University of Salerno
Department of Economics and Statistics

Ph.D Programme in

ECONOMICS AND POLICIES OF MARKETS AND FIRMS

Curriculum: Statistical Methods

Cycle XXXV

Doctoral Thesis in

COPULA LINK-BASED ADDITIVE MODELS FOR BIVARIATE TIME-TO-EVENT OUTCOMES WITH GENERAL CENSORING SCHEME: COMPUTATIONAL ADVANCES AND VARIABLE RANKING PROCEDURES

Supervisor
Prof.ssa Marcella Niglio

PhD Candidate
Danilo Petti

Co-Supervisor
Prof. Giampiero Marra

Coordinator
Prof.ssa Alessandra Amendola

A.Y. 2022/2023

Contents

Declaration

I certify that the thesis I have presented for the final examination for the PhD in Economics and Policies of markets and firms curriculum: Statistical Methods is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I confirm that chapter 1 was jointly co-authored with Professor Giampiero Marra, Professor Rosalba Radice and Alessia Eletti, while Chapter 2 was jointly co-authored with Professor Marcella Niglio and Professor Giampiero Marra. Chapter 1 has been published in Computational Statisics & Data Analysis *and part of it (section 1.2 and the content of Appendix A, B, E.) was submitted for the MSc in Statistics at University College London. While the full implementation in* `GJRM` *(appeared in April 2021 on* `CRAN`*), the extensive test phase (in Appendix C, D) carried out to assess the goodness of the results, the simulation study and the case study are completely original (published in Petti et al. (2022)) and have been developed thanks to the University of Salerno PhD Scholarship as well as the entirety of Chapter 2, which we plan to submit for publication soon.*

Acknowledgements

First I would like to show my gratitude to my main supervisor Professor Marcella Niglio and my co-supervisor Professor Giampiero Marra for their continuous support and invaluable guidance throughout my PhD study. It was an immense privilege to work with such fine researchers. I am deeply grateful for them having always encouraged me throughout this journey.

Special thanks go to Professor Marialuisa Restaino for encouraging me to apply as a lecturer at the University of Essex. If I got the position it is also thanks to you. I would like to thank all the staff and colleagues in the Department of Economics and Statistics and the Statlab for their continuous support.

I am also very thankful for the financial support of the University of Salerno PhD Scholarschip. Without it, this thesis could not have been undertaken nor completed. I would like to thank my parents for their support throughout my life, and being understandable about my decision to pursue PhD study.

Introduction

This thesis is organized in two intertwined parts whose second represents the natural prosecution of the other. In chapter one we extend a model appeared in Marra & Radice (2020) to then, in the second, propose a variable ranking method based on the same category of model(s).

In the first chapter, we extend the copula link-based time-to-event model originally proposed in Marra & Radice (2020) to a general censoring scheme, which includes mixtures of left-, right-, interval- and uncensored data. Among the features of the proposed model: The possibility to manage different mixtures of censoring; the baseline survival levels are modelled by monotonic P-splines; margins are included via transformations (e.g., proportional hazards, proportional odds); interactions among the covariates can be explored and splines can be applied if needed; a wide range of copula (rotation included) are implemented and ready to use. The algorithm is based on a computationally efficient and stable penalised maximum likelihood estimation approach, whose added valued is the analytical derivation of the gradient and Hessian matrix (expressions in Appendix A). This last aspect dramatically improved the procedure in terms of speed and stability, allowing any user to obtain an estimate in reasonable computational times with quantitative and qualitative outputs. The approach is illustrated via a simulation study, as well as using data from the Age-Related Eye Disease Study (AREDS), a multi-centered randomised clinical trial exploring the development and progression of Age-related Macular Degeneration (AMD), sponsored by the National Eye Institute. Finally, the modelling framework is incorporated into an `R` package named `GJRM`. In the second chapter, a variable ranking procedure based on copula bivariate time-to-event margins under right- and mixed censoring scheme is proposed. The procedure is based on a generalisation of the RBVS algorithm (Baranowski et al., 2020). The authors have extended the algorithm to include two sets of important variables and more importantly by proposing a new metric based on the Fisher information matrix. In brief, the algorithm uses subsampling to identify the sets of covariates which appear at the top of the variable ranking with a non-negligible probability. The method discussed in the second chapter is of particular interest because there is no such variable ranking proposal in the literature for copula bivariate time-to-event models. Furthermore, the method permits the identification of two non-necessarily overlapping sets of important variables for each margin, with an almost zero false positive rate. The potential of the approach is illustrated via a simulation study and an application based on a modified versione of the AREDS dataset. Finally, the algorithm has been made available in a free repository, allowing any practitioner to apply the method with minimal effort.

Chapter 1

Copula link-based Survival additive models extended to a general censoring scheme

Declaration

Part of the following chapter (sections 1.2, 1.2.1, 1.2.3-1.2.8) have been used in a MSc thesis at University College London, UK. Furthermore, part of the following chapter (1.3, 1.4) contains content published in *Copula link-based additive models for bivariate time-to-event outcomes with general censoring scheme, Computational Statistics & Data Analysis, Danilo Petti, Alessia Eletti, Giampiero Marra, Rosalba Radice, Volume 175, 2022, 107550, ISSN 0167-9473, https://doi.org/10.1016/j.csda.2022.107550.*

1.1 Motivation

Age-related Macular Degeneration (AMD) is an eye disease that results in blurred or even loss of vision in the centre of the visual field. Technically, it happens when the small central portion of the retina, called the macula, wears down (Salimiaghdam et al., 2019). AMD is currently the leading cause of blindness in wealthy countries and the third leading cause overall. Despite being characterised by a progressive loss of central vision, usually bilateral, degenerative maculopathy never leads to complete blindness, which is caused by the terminal phases of retinal maculopathy.

Among the numerous causes of blindness, maculopathy occupies a prominent place, and about 8-7% is associated with AMD (41% in wealthy countries). Recent studies confirmed that this pathology is particularly widespread in people older than 60 years. Not to motion that its prevalence is likely to increase as a consequence of exponential population ageing. Although there have been clear steps forward to counter this visual disease (e.g., anti-angiogenesis therapy, intravitreal injections, effective treatments that can prevent blindness), the treatments are often expensive and not available to all patients in many countries (Wong et al., 2014). AMD can be classified in the following variants, i) exudative maculopathy, ii) senile maculopathy, iii) myopia maculopathy, iv) diabetic maculopathy, v) cellophane maculopathy.

Exudative maculopathy can be classified into two distinct types, the dry and wet form. The former has a slow progression (Only 10% of people with AMD develop this form), while the latter (far more widespread) is characterised by a sudden onset of retinal haemorrhage, which causes loss of vision. Senile maculopathy, or macular degeneration related to the patient's age, is the most frequent form of maculopathy, that affects 25 to 30 million people in the western world. Myopia maculopathy is a schisis-like thickening of the retina in eyes with high myopia with posterior staphyloma. The pathologic features may also include lamellar or full-thickness macular holes, shallow foveal detachments, and inner retinal fluid. Estimated to affect between 9 and 34% of highly myopic eyes with posterior staphyloma, it is more prevelant in populations with high myopia and may be more prevalent in women, (Panozzo & Mercanti, 2004, 2007; Baba et al., 2003). Diabetic maculopathy happens when the blood vessels in the part of the eye called the macula become leaky or are blocked. It is usually encountered in older non-insulin-dependent diabetics with mainly nonproliferative diabetic retinopathy (Ivaniševic & Stanic, 1990). Cellophane maculopathy is characterised by fine membranes that grow on the surface of the retina. This membrane can distort and contort retinal structures resulting in vision being blurred and distorted. As these membranes grows they can affect the structure of the retina, resulting in blurred and distorted vision of variable degrees. Patients may also complain of differences in the size of objects between both eyes[1]

AMD, which affects the ocular system, is particularly notorious for having a tendency to man-

[1] source: Kirti M Jasani's blog

ifest itself without symptoms. This aspect not only makes diagnosis very difficult but also makes prevention a huge challenge in biomedical sciences. Usually in patients affected by the disease, blurred vision begins to manifest gradually. It is also fair to say that, when blurred vision occurs, the condition is irreversible in most cases. This brings to light the importance of prevention in the fight against AMD and in the development of statistical predictive tools.

Two significant aspects of the disease motivate the model formulation discussed in section 1.2: i) the lack of symptoms in affected patients; ii) the fact that it occurs mainly in the over 60s. Therefore, we believe that the model proposed in this chapter may actively contribute to a diagnosis of the disease. Typically, datasets on AMD are affected by censoring (e.g., left-censoring, interval-censoring, right-censoring and any combination of them). For this reason, we decided to extend the model presented in Marra & Radice (2020) in such a way to accommodate a more general censoring scheme. We recall that the content of this chapter has been published in *Computational Statistics & Data Analysis* (see, Petti et al., 2022).

1.1.1 Risk Factors

There are numerous factors that over the years have proved to be dominant in the development of AMD. The aim of this section is to discuss the main risk factors useful for predicting AMD, some of which are included in the Age-Related Eye Disease Studies dataset (AREDS), that we used in the analysis presented in section 1.3.2. We refer the reader to Schultz et al. (2021) for a recent literature overview.

Starting from age, Salimiaghdam et al. (2019) and Yonekawa et al. (2015) discussed the role played by age factors in AMD progression. Pooled data from seven population-based studies reveal that the prevalence of geographic atrophy in the US is 0.3% in 60- to 64-year-olds, 0.5% in 65- to 69-year-olds and 0.9% in 70- to 74-year-olds. Moreover, it is almost 7% in people 80 or older.

Smoking seems to play a role in disease progression, especially among women. It affects AMD progression both in active and passive smokers. The decrease in serum antioxidant levels, caused by smoke, affects the macula which is highly sensitive to oxidative stress. Salimiaghdam et al. (2019) suggests that as a result of the oxidative stress generated by smoking, mitochondrial DNA can be damaged, which induces retinal pigment epithelium degradation and contributes to the formation of drusen. Smoking enhances atherosclerosis susceptibility and induces damage to choroidal vessels. The choroid is the vascular layer of the eye, containing connective tissue and lying between the retina and the sclera.

Single nucleotide metamorphisms (SNPs) are genetic variations present in the human genome whose total number exceed 100 million, making them the most common type of genetic variation in the world population. The abundance of SNPs allowed them to become a reference point in studies involving sequence variations to phenotype changes. It has been proved that SNPs can be employed to spot genes associated to a particular disease, with this directly affecting the gene's function. Recent studies have also confirmed their role in the molecular mechanisms of sequence evolution. An organism's response to certain drugs and ability to track the inheritance of disease genes within families can be better understood by studying individual SNPs. The role of SNPs in the study of genetic diseases has had the opportunity to consolidate in recent years.

Studies on AREDS data have demonstrated the negative impacts of genetic factors on AMD progression (e.g., CFH, CFB, HTRA1/ARMS2). Wang et al. (2016) proved that multiple single-nucleotide polymorphisms (SNPs) were sufficient to assess the AMD lifetime risk. This was achieved through a new prediction model based on 25 highly associated SNPs from 15 genes. RJ et al. (2007) discussed the association of certain SNPs to the AMD development in patients, They successfully identified in CFH, BF-C2, PLEKHAI/ARMS2/HtrA1 genetic factors capable of playing a dominant role in AMD prediction. Chen et al. (2006) analysed six different genetic facotors and found differences in the association between the CFH gene and exudative AMD in Chinese from Caucasians and Japanese, detecting SNP rs3753394, rs800292 and rs1329428 associated to a significantly increased risk for exudative AMD.

1.1.2 State of the Art

From a historical point of view, the analysis of survival emerged with the creation of life tables, which date back to the 1600s. Since then, this approach has focused on obtaining more accurate tables. More modern analysis has been made possible by Aalen (1980), who adopted the theory of martingales.

Aalen's work not only allowed for the implementation of censored and truncated cases but also allowed for new methodologies in both Bayesian and Frequentist statistics. In this section, we present state-of-the-art multivariate survival models, which, during recent decades, have focused on three main groups of models. The first model types are founded on the marginal distributions basing their structure on the independence assumptions. Within this domain, joint and conditional distributions cannot be defined. The second category concerns the frailty models, which are mixed-effect models with a latent frailty variable applied to the conditional hazard functions. Finally, the third group refers to those based on copulas. The copula-based models directly connect the two marginal distribution using the so-called copula function.

Regarding the first method, Kim & Xue (2002) considered a marginal approach to model the effects of the covariates on multivariate interval-censoring survival data. To account for the dependence among the survivals, they proposed a robust estimate of the covariance matrix able to account for correlation between events. Wang et al. (2008) proposed a hazard model based on copula function for joint survival function. Furthermore, the score and the Hessian are derived analytically, which makes the implementation efficient. Chen et al. (2007) proposed a proportional odds model in which they specify covariates with multiplicative effects on the odds function. The authors applied the proportional odds model to multivariate interval-censored failure time data. The authors did not consider all possible censoring. Furthermore, the authors adopted the independence assumption, emphasising that the application can be unstable when involving a low sample size and a high percentage of censoring. Chen et al. (2014) proposed a multivariate model when several correlated survival times of interest exist. Moreover, they focused their work only on the interval-censoring case, and the estimation equations were derived using working independence assumption.

As for the second method, Oakes (1982) proposed a frailty model for survival based on Clayton (1978). A frailty model is a random effect model for time variables, where the random effect has a multiplicative effect on the hazard function. The peculiarity of the work concerns the information matrix that is derived explicitly. Furthermore, the author proved that the reparameterization successfully introduces orthogonality between the association parameter and the two-scale parameters. Chen et al. (2009) and Chen et al. (2014) proposed a full-likelihood approach based on the proportional hazards model. Concerning the estimation process, an expectation-maximization (EM) algorithm is used. In the simulation study, the algorithm works well in most situations despite revealing limitations, including hazard rates varying considerably among different types of failure time. Wen & Chen (2013) proposed a semiparametric maximum-likelihood estimation for the gamma frailty Cox model under interval censoring. The authors, extending their previous work, developed a computational algorithm utilising the self-consistency equations and contraction principle, which provides stability and efficiency to the convergence. Wang et al. (2015) proposed a gamma frailty PH model and an EM algorithm for the estimation task. The proposed algorithm is based on a three-stage process, which is easy to implement and stable in terms of convergence. Concerning the simulation study, the authors proved that the proposed method efficiently and quickly estimates the unknown parameters. Zhou et al. (2017) discussed a regression analysis of bivariate interval-censored failure time based on a flexible class of semiparametric transformation models. The use of a sieve maximum-likelihood approach based on Bernstein polynomials makes the model easy to implement. In addition, the authors discussed some theoretical results as asymptotically normal and the efficiency of the estimators. Zhang (2004) investigated the effects of possibly time-dependent covariates on multivariate failure times by considering a broad class of semiparametric transformation models with random effects. They developed an EM algorithm algorithm that is used for both parameter and variance estimation. The simulation study reveals how the algorithm performed well in all situations explored in the paper, encountering no non-convergence with any simulated or empirical data sets.

The third class of models is based on the copula. One of the pioneering works for this class of models is that of Clayton (1978). The copula function allows for modelling the association and the dependence between two survival. This allows for the modelling of the survival marginals and their association in a separate and highly flexible way. Cook & Tolusso (2009) proposed a method of estimating the baseline marginal distributions and association parameters for clustered status based on second-order generalized estimating equations, where the model is based on the assumption that the inspection time is conditionally independent of the underlying failure time. Hu et al. (2017) discussed a regression analysis of bivariate status in interval censoring, and the estimators proposed are proven to be strongly consistent, asymptotic normal and efficient. Kor et al. (2013) used Cox's proportional hazards model to analyse clustered interval-censored data.

The proposed model framework is more general and flexible as: i) any combination of censoring types can be accommodated, ii) we can adopt a variety of copula functions, iii) the model allows the user to specify all the model parameters as functions of flexible covariate effects via the penalized regression spline, iv) the margin of the copula are modelled via transformations of the survival functions. P-splines which are tractable and theoretically advantageous are employed in modelling the baseline survival. Not to mention that the model has been tediously implemented in `R` in the `GJRM` package.

1.2 Methodology

In this section we want to introduce some fundamental elements of the model presented in this work. The model discussed is an extension of the methodological framework presented in Marra & Radice (2020). For this reason the notation as well as the model specification remain consistent with what is discussed there. In this chapter we have extended the Copula link-based time-to-event model to a general censoring scenario, this has been achieved by writing from scratch the likelihood function and by the analytical derivation of the gradient and the hessian matrix. Finally, the model has been implemented in the `R` package `GJRM`.

First we will introduce in section 1.2.1, 1.2.2,and 1.2.3 the concept of survival analysis, censoring and copula function. These are the fundamental building blocks on which the model is built. Then we will discuss the proposed model in section 1.2.4 which has been already implemented in the package `GJRM` freely available in `R`. In 1.2.7 and 1.2.8 we will describe the predictor specification and the estimation procedure.

It is worth mentioning that the procedure is based on the analytical expressions of the gradient and the Hessian of the log-likelihood associated, which have been meticulously derived by the author with the aim of improving not only the instability of numerical integration, but also the computation time of these two quantities (see Appendix B and Appendix C for further discussion about the implementation's computational aspects and some tests that we carried out before to upload the package on CRAN). These quantities, as well as the full details of derivation, are presented in detail in Appendix A.

1.2.1 Survival function

The aim of this section is to briefly discuss the survival analysis framework and introduce some useful notation in order to make the reading of the following work as immediate as possible.

The *survival function*, which is commonly used to represent the probability that the time to event of interest T is not earlier than a specified time t, is traditionally denoted as

$$S(t) = P(T > t) = 1 - F(t) = \int_t^\infty f(s)ds,$$

where T is a continuous random variable which represents the time to event of interest. With regard to the properties, the survival function is non negative and non-increasing, when $t = 0$ we have that $S(0) = 1$. In other terms, at the beginning of the time under consideration, the patient is alive with probability equals to one. Meanwhile, at time $t = \infty$ we have that $S(\infty) = 0$, meaning that as the time goes to infinity the probability that no one survives is one. In mathematical terms, as time goes to infinity, the survival curve goes to 0. The *cumulative distribution function* represents the probability that the event of interest has occurred before time t. This can be denoted as

$$F(t) = P(T \leq t) = 1 - S(t).$$

The *death density function* can be defined as the variation of the cumulative distribution function over a small interval of time. If $F(t)$ is continuous, then the death density function will be given by

$$f(t) = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{d}{dt} F(t) = F'(t),$$

while if $F(t)$ is discrete, we can approximate the slope of the tangent line with a finite difference as follows

$$f(t) = \frac{F(t + \Delta t) - F(t)}{\Delta t},$$

where Δt denotes a small time interval. In survival analysis the *hazard function* and the *cumulative hazard function* are usually the objects of interest. The *hazard rate* then can be defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{1}{S(t)} \cdot \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{f(t)}{S(t)},$$

The hazard rate can be expressed with $h(t)$. It is a non-negative function. It can be proved to be monotonic and can take different shapes. Furthermore, it can be proved that the following relationships hold

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} S(t) \cdot \frac{1}{S(t)} = -\frac{d}{dt} [\log S(t)],$$

with $\log(\cdot)$ we denote the natural logarithm. Another quantity of interest is the *cumulative hazard function*, traditionally denoted as $H(\cdot)$

$$H(t) = \int_0^t h(s) ds,$$

one can simply verify that the relation $S(t) = \exp[-H(t)]$ exists between the cumulative hazard function and the survival function. An interesting characteristic of the survival analysis are the relationships that relate $S(t)$, $H(t)$,and $h(t)$. In fact, having even one of these quantities, we can immediately obtain the others. Using the quantities discussed so far, we can obtain the following relationships

$$
\begin{aligned}
h(t) &= -\frac{S'(t)}{S(t)}, \\
H(t) &= -\log S(t), \\
S(t) &= \exp\left(-H(t)\right).
\end{aligned}
$$

It is evident that the univariate case, although widely used, is completely ineffective in the contexts in which for each patient we want to observe two different quantities, in our case t_1 and t_2. The simplest way to extend the results discussed above would be to use a bivariate probability distribution. The *joint survival function* can be written as

$$
S(t_1, t_2) = P(T_1 \geq t_1, T_2 \geq t_2),
$$

As before, $S(t_1, t_2)$ can be viewed as the probability that both eyes of a given patient are healthy respectively at t_1 and t_2. The marginal survival function are

$$
\begin{aligned}
S(t_1) =& P(T_1 \geq t_1) = S(t_1, -\infty) \\
S(t_2) =& P(T_2 \geq t_2) = S(-\infty, t_2),
\end{aligned}
$$

Similarly to the univariate case, associated to each of these survival functions there is a *cumulative hazard function*

$$
\begin{aligned}
H(t_1) &= -\log S(t_1) \\
H(t_2) &= -\log S(t_2).
\end{aligned}
$$

The *joint hazard function* is

$$h(t_1, t_2) = \lim_{\Delta \mathbf{t}} \frac{P(t_1 < T_1 \leq t_1 + \Delta t, t_2 < T_2 \leq t_2 + \Delta t | T_1 \geq t_1, T_2 \geq t_2)}{\Delta t^2},$$

this quantity can be viewed as the instantaneous risk that in the first eye the symptom occurs at t_1, while at the second one the event of interest occurs at t_2 given that both had no symptoms just before the times t_1 and t_2. We can also define the *conditional hazard functions* as

$$h(t_1 | T_2 = t_2) = \lim_{\Delta t_1} \frac{P(t_1 < T_1 \leq t_1 + \Delta t_1 | T_1 \geq t_1, T_2 = t_2)}{\Delta t_1},$$

which denotes the risk for T_1, given that the other failed at t_2. Similarly

$$h(t_1 | T_2 \geq t_2) = \lim_{\Delta t_1} \frac{P(t_1 < T_1 \leq t_1 + \Delta t_1 | T_1 \geq t_1, T_2 \geq t_2)}{\Delta t_1},$$

denotes the risk for T_1, given that the other survived just before t_2.

1.2.2 Censoring

In the following section, we are going to introduce a very common problem affecting survival analysis, censoring. Readers unfamiliar with the following topic can consult two excellent texts Marubini & Valsecchi (2004) or Kleinbaum & Klein (2004). The following brief introduction is extracted from them.

One important problem in which any researcher incurs when dealing with survival analysis is censoring. In essence, censoring occurs when we have some information about individual survival time, but we don't know the survival time exactly. To better understand the censoring event, we can consider a simple example discussed in Kleinbaum & Klein (2004). Let's consider a leukaemia patient. She is followed until the remission which will be denoted by t_2. If for a given, but unknown, reason, the study ends while the patient is still in remission, meaning that the event is not observed, then the patient's survival time is considered censored. We know that for this person, the survival time is at least as long as the period that the person has been followed, but if the person goes out of remission after the study ends, we do not know the complete survival time.

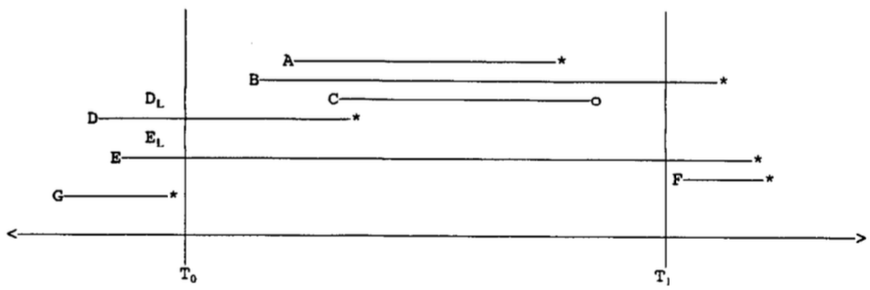


Figure 1.1: Types of point-censored observations. *source: Leung et al. (1997)*

In clinical and epidemiological studies, censoring is mainly caused by a time restriction. In clinical trials on chronic diseases the study continues until a pre-specified time point (cut off date). This means that the time to event is known only on those subjects whose events happened before the cut off date. For the remaining group, it is only known that the time to event is greater then the observation time. What has been just described takes the name of right censoring and the subjects are called withdrawn alive from the study.

In medical and epidemiological studies the censoring mechanism usually becomes even more complicated. Compared to the case described above, it is common to observe subjects unwilling or unable to continue participating in the study. These subjects are usually denoted as lost to follow-up (Marubini & Valsecchi, 2004).

Let us suppose having subjects under observation from T_0 to T_1 and that the censoring time is known exactly (Figure 1.1). The solid line denotes the risk period for each subject, the line ending with an asterisk is an occurrence of the event of interest, and the line ending with an open point is an occurrence of an event other than the event of interest. Subject A falls within the observation period and the time of occurrence of the event is known, no censoring here. For B the risk starts during the observation period and the event occurs after follow-up is terminated at T_1, right censored. For C the observation is right censored as well, but the motivation is different. It is so because an event other than the event of interest occurred during the observation period and takes the subject out of the risk set. Subject D is an example of left-truncation. Such a problem is common in AIDS studies when the patient is already HIV-1 seropositive prior to enrollment and the time variable of interest is the incubation period. As for subject E, the observation is both left and right censored (doubly censored). Finally, in most applications there are cases where the origin and the event both occur prior to the start of follow-up or after follow-up ends. Such cases are represented by subjects F and G and are known as completely right and completely left censored, (Leung et al., 1997). Finally, survival data can also be interval censored. This can occur if a subject's true survival time is within a certain known specified time interval. Interval censoring incorporates both right and left censoring as special cases.

1.2.3 Copula function

In this section we are going to introduce the concept of the copula function. In this section we do not claim to be exhaustive, the purpose is only to discuss the copula and its main properties at an introductory level. This section is extracted from Nelsen (2006).

Intuitively, the copulas are functions that join or "couple" multivariate distribution functions to their one-dimensional marginal distribution. In order to make the discussion coherent and consistent, we introduce some notation. We will let $\mathbb{R}$ denote the ordinary real line and $\bar{\mathbb{R}}$ denote the extended real plane $\bar{\mathbb{R}} \times \bar{\mathbb{R}}$. A rectangle in $\bar{\mathbb{R}}^2$ is the Cartesian product A of two closed intervals: $A = [x_1, x_2] \times [y_1, y_2]$. The vertices of the rectangle A are the points $(x_1, y_1), (x_1, y_2), (x_2, y_1)$, and (x_2, y_2). With $DomH$, we denote a subset of $\bar{\mathbb{R}}^2$ and whose range, denoted with $RanH$, is a subset of $\mathbb{R}$.

Before properly introducing the concept of copula, it is instructive to first discuss the concept of subcopulas. Subcopulas can be defined as a certain class of grounded 2-increasing functions with margins; then subcopulas can be defined as copulas with domain $\mathbf{I^2}$. Where $\mathbf{I^2}$ is the unit square $\mathbf{I} \times \mathbf{I}$, where $\mathbf{I} = [0, 1]$. A two-dimensional subcopula is a function C' with the following properties:

1. $DomC' = S_1 \times S_2$, where S_1 and S_2 are subsets of $\mathbf{I}$ containing 0 and 1;
2. C' is grounded and 2-increasing;
3. For every u in S_1 and every ν in S_2,

$$C'(u, 1) = u \text{ and } C'(1, \nu) = \nu.$$

Note that for every (u, ν) in $DomC'$, $0 \leq C'(u, \nu) \leq 1$, so that $RanC'$ is also a subset of **I**. A two-dimensional copula is a 2-subcopula C whose domain is $\mathbf{I^2}$. Or equivalently the definition just enunciated can be restated as: A copula is a function C from $\mathbf{I}^2$ to **I** with the following properties:

1. For every $u, \nu \in \mathbf{I}$

$$C(u, 0) = 0 = C(0, \nu),$$

   and

$$C(u, 1) = u \text{ and } C(1, \nu) = \nu.$$

2. For every $u_1, u_2, \nu_1, \nu_2 \in \mathbf{I}$ such that $u_1 \leq u_2$ and $\nu_1 \leq \nu_2$,

$$C(u_2, \nu_2) - C(u_2, \nu_1) - C(u_1, \nu_2) + C(u_1, \nu_2) \geq 0.$$

The distinction between copula and subcopula may appear to be a minor one, but it is of fundamental importance in Sklar's theorem, which is one of the milestone theorems concerning copula theory.

We can now provide the definition of n-copula and then state and discuss Sklar's theorem. Let $X_1, \ldots, X_n$ be $n \geq 2$ random variables with $F_n(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$ their joint cumulative distribution function and $F_i(x_i) = P(X_i \leq x_i)$ the marginal cumulative distribution function associated to random variable $i = 1, \ldots, n$. Let $\mathbf{I} = [0, 1]$ be the unit interval and $\mathbf{I}^n = [0, 1]^n$ be the unit n-cube.

An n-dimensional copula is an n-place real function $C : \mathbf{I}^n \rightarrow \mathbf{I}$ which satisfies the following conditions

1. $C(1, \ldots, 1, x_m, 1 \ldots, 1) = x_m \ \forall m = 1, \ldots, n \ \forall x_m \in \mathbf{I}$;

2. $C(x_1, \ldots, x_n) = 0$ if $\exists m = 1, \ldots, n$ s.t. $x_m = 0$;

3. C is n-increasing;

From the definition stated it follows that there is a single unique 1-copula and that it is the identity function on **I**.

An important result reported in Sklar (1973) is what is known as Sklar's theorem. This theorem defines the role that copulas play in the relationship between multivariate distribution functions and their univariate margins.

Theorem 1. *Let $F_n(\cdot)$ be the joint CDF with margins $F_1(\cdot), \ldots, F_n(\cdot)$. There then exists a copula C such that for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$*

$$F_n(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

If $F_i(\cdot)$ for $i = 1, \ldots, n$ are continuous then C is unique, otherwise C is uniquely determined on $Ran(F_1)\times, \ldots, \times Ran(F_n)$, where $Ran()$ represents the range of the function given as argument. Conversely. If C is a copula and $F_1(\cdot), \ldots, F_n(\cdot)$ are $CDFs$, then the function $F_n(\cdot)$ defined above is a joint CDF with margins $F_1(\cdot), \ldots, F_n(\cdot)$.

Finally, we can introduce the survival copulas so as to make clear the connection between the theory of copulas and the model presented in this work. Without loss of generality, we will refer to the 2-copula case, which in the following will be simply referred to as copula. Let $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$ be the joint survival function of the two random variables T_1 and T_2. The margins of $S(t_1, t_2)$ are $S(t_1, -\infty)$ and $S(-\infty, t_2)$ which are the univariate survival functions S_1 and S_2. Using Sklar's theorem we can define a relationship between univariate and join survival functions, after some algebraic manipulation it can be shown that.

$$\begin{aligned} S(t_1, t_2) &= 1 - F_1(t_1) - F_2(t_2) + F(t_1, t_2) \\ &= S_1(t_1) + S_2(t_2) - 1 + C(F_1(t_1), F_2(t_2)) \\ &= S_1(t_1) + S_2(t_2) - 1 + C(1 - S_1(t_1), 1 - S_2(t_2)), \end{aligned}$$

where it has been used the Sklar's theorem for the second to last equality. We can thus define a function.

$$\hat{C}(u, \nu) = u + \nu - 1 + C(1 - u, 1 - \nu),$$

which enables us to rewrite the survival as

$$S(t_1, t_2) = \hat{C}(S_1(t_1), S_2(t_2)),$$

It is immediate to verify that $\hat{C} : \mathbf{I}^2 \to \mathbf{I}$ and that it verifies the point 1-3 of the definition of copula functions given above. It follows that $\hat{C}$ is a copula, which is often referred to as a survival copula. By definition $\hat{C}$ couples the joint survival function to its univariate margins in a manner completely analogous to the way in which a copula connects the joint distribution function to its margins (Nelsen, 2006).

1.2.4 Model formulation

Let us consider the case of bivariate-censored data. For each individual i, let (C_{1i}, C_{2i}) be a vector of bivariate censoring times which by assumption is assumed to be independent of the pair of survival times (T_{1i}, T_{2i}) that are conditioned on a generic $\mathbf{x}_i$, which is the vector of non-informative covariates. We observe $(Y_{1i}, Y_{2i}) = (\min\{T_{1i}, C_{1i}\} \in \mathbb{R}^+, \min\{T_{2i}, C_{2i}\} \in \mathbb{R}^+)$ and the corresponding vector of censoring indicator $\{\delta_{l_{1i}}, \delta_{l_{2i}}\}_{i=1}^n$. Let also $\boldsymbol{\delta} \in \mathbb{R}^W$ be a generic vector of parameters of dimension W, and $i = 1, 2, \ldots, n$ where n represents the sample size.

Let T_{1i} and T_{2i} have conditional marginal survival functions generically defined 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 conditional joint survival function expressed as: $S(t_{1i}, t_{2i})|\mathbf{x}_i; \boldsymbol{\delta}) = P(T_{i1} > t_{1i}, T_{21} > t_{2i}|\mathbf{x}_i; \boldsymbol{\delta})$. We then assume that T_{1i} and T_{2i} are linked by a copula function expressed with the following formulation

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

where $\boldsymbol{\delta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \boldsymbol{\beta}_3^T)$, $\mathbf{x}_{1i}, \mathbf{x}_{2i},$ and $\mathbf{x}_{3i}$ are vectors of covariates (they can all be equal to $\mathbf{x_i}$ but not necessarily). We define the coefficient vectors as $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2,$ and $\boldsymbol{\beta}_3$ with respective dimensions of W_1, W_2 and W_3 such that $W = W_1 + W_2 + W_3$, the copula function can be defined as $C : (0, 1)^2 \to (0, 1)$ is a uniquely defined 2-dimensional copula function with coefficient $\theta_i = m\{\eta_{3i}(\mathbf{x}_{3i}; \boldsymbol{\beta_3})\}$ capturing the (possibly varying) conditional dependence of $(T_{1i}\ T_{2i})$ across observations (see, e.g., Marra & Radice (2017); Patton (2006), Sklar (1973)), $\eta_{3i}(\mathbf{x}_{3i}; \boldsymbol{\beta}_3) \in \mathbb{R}$ is a predictor which includes a generic additive covariate effect, and m is an inverse monotonic differentiable link function which ensures that the dependence parameter lies in a proper range. We can observe this from Table 1.1 at page 27. The table refers to the package `GJRM` in which the presented model has been implemented. The margins are modelled using generalized survival or link-based models (Liu et al., 2018; Royston & Parmar, 2002). That is, $S_\nu(t_{\nu i}|\mathbf{x}_{\nu i}; \boldsymbol{\beta}_\nu)$ is defined as $G_\nu\{\eta_{\nu i}(t_{\nu i}, \mathbf{x}_{\nu i}; \boldsymbol{\beta}_\nu)\}$, where G_ν is an inverse link function and the additive predictor $\eta_{\nu i}(t_{\nu i}, \mathbf{x}_{\nu i}; \boldsymbol{\beta}_\nu) \in \mathbb{R}$, for $\nu = 1, 2$, must include the baseline function of time or, as discussed in Marra & Radice (2020), a stratified set of functions of time, It is evident taking attention to the notation used. The set up used for the additive prediction is discussed in detail in the next section. It may not be immediate to understand the magnitude of the association between T_{1i} and T_{2i} from the knowledge of θ. In such a situation, the Kendall's τ, which takes values in $[-1, 1]$, can be employed. The above construction shows that the copula framework allows us to model jointly the survival function from the knowledge of the marginal survival function and with the help of a function C that binds them together. The model and derived expressions presented in Appendix A have been implemented in `GJRM` with the possibility of using the counter-clockwise rotated versions of coupulae, such as Clayton, Gumbel and Joe. These can be obtained using the following expressions:

$$
\begin{aligned}
C_{90} &= p_2 - C(1 - p_1, p_2) \\
C_{180} &= p_1 + p_2 - 1 + C(1 - p_1, 1 - p_2) \\
C_{270} &= p_1 - C(p_1, 1 - p_2),
\end{aligned}
$$

The subscript denotes the degree of rotation, p_1 and p_2 are margins and θ has been dropped to simplify the notation. More details on copulae and their theoretical foundation can be found in Nelsen (2006). Function $G_\nu\{\eta_{\nu i}(t_{\nu i}, \mathbf{x}_{\nu i}; \boldsymbol{\beta}_\nu)\}$ can e specified as showed in Table 1.2 The marginal cumulative hazard functions H_ν and h_ν for $\nu = 1, 2$, are given by

$$H_\nu(t_{\nu i}|\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu) = -\log[G_\nu\{\eta_{\nu i}(t_{\nu i},\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu)\}],$$

and

$$h_\nu(t|\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu) = -\frac{G'_\nu\{\eta_{\nu i}(t_{\nu i},\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu)\}}{G_\nu\{\eta_{\nu i}(t_{\nu i},\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu)}\frac{\partial\eta_{\nu i}(t_{\nu i},\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu)}{\partial t_{\nu i}}, \tag{1.1}$$

The joint function can be defined in a similar way.

1.2.5 Predictor specification

This section introduces some details on the set up of the three model's predictors. The main difference between $\eta_{\nu i}(t_{\nu i},\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu)$ for $\nu = 1, 2$ and $\eta_{3i}(\mathbf{x}_{3i};\boldsymbol{\beta}_3)$ is that the first two must include smooth functions of time. Apart from that, the design matrix setup is the same across the three additive predictors since $t_{\nu i}$ can be treated as a regressor. Let us consider a generic $\eta_{\nu i}$ ($\nu = 1, 2, 3$), where the dependence on the covariates and the parameters is momentarily dropped, and an overall covariate vector $\mathbf{z}_{\nu \mathbf{i}}$ made up of $\mathbf{x}_{\nu i}$ as well as $t_{\nu i}$ when $\nu = 1, 2$. For simplicity, the dimensions of $\mathbf{z}_{1i}$ and $\mathbf{z}_{2i}$ are assumed to be W_1 and W_2 since t_{1i} and t_{2i} can be treated as covariates.

The indisputable advantage of using additive predictors lies in the fact of being able to use different types of covariate effects. Furthermore, these effects do not need constraints regarding their forms to be applied (Wood, 2006). However, note that the additive assumption here involves that not all the interaction terms among covariates may be included in the predictor (e.g. Hastie & Tibshirani, 1993; Ruppert et al., 2003). An additive predictor can be defined as

$$\eta_{\nu i} = \beta_{\nu 0} + \sum_{k_\nu=1}^{K_\nu} s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i}),\ i = 1,\ldots,n, \tag{1.2}$$

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 coplete vector $\mathbf{z}_{\nu i}$ and the K_ν functions $s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i})$ are the generic effects which are chosen according to the type of covariates 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 $\boldsymbol{\beta}_{\nu k_\nu j_{\nu k_\nu}} \in \mathbb{R}$, that is (Wood, 2017).

Copula	$C(p_1, p_2, \theta)$	Range of θ	Link	Kendall's τ
AMH ("AMH")	$\frac{p_1p_2}{1-\theta(1-p_1)(1-p_2)}$	$\theta \in [-1, 1]$	$\tanh^{-1}(\theta)$	$-\frac{2}{3\theta^2}\{\theta + (1-\theta)^2 \log(1-\theta)\} + 1$
Clayton ("C0")	$(p_1^{-\theta} + p_2^{-\theta} - 1)^{-1/\theta}$	$\theta \in (0, \infty)$	$\log(\theta)$	$\frac{\theta}{\theta+2}$
FGM ("FGM")	$p_1p_2\{1 + \theta(1-p_1)(1-p_2)\}$	$\theta \in [-1, 1]$	$\tanh^{-1}(\theta)$	$\frac{2}{9}\theta$
Frank ("F")	$-\theta^{-1} \log\{1 + (\exp\{-\theta p_1\} - 1)(\exp\{-\theta p_2\} - 1)/(\exp\{-\theta\} - 1)\}$	$\theta \in \mathbb{R}\backslash\{0\}$	-	$1 - \frac{4}{\theta}[1 - D_1(\theta)]$
Gaussian ("N")	$\Phi_2(\Phi^{-1}(p_1), \Phi^{-1}(p_2); \theta)$	$\theta \in [-1, 1]$	$\tanh^{-1}(\theta)$	$\frac{2}{\pi} \arcsin(\theta)$
Gumbel ("G0")	$\exp[-\{(-\log p_1)^\theta + (-\log p_2)^\theta\}^{1/\theta}]$	$\theta \in [1, \infty]$	$\log(\theta - 1)$	$1 - \frac{1}{\theta}$
Joe ("J0")	$1 - \{(1-p_1)^\theta + (1-p_2)^\theta - (1-p_1)^\theta(1-p_2)^\theta\}^{1/\theta}$	$\theta \in (1, \infty)$	$\log(\theta - 1)$	$1 + \frac{4}{\theta^2}D_2(\theta)$
Plackett ("PL")	$(Q - \sqrt{R})/\{2(\theta - 1)\}$	$\theta \in (0, \infty)$	$\log(\theta)$	-
Student-t ("T")	$t_{2,\zeta}(t_\zeta^{-1}(p_1), t_\zeta^{-1}(p_2); \zeta, \theta)$	$\theta \in [-1, 1]$	$\tanh^{-1}(\theta)$	$\frac{2}{\pi} \arcsin(\theta)$

Table 1.1: In this table are defined the different copulae functions implemented in GJRM, with corresponding parameter ranges of association parameter θ, and relations between Kendall's τ and θ. $\Phi_2(\cdot, \cdot; \theta)$ denotes the cumulative distribution function (cdf) of a standard bivariate normal distribution with correlation coefficient θ, and $\Phi(\cdot)$ the cdf of a univariate standard normal distribution. $t_{2,\zeta}(\cdot, \cdot; \zeta, \theta)$ indicates the cdf of a 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 a univariate Student-t distribution with ζ degrees of freedom. $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{\exp(t)-1}dt$ is the Debye function and $D_2(\theta) = \int_0^1 t \log(t)(1-t)^{\frac{2(1-\theta)}{\theta}}dt$. Quantiles Q and R are given by $1 + (\theta - 1)(p_1 + p_2)$ and $Q^2 - 4\theta(\theta - 1)p_1p_2$, respectively. The Kendall's τ for "PL" is computed numerically as no analytical expression is available. Argument `BivD` of `gjrm()` in `GJRM` allows the user to employ the desired copula function and can be set to any of the values within brackets next to the copula names in the first column. For Clayton, Gumbel and Joe, the number after the capital letter indicates the degree of rotation required: the possible values are $0, 90, 180$ and 270.

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")	$-\log(S/(1-S))$	$\frac{\exp -\eta}{1+\exp(-\eta)}$	$-G^2(\eta)\exp(-\eta)$
probit("probit")	-$\Phi^{-1}(S)$	$\Phi(-\eta)$	-$\phi(-\eta)$

Table 1.2: In this table are shown the link functions implemented in `GJRM`. Argument margins of `gjrm()` in `GJRM` allow the user to define the desired marginal models and can be set to any of the values within brackets next to the models' names in the first column. Φ and ϕ are the cumulative distribution and density functions of a univariate standard normal distribution. The first two functions are typically known as log-log and -logit links, respectively.

$$\sum_{j_{\nu k_\nu}=1}^{J_{\nu k_\nu}} \beta_{\nu k_\nu j_{\nu k_\nu}} b_{\nu k_\nu j_{\nu k_\nu}}(\mathbf{z}_{\nu k_\nu i}). \tag{1.3}$$

The above formulation implies that the vector of evaluations $\left\{s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu 1}), \ldots, s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu n})\right\}^T$ can be written as $\mathbf{Z}_{\nu k_\nu}\, \boldsymbol{\beta}_{\nu k_\nu}$ with $\boldsymbol{\beta}_{\nu k_\nu} = (\beta_{\nu k_\nu,1}, \ldots, \beta_{\nu k_\nu J_{\nu k_\nu}})^T$ and the design matrix $\mathbf{Z}_{\nu k_\nu}[i, j_{\nu k_\nu}] = b_{\nu k_\nu j_{\nu k_\nu}}(\mathbf{z}_{\nu k_\nu i})$. This allows the predictor in equation (1.2) to be written as

$$\boldsymbol{\eta}_\nu = \beta_{\nu 0}\mathbf{1}_{\boldsymbol{n}} + \boldsymbol{Z}_{\nu 1}\boldsymbol{\beta}_{\nu 1} + \cdots + \boldsymbol{Z}_{\nu K_\nu}\boldsymbol{\beta}_{\nu k_\nu}, \tag{1.4}$$

where $\mathbf{1}_{\boldsymbol{n}}$ is a n-dimensional vector made up of ones. Equation (1.4) can be also be written in a more compact way as $\boldsymbol{\eta_\nu} = \boldsymbol{Z}_\nu\boldsymbol{\beta}_\nu$, where $\boldsymbol{Z}_\nu = (\mathbf{1}_n, \boldsymbol{Z}_{\nu 1}, \ldots, \boldsymbol{Z}_{\nu K_\nu})$ and $\boldsymbol{\beta}_\nu = (\beta_0, \boldsymbol{\beta}_{\nu 1}^T, \ldots, \boldsymbol{\beta}_{\nu K_\nu}^T)^T$.

Each $\boldsymbol{\beta}_{\nu k}$ is associated to a quadratic penalty $\lambda_{\nu k_\nu}\boldsymbol{\beta}_{\nu k_\nu}^T\boldsymbol{D}_{\nu k_\nu}\boldsymbol{\beta}_{\nu k_\nu}$, whose role is to enforce specific properties of the k_ν^{th} function, such as smoothness. Note that the $\boldsymbol{D}_{\nu k_\nu}$ only depends on the choice of the basis functions. Smoothing parameter $\lambda_{\nu k_\nu} \in [0, \infty)$ controls the trade-off between fit and smoothness, and has an essential role in determining the shape of the estimated smooth function $\hat{s}_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i})$. The overall penalty can be defined as $\boldsymbol{\beta}_\nu^T\boldsymbol{D}_\nu\boldsymbol{\beta}_\nu$, where $\boldsymbol{D}_\nu = diag(0, \lambda_{\nu 1}\boldsymbol{D}_{\nu 1}, \ldots, \lambda_{\nu k_\nu}\boldsymbol{D}_{\nu k_\nu})$. Finally, smooth functions are typically subject to centering constraints (see Wood (2017) for more details).

1.2.6 Penalized likelihood

In the case of bivariate censoring, there are sixteen possible combinations of censoring to deal with in the log-likelihood. Despite the length of the expression, it has an great advantage, its additive nature. Each component can be treated independently from the other, meaning that the model can be adapted to any type of dataset the user has to deal with. 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 $\gamma_{U_{vi}} = \mathbf{1}_{\{T_{vi}=l_{vi}=r_{vi}\}}$ and $\gamma_{I_{vi}} = \mathbf{1}_{\{T_{vi}\in(l_{vi},r_{vi})\}}$ which take a value of one when a give observation is U_{vi} (uncensored) or I_{vi} (interval-censored). Let us assume that a random *i.i.d.* sample $\{(l_{1i}, r_{1i}, l_{2i}, r_{2i}, \gamma_{I_{1i}}, \gamma_{U_{1i}}, \gamma_{I_{2i}}, \gamma_{U_{2i}}, \mathbf{x}_i)\}_{i=1}^{n}$ is available, that there are no competing risks and that censoring is independent and non-informative conditional on $\mathbf{x}_i$. Using a simplified notation to avoid clutter, the log-likelihood function can be written in an elegant way as

$$\ell(\boldsymbol{\delta}) = \gamma_{U_{1i}}\gamma_{U_{2i}} \sum_{i=1}^{n} \log f(t_{1i}, t_{2i}) + \gamma_{I_{1i}}\gamma_{I_{2i}} \sum_{i=1}^{n} \log P(T_{1i} \in (l_{1i}, r_{1i}], T_{2i} \in (l_{2i}, r_{2i}]) \tag{1.5}$$
$$+ \gamma_{U_{1i}}\gamma_{I_{1i}} \sum_{i=1}^{n} \log \left[\int_{l_{21}}^{r_{2i}} f(t_{1i}, y)dy \right] + \gamma_{I_{1i}}\gamma_{U_{1i}} \sum_{i=1}^{n} \log \left[\int_{l_{1i}}^{r_{1i}} f(y, t_{2i})dy \right].$$

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 reader is referred to Appendix A for the more explicit version of the log-likelihood. The likelihood is constructed in such a way as to take into account all possible combinations of censoring. The notation $\eta_{\nu i}$ stands for $\eta_{\nu i}(y_{\nu i}, \mathbf{x}_{\nu i}; \boldsymbol{\beta}_\nu)$. The terms $\{(U_{1i}, U_{2i})$, (U_{1i}, R_{2i}), (R_{1i}, U_{2i}), (U_{1i}, L_{2i}), $(L_{1i}, U_{2i})\}$ involve $\partial\eta_{\nu i}/\partial y_{\nu i}$ $(\nu = 1, 2)$ which can be calculated using $\mathbf{z}'_{\nu i}\boldsymbol{\beta}_\nu$ and must be positive to ensure that the hazard functions are positive. For the complete derivation of the presented likelihood, we refer the reader to Appendix A, where starting from the definition of bivariate probability, each piece is derived by developing all the theoretical steps necessary to arrive at the final result.

To this end, we model the time effects using B-splines with coefficients constrained such that the resulting smooth functions of time are monotonically increasing. Specifically, let $s_\nu(y_{\nu i}) = \sum_{j_\nu=1}^{J_\nu} \gamma_{\nu j_\nu} b_{\nu j_\nu}(y_{\nu i})$ where the $b_{\nu j_\nu}$ are B-spline basis functions of at least second order built over the interval [a,b], based on equally spaced knots, and $\gamma_{\nu j_\nu}$ are spline coefficients. A sufficient condition for $s'_\nu(y_{\nu i}) \geq 0$ over $[a, b]$ is that $\gamma_{\nu j_\nu} \geq \gamma_{\nu j_\nu - 1}$ $\forall j$ (e.g. Leitenstorfer & Tutz, 2007). Such condition can be imposed by re-parametrizing the spline coefficient vector so that $\boldsymbol{\gamma}_\nu = \Sigma_\nu \boldsymbol{\beta}_\nu$, where $\boldsymbol{\beta}_\nu^T = (\beta_{\nu 1}, \ldots, \beta_{\nu j_\nu})$, $\boldsymbol{\beta}_\nu^T = (\beta_{\nu 1}, \exp(\beta_{\nu 2}), \ldots, \exp(\beta_{\nu j_\nu}))$ and $\Sigma_\nu[t_{\nu 1}, t_{\nu 2}] = 0$ if $t_{\nu 1} < t_{\nu 2}$ and $\Sigma_\nu[t_{\nu 1}, t_{\nu 2}] = 1$ if $t_{\nu 1} \geq t_{\nu 2}$, with $t_{\nu 1}$ and $t_{\nu 2}$ denoting the row and the column entries of the respective matrix. When setting up penalty term we penalize the square differences between adjacent $\beta_{\nu j_\nu}$, starting from $\beta_{\nu 2}$, using $\mathbf{D}_\nu = \mathbf{D}_\nu^{*T} \mathbf{D}_\nu^*$ where $\mathbf{D}_\nu^*$ is a $(J_\nu - 2) \times J_\nu$ matrix made up of zeros except that $\mathbf{D}_\nu^*[t_\nu, t_\nu + 1] = -\mathbf{D}_\nu^*[t_\nu, t_\nu + 2] = 1$ for $t_\nu = 1, \ldots, J_\nu - 2$ (Pya & Wood, 2015). Matrix Σ_ν can be absorbed into $\mathbf{Z}_\nu$.

The proposed model allows for a high degree of flexibility in modeling data. If an unpenalized approach is employed to estimate $\boldsymbol{\delta}$ then the resulting smooth function estimates are likely to be unduly wiggly (Ruppert et al., 2003). To prevent overfitting we maximise

$$\ell_p(\boldsymbol{\delta}) = \ell(\boldsymbol{\delta}) - \frac{1}{2} \boldsymbol{\delta}^T \mathbf{S} \boldsymbol{\delta}, \tag{1.6}$$

where ℓ_p is the penalized log-likelihood, $\mathbf{S} = diag(\boldsymbol{D}_1, \boldsymbol{D}_2, \boldsymbol{D}_3)$, $\boldsymbol{D}_i$ $i = 1, 2, 3$ are overall penalties which contain $\lambda_1, \lambda_2, \lambda_3$, and $\lambda_\nu = (\lambda_{\nu 1}, \ldots, \lambda_{\nu K_\nu})^T$. The smoothing parameter vectors can be collected in the overall vector $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T, \boldsymbol{\lambda}_3^T)^T$.

1.2.7 Estimation δ

As it can bee seen from the formulation of the likelihood shown above, it is made up of sixteen main components. This because of the right, left and interval-censoring. This makes the structure of the score vector and Hessian matrix more involved and lengthy to compute analytically compared to the case of right-censoring only.

For this work, the entire analytical formulations were derived and implemented in the package GJRM. However, the time spent in the derivation and for the implementation of these quantities paid off. In fact, as can be appreciated in Appendix B, this procedure has made it possible to considerably reduce the time required for fitting, respective to the numerical integration.

The structure of the gradient and the Hessian matrix is made even more complicated by the non-linear dependence of γ_ν on the coefficients contained in β_ν that correspond to the B-spline bases of $y_{\nu i}$. This particular aspect creates the need to take into account terms like $\frac{\partial^2 \eta_{\nu i}(y_{\nu i}, \mathbf{x}_{\nu i}; \boldsymbol{\beta}_\nu)}{\partial y_{\nu i} \partial \boldsymbol{\beta}_\nu} = \mathbf{z}'^T_{\nu i} \mathbf{E}_\nu$ and $\frac{\partial \eta_{\nu i}(y_{\nu i}, \mathbf{x}_{\nu i}; \boldsymbol{\beta}_{\nu i})}{\partial \boldsymbol{\beta}_\nu} = \mathbf{z}^T_{\nu i} \mathbf{E}_\nu$, where $\mathbf{E}_\nu$ is a vector such that $\mathbf{E}_\nu[\nu k_\nu j_{\nu k_\nu}] = 1$ is $\tilde{\beta}_{\nu k_\nu j_{\nu k_\nu}} = \beta_{\nu k_\nu j_{\nu k_\nu}}$ and $\exp\left(\beta_{\nu k_\nu j_{\nu k_\nu}}\right)$ otherwise. Furthermore, the non-linear dependence of γ_ν on $\boldsymbol{\beta_\nu}$ makes the optimisation problem more difficult than in the case of unconstrained B-spline coefficients.

Preliminary experimentation revealed that the use of various optimisation schemes (Marra & Radice, 2020), including those based on derivative free and quasi-Newton methods, is generally problematic, even with more simple model specifications. Marra & Radice (2020) found that several gradient and Hessian components are poorly approximated by the numerical differentiation techniques. To make the fitting problem easier to deal with, we also experimented with a two-stage estimation approach as often seen in several copula contexts. In this case, the estimation of the marginal models and of the copula function is carried out in two separate steps; the use of a two-stage algorithm resulted in inefficient and (on occasion) unstable computations as compared to the joint approach. Eventually, we opted for a simultaneous estimation approach based on fully analytical first and second order derivatives. In practice, this was implemented using a trust region algorithm which was found to be efficient and well suited for the problem at hand. For a more detailed discussion of the topic, we refer the reader to Marra & Radice (2020), where Supplementary Material C reports some simulation-based evidence of the method's performance.

Holding $\boldsymbol{\lambda}$ fixed at a vector of values and for a given $\boldsymbol{\delta}^{[a]}$, where a is an iteration index, we maximise the equation (1.6) using

$$\boldsymbol{\delta}^{[a+1]} = \boldsymbol{\delta}^{[a]} + \underset{\mathbf{e}:||\mathbf{e}|| \leq \Delta^{[a]}}{\arg\min}\ \breve{\ell}_p(\boldsymbol{\delta}^{[\boldsymbol{a}]}), \tag{1.7}$$

where $\breve{\ell}_p^{[a]} = -\{\ell_p(\boldsymbol{\delta}^{[a]}) + \mathbf{e}^T \mathbf{g}_p(\boldsymbol{\delta}^{[a]}) + \frac{1}{2}\mathbf{e}^T \mathbf{H}_p(\delta^{[a]})\mathbf{e}\}$, $\mathbf{g}_p(\boldsymbol{\delta}^{[a]}) = \boldsymbol{g}(\boldsymbol{\delta}^{[a]}) - \mathbf{S}\boldsymbol{\delta}^{[\boldsymbol{a}]}$ and $\boldsymbol{H}_p(\boldsymbol{\delta}^{[a]}) = \mathbf{H}(\boldsymbol{\delta}^{[a]}) - \mathbf{S}$. Vector $\mathbf{g}(\boldsymbol{\delta}^{[a]})$ consists of $\mathbf{g}_1(\boldsymbol{\delta}^{[a]}) = \left.\frac{\partial \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_1}\right|_{\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{[a]}}, \ldots, \mathbf{g}_3(\boldsymbol{\delta}^{[a]}) = \left.\frac{\partial \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_3}\right|_{\boldsymbol{\beta}_3 = \boldsymbol{\beta}_3^{[a]}}$, the Hessian matrix has elements $\mathbf{H}(\boldsymbol{\delta}^{[a]})_{o,h} = \left.\frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_o \partial \boldsymbol{\beta}_h^T}\right|_{\boldsymbol{\beta}_o = \boldsymbol{\beta}_o^{[a]}, \boldsymbol{\beta}_h = \boldsymbol{\beta}_h^{[a]}}$ where $o, h = 1, 2, 3$, $||.||$ denotes the Euclidean norm, and $\Delta^{[a]}$ is the radius of the trust region which is adjusted through the iterations. The first line of (1.7) uses a quadratic approximation of $-\ell_p$ about $\boldsymbol{\delta}^{[a]}$ (the so-called model function) in order to choose the best $\mathbf{e}^{[a+1]}$ within the ball centered in $\boldsymbol{\delta}^{[a]}$ of a radius $\Delta^{[a]}$, the trust region. Note that, near the solution, the region method typically behaves as a classic Newton- Raphson unconstrained algorithm (e.g. Nocedal & Wright, 2006). The expressions of $\mathbf{g}(\boldsymbol{\delta})$ and $\mathbf{H}(\boldsymbol{\delta})$ are very tedious (due to censoring and the non-linear dependence of $\boldsymbol{\gamma}_\nu$ on $\boldsymbol{\beta}_\nu$) and have been analytically and modularity derived for all choices reported in Tables 1.1 and 1.2. Modularity here means that it is easy to extend our algorithm to other parametric copulae and marginal link functions.

1.2.8 Estimation of λ

As argued in Marra & Radice (2017), automatic multiple smoothing parameter estimation in the context of complex joint models is more successfully achieved if the smoothing criterion is based on $\mathbf{g}(\boldsymbol{\delta})$ and $\mathbf{H}(\boldsymbol{\delta})$.

For notation convenience, let us denote $\mathbf{g}_p^{[a]}$, $\mathbf{g}^{[a]}$, $\mathbf{H}_p^{[a]}$ and $\mathbf{H}^{[a]}$ the shorthand notations for $\mathbf{g}_p(\boldsymbol{\delta}^{[a]})$, $\mathbf{g}(\boldsymbol{\delta}^{[a]})$, $\mathbf{H}_p(\boldsymbol{\delta}^{[a]})$ and $\mathbf{H}(\boldsymbol{\delta}^{[a]})$ defined in the previous section. Let us first express the parameter estimator in terms of gradient and Hessian. The procedure is as follows. A first order Taylor expansion of $\mathbf{g}_p^{[a+1]}$ about $\boldsymbol{\delta}^{[a]}$ yields $\mathbf{0} = \mathbf{g}_p^{[a]} \approx \mathbf{g}_p^{[a]} + (\boldsymbol{\delta}^{[a+1]} - \boldsymbol{\delta}^{[a]})\mathbf{H}_p^{[a]}$. We then have $\mathbf{0} = \mathbf{g}_p^{[a]} + (\boldsymbol{\delta}^{[a+1]} - \boldsymbol{\delta}^{[a]})(\mathbf{H}^{[a]} - \mathbf{S})$ which leads to $\boldsymbol{\delta}^{[a+1]} = (-\mathbf{H}^{[a]} + \mathbf{S})^{-1}\sqrt{-\mathbf{H}^{[a]}}\mathbf{M}^{[a]}$, where $\mathbf{M}^{[a]} = \boldsymbol{\mu}_{\mathbf{M}}^{[a]} + \boldsymbol{\epsilon}^{[a]}$, $\boldsymbol{\mu}_{\boldsymbol{M}}^{[a]} = \sqrt{-\mathbf{H}^{[a]}}\boldsymbol{\delta}^{[a]}$ and $\boldsymbol{\epsilon}^{[a]} = \sqrt{-\mathbf{H}^{[a]}}^{-1}\mathbf{g}^{[a]}$, The square root of $-\boldsymbol{H}^{[a]}$ and its inverse are obtained by the eigen-value decomposition. From likelihood theory, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I})$ and $\mathbf{M} \sim N(\boldsymbol{\mu}_{\boldsymbol{M}}, \mathbf{I})$ where $\mathbf{I}$ is an identity matrix, $\boldsymbol{\mu}_{\boldsymbol{M}} = \sqrt{-\mathbf{H}}\boldsymbol{\delta}^0$ and $\boldsymbol{\delta}^0$ is the true parameter vector.

The predicted value vector for M is $\hat{\boldsymbol{\mu}}_{\mathrm{M}} = \sqrt{-\mathbf{H}}\hat{\boldsymbol{\delta}} = \mathbf{AM}$, where $\mathbf{A} = \sqrt{-\mathbf{H}}(-\mathbf{H}+\mathbf{S})^{-1}\sqrt{-\mathbf{H}}$. The main aim here is to estimate $\boldsymbol{\lambda}$ so that the smooth terms' complexity which is not supported by the data is suppressed. Therefore, we use the following criterion.

$$\mathbb{E}\bigg(||\boldsymbol{\mu}_{\boldsymbol{M}} - \hat{\boldsymbol{\mu}}_{\boldsymbol{M}}||^2\bigg) = \mathbb{E}\bigg(||\boldsymbol{M} - \boldsymbol{AM}||^2\bigg) - \breve{n} + 2\mathrm{tr}\left(\boldsymbol{A}\right), \tag{1.8}$$

where $\breve{n} = 3n$ and $\mathrm{tr}\left(\boldsymbol{A}\right)$ is the number of effective degrees of freedom of the penalized model. In practice, $\boldsymbol{\lambda}$ is estimated by minimising an estimate of (1.8)

$$||\widehat{\boldsymbol{\mu}_{\mathrm{M}} - \hat{\boldsymbol{\mu}}_{\mathrm{M}}}||^2 = ||\mathbf{M} - \mathbf{AM}||^2 - \breve{n} + 2\mathrm{tr}\left(\mathbf{A}\right). \tag{1.9}$$

The RHS of (1.9) depends on $\boldsymbol{\lambda}$ through $\mathbf{A}$ while M is associated with the un-penalized part of the model. Note that (1.9) is approximately equivalent to the Akaike information criterion (AIC, Akaike (1973)), as shown at the end of this section. This means that $\boldsymbol{\lambda}$ is estimated by minimising what is effectively the AIC with the number of parameters given by $\mathrm{tr}\left(\mathbf{A}\right)$. Holding the model's parameter vector value fixed at $\boldsymbol{\delta}^{[a+1]}$, we solve problem

$$\boldsymbol{\lambda}^{[a+1]} = \arg\min_{\boldsymbol{\lambda}} ||\mathbf{M}^{[a+1]} - \mathbf{A}^{[a+1]}\mathbf{M}^{[a+1]}||^2 - \breve{n} + 2\mathrm{tr}\left((\mathbf{A}^{[a+1]})\right), \tag{1.10}$$

using the automatic stable and efficient computation routine by Wood (2003). This approach is based on Newton's method and can evaluate in an efficient and stable way the components in (1.10) and their first and second derivatives respect to $\log(\boldsymbol{\lambda})$ (since the smoothing parameters can only take positive values).

The methods of estimating $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ are iterated until the algorithm satisfies the criterion $\frac{|\ell(\boldsymbol{\delta}^{[a+1]})-\ell(\boldsymbol{\delta}^{[a]})|}{0.1+\ell(\boldsymbol{\delta}^{[a+1]})} \leq$ 1e-07. The selection of starting values is crucial, as in the majority of optimisation problems.

In this case, values for the marginal models are obtained by employing the `gamlss()` function within `GJRM`. An initial value for the copula parameter is obtained by using a transformation of the empirical Kendall's association between the responses.

1.3 Simulation study and Case study

1.3.1 Simulation Study

In this section, we are going to discuss in detail the main findings of the simulation study carried out on the model structure proposed. The margins in this section will be denoted as $T_{\nu i}$, $\nu = 1, 2$. The first margin T_{1i} was generated from a proportional hazards (PH) model defined as $T_{1i} = \log[-\log S_{10}(t_{1i})] + \beta_{11} z_{1i} + s_{11}(z_{2i})$ where $S_{10i}(t_{1i}) = 0.9 \exp(-0.4 t_{1i}^{2.5}) + 0.1 \exp(-0.1 t_{1i})$. In the formulation expressed above we can note that for the covariate z_{2i} has been applied a smooth function. Time T_{2i} was generated from a proportional odds (PO) model defined as $T_{2i} = \log\left[\frac{\{1-S_{20}(t_{2i})\}}{S_{20}(t_{2i})}\right] + \beta_{21} z_{1i} + \beta_{22} z_{3i}$ where $S_{20}(t_{2i}) = S_{10}(t_{2i}) = 0.9 \exp(-0.4 t_{1i}^{2.5}) + 0.1 \exp(-0.1 t_{1i})$. The random censoring were obtained through uniform distributions so that censoring rates were about 42% and 33% for the first group of simulation and 75% and 50% for the second one. Observations were generated using the Brent's univariate-finding root. The two survival times were joined using a Clayton copula C0 where the predictor for the dependence parameter was specified as $\eta_{3i} = \beta_{31} z_{1i} + s_{31}(z_{2i})$. The specification of η_{3i} allowed the dependence to vary across the observations. In practice this was achieved using the conditional sampling approach. The setup 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 $s_{11}(z_i) = \sin(2\pi z_i)$, $s_{31} = 3\sin(\pi z_i)$. The parameters were defined as $\beta_{11} = -1.5$, $\beta_{21} = \beta_{22} = 1.2$, $\beta_{31} = -1.5$. The correlation structure among the covariates was generated using multivariate normal distribution with a correlation parameter $\rho = 0.5$, and then transformed using the distribution function of a standard normal distribution. As concern the covariate z_{1i}, this was rounded to obtain a dichotomous random variable. 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 1.3 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%.

The sample sizes were set up at $1000, 1500, 2000$. An example of a simulation study with $n = 300$ is discussed in Petti et al. (2022) Supplementary Material D. While the number of replicates to 1000. The models were fitted using `gjrm()` in `GJRM` using all the marginal links and copulae presented in Tables 1 and 2. The smooth components of the covariates were represented using penalized low rank thin plane splines with second order penalty and 10 basis function, and smooths of times using monotonic penalized B-splines with penalty and 10 basis. For each replicate, curve estimates were constructed using 200 equally spaced fixed values in the (0,8) 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 1.3: 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. Source: Petti et al. (2022)

The main findings of the simulation study are:

- *Parametric effects:* Figures 1.2 and 1.3 show that the bias and the variability computed across different sample sizes is very low for all the parameters estimated. Both the bias and the variability seem to decrease as the sample size increase. The $\beta_{31}(z_{1i})$ (the effect of z_{1i} contained in the additive predictor of the copula parameter) under the same sample size resulted to be the more variable. In (Marra & Radice, 2020) it is discussed how the profile likelihood tends to be less sharp around the optimum. This difficulty in estimating the dependence parameter is discussed in (Romeo et al., 2018) and reference therein who found the same difficulty in the estimation procedure.

- *Smooth effects:* Figures 1.4 and 1.5 with Tables 1.4 and 1.5 show that the true smooth functions are recovered well by the estimation method employed. Moreover, the results, in terms of bias and RMSE, improve as the sample size increases. The estimation of $s_{31}(z_{1i})$ is more challenging, doubling the sample size we observe a decrease in RMSE of 91% (mild censoring scenario) and 99% (high censoring scenario)

- *Impact of censoring rates:* From Figures 1.2 and 1.3 we can appreciate the impact of censoring on the estimation performance. The impact of high censoring affects mainly the estimation variability. The most effected are the copula's additive predictor. These results can be explained by the fact that as the censoring rates increase less and less observations contribute to the corresponding piece of the likelihood. As for β_{31}, passing from $n = 1000$ to $n = 2000$ we observe a decrease in terms of RMSE of 50% (mild censoring) and 99% (high censoring). As for the convergence rate, in all the simulations carried out it has always been greater than 90%.

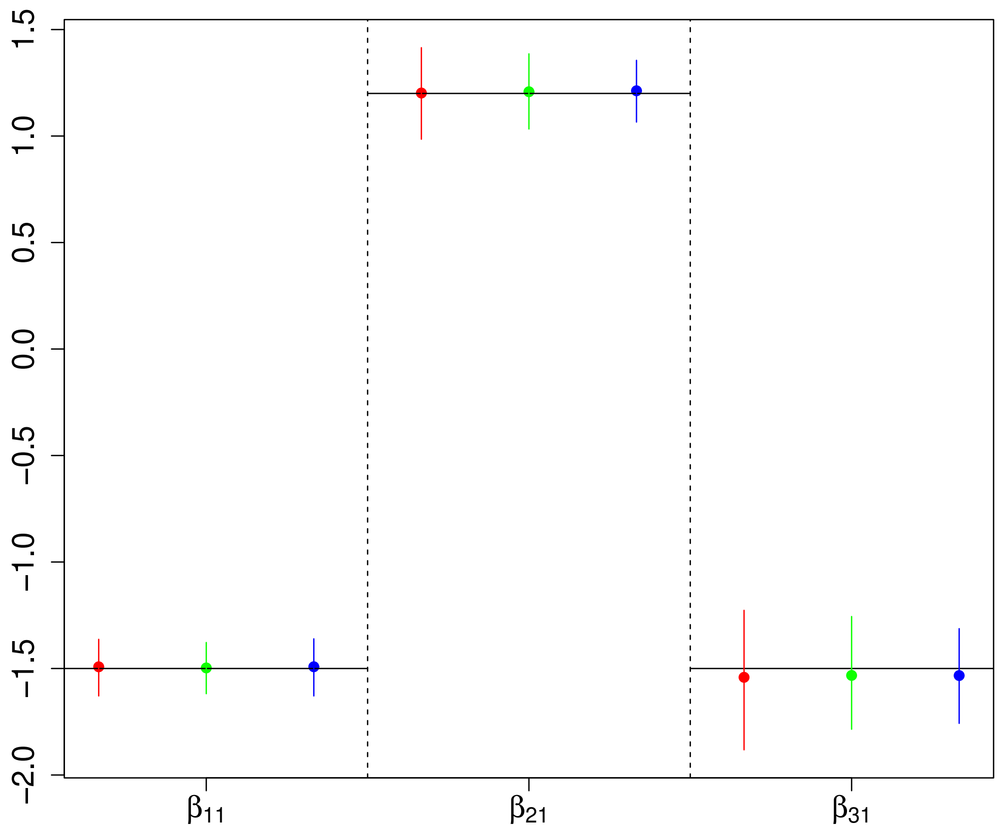


Figure 1.2: Linear coefficient estimates obtained by applying gjrm to bivariate survival data with mild censoring rates (about 42% and 33% for the two responses). Circles indicate mean estimate, bars are the estimates' range (5%-95% quantiles). True values are represented by black solid lines. Black circles and vertical bars refer to the results obtained for $n = 1000$, in dark grey and light grey are represented respectively the results for $n = 1500$ and $n = 3000$. Source: Petti et al. (2022)

	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 1.4: Bias and root mean squared error (RMSE) obtained by fitting with gjrm to bivariate survival data with mild censoring rates (about 42% and 33% for the two responses). 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 sized 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. Source: Petti et al. (2022)

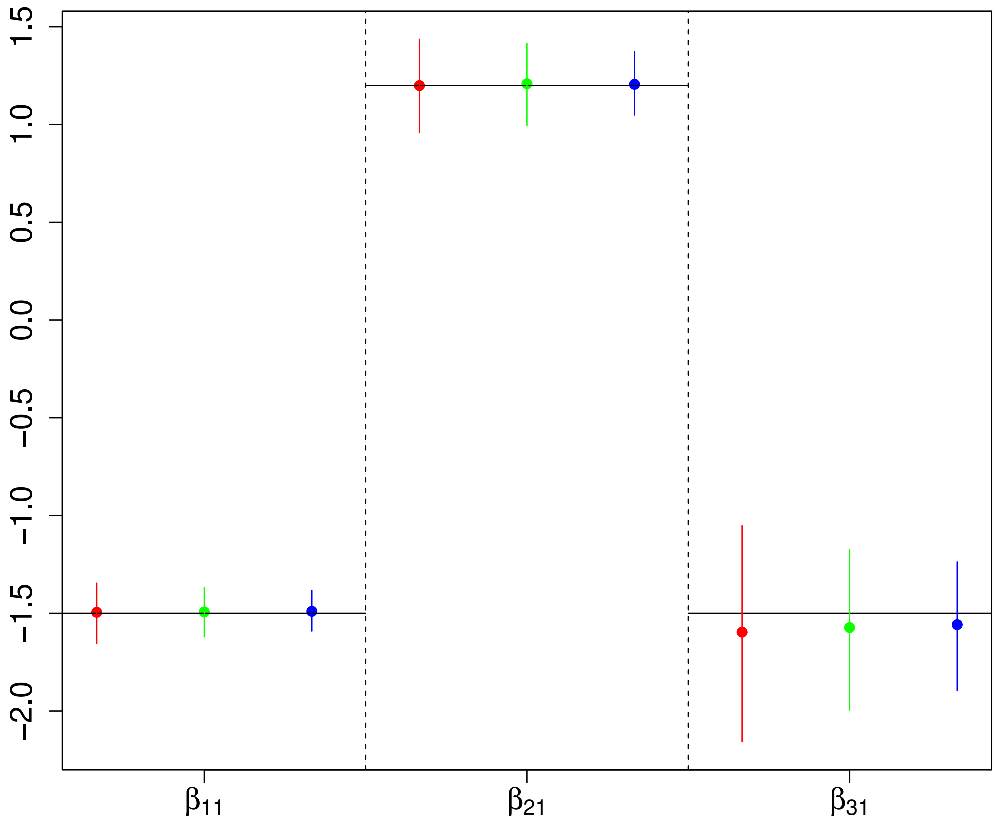


Figure 1.3: Linear coefficient estimates obtained by applying `gjrm` to bivariate survival data with high censoring rates (about 75% and 50% for the two responses). Further details are given in the caption of Figure 1.2. Source: Petti et al. (2022)

	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 1.5: Bias and root mean squared error (RMSE) obtained by fitting with `gjrm` to bivariate survival data with high censoring rates (about 75% and 50% for the two responses). Further details are given in the caption of Table 1.4. Source: Petti et al. (2022)

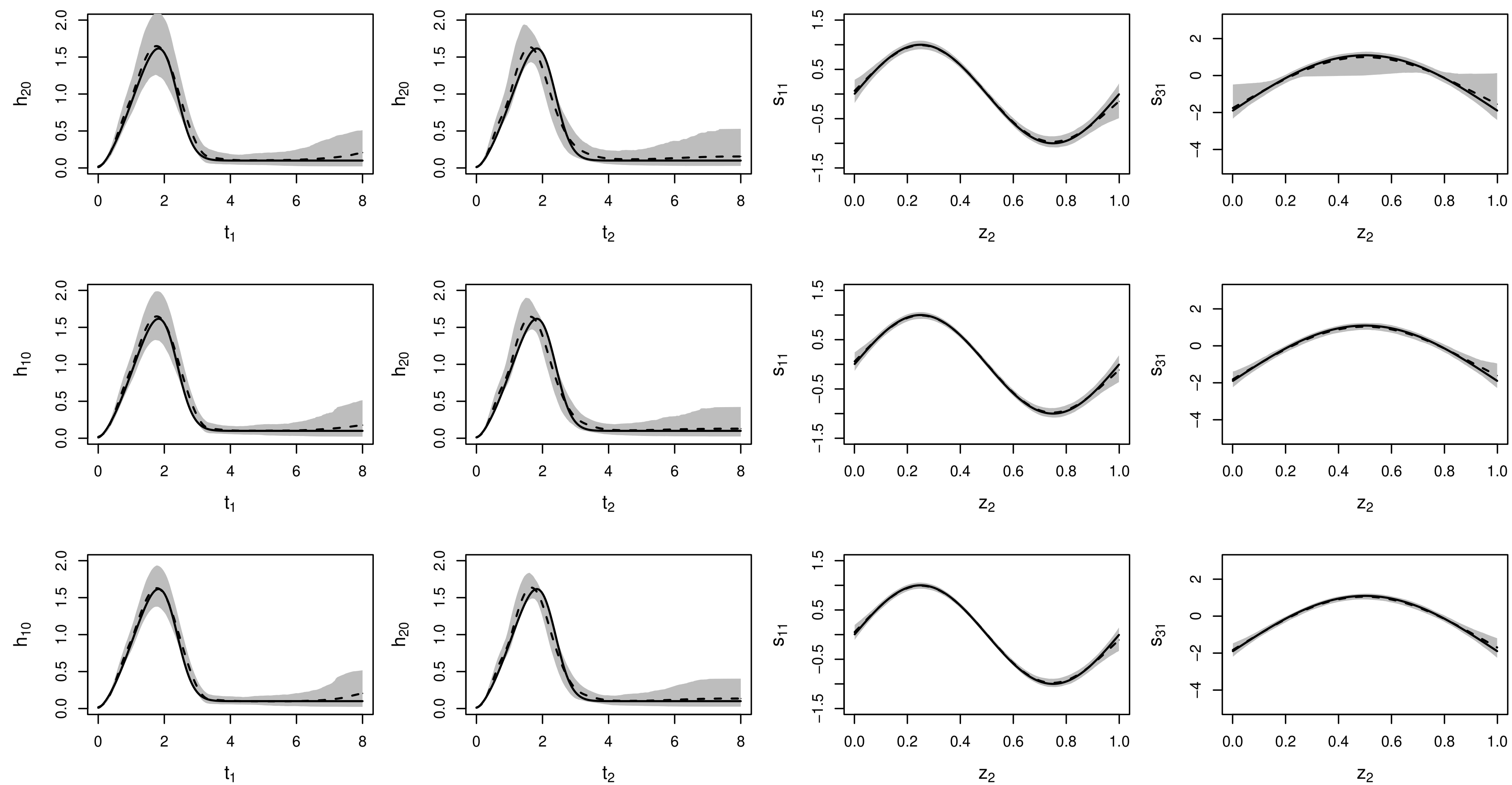


Figure 1.4: Smooth function estimates obtained by applying `gjrm` to bivariate survival simulated data with mild censoring (about 42% and 33% for the two responses). True functions are represented by black solid lines, mean estimates by dashed lines and pointwise ranges resulting from 5% and 95% quantities 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 $n = 3000$. Source: Petti et al. (2022)

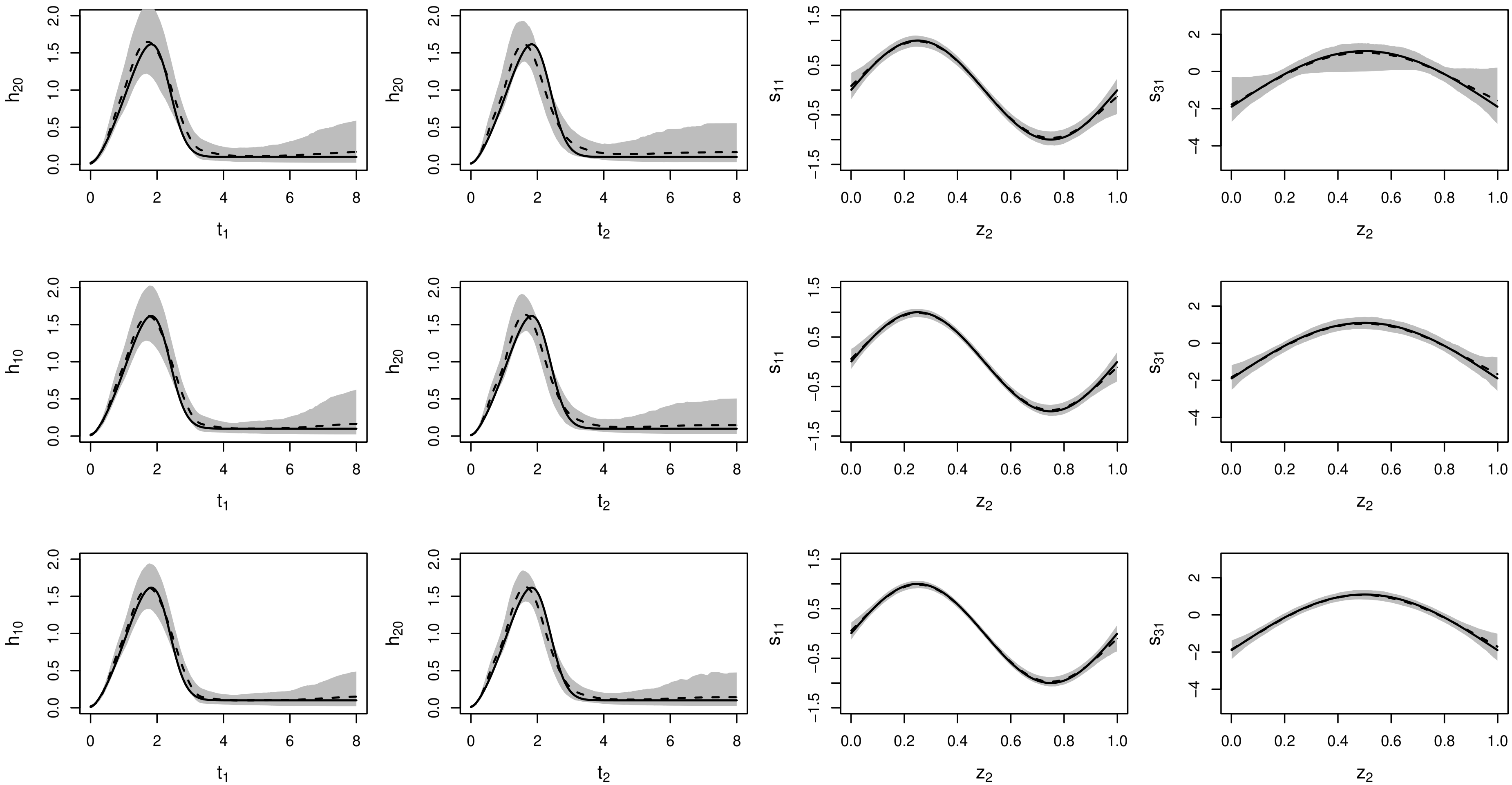


Figure 1.5: Smooth function estimates obtained by applying `gjrm` to bivariate survival simulated data with high censoring (about 72% and 50% for the two responses). Further details are given in the caption of Figure 1.4. Source Petti et al. (2022)

1.3.2 Application to the AREDS dataset

Exploratory Data Analysis

The AREDS dataset is formed by participants aged between 55 to 80 years old, with 650 observations. To be included in this study, participants were checked to be free from any pathology, disease or precarious health condition. An important feature of the AREDS study is that only patients who, at the time of inclusion in the study program, were using supplements are included. About 5% of the participants were taking a multivitamin with about half of them taking the recommended daily allowance (RDA). The RDAs are nothing more than guidelines issued by the Food and Nutrition Board. They define the essential levels of micronutrients that must be taken by a person to be healthy. The data discussed in this section are a sample of the full dataset, available in the package `CopulaCenR`.

The data used are affected by interval, right and mixed censoring (e.g., right eye interval-censored and left eye right-censored), the covariates are: `Age` (enrollment age), `SNP` (genetic factors highly associated with late AMD progression, coded as 0, 1, 2), `SScore1` (the severity score for the first eye. This is a measure of the progression of the disease and is scaled from 1 to 12. This was determined for each eye of the patients at every examination. The higher the score, the higher the progress of the disease), `SScore2` (severity score for the second eye) , `time` (two times for each eye namely, t_{11}, t_{12}, t_{21}, t_{22}).

A preliminary Exploratory Data Analysis (EDA), revealed that in 43% of the patients both eyes are interval-censored, in 35% both eyes are right-censored, in 10% the first eye is interval while the other is right-censored, the remaining 12% has the first eye right and the second interval-censored. Concerning the `SNP`, it takes value 1 in 47% of the cases, value 2 in 15% of the patients, and 38% assumes value 0. It seems that the two eyes do not progress perfectly synchronously. Not only, one eye, on average, is in a more advanced state of AMD progression on average than the other one, but also the eye with higher progression also presents greater variability. Finally, it is interesting to note that in 22% of cases, the two eyes are affected by different types of censoring. This increases the value from a scientific point of view of the model framework proposed in this chapter. Furthermore, there appears to be no interaction between the severity scores (`Sscore1`, Sscore2) and the SNP variable. The conditional distributions seem to overlap (density plot not shown), excluding a potential interaction between the two variables. Finally, regarding the severity score's distribution, it seems that one of the eyes is in a more advanced progression than the other.

Analysis

The proposed approach is applied to a dataset from the AREDS available through the R package CopulaCenR. Late AMD, the most common cause of blindness, is the fulcrum of the analysis presented in this section. Due to the intermittent assessment times (every 6 months up to the first 6 years and every 1 year thereafter) these data are affected by censoring. Furthermore, as discussed in the EDA, the two eyes do not have a simultaneous progression. Censoring combined with the intuition that the event is determined differently in the two eyes motivates the use of the model framework presented above. The dataset contains three covariates, whose impact on AMD has been discussed in section 1.1.1: 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 1.2.6), 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 1.2 were considered in the modelling. 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 Appendix H. The model specification used for this fit is

$$T_{1i} = \log(1 - S(t_{1i})/S(t_{1i})) + s(\texttt{ENROLLAGE}) + \beta_{11}\texttt{SEvScale1E}_{2i} + \beta_{12}\texttt{rs2284665}_{3i}$$
$$T_{2i} = \log(1 - S(t_{2i})/S(t_{2i})) + \beta_{21}\texttt{ENROLLAGE}_{1i} + \beta_{22}\texttt{SEvScale2E}_{2i} + \beta_{23}\texttt{rs2284665}_{3i}$$

All coefficients in the two marginal equations as well as the dependence parameter are significant. 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 the TT genotype group have the highest risk of developing the disease, followed by participants with the GT genotype group.

Figure 1.6 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.

The estimated Kendall's τ is 0.36 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 Appendix H).

Using the chosen model, we produced joint survival functions under several scenarios. The left panel of Figure 1.7 displays the joint progression-free probability contours for subjects who are

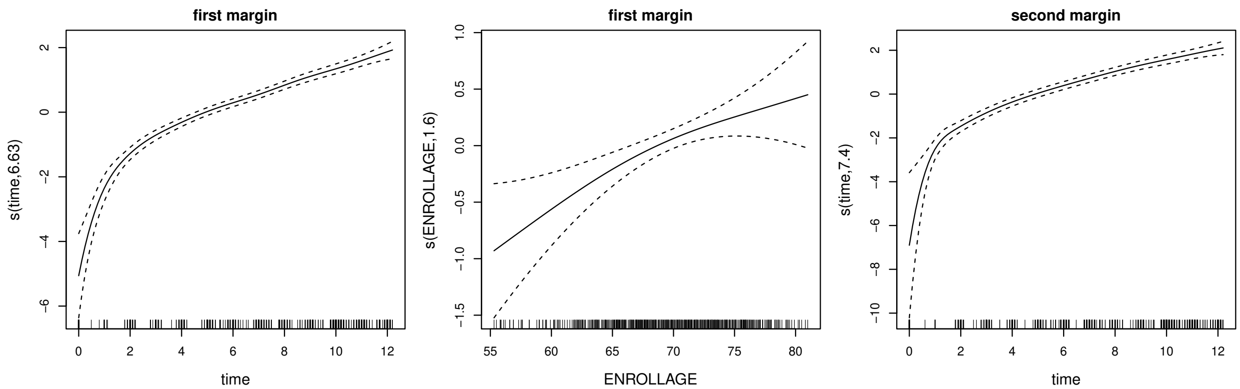


Figure 1.6: AREDS data. Baseline risks and smoothed effect of ENROLLAGE (for the first equation only). 95% point-wise intervals are based on the result discussed in Marra & Radice (2020). The rug plot, at the bottom of each graph, shows the values of the considered variable. The number in brackets in the y-axis caption of each plot represents the edf of the respective estimated smooth function. Source: Petti et al. (2022)

69 years old, with an AMD severity score equal to 6 for both eyes, but with different `rs2284665` genotypes. The middle panel of Figure 1.7 shows the joint progression-free probability contours for subjects who are 69 years old, with the 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 an 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 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.

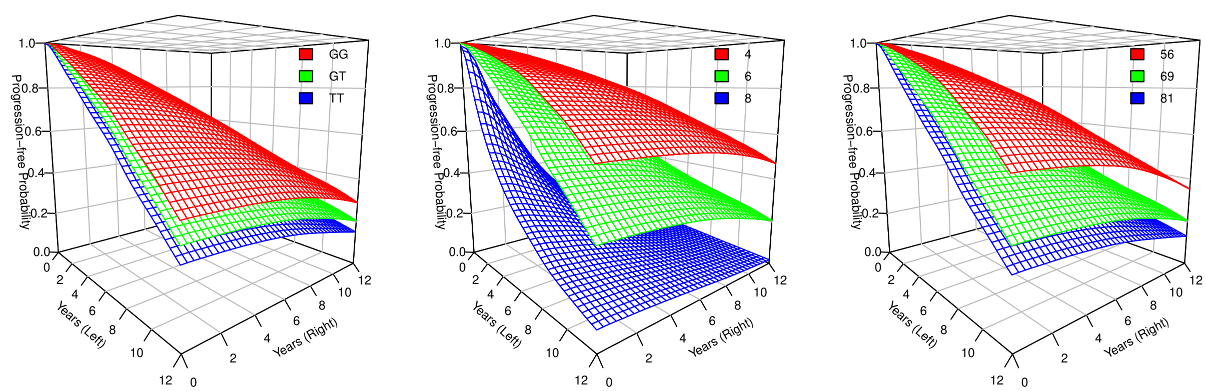


Figure 1.7: AREDS data. Joint progression-free probability contours under different scenarios. In the left panel, age is set to 69 and AMD severity score to 6 for both eyes. In the middle panel, age is set to 69 and genotype to GT. In the right panel, genotype is set to GT and AMD severity score to 6 in both eyes. Source: Petti et al. (2022)

1.4 Discussion

We have introduced a copula link-based additive model for bivariate time-to-event outcomes under a general domain in which all the bivariate censoring combinations can be employed. As has extensively been discussed in the methodology section, model fitting is structured around the simultaneous estimation of all model parameters and relies on a penalised maximum likelihood approach with integrated stable and efficient automatic multiple smoothing parameter selection. Inferential results are also readily available. Once the model has been fitted, a complete summary of the model can be obtained using the corresponding `R` command. 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. The simulation study has shown that the approach recovers the model parameters well and the real application using data on progression to late AMD has highlighted the methodological, computational and practical features of the model.

Although the literature in this area is reasonably ample, to the best of our knowledge, only Sun & Ding (2021) 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, models the baseline survival functions by means of monotonic P-splines which are theoretically and computationally advantageous, and conveniently characterises 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 (2021) does not simultaneously support all possible bivariate combinations of censoring types and forces the two sets of regression parameters to be the same.

Chapter 2

Bivariate Variable Ranking procedure for Copula Link-Based Survival additive models via Fisher Information metric

2.1 Motivation

Technologies have had a deep impact on society and on data collection in a wide range of scientific areas. With a relatively low cost we are able to collect massive amounts of information (and noise). This has led to the *high dimensional data* phenomenon. Pivotal examples are genome-wide association studies, biomedical imaging, tomography, and tumor classification. In disease classification tens of thousands of expressions of molecules or ions are potential predictors. In geno-wide association studies hundreds of thousands of SNPs are potential covariates. In the biomedical field, huge numbers of magnetic resonance images (MRI) and functional data are collected for each subject. Satellite imagery has been used in natural resource discovery and agriculture, collecting thousands of high resolution images. When interactions are considered the distinction between important and unimportant variables becomes even more dramatic.

In such areas, researchers have to deal with a ridiculous amount of information (with signals associated with noise). This makes not only the data exploration phase (see Gazis et al., 2010; Donoho et al., 2000; Giradi & Holzinger, 2018), but also the model building process far more challenging than in the past.

The ease of obtaining so much information with such minimal effort has contributed to reshaping statistical thinking and data analysis paradigms, pushing the existing statistical methods to the extreme of computational capabilities. This has driven researchers to think of new ways of making inferences in such high and ultra-high dimensional contexts. Donoho et al. (2000) convincingly demonstrated the need of developments in high dimensional data analysis, presenting the curse and blessing of dimensionality. Fan & Li (2006) give a comprehensive grasp of statistical challenges in high dimensionality.

2.1.1 State of the Art

Variable selection methods play a central role in contemporary statistical learning methods. During the model building process, the question of which variable suits better to explain the phenomenon often arises. This is even more true in the case of bivariate copula survival models under a censoring scheme (presence of two outcomes and missing information). Under this domain we are interested in identifying two sets of relevant covariates for the first and second random time to event respectively (T_{1i} and T_{2i}). This can intuitively be achieved by ranking, in order of importance, the covariates using a metric (say ω) to assess the contribution of each independent variable in the dataset. Variable selection procedures, in such a bivariate domain, take a different flavour. We have two distinct but dependent survival margins and we want to select two sets of relevant variables for each of them. Having said this, this is a different and potentially more complicated task than doing variable selection under a more conventional univariate domain. Furthermore, the challenge is made even more difficult because of a lack of solid bibliographic references. As far as the authors are aware there is no valuable variable selection or variable ranking method nor implementation available in the literature for the class of Bivariate Copula Survival models.

The main and the most used variable selection approaches (see Fan & Lv, 2010; Desboulets, 2018) usually rely on metrics. Metrics that have the role of evaluating the contribution that a single covariate X_j has on the outcome Y, for instance: Based on a standardised design matrix, Fan & Lv (2007) proposed in the linear regression context the correlation coefficient. Their application orthogonal design leads to $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ and as a consequence $\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}$, which can be interpreted as the vector consisting in the sample *Person correlation* coefficients. The screening procedure based on this metric is commonly known as Sure Independent Screening (SIS). Hall & Miller (2009) considered a *generalized version of correlation coefficient* which is able to capture non-linear dependencies. This is achieved by applying a functional $h(\cdot) : \mathbb{R} \rightarrow \mathcal{H}$ where $\mathcal{H}$ is a class of cubic splines applied at all the covariates. Following the same intuition, Fan & Song (2010) proposed a procedure based on the magnitude of *spline approximation* of Y over each X_j. In Cho & Fryzlewicz (2012), the variable screening is accomplished using the *tilted correlation*. It is obtained in practice by "tilting" each column $X_j^* \leftarrow g(X_j)$ where $g(\cdot)$ is a linear application that projects each variable onto a subspace chosen in the hard-thresholding step, such that the impact of other variables X_k, $k \neq j$ on the "tilted" correlation between X_j^* and Y is reduced and thus the relationship between the jth covariate and the response can be identified more accurately. Li et al. (2012) proposed to rank the covariates according to the *distance correlation* (Székely et al., 2007). He et al. (2013) applied a ranking procedure based on a *marginal quantile* of Y given $\mathbf{X} = (X_1, \ldots, X_p)^T$ such that $Q_\alpha = \inf\{y : \mathbb{P}(Y \leq y|X) \geq \alpha\}$, proving that the sure screening property remains valid when the response is subject to random right censoring. Shao & Zhang (2014) proposed a *martingale difference correlation ranking*.

In a bivariate survival copula domain, any metric measuring the association between outcomes and covariates is bound to fail, namely: i) these metrics capture the association between Y and X_j, in survival models we observe a time to event for each margin. This aspect makes the above measures (i.e. Person correlation, Kendall's τ, generalized correlation and generalization) completely ineffective; ii) most of the quoted methods are based on measuring only linear dependence between outcomes and covariates. This is too restrictive for our domain (i.e. splines, archimedian and non-archimedian copula structures) as we would not be able to exclusively handle linear dependencies; iii) the metrics mentioned above have not been derived taking into account the copula structure; iv) they do not allow us to obtain distinct sets of important variables for each margin. For all these reasons, in the bivariate survival copula context the selection of relevant variables can not be achieved by any metric so far discussed in literature.

In this chapter, we contribute in developing and implementing in R a first attempt of a variable ranking procedure by extending the Ranking-Based Variable Selection algorithm ($RBVS$) (Baranowski et al., 2020) in the scenario of two different sets of important variables. This has been practically achieved by generalizing the algorithm and more importantly by proposing a metric able to assess, in a specific way, which covariate is important for T_1 and T_2. The procedure discussed in this work has been implemented from scratch in GJRM package in R to facilitate and encourage the use of such a method in industry and academia and enhance reproducible research.

The following section is structured as follows. In section 2.2 the algorithm framework is introduced as well as our metric proposal based on the Fisher Information matrix. In section 2.3.1 a simulation study has been conducted to validate the variable ranking procedure discussed.

2.2 Methodology

The Ranking-Based Variable Selection $(RBVS)$ algorithm from Baranowski et al. (2020) is a screening procedure based on bootstrap and permutations of indexes. This technique is particularly suited when $p \gg n$, with the advantage that it does not rely on weak assumptions but on the properties of the metric employed.

The main idea is the following. Given the model $Y = f(X_1, \dots, X_p) + \epsilon$ and a given metric ω to assess the importance that each X_j for $j = 1, \dots, p$ has on Y, using the m out of n bootstrap of Bickel et al. (2012) constructs a permutation ranking $\mathbf{R} = \{R_1, \dots, R_p\}$ satisfying $\omega_{R_1} \geq \omega_{R_2} \geq \cdots \geq \omega_{R_p}$. The covariates that result on top of this ranking are the ones with the higher value of ω. The strength of this algorithm is that the metric can be any type (e.g., Person correlation, Lasso coefficient, generalized correlation) as long as it retains the k*- top ranked, locally-top-ranked and the top-ranked set properties* (see, section 2.2 Baranowski et al., 2020). The probability that a set of k variables $\mathcal{A}$ is ranked at the top is defined as $\pi(\mathcal{A}) = \mathbb{P}(\{R_1, \dots, R_p\} = \mathcal{A})$ where this value is computed using a bootstrap approach. Under this domain, a *top-ranked* set $\mathcal{A}$ is a set with not a negligible probability associated with it. Under the conditions stated in Baranowski et al. (2020) it can be proved to be unique.

The $RBVS$ procedure just introduced manifests some limitations if applied in a Bivariate

Copula Survival domain (see Marra & Radice, 2020; Petti et al., 2022). Since we are dealing with $S_\nu(t_{\nu i}|\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu) = P(T_{\nu i} > t_{\nu i}|\mathbf{x}_{\nu i};\boldsymbol{\beta}_\nu) \in (0,1)$ for $\nu = 1,2$ where each survival has its own set of covariates. We would like to obtain from the procedure two different rankings and two distinct *top-ranked* sets. The aim of this section is to extend the $RBVS$ procedure to the case of two not necessarily overlapping sets $\mathcal{A}^1$ and $\mathcal{A}^2$. Therefore we are not only interested in selecting the most informative covariates for the model, but we would like to assess which of the p (in)dependent variables contribute to explain one or both time to events. For this reason, the $RBVS$ algorithm as is originally presented produces results not appropriate in such a bivariate context. The generalization proposed in the following section does not alter the properties of the algorithm, k- top ranked, locally top ranked and the top ranked set property stated in Baranowski et al. (2020). In fact, we can always think of the set of important variables $\mathcal{A}$ as the union of $\mathcal{A}^1$ and $\mathcal{A}^2$.

Unfortunately the package that originally accommodated the $RBVS$ procedure is no longer available on `CRAN`. With the aim to facilitate the use of the procedure for practitioners, the authors decided to implement the extended procedure from scratch in `GJRM` package.

2.2.1 Notation

Let $\boldsymbol{X} = (X_1, \ldots, X_p)$ constitute a set of random variables which potentially influence either T_{1i} or T_{2i}, or even both of them. Unless specified, all quantities presented in this section are computed on n observations. Therefore, we observe $\mathbf{Z}_i = \{T_{1i}, T_{2i}, X_{i1}, \ldots, X_{ip}\}$, $i = 1, \ldots, n$, as independent copies of $\mathbf{Z} = \{T_1, T_2, X_1, \ldots, X_p\}$. Let $\mathcal{A}^\nu,\ \nu = 1,2$ be a set of covariates for the ν-th outcome and $|\mathcal{A}^\nu|$ its cardinality. For any of the k out of p covariates, we denote with $\Omega_k^\nu = \{\mathcal{A}^\nu \subset \{1, \ldots, p\} : |\mathcal{A}^\nu| = k\}$ the set of important covariates of size k selected using n observations. For notation convenience we have dropped the dependence of this set from p and thus from n. Furthermore, the k covariates in Ω_k^ν are ranked using a variable raking defined as a permutation of indexes $\mathbf{R}^\nu = (R_1^\nu, \ldots, R_k^\nu)$ computed on $\mathbf{Z}$ based on a known metric such that $\hat{\omega}^\nu_{R_1^\nu} \geq \cdots \geq \hat{\omega}^\nu_{R_k^\nu}$, where $\hat{\omega}^\nu$ is an estimate of ω^ν. This implies that both $\hat{\omega}$ and the variable ranking can vary with $\boldsymbol{Z}$. In other words, using a specified metric, we are assessing the importance that the k covariates, for each time to event.

This allows us to select which covariate is important for T_1 and which one contributes in explaining T_2. The approach just introduced turns out to be particularly valuable from a casual inference point of view as we are able to assess which of the covariates contribute in explaining marginally the outcomes.

In general, when $|\mathcal{A}^\nu| = k$, for any $\mathcal{A}^\nu \in \Omega_k^\nu$, $k = 1, \dots, p$, we are interested in obtaining the probability of being ranked at the top, that can be defined as

$$\pi^\nu(\mathcal{A}^\nu) = \mathbb{P}\Bigg(\{R_1^\nu, \dots, R_{|\mathcal{A}^\nu|}^\nu\} = \mathcal{A}^\nu \Bigg), \tag{2.1}$$

The quantity just presented allows us to associate to each column in the design matrix the probability of being exactly in the $j-$th position in the ranking. For convenience, we assume that when $k = 0$ then $\pi^\nu(\mathcal{A}^\nu) = \pi^\nu(\emptyset) = 1$. Under our framework we are interested in the probability of being ranked at the top using only a subset of m observations. This probability will be denoted as $\pi_m^\nu(\mathcal{A}^\nu)$.

To estimate $\pi_m^\nu(\mathcal{A}^\nu)$ Baranowski et al. (2020) used a bootstrap approach such that for each $b = 1, \dots, B$ (with B the number of bootstrap replicates) and given $r = \lfloor n/m \rfloor$, extracts from $\mathbf{Z}_i$, for $i = 1, \dots, n$, r independent subsets without replacement $(I_{b_1}, \dots, I_{b_r})$ and for each bootstrap replicate computes the empirical relative frequency of $\mathcal{A}^\nu$, given by $r^{-1} \sum_{j=1}^r \mathbf{1}(\mathcal{A}^\nu | I_{b_j})$, with $\mathbf{1}$ taking value one when $\mathcal{A}^\nu = \{R_1^\nu(\{\mathbf{Z}_i\}_{i \in I_{b_j}}), \dots, R_{|\mathcal{A}^\nu|}^\nu(\{\mathbf{Z}_i\}_{i \in I_{b_j}})\}$ and zero otherwise. Then, the estimate of $\pi_m^\nu(\mathcal{A}^\nu)$ is obtained from:

$$\hat{\pi}_m^\nu(\mathcal{A}^\nu) = B^{-1} \sum_{b=1}^{B} r^{-1} \sum_{j=1}^{r} \mathbf{1}(\mathcal{A}^\nu | I_{b_j}), \tag{2.2}$$

The probability introduced avoids the fact that the covariates are ordered according to a particularly lucky sample. This is by calculating for each of them the probability of being exactly in the j-th position in the ranking. Therefore, the procedure tends to select the set of variables that has the greatest value of (2.2), the top-ranked variables. In general, it is not equal to compute (2.2) on m rather than on n observations. Baranowski et al. (2020) showed that under the condition that m is not too small with respect to n there is a negligible difference between the probabilities computed on m with respect to the ones computed on n observations. In other words, their ratio tends asymptotically to one. With that said, in combination with some bounds on the estimation accuracy of (2.2), we can conclude that $\hat{\pi}^{\nu}_{m}(\mathcal{A}^{\nu})$ is a good candidate to estimate the k-top-ranked set. The variable selection procedure implemented from scratch in `R` in `GJRM` package can be divided in two distinct macro steps.

1. (*screening step*) given by:

$$\hat{\mathcal{A}^{\nu}}_{k,m} = \underset{\mathcal{A}^{\nu}\in\Omega^{\nu}_{k}}{\arg\max}\ \hat{\pi}^{\nu}_{m}(\mathcal{A}^{\nu}),$$

with Ω^{ν}_{k} being the set of all permutations of $\{1, \dots, k\}$. We need to detect the *important* variables for T_1 and T_2. Using `gjrm()` function in `GJRM` package, for each covariate in the dataset we fit a Copula Link-Based Survival model. The set of equations specified are:

$$\begin{aligned}
\eta_{1i} &= g(S_{10}(t_{1i})) + \beta_{10} + \mathbf{x}^{T}_{ij}\beta_{11},\\
\eta_{2i} &= g(S_{20}(t_{2i})) + \beta_{20} + \mathbf{x}^{T}_{ij}\beta_{21},\\
\eta_{3i} &= \beta_{30},\\
\text{for } & j = 1, \dots, p,\ \ b = 1, \dots, B, i = 1, \dots, n,
\end{aligned} \tag{2.3}$$

where with $g(\cdot) : (0, 1) \to (-\infty, \infty)$ we denote a differentiable and invertible link function that can be any type (table 1.2), while $S_{\nu 0}(t_{\nu i})$ is a background survival function estimated using monotonic splines and ten basis (see Tables 1.1 and 1.2 in section 1.2.4 for details about copulas and link functions implemented in the `GJRM` package). The set of equations defined in (2.3) allowed us to obtain two rankings for the p covariates based on $\hat{\omega}^{\nu}_{j}$ $j = 1, 2$. From (2.3) we note that a) we are evaluating the same $\mathbf{x}_j$ in η_1 and η_2 in the equations modelling the two time to event;. b) in the equation controlling the dependence parameter η_3 only the intercept is estimated; c) since we are evaluating the p covariates one by one, correlation should not impact the variable ranking algorithm. To better clarify the fitting procedure, the following simplified example is provided.

Example 1. *For the sake of simplicity let $\nu = 1$, our aim is to assess which of p columns is* *important in explaining T_1. With p different fits, we will end up with a ranking where each $\mathbf{x}_j$ is sorted in order of importance according to $\hat{\omega}_j(\mathbf{x}_j)$. After having repeated the procedure B times we finally have to verify the frequency that the j-th covariate resulted in the j'-th position using the probability in (2.2).*

Finally, this first step has the aim of associating each set of important variables a not null probability, such that to retain only the ones with the maximum values of $\hat{\pi}^{\nu}_{m}(\mathcal{A}^{\nu})$.

2. (*selection step*). The size of the top-ranked set $s^{\nu} = |S^{\nu}|$, $\nu = 1, 2$ is unknown so it should be estimated. It is common at this stage to introduce a threshold or in general a stopping rule. In Baranowski et al. (2020) the following ratio is proposed $(\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k+1,m}))^{\tau}/\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k,m})$ with $\tau \in (0, 1]$ such that the relevant covariates are the s top-ranked variables where:

$$\hat{s}^{\nu} = \underset{k=0,\ldots,k_{\max}-1}{\arg\min} \frac{\left(\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k+1,m})\right)^{\tau}}{\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k,m})}, \tag{2.4}$$

where τ is a tuning parameter. The intuition of this ratio can be caught from the following argument presented in $Baranowskiet\ al.$ (2020)

$$\frac{\left(\hat{\pi}^{\nu}_{n,m}(\hat{\mathcal{A}^{\nu}}_{k+1,m})\right)^{\tau}}{\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k,m})} = \left(\frac{\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k+1,m})}{\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k,m})}\right)^{\tau} \left(\frac{1}{\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k,m})}\right)^{1-\tau}.$$

For $\tau = 1$, we look for k where $\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k+1,m})$ declines drastically. For a general value of τ, we are looking for k that is a trade off between the most drastic decline in proportion and the hard thresholding rule.

In practice, given the estimated probabilities of $\hat{\pi}^{\nu}_{m}(\hat{\mathcal{A}^{\nu}}_{k,m})$, for $k = 0, \ldots, k_{\max} - 1$, with $k_{\max}$ a fixed large integer, the number of relevant variables is related to the evaluation of the magnitude of the estimated probability and $\hat{s}^{\nu}$ which corresponds to the case where the ratio in (2.4) has the greatest decrease. In other words, the set's sizes of relevant variables have not to be specified ex ante. In our experience the criterion expressed in (2.4) works well (evidence can be found in Baranowski et al. (2020)). Furthermore, we are assuming that $|S^{\nu}|$ is much smaller than p, it is computationally efficient to optimise over $\{0, \ldots, k_{max}\}$ instead of $\{0, \ldots, p-1\}$ in (2.4).

2.2.2 Metric proposal

From the discussion in section 2.2.1 emerged, in a clear way, the crucial role of the metric in the variable ranking procedure. Unfortunately as far as the authors are aware, there is no measure able to accommodate a bivariate survival copula setting. The aim of this section is to propose a metric able to perform the Bivariate Ranking-Based Variable Screening $(BRBVS)$ in an effective way without altering the screening properties discussed in Baranowski et al. (2020). The impatient reader can jump straight to section 2.3.1 for evidence of the effectiveness of the metric proposed.

Firstly, we want to present the reasoning that led us to the definition of our metric idea for the class of model(s) discussed in section 1.2.4 (see. Marra & Radice, 2020; Petti et al., 2022). A first attempt has been made by thinking about a metric able to capture the association between the two outcomes. The Mutual Information (MI) has several remarkable properties that motivate its use in a variable ranking procedure. The first one is the relation with the Copula Entropy (CE) and thus with the copula structure itself. It can be proved that MI is equivalent to negative CE. The second remarkable property is its monotonic increasing association with the copula dependence parameter. Finally, MI is zero if and only if we are under orthogonality condition (a discussion about the MI properties is in Appendix I). For the reasons just stated, MI seemed to be a suitable candidate to solve our task, as it captures the relation between two or more random variables, taking into account the copula structure. It turns out that we are under the conditions stated in section 2.3 of Baranowski et al. (2020). A sketch of proof that the concentration bound holds for MI is in Appendix I. The intuition is as follows: when a couple of noisy covariates (say X_j and $X_{j'}$) are added to the model setting, they should increase the entropy between the two dependent time to events. This implies that an MI-based variable ranking procedure has the potential to be far more effective than other metrics (e.g., Pearson correlation, tilted correlation, generalized correlation). The procedure has been practically achieved by computing MI on the fitted values. Let $\hat{\omega}_j(\mathbf{x}_j^{(1)}, \mathbf{x}_{j'}^{(2)}) = \hat{I}(\hat{y}_1, \hat{y}_2 | \mathbf{x}_j^{(1)}, \mathbf{x}_{j'}^{(2)})$ the MI estimate resulting form the following model setup

	$B = 1$		$B = 50$	
p	# of fit	time (hours)	# of fit	time (hours)
2	1	0.0016	50	0.083
5	10	0.0167	500	0.83
10	45	0.075	2250	3.75
100	4950	8.25	247500	412.5
1000	49950	83.25	2497500	4165.5

Table 2.1: MI-based variable ranking computational times obtained by considering an average fitting time of 6 seconds.

$$\eta_{1i} = g(S_{10}(t_{1i})) + \beta_{10} + \mathbf{x}_{ij}^T \beta_{11},$$

$$\eta_{2i} = g(S_{20}(t_{2i})) + \beta_{20} + \mathbf{x}_{ij'}^T \beta_{21},$$

$$\eta_{3i} = \beta_{30},$$

$$\text{for } j'j = 1, \ldots, p, \;\; i = 1 \ldots, n,$$

where $\mathbf{x}_j$ has been included in the equation modelling the first time to event, while $\mathbf{x}'_j$ models the second time to event. This setting requires to fit $\binom{p}{2}$ models to evaluate all the independent variables. Although this procedure has given satisfactory results, its computational cost is prohibitive, especially when the size of p is high. This considering that under a 2.60-GHz Intel(R) Core(TM) Windows 10 configuration the average computing time to fit a model using `gjrm()` function is about 6 seconds. Furthermore, MI-based variable ranking does not allow us to obtain $\mathcal{A}_1$ and $\mathcal{A}_2$. Some evidence of the MI-variable ranking procedure applied to a general case of Bivariate Copula model(s) is discussed in Appendix J. Despite the MI seeming to perform well, it has two main pitfalls. Firstly, it fails in defining two different and not necessarily overlapping rankings. Secondly, when applied to time to events it fails in recovering the set of important variables. Evidence can be found in Appendix J, where *selected* simulation studies results are presented. These aspects make the procedure inapplicable in the domain discussed in section 1.2. Table 2.1 shows an example of the computational time that the MI-based procedure requires.

The ideal metric, for the class of bivariate copula survival models, should not only be a measure capable of taking into account the likelihood and the copula probabilistic structure, but also a quantity able to marginally evaluate the contribution that each dependent variable has on T_1 and T_2. In other words, our goal is to obtain two non-necessarily overlapping sets. Under the model framework discussed in section 1.2, it seems reasonable to use a measure of goodness of fit as a starting point. Marra & Radice (2020) derived an expression of the Akaike Information Criterion (AIC) for the class of Bivariate Copula Survival models

$$AIC = 2\psi + ||\boldsymbol{M} - \sqrt{-\boldsymbol{H}}\hat{\boldsymbol{\delta}}||^2. \tag{2.5}$$

Where $\mathbf{M} = \boldsymbol{\mu}_{\mathbf{M}} + \boldsymbol{\epsilon}$, $\boldsymbol{\mu}_{\mathbf{M}} = \sqrt{-\mathbf{H}}\boldsymbol{\delta}^0$, $\boldsymbol{\epsilon} = \sqrt{-\mathbf{H}}^{-1}\mathbf{g}$, with $\mathbf{g}$ we denoted the gradient vector such that $\mathbf{g}_1 = \frac{\partial \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_1}\Big|_{\boldsymbol{\beta}_1 = \hat{\boldsymbol{\beta}}_1}$,$\mathbf{g}_2 = \frac{\partial \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_2}\Big|_{\boldsymbol{\beta}_2 = \hat{\boldsymbol{\beta}}_2}$, $\mathbf{g}_3 = \frac{\partial \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_3}\Big|_{\boldsymbol{\beta}_3 = \hat{\boldsymbol{\beta}}_3}$,and $\mathbf{H}$ is the Hessian matrix whose elements are $\mathbf{H}_{ij} = \frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_j}\Big|_{\boldsymbol{\beta}_i = \hat{\boldsymbol{\beta}}_i \boldsymbol{\beta}_j = \hat{\boldsymbol{\beta}}_j}$ $i, j = 1, 2, 3$, $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_3)^T$ is the estimated parameter vector, $\boldsymbol{\delta}^0$ is the vector of parameters such that minimises the Kullback-Lieber distance between the true likelihood and the observed one. Finally, ψ represents the effective degree of freedom. The expression (2.5) can be obtained in two steps. First, a Taylor series expansion of the likelihood around $\boldsymbol{\delta}^0 \in \Theta$, to then dropping all the terms not affected by $\boldsymbol{\lambda}$.

The Bivariate Ranking Based Variable Screening ($BRBVS$) procedure has been implemented using the model setting defined in Sections 1.2 and 2.2 (see equations (2.3)). Therefore, under this setting all the fitted models in the ranking algorithm are bound to have the same degree of freedom. In other words, ψ does not have a role in identifying the best model. Recalling some well-known quantities and asymptotic results explored in Supplementary Material B in Marra & Radice (2020), the expression (2.5) can be written as

$$\begin{aligned} AIC &\propto ||\boldsymbol{M} - \sqrt{-\boldsymbol{H}}\hat{\boldsymbol{\delta}}||^2 \\ &= ||\sqrt{-\mathbf{H}}\boldsymbol{\delta}^0 + \sqrt{-\mathbf{H}}^{-1}\mathbf{g} - \sqrt{-\boldsymbol{H}}\hat{\boldsymbol{\delta}}||^2 \\ &\propto ||\sqrt{-\mathbf{H}}\boldsymbol{\delta}^0 - \sqrt{-\boldsymbol{H}}\hat{\boldsymbol{\delta}}||^2. \end{aligned} \tag{2.6}$$

Noting that $\boldsymbol{\delta}, \boldsymbol{\delta}^0 \in \mathbb{R}^W$, where $W = W_1 + W_2 + W_3$ such that $\boldsymbol{\beta}_1 \in \mathbb{R}^{W_1}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{W_2}$ and $\boldsymbol{\beta}_3 \in \mathbb{R}^{W_3}$. The last line can be interpreted as the ℓ_2 norm between two W dimensional points adjusted for information contained in a given sample. Unless in some degenerative cases, the quantity $(\boldsymbol{\delta}^0 - \hat{\boldsymbol{\delta}})$ is unknown, meaning that we would never be able to evaluate how much our estimates are distant from the KL minimizer. Having said that, some interesting geometric considerations can be made on the Hessian matrix. It is a standard result to note that the accuracy of some estimates is measured by the sharpness of the underlying log-likelihood $\ell(\boldsymbol{\delta})$. In differential geometry, the curvature is related to the second derivatives of a function, the components of the main diagonal of the hessian matrix. The diagonal elements of the Hessian matrix can be seen as directional derivatives with respect to the vectors of the canonical basis. Using the plug-in principle, we can evaluate $||\sqrt{-\mathbf{H}}\hat{\boldsymbol{\delta}}||^2$ with respect to $-\mathbb{E}(\mathbf{H}) = \boldsymbol{\mathcal{I}}$. Furthermore, it is well known that $\mathcal{I}$ is symmetric, positive semi-definite and not singular. Then, by the spectral theorem there exists an orthogonal system such that $\boldsymbol{\mathcal{I}}$ can be diagonalized, in which the cross derivatives are zero. Finally, noting that every geometric quantity must be independent of the reference system, it follows that it seems reasonable to use only the information on the main diagonal of the Fisher information matrix. Recalling that $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T, \hat{\boldsymbol{\beta}}_3^T)$ and that the Fisher information matrix can be written as a 3×3 block matrix in the following way

$$\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\delta}}) = \begin{bmatrix} \mathcal{I}(\hat{\boldsymbol{\delta}})_{11} & \mathcal{I}(\hat{\boldsymbol{\delta}})_{12} & \mathcal{I}(\hat{\boldsymbol{\delta}})_{13} \\ \mathcal{I}(\hat{\boldsymbol{\delta}})_{21} & \mathcal{I}(\hat{\boldsymbol{\delta}})_{22} & \mathcal{I}(\hat{\boldsymbol{\delta}})_{23} \\ \mathcal{I}(\hat{\boldsymbol{\delta}})_{31} & \mathcal{I}(\hat{\boldsymbol{\delta}})_{32} & \mathcal{I}(\hat{\boldsymbol{\delta}})_{33} \end{bmatrix} \text{ where } \mathcal{I}(\hat{\boldsymbol{\delta}})_{ij} = -\mathbb{E}\left[\frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_j} \Big|_{\substack{\boldsymbol{\beta}_i = \hat{\boldsymbol{\beta}}_i \\ \boldsymbol{\beta}_j = \hat{\boldsymbol{\beta}}_j}} \right] \; i, j = 1, 2, 3.$$

After having estimated the vector of parameters and the standard errors, following the same notation used in (2.3), reasonable metrics would be

$$\begin{aligned} \hat{\omega}_j^1(\mathbf{x}_j) &= \hat{\beta}_{11}^2 i_{11}(\hat{\beta}_{11}) \\ \hat{\omega}_j^2(\mathbf{x}_j) &= \hat{\beta}_{21}^2 i_{22}(\hat{\beta}_{21}), j = 1, \ldots, p, \end{aligned} \tag{2.7}$$

where $\hat{\boldsymbol{\beta}}_\nu[J_{\nu k_\nu}+2] = \hat{\beta}_{\nu 1}$, $\nu = 1,2$ is the parametric effect extracted from the coefficients vector (containing smooth and parametric effects) associated with the $\nu-$th margin. While, $\mathcal{I}(\hat{\boldsymbol{\delta}})_{\nu\nu}[J_{\nu k_\nu}+2, J_{\nu k_\nu}+2] = i_{\nu\nu}(\hat{\beta}_{\nu 1})$ is the corresponding element of the Fisher Information matrix extracted from the appropriate block diagonal element. Intuitively, we are associating to our estimates $\hat{\beta}_{\nu 1}$ a measure of information given by $i_{\nu\nu}(\hat{\beta}_{\nu 1})$. A high value of $i_{\nu\nu}(\hat{\beta}_{\nu 1})$ denotes a sharp curvature of the likelihood in the direction of the estimated parameter. On the contrary, a low value denotes a lower sharpness. With (2.7) we are not only attributing a high score to the parameters that are different from zero, but we are also taking into account the associated log-likelihood sharpness. This leads the algorithm not to just deal with sparsity, but to make a more complete evaluation of the parameters taking into account the information extracted from the $b-$th bootstrap sample and used in the estimation process.

2.2.3 Bivariate RBVS and Computational aspects

The algorithm is based on four main steps, where step 1 in $BRBVS$ corresponds to step 1 in $RBVS$ (Baranowski et al., 2020). In **step 1** we draw B sub-samples from the data of size m. In **Step 2** for each sub-sample we estimate a Copula Link-based Survival model using the model specification just introduced. This leads us to obtain an estimate of $\omega_j^1(\mathbf{x}_j)$ and $\omega_j^2(\mathbf{x}_j)$ based on the I_{bl} sub-sample for both margins, such that the measures so computed $\{\hat{\omega}_j^\nu(\{\boldsymbol{Z}_{\boldsymbol{i}}\}_{i\in I_{bl}})\}_{j=1}^p$ are sorted in a non-increasing order to identify the set of permutation indexes $\boldsymbol{R}^\nu(\{\boldsymbol{Z}_{\boldsymbol{i}}\}_{i\in I_{bl}})$. Here the algorithm has been modified to accommodate not only the fitting with `gjrm()` using the setup presented in (2.3), but also with the main objective to obtain two distinct permutation rankings. In **step 3**, for $k = 1,\ldots,k_{max}$ we find $\hat{\mathcal{A}}^\nu{}_{k,m}$ the k-element set with the highest frequency in the top $\boldsymbol{R}_n^\nu(\{\boldsymbol{Z}_{\boldsymbol{i}}\}_{i\in I_{bl}})$ for all $b = 1,\ldots,B$ and $l = 1,\ldots,r$. In **step 4** $\hat{\pi}^\nu_{n,m}(\hat{\mathcal{A}}_{k,m})$ are employed to find $\hat{s}^\nu$. Also with regards to step 3 and 4, the algorithm has been modified to obtain two distinct sets of important variables as output.

As concerns the computational aspects of the procedure, we denote $c(p)$ the computational cost (time) in obtaining $\{\hat{\omega}_j^1(\mathbf{x}_j)\}_{j=1}^p$ and $\{\hat{\omega}_j^2(\mathbf{x}_j)\}_{j=1}^p$. Taking B times of random partition of n observations into r subsets takes $\mathcal{O}(Br)$ operations. Computing all the $\hat{\omega}_j^\nu$'s for all Br takes $c(p)\times$

Algorithm 1	Variable Ranking based on Copula Link-based additive models
Input	Standardised Random sample $\boldsymbol{Z_i}$ $i = 1, \ldots, n$ subsample of dimension m, positive integer k_{max}, number of bootstrap replicates B, predetermined value $\tau \in$ (0,1].
Output	The estimated set of important variables
Start	
Step 1	Set $r = \lfloor n/m \rfloor$, for $b = 1, \ldots, B$ draw uniformly without replacement from $\mathbf{z}_i$ $m-$ element subsets $I_{b1}, \ldots, I_{bl}$, $l = 1, \ldots, n$
Step 2	Calculate $\hat{\omega}_j^{\nu}(\{\boldsymbol{Z_i}\}_{i \in I_{bl}})$ and define the variable ranking $\mathbf{R}_n^{\nu}(\{\boldsymbol{Z}_i\}_{i \in I_{bl}})$ for all $b = 1, \ldots, B$, $l = 1, \ldots, r$, $j = 1, \ldots, p$, $\nu = 1, 2$
Step 3	for $k = 1, \ldots, k_{max}$, find $\hat{\mathcal{A}}^{\nu}{}_{k,m}$ and compute $\hat{\pi}_{n,m}(\hat{\mathcal{A}}^{\nu}{}_{k,m})$
Step 4	Find $\hat{s}^{\nu}$
return	$\hat{s}^1$ and $\hat{s}^2$
End procedure	

Br manipulations. Evaluating the rankings based on each subset takes $\mathcal{O}(p + p\log(p))$ operations via the selection algorithm and QuickSort partition scheme, doing so for Br subsets ends up with $\mathcal{O}((p + p\log(p))Br)$. In the $BRBVS$ algorithm implemented in `GJRM` we sort the ranking based on p covariates. Originally in Baranowski et al. (2020), once estimated ω the procedure cuts the ranking to k_{max}. We have modified the algorithm such that the sorting procedure takes place on the p covariates and not only a pre-specified number k_{max}. According to the authors, the choice to cut the ranking at a certain level k_{max} is sub optimal. In fact, there is a concrete risk of arbitrarily cutting off important covariates from the ranking procedure. Furthermore, we believe that the risk of ignoring valuable information from the data does not compensate the computational benefit. It is fair to say that the k_{max} value still plays a role in the definition of $\hat{s}^1$ and $\hat{s}^2$. Step 3 can be carried out using $\mathcal{O}(Brk_{max}^2)$ basic operations. The remaining step calls for $\mathcal{O}(k_{max})$ operations. Finally, the total computational cost is $c(p) \times Br + \mathcal{O}(\max\{p, k_{max}^2\}Br)$.

The $BRBVS$ requires the choice of the same number of tuning parameters as its univariate version, B, m, k_{max}, and τ. The parameter B plays a crucial role as it decreases the randomness of the method discussed. However, the computational complexity of $BRBVS$ increases linearly with B. In our simulation experiments we set $B = 50$ as in Baranowski et al. (2020). The problem of the choice of the subsample m is more challenging. As in a finite sample case m cannot be too small as this implies incapacity in selecting informative covariates. In practice Baranowski et al. (2020) proposed to set $m = \lfloor n/2 \rfloor$ that resulted to give satisfactory performance. As concerns

k_{max}, it turns out that this has a limited impact on the overall performance of the algorithm, as long as it is not too small. We recommend to set $k_{max} = \min\{n, p\}$. Finally, the proposed procedure seems not to be sensitive to the choice of τ, in our experiments, following what is stated in Baranowski et al. (2020) we set $\tau = 0.5$.

2.3 Simulation Study

2.3.1 Simulation Study

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a multivariate random variable partitioned in blocks such that $\mathbf{X} = [\mathbf{X}_{11} : \mathbf{X}_{12}]$, where $\mathbf{X}_{11} \in \mathbb{R}^{n \times 3}$ contains the covariates generated from a multivariate standard normal distribution with correlation $\rho_{\mathbf{X}_{11}} = 0.2$, and then transformed using the distribution function of a standard normal distribution, while the remaining (non-informative) independent variables $\mathbf{X}_{12} \in \mathbb{R}^{n \times (p-3)} \sim \mathcal{N}_{p-3}(\mathbf{0}, \mathbf{\Phi})$ have been generated using a multivariate normal distribution with null mean vector and a $\rho_{\mathbf{X}_{12}} = 0$. The first margin T_{1i} was generated from a proportional hazards (PH) using the following formulation

$$T_{1i} = \log[-\log S_{10}(t_{1i})] + \beta_{11}\mathbf{x}_{1i} + \beta_{21}\mathbf{x}_{2i}, \;\; i = 1, \ldots, n,$$

where $S_{10}(t_{1i}) = 0.9 \exp(-0.4 t_{1i}^{2.5}) + 0.1 \exp(-0.1 t_{1i})$. T_{2i} was generated from a proportional odds (PO) model

$$T_{2i} = \log\left[\frac{\{1 - S_{20}(t_{2i})\}}{S_{20}(t_{2i})}\right] + \beta_{21}\mathbf{x}_{1i} + \beta_{22}\mathbf{x}_{3i}, \;\; i = 1, \ldots, n,$$

where $S_{20}(t_{2i}) = S_{10}(t_{2i}) = 0.9 \exp(-0.4 t_{1i}^{2.5}) + 0.1 \exp(-0.1 t_{1i})$. With a vector of population parameters such that $\boldsymbol{\beta}_1^T = (-1, 5, 1.7)^T$ and $\boldsymbol{\beta}_2^T = (-1.5, 2.1)^T$. The right-censoring observations were obtained through uniform distributions so that the censoring rates were about 0.16% and 0.40%. 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. Observations were generated using the Brent's univariate-finding root. The two survival times were joined using a Clayton copula `C0`. In practice this was achieved using the conditional sampling approach, where the predictor for the dependence parameter was specified in two different configurations:

- (A) $\eta_{3i} = \beta_{30}$, with $\beta_{30} = 3$.

- (B) $\eta_{3i} = \beta_{30} + \beta_{31}\mathbf{x}_{1i} + \beta_{32}\mathbf{x}_{2i} + \beta_{33}\mathbf{x}_{3i}$, with $\boldsymbol{\beta}_3^T = (3, -1.5, 1.7, -1.3)$.

The setup of η_3 allowed dependence to vary across observations, with Kendall's τ values ranging approximately from 0.10 to 0.90. The smooth components of the times were represented using using monotonic penalized B-splines and 10 basis function. The set of tuning parameters was specified such that the number of replicates to be $n_{sub} = 100$, $B = 50$, $\tau = 0.5$, $k_{max} = 6$. The sample sizes was set as $n = \{500, 750\}$, while the number of covariates $p = \{20, 30\}$, to have a total of 8 simulation scenarios. Finally, the models were fitted using the function `gjrm()` in `GJRM` package in `R`. Results are shown in Figure 2.1 and in Table 2.2. The goodness of the Bivariate variable ranking procedure has been evaluated by averaging the false positive (FP) and negative (FN) obtained for each n_{sub}, $FP = n_{sub}^{-1}\sum_i FP_i$ and $FN = n_{sub}^{-1}\sum_i FN_i$ to obtain unconditional estimates.

The main findings of the simulation study are:

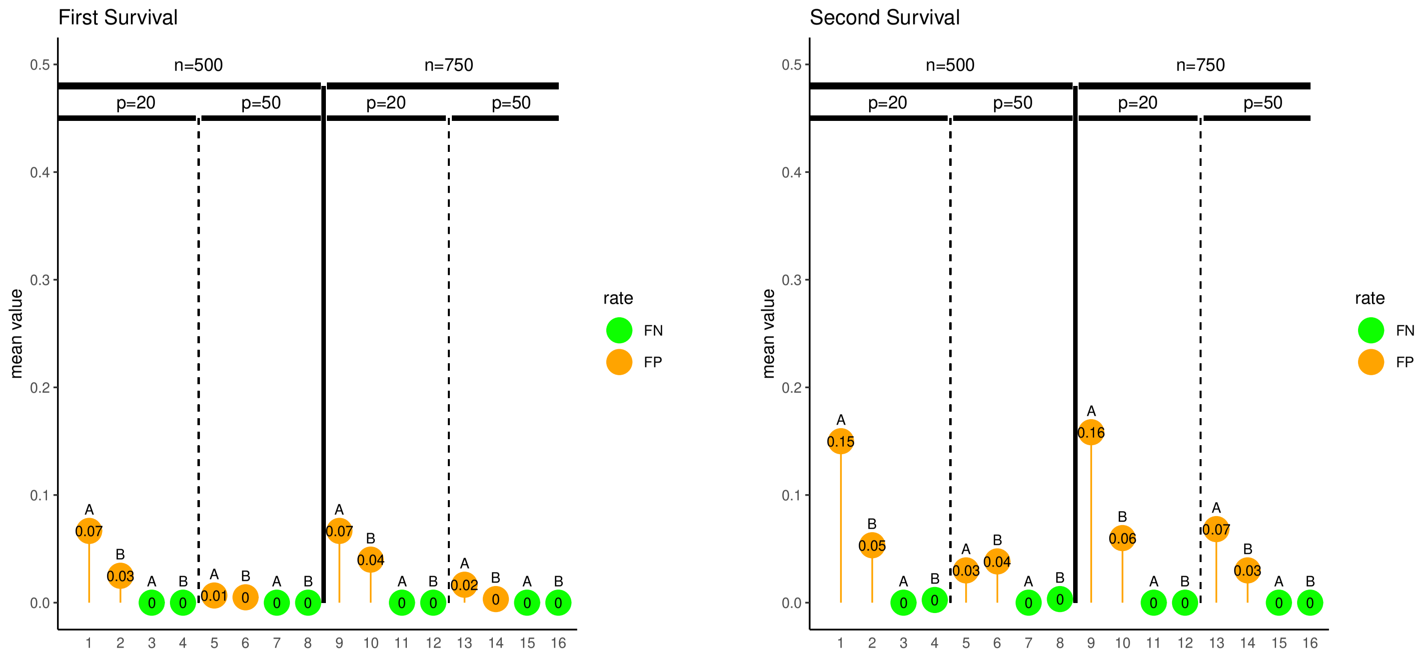
- *sample size*: Recalling that, in each estimate, the effective sample size was $m = \lfloor \frac{n}{2} \rfloor$, hence $m = \{250, 375\}$. There seems to be no noticeable difference between the two sample sizes. In other words, even with a sample size of 250 the metric seems to obtain satisfactory results in terms of selection. The reader should note that for each permutation, the algorithm requires an estimation of 22 parameters.

- *noisy information*: Figure 2.1 and Table 2.2 seem to show that with an increase in the amount of noisy information, from 20 to 50, the false positive rates and false negative rates do not seem to increase. Indeed, for both margins we notice a decrease in both

- *dependent parameter*: The way in which the dependency parameter is generated seems to play an important role in the ability of the metric to select the relevant covariates. In fact, the false positives and false negatives when the parameter is simulated using configuration B are smaller than A.

A	$n = 500$		$n = 750$	
	$p = 20$	$p = 50$	$p = 20$	$p = 50$
FP_{S_1}	0.07	0.01	0.07	0.02
FP_{S_2}	0.15	0.02	0.16	0.07
FN_{S_1}	0.00	0.00	0.00	0.00
FN_{S_2}	0.00	0.00	0.00	0.00
B	$n = 500$		$n = 750$	
	$p = 20$	$p = 50$	$p = 20$	$p = 50$
FP_{S_1}	0.03	0.01	0.04	0.00
FP_{S_2}	0.05	0.04	0.06	0.03
FN_{S_1}	0.00	0.00	0.00	0.00
FN_{S_2}	0.00	0.00	0.00	0.00

Table 2.2: False Positive (FP) and False Negative (FN) estimates for T_1 obtained by applying the (Bivariate)RBVS algorithm employing the metric discussed in section 2.7. Where $FP = n_{sub}^{-1} \sum_i FP_i$ and $FN = n_{sub}^{-1} \sum_i FN_i$. The fitting procedures within (Bivariate)RBVS have been computed using $m = \{250, 375\}$ observations, detailed in section 2.7.



(a) First Margin

(b) Second Margin

Figure 2.1: False Positive (FP) and False Negative (FN) estimates for T_2 obtained by applying the (Bivariate)RBVS algorithm. Details about the computation are in Table 2.2.

2.4 Application of the BRBVS algorithm to AREDS dataset

The global behavior of the $BRBVS$ using the Fisher Information metric is assessed in comparison with the naive metric (absolute values of the coefficients). The two metrics will be evaluated under different setups, one in which the covariates have been standardized while in the other they haven't. In both cases, 100 independent realizations of standard Gaussian random variables have been added to the original columns of the dataset.

The AREDS data is available through the R package from CopulaCenR (Sun & Ding, 2021). This dataset includes 629 Caucasian participants. The event of interest is the progression to late-AMD disease, which is currently considered the most common cause of blindness in developed countries. Due to intermittent assessment times (every 6 months up to the first 6 years and every 1 year thereafter), the data are affected by right-censored, interval-censored, and mixed-censored observations.

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); and more than one-third of the subjects did not develop late-AMD in both eyes (bivariate right-censored). The dataset contains some covariates that have been proven to be related to AMD, and recent examples of applications based on this dataset are discussed in Sun & Ding (2021); Pettifor et al. (2012).

The dataset includes the following variables: SevScaleBL for baseline AMD severity score (a variable with values between 4 and 8, with a higher value indicating more severe AMD). In detail, the patients underwent an ophthalmic exam, at the end of which the doctor assigned a score to both eyes, `SevScale1E` (right) and `SevScale2E` (left). `ENROLLAGE` represents baseline age (a numeric variable), and `rs2284665` represents a genetic variant (a factor variable with levels 0 (`GG`), 1 (`GT`), and 2 (`TT`)). A full description of the dataset is presented in Table 2.3

Notation	Characteristic	Allowed values	Description
$\mathbf{z}_1$	`SevScale1E`	$4-9$	Severity scale associated with the right eye
$\mathbf{z}_2$	`ENROLLAGE`	$1-99$	Age at baseline
$\mathbf{z}_3$	`rs2284665`	$0, 1, 2$	SNP covariate highly associated with late-AMD progression
$\mathbf{z}_4$	`SevScale2E`	$4-9$	Severity scale associated with the left eye
t_{11}	`enrollment_day`	decimal	Start of follow-up in days right eye
t_{12}	`AMD_recurrence_RightEye`	decimal	Time to recurrence or last follow-up in days
t_{21}	`enrollment_day`	decimal	Start of follow-up in days left eye
t_{22}	`AMD_recurrence_LeftEye`	decimal	Time to death or last follow-up in days
cens1	`censoring_status_RightEye`	recurrence, no recurrence	Recurrence censoring variable
cens2	`censoring_status_LeftEye`	recurrence, no recurrence	Overall survival censoring variable

Table 2.3: AREDS data, data description.

Before applying the BRBVS algorithm and comparing the sets of important variables obtained using the Fisher information metric $\omega j = \beta_j^2 i(\beta_j) j = 1^p$ and the naive metrics $\psi j = |\beta_j| j = 1^p$, the best combination of copula and margins had been selected (see., Table 1.1 and 1.2). This was achieved by fitting a full model $\mathcal{M}_{\text{full}}$ where ENROLLAGE, rs2284665, SevScale1E, and SevScale2E were included in the equations controlling the margins (η_1 and η_2) and the parameter of dependence (θ). The copulas employed in this experiment were (AMH 'AMH', Clayton 'C0', FGM 'FGM', Frank 'F', Gaussian 'N', Gumbel 'G0', Joe 'J0', Plackett 'PL', T-student 'T'), the margins (Proportional hazards 'PH', Proportional odds 'PO', Probit 'probit').

From these preliminary fittings, emerged that C0, POPO was the combination with the lowest BIC (4330.08). The results are shown in Table 2.4.

	POPO	PHPH	probprob	POPH	PHPO	probPO	POprob	probPH	PHprob
AMH	4338.05	4356.95	4350.49	4350.75	4344.96	4346.15	4343.08	4359.26	4350.15
C0	**4330.08**	4355.71	4343.66	4347.60	4339.84	4339.46	4334.97	4357.27	4344.96
FGM	4368.67	4386.19	4379.54	4379.82	4374.87	4374.82	4373.47	4385.84	4379.66
F	4333.73	4352.73	4342.50	4346.03	4341.40	4338.65	4337.96	4351.16	4345.91
N	4348.39	4369.10	4360.09	4362.75	4356.50	4354.78	4354.09	4369.87	4362.91
G0	4367.58	4383.04	4381.22	4378.91	4373.61	4375.21	4374.41	4387.28	4381.16
J0	4392.15	4406.13	4406.04	4402.34	4397.40	4400.00	4399.05	4410.89	4404.98
PL	4334.80	4353.36	4343.92	4347.00	4342.67	4340.01	4339.18	4352.44	4347.38
T	4353.31	4372.16	4365.00	4366.51	4361.68	4359.84	4359.29	4373.57	4368.27

Table 2.4: BIC values obtained by fitting gjrm() under the full model configuration $\mathcal{M}_{full}$ such that $\eta_1 = \eta_2 = \eta_3 = \beta_0 + \beta_1 \texttt{ENROLLAGE} + \beta_2 \texttt{rs2284665} + \beta_3 \texttt{SevScale1E} + \beta_4 \texttt{SevScale2E}$. In bold the lowest BIC.

Standardized AREDS perturbed with $\mathcal{N}(0,1)$

Under this scenario the covariates in AREDS had been standardized. In a variable selection procedure the domain of the covariates can have an impact on the magnitude of the coefficient and then invalidate the method. Standardizing is the solution to compare the coefficients within a model. We coded rs2284665 as $0/1$, such that to have three new covariates one for each rs2284665 level , $\mathbf{z}_{3_0} = \mathbf{1}_{\{\mathbf{z}_3=0\}}$, $\mathbf{z}_{3_1} = \mathbf{1}_{\{\mathbf{z}_3=1\}}$, $\mathbf{z}_{3_2} = \mathbf{1}_{\{\mathbf{z}_3=2\}}$. The other numeric inputs were centered and then divided by their standard deviations. The tuning parameters has been specified as follows: $k_{max} = 10$, $m = 328$, $\tau = 0.5$, $n.rep = 50$, Clayton copula (C0) and Proportional odds (PO,PO). The results of the BRBVS procedure are showed in Table 2.5.

	1st margin (right eye)	2nd margin (left eye)	BIC	AIC	LogLik
$\mathcal{M}_\omega$	$\{\mathbf{z}_1, \mathbf{z}_4, \mathbf{z}_{3_0}, \mathbf{z}_{3_2}\}$	$\{\mathbf{z}_{3_0}, \mathbf{z}_4, \mathbf{z}_{3_2}, \mathbf{z}_1, \mathbf{z}_{3_1}\}$	4325.849	4225.385	−2090.079
$\mathcal{M}_\psi$	$\{\mathbf{z}_1, \mathbf{z}_4\}$	$\{\mathbf{z}_4, \mathbf{z}_1, \mathbf{z}_{3_0}, \mathbf{z}_{3_2}, \mathbf{z}_{3_1}\}$	4324.518	4220.659	−2086.951

Table 2.5: BRBVS results using clayton copula, porportional hazard margins $k_{max} = 10$, $m = 328$, $\tau = 0.5$, $n.rep = 80$ using $\omega = \beta^2 i(\beta)$ and $\psi = |\beta|$ as metrics. The covariates are ordered according to their importance. The BIC, AIC and LogLik are obtained by applying `gjrm()` function to a non standardized AREDS.

Using ω, in order of importance we have that `SevScale1E`, `SevScale2E`, `GG` and `TT` resulted to be relevant variables for the progression of the right eye, while `GG`, `SevScale2E`, `TT`, `SevScale1E`,`GT`, are relevant covariates to explain the progression of the left eye. Using ψ we end up with a similar selection of covariates (`GG`, `TT` were not selected in the first margin). Also in terms of AIC and BIC we obtained comparable results.

Despite both the Fisher information metric (ω) and the naive metric (ψ) yielding almost the same values for BIC (4325.85 for ω and 4324.52 for ψ) and AIC (4225.38 for ω and 4220.66 for ψ) , the selection based on $\omega = \beta_j^2 i(\beta)$, resulted in a greater number of covariates in the set of important variables. Notably, the covariates included are known genetic factors that play a role in the progression of AMD, as highlighted in the literature (Sun & Ding, 2021; Petti et al., 2022). This finding provides compelling evidence of the effectiveness of the Fisher information metric in capturing important characteristics associated with the event of interest.

AREDS perturbed with $\mathcal{N}(0, 1)$

Under this framework we perturbed the AREDS dataset adding 100 realizations of independent standard Gaussian random variables $\mathbf{Z}_i \overset{ind.}{\sim} \mathcal{N}(0, 1)$ $i = 5, \ldots, 104$, leaving the covariates in AREDS (`ENROLLAGE`, `rs2284665`, `SevScale1E`, `SevScale2E`) in their original space. Using the same argument of the previous section where we have coded `rs2284665` as 0/1, such that to have three new covariates one for each `rs2284665` level , $\mathbf{z}_{3_0} = \mathbf{1}_{\{\mathbf{z}_3=0\}}$, $\mathbf{z}_{3_1} = \mathbf{1}_{\{\mathbf{z}_3=1\}}$, $\mathbf{z}_{3_2} = \mathbf{1}_{\{\mathbf{z}_3=2\}}$. .We then applied the BRBVS using both $\omega = \beta^2 i(\beta)$ and $\psi = |\beta|$ as metrics, the tuning parameters have been specified as follows: $k_{max} = 10$, $m = 328$, $\tau = 0.5$, $n.rep = 50$.

The combination Clayton copula (`C0`) and Proportional odds (`PO,PO`), this combination has been chosen using the procedure discussed in the previous section. The results of the BRBVS procedure are shown in Table 2.6.

Applying the $BRBVS$ based on ω, the set of important covariates are `SevScale1E`, `SevScale2E`, `ENROLLAGE` for the first margin, while `SevScale2E`, `SevScale1E`, `ENROLLAGE`, `GG`, `GT`, `TT` were included in the set of important covariates for the left eye. Using ψ, the genes `GG` and `GT` were incuded in the set of important covariates for the first margin, while for the second margin under the naive metric `ENROLLAGE` was not selected.

	1st margin (right eye)	2nd margin (left eye)	BIC	AIC	LogLik
$\mathcal{M}_\omega$	$\{\mathbf{z}_1, \mathbf{z}_4, \mathbf{z}_2\}$	$\{\mathbf{z}_4, \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_{3_0}, \mathbf{z}_{3_2}, \mathbf{z}_{3_1}\}$	4321.327	4218.106	-2085.819
$\mathcal{M}_\psi$	$\{\mathbf{z}_1, \mathbf{z}_4, \mathbf{z}_{3_0}, \mathbf{z}_{3_2}\}$	$\{\mathbf{z}_4, \mathbf{z}_{3_0}, \mathbf{z}_1, \mathbf{z}_{3_2}, \mathbf{z}_{3_1}\}$	4335.383	4222.823	-2086.074

Table 2.6: BRBVS results using clayton copula, porportional hazard margins $k_{max} = 10$, $m = 328$, $\tau = 0.5$, $n.rep = 80$ using $\omega = \beta^2 i(\beta)$ and $\psi = |\beta|$ as metrics. The covariates are ordered according to their importance. The BIC, AIC and LogLik were obtained by applying `gjrm()` function to the AREDS dataset perturbed using $\mathbf{Z}_i \overset{ind.}{\sim} \mathcal{N}(0,1)$ $i = 5, \ldots, 104$

2.5 Discussion

We have introduced a variable ranking procedure based on an information matrix for the class of Bivariate Copula Survival model(s). As discussed in the methodology section, the algorithm is based on a boostrap procedure executed to identify the sets of important covariates for the two (in)dependent time to events. The approach discussed makes a significant contribution in applied and methodological statistics not only as it represents the first concrete proposal of a variable selection procedure in the bivariate domain, but also for the introduction of a new metric. Considering that no references are available in the state of the art, the preliminary results discussed in this second chapter take on a completely different connotation. Overall both in simulation and real data setting, the Bivariate RBVS procedure performed surprisingly well, recovering the set of important variables for both time to events.

To the best of our knowledge, the Bivariate variable ranking proposed in this chapter is the first attempt to apply a ranking procedure to a bivariate domain. At the time of writing there is no package or implementation available in any programming language able to perform such task in a bivariate survival domain. This makes our proposal the only ready-to-use implementation of a bivariate survival variable ranking. Although the literature about the variable selection metrics is reasonably ample, as far as we know the Fisher information metric proposed is an absolute novelty in the literature, both in the univariate and multivariate domains.

Bibliography

Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Mathematical statistics and probability theory* (pp. 1–25). Springer.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *In: Petrov, B.N., Csaki, B.F. (eds.) Second International Symposium on Information Theory. Academiai Kiado, Budapest*.

Baba, T., Ohno-Matsui, K., Futagami, S., Yoshida, T., Yasuzumi, K., Kojima, A., Tokoro, T., & Mochizuki, M. (2003). Prevalence and characteristics of foveal retinal detachment without macular hole in high myopia. *American journal of ophthalmology*, 135(3), 338–342.

Baranowski, R., Chen, Y., & Fryzlewicz, P. (2020). Ranking-based variable selection for high-dimensional data. *Statistica Sinica*, 30(3), 1485–1516.

Bickel, P. J., Götze, F., & van Zwet, W. R. (2012). Resampling fewer than n observations: gains, losses, and remedies for losses. In *Selected works of Willem van Zwet* (pp. 267–297). Springer.

Chen, L. J., Liu, D., Tam, P., Chan, W. M., Liu, K., Chong, K., Lam, D., & Pang, C. P. (2006). Association of complement factor h polymorphisms with exudative age-related macular degeneration. *Mol Vis*, 12(5), 1536.

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.

Chen, M.-H., Tong, X., & Sun, J. (2007). The proportional odds model for multivariate interval-censored failure time data. *Statistics in medicine*, 26(28), 5147–5161.

Cho, H. & Fryzlewicz, P. (2012). High dimensional variable selection via tilting. *Journal of the Royal Statistical Society Series B*, 74(3), 593–622.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1), 141–151.

Cook, R. & Tolusso, D. (2009). Second-order estimating equations for the analysis of clustered current status data. *Biostatistics*, 10(4), 756–772.

Desboulets, L. D. D. (2018). A review on variable selection in regression analysis. *Econometrics*, 6(4), 45.

Donoho, D. L. et al. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture*, 1(2000), 32.

Fan, J. & Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *arXiv preprint math/0602133*.

Fan, J. & Lv, J. (2007). Sure independence screening for ultra-high dimensional feature space. *Journal of the Royal Statistical Society Series B*, 70.

Fan, J. & Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1), 101.

Fan, J. & Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics*, 38(6), 3567 – 3604.

Gazis, P. R., Levit, C., & Way, M. J. (2010). Viewpoints: A high-performance high-dimensional exploratory data analysis tool. *Publications of the Astronomical Society of the Pacific*, 122(898), 1518.

Giradi, D. & Holzinger, A. (2018). Dimensionality reduction for exploratory data analysis in daily medical research. In *Advanced Data Analytics in Health* (pp. 3–20). Springer.

Hall, P. & Miller, H. (2009). Using generalized correlation to effect variable selection in very high dimensional problems. *Journal of Computational and Graphical Statistics - J COMPUT GRAPH STAT*, 18.

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society Series B*, 55, 757–796.

He, X., Wang, L., & Hong, H. G. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*, 41(1), 342–369.

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.

Ivaniševic, M. & Stanic, R. (1990). Importance of fluorescein angiography in the early detection and therapy of diabetic retinopathy. *Ophthalmologica*, 201(1), 9–13.

Kim, M. Y. & Xue, X. (2002). The analysis of multivariate interval-censored survival data. *Statistics in Medicine*, 21(23), 3715–3726.

Kleinbaum, D. G. & Klein, M. (2004). *Survival analysis*. Springer.

Kor, C., Cheng, K., & Chen, Y. (2013). A method for analyzing clustered interval-censored data based on cox model. *Statistics in Medicine*, 32(5), 822–832.

Leitenstorfer, F. & Tutz, G. (2007). Generalized monotonic regression based on b-splines with an application to air pollution data. *Biostatistics*, 8(3), 654–673.

Leung, K.-M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. *Annual review of public health*, 18(1), 83–104.

Li, R., Zhong, W., & Zhu, L. (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107(499), 1129–1139. PMID: 25249709.

Liu, X.-R., Pawitan, Y., & Clements, M. (2018). Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, 27(5), 1531–1546.

Marra, G. & Radice, R. (2017). Bivariate copula additive models for location, scale and shape. *Computational Statistics & Data Analysis*, 112(C), 99–113.

Marra, G. & Radice, R. (2020). Copula link-based additive models for right-censored event time data. *Journal of the American Statistical Association*, 115(530), 886–895.

Marubini, E. & Valsecchi, M. G. (2004). *Analysing survival data from clinical trials and observational studies*, volume 15. John Wiley & Sons.

Nelsen, R. (2006). *An Introduction to Copulas*. Second Edition, Springer, New York.

Nocedal, J. & Wright, S. J. (2006). *Numerical Optimization*. Springer-Verlag, New York.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society: Series B*, 44(3), 414–422.

Panozzo, G. & Mercanti, A. (2004). Optical coherence tomography findings in myopic traction maculopathy. *Archives of ophthalmology*, 122(10), 1455–1460.

Panozzo, G. & Mercanti, A. (2007). Vitrectomy for myopic traction maculopathy. *Archives of Ophthalmology*, 125(6), 767–772.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2), 527–556.

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 and Data Analysis*, 175, 107550.

Pettifor, A., MacPhail, C., Nguyen, N., & Rosenberg, M. (2012). Can money prevent the spread of HIV? A review of cash payments for HIV prevention. *AIDS and Behavior*, 16(7), 1729–1738.

Pya, N. & Wood, S. N. (2015). Shape constrained additive models. *Statistics and computing*, 25(3), 543–559.

RJ, R., V, V., KI, R., CC, C., & J., T. (2007). Genetic markers and biomarkers for age-related macular degeneration. *Expert Rev Ophthalmol*, 2(3), 443–457.

Romeo, J., Meyer, R., & Gallardo, D. (2018). Bayesian bivariate survival snalysis using the power variance function copula. *Lifetime Data Analysis*, 24(2), 355–383.

Royston, P. & Parmar, M. (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.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, New York.

Salimiaghdam, N., Riazi-Esfahani, M., Fukuhara, P., Schneider, K., & Kenney, C. (2019). Age-related macular degeneration (amd): A review on its epidemiology and risk factors. *The Open Ophthalmology Journal*, 13, 90–99.

Schultz, N. M., Bhardwaj, S., Barclay, C., Gaspar, L., & Schwartz, J. (2021). Global burden of dry age-related macular degeneration: A targeted literature review. *Clinical Therapeutics*, 43(10), 1792–1818.

Shao, X. & Zhang, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association*, 109, 1302 – 1318.

Sklar, A. (1973). Random variables, joint distributions, and copulas. *Kybernetica*, 9(6), 449–460.

Sun, T. & Ding, Y. (2021). Copula-based semiparametric regression method for bivariate data under general interval censoring. *Biostatistics*, 22(2), 315–330.

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6), 2769 – 2794.

Wang, L., McMahan, C., Hudgens, M., & Qureshi, Z. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*, 72(1), 222–231.

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.

Wong, W. L., Su, X., Li, X., Cheung, C. M. G., Klein, R., Cheng, C.-Y., & Wong, T. Y. (2014). Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *The Lancet Global Health*, 2(2), e106–e116.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society Series B*, 65, 95–114.

Wood, S. N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics*, 48(4), 445–464.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction With R*. Second Edition, Chapman & Hall/CRC, London.

Yonekawa, Y., Miller, J. W., & Kim, I. K. (2015). Age-related macular degeneration: Advances in management and diagnosis. *Journal of clinical medicine*, 4(2), 343—359.

Zhang, J. (2004). A simple and efficient monotone smoother using smoothing splines. *Journal of Nonparametric Statistics*, 16(5), 779–796.

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.

Appendix:

Copula Link-Based Additive Models For Bivariate Time-To-Event Outcomes With General Censoring Scheme:

Computational Advances And Variable Ranking Procedures

Contents

Appendices Contents

Appendix A

It is discussed and presented the derivation of the likelihood, gradient and Hessian matrix of the model presented in chapter 1.

Appendix B

Insights about the procedures adopted and its implementation. Discussion bout the tools employed to double check the analytical quantities presented in Appendix A and some computational aspects are provided.

Appendix C

We discuss the use of mm() function, this with the help of some copula's theorems and proofs made by the author.

Appendix D

Tests and trust region debugs with the aim to evaluate the computational speed of the model under various copula settings.

Appendix E

Some asymptotic results needed to fully comprehend the asymptotic properties of the estimators presented in chapter 1.

Appendix F

Some model building procedure about the model presented in chapter 1 is discussed.

Appendix G

Some commented `R` code useful to fit the model discussed in chapter 1.

Appendix H

`R` code based on the AREDS analysis presented in chapter 1.

Appendix I

Mutual Information overview, properties and the proof of the concentration bound provided by the author.

Appendix J

BivariateRBVS Mutual Information based simulation study. (Note, the model(s) employed are the Flexible Copula Regression models).

Appendix K

Some properties of the RBVS stated in Baranowski et al.(2018) are presented.

Appendix A

Log-Likelihood, Gradient and Hessian Derivation

Declaration

This section has been used in a MSc thesis at University College London, UK. Furthermore, part of the following section it is published content in *Copula link-based additive models for bivariate time-to-event outcomes with general censoring scheme, Computational Statistics & Data Analysis, Danilo Petti, Alessia Eletti, Giampiero Marra, Rosalba Radice, Volume 175, 2022, 107550, ISSN 0167-9473, https://doi.org/10.1016/j.csda.2022.107550.* I would like to thank Alessia Eletti for having contributed in the derivation of the likelihood and Giampiero Marra, Rosalba Radice and Alessia Eletti for the feedback of my derivation of the gradient and hessian matrix

A.1 Complete log-likelihood function

In this section we will present the full derivation of the complete log-likelihood function. First the log-likelihood composed of its sixteen pieces will be presented, to then guide the reader towards its derivation. According to the standard notation, we will denote with ℓ the logarithm of the likelihood function.

$$\ell(\boldsymbol{\delta}) = \sum_{i=1}^{n} \delta_{U_{1i}} \delta_{U_{2i}} \log \left[\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i})) \partial G_2(\eta_{2i}(t_{2i}))} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}} \right] +$$
$$+ \delta_{R_{1i}} \delta_{R_{2i}} \log \left[C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\} \right] +$$
$$+ \delta_{L_{1i}} \delta_{L_{2i}} \log \left[1 - G_1(\eta_{1i}(l_{1i})) - G_2(\eta_{2i}(l_{2i})) + C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\} \right] +$$

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

A.1.1 Derivation

As per usual notation we have for $\nu = 1, 2$ and $i = 1, ..., n$

- $\delta_{U_{\nu i}} = 1$ if uncensored and $\delta_{U_{\nu i}} = 0$ otherwise
- $\delta_{L_{\nu i}} = 1$ if left-censored and $\delta_{L_{\nu i}} = 0$ otherwise
- $\delta_{R_{\nu i}} = 1$ if right-censored and $\delta_{R_{\nu i}} = 0$ otherwise
- $\delta_{I_{\nu i}} = 1$ if interval-censored and $\delta_{I_{\nu i}} = 0$ otherwise

In the bivariate case the log-likelihood function is made of 16 terms corresponding to the following combinations of the indicator functions

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

The derivation of each of these terms follows.

- T_{1i} uncensored and T_{2i} uncensored (note: 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

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

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

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

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

$$\int_{r_{2i}}^{+\infty} f(t_{1i}, y) dy = \int_0^{+\infty} f(t_{1i}, y) dy - \int_0^{r_{2i}} f(t_{1i}, y) dy =$$
$$= f_1(t_{1i}) - \frac{\partial}{\partial t_{1i}} F(t_{1i}, r_{2i}) = f_1(t_{1i}) - \frac{\partial}{\partial t_{1i}} [1 - S_1(t_{1i}) - S_2(t_{2i}) + S(t_{1i}, r_{2i})] =$$
$$= f_1(t_{1i}) - f_1(t_{1i}) - \frac{\partial}{\partial t_{1i}} S(t_{1i}, r_{2i}) = -\frac{\partial}{\partial t_{1i}} C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\} =$$
$$= -\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}$$

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

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

- T_{1i} uncensored and T_{2i} interval-censored (the symmetric case follows by switching the subscripts where needed)

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

- T_{1i} right-censored and T_{2i} left-censored (the symmetric case follows by switching the subscripts where needed)

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

- T_{1i} right-censored and T_{2i} interval-censored (the symmetric case follows by switching the subscripts where needed)

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

- T_{1i} left-censored and T_{2i} interval-censored (the symmetric case follows by switching the subscripts where needed)

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

A.2 The Gradient

In the following the derivatives of the complete log-likelihood with respect to the parameters of the survival of the first time-to-event, $\boldsymbol{\beta}_1$, to the the parameters of the survival of the second time-to-event, $\boldsymbol{\beta}_2$, and to the parameters of the copula coefficient, $\boldsymbol{\beta}_3$, are reported separately. The quantities of the type $D_{\zeta;l_{1i},l_{2i}}$, where $\zeta \in \mathbb{N}$ and $i = 1, \ldots, n$ have been introduced to make more compact and readable the expression of the Gradient.

Derivative with respect to $\boldsymbol{\beta}_1$

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

$$
\begin{aligned}
&\delta_{R_{1i}}\delta_{U_{2i}}D^{-1}_{1;\delta_{R_{1i}}\delta_{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\boldsymbol{\beta_1}}\right\}\\
&+\delta_{U_{1i}}\delta_{L_{2i}}\Bigg\{D^{-1}_{1;\delta_{U_{1i}}\delta_{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\boldsymbol{\beta_1}}+\\
&\qquad+D^{-1}_{2;\delta_{U_{1i}}\delta_{L_{2i}}}G_1''(\eta_{1i}(t_{1i}))\frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}}\\
&\qquad+D^{-1}_{3;\delta_{U_{1i}}\delta_{L_{2i}}}\frac{\partial^2\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\boldsymbol{\beta_1}}\Bigg\}+\\
&+\delta_{L_{1i}}\delta_{U_{2i}}\left\{D^{-1}_{1;\delta_{L_{1i}}\delta_{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\boldsymbol{\beta_1}}\right\}+\\
&+\delta_{U_{1i}}\delta_{I_{2i}}\Bigg\{D^{-1}_{1;\delta_{U_{1i}}\delta_{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}))^2}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\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\boldsymbol{\beta_1}}\Bigg]+\\
&\qquad+D^{-1}_{2;\delta_{U_{1i}}\delta_{I_{2i}}}\left[G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}}\right]+\\
&\qquad+D^{-1}_{3;\delta_{U_{1i}}\delta_{I_{2i}}}\left[\frac{\partial^2\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\boldsymbol{\beta_1}}\right]\Bigg\}+\\
&+\delta_{I_{1i}}\delta_{U_{2i}}D^{-1}_{1;\delta_{I_{1i}}\delta_{U_{2i}}}\Bigg\{\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\boldsymbol{\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\boldsymbol{\beta_1}}\Bigg\}+\\
&+\delta_{R_{1i}}\delta_{L_{2i}}D^{-1}_{\delta_{R_{1i}}\delta_{L_{2i}}}\left\{G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\boldsymbol{\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\boldsymbol{\beta_1}}\right\}+\\
&+\delta_{L_{1i}}\delta_{R_{2i}}D^{-1}_{\delta_{L_{1i}}\delta_{R_{2i}}}\left\{-\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\boldsymbol{\beta_1}}\right\}+\\
&+\delta_{R_{1i}}\delta_{I_{2i}}D^{-1}_{\delta_{R_{1i}}\delta_{I_{2i}}}\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\boldsymbol{\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\boldsymbol{\beta_1}}\Bigg\}+\\
&+\delta_{I_{1i}}\delta_{R_{2i}}D^{-1}_{\delta_{I_{1i}}\delta_{R_{2i}}}\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\boldsymbol{\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\boldsymbol{\beta_1}}\Bigg\}+\\
&+\delta_{L_{1i}}\delta_{I_{2i}}D^{-1}_{\delta_{L_{1i}}\delta_{I_{2i}}}\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\boldsymbol{\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\boldsymbol{\beta_1}}\Bigg\}
\end{aligned}
$$

$$\delta_{I_{1i}}\delta_{L_{2i}}D^{-1}_{\delta_{I_{1i}}\delta_{L_{2i}}}\left\{G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\boldsymbol{\beta_1}}-G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\boldsymbol{\beta_1}}+\right.$$
$$+\frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\boldsymbol{\beta_1}}+$$
$$\left.-\frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i}))}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\boldsymbol{\beta_1}}\right\}$$

Derivative with respect to $\boldsymbol{\beta}_2$

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

$$
\begin{aligned}
&+ \delta_{U_{1i}}\delta_{I_{2i}} D^{-1}_{1;\delta_{U_{1i}}\delta_{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 \boldsymbol{\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 \boldsymbol{\beta_2}} \Bigg\} \\
&+ \delta_{I_{1i}}\delta_{U_{2i}} \Bigg\{ D^{-1}_{1;\delta_{I_{1i}}\delta_{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}))^2} \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial \boldsymbol{\beta_2}} + \\
&\qquad - \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 \boldsymbol{\beta_2}} \Bigg] + \\
&\qquad + D^{-1}_{2;\delta_{I_{1i}}\delta_{U_{2i}}} G_2''(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial \boldsymbol{\beta_2}} \\
&\qquad + D^{-1}_{3;\delta_{I_{1i}}\delta_{U_{2i}}} \frac{\partial^2 \eta_{2i}(t_{2i})}{\partial t_{2i}\partial \boldsymbol{\beta_2}} \Bigg\} \\
&+ \delta_{R_{1i}}\delta_{L_{2i}} D^{-1}_{\delta_{R_{1i}}\delta_{L_{2i}}} \Bigg\{ - \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 \boldsymbol{\beta_2}} \Bigg\} \\
&+ \delta_{L_{1i}}\delta_{R_{2i}} D^{-1}_{\delta_{L_{1i}}\delta_{R_{2i}}} \Bigg\{ G_2'(\eta_{2i}(r_{2i})) \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\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 \boldsymbol{\beta_2}} \Bigg\} \\
&+ \delta_{R_{1i}}\delta_{I_{2i}} D^{-1}_{\delta_{R_{1i}}\delta_{I_{2i}}} \Bigg\{ \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 \boldsymbol{\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 \boldsymbol{\beta_2}} \Bigg\} + \\
&\delta_{I_{1i}}\delta_{R_{2i}} D^{-1}_{\delta_{I_{1i}}\delta_{R_{2i}}} \Bigg\{ \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 \boldsymbol{\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 \boldsymbol{\beta_2}} \Bigg\} + \\
&+ \delta_{L_{1i}}\delta_{I_{2i}} D^{-1}_{\delta_{L_{1i}}\delta_{I_{2i}}} \Bigg\{ G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} - G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\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 \boldsymbol{\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 \boldsymbol{\beta_2}} \Bigg\} + \\
&\delta_{I_{1i}}\delta_{L_{2i}} D^{-1}_{\delta_{I_{1i}}\delta_{L_{2i}}} \Bigg\{ \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 \boldsymbol{\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 \boldsymbol{\beta_2}} \Bigg\}
\end{aligned}
$$

Derivative with respect to $\boldsymbol{\beta}_3$

$$
\begin{aligned}
\frac{\partial \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_3} =& \delta_{U_{1i}}\delta_{U_{2i}} D^{-1}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \left\{ \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} \\
&+ \delta_{R_{1i}}\delta_{R_{2i}} D^{-1}_{\delta_{R_{1i}}\delta_{R_{2i}}} \left\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\}
\end{aligned}
$$

$$
\begin{aligned}
&+ \delta_{L_{1i}}\delta_{L_{2i}} D^{-1}_{\delta_{L_{1i}}\delta_{L_{2i}}} \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}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} + \\
&+ \delta_{I_{1i}}\delta_{I_{2i}} D^{-1}_{\delta_{I_{1i}}\delta_{I_{2i}}} \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}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\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}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\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}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\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}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} + \\
&+ \delta_{U_{1i}}\delta_{R_{2i}} D^{-1}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \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}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} \\
&+ \delta_{R_{1i}}\delta_{U_{2i}} D^{-1}_{1;\delta_{R_{1i}}\delta_{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 m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} \\
&+ \delta_{U_{1i}}\delta_{L_{2i}} D^{-1}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \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}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} + \\
&+ \delta_{L_{1i}}\delta_{U_{2i}} D^{-1}_{1;\delta_{L_{1i}}\delta_{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}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} + \\
&+ \delta_{U_{1i}}\delta_{I_{2i}} D^{-1}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \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}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} + \right. \\
&\qquad \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}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} + \\
&+ \delta_{I_{1i}}\delta_{U_{2i}} D^{-1}_{1;\delta_{I_{1i}}\delta_{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 m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} + \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}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} \\
&+ \delta_{R_{1i}}\delta_{L_{2i}} D^{-1}_{\delta_{R_{1i}}\delta_{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}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} + \\
&+ \delta_{L_{1i}}\delta_{R_{2i}} D^{-1}_{\delta_{L_{1i}}\delta_{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}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} + \\
&+ \delta_{R_{1i}}\delta_{I_{2i}} D^{-1}_{\delta_{R_{1i}}\delta_{I_{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}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\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}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} +
\end{aligned}
$$

$$
\begin{aligned}
&+ \delta_{I_{1i}}\delta_{R_{2i}} D^{-1}_{\delta_{I_{1i}}\delta_{R_{2i}}} \left\{ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} + \right. \\
&\qquad \left. - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} \\
&+ \delta_{L_{1i}}\delta_{I_{2i}} D^{-1}_{\delta_{L_{1i}}\delta_{I_{2i}}} \left\{ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} + \right. \\
&\qquad \left. - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\} + \\
&+ \delta_{I_{1i}}\delta_{L_{2i}} D^{-1}_{\delta_{I_{1i}}\delta_{L_{2i}}} \left\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} + \right. \\
&\qquad \left. - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial \eta_{3i}(\mathbf{x}_{3i}, \boldsymbol{\beta}_3)}{\partial \boldsymbol{\beta}_3} \right\}
\end{aligned}
$$

Symbols

$$
\begin{aligned}
D_{1;\delta_{U_{1i}}\delta_{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;\delta_{U_{1i}}\delta_{U_{2i}}} &= G_1'(\eta_{1i}(t_{1i})) \\
D_{3;\delta_{U_{1i}}\delta_{U_{2i}}} &= G_2'(\eta_{2i}(t_{2i})) \\
D_{4;\delta_{U_{1i}}\delta_{U_{2i}}} &= \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \\
D_{5;\delta_{U_{1i}}\delta_{U_{2i}}} &= \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}} \\
D_{\delta_{R_{1i}}\delta_{R_{2i}}} &= C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\} \\
D_{\delta_{L_{1i}}\delta_{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_{\delta_{I_{1i}}\delta_{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;\delta_{U_{1i}}\delta_{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;\delta_{U_{1i}}\delta_{R_{2i}}} &= -G_1'(\eta_{1i}(t_{1i})) \\
D_{3;\delta_{U_{1i}}\delta_{R_{2i}}} &= \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \\
D_{1;\delta_{R_{1i}}\delta_{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;\delta_{R_{1i}}\delta_{u_{2i}}} &= -G_2'(\eta_{2i}(t_{2i})) \\
D_{3;\delta_{R_{1i}}\delta_{U_{2i}}} &= \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}} \\
D_{1;\delta_{U_{1i}}\delta_{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;\delta_{U_{1i}}\delta_{L_{2i}}} &= G_1'(\eta_{1i}(t_{1i})) \\
D_{3;\delta_{U_{1i}}\delta_{L_{2i}}} &= \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \\
D_{1;\delta_{L_{1i}}\delta_{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) \\
D_{2;\delta_{L_{1i}}\delta_{U_{2i}}} &= G_2'(\eta_{2i}(t_{2i}))
\end{aligned}
$$

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

A.3 Hessian of complete log-likelihood

Symbols for the Hessian

In this section are introduced some quantities useful in order to express the Hessian in a more concise and clear way. In the following paragraphs, all the quantities useful for the derivation of the Hessian have been derived. In detail, the first and mixed derivatives of the quantities of the type $D_{\zeta;l_{1i},l_{2i}}$ were introduced, where $\zeta \in \mathbb{N}$ and $i = 1, \ldots, n$. The introduction of these quantities is necessary in order to make the writing of the Hessian much more compact and readable for the reader.

First Derivative respect with β_1

$$
\begin{aligned}
\frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\beta_1}}\right] \\
\frac{\partial D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_1}} &= \left[G_1''(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \boldsymbol{\beta_1}}\right] \\
\frac{\partial D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_1}} &= \left[\frac{\partial^2 \eta_{1i}(t_{1i})}{\partial t_{1i} \partial \boldsymbol{\beta_1}}\right] \\
\frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\beta_1}}\right\} \\
\frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta_1}} &= \left\{-G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial \eta_{1i}(l_{1i})}{\partial \boldsymbol{\beta_1}} + \right. \\
&\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 \boldsymbol{\beta_1}}\right\} \\
\frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\beta_1}} + \right. \\
&\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 \boldsymbol{\beta_1}} +
\end{aligned}
$$

$$- \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}} +$$

$$+ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}} \Bigg\}$$

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

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

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

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

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

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

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

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

$$\frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta_1}} = \Bigg[\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \boldsymbol{\beta_1}} +$$

$$- \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial \eta_{1i}(t_{1i})}{\partial \boldsymbol{\beta_1}} \Bigg]$$

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

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

$$\frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_1}} = \Bigg\{ \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))\partial G_2(\eta_{2i}(t_{2i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}} +$$

$$- \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(l_{1i})\partial G_2(\eta_{2i}(t_{2i}))} \cdot G_1'(\eta_{1i}(l_{1i}) \cdot \frac{\partial \eta_{1i}(l_{1i})}{\partial \boldsymbol{\beta_1}} \Bigg\}$$

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

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

$$\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta_1}} = \Bigg\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}} +$$

$$- \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}} \Bigg\}$$

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

First Derivative respect to β_2

$$
\begin{aligned}
\frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\beta_2}} \\
\frac{\partial D_{3;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_2}} &= G_2''(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial \boldsymbol{\beta_2}} \\
\frac{\partial D_{5;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_2}} &= \frac{\partial^2 \eta_{2i}(t_{2i})}{\partial t_{2i}\partial \boldsymbol{\beta_2}} \\
\frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\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_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} \right\} \\
\frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta_2}} &= \left\{ - G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\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 \boldsymbol{\beta_2}} \right\} \\
\frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\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 \boldsymbol{\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_1'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} + \\
&\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_1'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} \right\} \\
\frac{\partial D_{\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\beta_2}} \right\} \\
\frac{\partial D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\beta_2}} \right] \\
\frac{\partial D_{2;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_2}} &= \left[- G_2''(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial \boldsymbol{\beta_2}} \right] \\
\frac{\partial D_{3;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_2}} &= \frac{\partial^2 \eta_{2i}(t_{2i})}{\partial t_{2i}\partial \boldsymbol{\beta_2}} \\
\frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\beta_2}} \\
\frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\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 \boldsymbol{\beta_2}} \right] \\
\frac{\partial D_{2;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_2}} &= \left[G_2''(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial \boldsymbol{\beta_2}} \right] \\
\frac{\partial D_{3;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_1}} &= \left[\frac{\partial^2 \eta_{2i}(t_{2i})}{\partial t_{2i}\partial \boldsymbol{\beta_2}} \right]
\end{aligned}
$$

$$\frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta_2}} = \Bigg\{ \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} + \\ - \frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} \Bigg\}$$

$$\frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta_2}} = \Bigg[\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial \boldsymbol{\beta_2}} + \\ - \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial \eta_{2i}(t_{2i})}{\partial \boldsymbol{\beta_2}} \Bigg] +$$

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

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

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

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

$$\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta_2}} = \Bigg\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})} \cdot G_2'(\eta_{2i}(l_{2i}) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} + \\ - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} \Bigg\}$$

$$\frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta_2}} = \Bigg\{ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} + \\ - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} \Bigg\}$$

$$\frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta_2}} = \Bigg\{ G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} - G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} + \\ + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} + \\ - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} \Bigg\}$$

$$\frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta_2}} = \Bigg\{ \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} + \\ - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i}) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} \Bigg\}$$

First Derivative respect to β_3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Cross derivative: $\partial^2 D's/\partial\beta_1\partial\beta_1$

$$
\begin{aligned}
\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\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\boldsymbol{\beta_1}}\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}))^2\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1''(\eta_{1i}(t_{1i}))\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}}\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}))^2\partial G_2(\eta_{2i}(t_{2i}))}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}\\
\frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T} &= \Bigg\{G_1'''(\eta_{1i}(t_{1i}))\cdot\left(\frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}}\right)^2+G_1''(\eta_{1i}(t_{1i}))\cdot\frac{\partial^2\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}\Bigg\}\\
\frac{\partial^2 D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T} &= \frac{\partial^3\eta_{1i}(t_{1i})}{\partial t_{1i}\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}\\
\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\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\boldsymbol{\beta_1}}\right)^2+\\
&\quad+\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\boldsymbol{\beta_1}}\right)^2+\\
&\quad+\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\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}\Bigg\}+\\
\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T} &= \Bigg\{-G_1''(\eta_{1i}(l_{1i}))\cdot\left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\boldsymbol{\beta_1}}\right)^2-G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial^2\eta_{1i}(l_{1i})}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}+\\
&\quad+\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\boldsymbol{\beta_1}}\right)^2+\\
&\quad+\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\boldsymbol{\beta_1}}\right)^2+\\
&\quad+\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\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}\Bigg\}\\
\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\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\boldsymbol{\beta_1}}\right)^2+\\
&\quad+\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\boldsymbol{\beta_1}}\right)^2\\
&\quad+\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\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}+\\
&\quad-\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\boldsymbol{\beta_1}}\right)^2+\\
&\quad-\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\boldsymbol{\beta_1}}\right)^2+\\
&\quad-\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\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}+\\
&\quad-\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\boldsymbol{\beta_1}}\right)^2+
\end{aligned}
$$

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

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

$$
\frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta_1} \partial \boldsymbol{\beta_1}^T} = \Bigg\{ - G_1'''(\eta_{1i}(t_{1i})) \cdot \left(\frac{\partial \eta_{1i}(t_{1i})}{\partial \boldsymbol{\beta_1}}\right)^2 - G_1''(\eta_{1i}(t_{1i})) \cdot \frac{\partial^2 \eta_{1i}(t_{1i})}{\partial \boldsymbol{\beta_1} \partial \boldsymbol{\beta_1}} \Bigg\}
$$

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

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

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

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

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

$$\frac{\partial^2 D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T} = \Bigg\{ + \frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))^2} \cdot \left(G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial\eta_{1i}(l_{1i})}{\partial\boldsymbol{\beta_1}}\right)^2 + \\ + \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))} \cdot G_1''(\eta_{1i}(l_{1i})) \cdot \left(\frac{\partial\eta_{1i}(l_{1i})}{\partial\boldsymbol{\beta_1}}\right)^2 + \\ + \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))} \cdot G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial^2\eta_{1i}(l_{1i})}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_1}^T}\Bigg\}$$

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

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

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

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

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

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

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

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

$$- \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1''(\eta_{1i}(r_{1i}))\left(\frac{\partial \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}}\right)^2 + \\ - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(r_{1i}))} \cdot G_1'(\eta_{1i}(r_{1i}))\frac{\partial^2 \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}\partial \boldsymbol{\beta_1}^T}\Bigg\}$$

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

$$\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta_1}\partial \boldsymbol{\beta_1}^T} = \Bigg\{G_1''(\eta_{1i}(l_{1i})) \cdot \left(\frac{\partial \eta_{1i}(l_{1i})}{\partial \boldsymbol{\beta_1}}\right)^2 + G_1'(\eta_{1i}(l_{1i})) \cdot \frac{\partial^2 \eta_{1i}(l_{1i})}{\partial \boldsymbol{\beta_1}\partial \boldsymbol{\beta_1}^T} + \\ - G_1''(\eta_{1i}(r_{1i})) \cdot \left(\frac{\partial \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}}\right)^2 - G_1'(\eta_{1i}(r_{1i})) \cdot \frac{\partial^2 \eta_{1i}(r_{1i})}{\partial \boldsymbol{\beta_1}\partial \boldsymbol{\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 \boldsymbol{\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 \boldsymbol{\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 \boldsymbol{\beta_1}\partial \boldsymbol{\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 \boldsymbol{\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 \boldsymbol{\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 \boldsymbol{\beta_1}\partial \boldsymbol{\beta_1}^T}\Bigg\}$$

Cross derivative: $\partial^2 D's/\partial\beta_1\partial\beta_2$

$$\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_2}^T} = \frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2 \partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2'(\eta_{2i}(t_{2i}) \cdot \frac{\partial\eta_{2i}(t_{2i})}{\partial\boldsymbol{\beta_2}} \cdot G_1'(\eta_{1i}(t_{1i})) \cdot \frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Cross derivative: $\partial^2 D's/\partial\beta_1\partial\beta_3$

$$\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_3}^T}=\frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}}$$

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

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

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

$$\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_3}^T}=\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\cdot G_1'(\eta_{1i}(t_{1i})\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}}$$

$$\frac{\partial^2 D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_3}^T}=\frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(r_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\cdot G_1'(\eta_{1i}(r_{1i}))\cdot\frac{\partial\eta_{1i}(r_{1i})}{\partial\boldsymbol{\beta_1}}$$

$$\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_3}^T}=\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))^2\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial\boldsymbol{\beta_1}}$$

$$\frac{\partial^2 D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_1}\partial\boldsymbol{\beta_3}^T}=\frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial G_1(\eta_{1i}(l_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\cdot G_1'(\eta_{1i}(l_{1i}))\cdot\frac{\partial\eta_{1i}(l_{1i})}{\partial\boldsymbol{\beta_1}}$$

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

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

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

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

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

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

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

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

Cross derivative: $\partial^2 D's/\partial\beta_2\partial\beta_3$

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

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

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta_2} \partial \boldsymbol{\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}(\boldsymbol{x_{3i}}, \boldsymbol{\beta_3})}{\partial \boldsymbol{\beta_3}} \cdot G_2'(\eta_{2i}(l_{2i}) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\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}(\boldsymbol{x_{3i}}, \boldsymbol{\beta_3})}{\partial \boldsymbol{\beta_3}} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} \\
\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta_2} \partial \boldsymbol{\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}(\boldsymbol{x_{3i}}, \boldsymbol{\beta_3})}{\partial \boldsymbol{\beta_3}} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\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}(\boldsymbol{x_{3i}}, \boldsymbol{\beta_3})}{\partial \boldsymbol{\beta_3}} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\beta_2}} \\
\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta_2} \partial \boldsymbol{\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}(\boldsymbol{x_{3i}}, \boldsymbol{\beta_3})}{\partial \boldsymbol{\beta_3}} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial \eta_{2i}(r_{2i})}{\partial \boldsymbol{\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}(\boldsymbol{x_{3i}}, \boldsymbol{\beta_3})}{\partial \boldsymbol{\beta_3}} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}} \\
\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta_2} \partial \boldsymbol{\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}(\boldsymbol{x_{3i}}, \boldsymbol{\beta_3})}{\partial \boldsymbol{\beta_3}} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\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}(\boldsymbol{x_{3i}}, \boldsymbol{\beta_3})}{\partial \boldsymbol{\beta_3}} \cdot G_2'(\eta_{2i}(l_{2i}) \cdot \frac{\partial \eta_{2i}(l_{2i})}{\partial \boldsymbol{\beta_2}}
\end{aligned}
$$

Cross derivative: $\partial^2 D's/\partial\beta_3\partial\beta_3$

$$\frac{\partial^2 D_{\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T} = \Bigg\{ \frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})^2} \cdot \left(m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}} \right)^2 + \\ + \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m''(\eta_{3i}) \cdot \left(\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}} \right)^2 + \\ + \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial^2\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T} \Bigg\}$$

$$\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T} = \Bigg\{ \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})^2} \cdot \left(m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}} \right)^2 \\ + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m''(\eta_{3i}) \cdot \left(\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}} \right)^2 \\ + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial^2\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T} \Bigg\}$$

$$\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T} = \Bigg\{ \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})^2} \cdot \left(m'(\eta_{3i}) \cdot \frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}} \right)^2 \\ + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m''(\eta_{3i}) \cdot \left(\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}} \right)^2 \\ + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial m(\eta_{3i})} \cdot m'(\eta_{3i}) \cdot \frac{\partial^2\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T} \Bigg\}$$

$$\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T} + \\ - \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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\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}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T} +$$

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

$$\frac{\partial^2 D_{\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T}=\left\{\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\right)^2+\right.$$
$$+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\right)^2+$$
$$\left.+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T}\right\}$$

$$\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T}=\left\{\frac{\partial^3 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\right)^2+\right.$$
$$+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\right)^2+$$
$$\left.+\frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T}\right\}$$

$$\frac{\partial^2 D_{\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T}=\left\{\frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\right)^2+\right.$$
$$+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\right)^2+$$
$$\left.+\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T}\right\}$$

$$\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T}=\left\{\frac{\partial^3 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})^2}\cdot\left(m'(\eta_{3i})\cdot\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\right)^2+\right.$$
$$+\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m''(\eta_{3i})\cdot\left(\frac{\partial\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}}\right)^2$$
$$\left.+\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(t_{2i}))\}}{\partial G_2(\eta_{2i}(t_{2i}))\partial m(\eta_{3i})}\cdot m'(\eta_{3i})\cdot\frac{\partial^2\eta_{3i}(\boldsymbol{x_{3i}},\boldsymbol{\beta_3})}{\partial\boldsymbol{\beta_3}\partial\boldsymbol{\beta_3}^T}\right\}$$

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

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

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

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

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

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

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

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

Cross derivative: $\partial^2 D's/\partial\beta_2\partial\beta_2$

$$\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} = \left\{\frac{\partial^4 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^3} \cdot \left(G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial\eta_{2i}(t_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 + \right.$$
$$+ \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2''(\eta_{2i}(t_{2i})) \cdot \left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 +$$
$$\left. + \frac{\partial^3 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))\partial G_2(\eta_{2i}(t_{2i}))^2} \cdot G_2'(\eta_{2i}(t_{2i})) \cdot \frac{\partial^2\eta_{2i}(t_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}\right\}$$

$$\frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} = G_2'''(\eta_{2i}(t_{2i})) \cdot \left(\frac{\partial\eta_{2i}(t_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 + G_2''(\eta_{2i}(t_{2i})) \cdot \frac{\partial^2\eta_{2i}(t_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}$$

$$\frac{\partial^2 D_{5;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} = \frac{\partial^3\eta_{2i}(t_{2i})}{\partial t_{2i}\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}$$

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

$$\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} = \left\{ - G_2''(\eta_{2i}(l_{2i})) \cdot \left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 - G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial^2\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} + \right.$$
$$+ \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))^2} \cdot \left(G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 +$$
$$+ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2''(\eta_{2i}(l_{2i})) \cdot \left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 +$$
$$\left. + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial^2\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}\right\}$$

$$\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} = \left\{\frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))^2} \cdot \left(G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 + \right.$$
$$+ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2''(\eta_{2i}(l_{2i})) \cdot \left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 +$$
$$+ \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))} \cdot G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial^2\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} +$$
$$- \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2} \cdot \left(G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 +$$
$$- \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2''(\eta_{2i}(r_{2i})) \cdot \left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 +$$
$$- \frac{\partial C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))} \cdot G_2'(\eta_{2i}(r_{2i})) \cdot \frac{\partial^2\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} +$$
$$- \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i}))^2} \cdot \left(G_2'(\eta_{2i}(l_{2i})) \cdot \frac{\partial\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 +$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} = \Bigg\{ & G''_2(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 + G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}+ \\
& - \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 \\
& - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G''_2(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 \\
& - \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})^2}\cdot\left(G'_2(\eta_{2i}(l_{2i})\cdot\frac{\partial\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})}\cdot G''_2(\eta_{2i}(l_{2i})\cdot\left(\frac{\partial\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_2(\eta_{2i}(l_{2i})}\cdot G'_2(\eta_{2i}(l_{2i})\cdot\frac{\partial^2\eta_{2i}(l_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}+ \\
& - \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2+ \\
& - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G''_2(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2+ \\
& - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} = \Bigg\{ & \frac{\partial^2 C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G''_2(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2+ \\
& + \frac{\partial C\{G_1(\eta_{1i}(l_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T} \\
& - \frac{\partial^2 C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))^2}\cdot\left(G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 \\
& - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G''_2(\eta_{2i}(r_{2i}))\cdot\left(\frac{\partial\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}}\right)^2 \\
& - \frac{\partial C\{G_1(\eta_{1i}(r_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_2(\eta_{2i}(r_{2i}))}\cdot G'_2(\eta_{2i}(r_{2i}))\cdot\frac{\partial^2\eta_{2i}(r_{2i})}{\partial\boldsymbol{\beta_2}\partial\boldsymbol{\beta_2}^T}\Bigg\}
\end{aligned}
$$

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

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

The Hessian

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} =& \delta_{U_{1i}}\delta_{U_{2i}} \Bigg\{ (-1) D^{-2}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \\
& (-1) D^{-2}_{2;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{2;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \\
& (-1) D^{-2}_{4;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{4;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{R_{1i}}\delta_{R_{2i}} \Bigg\{ (-1) D^{-2}_{\delta_{R_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{\delta_{R_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{L_{1i}}\delta_{L_{2i}} \Bigg\{ (-1) D^{-2}_{\delta_{L_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{\delta_{L_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{I_{1i}}\delta_{I_{2i}} \Bigg\{ (-1) D^{-2}_{\delta_{I_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{\delta_{I_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{U_{1i}}\delta_{R_{2i}} \Bigg\{ (-1) D^{-2}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \\
& (-1) D^{-2}_{2;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{2;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \\
& (-1) D^{-2}_{3;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{3;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{3;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{R_{1i}}\delta_{U_{2i}} \Bigg\{ (-1) D^{-2}_{1;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{1;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{U_{1i}}\delta_{L_{2i}} \Bigg\{ (-1) D^{-2}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \\
& (-1) D^{-2}_{2;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{2;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{2;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \\
& (-1) D^{-2}_{3;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{3;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{3;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{L_{1i}}\delta_{U_{2i}} \Bigg\{ (-1) D^{-2}_{1;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{1;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} \\
& + \delta_{U_{1i}}\delta_{I_{2i}} \Bigg\{ (-1) D^{-2}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \\
& (-1) D^{-2}_{2;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{2;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{2;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} + \\
& (-1) D^{-2}_{3;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{3;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{3;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{I_{1i}}\delta_{U_{2i}} \Bigg\{ (-1) D^{-2}_{1;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{1;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} + \\
& + \delta_{R_{1i}}\delta_{L_{2i}} \Bigg\{ (-1) D^{-2}_{\delta_{R_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} \right)^2 + D^{-1}_{\delta_{R_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T} \Bigg\} +
\end{aligned}
$$

$$+ \delta_{L_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1}\right)^2 + D^{-1}_{\delta_{L_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T}\right\}+$$
$$+ \delta_{R_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1}\right)^2 + D^{-1}_{\delta_{R_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T}\right\}+$$
$$+ \delta_{I_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1}\right)^2 + D^{-1}_{\delta_{I_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T}\right\}+$$
$$+ \delta_{L_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1}\right)^2 + D^{-1}_{\delta_{L_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T}\right\}+$$
$$+ \delta_{I_{1i}}\delta_{L_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1}\right)^2 + D^{-1}_{\delta_{I_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1^T}\right\}$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} =& \delta_{U_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{2;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{2;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{4;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{4;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{R_{1i}}\delta_{R_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{\delta_{R_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{L_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{\delta_{L_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{I_{1i}}\delta_{I_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{\delta_{I_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{U_{1i}}\delta_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{2;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{2;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{3;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial D_{3;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{3;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{3;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{R_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{1;\delta R_{1i}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{1;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta R_{1i}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{U_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{2;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{2;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{3;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial D_{3;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{3;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{3;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{L_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{1;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} + \\
& \delta_{U_{1i}}\delta_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{2;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{2;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{3;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial D_{3;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{3;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{3;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{I_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{1;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} \\
& \delta_{R_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1} + D^{-1}_{\delta_{R_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2^T} \Bigg\} +
\end{aligned}
$$

$$\delta_{L_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_2}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{L_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2^T}\right\}+$$

$$\delta_{R_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_2}\cdot\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{R_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2^T}\right\}+$$

$$\delta_{I_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_2}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{I_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2^T}\right\}+$$

$$\delta_{L_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_2}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{L_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2^T}\right\}+$$

$$\delta_{I_{1i}}\delta_{L_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta}_2}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{I_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2^T}\right\}+$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_3^T} =&\delta_{U_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{1;\delta_{U_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}+\\
&(-1)D^{-2}_{2;\delta_{U_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{2;\delta_{U_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}+\\
&(-1)D^{-2}_{4;\delta_{U_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{4;\delta_{U_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial^2 D_{4;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{R_{1i}}\delta_{R_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{\delta_{R_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{L_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{\delta_{L_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{I_{1i}}\delta_{I_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{\delta_{I_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{U_{1i}}\delta_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{1;\delta_{U_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}+\\
&(-1)D^{-2}_{2;\delta_{U_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{2;\delta_{U_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}+\\
&(-1)D^{-2}_{3;\delta_{U_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial D_{4;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{4;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{3;\delta_{U_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{R_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{R_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{1;\delta R_{1i}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{1;\delta_{R_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\delta R_{1i}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{U_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{1;\delta_{U_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}+\\
&(-1)D^{-2}_{2;\delta_{U_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial D_{2;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{2;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{2;\delta_{U_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}+\\
&(-1)D^{-2}_{3;\delta_{U_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial D_{3;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{3;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{3;\delta_{U_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{L_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{L_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{1;\delta_{L_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}+\\
&\delta_{U_{1i}}\delta_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{1;\delta_{U_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}+\\
&(-1)D^{-2}_{2;\delta_{U_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial D_{2;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{2;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{2;\delta_{U_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial^2 D_{2;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}+\\
&(-1)D^{-2}_{3;\delta_{U_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial D_{3;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{3;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{3;\delta_{U_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{I_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{I_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{1;\delta_{I_{1i}}\delta_{U_{2i}}}\cdot\frac{\partial^2 D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}\\
&\delta_{R_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1}+D^{-1}_{\delta_{R_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_1\partial \boldsymbol{\beta}_3^T}\Bigg\}+
\end{aligned}
$$

$$
\begin{aligned}
&\delta_{L_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{L_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_3^T}\right\}+\\
&\delta_{R_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{R_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_3^T}\right\}+\\
&\delta_{I_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{I_{1i}}\delta_{R_{2i}}}\cdot\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_3^T}\right\}+\\
&\delta_{L_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{L_{1i}}\delta_{I_{2i}}}\cdot\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_3^T}\right\}+\\
&\delta_{I_{1i}}\delta_{L_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta}_3}\cdot\frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta}_1}+D^{-1}_{\delta_{I_{1i}}\delta_{L_{2i}}}\cdot\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_3^T}\right\}+
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} =& \delta_{U_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{3;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{3;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{3;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{5;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{5;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{5;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{5;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{R_{1i}}\delta_{R_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{R_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{L_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{L_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{I_{1i}}\delta_{I_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{I_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{U_{1i}}\delta_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{R_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{1;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{2;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{2;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{2;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{3;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{3;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{3;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{U_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} \\
& + \delta_{L_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{1;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{2;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{2;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{2;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{3;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{3;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{3;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{U_{1i}}\delta_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{I_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{1;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{2;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{2;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{2;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T} + \\
& (-1)D^{-2}_{3;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \left(\frac{\partial D_{3;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{3;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{R_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{R_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} +
\end{aligned}
$$

$$
\begin{aligned}
&+ \delta_{L_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{L_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\right\} + \\
&+ \delta_{R_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{R_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\right\} + \\
&+ \delta_{I_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{R_{2i}}} \cdot \left(\frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{I_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\right\} + \\
&+ \delta_{L_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{I_{2i}}} \cdot \left(\frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{L_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\right\} + \\
&+ \delta_{I_{1i}}\delta_{L_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{L_{2i}}} \cdot \left(\frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2}\right)^2 + D^{-1}_{\delta_{I_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} =& \delta_{U_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{1;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} + \\
& (-1)D^{-2}_{3;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{2;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{3;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{3;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} + \\
& (-1)D^{-2}_{5;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{5;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{5;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{5;\delta_{U_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{5;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} + \\
& + \delta_{R_{1i}}\delta_{R_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{R_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} + \\
& + \delta_{L_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{L_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} + \\
& + \delta_{I_{1i}}\delta_{I_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{I_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} + \\
& + \delta_{U_{1i}}\delta_{R_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{1;\delta_{U_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} + \\
& + \delta_{R_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{1;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} + \\
& (-1)D^{-2}_{2;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{2;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{2;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{2;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} + \\
& (-1)D^{-2}_{3;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{\delta_{3;R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{3;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{3;\delta_{R_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2^T}\Bigg\} + \\
& + \delta_{U_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{1;\partial D_{\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{1;\delta_{U_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} \\
& + \delta_{L_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{1;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} + \\
& (-1)D^{-2}_{2;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{2;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{2;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{2;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} + \\
& (-1)D^{-2}_{3;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{3;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{3;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{3;\delta_{L_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} + \\
& + \delta_{U_{1i}}\delta_{I_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{1;\delta_{U_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} + \\
& + \delta_{I_{1i}}\delta_{U_{2i}}\Bigg\{(-1)D^{-2}_{1;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{1;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} + \\
& (-1)D^{-2}_{2;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{2;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{2;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{2;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{2;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T} + \\
& (-1)D^{-2}_{3;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial D_{3;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{3;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{3;\delta_{I_{1i}}\delta_{U_{2i}}} \cdot \frac{\partial^2 D_{3;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} + \\
& + \delta_{R_{1i}}\delta_{L_{2i}}\Bigg\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{R_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\Bigg\} +
\end{aligned}
$$

$$+ \delta_{L_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{L_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\right\}+$$
$$+ \delta_{R_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{R_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{R_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\right\}+$$
$$+ \delta_{I_{1i}}\delta_{R_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{I_{1i}}\delta_{R_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\right\}+$$
$$+ \delta_{L_{1i}}\delta_{I_{2i}}\left\{(-1)D^{-2}_{\delta_{L_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{L_{1i}}\delta_{I_{2i}}} \cdot \frac{\partial^2 D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\right\}+$$
$$+ \delta_{I_{1i}}\delta_{L_{2i}}\left\{(-1)D^{-2}_{\delta_{I_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3} \cdot \frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2} + D^{-1}_{\delta_{I_{1i}}\delta_{L_{2i}}} \cdot \frac{\partial^2 D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3^T}\right\}$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_3 \partial \boldsymbol{\beta}_3^T} =& \delta_{U_{1i}}\delta_{U_{2i}}\left\{(-1)D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}^{-2}\cdot\left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}^{-1}\cdot\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\} \\
&+ \delta_{R_{1i}}\delta_{R_{2i}}\left\{(-1)D_{\delta_{R_{1i}}\delta_{R_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{R_{1i}}\delta_{R_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{L_{1i}}\delta_{L_{2i}}\left\{(-1)D_{\delta_{L_{1i}}\delta_{L_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{L_{1i}}\delta_{L_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{I_{1i}}\delta_{I_{2i}}\left\{(-1)D_{\delta_{I_{1i}}\delta_{I_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{I_{1i}}\delta_{I_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{U_{1i}}\delta_{R_{2i}}\left\{(-1)D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}^{-2}\cdot\left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}^{-1}\cdot\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\} \\
&+ \delta_{R_{1i}}\delta_{U_{2i}}\left\{(-1)D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}^{-2}\cdot\left(\frac{\partial D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}^{-1}\cdot\frac{\partial^2 D_{1;\delta_{R_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{U_{1i}}\delta_{L_{2i}}\left\{(-1)D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}^{-2}\cdot\left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}^{-1}\cdot\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\} \\
&+ \delta_{L_{1i}}\delta_{U_{2i}}\left\{(-1)D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}^{-2}\cdot\left(\frac{\partial D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}^{-1}\cdot\frac{\partial^2 D_{1;\delta_{L_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\} \\
&+ \delta_{U_{1i}}\delta_{I_{2i}}\left\{(-1)D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}^{-2}\cdot\left(\frac{\partial D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}^{-1}\cdot\frac{\partial^2 D_{1;\delta_{U_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\} \\
&+ \delta_{I_{1i}}\delta_{U_{2i}}\left\{(-1)D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}^{-2}\cdot\left(\frac{\partial D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}^{-1}\cdot\frac{\partial^2 D_{1;\delta_{I_{1i}}\delta_{U_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{R_{1i}}\delta_{L_{2i}}\left\{(-1)D_{\delta_{R_{1i}}\delta_{L_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{R_{1i}}\delta_{L_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{L_{1i}}\delta_{R_{2i}}\left\{(-1)D_{\delta_{L_{1i}}\delta_{R_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{L_{1i}}\delta_{R_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{R_{1i}}\delta_{I_{2i}}\left\{(-1)D_{\delta_{R_{1i}}\delta_{I_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{R_{1i}}\delta_{I_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{R_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{I_{1i}}\delta_{R_{2i}}\left\{(-1)D_{\delta_{I_{1i}}\delta_{R_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{I_{1i}}\delta_{R_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{R_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{L_{1i}}\delta_{I_{2i}}\left\{(-1)D_{\delta_{L_{1i}}\delta_{I_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{L_{1i}}\delta_{I_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{L_{1i}}\delta_{I_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}+ \\
&+ \delta_{I_{1i}}\delta_{L_{2i}}\left\{(-1)D_{\delta_{I_{1i}}\delta_{L_{2i}}}^{-2}\cdot\left(\frac{\partial D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3}\right)^2 + D_{\delta_{I_{1i}}\delta_{L_{2i}}}^{-1}\cdot\frac{\partial^2 D_{\delta_{I_{1i}}\delta_{L_{2i}}}}{\partial \boldsymbol{\beta}_3\partial \boldsymbol{\beta}_3^T}\right\}
\end{aligned}
$$

Appendix B

Computational Aspects

Declaration

This section has been used in a MSc thesis at University College London, UK. Furthermore, part of the following section it is published content in *Copula link-based additive models for bivariate time-to-event outcomes with general censoring scheme, Computational Statistics & Data Analysis, Danilo Petti, Alessia Eletti, Giampiero Marra, Rosalba Radice, Volume 175, 2022, 107550, ISSN 0167-9473, https://doi.org/10.1016/j.csda.2022.107550.*

A.1 Complete log-likelihood function

The scope of this section is to illustrate some relevant computational aspects of the wok. We first introduce the procedure used to verify analytical quantities derived in Appendix A. We finally discuss some computational advantages deriving from the implementation of the analytical quantities introduced in Appendix A,

The correctness of the analytical expressions for the gradien and Hessian have been checked through `numderiv` package in `R` which allowed to compute the numerical expressions and compare them with the analytical ones. For the data generating process, a proprietary function was used, a seed was set and a sample of 3000 was generated. Part of the code used for the experiment is the following

```
#fixing the seed
set.seed(24)

# Generating a sample of size 3000
output= datagenCopulaMixCens(n = 3000)

# Computing the relative frequencies
table(output$dataSim$cens)/nrow(output$dataSim)*100

 >        II       IL       IR       IU       LI
        12.4667  5.8667  7.2667  4.3667  8.1333
          LL       LR       LU       RI       RL
         5.6667  2.2667  3.3667 13.2000  5.8333
          RR       RU       UI       UL       UR
        14.7333  4.4333  5.0000  2.4333  2.6667
          UU
         2.3000
# Setting the two equations
eq1 <- t11 ~ s(t11, bs = "mpi") + z1 + z2
eq2 <- t21 ~ s(t21, bs = "mpi") + z3

#Allocation into a list
f.list = list(eq1, eq2)

# Separate censoring indicator as
# this is how the function takes it in ****
cens1 = as.factor(substr(as.character(dataSim$cens),
                        start = 1, stop = 1))
```

```
cens2 = as.factor(substr(as.character(dataSim$cens),
                         start = 2, stop = 2))


dataSim$cens1 = cens1
dataSim$cens2 = cens2



# Call gjrm function
out <- gjrm(f.list, data = dataSim, surv = TRUE,
            BivD = "T", margins = c("PH", "PH"),
            cens1 = cens1, cens2 = cens2, Model = "B",
            upperBt1 = 't12', upperBt2 = 't22')
```

To carry out this comparison we have generated a sample of size 3000, a t-student copula (`"T"`) and proportional hazards (`"PH"`) has been specified for both margins. The comparison analysis between numerical and analytical quantities was carried out individually on all 16 pieces and for each of the copula functions presented in Chapter 1. Either for the Gradient and the Hessian, the analysis has shown that the two quantities match over $1e-5$. The only exception is the t-student copula (`"T"`) whose, for the IU and UI pieces, showed some small discrepancies that we report in the table for completeness.

The discrepancies were found for the δ_{IU} and δ_{UI} pieces as we can see from the tables above.

(Intercept)	z1	z2	s(t11).1	s(t11).2	s(t11).3	s(t11).4	s(t11).5	s(t11).6	s(t11).7	s(t11).8	s(t11).9
-0.00010	-0.00010	-0.00005	-0.00210	-0.00007	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000
-0.00010	-0.00010	-0.00005	-0.00210	-0.00007	-0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000
-0.00005	-0.00005	-0.00003	-0.00109	-0.00004	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
-0.00210	-0.00210	-0.00109	-0.04258	-0.00138	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
-0.00007	-0.00007	-0.00004	-0.00138	-0.00005	-0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
-0.00015	-0.00015	-0.00008	-0.00306	-0.00010	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000
-0.00008	-0.00008	-0.00004	-0.00171	-0.00006	-0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000
-0.00329	-0.00329	-0.00173	-0.06680	-0.00219	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	0.00000
-0.00027	-0.00027	-0.00014	-0.00556	-0.00018	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000
-0.00001	-0.00001	-0.00001	-0.00025	-0.00001	-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000
-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
0.00022	0.00022	0.00012	0.00447	0.00015	0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000

Table B.1: This table shows the discrepancies between analytical and numerical Hessian for the IU piece fitted using as margin the proportional hazard (`"PH"`) for both of them and the t-student copula (`"T"`). Part 1

(Intercept)	z3	s(t21).1	s(t21).2	s(t21).3	s(t21).4	s(t21).5	s(t21).6	s(t21).7	s(t21).8	s(t21).9	V12
-0.00015	-0.00008	-0.00329	-0.00027	-0.00001	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00022
-0.00015	-0.00008	-0.00329	-0.00027	-0.00001	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00022
-0.00008	-0.00004	-0.00173	-0.00014	-0.00001	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00012
-0.00306	-0.00171	-0.06680	-0.00556	-0.00025	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00447
-0.00010	-0.00006	-0.00219	-0.00018	-0.00001	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00015
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
-0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
0.00173	0.00093	0.03775	0.00323	0.00017	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00079
0.00093	0.00052	0.02027	0.00168	0.00007	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00044
0.03775	0.02027	0.82532	0.07061	0.00363	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.01729
0.00323	0.00168	0.07061	0.00617	0.00035	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00145
0.00017	0.00007	0.00363	0.00035	0.00003	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00007
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00079	-0.00044	-0.01729	-0.00145	-0.00007	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00306

Table B.2: This table shows the discrepancies between analytical and numerical Hessian for the IU piece. Further details are given in the caption of Table B.1. Part 2

(Intercept)	z1	z2	s(t11).1	s(t11).2	s(t11).3	s(t11).4	s(t11).5	s(t11).6	s(t11).7	s(t11).8	s(t11).9
-0.00005	-0.00000	-0.00003	-0.00106	-0.00011	-0.00001	0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000
0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
-0.00003	0.00000	-0.00002	-0.00073	-0.00008	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000
-0.00106	0.00000	-0.00073	-0.02477	-0.00270	-0.00015	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00011	0.00000	-0.00008	-0.00270	-0.00029	-0.00002	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
-0.00001	-0.00000	-0.00000	-0.00015	-0.00002	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000
0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00001	-0.00000	0.00000	0.00015	0.00002	0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000
0.00000	-0.00000	0.00000	0.00008	0.00001	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
0.00012	0.00000	0.00008	0.00270	0.00029	0.00002	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000
0.00000	-0.00000	0.00000	0.00007	0.00001	0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00002	0.00000	0.00001	0.00039	0.00004	0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000

Table B.3: This table shows the discrepancies between analytical and numerical Hessian for the UI piece.Further details are given in the caption of Table B.1. Part 1

(Intercept)	z3	s(t21).1	s(t21).2	s(t21).3	s(t21).4	s(t21).5	s(t21).6	s(t21).7	s(t21).8	s(t21).9	V12
0.00001	0.00000	0.00012	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00002
0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00008	0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	0.00001
0.00015	0.00008	0.00270	0.00007	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00039
0.00002	0.00001	0.00029	0.00001	-0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00004
0.00000	0.00000	0.00002	0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000
0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00001	0.00000	0.00011	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00001
0.00000	0.00000	0.00006	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000
0.00011	0.00006	0.00204	0.00006	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00011
0.00000	0.00000	0.00006	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000
-0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000
-0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000
-0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000
-0.00001	-0.00000	-0.00011	-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00011

Table B.4: This table shows the discrepancies between analytical and numerical Hessian for the UI piece. Further details are given in the caption of Table B.1. Part 2

(Intercept)	z1	z2	s(t11).1	s(t11).2	s(t11).3	s(t11).4	s(t11).5	s(t11).6	s(t11).7	s(t11).8	s(t11).9
-0.00006	-0.00010	-0.00002	-0.00104	0.00005	0.00001	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00010	-0.00010	-0.00005	-0.00210	-0.00007	-0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	0.00000
-0.00002	-0.00005	-0.00001	-0.00036	0.00004	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000
-0.00104	-0.00210	-0.00036	-0.01781	0.00131	0.00015	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
0.00005	-0.00007	0.00004	0.00131	0.00025	0.00002	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
0.00001	-0.00000	0.00000	0.00015	0.00002	0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000
-0.00019	-0.00019	-0.00012	-0.00391	-0.00014	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000
-0.00012	-0.00012	-0.00008	-0.00244	-0.00008	-0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000
-0.00420	-0.00408	-0.00257	-0.08499	-0.00291	-0.00002	0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000
-0.00038	-0.00038	-0.00025	-0.00769	-0.00025	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000
-0.00006	-0.00006	-0.00006	-0.00128	-0.00004	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00005	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00001	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000
0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000
0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000
0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000
0.00020	0.00022	0.00011	0.00407	0.00010	-0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00000

Table B.5: This table shows the discrepancies between the full analytical and numerical Hessian Matrix. Further details are given in the caption of Table B.1. Part1

(Intercept)	z3	s(t21).1	s(t21).2	s(t21).3	s(t21).4	s(t21).5	s(t21).6	s(t21).7	s(t21).8	s(t21).9	V12
-0.00015	-0.00008	-0.00321	-0.00025	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00020
-0.00014	-0.00008	-0.00309	-0.00025	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	0.00022
-0.00007	-0.00004	-0.00161	-0.00012	0.00001	0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	0.00011
-0.00299	-0.00159	-0.06480	-0.00500	0.00007	0.00001	0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00408
-0.00009	-0.00004	-0.00183	-0.00010	0.00004	0.00000	0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00011
0.00001	0.00001	0.00021	0.00003	0.00002	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
0.00000	0.00000	0.00002	0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000
-0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000
-0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	-0.00000	0.00000	-0.00000	0.00000
0.00171	0.00091	0.03730	0.00318	0.00014	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00078
0.00091	0.00051	0.01990	0.00164	0.00005	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00043
0.03730	0.01990	0.81582	0.06956	0.00313	-0.00002	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.01707
0.00318	0.00164	0.06956	0.00603	0.00028	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00143
0.00014	0.00005	0.00313	0.00028	-0.00001	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00006
-0.00000	-0.00000	-0.00002	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	0.00000
0.00000	-0.00000	0.00000	0.00000	0.00000	-0.00000	0.00000	0.00000	-0.00000	0.00000	-0.00000	0.00000
-0.00080	-0.00045	-0.01746	-0.00148	-0.00009	-0.00000	-0.00000	-0.00000	-0.00000	0.00000	0.00000	-0.00317

Table B.6: This table shows the discrepancies between the full analytical and numerical Hessian Matrix. Further details are given in the caption of Table B.1. Part 2

Another fundamental part of this work concerns the reduction of computational times for the calculation of the parameters of the survival and hazard functions of the proposed model. For this reason we present a table showing the time required to compute the Hessian matrix for each of the sixteen pieces present in the log-likelihood. It can be clearly seen that the method based on the analytical calculation takes on average 6 seconds (time refers to a model with copula `T`, and margins `PH`, `PO`) for each of the pieces, a limited amount of time compared to the seconds required in the case of the numerical quantities. For this test we analyzed the pieces individually and calculated the computational time with the package `microbenchmark`, the results in table have been obtained running the function 10 times to then computing the computational time. The following table shows the minimum, first quartile, average, median, third quartile and maximum.

Table B.7: In this table are the result of the computing times (seconds) of the analytical (left) numerical (right) hessian derived using the function `jacobian` implemented in the package `numDeriv` in `R`. The results have been obtained using the the package `microbenchmark` in `R`, for each pieces the function computed the hessian 10 times. The values displayed are a summary of the runs. These computational times refer to a bivariate model fitted using the t-student ("`T`") as the copula, proportional hazard ("`PH`") as the first margin and proportional odds ("`PO`") as the second.

expr	min	lq	mean	median	uq	max
δ_{UU}	6.30	6.40	6.55	6.48	6.56	7.11
δ_{RR}	6.26	6.37	6.43	6.42	6.47	6.94
δ_{LL}	6.27	6.38	6.45	6.43	6.52	7.01
δ_{II}	6.29	6.38	6.47	6.46	6.52	7.03
δ_{UR}	6.20	6.32	6.41	6.38	6.43	7.28
δ_{RU}	6.21	6.31	6.37	6.36	6.43	6.92
δ_{UL}	6.21	6.32	6.38	6.36	6.43	6.94
δ_{LU}	6.23	6.29	6.36	6.33	6.40	6.95
δ_{UI}	6.20	6.30	6.37	6.36	6.43	6.96
δ_{IU}	6.19	6.31	6.38	6.36	6.42	6.96
δ_{RL}	6.20	6.29	6.38	6.35	6.43	7.25
δ_{LR}	6.19	6.28	6.36	6.34	6.40	6.88
δ_{RI}	6.21	6.30	6.37	6.35	6.41	6.98
δ_{IR}	6.20	6.29	6.45	6.38	6.52	7.36
δ_{LI}	6.35	6.52	6.70	6.62	6.75	8.07
δ_{IL}	6.35	6.73	6.88	6.89	7.03	7.43

Table B.8: Computational time in seconds Analytical Hessian

expr	min	lq	mean	median	uq	max
δ_{UU}	1226.32	1226.52	1255.69	1226.71	1270.37	1314.03
δ_{RR}	1227.28	1230.43	1232.60	1233.58	1235.27	1236.95
δ_{LL}	1234.69	1236.02	1236.56	1237.35	1237.50	1237.64
δ_{II}	1255.28	1256.15	1257.47	1257.03	1258.57	1260.10
δ_{UR}	1232.31	1232.85	1233.89	1233.38	1234.68	1235.99
δ_{RU}	1232.30	1232.64	1233.16	1232.98	1233.59	1234.19
δ_{UL}	1230.06	1231.70	1232.68	1233.34	1233.99	1234.64
δ_{LU}	1234.68	1236.05	1236.59	1237.43	1237.54	1237.66
δ_{UI}	1234.54	1239.75	1241.89	1244.97	1245.56	1246.15
δ_{IU}	1244.39	1244.59	1244.96	1244.78	1245.24	1245.70
δ_{RL}	1243.13	1243.59	1244.34	1244.06	1244.95	1245.84
δ_{LR}	1241.44	1241.66	1242.43	1241.88	1242.92	1243.96
δ_{RI}	1236.53	1238.15	1238.84	1239.78	1240.00	1240.22
δ_{IR}	1235.49	1235.50	1236.16	1235.51	1236.49	1237.48
δ_{LI}	1232.93	1233.78	1234.96	1234.63	1235.98	1237.34
δ_{IL}	1235.25	1236.70	1242.62	1238.15	1246.31	1254.46

Table B.9: Computational time in seconds Numerical Hessian

Appendix C

Detailed discussion about the mm() function

The aim of this section is to explore and discuss the application of the `mm()` function implemented in `GJRM` package. This function plays a crucial role in the correct log-likelihood, Gradien and Hessian imprementation in `GJRM`. In detail, `mm()` avoids situation in which specific quantities can cause an unexpected error. If not applied it can happen to have quantities like $\log(0)$ witch result in `Inf` values not allowing the function to work properly.

In the following we are going to discuss whether the application of `mm()` is theoretically founded or not, this will help us to not omit any sensible information in the log-likelihood. To do so, we need to introduce some important results about the copula functions, to then discuss the application for each piece in the log-likelihood presented in the previous chapter.

Theoretical results

Theorem 1. *(**Theorem 2.2.8** in Nelsen (2016)), Let C be a copula. if $\partial C(u,\nu)/\partial\nu$ and $\partial^2 C(u,\nu)/\partial u\partial\nu$ are continous on* $\mathbf{I}^2$ *and $\partial C(u,\nu)/\partial u$ exist for all $u \in (0,1)$ when $\nu = 0$, then $\partial C(u,\nu)/\partial u$ and $\partial^2 C(u,\nu)/\partial u\partial\nu$ exist in* $(0,1)^2$.

Theorem 2. *(**Theorem 2.2.3** in Nelsen (2016)), Let C a copula, then for every (u,ν) in $DomC$*

$$max\{u+\nu-1,0\} \leq C(u,\nu) \leq min\{u,\nu\}$$

Theorem 3. *(**lemma 2.1.3** in Nelsen (2016)) Let S_1 and S_2 be nonempty subsets of $\bar{\mathbb{R}}$, and let H be a 2-increasing function with domain $S_1 \times S_2$. Let x_1, x_2 be in S_1 with $x_1 \leq x_2$ and let y_1, y_2 be in S_2 with $y_1 \leq y_2$. Then the function $t \to H(t, y_2) - H(t, y_1)$ is nondecreasing on S_1, and the function $t \to H(x_2, t) - H(x_1, t)$ is nondecreasing on S_2.*

In the following we are going to discuss all the pieces of the log-likelihood function.

Uncensored- Uncensored

T_{1i} uncensored T_{2i} uncensored. As concern the part δ_{U_i,U_i} we have the following piece

$$\log P(T_{1i} = t_{1i}, T_{2i} = t_{2i}) = \log \left[\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i})), \partial G_2(\eta_{2i}(t_{2i}))} G_1'(\eta_{1i}(t_{1i})) G_2'(\eta_{2i}(t_{2i})) \frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}} \frac{\partial \eta_{2i}(t_{2i})}{\partial t_{2i}} \right]$$

The corresponding code implemented in `R` is

```
VC$indUU*( log(c.copula2.be1be2) + log(-dS1eta1) + log(-dS2eta2)
+ log(Xd1P) + log(Xd2P)
```

In order to prove some results we can use the **Theorem 2.2.8** in Nelsen (2016), using this theorem we can state that

$$\frac{\partial^2 C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i})), \partial G_2(\eta_{2i}(t_{2i}))} \in [0, 1]$$

we can conclude that this quantity is bounded $\in [0, 1]^2$. We can apply `mm()` with no damage.

Right censored - Right censored

T_{1i} right censored and T_{2i} right censored. As concern the part δ_{R_i,R_i} we have the following piece

$$\log(P(T_{1i} > r_{1i}, T_{2i} > r_{2i})) = \log S(r_{1i}, r_{2i}) = \log \left[C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\} \right]$$

The corresponding code in R is

```
VC$indRR*log(mm(p00))
```

In this case we are considering a probability which has been rewritten as a copula function. Copulas are by definition in $\in [0,1]^2$. Therefore,

$$C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\} \in [0,1]$$

We can apply `mm()` with no damage.

Left censored - Left censored

T_{1i} left censored and T_{2i} left censored. As concern the piece δ_{L_i,L_i} we have the following piece

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

The implemented code in R is

```
&VC$indLL*log(mm(1-p1-p2+p00))
```

Form some well known results in probability and the **Theorem 2.2.8** in Nelsen (2016) we know that $S_1 \in [0,1]$, $S_2 \in [0,1]$ and $C(S_1, S_2) \in [0,1] \times [0,1]$.

However, in order to define the domain of this quantity we need to discuss two different inequities:

The first inequality that we need to discuss is

$$1 - S_1 - S_2 + C(S_1, S_2) \geq 1$$

$$C(S_1, S_2) \geq S_1 + S_2$$

Meaning that to have the quantity on the left hand side greater than one, we need to have necessarily $C(S_1, S_2) > S_1 + S_2$

the second inequality

$$1 - S_1 - S_2 + C(S_1, S_2) \leq 0$$

$$C(S_1, S_2) \leq S_1 + S_2 - 1$$

At this point we can use the **Theorem 2.2.3** in Nelsen (2016) which states that for every (u, ν) in $DomC$

$$max\{u + \nu - 1, 0\} \leq C(u, \nu) \leq min\{u, \nu\}$$

This theorem makes the two inequalities false.

$$1 - G_1(\eta_{1i}(l_{1i})) - G_2(\eta_{2i}(l_{2i})) + C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\} \in (0, 1)$$

We can apply `mm()` with no damage.

Interval censored - Interval censored

T_{1i} interval censored and T_{2i} interval censored. As concern the piece δ_{I_i,I_i} we have

$$\log P(l_{1i} < T_{1i} < r_{1i}, l_{2i} < T_{2i} < r_{2i}) = \log \Big[C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\} + \\ - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} + C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\} \Big]$$

the `R` code implemented in `GJRM`

```
VC$indII*log( mm(p00-p00.mix1-p00.mix2+p00.2) )
```

In this part we are considering a bivariate probability, it is well known that $P(l_{1i} < T_{1i} < r_{1i}, l_{2i} < T_{2i} < r_{2i}) \in [0, 1]$. Therefore, no damage in applying `mm()`

Uncensored- Right censored

T_{1i} uncensored and T_{2i} right censored. As concern the piece δ_{U_i,R_i} we have

$$\log P(T_{1i}=t_{1i},T_{2i}>r_{2i})=\log\left[-\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}\right]$$

The R code is

```
VC$indUR*(log(c.copula.be1)+log(-dS1eta1)+log(Xd1P))
```

Using the **Theorem 2.2.8** in Nelsen (2016), we can state that $\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\in(0,1)$. Taking into account the fact that $G_1'(\eta_{1i}(t_{1i}))<0$ and the minus sign we can conclude that

$$-\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}\in(0,1)$$

Therefore, can apply `mm()` with no damage.

Uncensored-Left censored

T_{1i} uncensored and T_{2i} left censored. As concern the piece δ_{U_i,L_i} we have

$$\log\left[\left(\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}-1\right)\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial \eta_{1i}(t_{1i})}{\partial t_{1i}}\right]$$

In R code

```
VC$indUL*(log(mm( (c.copula.be1-1) * (dS1eta1) * Xd1P)))
```

Form the **Theorem 2.2.8** in Nelsen (2016) we know that $\partial C(u,\nu)/\partial u\in(0,1)$. Therefore,

$$\left[\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}-1\right]\in(-1,0)$$

Finally $\int_0^{l_{2i}} f(t_{1i},y)dy\in(0,1)$ as $G_1'(\eta_{1i}(t_{1i}))<0$
Therefore, can apply `mm()` with no damage.

Uncensored- Interval censored

T_{1i} uncensored and T_{2i} interval censored. As concern the piece δ_{U_i,I_i} we have

$$\log\left[\frac{\partial}{\partial t_{1i}}F(t_{1i},r_{2i})-\frac{\partial}{\partial t_{1i}}F(t_{1i},l_{2i})\right]=$$
$$\log\left[\left(\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}-\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\right)\cdot G_1'(\eta_{1i}(t_{1i}))\cdot\frac{\partial\eta_{1i}(t_{1i})}{\partial t_{1i}}\right]$$

The `R` code implemented in `GJRM` is

```
VC$indUI*( log( c.copula.be1.mix1-c.copula.be1)
                              + log(-dS1eta1)+ log(Xd1P))
```

We are mainly interested in get the range of the following quantity:

$$\left(\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}-\frac{\partial C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\right)$$
$$\frac{\partial}{\partial G_1(\eta_{1i}(t_{1i}))}\left[C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}-C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}\right]$$

From the theory of copulas we know that a copula is a 2-increasing function. Given $c\le d$ $a\le b$, $a,b,c,d\in\mathbf{I}$. Let $a=0$. Computing the volume and using the fact that $C(0,c)=0$ we end up with.

$$C(b,c)\le C(b,d)$$

in our case we have that $G_2(\eta_{2i}(r_{2i}))=S_2(r_{2i})=P(T_{2i}\ge r_{2i})$ and $G_1(\eta_{1i}(t_{1i}))=P(T_{2i}\ge l_{2i})$. Where $P(T_{2i}\ge l_{2i})>P(T_{2i}\ge r_{2i})$

$$C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}\le C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}$$

as $G_2(\eta_{2i}(r_{2i}))<G_2(\eta_{2i}(l_{2i}))$. Hence

$$C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(r_{2i}))\}-C\{G_1(\eta_{1i}(t_{1i})),G_2(\eta_{2i}(l_{2i}))\}\in[-1,0]$$

However, we are interested in the derivative of the argument in square brackets. Using the **Lemma 2.1.3** in Nelsen (2016) we have that the function $u \equiv C(u, \nu_2) - C(u, \nu_1) \geq 0$ is nondecresing. Hence $\frac{\partial(C(u,\nu_2)-C(u,\nu_1))}{\partial u}$ is defined and non-negative almost everywhere on **I**

According to the lemma we have that

$$\left[\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} - \frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\right] \geq 0$$

As consequence we have that

$$\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} \geq \frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}$$

For the **Theorem 2.2.8** we can assert that

$$\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}, \frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} \in [0, 1]$$

Therefore the **Theorem 2.2.8** and the **Lemma 2.1.3** stated before leads to

$$\left[\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} - \frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\right] \in [0, 1]$$

Finally

$$-\left[\frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))} - \frac{\partial C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i}))\}}{\partial G_1(\eta_{1i}(t_{1i}))}\right] \in [-1, 0]$$

In conclusion, we cannot apply `mm()` here as we are not dealing with a quantity in [0,1].

Right censored- Left censored

T_{1i} right censored and T_{2i} left censored. As concern the piece δ_{R_i,L_i} we have

$$\log P(T_{1i} > r_{1i}, T_{2i} < l_{2i}) = \log \Big[G_1(\eta_{1i}(r_{1i})) - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} \Big]$$

```
VC$indRL*log(mm(p1-p00))
```

From what as been discussed so far, it is well known that $G_1(\eta_{1i}(r_{1i})) \in [0,1]$ and that $C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i})) \in [0,1]$. In order to be sure that the quantity in square brackets is in [0,1], we need to prove that $C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} \leq G_1(\eta_{1i}(r_{1i}))$

From the **Theorem 2.2.3** in Nelsen (2016). It is well known that for every (u, ν) in $DomC$

$$max\{u + \nu - 1, 0\} \leq C(u, \nu) \leq min\{u, \nu\}$$

In this case $min\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} = G_1(\eta_{1i}(r_{1i}))$, $G_1(\eta_{1i}(r_{1i})) = P(T_1 \geq r_{1i})$ and $G_2(\eta_{2i}(l_{2i})) = P(T_2 \geq l_{2i})$, this means that $G_2(\eta_{2i}(l_{2i})) > G_1(\eta_{1i}(r_{1i}))$

we have that

$C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} \leq G_2(\eta_{2i}(r_{2i}))$. Therefore, the difference is $\in [0,1]$. We can apply safely `mm()` here.

Right censoring- Interval censoring

T_{1i} right censored and T_{2i} interval censored. As concern the piece δ_{R_i,I_i} we have

$$\begin{aligned}\log P(T_{1i} > r_{1i}, l_{2i} < T_{2i} < r_{2i}) &= \log \Big[F_2(r_{2i}) - F_2(l_{2i}) - F_1(r_{1i}, r_{2i}) + F_1(r_{1i}, l_{2i}) \Big] \\ &= \log \Big[C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i}))\} - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i}))\} \Big]\end{aligned}$$

In `R` code we have

```
VC$indRI*log(mm(p00-p00.mix1))
```

In this case inside the square bracket there is a probability in its own right. For this reason there is no need for further discussion. We can apply `mm()`.

Left censoring - Interval censoring

T_{1i} left censored and T_{2i} interval censored

$$\log P(T_{1i} < l_{1i}, l_{2i} < T_{2i} < r_{2i}) = \log\left[F(l_{1i}, r_{2i}) - F(l_{1i}, l_{2i})\right]$$

$$\log\left[G_2(\eta_{2i}(l_{2i})) - G_2(\eta_{2i}(r_{2i})) + C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i}))\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i}))\}\right]$$

In `R` code we have

```
VC$indLI*log(mm(p2-p2.2+p00.mix1-p00))
```

In this case inside the square bracket there is a probability in its own right. For this reason there is no need for further discussion. We can apply `mm()`.

Appendix D

Testing

In order to implement the Copula link based additive model in the `R` package `GJRM` we had to derive analitically the expression for the log-likelihood first, and for the Grandient and the Hessian then. These expressions have been also carefully implemented and tested. The aim of this section is to present a *selected* set of tests that have been meticulously carried out before the complete intergation into the `GJRM` package.

Testing: Computational time

The computational timing is the central part of this section. To test the code we set three different seeds $1, 24, 124$ to make the tests reproducible. For each copula implemented in the package `GJRM` we have fitted three types of models. The first one with proportional hazards for both margins (`PH,PH`), the second with proportional odds (`PO,PO`), the last one with probit (`probit, probit`). The sample size has been set to be $n = 1500$. The data have been simulated using the following set-up:

The time, denoted as T_{1i}, was generated from a proportional hazards model defined, on the survival function scale defined as

$$T_{1i} = \log[-\log S_{10}(t_{1i})] + \beta_{11} z_{1i} + s_{11}(z_{2i})$$

where $S_{10i}(t_{1i}) = 0.9 \exp(-0.4 t_{1i}^{2.5})$. Time T_{2i} was generated from a proportional odds model defined as

$$T_{2i} = \log\left[\frac{\{1 - S_{20}(t_{2i})\}}{S_{20}(t_{2i})}\right] + \beta_{21}z_{1i} + \beta_{22}z_{3i}$$

where $S_{20}(t_{2i}) = S_{10}(t_{2i}) = 0.9\exp\left(-0.4t_{1i}^{2.5}\right)$. The random censoring was obtained using uniform distributions in such a way to have a good proportion of censoring for each of the sixteen combinations in each simulation. Observations were generated using the Brent's univariate-finding root. The two survival times were joined using a Clayton copula. The predictor for the dependence parameter was specified as

$$\eta_{3i} = \beta_{31}z_{1i} + s_{31}(z_{2i})$$

The specification of η_{3i} allowed the dependence to vary across the observations. In practice this was achieved using the conditional sampling approach. The set up 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 $s_{11}(z_i) = \sin(2\pi z_i)$, $s_{31} = 3\sin(\pi z_i)$, $\beta_{11} = -1.5$, $\beta_{21} = \beta_{22} = 1.2$, $\beta_{31} = -1.5$.

The correlation structure among the covariates was generated using multivariate Normal distribution with a correlation parameter $\rho = 0.5$, and then transformed using the distribution function of a standard Normal distribution.

For all the models fitted in this section [1] we employed the following set of equations

$$\begin{aligned}
T_{1i} &= \psi(S_{10}(t_{1i})) + \beta_{11}z_{1i} + s_{11}(z_{2i})\\
T_{2i} &= \psi(S_{20}(t_{2i})) + \beta_{21}z_{1i} + \beta_{22}z_{3i}\\
\eta_{3i} &= \beta_{31}z_{1i} + s_{31}(z_{2i})
\end{aligned}$$

Where $\psi(\cdot) : [0,1] \rightarrow \mathbb{R}$ denotes a generic link function. In `R` code we have

[1] All the information in tables have been extracted using the function `conv.check(gjrm.fitted)`, where `gjrm.fitted` is a `gjrm` object

```
#  Model specification

eq1 <- t11 ~ s(t11, bs = "mpi") + z1 + s(z2)

eq2 <- t21 ~ s(t21, bs = "mpi") + z1 + z3

eq3 <- ~ s(z2)
```

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	22	7	26	None	19.9599
C0	14	10	37	None	22.45187
C90	22	1	6	None	15.02519
C180	37	7	80	None	36.54183
C270	22	1	9	None	15.71901
FGM	18	7	20	None	18.27255
F	15	8	33	None	21.55298
N	23	5	26	None	22.18613
G0	37	8	34	None	30.43247
G90	14	1	1	None	13.03001
G180	25	9	46	None	31.04879
G270	14	1	1	None	13.50349
J0	35	9	44	None	30.96297
J90	15	1	1	None	13.2069
J180	62	12	745	None	185.0906
J270	15	1	1	None	11.85245
PL	17	8	25	None	20.3481
T	29	8	29	None	474.675

Table D.1: Fitting results. Model `PH, PH` and seed=1, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	16	11	54	None	27.59977
C0	22	7	47	None	28.03658
C90	21	1	6	None	16.86469
C180	200	50	10000	None	2202.085
C270	21	1	6	None	16.57513
FGM	19	7	25	None	20.45774
F	13	10	55	None	26.62456
N	19	9	30	None	24.45876
G0	40	9	47	None	35.47162
G90	14	1	1	None	14.69865
G180	33	9	200	None	69.77589
G270	14	1	1	None	14.85373
J0	43	8	50	None	33.82469
J90	14	1	1	None	14.65545
J180	110	16	3063	None	688.2158
J270	14	1	1	None	14.5273
PL	11	10	41	None	26.75597
T	31	9	39	None	570.216

Table D.2: Fitting results. Model `PH, PH` and seed=24, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	200	18	1601	None	341.2036
C0	22	3	600	IM N PD	135.5123
C90	23	40	478	None	131.3181
C180	121	13	684	None	176.4275
C270	22	50	531	IM N PD	144.0548
FGM	200	50	10000	None	1725.037
F	15	21	199	None	58.18854
N	22	14	87	None	37.68124
G0	32	8	34	None	51.60512
G90	15	1	1	None	73.90189
G180	163	9	46	None	173.699
G270	21	1	1	None	75.4445
J0	38	9	44	None	41.37823
J90	16	1	1	None	68.26537
J180	19	12	754	None	20.56588
J270	16	1	1	None	67.63905
PL	17	8	25	None	126.0556
T	34	50	456	None	3655.559

Table D.3: Fitting results. Model `PH, PH` and seed=124, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	22	7	24	None	16.1143
C0	16	9	30	None	18.31315
C90	21	1	6	None	11.33788
C180	54	50	7613	None	1583.681
C270	21	1	6	None	12.05801
FGM	20	6	20	None	16.37437
F	16	6	26	None	17.80727
N	16	6	24	None	19.17481
G0	31	8	35	None	32.07566
G90	13	1	1	None	12.81056
G180	25	8	39	None	32.04177
G270	13	1	1	None	11.76854
J0	33	10	47	None	30.36208
J90	14	1	1	None	9.565875
J180	29	12	168	None	52.34302
J270	14	1	1	None	8.929115
PL	15	6	23	None	16.23262
T	27	7	27	None	426.07

Table D.4: Fitting results. Model `PO, PO` and seed=1, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	18	7	15	None	14.02403
C0	14	9	38	None	19.91496
C90	20	1	6	None	11.31615
C180	200	8	1271	None	304.9001
C270	23	1	7	None	12.46988
FGM	21	7	25	None	15.54671
F	16	7	38	None	18.15703
N	13	8	38	None	19.24094
G0	29	7	45	None	26.65045
G90	20	1	5	None	11.97756
G180	38	11	922	None	237.5011
G270	14	1	1	None	9.843499
J0	27	6	103	None	36.35251
J90	14	1	1	None	9.169414
J180	48	9	290	None	81.19933
J270	14	1	1	None	9.12127
PL	17	50	153	None	62.29843
T	21	8	28	None	430.891

Table D.5: Fitting results. Model `PO, PO` and seed=24, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	200	3	600	None	147.8308
C0	19	6	27	None	17.8257
C90	22	1	7	None	12.47949
C180	200	50	1786	None	423.1166
C270	20	1	7	None	11.78485
FGM	200	50	10000	None	1703.115
F	15	8	27	None	16.58173
N	14	9	33	None	18.99228
G0	25	9	30	None	23.1339
G90	14	1	1	None	10.19137
G180	26	10	58	None	30.8066
G270	14	1	1	None	10.21193
J0	28	8	32	None	22.03917
J90	15	1	1	None	9.884204
J180	180	10	77	None	64.63412
J270	15	1	1	None	9.847507
PL	13	9	31	None	19.25955
T	28	9	28	None	479.5751

Table D.6: Fitting results. Model `PO, PO` and seed=124, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	19	8	30	None	16.91084
C0	12	9	30	None	16.71748
C90	22	1	6	None	10.91304
C180	49	11	1803	None	376.8869
C270	22	1	6	None	11.18474
FGM	19	6	20	None	13.27417
F	15	6	23	None	14.32711
N	19	6	22	None	16.129
G0	26	8	30	None	22.01072
G90	14	1	1	None	9.130165
G180	20	9	37	None	22.68738
G270	14	1	1	None	9.214166
J0	28	10	47	None	25.54413
J90	15	1	1	None	8.933424
J180	31	12	342	None	89.0127
J270	15	1	1	None	9.039403
PL	13	6	22	None	16.15924
T	33	3	57	None	605.8231

Table D.7: Fitting results. Model `Probit, Probit` and seed=1, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	13	8	17	None	17.94738
C0	17	10	30	None	23.87917
C90	19	1	7	None	15.06983
C180	116	13	2600	None	576.5965
C270	24	1	7	None	15.47423
FGM	26	7	24	None	19.82227
F	17	8	42	None	24.161
N	7	3	40	None	19.93499
G0	19	9	53	None	29.59136
G90	15	1	1	None	13.10216
G180	31	12	666	None	179.262
G270	15	1	1	None	12.58143
J0	25	8	42	None	26.04726
J90	17	1	1	None	12.46699
J180	45	10	727	None	176.7257
J270	17	1	1	None	12.87111
PL	14	10	46	None	25.02405
T	16	10	31	None	436.2744

Table D.8: Fitting results. Model `Probit, Probit` and seed=24, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

Copula	Trust region iterations before smoothing parameter estimation	Loops for smoothing parameter estimation	Trust region iterations within smoothing loops	Problem	Average Fitting time
AMH	200	7	1383	None	289.5252
C0	21	3	442	None	100.6697
C90	22	1	8	None	12.01577
C180	160	50	10000	None	2064.035
C270	24	1	7	None	12.88474
FGM	200	50	10000	None	1713.78
F	10	9	31	None	16.6117
N	15	9	32	None	19.38031
G0	25	9	34	None	23.93471
G90	23	1	5	None	13.0579
G180	44	12	64	None	37.35579
G270	17	1	1	None	10.53458
J0	23	10	40	None	23.84282
J90	16	1	1	None	9.868283
J180	200	12	67	None	67.29565
J270	16	1	1	None	9.747149
PL	11	10	30	None	19.15415
T	25	10	32	None	501.6783

Table D.9: Fitting results. Model `Probit, Probit` and seed=124, in the problem colums we denote if any issue during the fitting process occurred. With (IM N PD) we denote Information Matrix not definite positive. The fitting time is expressed in seconds

From the table presented we can clearly appreciate the advantage, in terms of computational times, of having implemented the analytical expression of the Gradient and Hessian in `GJRM`. However, a careful reader will surely have noticed that the copula `T`, with the same model specification, is the most time consuming. This is particularly evident if we compare the `T` with the `AMH` that takes a comparable amount of steps to converge.

To investigate further, we decided to debug the function `trust` in order to evaluate the goodness of the starting values chosen when using a copula `T`. The results of these test will be discussed in the following paragraph.

Testing: Trust region debug

In this section we show the results of some debugging sessions of the `trust` function used to carry out the fitting procedure of the Copula link base additive model in `GJRM`. As the name of the function suggest `trust` implements the trust region algorithm.

In the following we present the debug of the models fitted for three different copula function T-student (`T`), Ali-Mikhail-Haq (`AMH`) and Clayton (`C0`). We have the `T` and `AMH` as they are comparable in terms of trust region steps. On the other hand, the Clayton `C0` is explored as is the copula used in the data generating process. Therefore, it can be considered as being a benchmark.

For sake of completeness, we report the part of the code debugged during these sessions.

```
for (iiter in 1:iterlim) {
            if (blather) {
                    theta.blather <- rbind(theta.blather, theta)
                    r.blather <- c(r.blather, r)
                    if (accept)
                            val.blather <- c(val.blather, out$value)
                    else val.blather <- c(val.blather, out.value.save)
            }
            if (accept) {
                    B <- out$hessian
                    g <- out$gradient
                    f <- out$value
                    out.value.save <- f
                    if (rescale) {
                            B <- B/outer(parscale, parscale)
                            g <- g/parscale
                    }
                    if (!minimize) {
```

```
                B <- (-B)
                g <- (-g)
                f <- (-f)
        }
        eout <- eigen(B, symmetric = TRUE)
        gq <- as.numeric(t(eout$vectors) %*% g)
}
is.newton <- FALSE
if (all(eout$values > 0)) {
        ptry <- as.numeric(
        -eout$vectors %*% (gq/eout$values))
        if (norm(ptry) <= r)
                is.newton <- TRUE
}
if (!is.newton) {
        lambda.min <- min(eout$values)
        beta <- eout$values - lambda.min
        imin <- beta == 0
        C1 <- sum((gq/beta)[!imin]^2)
        C2 <- sum(gq[imin]^2)
        C3 <- sum(gq^2)
        if (C2 > 0 || C1 > r^2) {
                is.easy <- TRUE
                is.hard <- (C2 == 0)
                beta.dn <- sqrt(C2)/r
                beta.up <- sqrt(C3)/r
                fred <- function(beep) {
                        if (beep == 0) {
```

```
                        if (C2 > 0)
                                return(-1/r)
                        else
                        return(sqrt(1/C1) - 1/r)
                }
                return(
                sqrt(1/sum((gq/(beta + beep))^2))
                - 1/r
                )
        }
        if (fred(beta.up) <= 0) {
                uout <- list(root = beta.up)
        }
        else if (fred(beta.dn) >= 0) {
                uout <- list(root = beta.dn)
        }
        else {
                uout <- uniroot(
                fred, c(beta.dn, beta.up)
                )
        }
        wtry <- gq/(beta + uout$root)
        ptry <- as.numeric(-eout$vectors %*% wtry)
}
else {
        is.hard <- TRUE
        is.easy <- FALSE
        wtry <- gq/beta
```

```
                    wtry[imin] <- 0
                    ptry <- as.numeric(-eout$vectors %*% wtry)
                    utry <- sqrt(r^2 - sum(ptry^2))
                    if (utry > 0) {
                            vtry <- eout$vectors[, imin,
                            drop = FALSE]
                            vtry <- vtry[, 1]
                            ptry <- ptry + utry * vtry
                    }
            }
    }
    preddiff <- sum(ptry * (g + as.numeric(B %*% ptry)/2))
    if (rescale) {
            theta.try <- theta + ptry/parscale
    }
    else {
            theta.try <- theta + ptry
    }
    out <- try(objfun(theta.try, ...))
    if (inherits(out, "try-error"))
            break
    check.objfun.output(out, minimize, d)
    ftry <- out$value
    if (!minimize)
            ftry <- (-ftry)
    rho <- (ftry - f)/preddiff
    if (ftry < Inf) {
            is.terminate <-
```

```
        abs(ftry - f) < fterm || abs(preddiff) < mterm
}
else {
        is.terminate <- FALSE
        rho <- (-Inf)
}
if (is.terminate) {
        if (ftry < f) {
                accept <- TRUE
                theta <- theta.try
        }
}
else {
        if (rho < 1/4) {
                accept <- FALSE
                r <- r/4
        }
        else {
                accept <- TRUE
                theta <- theta.try
                if (rho > 3/4 && (!is.newton))
                        r <- min(2 * r, rmax)
        }
}
if (blather) {
        theta.try.blather <-
        rbind(theta.try.blather, theta.try)
        val.try.blather <-
```

```
        c(val.try.blather, out$value)
        accept.blather <-
        c(accept.blather, accept)
        preddiff.blather <-
        c(preddiff.blather, preddiff)
        stepnorm.blather <-
        c(stepnorm.blather, norm(ptry))
        if (is.newton) {
                mytype <- "Newton"
        }
        else {
                if (is.hard) {
                        if (is.easy) {
                                mytype <- "hard-easy"
                        }
                        else {
                                mytype <- "hard-hard"
                        }
                }
                else {
                        mytype <- "easy-easy"
                }
        }
        type.blather <- c(type.blather, mytype)
        rho.blather <- c(rho.blather, rho)
}
if (is.terminate)
        break
```

In the following we present some interesting results based on some quantity extracted during the debug. In `argpath` and argtry are saved the set of parameters used and tried during each iteration. the quantity rho correspond to the expression $\rho_k = \frac{f(x_k)-f(x_k+p_k)}{m_k(0)-m_k(p_k)}$, where m_k is a Taylor-series expansion of f around x_k, which is

$$f(x_k + p) = f_k + g_k^T p + 1/2p^T \nabla^2 f(x_k + tp)p$$

where $f_k = f(x_k)$ and $g_k = \nabla f(x_k)$, and t is a scalar in (0,1). Using an approximation B_k to the Hessian we have m_k in the following way

$$m_k(p) = f_k + g_k^T p + 1/2p^T B_k P$$

Where B_k is some symmetric matrix. The difference between $m_k(p)$ and $f(x_k+p)$ is $\mathcal{O}(||p||^2)$. When $B_k = \nabla^2 f(x_k)$ leads to the trust-region Newton method. To obtain each step. we seek a solution of the subproblem

$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + 1/2p^T B_k P \;\; s.t.||p|| \leq \delta_k$$

where $\delta_k > 0$ is the trust region radius that in the code is stored within the object `r`. Where $|| \cdot ||$ is as usual the Euclidean norm.

One important ingrediant in the receip of the trust region is the strategy behind the choice of the trust-region radius δ_k. The choice is based on the agreement between the model function m_k and the objective function f at previous iterations. Given a step p_k we can define the ratio

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m(0) - m(p_k)}$$

This quantity, as the code above reveals, is stored at each step into the object called `rho`.

T Copula

In the following are presented convergence path of the rho, radius and parameters for the copula (T)

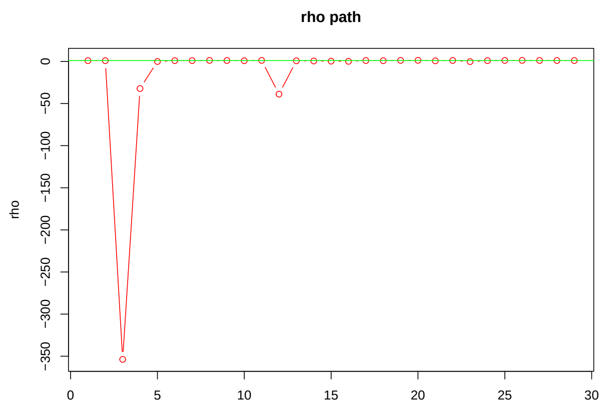


Figure D.1: Path of the $\rho_k = \frac{f(x_k)-f(x_k+p_k)}{m(0)-m(p_k)}$ during the trust region. Model with T, margins PH, PH, seed =1 and $n = 1500$.

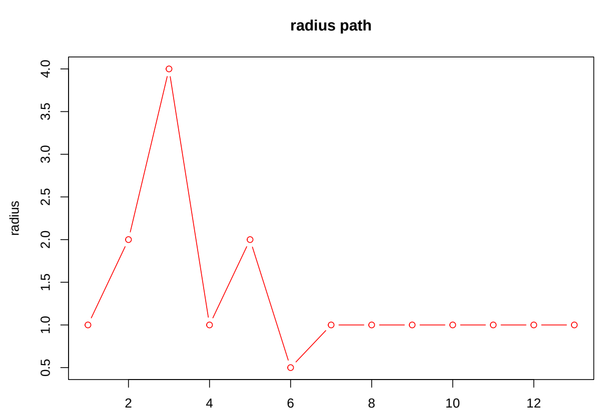


Figure D.2: Radius' path during the trust region. Model with T, margins PH, PH, seed =1 and $n = 1500$.

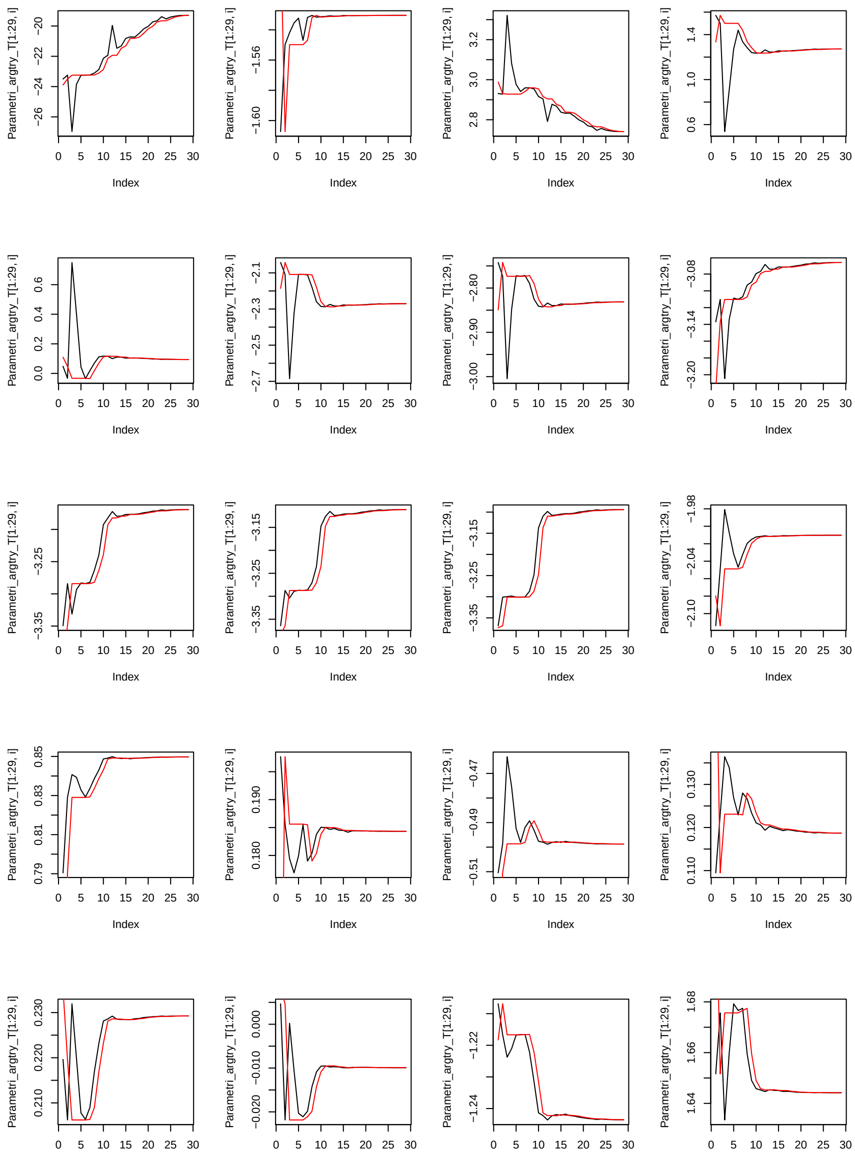


Figure D.3: Parameters' convergence path during the trust region. the red solid line represents the argument of the `argpath` object while in black solid line there is the `argtry` object. Model with `T`, margins `PH, PH`, seed =1 and $n = 1500$. pt.1

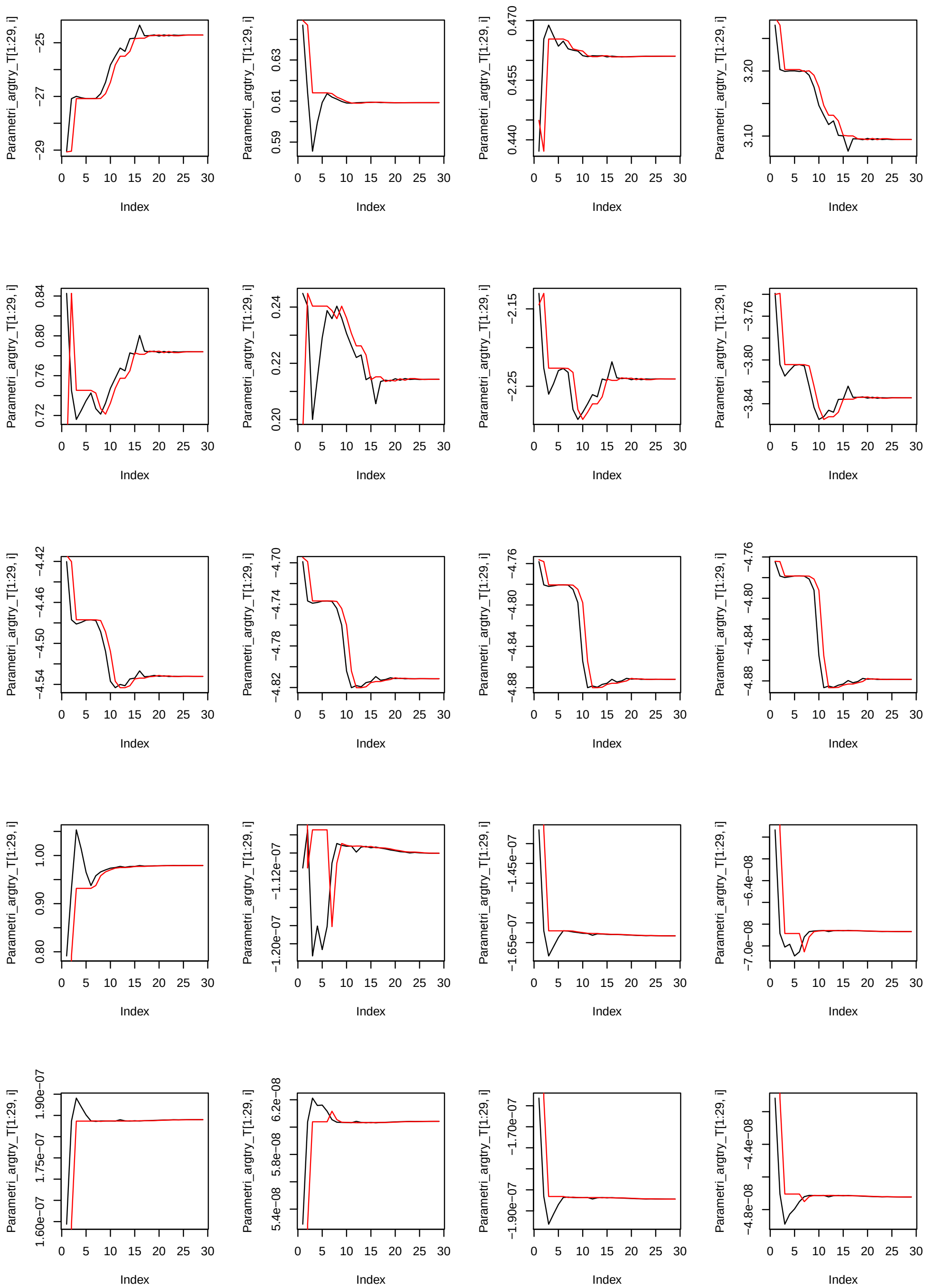


Figure D.4: Parameters' convergence path during the trust region. the red solid line represents the argument of the `argpath` object while in black solid line there is the `argtry` object. Model with `T`, margins `PH, PH`, seed =1 and $n = 1500$. pt.2

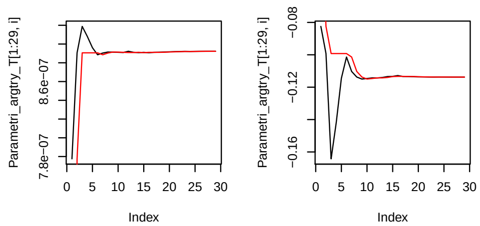


Figure D.5: Parameters' convergence path during the trust region. the red solid line represents the argument of the `argpath` object while in black solid line there is the `argtry` object. Model with `T`, margins `PH, PH`, seed =1 and $n = 1500$. pt.3

AMH Copula

In the following are presented convergence path of the rho, radius and parameters for the copula (AMH)

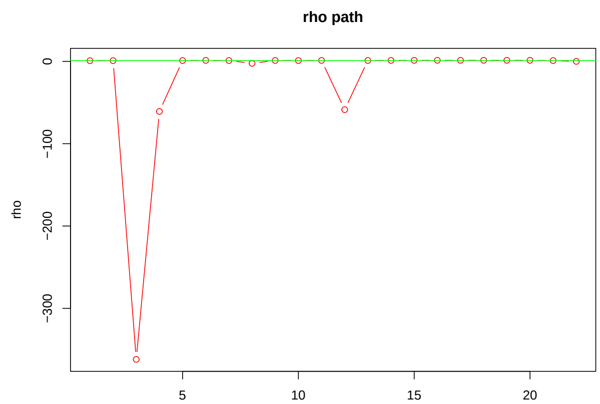


Figure D.6: Path of the $\rho_k = \frac{f(x_k)-f(x_k+p_k)}{m(0)-m(p_k)}$ during the trust region. Model with AMH, margins PH, PH, seed =1 and $n = 1500$.

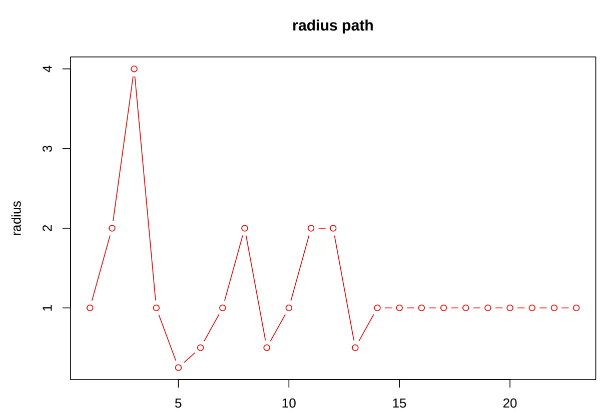


Figure D.7: Radius' path during the trust region. Model with AMH, margins PH, PH, seed =1 and $n = 1500$.

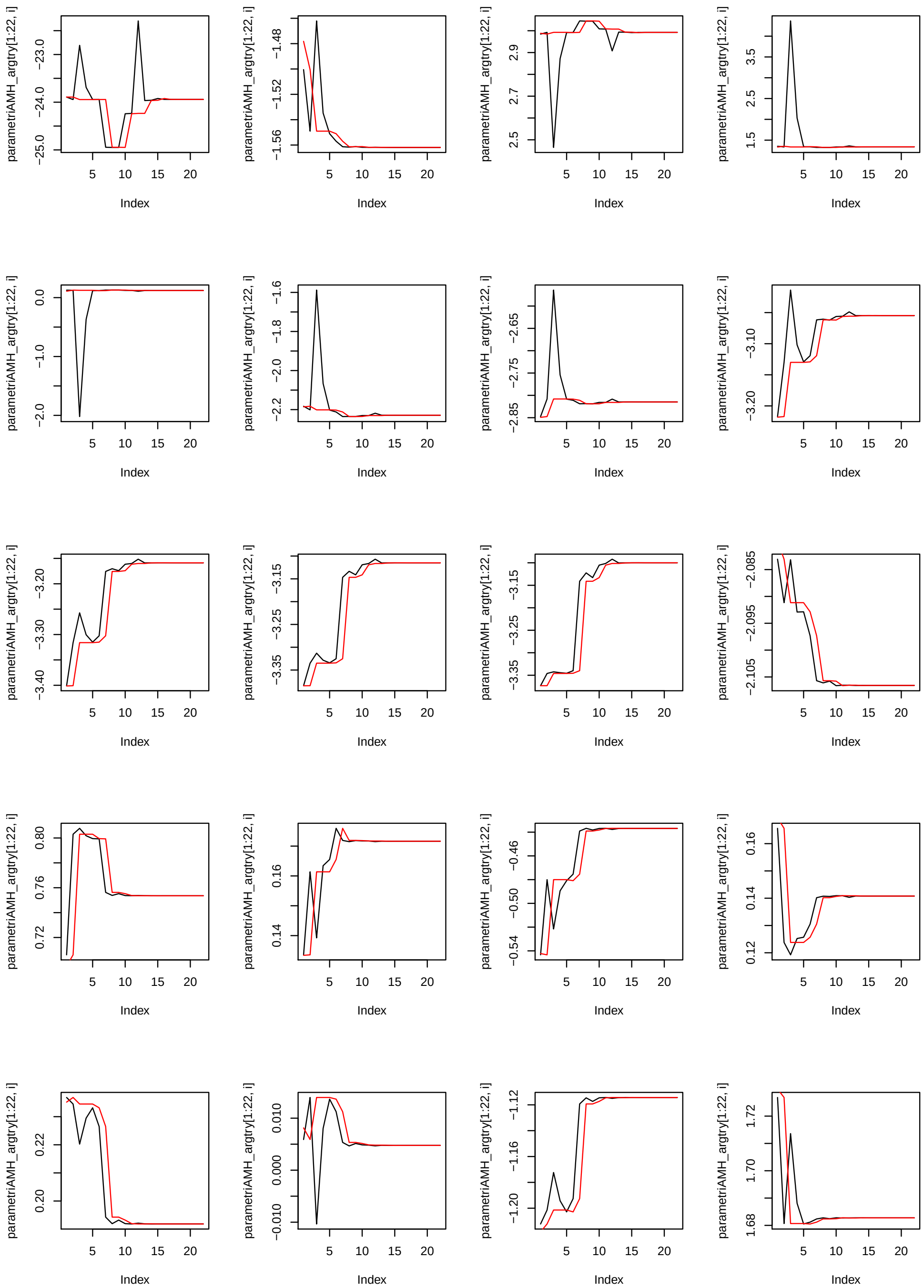


Figure D.8: Parameters' convergence path during the trust region. the red solid line represents the argument of `argpath` object while in black solid line there is the `argtry` object. Model with `AMH`, margins `PH, PH`, seed =1 and $n = 1500$. pt1

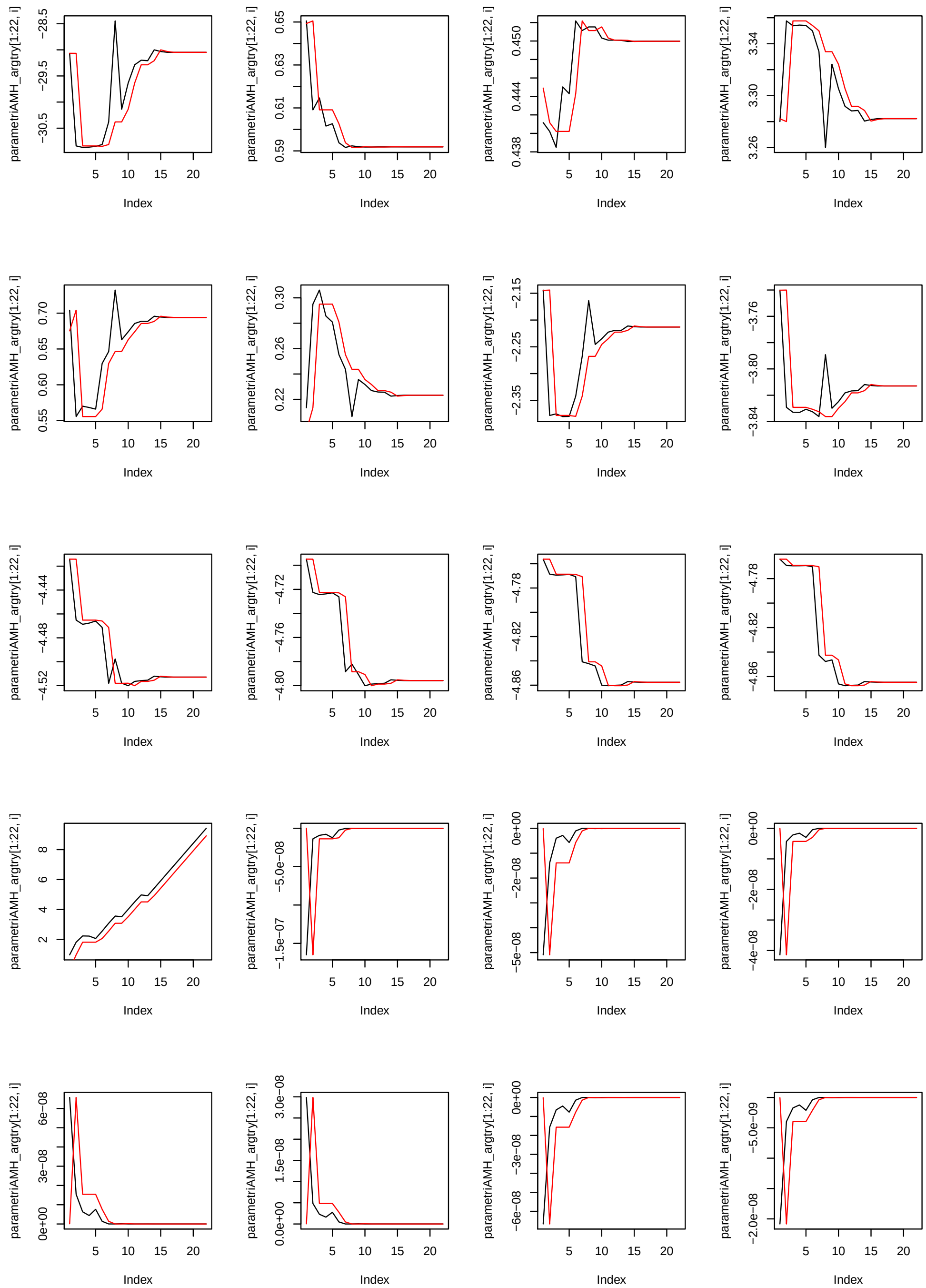


Figure D.9: Parameters' convergence path during the trust region. the red solid line represents the argument of `argpath` object while in black solid line there is the `argtry` object. Model with `AMH`, margins `PH, PH`, seed =1 and $n = 1500$. pt2

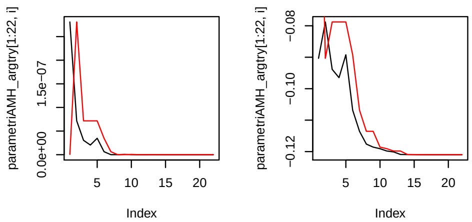


Figure D.10: Parameters' convergence path during the trust region. the red solid line represents the argument of `argpath` object while in black solid line there is the `argtry` object. Model with `AMH`, margins `PH, PH`, seed =1 and $n = 1500$. pt3

Clayton Copula

In the following are presented convergence path of the rho, radius and parameters for the copula (C0)

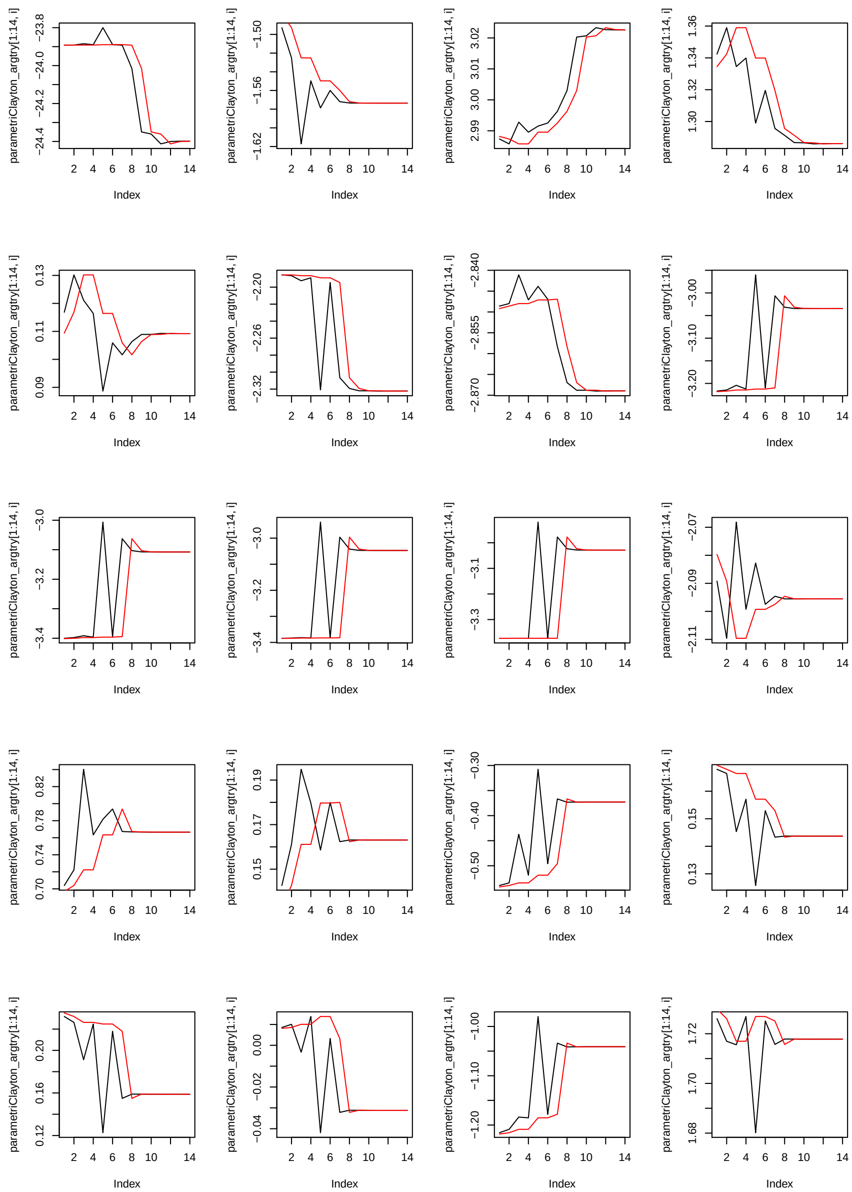


Figure D.11: Parameters' convergence path during the trust region. the red solid line represents the argument of `argpath` object while in black solid line there is the `argtry` object. Model with `C0` margins `PH, PH`, seed =1 and $n = 1500$. pt1

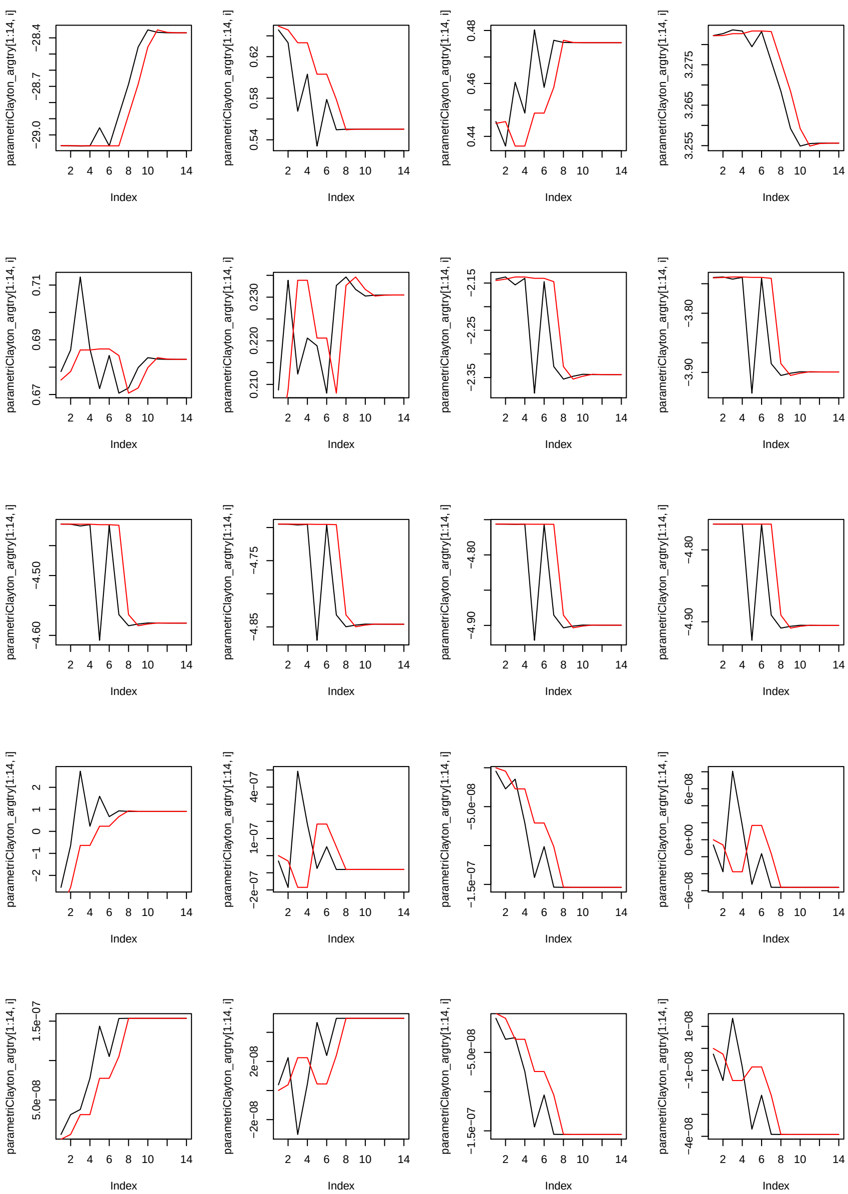


Figure D.12: Parameters' convergence path during the trust region. the red solid line represents the argument of `argpath` object while in black solid line there is the `argtry` object. Model with `C0` margins `PH, PH`, seed =1 and $n = 1500$. pt2

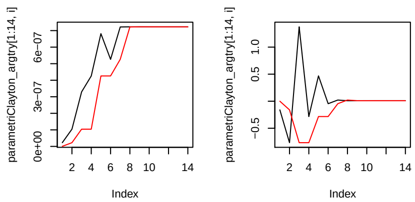


Figure D.13: Parameters' convergence path during the trust region. the red solid line represents the argument of `argpath` object while in black solid line there is the `argtry` object. Model with `C0` margins `PH, PH`, seed =1 and $n = 1500$. pt3

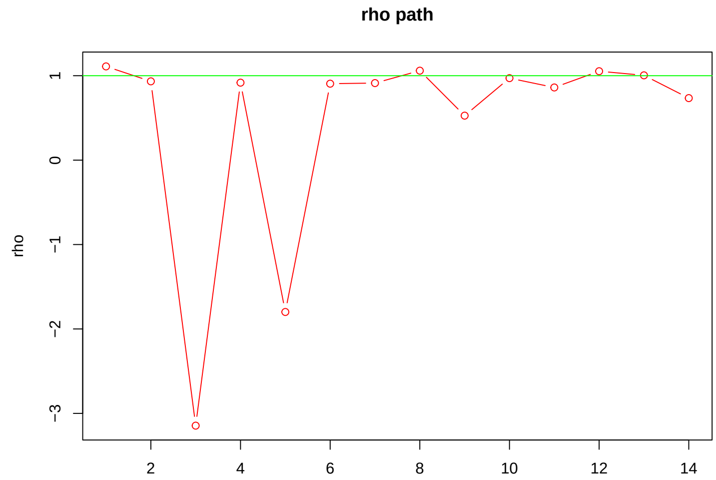


Figure D.14: Path of the $\rho_k = \frac{f(x_k)-f(x_k+p_k)}{m(0)-m(p_k)}$ during the trust region. Model with `C0`, margins `PH, PH`, seed =1 and $n = 1500$.

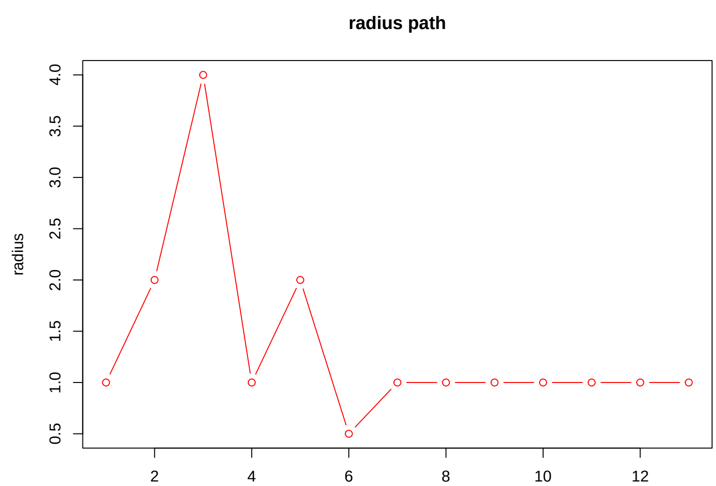


Figure D.15: Radius' path during the trust region. Model with copula `C0`, margins `PH, PH`, seed =1 and $n = 1500$.

In this table we summarize the quantities of the type $\Delta\hat{\boldsymbol{\delta}}_i = \hat{\boldsymbol{\delta}}_{start,i} - \hat{\boldsymbol{\delta}}_{end,i}, i \in \{\texttt{T}, \texttt{AMH}, \texttt{C0}\}$. Whith the symbol $\hat{\boldsymbol{\delta}}_{start}$ we denote the parameter vector tried by the trust region during the first iteration, while with $\hat{\boldsymbol{\delta}}_{end}$ we denote the parameter vector tried by the trust region during it last iteration. Concerning the model specification, we have that the following quantities have been obtained from a fitting with $n = 1500$, margins `PH,PH` and seed= 1.

The investigation of these quantities is particularly relevant as it helps to evaluate the range of values that every single parameter takes during the algorithm. An awkward range of value could be a symptom of poor convergence.

From the table does not emerge any strange behaviour of the parameters. This is a symptom that the starting values, which in `GJRM` are the parameters of two univariate gam models, are good enough to not cause an erratic change in the parameters' path.

	(Intercept)	z1	s(t11).1	s(t11).2	s(t11).3	s(t11).4	s(t11).5	s(t11).6	s(t11).7
T	-4.59	0.05	0.25	0.06	0.02	0.08	-0.02	-0.15	-0.23
A	0.05	0.08	-0.00	-0.00	-0.01	0.04	-0.03	-0.16	-0.24
C	0.51	0.10	-0.03	0.05	0.00	0.14	0.02	-0.18	-0.29
	s(t11).8	s(t11).9	s(z2).1	s(z2).2	s(z2).3	s(z2).4	s(z2).5	s(z2).6	s(z2).7
T	-0.27	-0.28	-0.07	-0.15	-0.05	-0.04	0.05	0.01	0.02
A	-0.27	-0.27	0.03	-0.06	-0.04	-0.11	0.03	0.04	0.00
C	-0.34	-0.34	0.02	-0.07	-0.03	-0.17	0.03	0.08	0.04
	s(z2).8	s(z2).9	(Intercept)	z1	z3	s(t21).1	s(t21).2	s(t21).3	s(t21).4
T	0.03	0.09	-4.35	0.04	-0.02	0.19	-0.11	-0.02	0.10
A	-0.10	0.05	-0.02	0.06	-0.01	-0.00	-0.02	-0.03	0.07
C	-0.18	0.01	-0.70	0.10	-0.03	0.03	-0.01	-0.04	0.20
	s(t21).4	s(t21).5	s(t21).6	s(t21).7	s(t21).8	s(t21).9	(Intercept)	s(z2).1	s(z2).2
T	0.10	0.09	0.12	0.12	0.12	0.11	-1.00	0.00	0.00
A	0.07	0.07	0.10	0.10	0.10	0.10	-8.94	0.00	-0.00
C	0.20	0.16	0.17	0.15	0.14	0.14	-4.43	0.00	0.00
	s(z2).2	s(z2).3	s(z2).4	s(z2).5	s(z2).6	s(z2).7	s(z2).8	s(z2).9	
T	0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	0.11	
A	-0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	0.12	
C	0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.01	

Table D.10: Let $\boldsymbol{\theta}$ the parameter vector, in table are shown the quantities $\Delta\hat{\boldsymbol{\delta}}_i = \hat{\boldsymbol{\delta}}_{start,i} - \hat{\boldsymbol{\delta}}_{end,i}, i \in \{\texttt{T}, \texttt{AMH}, \texttt{C0}\}$, where $\hat{\boldsymbol{\delta}}_{start}$ represents the parameter vector at the beginning of the trust region algorithm, while $\hat{\boldsymbol{\delta}}_{end}$ the one at the end. T=`T`, A=`AMH` e C=`C0`

Appendix E

Asymptotic Results

Declaration

This section has been used in a MSc thesis at University College London, UK.

A.1 Complete log-likelihood function

In this section we present some asymptotic results for the model discussed above. We will firstly introduce some notation. The following section is based on Jason (2010).

Big $\mathcal{O}$ and small o notation

This asymptotic notation concerns the rate at which the functions changes rather than the exact shape of them. Let give an informal intuition of the concept. Let f and g be two functions, $f(x) = \mathcal{O}(g(x))$ means that the two function are of the same order, i.e. they grow or shrink at the same rate. For instance, we can say that $f(x) = 2x^2 + 3x + 5$ is $\mathcal{O}(x^2)$ as in the long run, as x grows, the predominant term in this function will be the quadratic term. Conversely, $f(x) = o(g(x))$ means that f is negligible in asymptotic terms compared to g. In other words, g grows at a faster rate or shrinks at a slower rate. This concept are extremely important if we are interested to study the asymptotic behaviour of random sequence.

Definition 1. *Let b_n and a_n be two sequences, then $b_n = O(a_n)$ if there exist a constant C and n_0*

such that $b_n \leq Ca_n$ *for* $n \geq n_0$.

Definition 2. *Let* b_n *and* a_n *be two sequences, then* $b_n = o(a_n)$ *if for every* $\epsilon > 0$ *there exist* n_ϵ *such that* $b_n \leq \epsilon a_n$, *for* $n \geq n_\epsilon$

Big $\mathcal{O}_P$ and small o_P

In this section we will extend the above asymptotic concepts to the case where we are dealing with random variables. In the probabilistic setting we will indicate the above as $\mathcal{O}_P(.)$ and $o_P(.)$ respectively. Let start from the intuition first to then introduce the formal definitions. Given a sequence of r.v's X_n and a sequence of numbers a_n, $X_n = \mathcal{O}(a_n)$ means that the fraction X_n/a_n is bounded up to an exceptional event of arbitrarily small probability. For this reason it is also referred to as being bounded in probability. On the other hand, $X_n = o(a_n)$ means that the fraction X_n/a_n is negligible up to an exceptional event of arbitrarily small probability.

Definition 3. *Let* X_n *be a sequence of real valued random variables and* a_n *be a sequence of numbers.* $X_n = \mathcal{O}(a_n)$ *if for every* $\epsilon > 0$ *there exist constant* C_ϵ *and* n_ϵ *such that* $P(|X_n| \leq C_\epsilon a_n) > 1 - \epsilon$ *for every* $n \geq n_\epsilon$.

Definition 4. *Let* X_n *be a sequence of real valued random variables and* a_n *be a sequence of numbers.* $X_n = o(a_n)$ *if for every* $\epsilon > 0$ *there exist* n_ϵ *such that* $P(|X_n| \leq \epsilon a_n) > 1 - \epsilon$ *for every* $n \geq n_\epsilon$

Convergence of random variables

In this section we recall the notions of convergence in distribution, probability an the concept of almost sure convergence.

Definition 5. *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of real valued random variables and let* $F_n(x)$ *be the cumulative distribution function of* X_n. *The sequence is said to converge in distribution to a random variable* X *with cumulative distribution* $F(x)$ *if*

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for every $x \in \mathbf{R}$ at which F is continuous. We indicate this by $X_n \xrightarrow{d} X$

Definition 6. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of real valued random variables. The sequence is said to converge in probability to a random variables X if*

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \ \forall \epsilon > 0$$

We indicate this by $X_n \xrightarrow{p} X$

Definition 7. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of real valued random variables. The sequence is said to converge almost surely to a random variables X if*

$$P(\lim_{n\to\infty} X_n = X) = 1$$

We denote this convergence as $X_n \xrightarrow{a.s.} X$

Appendix F

Details on model building

For the model building procedure we encourage the use of AIC, BIC, Cox-Snell residuals and hypothesis testing. The AIC and BIC are given by $-2\ell(\hat{\boldsymbol{\delta}}) + 2edf$ and $-2\ell(\hat{\boldsymbol{\delta}}) + log(n)edf$, where the log-likelihood is evaluated at the penalized parameter estimates and $edf = tr(\hat{\mathbf{A}})$. The residuals are defined as $r_{vi} = -\log\{S_v(y_{vi}|\mathbf{x}_{vi};\hat{\boldsymbol{\beta}})\} \sim Exp(1)$ $v = 1, 2$ $i = 1, \ldots, n$. Assuming to have the observed cumulative hazard $\hat{H}_{rv}(r_{vi})$, which can be derived using a Kaplan-Meier estimate). If the model is correct then the plot $\{r_{vi,}, \hat{H}_{rv}(r_{vi})\}$ will have a 45^o slope, providing then an assessment of the model's goodness of fit. However, this procedure is not useful to suggest any clue about the mis-specification when the points do not follow the reference line.

A possible strategy that we suggest is to use the same set of covariates in all the three equations and choose the copula function using the AIC, BIC and Cox-Snell residuals. The same procedure can be use to select a relevant set of covariates (using stepwise, backward and/or forward selection). The model building can be simplified is the researcher wishes to include variables in the model based on some prior belief, or wishes to employ a particular set of link functions.

Appendix G

Software

As extensively discussed, the proposed model can be fitted using the `gjrm` function in the `R` package `GJRM`. A very basic example of syntax is as follows.

```
# set of three equations
eq1 <- t11 ~ s(t11, bs = "mpi") + z1 + s(z3)
eq2 <- t21 ~ s(t21, bs = "mpi") + z1 +  s(z3)
eq3 <- ~ z2 + s(z3)



f.list = list(eq1, eq2, eq3)

# call of gjrm on dataset dta
out <- gjrm(f.list, data = dta, surv = TRUE,
                    BivD = "C0", margins = c("probit", "probit"),
                    cens1 = cens1, cens2 = cens2, Model = "B",
                    upperBt1 = 't12', upperBt2 = 't22')
```

In the `R` code example the covariates that we intend to use to model the time are `z1`, `z2` and `z3`. The time variables for the first margin and the second margin are respectively denominated `t11`, `t12` and `t21`, `t22`, the three equations created are then allocated in `f.list` which is a list.While the first and second equations are dedicated to the two time margins, the third equation is for the copula dependence parameter θ. Symbols `s()` refers to the smooth function, whose have been specified with `bs='mpi'` as a monotonic smooth of time is needed, `k=10` the number of basis function by default are set to 10 and `m=2` the order of derivative is set to 2.

The dataset used in this fitting is stored in the object `dta`. the option `surv = TRUE` denotes that we are fitting a bivariate survival function. The option `BivD` is needed to specify the copula function (in this case `C0` stands for Clayton). `Model="B"` stands for bivariate model. The margins have to be specified using the option `margins` which requires a concatenation of two strings (in the example two probit link function were applied). In `upperBt1` and `upperBt2` the user have to specify the variable names of right bounds when interval censored observation are used.

Functions such as `AIC()`, `BIC()`, `summary()`, `predict()` can be used in the usual manner to extract the information criterion, to print a summary of the model fitted and make prediction. It can be either plotted the survival and the hazard function using `hazsurv.plot` in `GJRM`. Furthermore, to check the convergence `conv.check()` can be used.

Appendix H

Analysis AREDS data: R code

Declaration

This section it is published content in *Copula link-based additive models for bivariate time-to-event outcomes with general censoring scheme, Computational Statistics & Data Analysis, Danilo Petti, Alessia Eletti, Giampiero Marra, Rosalba Radice, Volume 175, 2022, 107550, ISSN 0167-9473, https://doi.org/10.1016/j.csda.2022.107550.*

The proposed modelling framework has been implemented within the `R` package `GJRM` (Marra & Radice, 2022), 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 is

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

Finally, 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`.

Appendix I

Mutual Information useful results

In this section some useful results about Mutual Information are presented. After having introduced MI, we will prove the concentration bound (condition $C1$ in Baranowski et al. (2020)), some well known MI's properties are discussed.

Mutual Information overview

The intuition behind this metric is the following: when an informative covariate is included in the estimation equations this should contribute in explaining the relation between the two time-to-event. On the contrary noisy information should not add any contribution in explaining the outcomes. The formal definition of Mutual Information is

Definition 8 (Mutual information)**.** *Let $\boldsymbol{X} = [\boldsymbol{X_1} : \boldsymbol{X_2} : \cdots : \boldsymbol{X_N}]$ such that $\boldsymbol{X_i} \in \mathbb{R}^n$, a set of random variables and let $f(\mathbf{x})$ be their joint density. The mutual information can be defined as*

$$I(\mathbf{x}) = \int_{\mathbf{x}} f(\mathbf{x}) \log \frac{f(\mathbf{x})}{\prod_{i=1}^{n} f_i(x_i)} d\mathbf{x},$$

where the integral sign have to be intended in a multivariate domain.

Most of the copula families are governed by a dependence parameter θ that spans the range between independence and full dependence. In the bivariate case, this is captured by the relationship between θ and the rank-based correlation measures, such as Spearman's rho or Kendall's tau

(Nelsen, 2006). In information theory, mutual information $I(Y, Y')$ measures the reduction in the uncertainty of Y due to the knowledge of Y'. Furthermore, it can be proved that $I(Y, Y')$ corresponds to the intersection of the information in Y with the information of Y' (Cover & Thomas, 2006). Therefore, MI quantifies the strength of the dependence between random variables. Furthermore, MI can be also viewed as the Kullback–Leibler divergence between the density in case of dependence and in case of independence $I(Y, Y') = D_{KL}(f(y, y')||f(y) \otimes f(y'))$.

It turns out that the MI is the entropy of the corresponding copula (Tenzer & Elidan, 2016). Therefore, we can define Copula Entropy (CE) as

Definition 9 (Copula entropy)**.** *Let* $\boldsymbol{X} = [\boldsymbol{X_1} : \boldsymbol{X_2} : \cdots : \boldsymbol{X_N}]$ *such that* $\boldsymbol{X_i} \in \mathbb{R}^n$*, a set of random variableS with a well defined marginal function* $\mathbf{u} = [F_1, \ldots, F_n]$ *with associated copula density* $c(\mathbf{u}) = \frac{\partial^n C(\mathbf{u})}{\partial \mathbf{u}_1 \cdots \partial \mathbf{u}_n}$*. Therefore, copula entropy can be defined as*

$$Ce(\mathbf{x}) = -\int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u}. \tag{I.1}$$

As stated in Ma & Sun (2008) the CE is equivalent to MI except for the minus out of the integral sign, a sketch of proof is provided, let MI be denoted $I(\mathbf{x})$.

$$\begin{aligned} I(\mathbf{x}) = &\int_{\mathbf{x}} f(\mathbf{x}) \log \frac{f(\mathbf{x})}{\prod_{i=1}^n f_i(x_i)} d\mathbf{x} \\ &\int_{\mathbf{x}} c(\mathbf{u_x}) \prod_{i=1}^n f_i(x_i) \log c(\mathbf{u_x}) d\mathbf{x} = -Ce(\mathbf{x}) \\ &= \mathbb{E}[\log c(\mathbf{u_x})]. \end{aligned}$$

This shows that the MI of random variables is similar to the negative entropy of their corresponding copula distributions. Furthermore, MI does not depend on the marginal distribution of the r.v.'s involved. CE and then MI are measures of association of two or more random variables (e.g., Sun et al., 2019; Embrechts et al., 2003; Jian, 2019; Blumentritt & Schmid, 2012). It turns out that CE or MI are zero if and only if two random variables are strictly independent. This can be proved by switching the joint density with the product of marginal densities in equation I.1.

According to the expressions derived, MI is not necessarily bounded in $[0,1]$, to overcome this problem Joe (1989) proposed

$$\mathcal{C} = \sqrt{1 - \exp(-2 \times I(\mathbf{x}))},$$

which is a normalizing index. The measure $\mathcal{C}$ is properly defined in $[0,1]$, taking values of 0 when two or more r.v. are independent and assuming value of 1 when they have maximal dependence. It is a generalization of the correlation, multiple correlation and partial correlation.

In order to compute expression I.1, we assume that that a r.v. X has a smooth density function and the integral below the curve exist. No further assumptions are needed to obtain CE.

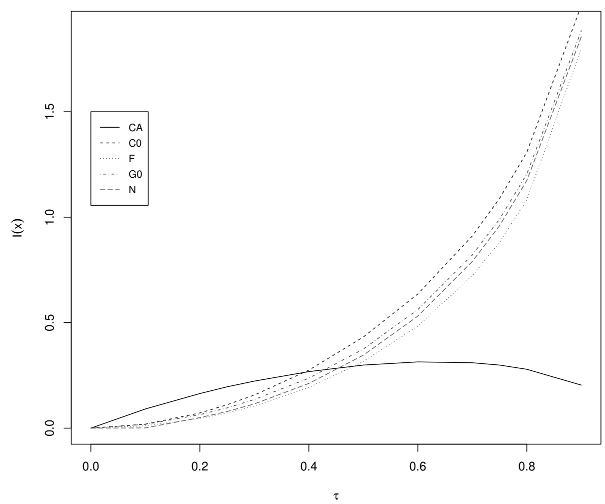


Figure I.1: MI computed for different copulas letting Kendall's τ varying in $[0,1]$. Caudras-Auge CA, Clayton C0, Frank F, Gumbel G0, Gaussian N.

CE and MI have several desirable properties, MI is independent respect to reparametrization. Invariant to monotonic transformations, if X' and Y' are homeomorphism with finite jacobian determinants. It can be proved that $I(X,Y) = I(X',Y')$. It also additive, which allows MI to be decomposed into hierarchical levels. By iterating it, one can decompose $\boldsymbol{I}(X_1,\ldots,X_n)\ \forall n > 2$ and for any partitioning of the set $\{X_1,\ldots,X_n\}$ into the MI between elements within one cluster and MI between clusters, (Appendix I for complete proofs of these results). Theoretically, CE has several advantages over traditional association measures (Jian, 2019; Tenzer & Elidan, 2016; Cover & Thomas, 2006; Nelsen, 2006).

Can be proved that there is a monotonicity dependence between MI and dependence parameter for bivariate copulas θ, with only exception for Caudras-Auge copula (Bolbolian, 2020), see Figure I.1. This conjecture has been formally stated by Tenzer & Elidan (2016). This means that the MI is monotonic increasing (decreasing) in θ if certain conditions hold for the copula family. This result has been generalized to the multivariate case.

Mutual Information for some bivariate distributions

The analytical derivation of MI is quite involved and some times simulation is required in order to obtain an estimate. Out of the domain of Caudras-Auge, Gaussian and T copulas, a closed form does not exist.

The Cuadras & Auge (C & A) copula, was defined in Cuadras & Augé (1981)

$$C(u,v) = [\min(u,v)]^{\theta}(uv)^{1-\theta}, \ \theta \in [0,1],$$

this copula family is obtained by considering a weighted geometric mean of the Independence distribution of the upper Frechet-Hoeffding bounds (Nelsen, 2006). The copula density is defined

$$c(u,v) = (1-\theta)[\max(u,v)]^{-\theta} + \theta u^{1-\theta}\mathbf{1}_{\{u=v\}},$$

in case of bivariate random variables, Mercier (2005) derived a closed form for MI

$$MI = -\frac{2(1-\theta)}{2-\theta}\left[\log(1-\theta) + \frac{\theta}{2-\theta}\right].$$

The Gaussian copula is defined by

$$C(u,v) = \Phi_{\theta}(\Phi^{-1}(u), \Phi^{-1}(v)),$$

where Φ_{θ} is the distribution function of a bivariate standard normal distribution with correlation parameter θ, Φ^{-1} denotes the inverse of the univariate normal distribution function (Meyer, 2013). The Gaussian copula can be expressed as

$$C(u,v)=\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{\exp\left(-\frac{t^2+s-2\theta ts}{2(1-\theta^2)}\right)}{2\pi\sqrt{1-\theta^2}}dtds,\ \ \theta\in[0,1]$$

The density is

$$c(u,v)=\sqrt{1-\theta^2}\exp\left(-\frac{x^2+y^2-2\theta xy}{2(1-\theta^2)}+\frac{x^2+y^2}{2}\right),$$

where $x=\phi^{-1}(u)$ and $y=\phi^{-1}(v)$. Kullback (1952) proved that MI is

$$I=-\frac{1}{2}\log(1-\rho^2).$$

The T-copula is defined as

$$C(u,v)=t_{\theta,\nu}(t_{\theta,\nu}^{-1}(u),t_{\theta,\nu}^{-1}(v)),$$

which is completely specified by the correlation parameter θ and ν degree of freedom. With $t_{\theta,\nu}$ and $t_{\theta,\nu}^{-1}$ we defined the distribution function and its inverse respectively. Therefore, the $T-$copula can be explicitly expressed as

$$C(u,v)=\int_{-\infty}^{t^{-1}(u)}\int_{-\infty}^{t^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\left[1+\frac{t^2+s^2-2\theta ts}{\nu(1-\theta^2)}\right]^{-\frac{\nu+2}{2}}dtds\ \theta\in[0,1],$$

the $T-$copula density is

$$c(u,v)=\frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{[\Gamma(\frac{\nu+1}{2})]^2\sqrt{1-\theta^2}}\frac{[1+\frac{a_\theta(t_\nu^{-1}(u),t_\nu^{-1}(v))}{\nu}]^{-\frac{\nu+2}{2}}}{[1+\frac{(t_\nu^{-1}(u))^2}{\nu}]^{-\frac{\nu+1}{2}}[1+\frac{(t_\nu^{-1}(v))^2}{\nu}]^{-\frac{\nu+1}{2}}},$$

where $a_\theta(x,y)=\frac{x^2+y^2-2\theta xy}{1-\theta^2}$ and $\Gamma(\cdot)$ denotes the Euler-Gamma function. It can be proved that that the MI can be decomposed as

$$I=I_{Gauss}(\theta)+I_{Excess}(\nu),$$

where $I_{Gauss}(\theta) = -0.5\log(1-\theta)$ and

$$I_{Excess}(\nu) = 2\log\left[\sqrt{\nu/2\pi}\beta(\frac{\nu}{2},\frac{1}{2})\right] - \frac{2+\nu}{\nu} + (1+\nu)\left[\psi(\frac{\nu+1}{2}) - \phi(\frac{\nu}{2})\right],$$

$\psi(\cdot)$ and $\beta(\cdot)$ are the Digamma and beta functions. For a complete proof of this result Guerrero-Cusumano (1996b,a); Calsaverini & Vicente (2009). $I_{Excess}(\nu)$ does not depend upon the dependence parameter.

Concentration bound

The aim of this section is to state formally that the concentration bound presented in Baranowski et al. (2020) holds for MI. As far as author knowledge, this result is a novelty in literature.

Proof. (*MI concentration bound*) Let Y and Y' be two random variables defined in Ω and Ω' respectively. And let $f(y,y')$, $f(y)$, $f(y')$ the joint density and the marginals, it a standard result to prove that the $I(Y,Y')$ has an upper bound. A similar proof can be found in Cover & Thomas (2006).

$$\begin{aligned} I(Y,Y') &= \int_{y\in\Omega}\int_{y'\in\Omega'} \log\frac{f(y,y')}{f(y)f(y')}dydy' \\ &= H(Y) + H(Y') - H(Y,Y') \\ &\text{where } \ H(Y) \ \text{ is Shannon Entropy} \\ &= H(Y) - H(Y|Y') \\ &\leq H(Y) = \mathbb{E}_y[\log f(y)^{-1}] \\ &\leq \log\left[\mathbb{E}[f(y)^{-1}]\right] \\ &= \log[\int_y f(y)^{-1}f(y)dy] = \log|\Omega|. \end{aligned}$$

Using the same argument on Y' we can conclude that

$$I(Y,Y') \leq \log min(|\Omega|,|\Omega'|).$$

With the above statement, we have now all the ingredients to prove that the concentration bound holds for MI

$$\begin{aligned}
\mathbb{P}\Bigg(|\hat{I}(Y,Y'|\mathbf{z}_i) - I(Y,Y')| \geq \gamma_\theta n^{-\theta}\Bigg) &\leq \frac{\mathbb{V}[\hat{I}(Y,Y'|\mathbf{z}_i)]}{\gamma_\theta^2 n^{-2\theta}} \\
&= \frac{\mathbb{E}(\hat{I}(Y,Y'|\mathbf{z}_i)^2) - I(Y,Y')^2}{\gamma_\theta^2 n^{-2\theta}} \\
&\leq \frac{\mathbb{E}(\hat{I}(Y,Y'|\mathbf{z}_i)^2)}{\gamma_\theta^2 n^{-2\theta}} \\
&\leq \frac{\mathbb{E}[(\log min(|\Omega|, |\Omega'|))]}{\gamma_\theta^2 n^{-2\theta}} \\
&= \frac{(\log min(|\Omega|, |\Omega'|))}{\gamma_\theta^2 n^{-2\theta}},
\end{aligned}$$

where the second line comes from $\mathbb{V}[\hat{I}(Y,Y'|\mathbf{z}_i)] = \mathbb{E}\{[\hat{I}(Y,Y'|\mathbf{z}_i) - I(Y,Y')]^2\}$, noting that $\mathbb{E}[\hat{I}(Y,Y'|\mathbf{z}_i)] = \mathbb{E}[\mathbb{E}(\log \frac{f_{Y,Y'}}{f_Y f_{Y'}}|\mathbf{z}_i)) = \mathbb{E}(\log \frac{f_{Y,Y'}}{f_Y f_{Y'}})$, thus we can write the variance as $\mathbb{V}[\hat{I}(Y,Y'|\mathbf{z}_i)] = \mathbb{E}[\hat{I}(Y,Y'|\mathbf{z}_i)^2] - I(Y,Y')^2$. The third line is justified by the mutual information property $I(Y,Y') \geq 0$. The fourth line comes from the mutual information lower bound. $\square$

Properties

In this section we are going to present some well known results about MI.

Proof. (*Invariant*) Let X and Y be two random variable with density $f(x, y)$ and X', Y' be homeomorphism (smooth and unique invertible maps. Let $|\nabla f_X| = ||\partial X/\partial X'||$ and $|\nabla f_Y| = ||\partial Y/\partial Y'||$ be the jacobian determinants. Therefore, $f'(x', y') = |\nabla f_X||\nabla f_Y|f(x, y)$, $f'(x') = |\nabla f_X|f(x)$ and $f'(y') = |\nabla f_Y|f(y)$. We can write

$$\begin{aligned} I(X', Y') &= \int\int \log \frac{f'(x', y')}{f'(x')f'(y')} f'(x', y') dx'dy' \\ &= \int\int \log \frac{f(x, y)}{f(x)f(y)} f(x, y) dxy \\ &= I(X, Y) \end{aligned}$$

□

Proof. (*Decomposition*) Let us consider an homeomorphism $(X', Y') = F(X, Y)$. Therefore

$$\begin{aligned} I(X', Y', Z) &= I[(X', Y'), Z] + I(X', Y') \\ &= I[(X, Y), Z] + I(X', Y') \\ &= I(X, Y, Z) - I(X, Y) + I(X', Y') \\ &= I(X, Y, Z) + [I(X', Y') - I(X, Y)] \end{aligned}$$

where the second line comes from the additivity property. □

Proof. (*Correlation*). Let $(X_1, \ldots, X_n)^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $diag(\boldsymbol{\Sigma}) = (\sigma_1^2, \ldots, \sigma_n^2)$, then MI can be written as $I(\mathbf{x}) = -0.5 \log |\boldsymbol{\Sigma}| / \prod_i \sigma_i^2$. For $n = 2$, $\mathcal{C} = |\rho|$, where ρ is the correlation coefficient.

(*Multiple Correlation*). if $(Y, \mathbf{X}) \sim \mathcal{N}\left[\begin{pmatrix}\mu_Y \\ \mu_{\mathbf{X}}\end{pmatrix}, \begin{pmatrix}\sigma_Y^2 & \mathbf{\Sigma}_{Y\mathbf{X}} \\ \mathbf{\Sigma}_{\mathbf{X}Y} & \mathbf{\Sigma}_{\mathbf{XX}}\end{pmatrix}\right]$, the $\mathcal{C} = \sqrt{[\mathbf{\Sigma}_{Y\mathbf{X}}\mathbf{\Sigma}_{\mathbf{XX}}^{-1}\mathbf{\Sigma}_{\mathbf{X},Y}/\sigma_Y^2]}$ which is the square root of the multiple correlation coefficient.

(*Partial Correlation*). if $(Y, X, \mathbf{Z}) \sim \mathcal{N}\left[\begin{pmatrix}\mu_Y \\ \mu_X \\ \mu_{\mathbf{Z}}\end{pmatrix}, \begin{pmatrix}\sigma_{YY} & \sigma_{YX} & \mathbf{\Sigma}_{Y\boldsymbol{Z}} \\ \sigma_{XY} & \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{X\boldsymbol{Z}} \\ \mathbf{\Sigma}_{\boldsymbol{Z}Y} & \mathbf{\Sigma}_{\boldsymbol{Z}X} & \mathbf{\Sigma}_{\boldsymbol{ZZ}}\end{pmatrix}\right]$, then $\mathcal{C}_{YX|\boldsymbol{Z}} = \sqrt{(a_2^2/a_1a_3)}$

where $a_1 = \sigma_{YY} - \mathbf{\Sigma}_{Y\boldsymbol{Z}}\mathbf{\Sigma}_{\boldsymbol{ZZ}}^{-1}\mathbf{\Sigma}_{\boldsymbol{Z}Y}$, $a_3 = \sigma_{XX} - \mathbf{\Sigma}_{X\boldsymbol{Z}}\mathbf{\Sigma}_{\boldsymbol{ZZ}}^{-1}\mathbf{\Sigma}_{\boldsymbol{Z}X}$, $a_2 = \sigma_{YX} - \mathbf{\Sigma}_{Y\boldsymbol{Z}}\mathbf{\Sigma}_{\boldsymbol{ZZ}}^{-1}\mathbf{\Sigma}_{\boldsymbol{Z}X}$

□

Appendix J

Simulation Study: Mutual Information

This section provides evidence that the MI can be a suitable option in the BivariateRBVS algorithm. The simulation experiments discussed refers to a general class of Flexible Copula Regression models (see. Stasinopoulos & Rigby (2008), Wood (2017)) fitted using `gjrm` function in `GJRM` package in `R`. This section does not have the ambition or the purpose of being comprehensive, it refers to some preliminary simulation studies reported for completeness. The following results concern a preliminary study, whose aim was to get the limits of MI as tool in variable selection procedure.

Let $\mathbf{X} \in \mathbb{R}^{n\times p}$ a partitioned blocks data matrix such that $\mathbf{X} = [\mathbf{X}_{11} : \mathbf{X}_{12}]$, where $\mathbf{X}_{11} \in \mathbb{R}^{n\times 3}$ contains the relevant covariates whose have been 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, while $\mathbf{X}_{12} \in \mathbb{R}^{n\times(p-3)} \sim \mathcal{N}_{p-3}(\mathbf{0}, \mathbf{\Phi})$ have been generated using a multivariate normal distribution with null mean vector and a covariance matrix $\mathbf{\Phi}$. Y_{i1} and Y_{i2} have been generated from an Inverse Gaussian distribution `iG` specified through the following set of equations.

$$\begin{aligned}
\boldsymbol{\eta}_{\mu_1} &= \beta_{10} + \beta_{11}\mathbf{x}_2 + \beta_{12}\mathbf{x}_3 \\
\boldsymbol{\eta}_{\mu_2} &= \beta_{20} + \beta_{21}\mathbf{x}_1 + \beta_{22}\mathbf{x}_2 \\
\boldsymbol{\eta}_{\sigma_1} &= \beta_{30} \\
\boldsymbol{\eta}_{\sigma_2} &= \beta_{40} \\
\boldsymbol{\eta}_{\theta} &= \beta_{50} + \beta_{51}\mathbf{x}_1 + \beta_{52}\mathbf{x}_2
\end{aligned}$$

where $\boldsymbol{\eta}_{\mu_v}$, $\boldsymbol{\eta}_{\sigma_v}$ $v = 1, 2$ are the linear predictors used to specify shape and scale parameters, while $\boldsymbol{\eta}_{\theta}$ specifies the dependence parameter. Y_1 and Y_2 were joined using a Joe `J0` copula. In practice this was achieved using iterative conditioning. The sets of population parameters are $\boldsymbol{\beta}_1^T = (0.5, -1.25, -0.8)^T$, $\boldsymbol{\beta}_2^T = (0.1, -0.9, 3)^T$, $\beta_{30} = 1.8$, $\beta_{40} = 0.1$ and $\boldsymbol{\beta}_5^T = (0.2, 0.7, -4.3)^T$.

The set of tuning parameters was specified such that $n_{sub} = 10$, $B = 50$, $\tau = 0.5$, $k_{max} = 6$, with a sample size of $n = 200$ and $p = 30$ covariates.

Finally the algoritm discussed in chapter 2 was carried out by fitting a Copula link Based Additive Model using `gjrm` with inverse Gaussian margins `iG` and Joe copula `J0`. The set of equations used internally have been specified in the following way

$$\begin{aligned}
\boldsymbol{\eta}_{\mu_1} &= \beta_{10} + \beta_{11}\mathbf{x}_j \\
\boldsymbol{\eta}_{\mu_2} &= \beta_{20} + \beta_{21}\mathbf{x}_{j'} \\
\boldsymbol{\eta}_{\sigma_1} &= \beta_{30} \\
\boldsymbol{\eta}_{\sigma_2} &= \beta_{40} \\
\boldsymbol{\eta}_{\theta} &= \beta_{50} \\
&\text{for } j'j = 1, \ldots, p
\end{aligned}$$

From the first simulation scenario we set $\rho_{X_{12}} = 0$, while for the second $\rho_{X_{12}} = 0.2$. The results are showed in Table J.1 and J.2. We can appreciate an overall good performance of the metric in recovering the important covariates ($\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$). For the first set up $\hat{FP} \approx 0.105$

and $\hat{FN} \approx 0$, in our second scenario the performance slightly worsened with $\hat{FP} \approx 0.156$ and $\hat{FN} \approx 0$.

$\hat{s}^{(1)}$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	0	0	0	0
Freq.	15	9	3	1	1	1
$\hat{s}^{(2)}$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	0	0	0	0
Freq.	27	21	4	1	1	1
$\hat{s}^{(3)}$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_2 : \mathbf{x}_3$	0	0	0
Freq.	24	26	15	1	1	1
$\hat{s}^{(4)}$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_13$	$\mathbf{x}_7 : \mathbf{x}_{22}$	0	0
Freq.	33	6	1	1	1	1
$\hat{s}^{(5)}$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_3$	0	0	0	0
Freq.	16	8	1	1	1	1
$\hat{s}^{(6)}$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_{10} : \mathbf{x}_{27}$	$\mathbf{x}_9 : \mathbf{x}_{10}$	0
Freq.	27	11	4	1	1	1
$\hat{s}^{(7)}$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_2 : \mathbf{x}_3$	0	0	0
Freq.	16	7	6	2	1	1
$\hat{s}^{(8)}$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	0	0	0
Freq.	19	8	8	1	1	1
$\hat{s}^{(9)}$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_2 : \mathbf{x}_3$	0	0	0
Freq.	19	12	6	1	1	1
$\hat{s}^{(10)}$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_{11} : \mathbf{x}_{13}$	$\mathbf{x}_{10} : \mathbf{x}_{30}$	$\mathbf{x}_{10} : \mathbf{x}_24$	0
Freq.	10	5	2	1	1	1

Table J.1: Best subsets with associated frequency results, setup: $n_{sub} = 10$, $B = 50$, $\tau = 0.5$, $k_{max} = 6$, $n = 200$, $p = 30$, $\rho_{X_{12}} = 0$.

$\hat{s}^{(1)}$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_{19} : \mathbf{x}_{23}$	$\mathbf{x}_1 : \mathbf{x}_2$	0	0
Freq.	15	4	1	1	1	1
$\hat{s}^{(2)}$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_{10} : \mathbf{x}_{24}$	$\mathbf{x}_{11} : \mathbf{x}_{12}$	0	0
Freq.	10	4	1	1	1	1
$\hat{s}^{(3)}$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_4 : \mathbf{x}_9$	$\mathbf{x}_1 : \mathbf{x}_2$	0	0	0
Freq.	9	1	1	1	1	1
$\hat{s}^{(4)}$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_2 : \mathbf{x}_3$	0	0	0	0
Freq.	17	7	1	1	1	1
$\hat{s}^{(5)}$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_2 : \mathbf{x}_{27}$	$\mathbf{x}_{14} : \mathbf{x}_{19}$	0
Freq.	10	4	2	1	1	1
$\hat{s}^{(6)}$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_2 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_{19} : \mathbf{x}_{27}$	$\mathbf{x}_{15} : \mathbf{x}_{23}$	0
Freq.	14	7	3	1	1	1
$\hat{s}^{(7)}$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_2 : \mathbf{x}_3$	0	0	0
Freq.	19	2	5	1	1	1
$\hat{s}^{(8)}$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_{12} : \mathbf{x}_{23}$	$\mathbf{x}_{13} : \mathbf{x}_{19}$	0	0
Freq.	21	3	1	1	1	1
$\hat{s}^{(9)}$	$\mathbf{x}_1 : \mathbf{x}_2$	$\mathbf{x}_1 : \mathbf{x}_3$	$\mathbf{x}_2 : \mathbf{x}_3$	0	0	0
Freq.	5	1	1	1	1	1
$\hat{s}^{(10)}$	$\mathbf{x}_2 : \mathbf{x}_3$	0	0	0	0	0
Freq.	25	1	1	1	1	1

Table J.2: Best subsets with associated frequency results, setup: $n_{sub} = 10$, $B = 50$, $\tau = 0.5$, $k_{max} = 6$, $n = 200$, $p = 30$, $\rho_{X_{12}} = 0.2$.

Appendix K

RBVS results

Proof of Proposition 2.1 in Baranowski et al.(2018)

Proposition 2.1 (Baranowski et al., 2020). Let $\mathbf{R}_n$ be a variable ranking based on $\hat{\omega}_j,\ j = 1\dots,p$. Where $\mathbf{R}_n = (R_{n1},\dots,R_{np})$ be a variable ranking computed on n observations based on $\hat{\omega}_1,\dots,\hat{\omega}_p$ satisfying $\hat{\omega}_1 > \cdots > \hat{\omega}_p$. And let $\omega_j^\nu(\mathbf{x}_j)$ with $j = 1,\dots,p,\ \nu = 1,2$ be the metrics obtained by considering the $j-$th covariate in the $\nu-$th margin. Without loss of generality, we can assume that the metric employed depend on $\mathbf{Z} = \{Y_1, Y_2, X_1, \dots, X_p\}$. Therefore, it change with n. Let $\hat{\omega}_j^\nu = \hat{\omega}_j^\nu(\mathbf{Z}_1,\dots,\mathbf{Z}_n)$ be an estimator of ω_j^ν. Under the following conditions

(C$_1$) Let $\mathbf{Z}_1,\dots,\mathbf{Z}_n$ be independent random samples, then $\exists\,\theta > 0$ and any $\gamma_\theta > 0$ such that;

$$\max_{j=1,\dots,p} \mathbb{P}\left(|\hat{\omega}_j^\nu - \omega_j^\nu| \geq \gamma_\theta n^{-\theta}\right) \leq \Gamma_\theta \exp\{-n^\psi\}\ \nu = 1,2,$$

where constants Γ_θ, ψ does not depend on n;

(C$_2$) The index sets of important variables is denoted as $S^\nu \subset \{1,\dots,p\}$. S^ν does not depend on n or p, and could potentially be an empty set;

(C$_3$) $\forall\ a \notin S^\nu\ \exists\,\mathcal{M}_a \subset \{1,\dots,p\} \setminus S^\nu : a \in \mathcal{M}_a$, the distribution of $\{\omega_j^\nu\}_{j\in\mathcal{M}_a}$is exchangeable and $|\mathcal{M}_a| \underset{n}{\rightarrow} \infty$;

(C$_4$) $\exists\,\eta \in (0,\theta]$, such that θ is as is (E_1), and $\gamma_\eta > 0 : \min_{j\in S^\nu} \omega_j^\nu - \max_{j\notin S^\nu} \omega_j^\nu \geq \gamma_\eta n^{-\eta}$;

(C$_5$) $p \leq C_1 \exp\{n^{b_1}\}$ $b_1 \in (0,\gamma)$ and γ is as in (E_1) ;

the unique top-ranked set stated in Definition 2.3 of Baranowski et al. (2020) exist and equals S.

Proof. Firstly, we want to show that $\pi_n(S) \to 1$. Let $\mathcal{E} = \{\min_{j\in S}\hat{\omega}_j > \max_{j\notin S}\hat{\omega}_j\}$. If no tie, $\mathcal{E} \equiv \{\{R_{n1},\ldots,R_{ns}\} = S\}$, meaning that all indeces from S are ranked in front of those do not belong to S. Using (E$_4$) we can state that

$$\pi_n(S) \geq \mathbb{P}(\mathcal{E}) \geq \mathbb{P}\left(\max_{j=1,\ldots,p}|\hat{\omega}_j - \omega_j| < \epsilon\right),$$

$\epsilon = \gamma_\eta n^{-\eta}/2$. Applying the Bonferroni's inequality it yields

$$\mathbb{P}\left(\max_{j=1,\ldots,p}|\hat{\omega}_j - \omega_j| < \epsilon\right) \geq (1-p)\sup_{j=1,\ldots,p}\mathbb{P}(|\hat{\omega}_j - \omega_j| \geq \epsilon),$$

the last term is of order $1 - \mathcal{O}(\exp(-n^{\psi}))$ as $(b_1 < \psi)$, which tends to 1 as $n \to \infty$. This proves that S is a $s-$top ranked set, where $s = |S|$.

Secondly, consider any $\mathcal{A} \in \Omega_{s+1}$. We will prove that $\pi_n(\mathcal{A}) \underset{n}{\to} 0$. Note that $\mathcal{E}$ implies that $S \subset \mathcal{A}$ as all indices from S are ranked in front of those who does not belong to S. Therefore, it is sufficient to consider the case of $S \subset \mathcal{A}$ such that $\mathcal{A}\backslash S$ has only one element, denoted by a. As in the previous point, we suppose that there is no tie in the ranking, on the event $\mathcal{E}$, we have

$$\begin{aligned}&\mathbb{P}(\{\min_{j\in\mathcal{A}}\hat{\omega}_j > \max_{j\notin\mathcal{A}}\hat{\omega}_j\} \cap \{\min_{j\in S}\hat{\omega}_j > \max_{j\notin S}\hat{\omega}_j\})\\ =&\mathbb{P}(\{\hat{\omega}_a > \max_{j\notin\mathcal{A}}\hat{\omega}_j\} \cap \mathcal{E}),\end{aligned}$$

we observe that $\mathbb{P}(\hat{\omega}_a > \max_{j\notin\mathcal{A}}\hat{\omega}_j) \leq \mathbb{P}(\hat{\omega}_a > \max_{j\in\mathcal{M}_a\backslash\{a\}}\hat{\omega}_j)$. Using E$_3$, we have that the probability of $\mathbb{P}(\hat{\omega}^*_j > \max_{j\in\mathcal{M}_a\backslash\{j^*\}}\hat{\omega}_j)$ is the same of any $j^* \in \mathcal{M}_a$ (i.e. any element of $\{\hat{\omega}_j\}_{j\in\mathcal{M}_a}$ are equally likely to be the largest). Observing that $\sum_{j^*\in\mathcal{M}_a}\mathbb{P}(\hat{\omega}^*_j > \max_{j\in\mathcal{M}_a\backslash\{j^*\}}\hat{\omega}_j) \leq 1$, we have that $\mathbb{P}(\hat{\omega}_a > \max_{j\in\mathcal{M}_a\backslash\{a\}}\hat{\omega}_j) \leq |\mathcal{M}_a|^{-1} \underset{n}{\to} 0$. Consequently we an argue that

$$\pi_n(\mathcal{A}) \leq \mathbb{P}\left(\hat{\omega}_a > \max_{j\notin\mathcal{A}}\hat{\omega}_j\right) + \mathbb{P}(\bar{\mathcal{E}}) \leq \mathbb{P}\left(\hat{\omega}_a > \max_{j\in\mathcal{M}_a\backslash a}\hat{\omega}_j\right) + \mathbb{P}(\bar{\mathcal{E}}) \underset{a}{\to} 0.$$

Where with $\bar{\mathcal{E}}$ we denote the negation event of $\mathcal{E}$. Otherwise, if there is a ties in the ranking, since we break the ties at random uniformly, it follows from the exchangeability assumption that are we are equally likely to pick any index from $\mathcal{M}_a$, given that we have picked one of them. Thus we car argue that is a similar manner to show that $\pi_n(\mathcal{A}) \leq 1/|\mathcal{M}_a| + \mathbb{P}(\bar{\mathcal{E}}) \underset{n}{\rightarrow} 0$, i.e. S is always locally top ranked.

Thirdly, $\forall k' = 1, \ldots, s-1$, we can show that $\exists \mathcal{A} \in \Omega_{k'} : \limsup_{n\to\infty} \pi_n(\mathcal{A}) > 0$. Note that

$$\sum_{\{\mathcal{A}:\mathcal{A}\in\Omega'_k \text{ and } \mathcal{A}\subset S\}} \pi_n(\mathcal{A}) \geq \mathbb{P}\left(\min_{j\in S} \hat{\omega}_j > \max_{j\notin S} \hat{\omega}\right) \underset{n}{\rightarrow} 1,$$

from the previous argument. However, there are $\binom{s}{k'}$ elements in $\{\mathcal{A} : \mathcal{A} \in \Omega'_k \text{ and } \mathcal{A} \subset S\}$, so

$$\max_{\{\mathcal{A}:\mathcal{A}\in\Omega'_k \text{ and } \mathcal{A}\subset S\}} \limsup_{n\to\infty} \pi_n(\mathcal{A}) \geq \frac{1}{\binom{s}{k'}}.$$

This implies that S is indeed a top ranked set.

Finally the uniqueness of S follows from the fact that $\pi_n(S) \underset{n}{\rightarrow} 1$ and $\sum_{\mathcal{A}\in\Omega_s} \pi_n(\mathcal{A}) = 1$ $\square$

Auxiliary lemmas

Lemma 1. *(Hoeffding (1963)). Let W be a binomial random variable with the probability of success π and r trials. For any $1 > t > \pi$, we have $\mathbb{P}(W \geq rt) \leq (\frac{\pi}{t})^{rt}(\frac{1-\pi}{t})^{r(1-t)}$. Moreover, for any $0 < t < \pi$, $\mathbb{P}(W \leq rt) \leq (\frac{\pi}{t})^{rt}(\frac{1-\pi}{1-t})^{r(1-t)}$*

Lemma 2. *Let $a_1, \ldots, a_l$ be non-negative numbers s.t. $\sum_{i=1}^{l} a_i \leq 1$ and $\max a_i \leq t$ for some $\frac{1}{l} \leq t \leq 1$. Let $N \in \mathbb{N}$ be the minimum integer such that there exist mutually sets $I_1, \ldots, I_n \subset \{1, \ldots, l\}$ with $\sum_{i \in I_j} a_i \leq t$ and $\bigcup_{j=1}^{N} I_j = \{1, \ldots, l\}$. Then, $N \leq \lfloor \frac{2}{t} \rfloor + 1$.*

Proof. Since N is the smallest possible integer, there must be at most one $j \in \{1, \ldots, N\}$ with $\sum_{i \in I_j} ai \leq \frac{t}{2}$. Otherwise, such two sets could be combined, leading to a smaller N. So for all other $j \in \{1, \ldots, N\}$, we have that $\sum_{i \in I_j} a_i > \frac{t}{2}$. Consequently, $(N-1)t/2 \leq \sum_{i=1}^{l} a_i \leq 1$. This implies that $N \leq \lfloor \frac{2}{t} \rfloor + 1$ □

Lemma 3. *Let $\Omega \subset \Omega_k$ for some $k = 1, \ldots, p$, $m \leq n$, $B \geq 1$, and t_1, t_2 satisfying $\max_{\mathcal{A} \in \Omega} \pi_{m,n}(\mathcal{A}) \leq t_2 < t_1 < 1$. Then $\mathbb{P}(\max_{\mathcal{A} \in \Omega} \hat{\pi}_{m,n}(\mathcal{A}) \geq t_1) \leq \frac{3B}{t_2} \left[\left(\frac{t_2}{t_1} \right)^{t_1} \left(\frac{1-t_2}{1-t_1} \right)^{1-t_1} \right]^r$.*

Proof. Denote $\mathcal{A}^1, \ldots, \mathcal{A}^l$ all the elements of ω. Applying Lemma 2 in Baranowski et al. (2020) we find a partition $I_1, \ldots, I_N$ such that $\max_{j=1,\ldots,N} \sum_{i \in I_j} \pi_{m,n}(\mathcal{A}^i) \leq t_2$ and $N \leq \frac{2}{t_2} + 1$. Using the union bound, we have that

$$\mathbb{P}\left(\max_{i=1,\ldots,l} \hat{\pi}_{m,n}(\mathcal{A}^i) \geq t_1 \right) \leq N \max_{j=1,\ldots,N} \mathbb{P}\left(\sum_{i \in I_j} \hat{\pi}_{m,n}(\mathcal{A}^i) \geq t_1 \right).$$

Note that when $B = 1$, we have that $r \sum_{i \in I_j} \hat{\pi}_{m,n}(\mathcal{A}^i)$ is a binomial random variable, which r is the number of trials with the probability of success denoted by $p_j^* = \sum_{i \in I_j} \pi_{m,n}(\mathcal{A}^i)$. From Lemma 1 we can conclude that

$$\mathbb{P}\left(\sum_{i \in I_j} \hat{\pi}_{m,n}(\mathcal{A}^i) \geq t_1 \right) \leq \left[\left(\frac{p_j^*}{t1} \right)^{t_1} \left(\frac{1-p_j^*}{1-t_1} \right)^{1-t_1} \right]^r \leq \left[\left(\frac{t_2}{t1} \right)^{t_1} \left(\frac{1-t_2}{1-t_1} \right)^{1-t_1} \right]^r,$$

this chain of inequality can be achieved by recognizing that fact that $(\frac{x}{t_1})^{t_1}(\frac{1-x}{1-t_1})^{(1-t_1)}$ is increasing for $x \in [0, t_1]$. When $B = 1$ and $N \leq 3/t_2$, lead us to rewrite the expression as

$$\mathbb{P}\left(\max_{i=1,\dots,l} \hat{\pi}_{m,n}(\mathcal{A}^i) \geq t_1\right) \leq \frac{3}{t_2}\left[\left(\frac{t_2}{t1}\right)^{t_1}\left(\frac{1-t_2}{1-t_1}\right)^{1-t_1}\right]^r.$$

Finally, when $B > 1$,$r\sum_{i\in I_j} \hat{\pi}_{m,n}(\mathcal{A}_j)$ is a sample average of B (not necessarily independent) binomial random variables. Since the average of a collection of non-negative numbers is always no greater that its maximum, we could simply use the union bound again to establish that

$$\mathbb{P}\left(\max_{i=1,\dots,l} \hat{\pi}_{m,n}(\mathcal{A}^i) \geq t_1\right) \leq \frac{3B}{t_2}\left[\left(\frac{t_2}{t1}\right)^{t_1}\left(\frac{1-t_2}{1-t_1}\right)^{1-t_1}\right]^r.$$

$\square$

Proof of Theorem 2.1 in Baranowski et al.(2018)

Proof. We define $\hat{\omega}_{j,m} = \hat{\omega}(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$, $\delta = \pi_{m,n}(S)$ and $\theta = \max_{\mathcal{A}\not\subset S, |\mathcal{A}|\leq k_{max}} \pi_{m,n}(\mathcal{A})$. The proof start showing that θ and δ are well separated for sufficiently large n.

Take $\epsilon = \frac{\gamma_\eta m^{-\eta}}{2}$. using ($A_1$) and E_5 combined with Bonferroni's inequality, we get $\delta \geq \mathbb{P}(\max_{j=1,\ldots,p} |\hat{\omega}_{j,m} - \omega_j| < \epsilon) \geq 1 - \Gamma_\epsilon p e^{-m^\psi}$. Considering ($A_2$) and ($A_3$), since we assume that $\psi b2 > b1$, we gen that $\delta = 1 - \mathcal{O}(e^{-n^{\psi b2}})$, which tends to one as $n \to \infty$.

For every $A \in \Omega_k$ with $k \leq k_{max}$ that contains at least one $a \in \mathcal{A}\backslash S$, if there is no tie in the ranking of $\{\hat{\omega}_{j,m}\}_{j\in[1,p]}$, we have that

$$\begin{aligned}\pi_{m,n}(\mathcal{A}) =& \mathbb{P}\left(\min_{j\in\mathcal{A}} \hat{\omega}_{j,m} > \max_{æ\notin\mathcal{A}} \hat{\omega}_{j,m}\right) \leq \mathbb{P}\left(\hat{\omega}_{a,m} > \max_{j\in\mathcal{M}_a\backslash\mathcal{A}} \hat{\omega}_{j,m}\right) \\ &\leq \frac{1}{|\mathcal{M}_a| - k_{max}} \leq \frac{1}{\min_{a\notin S} |\mathcal{M}_a| - k_{max}} \leq \frac{1}{\Gamma_3 n^{b_3} - \Gamma_4 n^{b_4}},\end{aligned} \tag{K.1}$$

Here we can use the exchangeability of $\{\hat{\omega}_{j,m j\in\mathcal{M}_a\backslash S}\}$ together with (A_4) and (A_7). Even if there are ties, we still have that $\pi_{n,m}(\mathcal{A}) \leq 1/(\Gamma_3 n^{b_3} - \Gamma_4 n^{b_4})$ due to exchangeability and since we break the ties uniformly at random. Notice that K.1 does not depend on $\mathcal{A}$ or a, so the inequality $\pi_{n,m}(\mathcal{A}) \leq 1/(\Gamma_3 n^{b_3} - \Gamma_4 n^{b_4})$ is valid for $\forall \mathcal{A} \in \Omega_k$ with $k \leq k_{max}$ and $\mathcal{A}\backslash S \neq \emptyset$. As such, we conclude that $\theta = max_{\mathcal{A}\notin S, |\mathcal{A}|\leq k_{max}} \pi_{m,n}(\mathcal{A}) = \mathcal{O}(n^{-b_3})$.

Next, to fix the ideas, take $\delta = (b_2 + b_3 - 1)/2$, $t_1 = n^(-b_3 + \delta)/2$ and $t_2 = t_1^2$. For sufficiently large n we always have $\theta \leq t_1^2 \leq t_1 \leq 1/2 \leq \delta$. Then we can define the events

$$\begin{aligned}\mathcal{E}_k &= \left\{\max_{\mathcal{A}\in\Omega_k, \mathcal{A}\notin S} \hat{\pi}_{m,n}(\mathcal{A}) < t_1\right\}, \text{ for } k = 1, \ldots, k_{max} \\ \mathcal{B} &= \left\{\hat{\pi}_{m,n}(S) > 0.5\right\}, \\ \mathcal{E} &= \mathcal{B} \bigcap_{k=1}^{k_{max}} \mathcal{E}_k,\end{aligned}$$

we are going to prove that $\mathbb{P}(\mathcal{E}) \underset{n}{\to} 1$ at an exponential rate, and with $\hat{\mathcal{A}}_{\hat{s},m} = S$ on the event $\mathcal{E}$.

To prove the first claim when $B = 1$, for sufficiently large n, we culd use Lemma 1 and the fact that $1 - \delta = \mathcal{O}(e^{-n^{\psi b_2}}) \underset{n}{\rightarrow}$ to bound $\mathbb{P}(\bar{\mathcal{B}})$ by

$$\mathbb{P}(\bar{\mathcal{B}}) \leq \left[\left(\frac{\delta}{0.5}\right)^{0.5}\left(\frac{1-\delta}{0.5}\right)^{0.5}\right]^r \leq [2(1-\delta)]^{0.5r} \leq \exp\left(-\Gamma' n^{\psi b_2(1-b_2)/2}\right), \tag{K.2}$$

for some $\Gamma' \in (0,1)$. When $B > 1$, since $\hat{\pi}_{m,n}(S)$ is the average of B copies of the that with $B = 1$, using once again the Bonferroini bound, we have that $\mathbb{P}(\bar{\mathcal{B}}) \leq B\exp\left(-\Gamma' n^{\psi b_2(1-b_2)/2}\right)$. Moreover by Lemma 3,

$$\mathbb{P}(\bar{\mathcal{E}}_k) \leq \frac{3B}{t_2}\left[\left(\frac{t_2}{t_1}\right)_1^t\left(\frac{1-t_2}{1-t_1}\right)^{1-t_1}\right]^r = \frac{3B}{t_1^2}\left[\left(\frac{t_1}{1+t_1}\right)^{t_1}(1+t_1)\right]^r. \tag{K.3}$$

Taking the logarithm, using A6 (for details see. Baranowski et al. (2020)), we have

$$\mathbb{P}(\bar{\mathcal{E}}_k) \leq \frac{3B}{t_1^2}\exp\left(\frac{-rt_1}{6}\right) \leq \exp\left(-\Gamma'' n^{(1-b_2-b_3/2)}\right),$$

$\Gamma'' > 0$ and for sufficiently large n. We know from A_4 that the right hand side of the above inequality goes to 0.

The remaining arguments used in this proof are valid on $\mathcal{E}$ with a sufficiently large n. Notice that from $1/2 > t_1$ one concludes that$\hat{}$

$A_{s,m}$ is given by $\hat{\mathcal{A}}_{k,m} = \arg\max_{\mathcal{A}\in\Omega_k} \hat{\pi}_{m,n}(\mathcal{A})$, hence showing that $\hat{s} = s$ proved $\hat{S} = S$. Denote $T_k = \frac{\hat{\pi}^{\tau}_{m,n}(\hat{\mathcal{A}}_{k+1,m})}{\hat{\pi}_{m,n}(\hat{\mathcal{A}}_{k,m})}$, then from definition, $\hat{s} = \arg\min_{k=0,1,\ldots,k_{max}} T_k$. Three cases are considered

- For every $k = 0,\ldots,s-1$, the event $\{\{R_n(\mathbf{Z}_1,\ldots,\mathbf{Z}_m),\ldots,R_{n,s}(\mathbf{Z}_1,\ldots,\mathbf{Z}_m)\} = S\}$ implying that the index set $\{R_n(\mathbf{Z}_1,\ldots,\mathbf{Z}_m),\ldots,R_{n,k+1}(\mathbf{Z}_1,\ldots,\mathbf{Z}_m)\}$ (i.e. of size $k+1$)

must be one of the elements in $\{\mathcal{A} \in \Omega_{k+1} : \mathcal{A} \subset S\}$. Therefore,

$$\sum_{\{\mathcal{A} \in \Omega_{k+1} : \mathcal{A} \subset S\}} \hat{\pi}_{m,n}(\mathcal{A}) \geq \hat{\pi}_{m,n}(S).$$

Noting that $|\{\mathcal{A} \in \Omega_{k+1} : \mathcal{A} \subset S\}| = \binom{s}{k+1}$ and $\hat{\pi}_{m,n}(S) > \frac{1}{2}$ imply that

$$\hat{\pi}_{m,n}(\hat{\mathcal{A}}_{k+1,m}) \geq \max_{\{\mathcal{A} \in \Omega_{k+1} : \mathcal{A} \subset S\}} \hat{\pi}_{m,n}(\mathcal{A}) \geq \frac{\hat{\pi}_{m,n}(S)}{\binom{s}{k+1}} \geq \frac{1}{2\binom{s}{k+1}},$$

hence $T_k \geq \frac{1}{2\binom{s}{k+1}}$, for $k = 0, \ldots, s-1$.

- Directly from the definition of the event $\mathcal{E}_s$ and $\mathcal{B}$, we bound $T_s \leq 2t_1^\tau$.

- $\hat{\pi}_{m,n}(\hat{\mathcal{A}}_{k+1,m}) \geq (Br)^{-1}$ for any k. To see this note that $\sum_{\mathcal{A} \in \Omega_{k+1}} \hat{\pi}_{m,n}(\mathcal{A}) = 1$. Picking $\hat{\mathcal{A}}_{k+1,m} \in \arg\max_{\mathcal{A} \in \Omega_{k+1}} \hat{\pi}_{m,n}(\mathcal{A})$ would mean that $\hat{\pi}_{m,n}(\hat{\mathcal{A}}_{k+1,m}) > 0$, because otherwise it would imply that $\sum_{\mathcal{A} \in \Omega_{k+1}} \hat{\pi}_{m,n}(\hat{\mathcal{A}}) = 0$, leading to a contradiction. Now that $\hat{\pi}_{m,n}(\hat{\mathcal{A}}_{k+1,m}) > 0$, it must be the case that $\hat{\pi}_{m,n}(\hat{\mathcal{A}}_{k+1,m}) > (Br)^{-1}$. Thus $T_k \geq \frac{1}{t_1 (Br)^\tau}$ for every $k = s+1, \ldots, k_{max}$.

To prove that $T_k > T_s$ for $k = 0, \ldots, s-1$, it is sufficient to demonstrate that $\frac{1}{2\binom{s}{k+1}} > 2t_1^\tau$, which is true for sufficiently large n, as $t_1 \underset{n}{\to} 0$ and $\max_{k=0,\ldots,s-1} \binom{s}{k+1}$ is bounded. Similarly, to claim that $T_s < T_k$ for $k = s+1, \ldots, k_{max}$, we need to show $2t_1^\tau < \frac{1}{t_1(Br)^\tau}$, which amounts to $2t_1^{1+\tau} < (Br)^{-1}$, or $2^{1/\tau} t_1^{1+1/\tau} < (Br)^{-1}$. This is true for sufficiently large n, because $t_1^2 = n^{-b_3+\delta}$, $Br = \mathcal{O}(n^{1-b_2})$ and $b_2 + b_3 - \delta > 1$ from A_4.

Therefore T_k is necessarily minimized at $k = s$ over $\mathcal{E}$ for sufficiently large n, meaning that $\hat{s} = s$, which completes the proof. □

Bibliography

Baranowski, R., Chen, Y., & Fryzlewicz, P. (2020). Ranking-based variable selection for high-dimensional data. *Statistica Sinica*, 30(3), 1485–1516.

Blumentritt, T. & Schmid, F. (2012). Mutual information as a measure of multivariate association: Analytical properties and statistical estimation. *Journal of Statistical Computation and Simulation*, 82, 1257–1274.

Bolbolian, M. (2020). Relationship between kendall's tau correlation and mutual information. *Revista Colombiana de Estadística*, 43, 3–20.

Calsaverini, R. S. & Vicente, R. (2009). An information-theoretic approach to statistical dependence: Copula information. *EPL (Europhysics Letters)*, 88(6), 68003.

Cover, T. M. & Thomas, J. A. (2006). *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience.

Cuadras, C. M. & Augé, J. (1981). A continuous general multivariate distribution and its properties. *Communications in Statistics - Theory and Methods*, 10(4), 339–353.

Embrechts, P., Lindskog, F., & McNeil, A. J. (2003). Chapter 8 – modelling dependence with copulas and applications to risk management.

Guerrero-Cusumano, J.-L. (1996a). An asymptotic test of independence for multivariate t and cauchy random variables with applications. *Information Sciences*, 92(1), 33–45.

Guerrero-Cusumano, J.-L. (1996b). A measure of total variability for the multivariate t distribution with applications to finance. *Information Sciences*, 92(1), 47–63.

Jian, M. (2019). Discovering association with copula entropy. *CoRR*, abs/1907.12268.

Joe, H. (1989). Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association*, 84(405), 157–164.

Kullback, S. (1952). An Application of Information Theory to Multivariate Analysis. *The Annals of Mathematical Statistics*, 23(1), 88 – 102.

Ma, J. & Sun, Z. (2008). Mutual information is copula entropy. *CoRR*, abs/0808.0845.

Marra, G. & Radice, R. (2022). *GJRM: Generalised Joint Regression Modelling*. R package version 0.2-6.

Mercier, G. (2005). Measures de dependance entre images rso. *Technical report*.

Meyer, C. (2013). The bivariate normal copula. *Communications in Statistics - Theory and Methods*, 42(13), 2402–2422.

Nelsen, R. (2006). *An Introduction to Copulas*. Second Edition, Springer, New York.

Stasinopoulos, D. M. & Rigby, R. A. (2008). Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23, 1–46.

Sun, F., Zhang, W., Wang, N., & Zhang, W. (2019). A copula entropy approach to dependence measurement for multiple degradation processes. *Entropy*, 21(8).

Tenzer, Y. & Elidan, G. (2016). On the monotonicity of the copula entropy.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction With R*. Second Edition, Chapman & Hall/CRC, London.