	0.016	0.014	0.073	0.058	0.050
s_{31}	0.087	0.060	0.045	0.279	0.196	0.146

Table A.2: Bias and root mean squared error (RMSE) obtained by applying `gjrm` to bivariate survival data with mild censoring rates. Bias and RMSE for the smooth terms are calculated using the following expressions: Bias=$n_s^{-1}\sum_{i=1}^{n_s}|\bar{\hat{s}}_i - s_i|$ and RMSE=$n_s^{-1}\sum_{i=1}^{n_s}\sqrt{n_{rep}^{-1}\sum_{rep=1}^{n_{rep}}(\hat{s}_{rep,i} - s_i)^2}$, where $\bar{\hat{s}}_i = n_{rep}^{-1}\sum_{rep=1}^{n_{rep}}\hat{s}_{rep,i}$, n_s is the number of equally spaced fixed values in the (0,8) or (0,1) range, and n_{rep} is the number of simulation replicates. The bias for the smooth terms is based on absolute differences in order to avoid compensating effects when taking the sum.

	Bias			RMSE		
	n = 1000	n = 1500	n = 2000	n = 1000	n = 1500	n = 2000
β_{11}	0.006	0.007	0.010	0.095	0.077	0.066
β_{21}	0.001	0.009	0.006	0.146	0.123	0.099
β_{31}	-0.100	-0.073	-0.055	0.344	0.259	0.204
h_{10}	0.051	0.036	0.030	0.149	0.122	0.105
h_{20}	0.038	0.027	0.019	0.164	0.137	0.124
s_{11}	0.022	0.017	0.017	0.086	0.068	0.059
s_{31}	0.075	0.044	0.036	0.379	0.254	0.190

Table A.3: Bias and root mean squared error (RMSE) obtained by applying `gjrm` to bivariate survival data with high censoring rates. Further details are given in the caption of Table A.2.

A.5 Model fitting using `GJRM`

The proposed modelling framework has been implemented within the `R` package `GJRM` (Marra & Radice, 2024), which required extending the `gjrm()` function. This package has been created to enhance reproducible research and to disseminate results in a straightforward and transparent way. The function is generally very easy to use, especially if the user is already familiar with the syntax of (generalized) linear and additive models in `R`. For instance, one of the calls used for modelling data from the AREDS, available through the `R` package `CopulaCenR`, is

```
eq1 <- t11 ~ s(t11, bs = "mpi") + s(ENROLLAGE) + SevScale1E + rs2284665
eq2 <- t21 ~ s(t21, bs = "mpi") + s(ENROLLAGE) + SevScale2E + rs2284665
```

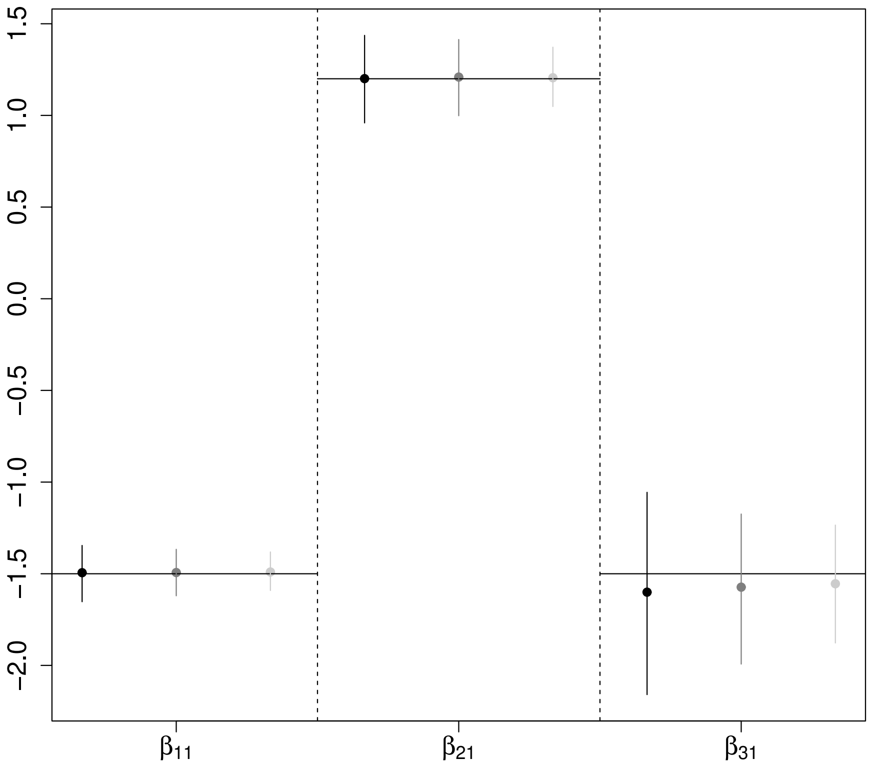

Figure A.2: Linear coefficient estimates obtained by applying `gjrm` to bivariate survival data with high censoring rates. Further details are given in the caption of Figure A.1.

```
eq3 <-     ~                          s(ENROLLAGE)              + rs2284665
f.list <- list(eq1, eq2, eq3)
out <- gjrm(f.list, data = AREDS, surv = TRUE, BivD = "PL",
       margins = c("PO", "PO"), cens1 = cens1, cens2 = cens2,
       Model = "B", upperBt1 = "t12", upperBt2 = "t22")
```

where `t11` and `t12` represent the lower and upper bounds, respectively, of the time interval where the left eye progressed to late-AMD. If `t12 = NA`, then the left eye did not progress to late-AMD by the end of the study and hence the outcome is not observed (right censoring). `cens1` is a factor variable indicating the type of censoring (in this case, either interval or right in accordance with `t12`). Similarly, `t21` and `t22` represent the lower and upper bounds of the time interval for the right eye and `cens2` is the censoring indicator. `AREDS` is a data frame containing the variables, including the three covariates, `ENROLLAGE`, `SevScale1E` and `rs2284665`, already defined in

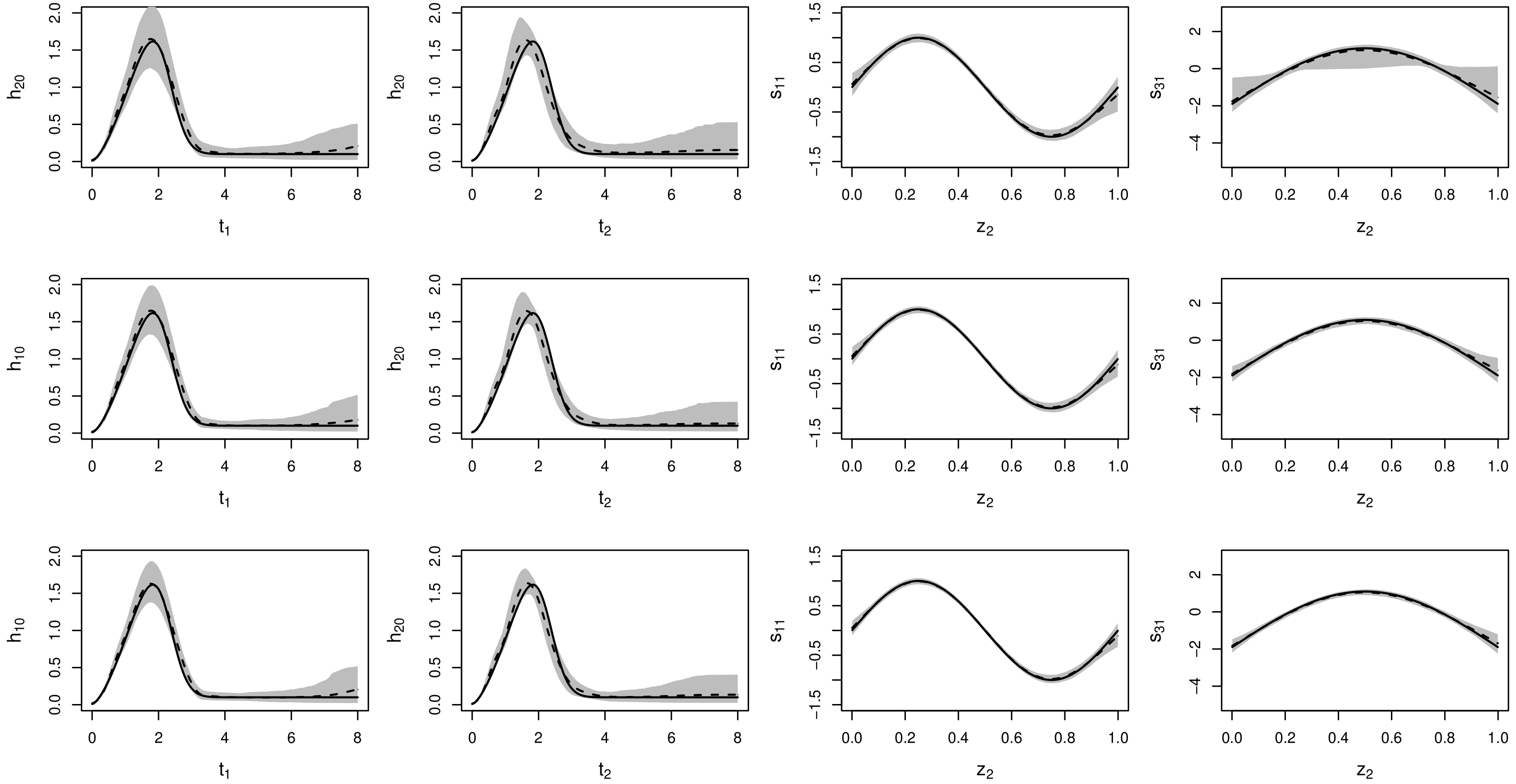

Figure A.3: Smooth function estimates obtained by applying `gjrm` to bivariate survival simulated data with mild censoring. True functions are represented by black solid lines, mean estimates by dashed lines and point-wise ranges resulting from 5% and 95% quantities by shaded areas. The results in the first row refer to $n = 1000$, whereas those in the second and third rows to $n = 1500$ and $n = 2000$, respectively.

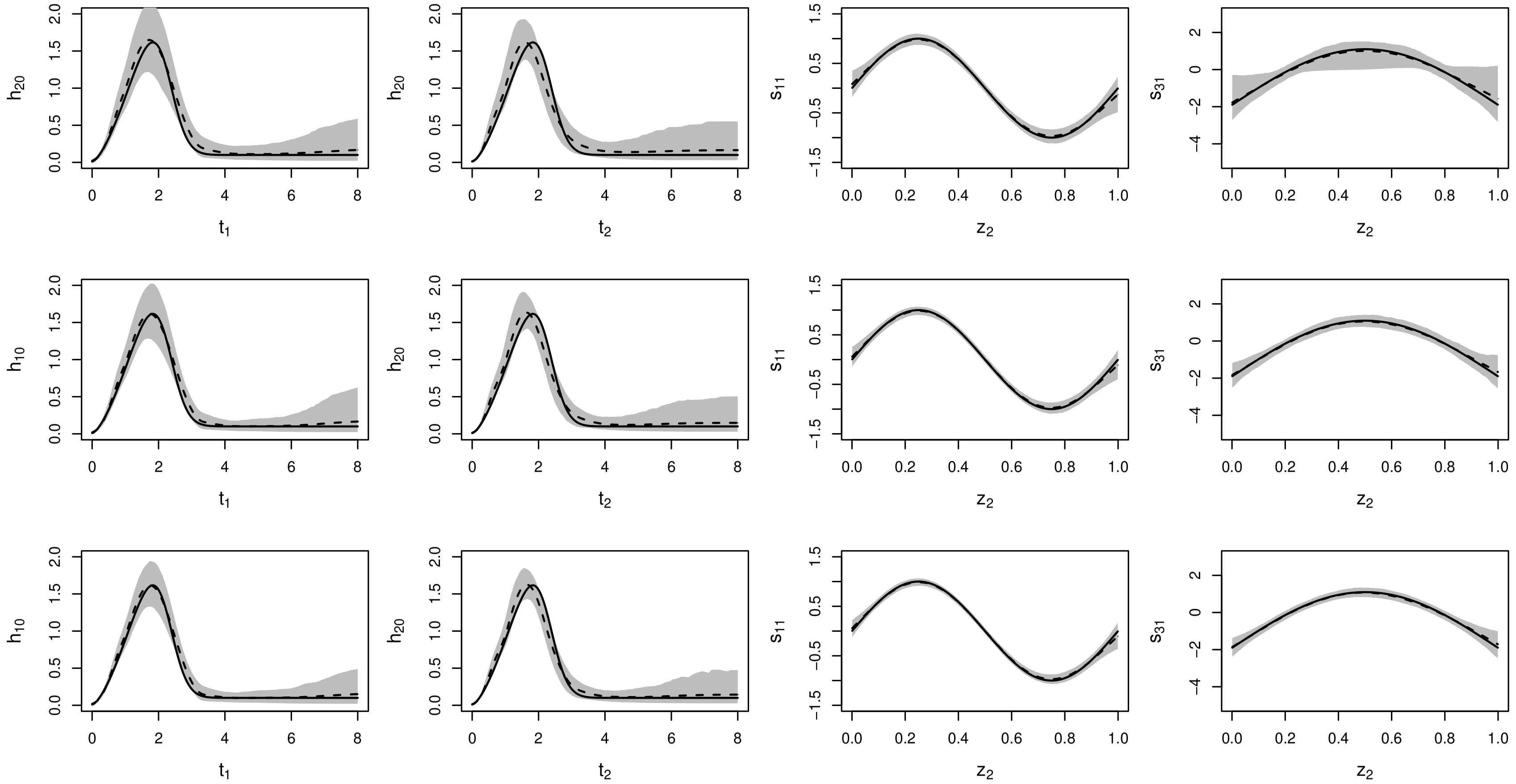

Figure A.4: Smooth function estimates obtained by applying `gjrm` to bivariate survival simulated data with high censoring. Further details are given in the caption of Figure A.3.

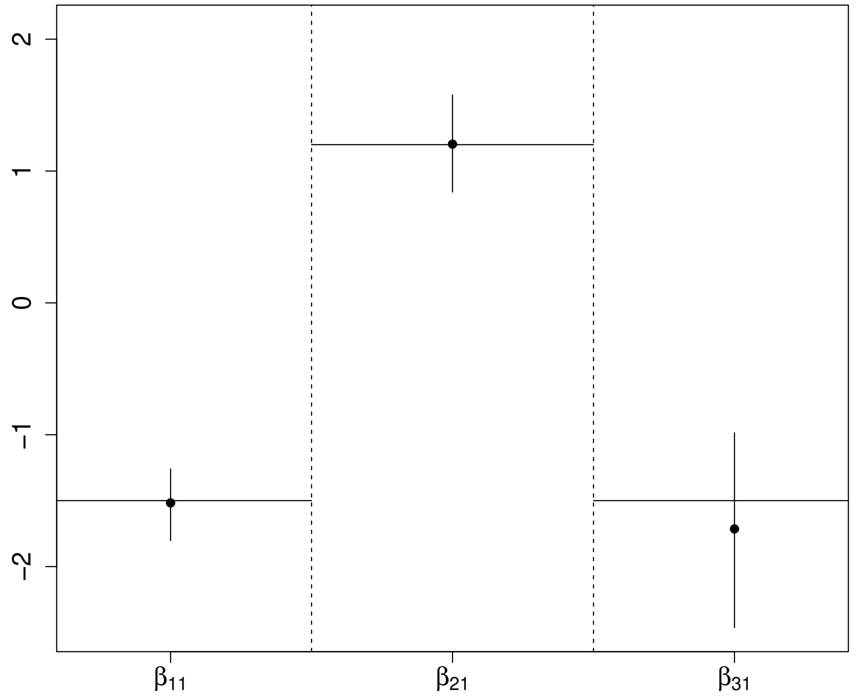

Figure A.5: Linear coefficient estimates obtained by applying `gjrm` to bivariate survival data with mild censoring rates and $n = 300$. Circles indicate mean estimates, whereas vertical bars represent estimate ranges (5%-95% quantiles). True values are denoted with black horizontal solid lines.

Section 4 of the main paper. `Model` must be set to = ”B” and `surv` to `TRUE` in order to employ a joint bivariate survival model. The possible choices for `BivD` and `margins` are given in Section 2 of the paper, `f.list` is a list of equations for the survival outcomes and the copula dependence parameter, `s` denotes the use of a smooth term and argument `bs` specifies the type of spline basis (e.g., `tp` for thin plate regression spline (the default) and `mpi` for monotonic P-spline). Monotonic P-splines must always be used for the smooth terms of the time variables, otherwise the program will produce an error message. After fitting the model, function `conv.check()` can be used to check that convergence has been achieved.

```
conv.check(out)
```

```
Largest absolute gradient value: 2.646634e-05
```

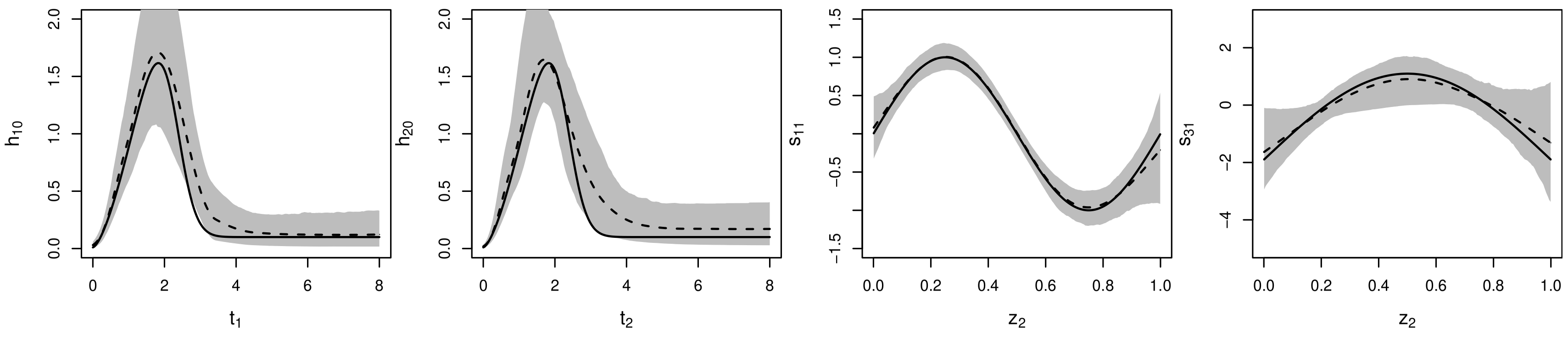

Figure A.6: Smooth function estimates obtained by applying `gjrm` to bivariate survival simulated data with mild censoring rates and $n = 300$. True functions are represented by black solid lines, mean estimates by dashed lines and point-wise ranges resulting from 5% and 95% quantities by shaded areas.

```
Observed information matrix is positive definite
Eigenvalue range: [0.008631199,1.983189e+13]

Trust region iterations before smoothing parameter estimation: 55
Loops for smoothing parameter estimation: 9
Trust region iterations within smoothing loops: 22
Estimated overall probability range: 0.0209511 0.9999631
Estimated overall density range: 3.687044e-05 5.944891
```

The function provides various information about the estimation process. Convergence is assessed by checking that the maximum of the absolute value of the score vector is virtually equal to 0 and that the observed information matrix is positive definite.

To obtain summary statistics, we can use summary() which works in a similar fashion as that of (generalised) linear and additive models.

```
summary(out)
COPULA:   Plackett
MARGIN 1: survival with -logit link
MARGIN 2: survival with -logit link

EQUATION 1
Formula: t11 ~ s(t11, bs = "mpi") + s(ENROLLAGE) + SevScale1E + rs2284665

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -18.0019     4.3929  -4.098 4.17e-05 ***
SevScale1E5   0.6709     0.2413   2.781  0.00542 **
SevScale1E6   0.9975     0.2226   4.482 7.40e-06 ***
SevScale1E7   1.9248     0.2303   8.358  < 2e-16 ***
SevScale1E8   2.8320     0.3163   8.954  < 2e-16 ***
rs22846651    0.3196     0.1667   1.918  0.05517 .
rs22846652    0.5950     0.2337   2.546  0.01090 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Smooth components' approximate significance:
               edf Ref.df  Chi.sq p-value
s(t11)       6.680  7.697 1867.11 < 2e-16 ***
s(ENROLLAGE) 1.545  1.923   14.46 0.00173 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



EQUATION 2
Formula: t21 ~ s(t21, bs = "mpi") + s(ENROLLAGE) + SevScale2E + rs2284665


Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -30.7363    10.7534  -2.858 0.004260 **
SevScale2E5   0.7855     0.2555   3.075 0.002107 **
SevScale2E6   1.1900     0.2383   4.994 5.92e-07 ***
SevScale2E7   2.4208     0.2527   9.578  < 2e-16 ***
SevScale2E8   3.2760     0.3284   9.977  < 2e-16 ***
rs22846651    0.4452     0.1689   2.635 0.008403 **
rs22846652    0.7772     0.2263   3.434 0.000595 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Smooth components' approximate significance:
               edf Ref.df   Chi.sq p-value
s(t21)       7.452  8.266 3933.136 < 2e-16 ***
s(ENROLLAGE) 1.000  1.000    6.714 0.00957 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
EQUATION 3
Link function for theta: log
Formula: ~s(ENROLLAGE) + rs2284665

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.4190     0.2387   5.946 2.75e-09 ***
rs22846651    0.3915     0.3058   1.280    0.200
rs22846652    0.3023     0.4032   0.750    0.453
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Smooth components’ approximate significance:
             edf Ref.df Chi.sq p-value
s(ENROLLAGE)   1      1  0.003   0.954

theta = 5.27(3.31,9.12)  tau = 0.353(0.253,0.456)
n = 628  total edf = 34.7
```

Since the copula parameter does not seem, on this instance, to be influenced by covariates and the effect of ENROLLAGE is linear (edf= 1) in the second margin, a more parsimonious model, the one reported in Section 4 of the manuscript, was specified:

```
eq1 <- t11 ~ s(t11, bs = "mpi") + s(ENROLLAGE) + SevScale1E + rs2284665
eq2 <- t21 ~ s(t21, bs = "mpi") +   ENROLLAGE  + SevScale2E + rs2284665
f.list <- list(eq1, eq2)
out <- gjrm(f.list, data = AREDS, surv = TRUE, BivD = "PL",
       margins = c("PO", "PO"), cens1 = cens1, cens2 = cens2,
       Model = "B", upperBt1 = "t12", upperBt2 = "t22")
summary(out)

COPULA:   Plackett
MARGIN 1: survival with -logit link
MARGIN 2: survival with -logit link


EQUATION 1
Formula: t11 ~ s(t11, bs = "mpi") + s(ENROLLAGE) + SevScale1E + rs2284665


Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -18.0368     4.3965  -4.103 4.09e-05 ***
SevScale1E5   0.6707     0.2419   2.773  0.00556 **
SevScale1E6   1.0049     0.2235   4.497 6.90e-06 ***
SevScale1E7   1.9255     0.2309   8.338  < 2e-16 ***
SevScale1E8   2.8208     0.3165   8.914  < 2e-16 ***
rs22846651    0.3269     0.1665   1.963  0.04966 *
rs22846652    0.6058     0.2328   2.602  0.00927 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Smooth components’ approximate significance:
               edf Ref.df  Chi.sq p-value
s(t11)       6.633  7.658 1879.02 < 2e-16 ***
s(ENROLLAGE) 1.604  2.007   12.89 0.00159 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



EQUATION 2
Formula: t21 ~ s(t21, bs = "mpi") + ENROLLAGE + SevScale2E + rs2284665


Parametric coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -33.28111   10.89158  -3.056 0.002246 **
ENROLLAGE     0.03643    0.01443   2.524 0.011592 *
```

```
SevScale2E5   0.81869    0.25569   3.202 0.001365 **
SevScale2E6   1.20579    0.23953   5.034 4.81e-07 ***
SevScale2E7   2.42703    0.25287   9.598  < 2e-16 ***
SevScale2E8   3.27930    0.32983   9.942  < 2e-16 ***
rs22846651    0.45890    0.16852   2.723 0.006467 **
rs22846652    0.78741    0.22556   3.491 0.000481 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


Smooth components’ approximate significance:
         edf Ref.df Chi.sq p-value
s(t21) 7.404  8.227   3872  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


theta = 5.26(4.06,6.87)  tau = 0.356(0.304,0.408)
n = 628  total edf = 31.6
```

Refer to Section 4 of the manuscript for the interpretation of these summaries. Function `plot()` can be used to visualise results.

```
par(mfrow = c(1, 3), mar = c(4, 5, 2, 0) + 0.1  )
plot(out, eq = 1, scale = 0, select = 1)
plot(out, eq = 1, scale = 0, select = 2)
plot(out, eq = 2, scale = 0, select = 1)
```

They correspond to the three estimated smooth functions reported in Figure 1 of Section 4.

To obtain 3D plots, such as the one reported in the left panel of Figure 2 of Section 4 which represents the joint progression-free probability contours for subjects who are 69 years old with AMD severity score equal to 6 for both eyes but with different genotypes of `rs2284665`, the following `R` code chunk was used

```
size <- 40
```

```
x  <-  y  <- seq(from = 0, to = 12, length.out = size)
t11 <- rep(x = x, each = size)
t21 <- rep(x = y, times = size)

newd0 <- data.frame(t11 = t11, t21 = t21, ENROLLAGE = 69, SevScale1E = 4,
                    SevScale2E = 4, rs2284665 = 1)
newd1 <- data.frame(t11 = t11, t21 = t21, ENROLLAGE = 69, SevScale1E = 6,
                    SevScale2E = 6, rs2284665 = 1)
newd2 <- data.frame(t11 = t11, t21 = t21, ENROLLAGE = 69, SevScale1E = 8,
                    SevScale2E = 8, rs2284665 = 1)

res0 <- jc.probs(out, type = "joint", newdata = newd0)
res1 <- jc.probs(out, type = "joint", newdata = newd1)
res2 <- jc.probs(out, type = "joint", newdata = newd2)

z0  <- matrix(data = res0$p12, nrow = size, byrow = TRUE)
z1  <- matrix(data = res1$p12, nrow = size, byrow = TRUE)
z2  <- matrix(data = res2$p12, nrow = size, byrow = TRUE)

persp3D(x = x, y = y, z = z0, zlim = c(0, 1), box = TRUE, plot = TRUE,
        theta = 50, phi = 10, expand = 1, col = "grey90",
        xlab = "Years (Left)", ylab = "Years (Right)",
        zlab = "Progression-free Probability", ticktype = "detailed",
        facets = FALSE, bty = "b2")

persp3D(x = x, y = y, z = z1, zlim = c(0, 1), box = FALSE, plot = TRUE,
        add = TRUE, theta = 50, phi = 10, expand = 1, col = "grey50",
        facets = FALSE)
persp3D(x = x, y = y, z = z2, zlim = c(0, 1), box = FALSE, plot = TRUE,
        add = TRUE, theta = 50, phi = 10, expand = 1, col = "grey5",
        facets = FALSE)
```

```
legend(x = 0.2, y = 0.3, legend = c("4", "6", "8"),
       fill = c("grey90","grey50","grey5"), bty = "n")
```

The remaining 3D plots of Figure 2 were obtained by modifying the above code accordingly.

Interaction terms can also be included in the model by using the same syntax employed for generalised linear and additive models. One example is give below

```
eq1 <- t11 ~ s(t11, bs = "mpi") + s(ENROLLAGE) + ti(t11, ENROLLAGE) +
             SevScale1E * rs2284665
eq2 <- t21 ~ s(t21, bs = "mpi") +   ENROLLAGE * rs2284665 +
             SevScale2E
f.list <- list(eq1, eq2)
out <- gjrm(f.list, data = AREDS, surv = TRUE, BivD = "PL",
       margins = c("PO", "PO"), cens1 = cens1, cens2 = cens2,
       Model = "B", upperBt1 = "t12", upperBt2 = "t22")
```

Other familiar functions such as `AIC()`, `BIC()`, `predict()` can be used in the usual manner to extract the information criteria and to make prediction. Further details can be found in the documentation of the `GJRM` package in `R`.

Following a referee suggestion, we also fitted models with alternative copulae (here we used the second and third best ones).

```
outF <- gjrm(f.list, data = AREDS_formatted, surv = TRUE,
            BivD = "F", margins = c("PO", "PO"),
            cens1 = cens1, cens2 = cens2, Model = "B",
            upperBt1 = 't12', upperBt2 = 't22')


outG180 <- gjrm(f.list, data = AREDS_formatted, surv = TRUE,
            BivD = "F", margins = c("PO", "PO"),
            cens1 = cens1, cens2 = cens2, Model = "B",
            upperBt1 = 't12', upperBt2 = 't22')
```

The substantive conclusions did not change. Below, we report the summary results from the former model.

```
COPULA:   Frank
MARGIN 1: survival with -logit link
MARGIN 2: survival with -logit link

EQUATION 1
Formula: t11 ~ s(t11, bs = "mpi") + s(ENROLLAGE) + SevScale1E + rs2284665

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -17.9941     4.4087  -4.081 4.47e-05 ***
SevScale1E5   0.6737     0.2435   2.767  0.00566 **
SevScale1E6   0.9966     0.2238   4.453 8.46e-06 ***
SevScale1E7   1.9478     0.2309   8.435  < 2e-16 ***
SevScale1E8   2.8656     0.3141   9.125  < 2e-16 ***
rs22846651    0.3152     0.1666   1.891  0.05859 .
rs22846652    0.5643     0.2345   2.407  0.01610 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Smooth components’ approximate significance:
               edf Ref.df  Chi.sq p-value
s(t11)       6.620  7.648 1849.16 < 2e-16 ***
s(ENROLLAGE) 1.591  1.988   13.97 0.00127 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


EQUATION 2
Formula: t21 ~ s(t21, bs = "mpi") + s(ENROLLAGE) + SevScale2E + rs2284665

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept) -30.5894    10.7055  -2.857 0.004272 **
SevScale2E5   0.7991     0.2562   3.120 0.001810 **
SevScale2E6   1.2219     0.2393   5.106 3.29e-07 ***
SevScale2E7   2.4428     0.2540   9.618  < 2e-16 ***
SevScale2E8   3.2842     0.3286   9.994  < 2e-16 ***
rs22846651    0.4644     0.1690   2.747 0.006010 **
rs22846652    0.7522     0.2246   3.350 0.000808 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Smooth components’ approximate significance:
              edf Ref.df   Chi.sq p-value
s(t21)       7.389  8.215 3857.456  <2e-16 ***
s(ENROLLAGE) 1.000  1.000    5.993  0.0144 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


EQUATION 3
Link function for theta: identity
Formula: ~s(ENROLLAGE) + rs2284665

Parametric coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.9658     0.5493   5.399  6.7e-08 ***
rs22846651    1.0241     0.7345   1.394    0.163
rs22846652    1.0612     1.0168   1.044    0.297
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Smooth components’ approximate significance:
              edf Ref.df Chi.sq p-value
```

```
s(ENROLLAGE) 1.137  1.262  0.259   0.825


theta = 3.6(2.21,4.97)  tau = 0.355(0.233,0.452)
n = 628  total edf = 34.7
```

A.6 R code to simulate data

The code used to simulate data in the simulation study is

```
datagenCopulaMixCens <- function(n, my.seed = 1){


  set.seed(my.seed)
  cor.cov <- matrix(0.5, 3, 3); diag(cor.cov) <- 1
  cov <- rMVN(n, rep(0,3), cor.cov)
  cov <- pnorm(cov)
  z1  <- round(cov[, 1])
  z2  <- cov[, 2]
  z3  <- cov[, 3]

  s11 <- function(x) sin(2*pi*x)
  s31 <- function(x) 3*sin(pi*x)

  beta11 <- -1.5
  beta21 <-  1.2
  beta31 <- -1.5

  f1 <- function(t, beta, sm.fn, u, z1, z2, z3){
    S_0 <- 0.9 * exp(-0.4*t^2.5) + 0.1*exp(-0.1*t^1)
    exp( -exp(log(-log(S_0)) + beta*z1 + sm.fn(z2)) ) - u
  }

  f2 <- function(t, beta1, beta2, sm.fn, u, z1, z2, z3){
    S_0 <- 0.9 * exp(-0.4*t^2.5) + 0.1*exp(-0.1*t^1)
    1/(1 + exp(log((1-S_0)/S_0) + beta1*z1 + beta2*z3 )) - u
```

```
}

u    <- runif(n, 0, 1)
t    <- rep(NA, n)

for (i in 1:n){
  t[i] <- uniroot(f1, c(0, 8), tol = .Machine$double.eps^0.5,
                  beta = beta11, sm.fn = s11, u = u[i],
                  z1 = z1[i], z2 = z2[i], z3 = z3[i], extendInt = "yes")$root
}

c1 <-      runif(n, 0, 2)
c2 <- c1 + runif(n, 0, 6)

dataSim <- data.frame(t.true1 = t, c11 = c1, c12 = c2, t11 = NA, t12 = NA,
                    z1, z2, z3, cens = character(n),
                      surv1 = u, stringsAsFactors = FALSE)

for (i in 1:n){

  if(t[i] <= c1[i]) {
    dataSim$t11[i] <- c1[i]
    dataSim$t12[i] <- NA # redundant but nice for clarity
    dataSim$cens[i] <- "L"
  } else if (c1[i] < t[i] && t[i] <= c2[i]){
    dataSim$t11[i] <- c1[i]
    dataSim$t12[i] <- c2[i]
    dataSim$cens[i] <- "I"
  } else if (t[i] > c2[i]){
    dataSim$t11[i] <- c2[i]
    dataSim$t12[i] <- NA # redundant but nice for clarity
    dataSim$cens[i] <- "R"}
```

```
}
%p=0.25 high
%p=0.60 mild censoring
uncens <- (dataSim$cens %in% c("L", "I")) + (rbinom(n, 1, 0.60) == 1) == 2

dataSim$t11[uncens]  <- t[uncens]
dataSim$t12[uncens]  <- NA
dataSim$cens[uncens] <- "U"

eta.theta <- beta31*z1 + s31(z2) # this is for Clayton
theta     <- exp(eta.theta)

u2      <- runif(n, 0, 1)
u_prime <- ((u2^(-theta/(1 + theta)) - 1) * u^(-theta) + 1)^(-1/theta)
t       <- rep(NA, n)

for (i in 1:n){
  t[i] <- uniroot(f2, c(0, 8), tol = .Machine$double.eps^0.5,
                  beta1 = beta21, beta2 = beta21, sm.fn = s21, u = u_prime[i],
                  z1 = z1[i], z2 = z2[i], z3 = z3[i], extendInt = "yes")$root
}

dataSim$t.true2 <- t
c1 <-      runif(n, 0, 2)
c2 <- c1 + runif(n, 0, 6)

dataSim$c21 = c1
dataSim$c22 = c2

for (i in 1:n){
  if(t[i] <= c1[i]) {
    dataSim$t21[i] <- c1[i]
```

```
      dataSim$t22[i] <- NA
      dataSim$cens[i] <- paste(dataSim$cens[i], "L", sep = "")
    } else if (c1[i] < t[i] && t[i] <= c2[i]){
      dataSim$t21[i] <- c1[i]
      dataSim$t22[i] <- c2[i]
      dataSim$cens[i] <- paste(dataSim$cens[i], "I", sep = "")
    } else if (t[i] > c2[i]){
      dataSim$t21[i] <- c2[i]
      dataSim$t22[i] <- NA
      dataSim$cens[i] <- paste(dataSim$cens[i], "R", sep = "")
    }
  }

  %p=0.25 high
  %p=0.60 mild censoring
  uncens <- (substr(dataSim$cens, 2, 2) %in% c("L", "I")) +
          (rbinom(n, 1, 0.60) == 1) == 2

  dataSim$t21[uncens]  <- t[uncens]
  dataSim$t22[uncens]  <- NA
  dataSim$cens[uncens] <- paste(substr(dataSim$cens[uncens], 1, 1), "U", sep = "")
  dataSim$surv2 = u_prime
  dataSim$surv.joint = u2

  list(dataFull = dataSim,
       dataSim = dataSim[, c('t11', 't12', 't21', 't22',
                             'z1', 'z2', 'z3', 'cens')])
}
```

Data can then be simulated and a model fitted in the following way

```
library(GJRM)

n   <- 1000
```

```
eq1 <- t11 ~ s(t11, bs = "mpi") + z1 + s(z2)
eq2 <- t21 ~ s(t21, bs = "mpi") + z1 +   z3
eq3 <-     ~                      z1 + s(z2)
f.l  <- list(eq1, eq2, eq3)

dataSim       <- datagenCopulaMixCens(n, my.seed = 1)$dataSim
dataSim$cens  <- as.factor(dataSim$cens)
dataSim$cens1 <- as.factor(substr(as.character(dataSim$cens), start = 1, stop = 1))
dataSim$cens2 <- as.factor(substr(as.character(dataSim$cens), start = 2, stop = 2))

out <- gjrm(f.l, data = dataSim, surv = TRUE, BivD = "C0",
            margins = c("PH", "PO"), cens1 = cens1, cens2 = cens2,
Model = "B", upperBt1 = 't12', upperBt2 = 't22')
```

Appendix B

Supplementary Material B

B.1 The P matrix in the continuously observed setting: an overview

When the process is assumed to be time-dependent, computing the transition probabilities from the estimated transition intensities is a nontrivial problem since closed form expressions of the former as functions of the latter are not available. Two main approaches can be identified in the literature of continuously observed processes to address this problem.

The first approach is to solve, by means of packages such as `deSolve` in `R` (Titman, 2011), the ordinary differential equations that tie the transition probability matrix to the transition intensity matrix, when the process is assumed to be Markov. This method is appealing in that it provides the entire transition probability matrix in one step and is the technique implemented in the `R` packages `rstpm2` (Clements et al., 2021), through the function `markov_msm()`, and `flexsurv` (Jackson, 2021), through the function `pmatrix.fs()`. In both cases, the main required inputs are the fitted transition intensities. In the former case the transition intensities can be specified in a number of ways using a handful of existing survival modelling `R` packages, with the most flexible options provided by the `stpm2()` function present within the `rstpm2` package and the `survPen()` function from the `R` package `survPen` (Fauvernier et al., 2020). With regard to `flexsurv`, the most common parametric forms found in survival analysis, e.g. Weibull, can be assumed for the transition intensities through the function `flexsurvreg()` as well as the Royston-Parmar model through the function `flexsurvspline()`. Overall, the drawback of this approach is that it is difficult to generalise to the case in which the process is not assumed to be Markov, e.g. when it is semi-Markov, another common type of dependence on past history. Confidence intervals can then be obtained by using the covariance matrix computed from the knowledge of the first derivative of the transition probability matrix, obtained by simultaneously solving these ODEs mentioned above and an augmented version of them obtained by taking the derivative of the left and right hand-side with respect to time.

The second approach, and indeed the one that we adopt, is a simulation-based approach which allows one to estimate the transition probabilities by simulating a number M of paths through the assumed multistate process and counting the number of individuals experiencing each transition (Iacobelli & Carstensen, 2013; Touraine et al., 2016). This method benefits of the generality lacking in the previous one, i.e. both Markov and semi-Markov processes are supported, thus tying nicely with the flexibility available for the transition-specific modelling. Indeed, it is the only such approach which is general in this respect. It was proposed in Fiocco et al. (2008) and implemented in the `Stata` package `multistate` (Crowther & Lambert, 2016) and in the `R` packages `flexsurv` and `mstate` (Putter et al., 2020). In the following we will focus only on `R` packages. In particular, we will use the latter as it can be seamlessly integrated with our estimation approach, implemented in the `R` package `GJRM`. Indeed, this package allows the user to obtained simulation-based estimates of the transition probability matrix at any vector of time points through the function `mssample()` by providing the estimated transition-specific cumulative hazards computed at the time points of interest. Whether the process is Markov or semi-Markov is then simply accounted for by specifying the argument `clock = 'forward'` in the former case and `clock = 'reset'` in the latter. Note that the estimated cumulative hazards can in turn be straightforwardly obtained through the function `hazsurv.plot()` from the `GJRM` package. Confidence intervals can then be obtained by simulation from the asymptotic distribution of the maximum likelihood estimates of the model parameters. This is what is done in `flexsurv` and `mstate`. We too adopt this approach by exploiting the fact that `hazsurv.plot()` already has a built-in way of simulating cumulative hazard functions given the asymptotic distribution of the model parameters. These can then be used as one would with a single cumulative hazard curve to obtain many corresponding transition probability matrices and thus compute the quantiles on these, as explained in Section 3.5. For more details on how to fit the transition intensities and then obtain transition probabilities and the related confidence intervals for a profile of interest, we refer the reader to Supplementary Material B.3 and to the code accessible in the public repository `https://github.com/AlessiaEletti/ContinObsMultistateProcesses`, through which the results reported in the case study from Section 3.7 can be reproduced.

As an aside, in a nonparametric setting, one may also obtain the estimated transition probabilities through the Aalen-Johansen estimator which provides a way to compute the product integral tying the transition probability matrix to the matrix containing the transition-specific cumulative hazard functions, when the process is assumed to Markov (De Wreede et al., 2010). This is one of the approaches implemented in the `R` package `mstate`. Indeed, the transition specific cumulative hazard functions are computed through the `msfit()` function either via the Aalen estimator (by specifying `vartype = 'Aalen'`) or the Greenwood estimator (by specifying `vartype = 'Greenwood'`). These estimates are then used in the `probtrans()` function to compute the transition probability matrix via the mentioned Aalen-Johansen estimator.

B.2 Rewriting the model log-likelihood when only exact transitions are observed

For a multistate survival process assumed to be observed continuously and for individual $i = 1,\dots,N$, where N represents the sample size, let $T_i^{(rs)}$ be the transition-specific true event time. This can be either uncensored, i.e. exactly observed, or right-censored if the transition $r \to s$ did not occur prior to the maximum follow-up time $T_{\max}$, in which case the transition is only known to have occurred after this time. In either case, the time may also be left-truncated if the event it relates to is an intermediate one, i.e one which requires the individual to have transition to the starting state considered prior to the current observation time. Indeed, left-truncation of survival data occurs when only individuals whose event time lies within a window (T_i^{td}, ∞) are observed, otherwise no information on the individuals is available and thus the subjects are not considered for inclusion into the study. This is precisely the case here. Indeed, given an intermediate state r, an individual is at risk of experiencing the transition $r \rightarrow s$ at a given time only if they are in state r at that time. In particular, if they are known to have transitioned to state r at time T_i^{td}, then they are at risk of the transition $r \rightarrow s$ only after this time, i.e. in the window (T_i^{td}, ∞), which is thus the left-truncation time associated with the transition.

We will now sketch the steps which show how one can pass from the general overall log-likelihood associated with a Markov multistate process to the reformulation of it in terms of a sum of log-likelihoods, each associated with a specific transition, when the exact transition times are known. Showing this for the semi-Markov case is outside of the scope of this paper.

Let us assume that a random *i.i.d.* sample of size N and let $0 = t_{i0} < t_{i1} < \cdots < t_{in_i}$ be the observed transition times for individual i. At these times the process is observed to be in states $z_{i0}, z_{i1}, \ldots, z_{in_i}$. If $\ell_i(\theta)$ is the likelihood contribution of individual i, $\mathcal{A} = \{(r,s) \in \mathcal{S} \times \mathcal{S} \mid r \neq s \wedge q^{(rs)}(\cdot) \neq 0\}$ is the set of the pairs of states corresponding to allowed transitions and $\theta = \{\beta^{(rs)} \mid (r,s) \in \mathcal{A}\}$ is an overall model parameter vector, the full log-likelihood is given by

$$\ell(\theta) = \sum_{i=1}^{N} \ell_i(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \ell_{ij}(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \log(L_{ij}),$$

where

$$L_{ij} = \exp\left[\int_{t_{ij-1}}^{t_{ij}} q_{z_{ij-1}, z_{ij-1}}(u;x)du\right] q_{z_{ij-1}, z_{ij}}(t_{ij};x).$$

We will now clarify this by specialising it to a simple example which will allow us to

ID	t_i^{start}	t_i^{stop}	Transition	Status
i	0	t_{12}	$1 \rightarrow 2$	1
i	0	t_{12}	$1 \rightarrow 3$	0
i	t_{12}	t^*	$2 \rightarrow 3$	0

Table B.1: i^{th} individual in the dataset.

write out each term explicitly. In particular, we will do this for a time-homogeneous IDM for simplicity but the same reasoning can be extended to more general contexts as settings. Let us assume we have a dataset with the i^{th} individual characterised by the observed transitions described in Table B.1. The i^{th} likelihood contribution associated with the process at hand will have the following form

$$\begin{aligned} L_i &= \overbrace{q_{12}\exp[t_{12}q_{11}]}^{1^{st}\text{ term}} \cdot \overbrace{p_{22}(t^* - t_{12})}^{2^{nd}\text{ term}} \\ &= q_{12}\exp[t_{12}q_{11}] \cdot \exp[-(t^* - t_{12})q_{23}], \end{aligned} \tag{B.1}$$

i.e. is the product of the contributions associated with the two observation times t_{12} and t^*. In particular, the first term refers to the the exactly observed transition at time t_{12}. Recall that we assume the process stays in state 1 throughout time interval $(0, t_{12})$, hence the term q_{11} in the exponential, and then jumps to state 2 at time t_{12}. The second term, instead, refers to the fact that the process is observed to be in state 2 at time t_{12} and to still be in state 2 at time t^*, the maximum follow-up time. This can be re-written in the following way

$$\begin{aligned} L_i &= q_{12}\exp[t_{12}q_{11}] \cdot \exp[-(t^* - t_{12})q_{23}] \\ &= q_{12}\exp[-t_{12}(q_{12}+q_{13})] \cdot \exp[-(t^* - t_{12})q_{23}] \\ &= q_{12}\exp[-t_{12}q_{12}] \cdot \exp[-t_{12}q_{13}] \cdot \exp[-(t^* - t_{12})q_{23}] \\ &= f_{12}(t_{12}) \cdot S_{13}(t_{12}) \cdot \frac{S_{23}(t^*)}{S_{23}(t_{12})}. \end{aligned} \tag{B.2}$$

In other terms, we broke up the likelihood in the product of terms each corresponding to specific transitions and hence which are functions of nonoverlapping sets of parameters. The usefulness of writing the likelihood contribution as a product of densities and survival functions associated to each transition, rather than as the product of transition probabilities, comes from the fact that one can then group all of the terms relating to the transition $r \to s$ and obtain a transition specific likelihood contribution $L_i^{(rs)}$. In this way we thus have

$$L_i = f_{12}(t_{12}) \cdot S_{13}(t_{12}) \cdot \frac{S_{23}(t^*)}{S_{23}(t_{12})} = L_i^{(12)} \cdot L_i^{(13)} \cdot L_i^{(23)},$$

where each transition specific likelihood can be optimised as a standalone likelihood associated to what then becomes a univariate survival analysis problem. Note, further, that the left-truncation for the $2 \to 3$ transition is apparent in that we have a conditional survival function.

B.3 Further details on the case study and code

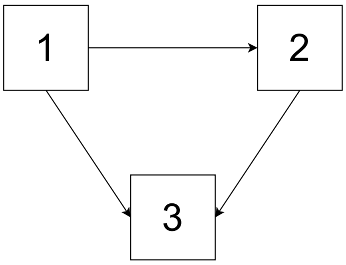

Figure B.1: Graphical representation of the IDM assumed to model the data.

The results presented in the case study have been obtained by combining the R packages GJRM and mstate, as mentioned above. Note, however, that a small bug was found in the sampling function mssample() of the latter package, which thus had to be modified. To allow for the full reproducibility of the analysis carried out this paper, we therefore not only provide the code used for the case study, but also the modified function. In this way the code provided is entirely self contained. This can be found in the public repository https://github.com/AlessiaEletti/ContinObsMultistateProcesses. In the following, instead, we report only a few code snippets to exemplify the usage of the main functions needed for the fitting of the model through our framework, implemented in the GJRM package, and the computation of the estimated transition probabilities via the simulation based procedure implemented in the mstate package. In particular, the transition-specific models are fitted using the gamlss() function from the GJRM package, as shown in the following code snippet for the 2 → 3 transition, so as to also show how we account for left-truncation.

```
out.rd = gamlss(list(Tstop ~ s(log(Tstop), bs = 'mpi')
                            +  size2 + size3 + hormon
                            + s(age) + s(nodes) + s(pr_1)
                            + ti(log(Tstop), pr_1)),
                surv = TRUE, margin = 'PH',
                data = mex[mex$trans == 3, ],
                truncation.time = 'Tstart',
                type = 'mixed',
                cens = status.factor[mex$trans == 3])
```

The plots of the smooths included in the three models, and reported in Figure 3.2 of Section 3.7 and in Figures B.2 and B.3 below, can be obtained by using the `plot()` command on the fitted model output.

We then obtain the estimated transition probabilities by combining the predicted cumulative hazards obtained using function `hazsurv.plot()` from the `GJRM` package with the (modified) function `mssample()` from the `mstate` package. Indeed, the latter takes the estimated transition-specific cumulative hazards as an input and samples paths through the multistate model outputting either the sampled paths or the estimated transition probabilities depending on the user choice; this is controlled by argument `output`. In particular, for this application, we used $M = 10000$ sampled paths through the multistate model. We show this in the following code snippet for one of the three transitions, as the others are then identical.

```
# 1-3 transition
pred.rd.test =  hazsurv.plot(out.rd, eq = 1, t.vec = times,
                              newdata = newdata, type = 'cumhaz',
                              plot.out = F)


CH13 = pred.rd.test$ch # 1-3 cumulative hazard
# ... (similarly for the others)


Hazprep = data.frame(time = rep(times, 3),
```

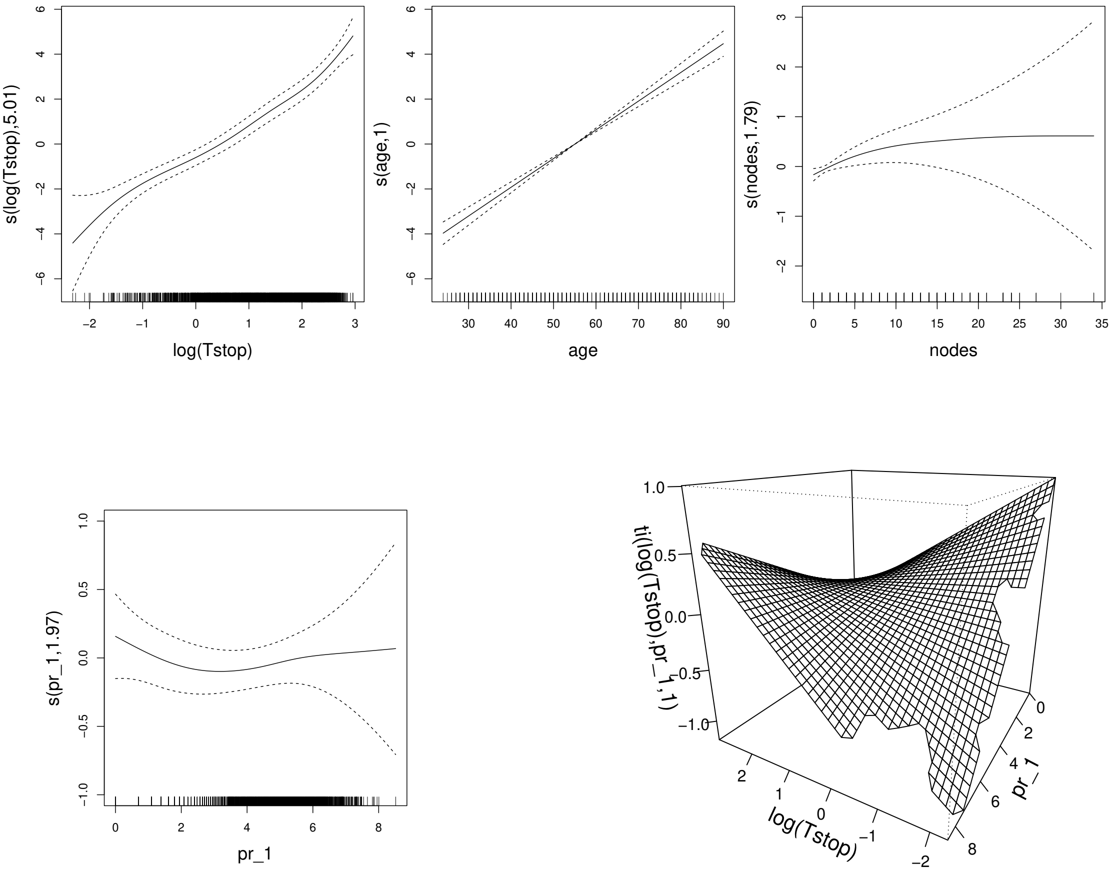

Figure B.2: Smooth of log-time (top left), smooth of age (top middle), smooth of the number of positive nodes (top right), smooth of the progesterone level (bottom left) and smooth interaction between log-time and progesterone level (bottom right) for the transition *health* → *death*.

```
          Haz = c(CH12, CH13, CH23),
          trans = c(rep(1, length(times)),
                    rep(2, length(times)),
                    rep(3, length(times))))

probs = mssample(Haz = Hazprep, trans = transmat, tvec = times, M = 10000)
```

As mentioned above, this approach allows us to model both Markov and semi-Markov processes in the exact same way, with the only exception that a different time scale needs to be defined for the latter. In particular, when fitting the model using function `gamlss()` we

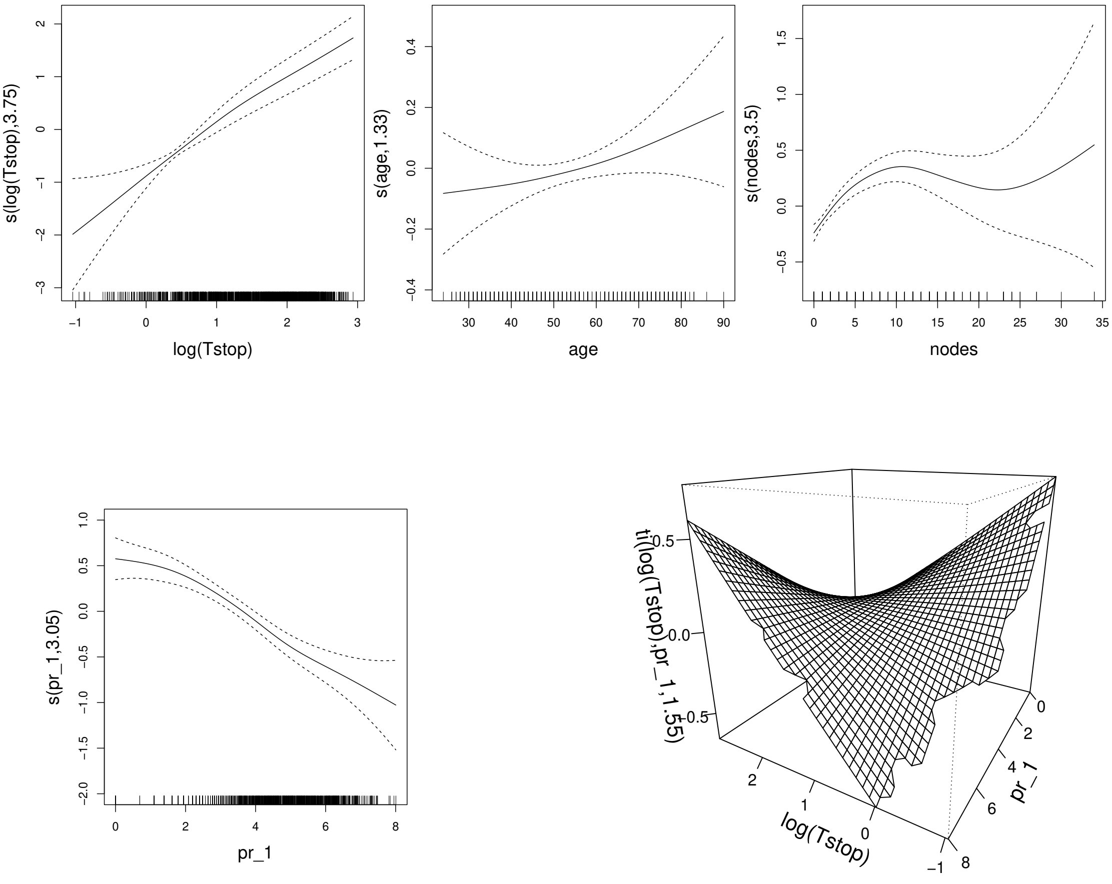

Figure B.3: Smooth of log-time (top left), smooth of age (top middle), smooth of the number of positive nodes (top right), smooth of the progesterone level (bottom left) and smooth interaction between log-time and progesterone level (bottom right) for the transition *relapse* → *death*.

would reset the time at the moment of entry to each state. When calling function `mssample()` to obtain the transition probabilities one needs to set argument `clock = 'reset'` to specify that the time-scale of the cumulative hazards is the duration in the present state. In this way we are able to fully harness the flexibility allowed when the multistate process is observed continuously through time by combining existing tools. Note that, when the process is observed only intermittently, it becomes considerably more difficult to allow for this degree of flexibility in the assumptions made on the dependence on time and past history. An attempt assuming semi-Markovianity was also made but resulted in inferior AIC values, we thus omit the results here.

We can now obtain confidence intervals for the estimated transition probabilities by simulation. In particular, we already have simulated transition-specific cumulative hazard functions from the previous calls to the `hazsurv.plot()` function. Each of these can thus be used as inputs to obtain the corresponding transition probabilities through the simulation-based procedure, i.e. by iteratively repeating the computation shown in the code snippet above for each simulated transition specific cumulative hazard. Note that the simulated transition-specific cumulative hazard can be extracted through the command `pred.rd.test$s.sim`. The quantiles of the resulting set of transition probabilities extracted in this way can thus be computed to find the 95% confidence intervals.

Finally, the plots in Figure 4.2 can be obtained using the `hazsurv.plot()` function, specifying that the curve of interest is the hazard through argument `type = ’hazard’`, as shown below for the 2 → 3 transition intensity.

```
# Transition n. 3 (2-3)
q23 = hazsurv.plot(out.rd, eq = 1,
                   t.vec = seq(min(mex$Tstop), max(mex$Tstop),
                               length.out = 1000),
                    newdata = data.frame(age = 54, size2 = 0, size3 = 1,
                                           nodes = 10, pr_1 = 3,
                                           hormon = 1),
                    type = ’hazard’,
                    ylab = ’Relapse to Death transition intensity’,
                    xlab = ’Time since surgery (years)’)
```

In conclusion, note that when fitting the time-only model with our splines-based approach, i.e. when the transition intensities are specified with no covariates, we indeed recover the estimated transition-specific cumulative hazard functions reported in Crowther & Lambert (2017). We report these in Figure B.4. These plots can be straightforwardly obtained using function `hazsurv.plot()` from the `R` package `GJRM` by specifying the argument `type = ’cumhaz’`. It can, for instance, be seen that when no covariates are considered, the risk of transitioning to the death state is considerably higher given that relapse occurred compared to the relapse-free setting.

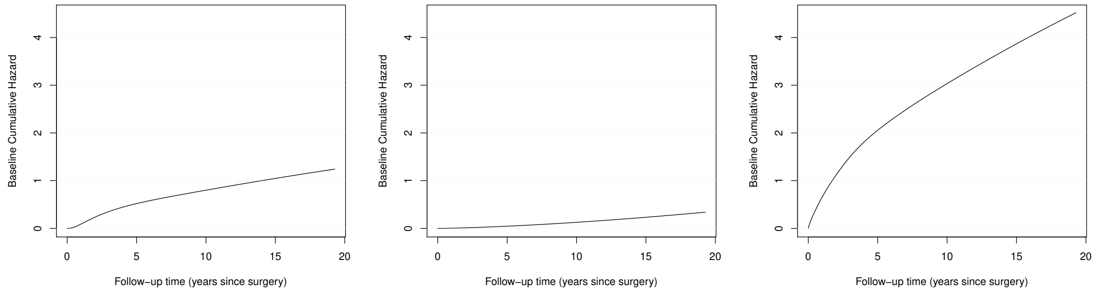

Figure B.4: Estimated baseline cumulative hazard functions associated with the health to relapse (left), health to death (middle) and relapse to death (right) transitions with their 95% confidence intervals.

B.4 Further details on the continuously observed setting

Longitudinal data are characterised by multiple observations through time of at least one quantity of interest, for the same individual and generally come in one of two forms, referred to as *stacked* (or long) and *unstacked* (or wide), respectively. In the unstacked (or wide) data format, a subject's repeated responses will be displayed in a single row, i.e. each response is in a separate column. In the stacked (or long) data format, each row represents a single time point per subject. So each subject will have data in multiple consecutive rows. In the continuously observed setting, the data will typically be formatted in the latter form. In particular, assuming an IDM like the one considered in the case study of Section 3.7, each of the rows corresponding to a given patient will look like either of the four combinations in Tables B.2-B.5. Note that the only type of censoring possible in this setting is right censoring, i.e. the transition has not taken place by the maximum follow-up time. Note also that the start time for transitions exiting the first state will usually be 0. An exception to this is had when the first transition is itself left-truncated, e.g., as a consequence of the nature of the phenomenon of interest. We refer the reader to the tutorial by Putter (2011) as well for further examples on the setup for continuously observed multistate survival processes.

trans	start	stop	event
$1 \to 3$	0	t^{max}	0
$1 \to 2$	0	t^{max}	0

Table B.2: The patient does not experience any transition between 0 and t^{max}, the maximum observed follow-up time. The patient is right censored at t^{max} for both transitions.

trans	start	stop	event
$1 \to 3$	0	t_{13}	1
$1 \to 2$	0	t_{13}	0

Table B.3: The patient experiences a transition to the absorbing state at time t_{13}. Transition $1 \to 3$ is thus uncensored. For transition $1 \to 2$ this represents a right censoring time.

trans	start	stop	event
$1 \to 3$	0	t_{12}	0
$1 \to 2$	0	t_{12}	1
$2 \to 3$	t_{12}	t^{max}	0

Table B.4: The patient experiences a transition to the intermediate state at time t_{12} but does not transition to the following state between t_{12} and t^{max}, i.e. $1 \to 2$ is uncensored. For $2 \to 3$ t_{12} is a left truncation time while t^{max} is a right censoring time. For $1 \to 3$, t_{12} represents a right censoring time.

trans	start	stop	event
$1 \to 3$	0	t_{12}	0
$1 \to 2$	0	t_{12}	1
$2 \to 3$	t_{12}	t_{23}	1

Table B.5: The patient experiences a transition to the intermediate state at time t_{12} and then transitions to the absorbing state at time t_{23}. Transitions $1 \to 2$ and $2 \to 3$ are thus uncensored. For transition $2 \to 3$, t_{12} represents a left truncation time. For transition $1 \to 3$, t_{12} represents a right censoring time.

Appendix C

Supplementary Material C

C.1 Log-likelihood contributions

This section follows Jackson et al. (2011). For a time-inhomogeneous Markov process, the likelihood contribution for the j^{th} observation of the i^{th} unit can take any of the following forms

$$L_{ij}(\theta) = \begin{cases} p^{(z_{ij-1}z_{ij})}(t_{ij-1}, t_{ij}), & \text{if } z_{ij} \text{ is an interval censored state} \\ \exp\left[\int\limits_{t_{ij-1}}^{t_{ij}} q^{(z_{ij-1}z_{ij-1})}(u)du\right] q^{(z_{ij-1}z_{ij})}(t_{ij}), & \text{if } z_{ij} \text{ is an exactly observed state} \\ \sum\limits_{c \in \tilde{\mathcal{S}} \subset \mathcal{S}} p^{(z_{ij-1}c)}(t_{ij-1}, t_{ij}), & \text{if } z_{ij} \text{ is a censored state} \\ \sum\limits_{\substack{c=1 \\ c \neq z_{ij}}}^{C} p^{(z_{ij-1}c)}(t_{ij-1}, t_{ij}) q^{(cz_{ij})}(t_{ij}), & \text{if } z_{ij} \text{ is an exactly observed death state} \end{cases} .$$

for $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, with N the total number of statistical units and n_i the number of observations for unit i and where $p^{(z_{ij-1}z_{ij})}(t_{ij-1}, t_{ij}) = P(Z(t_{ij}) = z_{ij} \mid Z(t_{ij-1}) = z_{ij-1})$. In other words, each pair of consecutively observed states contributes one term to the likelihood. Specifically, if a transition between two transient states is observed and the transition time is interval-censored then the contribution is

$$L_{ij}(\theta) = P(Z(t_{ij}) = z_{ij} \mid Z(t_{ij-1}) = z_{ij-1}),$$

If, due to the nature of the process, the transitions to some living states are exactly observed, the contribution is

$$L_{ij}(\theta) = \exp\left[\int\limits_{t_{ij-1}}^{t_{ij}} q^{(z_{ij-1}z_{ij-1})}(u)du \right] q^{(z_{ij-1}z_{ij})}(t_{ij}),$$

since the process is known to have stayed in state z_{ij-1} between t_{ij-1} and t_{ij} and then jumped from state z_{ij-1} to state z_{ij} at exactly t_{ij}. The first term can be explained by observing that

$$\exp\left[\int\limits_{t_{ij-1}}^{t_{ij}} q^{(z_{ij-1}z_{ij-1})}(u)du \right] = \exp\left[-\int_{t_{ij-1}}^{t_{ij}} \sum_{c \neq z_{ij-1}} q^{(z_{ij-1}c)}(u)du \right] = \prod_{c \neq z_{ij-1}} \frac{\exp\left[-\int_0^{t_{ij}} q^{(z_{ij-1}c)}(u)du \right]}{\exp\left[-\int_0^{t_{ij-1}} q^{(z_{ij-1}c)}(u)du \right]},$$

which implies that no transition exiting state z_{ij-1} has occurred at time t_{ij} given that it had not occurred by time t_{ij-1} either.

If the state occupied at a given time is unknown then it is said to be censored. In this case, the contribution to the likelihood has to account for all the possible trajectories that may have occurred from the last known occupied state to the current observation time. Therefore, the sum over the various probabilities is taken, which will be null if the transition is not allowed. In particular,

$$L_{ij}(\theta) = \sum_{c=1}^{C} P(Z(t_{ij}) = c \mid Z(t_{ij-1}) = z_{ij-1}).$$

Finally, if the last observed state is an absorbing one then the time at which the transition occurred is generally assumed to be known. In this case, one needs to account for the possibility that the state occupied before the absorbing state is unknown and thus the contribution to the likelihood is summed over the possible states occupied by the process. The information of the exact observation time t_{in_i} is included through the transition intensity computed in that time. Here, we have

$$L_{ij}(\theta) = \sum_{\substack{c=1 \\ c \neq z_{ij}}}^{C} p^{(z_{ij-1}c)}(t_{ij-1}, t_{ij}) q^{(cz_{ij})}(t_{ij}).$$

C.2 R package flexmsm

To support applicability and reproducibility, the proposed modelling framework has been implemented in the R package flexmsm. The package is straightforward to use, especially if the user is already familiar with the syntax of generalised linear models and generalised additive models (GAMs) in R. The key function is fmsm(), which carries out model fitting and inference, and is exemplified with some of its main arguments in the following code snippet

```
out <- fmsm(formula = formula, data = df,
            id = ID, state = state,
            death = TRUE, living.exact = NULL, cens.state = -99,
            sp.method = 'perf',
            constraint = NULL, parallel = TRUE, ...)
```

where the user specifies the model through the argument formula as a list() containing the model specifications for the transition intensities, and the dataset has to be provided through the argument data. This will always have at least three columns: the state column (whose name is provided through the argument state), the column containing the unique IDs (whose name is provided through the argument id) identifying each individual, and a column containing the (intermittent) observation times. The arguments death, living.exact and cens.state allow the user to specify the observation type. If the last state in the process is an exactly observed death state then the user must specify death = TRUE; if there are exactly observed living states then the dataset must contain an additional column with TRUE (or 1) if the data point is exactly observed and FALSE (or 0) otherwise; the name of this column must be passed through the argument living.exact, which defaults to NULL. If there are any censored states then the user must specify the code used to indicate this through the argument cens.state, which defaults to -99. The sp.method argument specifies the method employed for multiple smoothing parameter estimation (this can be set to 'perf' or 'efs'). The argument constraint allows the user to specify equality constraints on the covariates. The parallel argument allows the user to exploit parallel computing, in Windows, for the likelihood, gradient and Hessian, thus cutting the run-time of the algorithm by factor proportional to the number of cores on the computer.

The formula is a list() object whose elements are the off-diagonal elements of the transition intensity matrix. The order of the elements is that given by reading the **Q** matrix from the first row to the last and from left to right. The equation corresponding to each non-zero transition intensity has to be specified with syntax similar to that used for GAMs, with the response given by the time-to-event variable. Trivially, zero elements have to be specified with a `0`. For instance, we may consider the following model, with a smooth effect of time t and two covariates x_1 and x_2, one included linearly and the other as a time-dependant flexible effect, for a transition $r \to r'$

$$q^{(rr')}(t_{ij}) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t_{ij}) + \beta_2^{(rr')} x_{1ij} + s_3^{(rr')}(x_{2ij}) + s_4^{(rr')}(t_{ij}, x_{2ij})\right].$$

This will be specified, in the correct position, as part of the list

```
formula <- list(...,
                t ~ s(t) + x1 + s(x2), ti(t, x2), # r -> r' trans.
                ...)
```

where `...` represent other possible transition-specific equations or `0`s for transitions not allowed by the process. The model specified here is only an example and many types of effects are supported. For instance, as the above example shows, time-dependent effects are modelled by using a tensor interaction function `ti()` on the covariate of interest and time.

Functions `summary()` and `plot()` can be used in the usual way to obtain post-estimation summaries for each non-zero transition intensity and the plots of the smooths. In the example above there is a two-dimensional spline, thus `plot()` will also automatically produce a three-dimensional plot of the surface representing this time-dependent effect.

Function `conv.check()` allows the user to check the convergence of the fitted model by providing information on whether the gradient is zero and the Hessian is positive definite. It also provides information on the values taken by the **Q** matrix since, in practice, we have found that particularly large values are red flags for ill-defined problems, for instance.

Prediction and plotting of the **P** and the **Q** matrices can be carried out through the functions `P.pred()` and `Q.pred()`, respectively. For instance, the specification

```
P.hat <- P.pred(out, newdata = newdata, plot.P = TRUE
                 get.CI = TRUE, prob.lev = 0.05)
```

will provide an object `P.hat` containing the estimated transition probability matrix corresponding to the time interval and profile of interest, specified through argument `newdata`.

The intermediate transition probabilities corresponding to each sub-interval specified in `newdata` are also provided. The $100(1-\texttt{prob.lev})\%$ confidence intervals can be obtained by setting `get.CI = TRUE`. When `plot.P = TRUE` the transition probabilities are also plotted as function of time over the interval considered, otherwise the plots are suppressed. The analogous output can be obtained for the **Q** matrix through function `Q.pred()` with similar syntax.

To exemplify the usage of the software, we report the code used to fit the models presented in Section 4.6. We recall that the IDM specified in Section 4.6.1 is given by

$$q^{(rr')}(t_{ij}) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t_{ij}) + \beta_2^{(rr')}\texttt{dage}_{ij} + \beta_3^{(rr')}\texttt{pdiag}_{ij}\right].$$

This can be fitted in the following way:

```
formula <- list(t ~ s(t, bs = 'cr', k = 10) + dage + pdiag, # 1-2
                t ~ s(t, bs = 'cr', k = 10) + dage + pdiag, # 1-3
                0,                                          # 2-1
                t ~ s(t, bs = 'cr', k = 10) + dage + pdiag, # 2-3
                0,                                          # 3-1
                0)                                          # 3-2


fmsm.out <- fmsm(formula = formula, data = Data,
              id = PTNUM, state = state, death = TRUE,
              sp.method = 'perf', parallel = TRUE)
```

Here `bs = 'cr'` and `k = 10` imply that the smooths of time are specified through cubic regression splines with ten basis functions. We will omit this in the following to avoid redundancies. To obtain the two-dimensional spline based model, it suffices to swap the `formula` reported above with the following

```
formula <- list(t ~ s(t) + s(dage) + ti(t, dage) + pdiag, # 1-2
                t ~ s(t) + s(dage) + ti(t, dage) + pdiag, # 1-3
                0,                                        # 2-1
                t ~ s(t) + s(dage) + ti(t, dage) + pdiag, # 2-3
                0,                                        # 3-1
                0)                                        # 3-2
```

For the five-state model described in Section 4.6.2, the first model explored was

$$q^{(rr')}(t_{ij}) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t_{ij})\right].$$

This can be implemented in the following way:

```
formula <- list(t ~ s(t) + sex + edu, # 1-2
                0,                    # 1-3
                0,                    # 1-4
                t ~ s(t) + sex + edu, # 1-5
                t ~ s(t) + sex + edu, # 2-1
                t ~ s(t) + sex + edu, # 2-3
                0,                    # 2-4
                t ~ s(t) + sex + edu, # 2-5
                0,                    # 3-1
                t ~ s(t) + sex + edu, # 3-2
                t ~ s(t) + sex + edu, # 3-4
                t ~ s(t) + sex + edu, # 3-5
                0,                    # 4-1
                0,                    # 4-2
                t ~ s(t) + sex + edu, # 4-3
                t ~ s(t) + sex + edu, # 4-5
                0,                    # 5-1
                0,                    # 5-2
                0,                    # 5-3
                0)                    # 5-4


fmsm.out <- fmsm(formula = formula, data = ELSA.df,
              id = idauniq, state = state, death = TRUE,
              sp.method = 'efs')
```

C.3 Parameter estimation

The algorithm employed for model fitting is characterised by two steps. In the first step, $\boldsymbol{\lambda}$ is held fixed at a vector of values and for a given $\theta^{[a]}$, where a is an iteration index, equation (4.4) is maximised using

$$\theta^{[a+1]} = \theta^{[a]} + \underset{\mathbf{e}:\|\mathbf{e}\|\leq\Delta^{[a]}}{\arg\min}\ \breve{\ell}_p(\theta^{[a]}), \tag{C.1}$$

where $\breve{\ell}_p(\theta^{[a]}) = -\left\{\ell_p(\theta^{[a]}) + \mathbf{e}^\mathsf{T}\mathbf{g}_p(\theta^{[a]}) + \frac{1}{2}\mathbf{e}^\mathsf{T}\mathbf{H}_p(\theta^{[a]})\mathbf{e}\right\}$, $\mathbf{g}_p(\theta^{[a]}) = \mathbf{g}(\theta^{[a]}) - \mathbf{S}_\lambda\theta^{[a]}$, and $\mathbf{H}_p(\theta^{[a]}) = \mathbf{H}(\theta^{[a]}) - \mathbf{S}_\lambda$. $\mathbf{g}(\theta^{[a]}) = \partial\ell(\theta)/\partial\theta|_{\theta=\theta^{[a]}}$ and $\mathbf{H}(\theta^{[a]}) = \partial^2\ell(\theta)/\partial\theta\partial\theta^\mathsf{T}|_{\theta=\theta^{[a]}}$ are given in Section 4.4, $\|\cdot\|$ denotes the Euclidean norm, and $\Delta^{[a]}$ is the radius of the trust region which is adjusted through the iterations. The first line of (C.1) uses a quadratic approximation of $-\ell_p$ about $\theta^{[a]}$ (the so-called model function) to choose the best $\mathbf{e}^{[a+1]}$ within the ball centered in $\theta^{[a]}$ of radius $\Delta^{[a]}$, the trust-region. Throughout the iterations, a proposed solution is accepted or rejected and the trust region adjusted (i.e., expanded or shrunken) based on the ratio between the improvement in the objective function when going from $\theta^{[a]}$ to $\theta^{[a+1]}$ and that predicted by the approximation. The use of the observed information matrix gives global convergence guarantees due to Moré & Sorensen (1983). Importantly, convergence to a point satisfying the second-order sufficient conditions (i.e., a local strict minimiser) is super-linear. Near the solution, the algorithm proposals become asymptotically similar to Newton-Raphson steps, hence benefitting from the resulting fast convergence rate. Trust region algorithms are also generally more stable and faster compared to in-line search methods. See Nocedal & Wright (Chapter 4, 2006) for proofs and further details.

In the second step, at $\theta^{[a+1]}$, there are two options to estimate the smoothing parameter vector: the stable and efficient multiple smoothing parameter approach adopted by Marra & Radice (2020), and the generalised Fellner-Schall method of Wood & Fasiolo (2017). Both techniques can be employed for fitting penalised likelihood-based models, and require the availability of the analytical score and information matrix. In the former, the following problem is solved

$$\boldsymbol{\lambda}^{[a+1]} = \underset{\boldsymbol{\lambda}}{\arg\min}\ \|\mathbf{M}^{[a+1]} - \mathbf{O}^{[a+1]}\mathbf{M}^{[a+1]}\|^2 - \check{n} + 2\mathrm{tr}(\mathbf{O}^{[a+1]}). \tag{C.2}$$

The idea is to estimate $\boldsymbol{\lambda}$ so that the complexity of the smooth terms not supported by the data is suppressed. This is formalised as $\mathbb{E}\left(\|\boldsymbol{\mu}_{\mathbf{M}} - \widehat{\boldsymbol{\mu}}_{\mathbf{M}}\|^2\right) = \mathbb{E}\left(\|\mathbf{M} - \mathbf{OM}\|^2\right) - \check{n} + 2\text{tr}(\mathbf{O})$, where $\mathbf{M} = \boldsymbol{\mu}_{\mathbf{M}} + \boldsymbol{\varepsilon}$, $\boldsymbol{\mu}_{\mathbf{M}} = \sqrt{-\mathbf{H}(\theta)}\theta$, $\boldsymbol{\varepsilon} = \sqrt{-\mathbf{H}(\theta)}^{-1}\mathbf{g}(\theta)$, $\mathbf{O} = \sqrt{-\mathbf{H}(\theta)}(-\mathbf{H}(\theta) + \mathbf{S}_\lambda)^{-1}\sqrt{-\mathbf{H}(\theta)}$, and $\text{tr}(\mathbf{O})$ is defined in Section 5 of the main paper. It can be proved that (C.2) is approximately equivalent to the AIC with number of parameters given by $\text{tr}(\mathbf{O})$. Iteration (C.2) is implemented via the routine by Wood (2004), which is based on the Newton method and can evaluate in an efficient and stable manner the terms in (C.2), their scores and Hessians, with respect to $\log(\boldsymbol{\lambda})$.

The approach proposed in Wood & Fasiolo (2017) is based on a different principle. The starting point is the well established stance that smoothing penalties can be viewed as resulting from improper Gaussian prior distributions on the spline coefficients. This is also the Bayesian viewpoint taken for the inferential result discussed in Section 4.5, and implies the following improper joint log-density, where the dependence on the smoothing parameter has been made explicit,

$$\log L(\theta;\lambda) = \ell(\theta) - \frac{1}{2}\theta^{\mathsf{T}}\mathbf{S}_\lambda\theta + \frac{1}{2}\log|\mathbf{S}_\lambda|.$$

The idea is to develop an update for λ that maximises the restricted marginal likelihood $L(\lambda)$, obtained integrating θ out of the likelihood $L(\theta;\lambda)$. It is, however, more computationally efficient and equally theoretically founded to maximise the log Laplace approximation

$$\ell_{LA}(\lambda) = \ell(\hat{\theta}) - \frac{1}{2}\hat{\theta}^{\mathsf{T}}\mathbf{S}_\lambda\hat{\theta} + \frac{1}{2}\log|\mathbf{S}_\lambda| - \frac{1}{2}\log|-\mathbf{H}(\hat{\theta}) + \mathbf{S}_\lambda|,$$

where $\hat{\theta} = \arg\max_\theta L(\theta;\lambda)$ for a given λ. At $\theta^{[a+1]}$, the update for the k^{th} element of $\lambda^{(rr')}$ for all $(r,r') \in \mathcal{A}$ is

$$\lambda_k^{(rr')[a+1]} = \lambda_k^{(rr')[a]} \times \frac{\text{tr}\left\{\mathbf{S}_{\lambda^{[a]}}^{-1}\frac{\partial\mathbf{S}_\lambda}{\lambda_k^{(rr')}}\bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{[a]}}\right\} - \text{tr}\left\{[-\mathbf{H}(\hat{\theta}) + \mathbf{S}_{\lambda^{[a]}}]^{-1}\frac{\partial\mathbf{S}_\lambda}{\partial\lambda_k^{(rr')}}\bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{[a]}}\right\}}{\hat{\theta}^\top\left(\frac{\partial\mathbf{S}_\lambda}{\partial\lambda_k^{(rr')}}\bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{[a]}}\right)\hat{\theta}}, \tag{C.3}$$

with $k = 1,\ldots,K^{(rr')}$. The two steps, (C.1) and either (C.2) or (C.3), are iterated until the algorithm satisfies the stopping rule $\frac{|\ell(\theta^{[a+1]}) - \ell(\theta^{[a]})|}{0.1 + |\ell(\theta^{[a+1]})|} < 1e-07$, and convergence is assessed by checking that the maximum of the absolute value of the gradient vector is numerically equivalent to 0 and that the observed information matrix is positive definite. In practice, we found the two smoothing methods to yield similar smooth term estimates.

C.4 Simulation study

To exemplify the empirical effectiveness of the proposed approach in recovering the true values of key quantities of interest (e.g., transition intensity curves), we carried out two simulation studies. The first one replicates that designed in Mariano Machado et al. (2021) and uses an IDM set-up. The second study is about a five-state Markov process and serves to illustrate the performance of the proposal in a setting that is more complex than those supported by the methods available in the literature.

C.4.1 IDM based simulation

We consider a progressive IDM, assuming a different time-dependent shape for each of the three allowed transitions. The time-to-events relating to transition $1 \rightarrow 2$ are simulated from a log-normal distribution with location 1.25 and scale 1. This implies that the hazard increases first and then decreases at a later time. For $1 \rightarrow 3$, an exponential distribution with rate $\exp(-2.5)$ is employed. For $2 \rightarrow 3$, we assume a strictly increasing hazard by simulating the time-to-events from a conditional Gompertz distribution with rate $\exp(-2.5)$ and shape 0.1. For this transition, we have to condition on the event that the individual transitions to state 2 to ensure that the simulated time is larger than the $1 \rightarrow 2$ transition time. As in Mariano Machado et al. (2021), we simulate $N = 500$ trajectories (i.e., individuals) $\mathcal{M} = 100$ times.

More specifically, let $T_{rs} = T_{rs|u}$ represent the time of the transition to state r' conditional on being in state r at time $u > 0$. If the state at u is 1 then the time of transition to the next state can be obtained by taking $T = \min\{T_{12}, T_{13}\}$. If $T = T_{12}$ then the next state is 2, otherwise the next state is 3. If the state is 2 then the time of the next state is T_{23}. Censoring needs to be imposed to render the data intermittently observed; we assume a yearly time-grid spanning over 15 years, i.e. $(t_{i0}, t_{i1}, \dots, \min\{t_{i15}, T_{13}\}) = (0, 1, \dots, \min\{t_{i15}, T_{13}\})$ for $i = 1, \dots, N$. The reader is referred to Van Den Hout (2016) for further details on how to simulate intermittently-observed multi-state survival data. The transition intensities are specified as $q^{(rr')}(t) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t)\right]$ for $(r, r') \in \{(1,2), (1,3), (2,3)\}$, where $s_1^{(rr')}(t)$ is represented using a cubic regression spline with $J_1^{(rr')} = 10$ and second order penalty.

In line with Mariano Machado et al. (2021), Figure C.1 shows the estimated median and true hazards as well as all the $\mathcal{M}$ estimated hazards. Note that the large variation observed towards the end of the study time is due to scarceness of data at later years. Overall, the plots show that the proposed approach is able to recover well the true transition intensity curves for each allowed transition, and that the performance is similar across the two methods. The discrepancy between fitted median and true hazards for transition 1 → 2 is due to definition of interval censoring adopted in the simulation study: the sampling design implies that the living states are observed at intervals of one year; for the first two years after baseline, this design does not work well.

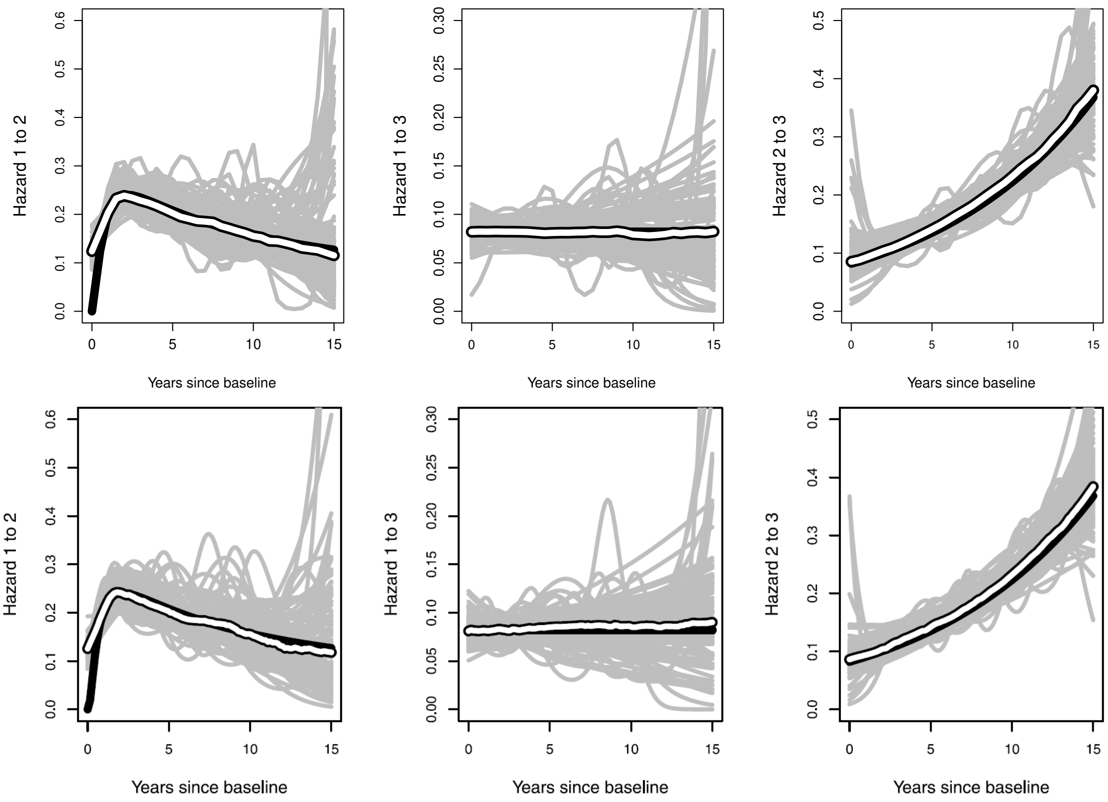

Figure C.1: True (black), estimated (grey, $\mathcal{M} = 100$ replicates) and median estimated (white) hazard functions for transitions $1 \to 2$ (left), $1 \to 3$ (middle) and $2 \to 3$ (right) obtained by `flexmsm` (top row) and Mariano Machado et al. (2021) (bottom row).

We also evaluated our approach on the transition probability scale. In particular, Table C.1 reports the true, average and median ten-year estimated transition probabilities, where the average is taken over the $\mathcal{M}$ simulations. The biases are also reported and are defined as $\text{Bias}^{(rr')}(t) = \frac{1}{\mathcal{M}}\left(\sum_{\nu=1}^{\mathcal{M}} p^{(\nu,rr')}(0,10) - p^{(rr')}(0,10)\right)$, where $p^{(\nu,rr')}(0,10)$ denotes the estimated ten-year probability of transitioning from state r to state r' for the ν^{th} simulated dataset. Our methodology recovers well the true ten-year transition probabilities and consistently outperforms the approach of Mariano Machado et al. (2021).

True	flexmsm		M. et al. (2021)	
	Mean	Bias	Mean	Bias
$p^{(11)}(0,10) = 0.065$	0.063	−0.002	0.060	0.004
$p^{(12)}(0,10) = 0.231$	0.232	0.001	0.222	0.009
$p^{(13)}(0,10) = 0.704$	0.705	0.001	0.718	−0.014
$p^{(22)}(0,10) = 0.245$	0.242	−0.003	0.231	0.014
$p^{(23)}(0,10) = 0.755$	0.758	0.003	0.769	−0.014

Table C.1: Ten-year true and average estimated transition probabilities, and bias for $\mathcal{M} = 100$ replicates.

Finally, we explored the effect that the length of the gap occurring between two successive observations has on estimation performance; it is known that when such gap is large, identifiability issues may arise. To this end, we additionally considered two-, three-, four- and five-yearly time-grids. As expected, the performance deteriorated as the gap increased, with reasonable results (not reported here, but available upon request) still attainable for two- and three-yearly time-grids.

C.4.1.1 Approximate information matrix

This section provides some evidence on the convergence performance of the proposed approach when employing an information matrix approximated via first order analytical derivatives. When comparing the results for the analytic Hessian based estimation (M1) and approximate one (M2), we found that the proportions of simulated replicates M1 was better than M2 were:

- 94.5% when analysing the total numbers of iterations for the trust region and smoothing steps discussed in Section C.3;
- 81.5% when examining the log-likelihoods of M1 and M2;
- 61.5%s when comparing the gradients of M1 and M2.

Note that, in many of the simulated replicates, M2 exhibited a behavior similar to that depicted in Figure C.2, hence highlighting the importance of exploiting in model fitting the information provided by the analytical Hessian matrix of the log-likelihood.

Finally, we would like to point out that the model specification employed for the simulation set up explored here is simpler than those investigated in the CAV and ELSA studies; as mentioned in the previous section, this was done for comparability with the results

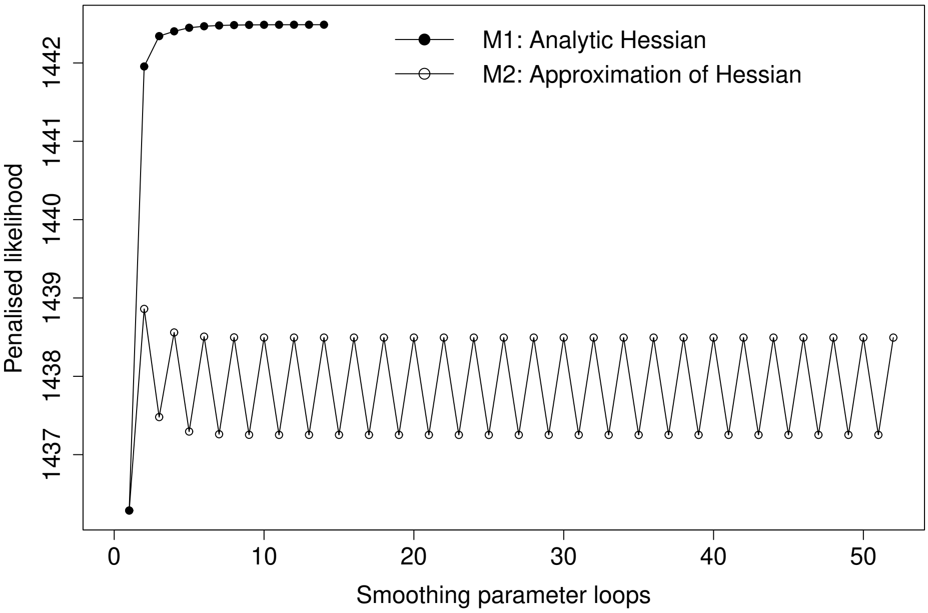

Figure C.2: Penalized log-likelihood at each iteration of the proposed estimation approach, based on M1 and M2, for model (4.10) in the CAV case study. The run-times on a laptop with Windows 10, Intel 2.20 GHz processor, 16 GB of RAM and eight cores, were 17 minutes for M1 and over 2 hours for M2.

of Mariano Machado et al. (2021). The difference in performance between M1 and M2 becomes even starker as the complexity of the model specification increases. In fact, in the case studies, it was not possible to estimate the models of interest when basing estimation on the approximate Hessian, because of convergence failures of the type displayed in Figure C.2.

C.4.2 Five-state process based simulation

We consider a progressive five-state survival process with an absorbing state, and seven transitions whose parameters were chosen to produce intensities similar to those found in the ELSA case study described in Section 4.6.2. In particular, we simulate the time-to-events from (conditional) Gompertz distributions with rates and shapes provided for each transition in Table C.2. We simulate $N = 500$ trajectories $\mathcal{M} = 100$ times, which are observed for 40 semesters. An intermittently observation scheme is imposed by assuming that individuals are visited every 4 semesters. The time is then brought back to the year scale. This gives counts of pairs of consecutively observed states that are similar to those found in the ELSA case study.

	$1 \to 2$	$1 \to 5$	$2 \to 3$	$2 \to 5$	$3 \to 4$	$3 \to 5$	$4 \to 5$
log(rate)	-2.25	-5	-2.20	-5	-2	-5	-3
shape	0.06	0.02	0.05	0.09	0.01	0.02	0.04

Table C.2: Rates and shapes for the (conditional) Gompertz distributions generating the transition times in the five-state process based simulation.

The transition intensities are specified as $q^{(rr')}(t) = \exp\left[\beta_0^{(rr')} + s_1^{(rr')}(t)\right]$ for $(r,r') \in \{(1,2),(1,5),(2,3),(2,5),(3,4),(3,5),(4,5)\}$, where $s_1^{(rr')}(t)$ is represented using a cubic regression spline with $J_1^{(rr')} = 10$ and second order penalty.

In Figure C.3, we report the median estimated transition intensities obtained for the $\mathcal{M}$ simulations with our framework, alongside the true curve $q^{(rr')}(t)$, for each of the seven allowed transitions. Overall, the proposed approach recovers adequately the true transition intensity curves.

As done for the three-state simulated process, we also evaluate our approach on the transition probabilities scale. In Table C.3, we report the true and average ten-year estimated transition probabilities, where the average is taken over the $\mathcal{M}$ simulations, and the corresponding biases. The method is able to recover the true ten-year transition probabilities reasonably well, exhibiting consistently small biases. This is reassuring considering the multi-state process adopted here, which is more involved and complex that those commonly explored and used in the literature.

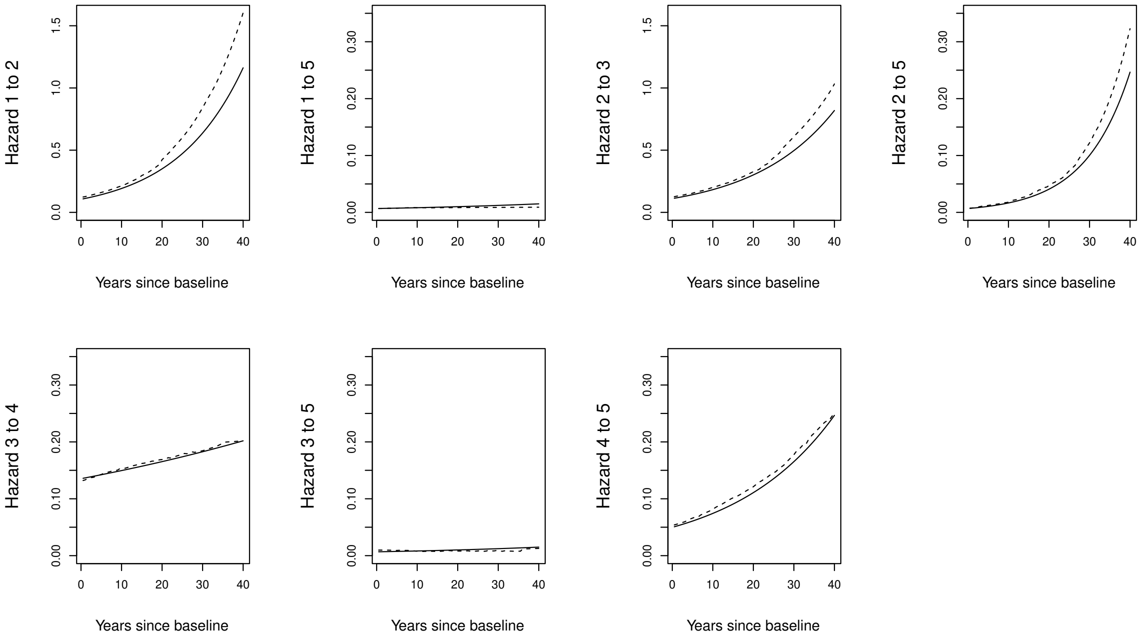

Figure C.3: True (black) and median estimated (dashed) hazard functions for each transition in the simulated five-state process.

True	Mean	Bias
$p^{(11)}(0,10) = 0.229$	0.192	-0.037
$p^{(12)}(0,10) = 0.318$	0.300	-0.018
$p^{(13)}(0,10) = 0.230$	0.255	0.025
$p^{(14)}(0,10) = 0.121$	0.137	0.016
$p^{(15)}(0,10) = 0.102$	0.116	0.014
$p^{(22)}(0,10) = 0.222$	0.186	-0.036
$p^{(23)}(0,10) = 0.330$	0.333	0.003
$p^{(24)}(0,10) = 0.294$	0.299	0.006
$p^{(25)}(0,10) = 0.154$	0.181	0.027
$p^{(33)}(0,10) = 0.225$	0.222	-0.003
$p^{(34)}(0,10) = 0.508$	0.481	-0.027
$p^{(35)}(0,10) = 0.267$	0.297	0.03
$p^{(44)}(0,10) = 0.549$	0.527	-0.021
$p^{(45)}(0,10) = 0.451$	0.473	0.021

Table C.3: Ten-year true, average and median transition probabilities for our framework. The order is that found when reading the transition probability matrix row-wise.

C.5 List of symbols

Covariates and functions or longer terms

- age_ι covariate in model

- $\text{Bias}^{(rr')}(t)$ bias relating to the $r \to r'$ transition at time t in the simulation study

- $\texttt{dage}_{ij}$ covariate in CAV model

- $\texttt{pdiag}_{ij}$ covariate in CAV model

- sex_ι covariate in model

- $\texttt{sex}_{ij}$ covariate in ELSA model

- $\texttt{higherEdu}_{ij}$ covariate in ELSA model

- *edf* for effective degrees of freedom

- $\text{tr}(\cdot)$ trace function

- $\mathbf{1}_{\check{n}}$ vector of 1s of length $\check{n}$.

Latin letters

- a estimation algorithm iteration index.

- $\mathbf{A}$ matrix of eigenvectors.

- $\mathcal{A}$ set of allowed transitions

- $\mathbf{b}_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota})$ bases function vector for the k^{th} term in the (r,r') transition intensity.

- c indexing for likelihood contributions (censored state contribution and for exactly observed absorbing state).

- C total number of states.

- d_υ difference of knots in the construction of the cubic regression spline.

- $\mathbf{D}_k^{(rr')}$ penalty matrix for the k^{th} term in the (r,r') transition intensity.

- $\mathbf{e}$ vector in the Taylor approximation.

- $\mathbf{E}$ matrix found in the closed-form expressions of the first and second derivatives of the transition probability matrix.

- $\mathbb{E}$ expectation function.

- f_θ prior on the model parameter θ.

- $G_{lm}^{(w)}$ the (l,m) element of $\mathbf{G}^{(w)}$.

- $G_{lm}^{(ww')}$ the (l,m) element of $\mathbf{G}^{(ww')}$.

- $\mathbf{g}(\theta)$ gradient vector.

- $\mathbf{G}^{(w)}$ matrix needed for the closed form expression of $\partial^2\mathbf{P}$ (transformation of first derivative of Q matrix).

- $\mathbf{G}^{(ww')}$ matrix needed for the closed form expression of $\partial^2\mathbf{P}$ (transformation of second derivative of Q matrix).

- h infinitesimal time in the limit-based definition of the transition intensity.

- $\mathbf{H}(\theta)$ hessian matrix.

- $\mathbf{H}_p(\theta)$ penalized hessian matrix.

- i indexing for the statistical units when defining the likelihood. Here $i = 1,\dots,N$.

- j indexing for the observations of a specific statistical unit.

- $J_k^{(rr')}$ number of basis functions for the k^{th} term in (r,r') transition intensity.

- k indexing for overall covariate/parameter vector, with $k = 1,\dots,K^{(rr')}$.

- $K^{(rr')}$ total number of terms in additive predictor $\eta_i^{(rr')}(t_i,\mathbf{x}_i;\beta^{(rr')})$, excluding the intercept.

- l indexing for the (l,m) element of the matrices needed for the closed form expression of $\partial^2\mathbf{P}$.

- ℓ_{LA} log Laplace approximation of $L(\lambda)$.

- $\ell(\theta)$ model log-likelihood.

- $\ell_p(\theta)$ penalized log-likelihood.

- $\breve{\ell}_p(\theta)$ second order approximation of the model log-likelihood.

- $L_{ij}(\theta)$ likelihood contribution for j^{th} observation of i^{th} individual.

- $L(\theta;\lambda)$ joint log density (used to explain efs smoothing approach).
- $L(\lambda)$ joint log density when integrating out θ (used to explain efs smoothing approach).
- m indexing for the (l,m) element of the matrices needed for the closed form expression of $\partial^2\mathbf{P}$.
- $\mathcal{M}$ number of simulations in the simulation study.
- $\mathbf{M}$ matrix appearing in the update of the smoothing parameter.
- N total number of statistical units.
- $\check{n}$ total number of observations in the dataset.
- n_i number of observations for the i^{th} statistical unit with $i = 1,\dots,N$.
- n_{sim} number of simulations used to obtain confidence intervals.
- $\mathbf{O}$ quantity appearing in the smoothing parameter update and *edf* definition.
- $p^{(rr')}(t,t')$ transition probabilities referring to time interval (t,t').
- $p^{(\nu,rr')}(t,t')$ the ν^{th} simulated transition probability referring to time interval (t,t'), with $\nu = 1,\dots,\mathcal{M}$.
- $\mathbf{P}(t,t')$ transition probability matrix referring to time interval (t,t').
- $\hat{\mathbf{P}}(t,t')$ estimated transition probability matrix referring to time interval (t,t').
- $q^{(rr')}(t)$ transition intensity at time t.
- $q^{(n_{sim},rr')}$ the n_{sim}^{th} simulated transition intensity (for confidence interval construction).
- $\mathbf{Q}(t)$ transition intensity matrix at time t.
- $\hat{\mathbf{Q}}(t)$ estimated transition intensity matrix at time t.
- $\mathbf{Q}_j(\theta)$ transition intensity matrix at the j^{th} observation of a generic individual.
- r starting state.
- r' arrival state.

- $\mathbb{R}$ real numbers set.

- $s_k^{(rr')}(\tilde{\mathbf{x}}_{k\iota})$ k^{th} smooth for the (r,r') transition intensity.

- $\mathcal{S}$ state space of process.

- $\mathbf{S}^{(rr')}_{\lambda^{(rr')}}$ penalty term for the (r,r') transition intensity.

- $\mathbf{S}_\lambda$ overall penalty term.

- t and t' generic time.

- t_{ij} with $i=1,\dots,N$ and $j=1,\dots,n_i$ is the j^{th} observed time for the i^{th} statistical unit.

- t_j used as shorthand of t_{ij} for the generic statistical unit (i.e. when dropping i for simplicity).

- δt time interval in the definition of the closed form expression of $\mathbf{P}$

- T_{rs} time of the $r \rightarrow r'$ transition

- $T_{rs|u}$ time of the $r \rightarrow r'$ transition conditional on being in state r at time u

- u integration variable when integrating transition intensity.

- u_υ knot for the example in the (cubic regression) smooth of time.

- $\check{\mathbf{U}}_{ww'}$ one of the matrices of the closed form expression of $\frac{\partial^2}{\partial\theta_w\partial\theta_{w'}}\mathbf{P}$.

- $\dot{\mathbf{U}}_w$ one of the matrices of the closed form expression of $\frac{\partial}{\partial\theta_w}\mathbf{P}$.

- $\dot{\mathbf{U}}_{ww'}$ one of the matrices of the closed form expression of $\frac{\partial^2}{\partial\theta_w\partial\theta_{w'}}\mathbf{P}$.

- $\mathbf{V}_\theta$ estimated negative inverse penalized Hessian.

- w and w' indexing for gradient vector and Hessian, with $w,w'=1,\dots,W$.

- W total number of parameters

- $\mathbf{x}_i$ covariate vector (without time).

- $\tilde{\mathbf{x}}_\iota$ overall covariate vector (with time).

- $\tilde{\mathbf{x}}_{k\iota}$ is the k^{th} sub-vector of the overall covariate vector $\mathbf{z}_i$.

- $\tilde{\mathbf{X}}_k^{(rr')}$ the design matrix corresponding to the k^{th} term in the (r,r') transition intensity.
- $\tilde{\mathbf{X}}^{(rr')}$ overall design matrix for the (r,r') transition intensity.
- y indexing of the eigenvalues.
- Y number of eigenvalues.
- z_{ij} with $i=1,\dots,N$ and $j=1,\dots,n_i$ is the j^{th} state occupied by the i^{th} statistical unit.
- $Z(t)$ multi-state process.

Greek letters

- α confidence level.
- $\beta_0^{(rr')}$ intercept parameter for (r,r') transition intensity.
- $\beta_k^{(rr')}$ parameter vector for the k^{th} term in the (r,r') transition intensity. Its length is $J_k^{(rr')}$.
- $\beta^{(rr')}$ parameter vector for (r,r') transition intensity. Its length is $\sum_{k=1}^{K^{(rr')}} J_k^{(rr')}$.
- $\hat{\beta}^{(rr')}$ estimated parameter vector of $\beta^{(rr')}$.
- $\beta^{(n_{sim},rr')}$ the n_{sim}^{th} simulated parameter vector for the (r,r') transition intensity.
- γ_y the y^{th} eigenvalue, with $y=1,\dots,Y$.
- Γ matrix of eigenvalues.
- δt time interval in the definition of the closed form expression of the transition probability matrix (and its derivatives).
- $\Delta^{[a]}$ radius of the trust region at the a^{th} iteration.
- ε quantity appearing in the smoothing parameter update.
- ζ indexing for the series representing the exponential.
- $\eta_\iota^{(rr')}(t_\iota,\mathbf{x}_\iota;\beta^{(rr')})$ additive predictor.
- $\eta^{(rr')}$ overall additive predictor for the (r,r') transition intensity.

- θ overall parameter vector.

- $\hat{\theta}$ estimated overall parameter vector.

- $\theta^{[a]}$ overall parameter vector at the a^{th} iteration of the estimation algorithm.

- ι indexing of the observations when defining the additive predictor. Here $i = 1, \dots, \check{n}$.

- κ indexing for the summations appearing in the proof of the $\partial^2\mathbf{P}$ expression.

- $\lambda_k^{(rr')}$ smoothing parameter for the k^{th} term in the (r, r') transition intensity.

- $\lambda^{(rr')}$ smoothing parameter vector in the (r, r') transition intensity. It's length is $K^{(rr')}$.

- λ overall smoothing parameter vector.

- $\mu_{\mathbf{M}}$ and $\hat{\mu}_{\mathbf{M}}$ quantity appearing in the smoothing parameter update.

- ν indexing for simulated probabilities to compute the bias in the simulation study.

- ρ indexing for the summations appearing in the proof of the $\partial^2\mathbf{P}$ expression.

Bibliography

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer.

Barthel, N., Geerdens, C., Killiches, M., Janssen, P., & Czado, C. (2018). Vine copula based likelihood estimation of dependence patterns in multivariate event time data. *Computational Statistics & Data Analysis*, 117, 109–127.

Brechmann, E. C. & Schepsmeier, U. (2013). Modeling dependence with c- and d-vine copulas: The R package CDVine. *Journal of Statistical Software*, 52(3), 1–27.

Cadar, D., Robitaille, A., Clouston, S., Hofer, S. M., Piccinin, A. M., & Muniz-Terrera, G. (2017). An international evaluation of cognitive reserve and memory changes in early old age in 10 european countries. *Neuroepidemiology*, 48(1-2), 9–20.

Chen, M., Chen, L., Lin, K., & Tong, X. (2014). Analysis of multivariate interval censoring by diabetic retinopathy study. *Communications in Statistics-Simulation and Computation*, 43(7), 1825–1835.

Chen, M., Tong, X., & Sun, J. (2009). A frailty model approach for regression analysis of multivariate current status data. *Statistics in Medicine*, 28(27), 3424–3436.

Clements, M., Liu, X.-R., & Christoffersen, B. (2021). *rstpm2: Smooth Survival Models, Including Generalized Survival Models*. R package version 1.5.1.

Cook, R. & Tolusso, D. (2009). Second-order estimating equations for the analysis of clustered current status data. *Biostatistics*, 10(4), 756–772.

Cook, R. J. & Lawless, J. F. (2014). Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences*, 6(1), 127–161.

Cook, R. J. & Lawless, J. F. (2018). *Multistate models for the analysis of life history data*. CRC Press.

Cook, R. J., Yi, G. Y., Lee, K.-A., & Gladman, D. D. (2004). A conditional markov model for clustered progressive multistate processes under incomplete observation. *Biometrics*, 60(2), 436–443.

Cox, D. R. & Miller, H. D. (1977). *The theory of stochastic processes*, volume 134. CRC press.

Crowther, M. J. & Lambert, P. (2016). MULTISTATE: Stata module to perform multi-state survival analysis. Statistical Software Components, Boston College Department of Economics.

Crowther, M. J. & Lambert, P. C. (2017). Parametric multistate survival models: flexible modelling allowing transition-specific distributions with application to estimating clinically useful measures of effect differences. *Statistics in medicine*, 36(29), 4719–4742.

Datta, S. & Satten, G. A. (2001). Validity of the aalen–johansen estimators of stage occupation probabilities and nelson–aalen estimators of integrated transition hazards for non-markov models. *Statistics & probability letters*, 55(4), 403–411.

De Wreede, L. C., Fiocco, M., & Putter, H. (2010). The mstate package for estimation and prediction in non-and semi-parametric multi-state and competing risks models. *Computer methods and programs in biomedicine*, 99(3), 261–274.

Dettoni, R., Marra, G., & Radice, R. (2020). Generalized link-based additive survival models with informative censoring. *Journal of Computational and Graphical Statistics*, 29(3), 503–512.

Diao, L. & Cook, R. J. (2014). Composite likelihood for joint analysis of multiple multistate processes via copulas. *Biostatistics*, 15(4), 690–705.

Eilers, P. H. & Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, 11(2), 89–121.

Eletti, A., Marra, G., Quaresma, M., Radice, R., & Rubio, F. J. (2022). A unifying framework for flexible excess hazard modelling with applications in cancer epidemiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.

Eletti, A., Marra, G., & Radice, R. (2023a). *flexmsm: A General Framework for Flexible Multi-State Survival Modelling*. R package version 0.1.0.

Eletti, A., Marra, G., & Radice, R. (2023b). A spline-based framework for the flexible modelling of continuously observed multistate survival processes. *Statistical Modelling*, 23(5-6), 495–509.

Emura, T., Sofeu, C. L., & Rondeau, V. (2021). Conditional copula models for correlated survival endpoints: Individual patient data metaanalysis of randomized controlled trials. *Statistical methods in medical research*, 12(30), 2634–2650.

Eryilmaz, S. (2014). Modeling dependence between two multi-state components via copulas. *IEEE Transactions on Reliability*, 63(3), 715–720.

Fauvernier, M., Roche, L., & Remontet, L. (2020). *survPen: Multidimensional Penalized Splines for Survival and Net Survival Models*. R package version 1.5.1.

Filippou, P., Kneib, T., Marra, G., & Radice, R. (2019). A trivariate additive regression model with arbitrary link functions and varying correlation matrix. *Journal of Statistical Planning and Inference*, 199, 236–248.

Fiocco, M., Putter, H., & van Houwelingen, H. C. (2008). Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistics in Medicine*, 27(21), 4340–4358.

Ford, K. J. & Robitaille, A. (2023). How sweet is your love? disentangling the role of marital status and quality on average glycemic levels among adults 50 years and older in the english longitudinal study of ageing. *BMJ Open Diabetes Research and Care*, 11(1), e003080.

Geerdens, C., Acar, E. F., & Janssen, P. (2018). Conditional copula models for right-censored clustered event time data. *Biostatistics*, 19(2), 247–262.

Gorfine, M., Keret, N., Ben Arie, A., Zucker, D., & Hsu, L. (2021). Marginalized frailty-based illness-death model: application to the uk-biobank survival data. *Journal of the American Statistical Association*, 116(535), 1155–1167.

Group, A. (1999). The age-related eye disease study (areds): Design implications. *AREDS report no. 1. Controlled Clinical Trials*, 20(6), 573–600.

Hu, T., Zhou, Q., & Sun, J. (2017). Regression analysis of bivariate current status data under the proportional hazards model. *Canadian Journal of Statistics*, 45(4), 410–424.

Iacobelli, S. & Carstensen, B. (2013). Multiple time scales in multi-state models. *Statistics in medicine*, 32(30), 5315–5327.

Jackson, C. (2019). *msm: Multi-State Markov and Hidden Markov Models in Continuous Time*. R package version 1.6.8.

Jackson, C. (2021). *flexsurv: Flexible Parametric Survival and Multi-State Modelsflexsurv: Flexible Parametric Survival and Multi-State Models*. R package version 2.0.

Jackson, C. H. et al. (2011). Multi-state models for panel data: the msm package for R. *Journal of Statistical Software*, 38(8), 1–29.

Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., & Couto, E. (2003). Multistate markov models for disease progression with classification error. *Journal of the Royal Statistical Society Series D: The Statistician*, 52(2), 193–209.

Jiang, S. & Cook, R. J. (2020). Composite likelihood for aggregate data from clustered multistate processes under intermittent observation. *Communications in Statistics-Theory and Methods*, 49(12), 2913–2930.

Joly, P., Commenges, D., Helmer, C., & Letenneur, L. (2002). A penalized likelihood approach for an illness–death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics*, 3(3), 433–443.

Jones, R. H. (2011). Bayesian information criterion for longitudinal and clustered data. *Statistics in medicine*, 30(25), 3050–3056.

Kalbfleisch, J. D. & Lawless, J. F. (1985). The analysis of panel data under a markov assumption. *Journal of the American Statistical Association*, 80(392), 863–871.

Klein, J. & Moeschberger, M. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Kosorok, M. R. & Chao, W.-H. (1995). *Further Details On The Analysis of Longitudinal Ordinal Response Data in Continuous Time*. Technical Report 92, University of Wisconsin, Madison, Dept. of Biostatistics.

Kosorok, M. R. & Chao, W.-H. (1996). The analysis of longitudinal ordinal response data in continuous time. *Journal of the American Statistical Association*, 91(434), 807–817.

Kwon, S., Ha, I. D., Shih, J.-H., & Emura, T. (2021). Flexible parametric copula modeling approaches for clustered survival data. *Pharmaceutical Statistics*, 21(1), 69 – 88.

Leitenstorfer, F. & Tutz, G. (2006). Generalized monotonic regression based on B-splines with an application to air pollution data. *Biostatistics*, 8(3), 654–673.

Lintu, M., Shreyas, K., & Kamath, A. (2022). A multi-state model for kidney disease progression. *Clinical Epidemiology and Global Health*, 13, 100946.

Liu, X.-R., Pawitan, Y., & Clements, M. (2018). Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, 27(5), 1531–1546.

Lo, S. M. S., Mammen, E., & Wilke, R. A. (2020). A nested copula duration model for competing risks with multiple spells. *Computational Statistics & Data Analysis*, 150, 106986.

Mariano Machado, R. J., Van den Hout, A., & Marra, G. (2021). Penalised maximum likelihood estimation in multi-state models for interval-censored data. *Computational Statistics & Data Analysis*, 153, 107057.

Marra, G. & Radice, R. (2020). Copula link-based additive models for right-censored event time data. *Journal of the American Statistical Association*, 115, 886–895.

Marra, G. & Radice, R. (2024). *GJRM: Generalised Joint Regression Modelling*. R package version 0.2-6.5.

Martins, A., Aerts, M., Hens, N., Wienke, A., & Abrams, S. (2019). Correlated gamma frailty models for bivariate survival time data. *Statistical Methods in Medical Research*, 28(10-11), 3437–3450.

Moll, J. & Wright, V. (1973). Psoriatic arthritis. In *Seminars in arthritis and rheumatism*, volume 3 (pp. 55–78).: Elsevier.

Moré, J. J. & Sorensen, D. C. (1983). Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 4(3), 553–572.

Nießl, A., Allignol, A., Beyersmann, J., & Mueller, C. (2023). Statistical inference for state occupation and transition probabilities in non-markov multi-state models subject to both random left-truncation and right-censoring. *Econometrics and Statistics*, 25, 110–124.

Nocedal, J. & Wright, S. J. (2006). *Numerical Optimization*. Springer-Verlag, New York.

O'Keeffe, A. G., Tom, B. D., & Farewell, V. T. (2011). A case-study in the clinical epidemiology of psoriatic arthritis: multistate models and causal arguments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(5), 675–699.

O'Keeffe, A. G., Su, L., & Farewell, V. T. (2018). Correlated multistate models for multiple processes: an application to renal disease progression in systemic lupus erythematosus. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(4), 841–860.

Petti, D., Eletti, A., Marra, G., & Radice, R. (2022). Copula link-based additive models for bivariate time-to-event outcomes with general censoring scheme. *Computational Statistics & Data Analysis*, 175, 107550.

Putter, H. (2011). Tutorial in biostatistics: Competing risks and multi-state models analyses using the mstate package. *Companion file for the mstate package*.

Putter, H., de Wreede, L. C., & Fiocco, M. (2020). *mstate: Data Preparation, Estimation and Prediction in Multi-State Models*. R package version 0.3.1.

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, 26(11), 2389–2430.

Putter, H. & Spitoni, C. (2018). Non-parametric estimation of transition probabilities in non-markov multi-state models: the landmark aalen–johansen estimator. *Statistical methods in medical research*, 27(7), 2081–2092.

Pya, N. & Wood, S. (2015). Shape constrained additive models. *Statistics and Computing*, 25(3), 543–559.

Pyke, R. (1961). Markov renewal processes: definitions and preliminary properties. *The Annals of Mathematical Statistics*, (pp. 1231–1242).

R Development Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reid, N. (1994). A conversation with sir david cox. *Statistical Science*, 9(3), 439–455.

Romeo, J., Meyer, R., & Gallardo, D. (2018). Bayesian bivariate survival analysis using the power variance function copula. *Lifetime Data Analysis*, 24(2), 355–383.

Ross, S. M., Kelly, J. J., Sullivan, R. J., Perry, W. J., Mercer, D., Davis, R. M., Washburn, T. D., Sager, E. V., Boyce, J. B., & Bristow, V. L. (1996). *Stochastic processes*, volume 2. Wiley New York.

Royston, P. & Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in medicine*, 21(15), 2175–2197.

Sauerbrei, W. & Royston, P. (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1), 71–94.

Sauerbrei, W., Royston, P., & Look, M. (2007). A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biometrical Journal*, 49(3), 453–473.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, (pp. 461–464).

Sujica, A. & Van Keilegom, I. (2018). The copula-graphic estimator in censored nonparametric location-scale regression models. *Econometrics and Statistics*, 7, 89–114.

Sun, T. & Ding, Y. (2021a). Copula-based semiparametric regression method for bivariate data under general interval censoring. *Biostatistics*, 22(2), 315–330.

Sun, T. & Ding, Y. (2021b). *CopulaCenR: Copula-Based Regression Models for Bivariate Censored Data*. R package version 1.1.3.

Swaroop, A., Chew, E., Rickman, C., & Abecasis, G. (2009). Unraveling a multifactorial late-onset disease: from genetic susceptibility to disease mechanisms for age-related macular degeneration. *Annual Review of Genomics and Human Genetics*, 10(1), 19–43.

Titman, A. (2023). *nhm: Non-Homogeneous Markov and Hidden Markov Multistate Models*. R package version 0.1.1.

Titman, A. C. (2008). *Model diagnostics in multi-state models of biological systems*. PhD thesis, University of Cambridge.

Titman, A. C. (2009). Computation of the asymptotic null distribution of goodness-of-fit tests for multi-state models. *Lifetime Data Analysis*, 15(4), 519–533.

Titman, A. C. (2011). Flexible nonhomogeneous markov models for panel observed data. *Biometrics*, 67(3), 780–787.

Touraine, C., Helmer, C., & Joly, P. (2016). Predictions in an illness-death model. *Statistical methods in medical research*, 25(4), 1452–1470.

Van Den Hout, A. (2016). *Multi-state survival models for interval-censored data*. CRC Press.

Van den Hout, A. & Matthews, F. E. (2008). Multi-state analysis of cognitive ability data: a piecewise-constant model and a weibull model. *Statistics in Medicine*, 27(26), 5440–5455.

Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1), 1.

Vatter, T. & Chavez-Demoulin, V. (2015). Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141(C), 147–167.

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society. Series B*, 45(1), 133–150.

Wang, L., Sun, J., & Tong, X. (2008). Efficient estimation for the proportional hazards model with bivariate current status data. *Lifetime Data Analysis*, 14(2), 134–153.

Wang, N., Wang, L., & McMahan, C. (2015). Regression analysis of bivariate current status data under the gamma-frailty proportional hazards model using the em algorithm. *Computational Statistics & Data Analysis*, 83(C), 140–150.

Wen, C. & Chen, Y. (2013). A frailty model approach for regression analysis of bivariate interval-censored survival data. *Statistica Sinica*, 23(1), 383–408.

Williams, J. P., Storlie, C. B., Therneau, T. M., Jr, C. R. J., & Hannig, J. (2020). A bayesian approach to multistate hidden markov models: application to dementia progression. *Journal of the American Statistical Association*, 115(529), 16–31.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99(467), 673–686.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction With R*. Second Edition, Chapman & Hall/CRC, London.

Wood, S. N. & Fasiolo, M. (2017). A generalized fellner-schall method for smoothing parameter optimization with application to tweedie location, scale and shape models. *Biometrics*, 73(4), 1071–1081.

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563.

Yang, Y. & Nair, V. N. (2011). Parametric inference for time-to-failure in multi-state semi-markov models: A comparison of marginal and process approaches. *Canadian Journal of Statistics*, 39(3), 537–555.

Yiu, S., Farewell, V. T., & Tom, B. D. (2017). Exploring the existence of a stayer population with mover–stayer counting process models: application to joint damage in psoriatic arthritis. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 66(4), 669.

Younes, N. & Lachin, J. (1997). Link-based models for survival data with interval and continuous time censoring. *Biometrics*, (pp. 1199–1211).

Zeng, D., Gao, F., & Lin, D. (2017). Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. *Biometrika*, 104(3), 505–525.

Zhou, Q., Hu, T., & Sun, J. (2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112(518), 664–672.