UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-graduação em Estatística

Walmir dos Reis Miranda Filho

Semiparametric modeling for multivariate survival data via copulas

Belo Horizonte
2022

Walmir dos Reis Miranda Filho

Semiparametric modeling for multivariate survival data via copulas

Versão final

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutor em Estatística.

Orientador: Prof. Dr. Fábio Nogueira Demarqui

Belo Horizonte
2022

Miranda Filho, Walmir dos Reis.

M672s Semiparametric modeling for multivariate survival data via copulas [manuscrito] / Walmir dos Reis Miranda Filho. – 2022.
152 f. il.

Orientador: Fábio Nogueira Demarqui.
Tese (doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.
Referências: f. 108-113

1. Estatística – Teses. 2. Análise de sobrevivência (Biometria) – Teses. 3. Análise funcional – Teses. 4. Análise de regressão – Teses. I. Demarqui, Fábio Nogueira. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III.Título.

CDU 519.2(043)

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa
CRB 6/1510 Universidade Federal de Minas Gerais - ICEx

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

ATA DA DEFESA DE TESE DE DOUTORADO DO ALUNO **WALMIR DOS REIS MIRANDA FILHO**, MATRICULADO, SOB O Nº 2018.657.679, NO PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA, DO INSTITUTO DE CIÊNCIAS EXATAS, DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, REALIZADA NO DIA 24 DE MARÇO DE 2022.

Aos 24 dias do mês de Março de 2022, às 08h00, em reunião pública virtual 74 (conforme orientações para a atividade de defesa de tese durante a vigência da Portaria PRPG nº 1819) OU na sala do Instituto de Ciências Exatas da UFMG, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pelo Colegiado do Programa de Pós-Graduação em Estatística, para julgar a defesa de tese do aluno **WALMIR DOS REIS MIRANDA FILHO**, nº matrícula 2018.657.679, intitulada: *"Semiparametric modeling for multivariate survival data via copulas"*, requisito final para obtenção do Grau de doutor em Estatística. Abrindo a sessão, o Senhor Presidente da Comissão, Prof. Fábio Nogueira Demarqui, passou a palavra ao aluno para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do aluno. Após a defesa, os membros da banca examinadora reuniram-se reservadamente sem a presença do aluno e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação:

(x) Aprovada.
( ) Reprovada com resubmissão do texto em ____ dias.
( ) Reprovada com resubmissão do texto e nova defesa em ____ dias.
( ) Reprovada.

Prof. Fabio Nogueira Demarqui
Orientador – (EST/UFMG)

Prof. Antonio Carlos Pedroso de Lima (IME/USP)

Prof. Dani Gamerman
(DME/UFRJ)

Prof(a). Marcos Oliveira Prates
(EST/UFMG)

Prof. Leonardo Soares Bastos
(FIOCRUZ).

O resultado final foi comunicado publicamente ao aluno pelo Senhor Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 24 de Março de 2022.

Observações:

1. No caso de aprovação da tese, a banca pode solicitar modificações a serem feitas na versão final do texto. Neste caso, o texto final deve ser aprovado pelo orientador da tese. O pedido de expedição do diploma do candidato fica condicionado à submissão e aprovação, pelo orientador, da versão final do texto.
2. No caso de reprovação da tese com resubmissão do texto, o candidato deve submeter o novo texto dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca que então decidirão pela aprovação ou reprovação da tese.
3. No caso de reprovação da tese com resubmissão do texto e nova defesa, o candidato deve submeter o novo texto com a antecedência à nova defesa que o orientador julgar adequada. A nova defesa, mediante todos os membros da banca, deve ser realizada dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca. Baseada no novo texto e na nova defesa, a banca decidirá pela aprovação ou reprovação da tese.

Aos meus pais, Maria e Walmir

Agradecimentos

Em primeiro lugar, agradeço a Deus por esta grande conquista e por me permitir conhecer pessoas maravilhosas em minha vida que, direta ou indiretamente, me ajudaram a chegar até aqui.

Aos meus pais, Maria Aparecida de Campos e Walmir dos Reis Miranda, com os quais pude contar sempre, principalmente durante os dois últimos anos, marcados por uma pandemia, quando precisei transferir minha rotina acadêmica para dentro de casa. Nunca deixaram de me apoiar e sempre foram compreensivos e afetuosos.

Ao meu orientador, Fábio Nogueira Demarqui, por aceitar a supervisão desta tese. Foi bastante paciente quando tive dificuldades para assimilar alguns conceitos ou resolver os problemas que apareciam. Além disso, tem participado ativamente do meu crescimento acadêmico para além da tese, com muitos ensinamentos e conselhos. É uma honra poder trabalharmos juntos. Muito obrigado!

A todos os colegas estudantes do Programa de Pós-Graduação em Estatística (PPGE) da UFMG pelo companheirismo e afeto, dentro e fora da sala de aula. Passamos por muitos momentos alegres, e também de superação, juntos. Formei grandes amizades para a vida toda.

Aos meus orientadores de graduação e mestrado, Marcel de Toledo Vieira e Cristiano Ferraz, por terem sido fundamentais durante a minha trajetória acadêmica e profissional como estatístico amostrista, bem como pela grande amizade pessoal.

Aos professores do PPGE/UFMG, pela formação adquirida no decorrer do Doutorado. Com eles aprendi bastante e é possível sentir o amor pelo conhecimento que transmitem. Sempre foram didáticos, tranquilos e abertos quando necessitava tirar dúvidas.

Aos amigos que fiz em Belo Horizonte fora da pós-graduação, dentro ou fora da UFMG, muitos deles colegas de quarto ou externos à UFMG, com os quais dividi experiências e momentos de aprendizado e lazer.

Aos amigos e familiares residentes em Juiz de Fora – MG, cidade onde resido, pelo apoio durante o Doutorado. Em particular, agradeço ao médico oftalmologista Gilmar Antônio Nogueira e família, pelo apoio e amizade desde a época em que eu cursava o colegial.

A todos os professores e educadores que, em algum momento, fizeram parte da minha trajetória desde quando eu era bem pequeno na creche até a conclusão do ensino médio. Tenho um carinho enorme e não seria vitorioso sem o incentivo de vocês!

Aos mestres e doutores com os quais tive a oportunidade de aprender muito do que sei sobre a área que escolhi, Estatística, e outras correlatas, ao longo da graduação (UFJF) e mestrado (UFPE).

Ao professor Fernando de Souza Bastos, por me compartilhar um modelo (*template*) *Latex* muito bem editado e apropriado para o desenvolvimento do corpo textual desta tese.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro prestado durante todo o período de Doutorado que deu origem a esta tese.

Abstract

Clustered survival data can arise if the event of interest (the *failure*) is recurrent and more than one observed time is registered for each subject (which forms a cluster) under study, and the number of observed times is fixed for all subjects. Since survival data associated with the same cluster is expected to be correlated, it should be modeled in order to account for that dependence. Copula models became an appropriate framework to model clustered survival data, linking marginal survival functions to form a joint survival distribution. Much of the literature on survival copula models is concentrated on results marginally using only the Weibull model as the baseline distribution and the Proportional Hazards model as the regression structure when working with clustered survival data or supposing an informative censoring model for univariate survival data. This work proposes new survival copula models under a random and independent right-censoring assumption, addressing a variety of marginal baseline distributions (Weibull, Bernstein Polynomials, and Piecewise Exponential models) and regression model classes (Proportional Hazards, Proportional Odds, and Yang-Prentice models). Concerning the copulas themselves, each one among those treated in this work belongs to the Archimedean copula class, a family of copulas widely used in survival analysis due to some important properties. Five Archimedean copula models were addressed in this work: Ali-Mikhail-Haq; Clayton; Frank; Gumbel-Hougaard, and Joe. To evaluate and compare the proposed survival copula models, results for an extensive simulation study and a real data application were obtained. For the simulated data, variations can occur on the copula function and on the marginal baseline distribution or regression model class used for generation.Also, simulation scenarios were divided by true Kendall's τ correlation values for the copula model chosen for generation. When fitting the simulated data, better results are obtained for fitted models with the correct copula, given a specification of baseline distribution and regression structure. Moreover, even generating marginally from the Weibull model, results for fitted semiparametric models follow closely those obtained when fitting the Weibull model, being better (in general) for marginally generated data from the Exponentiated Weibull distribution, among the models fitted with the correct copula. For all survival copula models presented in this work, an R package is currently in development, containing specific functions for fitting and analysis.

Keywords: Archimedean copulas, Marginal survival functions, Baseline distributions, Regression model classes.

Resumo

Dados de sobrevivência *clusterizados* podem surgir se o evento de interesse (a *falha*) é recorrente e mais de um tempo observado é registrado para cada indivíduo (o qual forma um *cluster*) sob estudo, e a quantidade de tempos observados é fixa para todos os indivíduos. Como se espera que dados de sobrevivência associados a um mesmo cluster estejam correlacionados, a modelagem dos mesmos deve considerar esta dependência. Modelos de cópula se tornaram uma estrutura útil para a modelagem de dados de sobrevivência clusterizados, conectando funções de sobrevivência marginais para construir uma distribuição conjunta de sobrevivência. Muito da literatura sobre modelos de cópula de sobrevivência está restrita a resultados para o uso do modelo Weibull como a distribuição marginal da linha de base e do modelo de Riscos Proporcionais como a estrutura marginal de regressão ao se trabalhar com dados de sobrevivência clusterizados, ou a resultados para modelos de censura informativa aplicados a dados de sobrevivência univariados. Este trabalho propõe, sob o pressuposto de censura à direita aleatória e independente, novos modelos de cópula de sobrevivência abordando uma variedade de distribuições para a linha de base marginal (modelos Weibull, Polinômios de Bernstein e Exponencial por Partes) e de classes de modelos de regressão (Riscos Proporcionais, de Chances Proporcionais e Yang-Prentice). Com respeito às cópulas, cada uma dentre as tratadas neste trabalho pertence à classe de cópulas arquimedianas, uma família de cópulas amplamente utilizada em análise de sobrevivência devido a propriedades importantes. Cinco cópulas arquimedianas foram abordadas neste trabalho: Ali-Mikhail-Haq; Clayton; Frank; Gumbel-Hougaard e Joe. Para avaliar e comparar os modelos de cópula de sobrevivência propostos, foram obtidos resultados para um estudo extensivo de simulação, bem como para uma aplicação de dados reais. Para os dados simulados, variações podem ocorrer na cópula e na classe de modelos de regressão marginal. Além disso, os cenários para simulação foram divididos por valores verdadeiros supostos para a correlação τ de Kendall, dado o modelo de cópula escolhido para a geração. Ao ajustar os dados simulados, resultados melhores são obtidos para modelos ajustados com a cópula correta, dada uma especificação da distribuição para a linha de base e da estrutura de regressão. Além disso, mesmo gerando do modelo Weibull, resultados para ajustes de modelos semiparamétricos seguem de perto os obtidos ao ajustar o modelo Weibull, dentre os modelos ajustados com a cópula correta, sendo melhores (em geral) para dados marginalmente gerados da distribuição Weibull Exponenciada. Para todos os modelos de cópula de sobrevivência apresentados neste trabalho, um pacote R está atualmente em desenvolvimento, contendo funções específicas para ajuste e análise.

Palavras-chave: Cópulas arquimedianas, Funções de sobrevivência marginais, Distribuições da linha de base, Classes de modelos de regressão.

List of Figures

List of Tables

List of Abbreviations

AE	Average Estimate
AIC	Akaike Information Criterion
ALB	Average Lower Bound
ARB	Average Relative Bias
ASE	Average Standard Error
AUB	Average Upper Bound
AMH	Ali-Mikhail-Haq copula model
BFGS	Broyden-Fletcher-Goldfarb-Shanno algorithm
BP	Bernstein Polynomials approximation/model
CR	Coverage Rate
DF	Degrees of Freedom
EW	Exponentiated Weibull model
GH	Gumbel-Hougaard copula model
HMC	Hamiltonian Monte Carlo algorithm
HR	Hazard Ratio
LR	Likelihood Ratio test
LS	Laplace-Stieltjes transform
MC	Monte Carlo
OR	Odds Ratio
PE	Piecewise Exponential model
PH	Proportional Hazards regression model class
PO	Proportional Odds regression model class
SDE	Standard Deviation Estimate
SE	Standard Error
WAT	Weierstrass Approximation Theorem
YP	Yang-Prentice model

Contents

CHAPTER 1

Introduction

Survival analysis is a wide field in Statistics that deals with time data ranging from a well-defined start point until the occurrence of some particular event of interest, taken as the endpoint criteria (Collett, 2015). For instance, in medical research, the start point can correspond to the recruitment of a subject diagnosed with a particular condition, such as a disease, into a clinical trial to compare two or more treatments. If the endpoint is the death of a subject, the collected time data is its survival time. However, similar data can be obtained when the endpoint is not fatal, such as the relief of pain or the recurrence of a disease symptom. More generally, observed times are referred to as time to event data. Survival times are not liable to be analyzed by standard statistical procedures, since in general they are not normally or even symmetrically distributed. This difficulty could be remedied by applying a transformation over the data in order to achieve a normally-like distribution, like the famous Box and Cox (1964) transformation. However, using a transformation frequently implies the loss of interpretation in the inference made on parameters of a model for the original data, but given the transformed data.

Another important feature of survival analysis that requires modeling methods different from the standard ones is censoring (Collett, 2015). Among the different censoring types in the literature, special attention will be given for the right-censoring. An observed time is said to be right-censored if the event of interest has not been observed for a subject, but it is known that the subject has survived up to the observed time. A right-censored time can arise if a subject has been lost to follow-up before the event occurrence (a patient moves to another country and can no longer be tracked), the event has not been observed at the end of the study for all subjects, or death has been caused for a reason completely unrelated to the event of interest. A crucial assumption that will be made in the present work is that the actual survival time, whether it is observed or not, does not depend on any censoring mechanism. Then, censoring is said to be independent, operating randomly over all subjects under study. In other words, when considering a group of subjects who

have the same values on a set of relevant explanatory variables, a subject whose survival time is censored must be representative of all other subjects in that group who have survived up to the censoring time.

Depending on the study, survival data can be clustered by a grouping variable or even by subjects. This last can occur when the event of interest is observed more than one time for the same subject, and for all subjects under study. Thus, each cluster is composed by two or more observed times. Since data from the same cluster are expected to be correlated, it cannot be modeled as if all observed times were independent. A theoretical and well-suited framework to model clustered data is the copula, a multidimensional distribution function over a fixed number of standard uniform univariate margins, equal to the size of each cluster. Then, the copula can link marginal survival functions to form a joint survival distribution to model clustered data (Marra and Radice, 2020). In general, not only survival or distribution functions but any function restricted to the unit interval can be used as a marginal component of a multidimensional copula model.

1.1 A Brief Literature Review

Survival analysis is one of the oldest fields of study in Statistics, with roots in demography and actuarial science, dating back to the 17$^{\text{th}}$ century. The first results involved a large set of inferences based on mortality records, which were published by John Graunt, a London merchant, in 1662 (Hacking, 2006, *apud* Rickert, 2017). Even before the 1700s, basic-life table methods have already comprised techniques for dealing with the delayed entry (left truncation) and right-censored data when estimating a survival function (Andersen and Keiding, 1998). In the early 18$^{\text{th}}$ century, Abraham de Moivre and Daniel Bernoulli developed the modern foundations of the survival analysis when working on annuities and competing risks for smallpox inoculation, respectively. De Moivre proposed a linear approximation for the survival function, the first example of a parametric model. In the 19$^{\text{th}}$ century, proposals of parametric models were also made for the hazard function. Bernoulli, for his part, estimated the expected number of deaths given a standard set of death rates. In the early 20$^{\text{th}}$ century, two themes of research in survival analysis arose: multistate modeling and nonparametric estimation of survival functions in continuous time.

Although those themes were part of a more modern survival analysis, they still were studied in the fields of actuarial mathematics and demography without integrating theoretical statistics (Andersen and Keiding, 1998). Until the first half of the 20$^{\text{th}}$ century, the main contributions to the statistical theory of survival analysis were still limited to simple parametric models. Inference results started to appear from two key articles. The first is the paper by Kaplan and Meier (1958), who proposed a nonparametric estimator of the survival function. The second one is the work of Cox (1972), who introduced a regression model class for the hazard function, depending arbitrarily (in a nonparametric way) on the observed times and parametrically on covariates: the semiparametric Proportional Hazards (PH) model class. However, until the late 1970s, much of the work in survival analysis still involved only constructing better life tables and enhancements of non-censored nonparametric estimators (Klein et al., 2013). At that time, the counting process martingales theory was adapted by Aalen (1978) to survival analysis problems. This paved a way for the development of easier techniques for censored and truncated data, allowing both frequentist and Bayesian approaches of inference

on a wide range of problems. Since then, survival analysis modeling has been widely used in a variety of disciplines, including Medicine; Epidemiology; Engineering; Insurance; Economy; Marketing, and many other applications.

Copula modeling is, on the other hand, a more recent topic. The first investigation of a multivariate standardized distribution function was made by Hoeffding (1940), who worked with bivariate standardized distributions whose support was given by the square $[-1/2, 1/2]^2$, with uniform margins on the interval $[-1/2, 1/2]$. He also studied measures of dependence that are invariant under strictly increasing transformations. As Schweizer (1991), *apud* Nelsen (2006), pointed out, Hoeffding could have discovered copulas if he had chosen the unit square $[0, 1]^2$ instead. Féron (1956) considered standardized distribution functions defined on the unit cube $[0, 1]^3$, but the term "copula" (a Latin word for link) was introduced by Sklar (1959). He was also the first to establish a result on the connection between multivariate distribution functions and their associated one-dimensional margins, posteriorly known as the Sklar's Theorem (Nelsen, 2006). Until the early 1980s, results for the copula theory were mostly limited in the context of probabilistic metric spaces. At that time, Schweizer and Wolff (1981) studied different criteria for measures of dependence. They concluded that copulas provide a simple tool for analyzing the dependence among random variables. They also showed that copulas are invariant under strictly increasing transformations. Thereby, it was proved that copulas resume all the information about the dependence structure among random variables. Some recent references on copula theory and applications are the books of Nelsen (2006); Joe (2014); Durante and Sempi (2015); Flores et al. (2017) and Hofert et al. (2018).

Once copulas allow building multivariate distribution functions with margins defined on the unit interval $[0, 1]$, those same margins could be represented each one by a survival function, which would form a joint survival distribution. The first use of copulas as multivariate survival models dates back to the work of Clayton (1978) and Oakes (1982). Clayton noted that, when adjusting for covariates, marginal survival functions and the copula dependence parameter can unveil underlying factors (not specified as covariates) influencing the probability of event times simultaneously. Oakes, in turn, reparameterized the model introduced by Clayton and proposed a correction for his likelihood expression and variance estimation. The Clayton, or Clayton-Oakes, copula is an example of a wide class: the Archimedean copula models, which include Frank, Gumbel-Hougaard (GH), Joe, and Ali-Mikhail-Haq (AMH) copulas (see Nelsen (2006) and references therein). Special attention will be given for those aforementioned copulas.

1.2 Main Functions in Survival Analysis

The following definitions will be mostly based on Collett (2015). Let $T \geqslant 0$ be a non-negative random variable for the observable (survival) time of a subject and t the value of its actual observed time. Suppose that T has a probability distribution with a continuous density function $f(t|\boldsymbol{\kappa})$, where $\boldsymbol{\kappa}$ is the vector of parameters from the distribution of T. Then, the cumulative distribution function of T is given by

$$F(t|\boldsymbol{\kappa}) = P(T < t|\boldsymbol{\kappa}) = \int_0^t f(u|\boldsymbol{\kappa})du, \tag{1.1}$$

which represents the probability that a survival time is less than a value t. The survival (or survivor) function is the probability that a survival time is greater than or equal to a value t, so

$$S(t|\boldsymbol{\kappa}) = P(T \geqslant t|\boldsymbol{\kappa}) = 1 - F(t|\boldsymbol{\kappa}) \tag{1.2}$$

is the complement of the cumulative distribution function in (1.1), and represents the probability that a subject will survive beyond a given time t.

One fundamental concept in survival analysis is the hazard function, used to express the hazard of an event of interest at a time t. The hazard function is obtained from the probability that the event occurs at t, conditional on the subject having survived up to t. For its formal definition, let $P(t \leqslant T < t + \varepsilon t|T \geqslant t; \boldsymbol{\kappa})$ be the probability that T lies between t and $t + \varepsilon t$, where $\varepsilon > 0$. This conditional probability can be divided by the time interval, εt, giving a rate that expresses a probability per unit of time. If εt tends to zero, then

$$h(t|\boldsymbol{\kappa}) = \lim_{\varepsilon t \to 0} \left\{ \frac{P(t \leqslant T < t + \varepsilon t|T \geqslant t; \boldsymbol{\kappa})}{\varepsilon t} \right\} \tag{1.3}$$

is the hazard function, also referred to as the hazard rate function.

The hazard function defined in (1.3) allows constructing relationships among the cumulative distribution, density, survival, and hazard functions. Using the definition of conditional probability, it is possible to rewrite

$$P(t \leqslant T < t + \varepsilon t|T \geqslant t; \boldsymbol{\kappa}) = \frac{P(t \leqslant T < t + \varepsilon t|\boldsymbol{\kappa})}{P(T \geqslant t|\boldsymbol{\kappa})} = \frac{F(t + \varepsilon t|\boldsymbol{\kappa}) - F(t|\boldsymbol{\kappa})}{S(t|\boldsymbol{\kappa})}$$

and, for the hazard function

$$h(t|\boldsymbol{\kappa}) = \lim_{\varepsilon t \to 0} \left\{ \frac{F(t + \varepsilon t|\boldsymbol{\kappa}) - F(t|\boldsymbol{\kappa})}{\varepsilon t} \right\} \frac{1}{S(t|\boldsymbol{\kappa})} = \frac{f(t|\boldsymbol{\kappa})}{S(t|\boldsymbol{\kappa})}. \tag{1.4}$$

Thus, given any of the equations (1.1), (1.2) or (1.4), the other two can be determined. Following from (1.4), it is possible to define the cumulative or integrated hazard function $H(t|\boldsymbol{\kappa})$ since

$$h(t|\boldsymbol{\kappa}) = \frac{f(t|\boldsymbol{\kappa})}{S(t|\boldsymbol{\kappa})} = -\frac{d}{dt}\{\log[S(t|\boldsymbol{\kappa})]\},$$

and by definition

$$H(t|\boldsymbol{\kappa}) = \int_0^t h(u|\boldsymbol{\kappa})du = -\log[S(t|\boldsymbol{\kappa})]. \tag{1.5}$$

Therefore, while the hazard function at the time t can be interpreted as the expected number of events experienced by a subject in a unit of time (constant over that period), given that the event has not yet occurred until t, the cumulative hazard function can be regarded as the expected number of events that occur in the time interval from the origin up to time t. The hazard function has a relevant role in survival modeling. In general, it is analytically easier to deal with the hazard function associated with a parametric survival model rather than this last itself. As will be seen later, if the hazard ratio is about the same for two subgroups of subjects in the study, a Proportional Hazards (PH) model can be used to model the survival times.

Another function in survival analysis that has gained more attention recently is the odds function (and its associated derivative). Although the odds definition is traditional in epidemiological case-control studies, its use in survival models appeared only with the work of Bennett (1983), who proposed the Proportional Odds (PO) model class. The odds function express, for a fixed time t, how much an event of interest is more likely to occur than not to occur. Thus, denoting the odds function by $R(t|\boldsymbol{\kappa})$, its mathematical expression is defined by the ratio $F(t|\boldsymbol{\kappa})/S(t|\boldsymbol{\kappa})$, which can also be rewritten using (1.5) as

$$R(t|\boldsymbol{\kappa}) = \frac{F(t|\boldsymbol{\kappa})}{S(t|\boldsymbol{\kappa})} = \frac{1-\exp\left[-H(t|\boldsymbol{\kappa})\right]}{\exp\left[-H(t|\boldsymbol{\kappa})\right]} = \exp\left[H(t|\boldsymbol{\kappa})\right] - 1. \tag{1.6}$$

Equation 1.6 establishes a relationship between the odds and the cumulative hazard functions. Similarly, another can be set for the derivative of the odds function, denoted by $r(t|\boldsymbol{\kappa})$

$$r(t|\boldsymbol{\kappa}) = \frac{d}{dt}\left[R(t|\boldsymbol{\kappa})\right] = \frac{d}{dt}\left\{\exp\left[H(t|\boldsymbol{\kappa})\right] - 1\right\} = h(t|\boldsymbol{\kappa})\exp\left[H(t|\boldsymbol{\kappa})\right] = \frac{f(t|\boldsymbol{\kappa})}{\left[S(t|\boldsymbol{\kappa})\right]^2}. \tag{1.7}$$

1.2.1 The Right-censoring Role

As stated early, not always an observed time will be a survival time: the subject has survived up to a given time and is no longer followed up for a reason unrelated to the event occurrence. This is an example of a right-censored observed time, the most common type of censoring, which shall be considered in the present work. Although there are different right-censoring mechanisms, they lead to the same survival likelihood functions (Duchateau and Janssen, 2007). Assuming also that the survival and censoring times are independent random variables for all subjects (random censoring), and the censoring times do not depend on any parameter related to the survival function (non-informative censoring), this guarantees the identifiability for the distribution of the observed times (Fleming and Harrington, 1991).

Based on these assumptions, a generic expression for the survival likelihood function can be constructed. Suppose that for each subject i, $1 \leqslant i \leqslant n$, a survival time $T_i = t_i$ or a censored time $A_i = a_i$ are registered. Assume also that survival (censoring) times are independent among all subjects, i. e., $T_1, \ldots, T_n \sim F_T(t|\boldsymbol{\kappa}_T)$ ($A_1, \ldots, A_n \sim F_A(a|\boldsymbol{\kappa}_A)$). The actual observable time is defined by $Y_i = \min(T_i, A_i)$, whose distribution is indexed by a vector $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_T, \boldsymbol{\kappa}_A)$ of parameters. Then, the information of a subject i is given by the pair (Y_i, δ_i), with $\delta_i = \mathbb{I}_{T_i<A_i}$ being the censoring indicator random variable. For a pair $(Y_i = t_i, \delta_i = 1)$ (a survival observed time), the likelihood contribution is given by (Duchateau and Janssen, 2007, p. 19)

$$\begin{aligned}
\lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}P(y_i-\varepsilon<Y_i<y_i+\varepsilon,\delta_i=1|\boldsymbol{\kappa}) &= \lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}P(y_i-\varepsilon<T_i<y_i+\varepsilon,T_i\leqslant A_i|\boldsymbol{\kappa}) \\
&= \lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}\int_{y_i-\varepsilon}^{y_i+\varepsilon}\int_t^{\infty} dF_A(a|\boldsymbol{\kappa}_A)dF_T(t|\boldsymbol{\kappa}_T) \quad \text{(independence)} \\
&= \lim_{\varepsilon\to 0^+}\frac{1}{2\varepsilon}\int_{y_i-\varepsilon}^{y_i+\varepsilon}[1-F_A(a|\boldsymbol{\kappa}_A)]dF_T(t|\boldsymbol{\kappa}_T) \\
&= [1-F_A(y_i|\boldsymbol{\kappa}_A)]f_T(y_i|\boldsymbol{\kappa}_T).
\end{aligned}$$

On the other hand, the likelihood contribution for a pair $(Y_i = a_i, \delta_i = 0)$ (right-censored observed time) is given by

$$\begin{aligned}\lim_{\varepsilon\to 0^+} \frac{1}{2\varepsilon} P(y_i - \varepsilon < Y_i < y_i + \varepsilon, \delta_i = 0|\boldsymbol{\kappa}) &= \lim_{\varepsilon\to 0^+} \frac{1}{2\varepsilon} P(y_i - \varepsilon < C_i < y_i + \varepsilon, T_i > C_i|\boldsymbol{\kappa}) \\ &= [1 - F_T(y_i|\boldsymbol{\kappa}_T)] f_A(y_i|\boldsymbol{\kappa}_A).\end{aligned}$$

Thus, under a random right censoring, the survival likelihood function for a sample $\boldsymbol{y}$ of size n has the following expression

$$L(\boldsymbol{\kappa}|\boldsymbol{y}) = \prod_{i=1}^{n} \{[1 - F_A(y_i|\boldsymbol{\kappa}_A)] f_T(y_i|\boldsymbol{\kappa}_T)\}^{\delta_i} \{[1 - F_T(y_i|\boldsymbol{\kappa}_T)] f_A(y_i|\boldsymbol{\kappa}_A)\}^{1-\delta_i}. \tag{1.8}$$

Assuming also that censoring is non-informative, i. e., the distribution of the censoring times does not depend on the parameters $\boldsymbol{\kappa}_T$ from the survival function, the factors $[1 - F_A(y_i\boldsymbol{\kappa}_A)]^{\delta_i}$ and $f_A(y_i\boldsymbol{\kappa}_A)^{1-\delta_i}$ do not give any information for inference and can be dropped from (1.8). Thereby, $\boldsymbol{\kappa} = \boldsymbol{\kappa}_T$ and a simpler survival likelihood function is given by

$$L(\boldsymbol{\kappa}|\boldsymbol{y}) \propto \prod_{i=1}^{n} [f(y_i|\boldsymbol{\kappa})]^{\delta_i} [1 - F(y_i|\boldsymbol{\kappa})]^{1-\delta_i} = \prod_{i=1}^{n} [f(y_i|\boldsymbol{\kappa})]^{\delta_i} [S(y_i|\boldsymbol{\kappa})]^{1-\delta_i} = \prod_{i=1}^{n} [h(y_i|\boldsymbol{\kappa})]^{\delta_i} S(y_i|\boldsymbol{\kappa}). \tag{1.9}$$

Expression (1.9) is useful for the modeling of a hazard function, like in a PH model. An alternative version can be obtained only in terms of the odds function and its derivative as

$$\begin{aligned}L(\boldsymbol{\kappa}|\boldsymbol{y}) &\propto \prod_{i=1}^{n} [h(y_i|\boldsymbol{\kappa})]^{\delta_i} S(y_i|\boldsymbol{\kappa}) = \prod_{i=1}^{n} \{r(y_i|\boldsymbol{\kappa}) \exp[-H(y_i|\boldsymbol{\kappa})]\}^{\delta_i} \frac{F(y_i|\boldsymbol{\kappa})}{R(y_i|\boldsymbol{\kappa})} \\ &= \prod_{i=1}^{n} [r(y_i|\boldsymbol{\kappa}) S(y_i|\boldsymbol{\kappa})]^{\delta_i} \frac{R(y_i|\boldsymbol{\kappa})}{1 + R(y_i|\boldsymbol{\kappa})} \frac{1}{R(y_i|\boldsymbol{\kappa})} \\ &= \prod_{i=1}^{n} \left[\frac{r(y_i|\boldsymbol{\kappa})}{1 + R(y_i|\boldsymbol{\kappa})}\right]^{\delta_i} \frac{1}{1 + R(y_i|\boldsymbol{\kappa})},\end{aligned} \tag{1.10}$$

which is useful if the odd function is modeled instead, as done in a PO model.

Presented all the basic concepts in survival analysis, they can be extended to a great variety of problems. The interested reader can consult the following references for more details. Nonparametric estimators for the survival function and related quantities are well described in Section 1.4 of Kalbfleisch and Prentice (2002) and Chapter 4 of Klein and Moeschberger (2005). Parametric models can be viewed in extension on Chapter 2 of Klein and Moeschberger (2005), with diagnostics and inference addressed in Chapters 11 and 12. Censoring types are widely discussed by Lawless (2011) and Schneider (2017). Finally, a Bayesian approach is covered by Ibrahim et al. (2001). All those references also address more sophisticated survival models, such as additive and competing risk models.

1.3 Research Overview and Contributions

Modeling dependence on clustered survival data with copulas has some advantages in comparison with the traditional frailty models when applied for the same context (called in this case shared frailty models). When estimating a dependence parameter under a frequentist approach, in a copula model it is not necessary to integrate out any random effect, a difficulty that can arise for some shared frailty models. Moreover, a shared frailty model can require some restrictions on the random effect distribution moments to ensure identifiability. On the other hand, a copula model does not allow to rank each cluster by its heterogeneity, since a unique parameter (or set of parameters) will contain all information on the dependence of observed times for the same cluster, which is possible on shared frailty models. However, if there is no interest in any particular cluster, this is not a drawback of concern. Since shared frailty models will not be covered in this work, the interested reader can consult Duchateau and Janssen (2007) for a detailed historical review on frailty models for clustered data and their main properties.

The study of survival models as marginal components of a copula model still has an open avenue for development, even for Archimedean copulas and under an independent and non-informative right-censoring scheme for an event of interest. Examples of what have been done recently are the works of Goethals et al. (2012), who fitted copula models with Weibull Proportional Hazards (PH) components for bivariate clustered data and compared it with shared frailty models; Louzada et al. (2013) for bivariate copulas with survival mixture models for each of their components; Prenen et al. (2017) for multivariate clusters with variable size, and Marra and Radice (2020) for additive copula models with monotonic splines. Several regression model classes of practical importance, however, have not been addressed yet. One example is the aforementioned Proportional Odds (PO) model class proposed by Bennett (1983). Therefore, this work will provide not only a theoretical framework, but also simulation studies and an application for a real data set, for inference on Archimedean copula models with a PO model class for both survival marginal components. It will include comparisons with similar procedures using the PH model class and the wider Yang-Prentice (YP) model class, which comprises the earlier two as particular cases (Yang and Prentice, 2005). Each regression model class goes along with a baseline function that can also be estimated. Three models will be considered in this work for the fitting of a baseline function, also being the same for both copula margins.

The first model addressed for the baseline function is the Weibull distribution, a traditional parametric survival model whose main attractive feature is its flexibility for the baseline hazard function (which can increase, decrease or remain constant), while maintaining a simple expression. Many survival analysis studies work with Weibull models, or its Exponential particular case, when a comparison is necessary for a proposed model that includes a different baseline function. The second model addressed is the Bernstein Polynomials (BP) model for a target function with a baseline component, be it the cumulative hazard function (under a PH regression model class) or the odds function (under a PO regression model class). The last one, which follows a similar idea, is the Piecewise Exponential (PE) model. Both BP and PE models are also classified as semiparametric in the literature due to their flexibility for the number of baseline parameters.

The Bernstein Polynomials model was first introduced in survival analysis by Chang et al. (2005) for the

cumulative hazard function of homogeneous populations (without any regression structure). They treated the polynomial degree as a random quantity, imposing a Beta prior process to it. However, the chosen framework for the BP modeling in this work is the one developed by Osman and Ghosh (2012), who used the BP to approximate both cumulative hazard and hazard functions, since the BP modeling can be used to obtain another approximation for a derivative function if it exists (Lorentz, 1986). They obtained a semiparametric model for baseline functions, although fixing the polynomial degree, and also established a proof for the likelihood log-concavity property of their BP modeling. It led to faster computational routines to find Bayesian estimators, besides being a necessary condition to the uniqueness of the maximum likelihood estimator.

Naturally, proposals combining the BP modeling with survival regression model classes have arisen in the literature, initially for univariate data (see Panaro (2020) for a brief review). Hu et al. (2017) were the first to use Bernstein Polynomials to model a baseline function, jointly with a PH model for the regression structure, as a marginal component of a copula model for bivariate interval-censored data. Xu et al. (2019) also worked with bivariate copula models, but their marginal components were given by the survival and censoring times of a same subject. They assumed informative interval censoring, using the BP modeling for the associated baseline hazard function. Although dealing with more complex forms of censoring while introducing the copula modeling, these two works stick to PH model classes, limiting interpretation for the associated coefficients.

Having been introduced shortly after the Cox's regression model by Kalbfleisch and Prentice (1973), the Piecewise Exponential model arose as a discrete model for the hazard function, being an alternative for grouped survival data or continuous survival data in the presence of ties. The main idea is to assume that distinct survival times have a constant hazard function between them. Then, the true hazard function is approximated by a sum of constant hazard functions, each one fitted by an Exponential model. Later, Friedman (1982) proved the asymptotic properties for the maximum likelihood estimators of PE model parameters. As an example of using as a marginal component of a copula model, Prenen et al. (2017) fitted PE models for the margins of multivariate Archimedean copula models with independent random censoring and variable cluster size, but only using the PH model for the regression structure. For dependent censoring, Emura and Michimae (2017) applied copula models with PE margins in the context of bivariate competing risks modeling, although without imposing any regression structure.

To evaluate the fitting of distinct Archimedean copula models with survival marginal components (also called survival copula models), a simulation study will be performed in this work. Moreover, a real data set will be fitted: the study of a data collection of patients with ovarian cancer described by Ganzfried et al. (2013). The computational framework consists of programmed functions in Stan (2020b), an open-source language designed to define custom likelihood functions and, for a Bayesian analysis, prior specifications. Stan has a broad supporting material available online, such as documentation (including for integrated modules with R and Python), articles, and books for users and developers. The present work uses the integrated version of Stan with R, the `rstan` (2020a). Under a frequentist approach of inference, likelihood maximization is done through the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Nocedal and Wright, 2006). Under a full Bayesian inference approach, Stan draws posterior samples through the Hamiltonian Monte Carlo (HMC) algorithm extended with the No-U-Turn Sampler, preventing a random walk path and greatly reducing the

sensitivity to correlated parameters, when compared to traditional Markov Chain Monte Carlo methods, such as random walk Metropolis and Gibbs Sampling. Thereby, their chains converge faster to the target distribution (Hoffman and Gelman, 2014). In summary, the main contributions of this work are:

- Provide a novel theoretical framework on a fully likelihood-based approach to handle bivariate survival data, allowing a wide range of behaviors for marginally fitted survival functions. This is done through the choice of nonparametric models for the baseline distribution, combined with the YP regression model class, which allows crossing survival times and contains both PH and PO classes as special cases.
- Establishment of inference results on regression parameter estimates and information criterion for distinct survival copula models under random and independent right-censoring, given a combination of baseline function (when involving the BP and PE models) and regression model class (for PH, PO and YP);
- Comparison of results, given a combination of baseline function and regression model class, for different fitted copula models given each one of the five Archimedean copulas used for generation;
- Comparison of inference results, given distinct baseline functions, concerning the regression and correlation parameter estimates, and information criterion, given a combination of copula model (only when the fitted copula and the one used for generation are the same) and regression model class;
- Evaluation of nested fitted models with respect to their regression model classes, for any given class used for generation, given a combination of (correctly) fitted copula model and baseline distribution;
- Given (correctly) fitted survival copula models with the YP regression model class, crossing time estimation for some specific scenarios of correlation and baseline distribution used for generation, comparing all fitted baseline distributions;
- Development of an R package (named `copSurv`), integrated with Stan, containing all survival copula models treated here and their associated methods for inference (see the Appendix for more details).

1.4 Thesis Structure

In the earlier sections, a brief introduction to survival analysis and copula theory was presented, followed by a succinct literature review. Definitions of basic concepts in survival analysis were made. Finally, an outline of the main goals of this work and their related procedures was described. The development of this work is structured as follows: Chapter 2 presents the Weibull; BP, and PE survival models for the baseline functions as well as the PH; PO and YP regression model classes. Chapter 3 addresses the basics of copula theory, giving special attention to the Archimedean copula class and its integration with survival models for the observed times of a given cluster as the copula marginal components. Dependence measures are also covered, highlighting some important concepts for survival analysis. Chapter 4 integrates the theory presented in the earlier two chapters and describes the proposed survival copula models. Numerical results for an intensive simulation study and an application for real clustered data are discussed in Chapters 5 and 6, respectively. Finally, Chapter 7 summarizes the main conclusions and appoints possible directions for future work.

CHAPTER 2

Parametric Survival Models and Regression Model Classes

This chapter is divided into two parts. The first part covers the definition of a parametric survival model to a baseline function, focusing on the Weibull; Bernstein Polynomial (BP), and Piecewise Exponential (PE) models. While the Weibull model is extensively used in the literature due to allowing some flexibility for the hazard function (monotone curves) with few parameters, the BP and PE models have a nonparametric appeal: the polynomial degree or the number of intervals for the hazard partition can be as high as needed if there is no interest on the baseline parameter estimates. Also, both BP and PE models do not impose any restriction on the form of hazard or odds functions. The second part of the chapter addresses ways to offer more flexibility in the presence of covariates for the subjects under study: the specification of a model class containing a particular regression structure. Here, special attention is given to the Proportional Hazards (PH); Proportional Odds (PO), and Yang-Prentice (YP) regression model classes, since they allow practicable interpretation in terms of hazards and odds (respectively) from the regression coefficient estimates. When combined, each particular choice of baseline and regression model class will enter as a marginal component of the proposed Archimedean survival copula models in this work.

2.1 Parametric Survival Models

Given a set of (positive) observed times, any family of probabilistic models with support on the positive real line is a candidate for a parametric survival model. It is said family in the sense that a set of parameters are not necessarily fixed and each probabilistic model is defined when these parameters (denoted here by the vector $\boldsymbol{\kappa}$) assume a particular value. However, the chosen model cannot be arbitrary: it should represent well many plausible possibilities for the empirical survival curve behavior.

2.1.1 The Weibull Model

Although he was not the first to use the model that bears his name, Weibull (1951) proposed a family of probability distributions for describing the life length of materials. Its use in engineering and other areas of applied research quite increased over the years. To understand the popularity of the Weibull distribution, let's start from its cumulative distribution and probability density functions. If the random variable for the survival times T follows a Weibull distribution with parameters $\boldsymbol{\kappa} = (\alpha, \lambda)$, then

$$F(t|\alpha,\lambda) = 1 - \exp(-\lambda t^{\alpha}), \quad \lambda > 0,\ \alpha > 0,$$
$$f(t|\alpha,\lambda) = \lambda\alpha t^{\alpha-1}\exp(-\lambda t^{\alpha}),$$

where λ and α are the scale and shape parameters, respectively. Note that the classical exponential distribution is obtained if $\alpha = 1$. Regarding the hazard function, its expression is a bit simpler: since $S(t|\alpha,\lambda) = 1 - F(t|\alpha,\lambda) = \exp(-\lambda t^{\alpha})$, then the hazard function and its cumulative are given by

$$h(t|\alpha,\lambda) = \frac{f(t|\alpha,\lambda)}{S(t|\alpha,\lambda)} = \frac{\lambda\alpha t^{\alpha-1}\exp(-\lambda t^{\alpha})}{\exp(-\lambda t^{\alpha})} = \lambda\alpha t^{\alpha-1}, \tag{2.1}$$
$$H(t|\alpha,\lambda) = -\log\left[S(t|\alpha,\lambda)\right] = \lambda t^{\alpha}. \tag{2.2}$$

Expression (2.1) is attractive on a practical point of view because it accommodates increasing ($\alpha > 1$), decreasing ($\alpha < 1$) and constant ($\alpha = 1$, backing to the exponential case) behaviors for the hazard function of t. Even when increasing, (2.1) can be concave ($\alpha \in (1,2)$), linear ($\alpha = 2$) or convex ($\alpha > 2$). However, note that it does not allow non-monotonicity, such as unimodal and "bathtub" forms. Nevertheless, the Weibull hazard function is a good start point for the development and comparison of any proposed survival model, being the most used parametric survival model for the hazard function in the literature. Regarding the odds function and its derivative, they are expressed as

$$R(t|\alpha,\lambda) = \exp\left[H(t|\alpha,\lambda)\right] - 1 = \exp(\lambda t^{\alpha}) - 1, \tag{2.3}$$
$$r(t|\alpha,\lambda) = h(t|\alpha,\lambda)\exp\left[H(t|\alpha,\lambda)\right] = \lambda\alpha t^{\alpha-1}\exp(\lambda t^{\alpha}). \tag{2.4}$$

2.1.2 The Bernstein Polynomial (BP) Model

The Bernstein Polynomials (BP) were originally proposed by Bernstein (1913) as a proof for the Weierstrass Approximation Theorem (WAT) in the unit interval (Lorentz, 1986). Following p. 148 of Bartle and Sherbert (2011), the WAT states

Theorem 2.1 (Weierstrass Approximation Theorem)**.** *Let $I = [a,b]$ and let $v : I \mapsto \mathbb{R}$ continuous over I. If $\varepsilon > 0$ is given, then there exists a polynomial function p_{ε} such that $|v(x) - p_{\varepsilon}(x)| < \varepsilon$ for all $x \in I$.*

To understand the BP approximation, first consider an event A such as $\mathbb{P}(A) = x$, where $\mathbb{P}$ is a probability measure. Then, suppose that an experiment with m trials will be performed in such a way that, if the event

A occurs k times, $0 \leqslant k \leqslant m$, a monetary amount equal to $v(k/m)$ will be paid to a hypothetical gambler. Thereby, a random variable K defined as the number of successes (the event A has happened) in m trials has a binomial distribution: $K \sim Bin(x, m)$, where $x \in [0, 1]$. Therefore, the probability of k occurrences for the event A and the expected value for a random variable $Q = v(K/m)$ representing the amount received by the gambler are given respectively by

$$\mathbb{P}(K = k) = \binom{m}{k} x^k (1-x)^{m-k}, \tag{2.5}$$

$$\mathbb{E}_m(Q) = B_{(m)}(x) = \sum_{k=0}^{m} v\left(\frac{k}{m}\right) \binom{m}{k} x^k (1-x)^{m-k}, \quad x \in [0, 1]. \tag{2.6}$$

From the relations in (2.5) and (2.6), combined with the Theorem 2.1, Bernstein proved that, given $\varepsilon > 0$, $|v(x) - \mathbb{E}_m(Q)| < \varepsilon$. In other words

$$v(x) = \lim_{m\to\infty} \mathbb{E}_m(Q) = \lim_{m\to\infty} \sum_{k=0}^{m} v\left(\frac{k}{m}\right) \binom{m}{k} x^k (1-x)^{m-k} = \lim_{m\to\infty} B_{(m)}(x).$$

Thus, the Bernstein Polynomial of degree m that approximates $v(x)$ is given by $B_m(x)$, where

$$b_{(k,m)}(x) = \binom{m}{k} x^k (1-x)^{m-k}$$

is the Bernstein basis. Note that each basis can be seen as a weight since, given the degree m, $b_{(k,m)}(x) \in (0, 1)$ for all k and

$$\sum_{k=0}^{m} b_{(k,m)}(x) = \sum_{k=0}^{m} \binom{m}{k} x^k (1-x)^{m-k} = 1.$$

To accommodate functions restricted to any compact interval $[a, b]$, $a < b \in \mathbb{R}$, the result in (2.6) can be extended as (Farouki and Rajan, 1987, p. 191)

$$B_{(m)}(x) = \sum_{k=0}^{m} v\left[a + \frac{k}{m}(b-a)\right] b_{(k,m)}\left(\frac{x-a}{b-a}\right), \quad x \in [a, b]. \tag{2.7}$$

According to Carnicer and Peña (1993), the BP approximation has optimal shape-preserving property when compared to other polynomial approximations. The Section 5 of Farouki (2012) review paper lists many properties and algorithms associated with the Bernstein bases (a total of 18 topics), but four are of major concern for the construction of a BP survival model:

1. **Symmetry**: $b_{(k,m-k)}(x) = b_{(k,m)}(1-x)$;
2. **Recursion**: $b_{(k,m+1)}(x) = x b_{(k-1,m)}(x) + (1-x) b_{(k-1,m)}(x)$;
3. **Non-negativity**: $b_{(k,m)}(x) \geqslant 0$, $\forall\ x \in [0, 1]$, if $0 \leqslant k \leqslant m$;
4. **Basis Derivative**: $\dfrac{d}{dx} b_{(k,m)}(x) = m\left[b_{(k-1,m-1)}(x) - b_{(k-1,m)}(x)\right]$.

Following Panaro (2020), the properties above allow the construction of an approximation for the derivative of $B_{(k,m)}$ in (2.6) with respect to x, which provides

$$\begin{aligned}
\frac{d}{dx}B_{(m)}(x;v) &= \sum_{k=0}^{m} v\left(\frac{k}{m}\right)\binom{m}{k}\left\{kx^{k-1}(1-x)^{m-k} - (m-k)x^{k}(1-x)^{m-k-1}\right\} \\
&= m\sum_{k=0}^{m} v\left(\frac{k}{m}\right)\left[\binom{m-1}{k-1}x^{k-1}(1-x)^{m-k} - \binom{m-1}{k}x^{k}(1-x)^{m-k-1}\right] \\
&= m\sum_{k=0}^{m} v\left(\frac{k}{m}\right) b_{(k-1,m-1)}(x) - m\sum_{k=0}^{m} v\left(\frac{k}{m}\right) b_{(k,m-1)}(x) \\
&= m\sum_{i=-1}^{m-1} v\left(\frac{i+1}{m}\right) b_{(i,m-1)}(x) - m\sum_{k=0}^{m} v\left(\frac{k}{m}\right) b_{(k,m-1)}(x), \qquad (2.8)
\end{aligned}$$

where $i = k-1$. By definition (Farouki, 2012), consider $b_{-1,m-1}(x) = b_{m,m-1}(x) = 0$. Then, (2.8) can be rewritten as

$$\frac{d}{dx}B_{(m)}(x;v) = m\sum_{i=0}^{m-1}\left\{v\left(\frac{i+1}{m}\right) - v\left(\frac{i}{m}\right)\right\} b_{(i,m-1)}(x) = m\sum_{i=0}^{m-1}\Delta v_i^{(1)} b_{(i,m-1)}(x), \qquad (2.9)$$

where $\Delta v_i^{(1)} = v[(i+1)/m] - v[i/m]$ is the first-order difference of $v(x)$ at $x = i/m$.

Chang et al. (2005) noted that the finite BP approximation could be used to estimate both hazard and cumulative hazard functions of a survival model, since this last is positive and bounded. Assuming $t \in [0,\tau]$, where $\tau = \inf\{t : S(t) = 0\} < \infty$, let $H(t)$ be the target function for the BP approximation. Thereby, rewriting (2.7) with $a = 0$ and $b = \tau$, the BP approximation for the cumulative hazard function is expressed as

$$B_{(m)}(t;H) = \sum_{k=0}^{m} H\left(\frac{k}{m}\tau\right) b_{(k,m)}\left(\frac{t}{\tau}\right), \quad t \in [0,\tau], \qquad (2.10)$$

and its first derivative with respect to the time t (approximating the hazard function), using (2.9), as

$$\begin{aligned}
\frac{d}{dt}B_{(m)}(t;H) &= \frac{m}{\tau}\sum_{i=0}^{m-1}\left\{H\left(\frac{i+1}{m}\tau\right) - H\left(\frac{i}{m}\tau\right)\right\} b_{(i,m-1)}\left(\frac{t}{\tau}\right) \\
&= \frac{m}{\tau}\sum_{k=1}^{m}\left\{H\left(\frac{k}{m}\tau\right) - H\left(\frac{k-1}{m}\tau\right)\right\}\binom{m-1}{k-1} b_{(k-1,m-1)}\left(\frac{t}{\tau}\right) \\
&= \frac{m}{\tau}\sum_{k=1}^{m}\left\{H\left(\frac{k}{m}\tau\right) - H\left(\frac{k-1}{m}\tau\right)\right\}\binom{m-1}{k-1}\left(\frac{t}{\tau}\right)^{k-1}\left(1-\frac{t}{\tau}\right)^{(m-1)-(k-1)} \\
&= \frac{1}{\tau}\sum_{k=1}^{m}\left\{H\left(\frac{k}{m}\tau\right) - H\left(\frac{k-1}{m}\tau\right)\right\}\frac{\Gamma(m+1)}{\Gamma(m-k+1)\Gamma(k)}\left(\frac{t}{\tau}\right)^{k-1}\left(1-\frac{t}{\tau}\right)^{m-k} \\
&= \sum_{k=1}^{m}\left\{H\left(\frac{k}{m}\tau\right) - H\left(\frac{k-1}{m}\tau\right)\right\}\left(\frac{1}{\tau}\right) f_B\left(\frac{t}{\tau};k,m-k+1\right), \qquad (2.11)
\end{aligned}$$

where B denotes the Beta distribution with parameters $\alpha = k$ and $\beta = m-k+1$. For simplicity, the cumulative hazards differences between braces and the Bernstein bases in (2.11) will be rewritten, respectively, as

$$\gamma_k = \left\{ H\left(\frac{k}{m}\tau\right) - H\left(\frac{k-1}{m}\tau\right) \right\}, \quad g_{(k,m)}(t) = \left(\frac{1}{\tau}\right) f_B\left(\frac{t}{\tau}; k, m-k+1\right).$$

Note that $\gamma_k > 0$, $k \in \{1, \ldots, m\}$, since $H(\cdot)$ is monotone increasing. As the coefficients γ_k do not depend on t, no information is given on the true cumulative hazard function and they should be estimated, compounding a vector $\boldsymbol{\kappa} = \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)'$ of BP parameters. Given a time t, its Bernstein bases can also be defined on a vector $\boldsymbol{g}_m(t) = (g_{(1,m)}(t), \ldots, g_{(m,m)}(t))'$ of fixed non-negative quantities. Then, the hazard and cumulative functions are modeled as (Osman and Ghosh, 2012, p. 561)

$$h(t|\boldsymbol{\gamma}) = \boldsymbol{\gamma}' \boldsymbol{g}_m(t), \tag{2.12}$$

$$H(t|\boldsymbol{\gamma}) = \int_0^t h(u, \boldsymbol{\gamma}) du = \boldsymbol{\gamma}' \boldsymbol{G}_m(t), \tag{2.13}$$

where $\boldsymbol{G}_m(t) = (G_{(1,m)}(t), \ldots, G_{(m,m)}(t))'$, with

$$G_{(k,m)}(t) = \int_0^t g_{(k,m)}(u) du = \int_0^t f_B\left(\frac{u}{\tau}; k, m-k+1\right) d\left(\frac{u}{\tau}\right) \geqslant 0, \quad k \in \{1, \ldots, m\}.$$

Since the true cumulative hazard function is unknown, a finite value should be taken for m on the estimation of BP parameters. Osman and Ghosh (2012) suggest a value m such that $n^{2/5} < m < n^{2/3}$ for the polynomial degree. As BP models are computationally intensive, this work will choose the smallest possible integer, $m = \lceil n^{2/5} \rceil$, for the simulation results and applied data in this work.

2.1.3 The Piecewise Exponential (PE) Model

Proposed by Kalbfleisch and Prentice (1973) as an alternative to the Cox's regression model in the presence of ties for the survival times or grouped survival data, the Piecewise Exponential (PE) model assumes that the hazard function is constant (*i.e.*, an Exponential model) between consecutive distinct survival times. In the end, the true hazard function is approximated by "steps" of constant hazard functions.

The formal definition of a PE model starts from a finite partition of the time axis, *i.e.*, a time grid $E = \{e_0, e_1, \ldots, e_p\}$, with $0 = e_0 < e_1 < \cdots < e_p < \infty$. That way, there are p intervals $E_k = (e_{k-1}, e_k]$, $k = 1, \ldots, p$. For each interval induced by E, a constant hazard function is assumed, that is

$$h(t) = \lambda_k, \quad t \in E_k, \quad k = 1, \ldots, p. \tag{2.14}$$

Therefore, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$ is the vector of constant hazard rates. If $p = 1$, the Exponential model is obtained as particular case. To obtain the cumulative hazard and odds functions, for $k = 1, \ldots, p$, define

$$t_k = \begin{cases} e_{k-1}, & t \leqslant e_{k-1}; \\ t & t \in E_k; \\ e_k & t > e_k. \end{cases} \tag{2.15}$$

Then, the cumulative hazard function is computed from (2.14) as

$$H(t|\boldsymbol{\lambda}, E) = \sum_{k=1}^{p} \lambda_k (t_k - e_{k-1}) = \boldsymbol{\lambda}' (\boldsymbol{t} - \boldsymbol{e}) \,. \tag{2.16}$$

The choice of how many intervals should be used for the time grid E will impact on the goodness-of-fit of a PE model with respect to the target model, *i.e.*, the true hazard function. For the purpose of this work, it is assumed that the time grid E is random and composed by a subset of size p drawn from all distinct ordered observed failure times Demarqui et al. (2014). To ensure that at least one failure time falls at each interval, the time grid associated to a fully nonparametric approach (using all distinct failure times) is initially imposed as the finest possible time grid for the PE model. Then, such intervals are clustered by the structure of a product partition model (see Barry and Hartingan (1992) for more details). By allowing that these intervals may contain more than one failure time, this enables controlling the maximum number of intervals and therefore the maximum number of parameters in the PE model. It is important to note that intervals near the end of a follow-up tend to contain more observed times, since censored times appear more frequently as close a study reaches its end (specially if there are right-censored subjects). An advantage of this approach is that the interval length is smaller for intervals with a large number of observed times, and larger for intervals where few time points are observed (Schneider et al., 2020).

Like the Weibull model, for univariate data the PE model has been used as a benchmark for comparison with other models for a baseline function, whether they also have a nonparametric appeal or not (Ibrahim et al., 2001). Although simple, the PE model can accommodate various possible shapes across the extension of E, since the number of intervals p is (in general) arbitrary and can be as large as needed. Older than the BP model, the choice of a fixed p for the PE model has been widely investigated in the literature (see Mello e Silva (2016) and references therein for a discussion). To retain comparability with the number m of polynomial degrees from the BP model, this work will fix $p = \lceil n^{2/5} \rceil$ for the simulation results and applied data.

2.2 Regression Model Classes

It was previously seen that survival data is in general highly skewed and can be censored for some or even most of the subjects. Thus, classical linear regression models are not suitable to fit them and also limit the interpretation of regression coefficients to the mean of times. However, survival data can be modeled in alternative ways to achieve other forms of interpretation. This is done by using functions of the observed times than themselves. In particular, two functions that are of great practical importance in survival analysis are the hazard and the odds functions. Both of them account for a baseline function (the hazard or odds for a reference level) associated with a regression structure, which is represented in this work by the multiplicative term $\exp(\boldsymbol{x}_i\boldsymbol{\beta})$. The vector $\boldsymbol{x}_i$ represents covariate values, given a design matrix X, for a subject i, $i \in \{1, \ldots, n\}$. The vector $\boldsymbol{\beta}$ represents the parameters associated to each covariate. Finally, the reference level is represented by a subject i whose covariate values are all equal to zero ($\boldsymbol{x}_i = 0$).

2.2.1 Proportional Hazards (PH) Model

The Proportional Hazards (PH) model is the most used regression model class in survival analysis. As the name says, it is assumed that the hazard functions for any pair of subjects are proportional between themselves: when taking the ratio of one hazard over the other, it is constant for any value of T. In a PH model, covariates are included to model the hazard function. Then, the hazard function for all subjects is always proportional to the baseline hazard function associated with a reference level.

One of the most traditional and known PH models is the Cox (1972) semiparametric model. Taking advantage of the proportionality assumption, the Cox model does not estimate the hazard function itself: the baseline term (the nonparametric part of the model) is dropped to estimate directly only the regression parameters. Formally, the hazard function for every subject i is modeled as

$$h(t|\boldsymbol{x}_i, \boldsymbol{\beta}) = h_0(t) \exp(\boldsymbol{x}_i \boldsymbol{\beta}), \tag{2.17}$$

where $h_0(t)$ denotes the baseline hazard function. Due to the presence of a baseline term, there is no intercept in (2.17). Since the PH regression model class assumes the hazard function ratio is constant over time for any pair of subjects i_1 and i_2, they can be compared through their Hazard Ratio (HR), which is given by

$$\text{HR}(i_1, i_2) = \frac{h_0(t) \exp(\boldsymbol{x}_{i_1} \boldsymbol{\beta})}{h_0(t) \exp(\boldsymbol{x}_{i_2} \boldsymbol{\beta})} = \exp\left[(\boldsymbol{x}_{i_1} - \boldsymbol{x}_{i_2})\boldsymbol{\beta}\right], \tag{2.18}$$

where $\boldsymbol{\beta}$ is estimated through a partial likelihood function calculated only over subjects whose times were not censored, (Collett, 2015). Then, the interpretation the Cox model is done in terms of the HR for two subjects.

On the other hand, when taking a parametric specification for the baseline hazard function, such as the Weibull; BP, or the PE models, a fully likelihood-based inference that also accounts for the right-censored times becomes possible. Thus, given a PH regression model with baseline function $h_0(t|\boldsymbol{\kappa})$ depending on a set of parameters $\boldsymbol{\kappa}$, the survival likelihood expression in (1.9) can be rewritten as (Panaro, 2020, p. 10)

$$\begin{aligned} L_{\text{PH}}(\boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) &\propto \prod_{i=1}^{n} \left[h(y_i)\right]^{\delta_i} S(y_i) = \prod_{i=1}^{n} \left[h(y_i, \boldsymbol{\beta}, \boldsymbol{\kappa}|\boldsymbol{x}_i)\right]^{\delta_i} S(y_i, \boldsymbol{\beta}, \boldsymbol{\kappa}|\boldsymbol{x}_i) \\ &= \prod_{i=1}^{n} \left[h_0(y_i, \boldsymbol{\kappa}) \exp(\boldsymbol{x}_i \boldsymbol{\beta})\right]^{\delta_i} S_0(y_i, \boldsymbol{\kappa}) \exp(\boldsymbol{x}_i \boldsymbol{\beta}), \end{aligned} \tag{2.19}$$

where $S_0(y_i, \boldsymbol{\kappa})$ is the baseline survival function. Using the Weibull ($\boldsymbol{\kappa} = (\alpha, \lambda)$); BP ($\boldsymbol{\kappa} = \boldsymbol{\gamma}$), and PE ($\boldsymbol{\kappa} = \boldsymbol{\lambda}$) model specifications for the baseline function, expression (2.19) is replaced by

$$L_{\text{PH}}(\alpha, \lambda, \boldsymbol{\beta}|\boldsymbol{y}, X) \propto \prod_{i=1}^{n} \left[\lambda \alpha y_i^{\alpha-1} \exp(\boldsymbol{x}_i \boldsymbol{\beta})\right]^{\delta_i} \lambda y_i^{\alpha} \exp(\boldsymbol{x}_i \boldsymbol{\beta}), \tag{Weibull}$$

$$L_{\text{PH}}(\boldsymbol{\gamma}, \boldsymbol{\beta}|\boldsymbol{y}, X) \propto \prod_{i=1}^{n} \left[\boldsymbol{\gamma}' \boldsymbol{g}_m(y_i) \exp(\boldsymbol{x}_i \boldsymbol{\beta})\right]^{\delta_i} \exp\left[-\boldsymbol{\gamma}' \boldsymbol{G}_m(y_i) \exp(\boldsymbol{x}_i \boldsymbol{\beta})\right], \tag{BP}$$

$$L_{\text{PH}}(\boldsymbol{\lambda}, \boldsymbol{\beta}|\boldsymbol{y}, X) \propto \prod_{i=1}^{n} \left\{ \left[\sum_{k=1}^{p} \lambda_k \mathbb{I}_{E_k}(y_i) \exp\left(\boldsymbol{x}_i \boldsymbol{\beta}\right)\right]^{\delta_i} \exp\left[-\boldsymbol{\lambda}' \left(\boldsymbol{t} - \boldsymbol{e}\right) \exp\left(\boldsymbol{x}_i \boldsymbol{\beta}\right)\right] \right\}. \tag{PE}$$

Although providing an easy interpretation, not always the assumption of proportional hazards is suitable. In the Section 4.4, Collett (2015) discuss some *ad hoc* diagnostic techniques and a hypothesis test to check this assumption for a set of univariate data. If those methods fail to ensure hazard proportionality, a possible alternative is to model the odds function.

2.2.2 Proportional Odds (PO) Model

Introduced by Bennett (1983), the Proportional Odds (PO) regression model class was designed for studies in which the mortality rates for different groups of subjectsconverge in time (without crossing). The regression structure is also given by the multiplicative term $\exp(\boldsymbol{x}_i\boldsymbol{\beta})$, but now modeling the odds function $R(t)$ with its associated baseline $R_0(t)$ for a reference level. The assumption of convergence for the mortality rates given two subjects with distinct values on covariates is similar to suppose a constant Odds Ratio (OR) for these same subjects. In other words, this is a proportional odds assumption. Thus, the odds function is modeled as

$$R(t|\boldsymbol{x}_i,\boldsymbol{\beta}) = R_0(t)\exp(\boldsymbol{x}_i\boldsymbol{\beta}), \tag{2.20}$$

and the OR for a pair of subjects i_1 and i_2 is given by

$$\text{OR}(i_1,i_2) = \frac{R_0(t)\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{R_0(t)\exp(\boldsymbol{x}_i\boldsymbol{\beta})} = \exp\left[(\boldsymbol{x}_{i_1}-\boldsymbol{x}_{i_2})\boldsymbol{\beta}\right], \tag{2.21}$$

For a PO model, the survival likelihood can be expressed in terms of the baseline odds function. In that case, taking now the expression in (1.10) and replacing the odds function and its derivative for their corresponding baselines, the survival likelihood for a PO model is given by (Panaro, 2020, p. 12)

$$\begin{aligned} L_{\text{PO}}(\boldsymbol{\kappa},\boldsymbol{\beta}|\boldsymbol{y},X) &\propto \prod_{i=1}^{n}\left[\frac{r(y_i)}{1+R(y_i)}\right]^{\delta_i}\frac{1}{1+R(y_i)} = \prod_{i=1}^{n}\left[\frac{r(y_i,\boldsymbol{\beta},\boldsymbol{\kappa}|\boldsymbol{x}_i)}{1+R(y_i,\boldsymbol{\beta},\boldsymbol{\kappa}|\boldsymbol{x}_i)}\right]^{\delta_i}\frac{1}{1+R(y_i,\boldsymbol{\beta},\boldsymbol{\kappa}|\boldsymbol{x}_i)} \\ &= \prod_{i=1}^{n}\left[\frac{r_0(y_i,\boldsymbol{\kappa})\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+R_0(y_i,\boldsymbol{\kappa})\exp(\boldsymbol{x}_i\boldsymbol{\beta})}\right]^{\delta_i}\frac{1}{1+R_0(y_i,\boldsymbol{\kappa})\exp(\boldsymbol{x}_i\boldsymbol{\beta})}. \end{aligned} \tag{2.22}$$

When specifying a Weibull ($\boldsymbol{\kappa}=(\alpha,\lambda)$); a BP model (now, $\boldsymbol{\kappa}=\boldsymbol{\zeta}$, applying the BP approximation to model the odds function instead of the hazard), or a PE model ($\boldsymbol{\kappa}=\boldsymbol{\lambda}$) as the baseline function, the following survival likelihood expressions are obtained replacing the terms from (2.22)

$$L_{\text{PO}}(\alpha,\lambda,\boldsymbol{\beta}|\boldsymbol{y},X) \propto \prod_{i=1}^{n}\left\{\frac{\lambda\alpha y_i^{\alpha-1}\exp(\lambda y_i^{\alpha}+\boldsymbol{x}_i\boldsymbol{\beta})}{1+[\exp(\lambda y_i^{\alpha})-1]\exp(\boldsymbol{x}_i\boldsymbol{\beta})}\right\}^{\delta_i}\frac{1}{1+[\exp(\lambda y_i^{\alpha})-1]\exp(\boldsymbol{x}_i\boldsymbol{\beta})}, \tag{Weibull}$$

$$L_{\text{PO}}(\boldsymbol{\zeta},\boldsymbol{\beta}|\boldsymbol{y},X) \propto \prod_{i=1}^{n}\left[\frac{\boldsymbol{\zeta}'\boldsymbol{g}_m(y_i)\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\boldsymbol{\zeta}'\boldsymbol{G}_m(y_i)\exp(\boldsymbol{x}_i\boldsymbol{\beta})}\right]^{\delta_i}\frac{1}{1+\boldsymbol{\zeta}'\boldsymbol{G}_m(y_i)\exp(\boldsymbol{x}_i\boldsymbol{\beta})}, \tag{BP}$$

$$\begin{aligned} L_{\text{PO}}(\boldsymbol{\lambda},\boldsymbol{\beta}|\boldsymbol{y},X) \propto \prod_{i=1}^{n}&\left\{\frac{\sum_{k=1}^{p}\lambda_k\mathbb{I}_{E_k}(y_i)\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{1+\left\{\exp\left[\boldsymbol{\lambda}'(\boldsymbol{t}-\boldsymbol{e})\right]-1\right\}\exp(\boldsymbol{x}_i\boldsymbol{\beta})}\right\}^{\delta_i}\times \\ &\frac{1}{1+\left\{\exp\left[\boldsymbol{\lambda}'(\boldsymbol{t}-\boldsymbol{e})\right]-1\right\}\exp(\boldsymbol{x}_i\boldsymbol{\beta})}. \end{aligned} \tag{PE}$$

Note that all the likelihoods above are valid for univariate data (without clustering). In Chapter 3 it will be seen that survival copula likelihood expressions depend on the combination of censoring values for each cluster. Marginally, it will be necessary to know the survival and density functions in terms of the hazard and its cumulative functions or of the odds and its derivative functions, depending on the assumption made for the proportionality (and consequently on the regression model class chosen).

2.2.3 Yang-Prentice (YP) Model

Yang and Prentice (2005) proposed a new regression model class in order to accommodate crossing survival curves. These can occur when, for instance, a treatment can be effective in the long run but may present adverse effects to the subjects in early stages of the follow-up, or, in clinical trials of a curable disease, proportions of long-term survival (*i.e.*, near the end of the study) for treatment and control groups may be equal, but deaths (the event of interest) may tend to appear earlier among the control patients. Isolated, the aforementioned PH and PO models are not suitable for survival data with crossing curves.

The YP model defines two vectors of short and long-term hazard ratio parameters to allow intersection among survival curves. Containing the PH and PO models as particular cases, the YP model can be used for model-fitting checking and may provide a more accurate inference when either the assumption of proportional hazards or odds are violated. Let T be a nonnegative random variable that represents the time until the occurrence of an event of interest; $\boldsymbol{x}_i$ the vector of covariate values for a subject i and $\Phi_i = (\phi_i^{(S)}, \phi_i^{(L)})$, with $\phi_i^{(S)} = \exp(\boldsymbol{x}_i\boldsymbol{\beta}^{(S)})$; $\phi_i^{(L)} = \exp(\boldsymbol{x}_i\boldsymbol{\beta}^{(L)})$, and $\boldsymbol{\beta}^{(S)}$, $\boldsymbol{\beta}^{(L)}$ are vectors of regression parameters with same length, neither of them including an intercept. Following Demarqui and Mayrink, 2021 and references therein, the survival function is given by

$$S(t|\Phi_i) = \left[1 + \frac{\phi_i^{(S)}}{\phi_i^{(L)}} R_0(t)\right]^{-\phi_i^{(L)}}, \tag{2.23}$$

where $R_0(t) = \exp[H_0(t)] - 1$ is the baseline odds function. If $\boldsymbol{x}_i = \mathbf{0}$, then (2.23) reduces to the baseline survival function $S_0(t) = 1/[1 + R_0(t)]$.

The hazard function associated with (2.23) can be expressed as

$$h(t|\Phi_i) = \frac{\phi_i^{(S)}\phi_i^{(L)} r_0(t)}{\phi_i^{(S)} + \phi_i^{(L)} R_0(t)} = \frac{\phi_i^{(S)}\phi_i^{(L)}}{\phi_i^{(S)} F_0(t) + \phi_i^{(L)} S_0(t)} h_0(t). \tag{2.24}$$

From (2.24), it follows that

$$\lim_{t\to 0} \frac{h(t|\Phi_i, \boldsymbol{x}_i)}{h(t|\Phi_i, \mathbf{0})} = \exp(\boldsymbol{x}_i\boldsymbol{\beta}^{(S)}) = \phi_i^{(S)}, \quad \lim_{t\to\infty} \frac{h(t|\Phi_i, \boldsymbol{x}_i)}{h(t|\Phi_i, \mathbf{0})} = \exp(\boldsymbol{x}_i\boldsymbol{\beta}^{(L)}) = \phi_i^{(L)}.$$

Therefore, $\phi_i^{(S)}$ and $\phi_i^{(L)}$ can be interpreted as the short and long-term hazard ratios for a subject i, respectively, and $\boldsymbol{\beta}^{(S)}$, $\boldsymbol{\beta}^{(L)}$ are the correspondent vectors of short and long-term coefficients. If $\boldsymbol{\beta}^{(S)} = \boldsymbol{\beta}^{(L)}$, the PH model is obtained. On the other hand, if $\boldsymbol{\beta}^{(L)} = \mathbf{0}$, then the YP model is reduced to the PO model. It can be shown that crossing survival curves are present if $\beta_l^{(S)}\beta_l^{(L)} < 0$ for any $l = 1, \ldots, q$, where q is the

number of covariates (Yang and Prentice, 2005). Given a baseline function with a vector of parameters $\boldsymbol{\kappa}$, the survival likelihood function can be expressed as

$$L_{\mathrm{YP}}(\boldsymbol{\kappa}, \boldsymbol{\beta}^{(S)}, \boldsymbol{\beta}^{(L)} | \boldsymbol{y}, X) \propto \prod_{i=1}^{n} \left[\frac{\phi_i^{(S)} \phi_i^{(L)}}{\phi_i^{(S)} F_0(y_i|\boldsymbol{\kappa}) + \phi_i^{(L)} S_0(y_i|\boldsymbol{\kappa})} h_0(y_i|\boldsymbol{\kappa}) \right]^{\delta_i} \left[1 + \frac{\phi_i^{(S)}}{\phi_i^{(L)}} R_0(y_i|\boldsymbol{\kappa}) \right]^{-\phi_i^{(L)}}.$$

When combined to a baseline distribution for the survival times, given a covariate representing the treatment and control indicator, the YP model can provide continuous crossing survival functions given both values on that covariate. If this is the case, due to that continuity, there exists a time point at which the survival curves intersect each other. Although the observed Fisher information matrix allows to obtain point and interval estimates for the YP model parameters, it is not straightforward to find an interval estimate for a crossing survival time t^*, since there is no closed form expression for the standard error of its estimator $\widehat{t^*}$ (Demarqui and Mayrink, 2021). A viable solution is to apply a numerical procedure to find the root that solves the nonlinear equation $S_C(t^*) - S_T(t^*) = 0$, where $S_C(\cdot)$ and $S_T(\cdot)$ denote the survival functions for control and treated subjects, respectively, maintaining all other covariates constant. This issue can be overcome by applying a resampling method that enables inference for t^*.

CHAPTER 3

Copulas

As stated in Chapter 1, copulas are a way to model dependence among observed values from the same cluster. The three most studied copula families in the literature are the elliptical copulas; the Archimedean copulas and the extreme-value copulas (see Chapter 3 of Hofert et al. (2018) and references therein to more details). In this chapter, the main topics of copula theory will be covered, highlighting additional properties for the family of Archimedean copulas, since some copula models of greater importance in the literature of survival analysis (and hence treated in this work) belong to this family. Measures to evaluate the strength of dependence given the estimate for a copula parameter are also discussed. For the copula theory development below, each marginal component is always supposed to be an absolutely continuous function associated with a random variable with continuous support. A more general version, including random variables with discrete support, is treated by Nelsen (2006).

3.1 Basics of the Copula Theory

In the literature, copulas are generally referred to as "functions that assemble or couple one-dimensional distribution functions to a multivariate distribution function" or as "distribution functions whose one-dimensional margins are all uniform" (Nelsen, 2006). Not all multivariate distributions are a copula itself, but each one can be reparameterized to a copula (see Section 1.2 of Joe (2014) for some examples). Before presenting the main definitions, let d be a positive integer value and $I = [0, 1]$ be the unit interval.

Definition 3.1 (Flores et al., 2017, p. 2-3). *A d-dimensional copula is a function $C : I^d \mapsto I$ that attends the following properties:*

(a) $C(u_1, \ldots, u_{j-1}, 0, u_{j+1}, \ldots, u_d) = 0$ *for all* $j = 1, \ldots, d$;

(b) $C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j$ *for all* $j = 1, \ldots, d$;

(c) For each rectangle $[\boldsymbol{a}, \boldsymbol{b}] = \prod_{j=1}^{d} [a_j, b_j] \subset I^d$, *with* $a_j \leqslant b_j$ *for all* $j = 1, \ldots, d$ *in the Cartesian product of intervals on the right-hand side, its volume is given by*

$$Vol_C([\boldsymbol{a}, \boldsymbol{b}]) = \sum_{j_1=1}^{2} \cdots \sum_{j_d=1}^{2} (-1)^{j_1 + \cdots + j_d} C\left(u_{1_{j_1}}, \ldots, u_{d_{j_d}}\right) \geqslant 0,$$

where $u_{j_1} = a_j$ *and* $u_{j_2} = b_j$ *for all* $j = 1, \ldots, d$.

In the particular case of bivariate copulas ($d = 2$), (3.1) is rewritten as

Definition 3.2 (Flores et al., 2017, p. 2). *A two-dimensional or bivariate copula ($d = 2$) is a function $C : I^2 \mapsto I$ attending the following properties:*

(a) For all $u_1, u_2 \in I$, $C(u_1, 0) = 0 = C(0, u_2)$;

(b) For all $u_1, u_2 \in I$, $C(u_1, 1) = u$ *and* $C(1, u_2) = u_2$;

(c) For all $a_1, b_1, a_2, b_2 \in I$ *such that* $a_1 \leqslant b_1$ *and* $a_2 \leqslant b_2$:

$$Vol_C([\boldsymbol{a}, \boldsymbol{b}]) = C(b_1, b_2) - C(b_1, a_2) - C(a_1, b_2) + C(a_1, a_2) \geqslant 0.$$

The most important result on copula theory was established by Sklar (1959). It allowed a representation for the joint distribution of d random variables as a function of the marginal distribution functions. Let $F_1, \ldots, F_d$ be a set of cumulative continuous distribution functions with range (*i.e.*, all values that F_j can assume) $\text{ran}F_j = I, j = 1, \ldots, d$, and F_j^{-1} the corresponding inverse function. The Sklar's Theorem follows below.

Theorem 3.1 (Sklar, 1959, *apud* Hofert et al., 2018, p. 23). *Let F be the continuous joint distribution function of a d-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$ with marginal continuous distribution functions $F_1, \ldots, F_d$. Then, there exists a d-dimensional copula C such that*

$$F(\boldsymbol{x}) = C\left[F_1(x_1), \ldots, F_d(x_d)\right], \quad \boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{3.1}$$

Moreover, C is uniquely defined by the Cartesian product of the d ranges, $\prod_{j=1}^{d} ranF_j = I^d$, and there given by

$$C(\boldsymbol{u}) = F\left[F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right], \quad \boldsymbol{u} \in I^d. \tag{3.2}$$

Conversely, given a d-dimensional copula C and a sequence of univariate distribution functions $F_1, \ldots, F_d$, F defined in (3.1) is a d-dimensional joint distribution function.

Note that the second result in (3.2) gives a method for construction of copulas from joint distributions. Thereby, many statistical applications of copulas consist of modeling continuous random vectors (Hofert et al., 2018). From (3.1), it follows that F is absolutely continuous if and only if all marginal distribution functions $F_1, \ldots, F_d$ and the copula C are absolutely continuous. To attend this property, the copula should admit a density c such that

$$c(\boldsymbol{u}) = \frac{\partial^d}{\partial u_d \cdots u_1} C(u_1, \ldots, u_d), \boldsymbol{u} \in I^d \tag{3.3}$$

exists and is integrable. Each margin of a copula function can be seen as a continuous standard uniform random variable, *i.e.*, $U_j \sim \mathrm{U}[0,1], j = 1, \ldots, d$. The copula itself is a multivariate cumulative distribution function; its d first-order partial derivatives are the joint conditional distribution functions for $d-1$ margins given the remaining margin (when $d = 2$, the conditional distribution for one margin given the another); its $d(d-1)$ second-order partial derivatives are the joint conditional distribution functions for $d-2$ margins given the remaining 2 margins, and so on. Differentiating d times, the copula joint density function is finally reached.

Taking the bivariate case ($d = 2$) for instance, the first-order derivative $\partial C(u_1, u_2)/\partial u_1$ is the conditional distribution for the margin U_2 given that $U_1 = u_1$, also denoted by $C_{2|1}(u_2|u_1)$. Similarly, $\partial C(u_1, u_2)/\partial u_2 = C_{1|2}(u_1|u_2)$ is the conditional distribution for the margin U_1 given that $U_2 = u_2$. Finally, $\partial^2 C(u_1, u_2)/\partial u_2 \partial u_1 = c(u_1, u_2)$ is the density function associated with the copula C.

A copula can specify a vector of parameters $\boldsymbol{\theta}$ to govern the dependence among its margins. This work will only deal with the case in which $\boldsymbol{\theta} = \theta$ is a single parameter of dependence. A watchful reader has already noted from Theorem 3.1 that a similar dependence relationship can be constructed for a joint survival function starting from marginal survival functions instead of the distribution ones since the survival function is defined as the complement of a distribution function. In fact, the Sklar's Theorem can be formulated using survival functions as well.

Theorem 3.2 (Sklar's Theorem for Survival Functions, Hofert et al., 2018, p. 41)**.** *Let S be the joint survival function of a d-dimensional random vector $\boldsymbol{T} = (T_1, \ldots, T_d)$ with marginal continuous survival functions $S_1, \ldots, S_d$. Then, there exists a d-dimensional survival copula $\overline{C}$ such that*

$$S(\boldsymbol{t}) = \overline{C}\left[S_1(t_1), \ldots, S_d(t_d)\right], \quad \boldsymbol{t} = (t_1, \ldots, t_d) \in \mathbb{R}^d. \tag{3.4}$$

The survival copula $\overline{C}$ is uniquely defined by the Cartesian product of the d ranges, $\prod_{j=1}^{d} ran S_j = I^d$, and there given by

$$\overline{C}(\boldsymbol{u}) = S\left[F_1^{-1}(1-u_1), \ldots, F_d^{-1}(1-u_d)\right], \quad \boldsymbol{u} \in I^d. \tag{3.5}$$

Conversely, given a d-dimensional survival copula $\overline{C}$ and a sequence of univariate survival functions $S_1, \ldots, S_d$, S defined in (3.4) is a d-dimensional survival function.

The survival copula $\overline{C}$ in (3.5) is also a distribution function, despite neither S nor $S_1, \ldots, S_d$ being. To understand this claim, note that for all $j \in 1, \ldots, d, u_j = S_j(t_j) = 1 - F_j(t_j)$. Since F_j and S_j are absolutely

continuous for all j, then $F_j^{-1}(1-u_j) = S_j^{-1}(u_j) = t_j$, and

$$\begin{aligned}\overline{C}(\boldsymbol{u}) &= S\left[F_1^{-1}(1-u_1), \ldots, F_d^{-1}(1-u_d)\right] = S\left[S_1^{-1}(u_1), \ldots, S_d^{-1}(u_d)\right] \\ &= \mathbb{P}\left[S_1^{-1}(U_1) \geqslant S_1^{-1}(u_1), \ldots, S_d^{-1}(U_d) \geqslant S_d^{-1}(u_d)\right] \\ &= \mathbb{P}\left[S_1(T_1) \leqslant S_1(t_1), \ldots, S_d(T_d) \leqslant S_d(t_d)\right] \\ &= \mathbb{P}\left[U_1 \leqslant u_1, \ldots, U_d \leqslant u_d\right].\end{aligned}$$

The Clayton, Frank and Gumbel-Hougaard (GH) Archimedean copulas are prominent examples of survival copula models with a single dependence parameter θ (Hofert et al., 2018). Now, let $i \in \{1, \ldots, n\}$ be the cluster (subject) index and $\boldsymbol{y}_i$ be a d-dimensional vector of observed times. Suppose that all marginal times $y_{i;1}, \ldots, y_{i;j}$, $j = 1, \ldots, d$ follow the same univariate parametric survival model with a set $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d$ of parameters for each margin, fixed for all i. Let $d_i = \sum_{j=1}^{d} \delta_{i;j}$ the number of survival times. Following p. 96–97 of Duchateau and Janssen (2007) and p. 486 of Prenen et al. (2017), but supposing that all clusters have the same size d for the multivariate case, the contribution of each cluster i to the likelihood function for the survival copula model in (3.4) with is given by

$$\begin{aligned}L_i(\theta, \boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d | \boldsymbol{y}_i) &\propto (-1)^{d_i} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} S(\boldsymbol{y}_i | \theta, \boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d) \\ &= \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \overline{C}\left[S_1(y_{i;1}|\boldsymbol{\kappa}_1), \ldots, S_d(y_{i;d}|\boldsymbol{\kappa}_d)|\theta\right] \prod_{j=1}^{d} f_j(y_{i;j}|\boldsymbol{\kappa}_j)^{\delta_{i;j}}, \qquad (3.6)\end{aligned}$$

where $\delta_{i;j}$ is the censoring indicator random variable for the observed time $y_{i;j}$. Therefore, the survival likelihood function is given by the product over all subjects i, $i = 1, \ldots, n$, of expression (3.6). For the bivariate case ($d = 2$), the survival copula likelihood function is expressed as

$$\begin{aligned}L_{\overline{C}}(\theta, \boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2 | \boldsymbol{y}) \propto \prod_{i=1}^{n} &\left\{\overline{C}\left[S_1(y_{i;1}|\boldsymbol{\kappa}_1), S_2(y_{i;2}|\boldsymbol{\kappa}_2)|\theta\right]\right\}^{(1-\delta_{i;1})(1-\delta_{i;2})} \\ &\times \left\{\frac{\partial \overline{C}\left[S_1(y_{i;1}|\boldsymbol{\kappa}_1), S_2(y_{i;2}|\boldsymbol{\kappa}_2)|\theta\right]}{\partial y_{i;1}}\right\}^{\delta_{i;1}(1-\delta_{i;2})} \\ &\times \left\{\frac{\partial \overline{C}\left[S_1(y_{i;1}|\boldsymbol{\kappa}_1), S_2(y_{i;2}|\boldsymbol{\kappa}_2)|\theta\right]}{\partial y_{i;2}}\right\}^{(1-\delta_{i;1})\delta_{i;2}} \\ &\times \left\{\frac{\partial^2 \overline{C}\left[S_1(y_{i;1}|\boldsymbol{\kappa}_1), S_2(y_{i;2}|\boldsymbol{\kappa}_2)|\theta\right]}{\partial y_{i;1} \partial y_{i;2}}\right\}^{\delta_{i;1}\delta_{i;2}} f_1(y_{i;1}|\boldsymbol{\kappa}_1)^{\delta_{i;1}} f_2(y_{i;2}|\boldsymbol{\kappa}_2)^{\delta_{i;2}}, \qquad (3.7)\end{aligned}$$

which is a product involving the survival copula $\overline{C}(u_1, u_2)$ in the first line; the conditional distribution functions $\overline{C}_{2|1}(u_2|u_1)$ and $\overline{C}_{1|2}(u_1|u_2)$ in the second and third lines, respectively, and the joint density function $\overline{c}(u_1, u_2)$, multiplied by the marginal densities, in the fourth line.

Initially, inferential procedures in survival copula models were made in two stages, by first estimating the marginal survival and density functions, and then maximizing the survival likelihood function after replacing the copula function and its derivatives by their estimated versions from the first stage (Shih, J. H. and Louis, T. A., 1995). This is similar to adopt a profile likelihood estimation, treating θ as a nuisance parameter. In this

work, however, the survival copula likelihood will be maximized at once, estimating all parameters from copula and marginal survival functions simultaneously. This allows incorporating all covariance structure among the parameters in the inference, avoiding underestimation on their standard errors.

3.2 Archimedean Copulas

According to Nelsen (2006), the class of Archimedean copulas find a broad range of applications due to its easy construction and attractive mathematical properties. Originally, Archimedean copulas appeared as a part of the development of a probabilistic version for the triangle inequality (Schweizer, 1991). Archimedean copulas are constructed through an additive generator function and its pseudo-inverse.

Definition 3.3 (Additive Generator, Durante and Sempi, 2015, p. 196)**.** *A function $\psi : [0, \infty) \mapsto I$, $I = [0, 1]$, is said to be additive generator if:*

(a) ψ is continuous and decreasing;

(b) $\psi(0) = 1$ and $\lim_{w \to \infty} \psi(w) = 0$;

(c) ψ is strictly decreasing on the interval $[0, w_0]$, where $w_0 = \inf\{w > 0 : \psi(w) = 0\}$.

Unless when said otherwise, whenever speaking of a generator ψ, it is meant to be an additive generator. The pseudo-inverse of the generator ψ, denoted as $\psi^{(-1)}(w)$, is equal to the inverse function $\psi^{-1}(w)$ if $w \in (0, 1]$ and equal to w_0 if $w = 0$.

Definition 3.4 (Archimedean Copula, Durante and Sempi, 2015, p. 196)**.** *A d-dimensional copula is said to be Archimedean if*

$$C(\boldsymbol{u}) = \psi\left[\psi^{(-1)}(u_1) + \ldots + \psi^{(-1)}(u_d)\right] = \psi\left[\sum_{j=1}^{d} \psi^{(-1)}(u_j)\right], \quad \boldsymbol{u} \in I^d. \tag{3.8}$$

Since Archimedean copulas are defined under a generator ψ, the notation C_ψ can be used to replace $C(\boldsymbol{u})$ in (3.8). Not all generators are suitable to construct an Archimedean copula: a generator ψ needs also to be a d-monotone function for some value d.

Definition 3.5 (d-Monotone and Completely Monotone Functions, Durante and Sempi, 2015, p. 197)**.** *A function $\psi : (a, b) \mapsto \mathbb{R}$ is said to be d-monotone in (a, b), where $-\infty \leqslant a < b \leqslant +\infty$ and $d \geqslant 2$, if:*

(a) ψ admits derivatives $\psi^{(k)}$ up to the order $k = d - 2$;

(b) For all $w \in (a, b)$, $(-1)^k \psi^{(k)}(w) \geqslant 0$, for $k \in \{0, 1, \ldots, d - 2\}$;

(c) $(-1)^{d-2}\psi^{(d-2)}$ is decreasing and convex in the interval (a, b).

Moreover, if ψ has derivatives of any order in (a, b) and if $(-1)^k \psi^{(k)}(w) \geqslant 0$ for all $w \in (a, b)$ and for all $k \in \mathbb{Z}_+$ (the set of the non-negative integers), then ψ is also said to be completely monotone.

The following result yields a characterization for the Archimedean copulas C_ψ.

Theorem 3.3 (Durante and Sempi, 2015, p. 198)**.** *Let $\psi(w) : [0, \infty)$ be a generator and $d \geqslant 2$ a fixed integer. The following statements are equivalent:*

(a) ψ is d-monotone on $[0, \infty)$;

(b) The function $C_\psi : I^d \mapsto I$ is a d-dimensional copula.

An immediate consequence of this result is that, when $d = 2$, C_ψ is a copula if and only if ψ is a convex function. Within the class of Archimedean copulas, there is a subclass whose members are generated by completely monotone functions. The GH and Joe copulas are examples of that subclass. Completely monotone generators can be written as a Laplace-Stieltjes (LS) transform.

Definition 3.6 (Joe, 2014, p. 33, Hofert et al., 2018, p. 99)**.** *Let V be a non-negative random variable with distribution function F_V. The Laplace-Stieltjes (LS) transform of F_V is given by*

$$\mathcal{LS}_V(w) = \mathbb{E}[\exp(-wV)] = \int_0^\infty \exp(-wv) dF_V(v), \quad w \in [0, \infty). \tag{3.9}$$

Since $|\exp(-wv)| \leqslant 1$ for all $v, w \in [0, \infty)$, the LS transform always exist for non-negative variables. In that case, it is more convenient to use than the moment generating function. Some useful properties of the LS transform for an Archimedean copula generation are (see Joe, 2014, p. 33, and references therein):

1. $\mathcal{LS}_V(0) = 1$; $\mathcal{LS}_V(w) \leqslant 1$ for all $w \in [0, \infty)$ and $\mathcal{LS}_V$ is strictly decreasing;
2. A probability distribution with support on $[0, \infty)$ is uniquely determined by its LS transform;
3. $\mathcal{LS}_V$ is completely monotone;
4. Bernstein's theorem (Bernstein, 1928) for the special case of an expected value as a weighted average: A function $\mathcal{LS}_V : (0, \infty) \mapsto [0, 1]$ is the LS transform of a non-negative random variable V if and only if it is completely monotone and $\mathcal{LS}_V(0) = 1$;
5. If V has positive mass p_0 at 0, then (3.9) is rewritten as

$$\mathcal{LS}_V(w) = p_0 + \int_0^\infty \exp(-wv) dF_V(v), \quad w \in (0, \infty)$$

   and $\mathcal{LS}_V(\infty) = \lim_{w \to \infty} \mathcal{LS}_V(w) = p_0$. Therefore, $\mathcal{LS}_V(\infty) = 0$ if and only if $p_0 = 0$;
6. If $\mathcal{LS}_V(\infty) = p_0$, its functional inverse $\mathcal{LS}_V^{-1} : [p_0, 1] \mapsto [0, \infty)$ is strictly decreasing and satisfies $\mathcal{LS}_V^{-1}(p_0) = \infty$, $\mathcal{LS}_V^{-1}(1) = 0$.

To define a LS transform as generator of an Archimedean copula ($\mathcal{LS}_V = \psi$), it is necessary that $\mathcal{LS}_V(\infty) = p_0 = 0$ (V cannot have any positive mass at 0). Thus, $\mathcal{LS}_V^{(-1)} = \psi^{(-1)}$ has a closed domain $[0, 1]$, with $\mathcal{LS}_V^{(-1)}(0) = \infty$ (hence, $\mathcal{LS}_V^{-1} = \psi^{-1}$). When $\psi = \psi_\theta$ is indexed by a single parameter of dependence θ, the same parameter indexes the distribution $F_V = F_{V|\theta}$.

Knowing that each Archimedean copula generator can be represented by a LS transform, an alternative stochastic representation can be used to generate from the marginal random variables $U_1, \ldots, U_d$ (Durante

and Sempi, 2015). Let $E_1, \ldots, E_d$ be independent and identically distributed random variables following an exponential distribution with mean 1, and V a positive and independent random variable with LS transform $\mathcal{LS}_V = \psi$. Then, the random vector

$$(U_1, \ldots, U_d) = \left[\psi\left(\frac{E_1}{V}\right), \ldots, \psi\left(\frac{E_d}{V}\right)\right] \tag{3.10}$$

has joint distribution function given by the Archimedean copula C_ψ. Then, a sample can be drawn from an Archimedean copula generated by a completely monotone ψ under the following steps (Marshall and Olkin, 1988):

1. Sample V from $F_V = \mathcal{LS}^{-1}[\psi]$;
2. Sample $E_1, \ldots, E_d \overset{\text{iid}}{\sim} \text{Exp}(1)$, independently of V;
3. Return $\boldsymbol{U} = [\psi(E_1/V), \ldots, \psi(E_d/V)]$ as a sample from $\overline{C}$.

Before its use for Archimedean copulas, LS transforms have already been explored in the context of shared frailty models belonging to the power variance function family (see Chapter 4 of Duchateau and Janssen, 2007 for more details), since the dependence induced by distributions of those frailty models can also be expressed in terms of a LS transform. For that reason, F_V is referred as a frailty distribution in the literature.

A minor drawback of the LS transform as a representation for a (completely monotone) generator ψ is that its validity is restricted to cases where $\theta \in [0, \infty)$ (positive dependence). However, it is possible to define Archimedean copulas that also allow negative dependence modeling, although d-monotonicity for those models is guaranteed only for $d = 2$ (see Joe, 2014, p. 48, for further details). Examples of Archimedean copula models accommodating negative dependence are the AMH and Frank copulas. Section 4.4 of Joe (1997) discuss how to generate from Archimedean copulas with negative dependence.

If $\overline{C}_\psi$ is a completely monotone Archimedean copula over marginal survival functions $S(t_j|\boldsymbol{\kappa}_j), j = 1, \ldots, d$, with generator $\psi_\theta = \mathcal{LS}_{V|\theta}$ (then, $p_0 = 0$), the joint survival function in (3.4) for each cluster i can be rewritten as (Prenen et al., 2017, p. 486)

$$\begin{aligned} S_i(\boldsymbol{t}|\theta, \boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d) &= \overline{C}\left[S_1(t_{i;1}), \ldots, S_d(t_{i;d})|\theta, \boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d\right] = \psi_\theta\left\{\psi_\theta^{-1}\left[\sum_{j=1}^{d} S_j(t_{i;j}|\boldsymbol{\kappa}_j)\right]\right\} \\ &= \int_0^\infty \exp\left\{-v\sum_{j=1}^{d}\psi_\theta^{-1}\left[S_j(t_{i;j}|\boldsymbol{\kappa}_j)\right]\right\} dF_{V|\theta}(v|\theta). \end{aligned} \tag{3.11}$$

Hence, the contribution of each cluster i to the likelihood function in (3.6) is given by

$$L_i(\theta, \boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d|\boldsymbol{y}_i) \propto \int_0^\infty \exp\left\{-v\sum_{j=1}^{d}\psi_\theta^{-1}\left[S_j(y_{i;j}|\boldsymbol{\kappa}_j)\right]\right\} \prod_{j=1}^{d}\left\{\frac{-v f_j(y_{i;j}|\boldsymbol{\kappa}_j)}{\psi_\theta^{(1)}\left[\psi_\theta^{-1}\left(S_j(y_{i;j}|\boldsymbol{\kappa}_j)\right)\right]}\right\}^{\delta_{i;j}} dF_{V|\theta}(v|\theta), \tag{3.12}$$

where $\psi_\theta^{(1)}$ is the first-order derivative of ψ_θ. Before obtain the survival copula likelihood expression for all clusters, let $d_i = \sum_{j=1}^{d}\delta_{i;j}$ and consider the following result for the g-order derivatives of $\psi = \mathcal{LS}_V$ (Prenen

et al., 2017, p. 486)

$$\psi^{(g)}(w) = \int_0^\infty (-v)^g \exp(-wv) dF_V(v). \tag{3.13}$$

Taking the contribution in (3.12) over all clusters, the following survival copula likelihood is obtained

$$\begin{aligned} L_{\overline{C}_\psi}(\theta, \boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d | \boldsymbol{y}) &\propto \prod_{i=1}^n L_i(\theta, \boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d | \boldsymbol{y}_i) \\ &= \prod_{i=1}^n \int_0^\infty \exp\left\{-v \sum_{j=1}^d \psi_\theta^{-1}\left[S_j(y_{i;j}|\boldsymbol{\kappa}_j)\right]\right\} \prod_{j=1}^d \left\{ \frac{-v f_j(y_{i;j}|\boldsymbol{\kappa}_j)}{\psi_\theta^{(1)}\left[\psi_\theta^{-1}\left(S_j(y_{i;j}|\boldsymbol{\kappa}_j)\right)\right]} \right\}^{\delta_{i;j}} dF_{V|\theta}(v|\theta) \\ &= \prod_{i=1}^n \int_0^\infty \prod_{j=1}^d \exp\left\{-v\psi_\theta^{-1}\left[S_j(y_{i;j}|\boldsymbol{\kappa}_j)\right]\right\} \left\{ \frac{-v f_j(y_{i;j}|\boldsymbol{\kappa}_j)}{\psi_\theta^{(1)}\left[\psi_\theta^{-1}\left(S_j(y_{i;j}|\boldsymbol{\kappa}_j)\right)\right]} \right\}^{\delta_{i;j}} dF_{V|\theta}(v|\theta) \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^d \left[\frac{f_j(y_{i;j}|\boldsymbol{\kappa}_j)}{\psi_\theta^{(1)}\left(\psi_\theta^{-1}\left(S_j(y_{i;j}|\boldsymbol{\kappa}_j)\right)\right)} \right]^{\delta_{i;j}} \right\} \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^d \psi_\theta^{-1}\left[S_j(y_{i;j}|\boldsymbol{\kappa}_j)\right] \right\}. \end{aligned} \tag{3.14}$$

In the sequel, five Archimedean copula models for the development of survival copulas in this work will be presented through their multivariate representations for the copula function and the associated generator, following mainly Joe (2014) for the copula model expressions and Hofert et al. (2018) for the associated frailty distributions when the generator is a completely monotone function. All of them specify a unique parameter θ to model the copula dependence.

Ali-Mikhail-Haq (AMH) Copula

The Ali-Mikhail-Haq (AMH) copula (Ali et al., 1978) was originally proposed as a bivariate extension for the univariate logistic distribution. Starting from the generator $\psi_\theta(w) = (1-\theta)/(\exp(w) - \theta)$, where $\theta \in [0, 1]$, the general expression for the d-dimensional AMH copula is given by

$$C(\boldsymbol{u}|\theta) = \frac{1-\theta}{\prod_{j=1}^d \left[\dfrac{1-\theta(1-u_j)}{u_j}\right] - \theta}, \quad \boldsymbol{u} \in I^d \tag{3.15}$$

The distribution $F_{V|\theta}$ associated with ψ_θ as a LS transform is the Geometric distribution Geo(p) with probability mass function $p_k = p(1-p)^{k-1}, k \in \mathbb{N}$, where $p = 1 - \theta$.

If $d = 2$, expression (3.15) can be extended to accommodate negative dependence (Joe, 1997; Joe, 2014). In that case, it reduces to $C(u_1, u_2) = u_1 u_2/[1 - \theta(1-u_1)(1-u_2)]$, $\theta \in [-1, 1]$.

Clayton Copula

The Clayton copula (Clayton, 1978) has born as a survival model to the problem of demonstrating association (dependence) between pairs of subjects for a disease incidence. Its parameter of association θ can assume any value in the interval $[-1/(d-1), \infty)$, with generator $\psi_\theta(w) = [\max(0, 1 + \theta w)]^{-1/\theta}$. However, taking the bivariate case for instance, when $\theta \in [-1, 0)$, the copula support is restricted to a region that satisfies $\psi_\theta^{-1}(u_1) + \psi_\theta^{-1}(u_2) < \psi_\theta^{-1}(0) = -1/\theta$ (Cooray, 2018). Thus, it depends on θ and the density is 0 on the set $\{(u_1, u_2) : u_1^{-\theta} + u_2^{-\theta} < 1\}$ (Joe, 2014). To avoid this problem, only the non-negative part of the dependence parameter support will be addressed in this work, *i.e.*, $\theta \in [0, \infty)$. In that case, the general expression for the

d-dimensional Clayton copula is given by

$$C(\boldsymbol{u}|\theta) = \left\{ \max \left[\sum_{j=1}^{d} u_j^{-\theta} - (d-1),\, 0 \right] \right\}^{-1/\theta}, \quad \boldsymbol{u} \in I^d. \tag{3.16}$$

The distribution $F_{V|\theta}$ associated with ψ_θ as a LS transform is the Gamma distribution Gamma(a, b) with probability density function $p(k) = b^a k^{a-1} \exp(-bk)[\Gamma(a)]^{-1}$, setting $a = 1/\theta$ and $b = 1$.

Frank Copula

The Frank copula (Frank, 1979) arose on a purely mathematical context, but with many properties already discovered in its debut. Starting from the generator $\psi_\theta(w) = -\log[1-(1-\exp^{-\theta})\exp^{-w}]/\theta$, where $\theta \in (0, \infty)$, the general expression for the d-dimensional Frank copula is given by

$$C(\boldsymbol{u}|\theta) = -\frac{1}{\theta} \log \left\{ 1 - \frac{\prod_{j=1}^{d} [1 - \exp(-\theta u_j)]}{[1 - \exp(-\theta)]^{d-1}} \right\}, \quad \boldsymbol{u} \in I^d. \tag{3.17}$$

The distribution $F_{V|\theta}$ associated with ψ_θ as a LS transform is the Logarithmic distribution Log(p) with probability mass function $p_k = p^k[-k\log(1-p)]^{-1}, k \in \mathbb{N}$, where $p = 1 - \exp(-\theta)$.

Like in the AMH copula, the expression in (3.16) can be extended to allow negative dependence for on the whole negative part of the real line if $d = 2$. In that case, it reduces to $C(u_1, u_2) = -(1/\theta)$ $\log\left\{\left[1 - \exp^{-\theta} - \left(1 - \exp^{-\theta u_1}\right)\left(1 - \exp^{-\theta u_2}\right)\right] / \left(1 - \exp^{-\theta}\right)\right\}$, $\theta \in \mathbb{R} \setminus \{0\}$. The range of $\theta < 0$ decreases as d increases, leading to a limited negative dependence if $d \geqslant 3$ (Joe, 2014).

Among the Archimedean copula models, the Frank copula is the only one that has the reflection symmetry property (Frank, 1979), i.e., $C(u_1, \ldots, u_d) = C(1-u_1, \ldots, 1-u_d)$. As consequence, a survival Frank copula is equal to a Frank copula over marginal distribution functions.

Gumbel-Hougaard (GH) Copula

The Gumbel-Hougaard (GH) copula (Nelsen, 2006), sometimes referred as Bivariate Gumbel (Joe, 2014), was originally proposed by Gumbel (1960) and later discussed by Hougaard (1986) in the survival analysis context. This copula model generator is given by $\psi_\theta(w) = \exp(-w^{1/\theta})$, where $\theta \in [1, \infty)$. Such generator is also a positive stable LS transform Joe (2014). This unique property among the Archimedean copulas (see Theorem 4.5.2, p. 143, from Nelsen, 2006) also makes the GH copula a member of the extreme value copula class. The general expression for the d-dimensional GH copula is given by

$$C(\boldsymbol{u}|\theta) = \exp\left\{ -\left[\sum_{j=1}^{d} (-\log(u_j))^\theta \right]^{1/\theta} \right\}, \quad \boldsymbol{u} \in I^d. \tag{3.18}$$

The distribution $F_{V|\theta}$ associated with ψ_θ as a LS transform is the 1-parameterization of the Stable distribution S$(1/a, 1, \cos^{1/a}(a\pi/2), 1(a=1); 1)$ with characteristic function $\phi(k) = \exp[-(-ik)^a][\Gamma(a)]^{-1}$, setting $a = \theta$ (see Nolan, 2020 for more details on the general family of Stable distributions).

Joe Copula

The Joe copula first appeared in another work of Frank (1981), but its properties have been investigated in depth by Joe (1993), initially specifying two parameters to model the copula dependence. For the purposes of this work, the copula specification considered is the one in Section 3.1 of the second paper, but taking only θ as the dependence parameter (the other being fixed at 1). Taking the generator $\psi_\theta(w) = 1-[1-\exp(-w)]^{1/\theta}$, where $\theta \in [1, \infty)$, the general expression for the d-dimensional Joe copula is given by

$$C(\boldsymbol{u}|\theta) = 1 - \left\{ 1 - \prod_{j=1}^{d} \left[1 - (1-u_j)^{\theta} \right] \right\}^{1/\theta}, \quad \boldsymbol{u} \in I^d. \tag{3.19}$$

The distribution $F_{V|\theta}$ associated with ψ_θ as a LS transform is the Sibuya distribution $\text{Si}(p)$ with probability mass function $p_k = \binom{p}{k}(-1)^{k-1}, k \in \mathbb{N}$, where $p = 1/\theta$.

3.3 Measuring Dependence

In general, it is desirable from a practical point of view to summarize the dependence among components of a random vector by a real number (Hofert et al., 2018). In the bivariate case, the two most common ways to summarize dependence between two random variables X_1 and X_2 are through measures of monotone association (bivariate concordance) or tail dependence (Joe, 2014).

A monotone association between random variables exists if whenever one variable increases (decreases) the other also increases (decreases). In other words, there is a monotone (but not necessarily linear) relationship in terms of the conditional expectation or median of one variable given the another. As this relationship approaches a monotone function in probability, the measure increases in absolute value. In the copula theory, a monotone measure of association must be a function $q(\theta; X_1, X_2)$ of the dependence parameter and random variables that satisfies the following properties (see Joe, 2014, p. 54 for more details):

1. **Domain**: $q(\theta; X_1, X_2)$ can be defined for all pairs of random variables;
2. **Symmetry**: $q(\theta; X_1, X_2) = q(\theta; X_2, X_1)$;
3. **Range**: $q(\theta; X_1, X_2) \in [-1, 1]$;
4. **Independence**: If X_1 and X_2 are independent, then $q(\theta; X_1, X_2) = 0$;
5. **Invariance**: If h_1, h_2 are strictly increasing functions, then $q[\theta; h_1(X_1), h_2(X_2)] = q(\theta; X_1, X_2)$.

As Hofert et al. (2018) pointed out, the Pearson's coefficient does not attend all the properties above, although being a simple measure of monotone (linear) association. The domain and independence can only be evaluated if both X_1 and X_2 have finite 2nd order moments. Moreover, this measure is not invariant under strictly increasing nonlinear transformations. By only depending on a dependence parameter from a underlying copula, measures of rank correlations such as the Kendall's τ and Spearman's ρ coefficients fulfill the aforesaid properties.

Definition 3.7 (Joe, 2014, p. 55). *Let (X_1, X_2) and (X_1^*, X_2^*) be two independent random pairs with a common joint continuous distribution F_{12} and copula C. The Kendall's τ measure is given by*

$$\begin{aligned}\tau &= \mathbb{P}\left[(X_1 - X_1^*)(X_2 - X_2^*) > 0\right] - \mathbb{P}\left[(X_1 - X_1^*)(X_2 - X_2^*) < 0\right] \\ &= 2\mathbb{P}\left[(X_1 - X_1^*)(X_2 - X_2^*) > 0\right] - 1 \\ &= 4\int F_{12} dF_{12} - 1 \\ &= 4\int_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1. \end{aligned} \tag{3.20}$$

Since the Kendall's τ will depend on a known copula with well defined conditional distributions for each of its margins, the coefficient can be directly estimated by plugging the estimate for θ from the survival copula modeling in (3.20). Note that Kendall's τ is defined as the difference between probabilities of concordance (two pairs (X_1, X_2) and (X_1^*, X_2^*) are concordant if $(X_1 - X_1^*)(X_2 - X_2^*) > 0$) and discordance (they are discordant if $(X_1 - X_1^*)(X_2 - X_2^*) < 0$). For Archimedean copulas, the Kendall's τ can be rewritten in terms of the corresponding generator as (Nelsen, 2006, p. 163 and 166)

$$\tau = 4\int_0^1 \frac{\psi^{-1}(w)}{\psi^{-1;(1)}(w)} dw + 1 = 1 - 4\int_0^\infty w \left[\frac{d\psi(w)}{dw}\right]^2 dw.$$

However, the same cannot be said for the Spearman's ρ, since its integral (defined below) does not have a closed form for some Archimedean copula models, such as the Clayton and Joe copulas.

Definition 3.8 (Joe, 2014, p. 56). *Let (X_1, X_2), (X_1^*, X_2^*) and (X_1^{**}, X_2^{**}) be three independent random pairs with a common joint continuous distribution F_{12} (with marginals F_1 and F_2) and copula $\overline{C}$. The Spearman's ρ measure is given by*

$$\begin{aligned}\rho &= 3\left\{\mathbb{P}\left[(X_1 - X_1^*)(X_2 - X_2^{**}) > 0\right] - \mathbb{P}\left[(X_1 - X_1^*)(X_2 - X_2^{**}) < 0\right]\right\} \\ &= 3\left\{2\mathbb{P}\left[(X_1 - X_1^*)(X_2 - X_2^{**}) > 0\right] - 1\right\} \\ &= 3\left(4\int F_{12} dF_{12} - 1\right) \\ &= 3\left(4\int F_1 F_2 dF_{12} - 1\right) \quad (\textit{by independence of } X_1^* \textit{ and } X_2^{**}) \\ &= 3\left(4\int_{[0,1]^2} u_1 u_2 dC(u_1, u_2) - 1\right). \end{aligned} \tag{3.21}$$

The result in (3.21) can also be expressed in terms of a Riemann integral or using one of the marginal conditional distributions as $\rho = 12\int_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3$ (Joe, 2014, p. 56). Like the Kendall's τ coefficient, Spearman's ρ can also be written as a function of θ. However, the meaning behind its formal definition is slightly different since it allows for independence between the second margin components, i. e., the random vector (X_1, X_2) has joint distribution function F_{12}, but the vector (X_1^*, X_2^{**}) has joint distribution function given by the product $F_1 F_2$, which is not guaranteed to be equal to F_{12}. Table 3.1 presents the Kendall's τ and Spearman's ρ measures as functions of θ for the AMH, Clayton, Frank, GH and Joe copulas. Computing of

those coefficients will be done through functions of the R `copula` package (Hofert et al., 2020), including the cases where there is no closed form or a special function is involved.

Table 3.1: Kendall's τ and Spearman's ρ coefficients for some bivariate Archimedean copula models

Copula	Kendall's τ	Spearman's ρ
AMH[1]	$\frac{3\theta-2}{3\theta} - \frac{2(1-\theta)^2\log(1-\theta)}{3\theta^2}$	$\frac{12(1+\theta)\mathrm{dl}(1-\theta) - 24(1-\theta)\log(1-\theta)}{\theta^2} - \frac{3(\theta+12)}{\theta}$
Clayton[2]	$\frac{\theta}{\theta+2}$	$12\frac{\theta+1}{(\theta+2)^2}\,{}_3F_2\left(\frac{\theta+1}{\theta}, 1, 1, 2\frac{\theta+1}{\theta}, \frac{\theta+1}{\theta}; 1\right)$
Frank[3]	$1 + \frac{4}{\theta}\left[D_1(\theta) - 1\right]$	$1 + \frac{12}{\theta}\left[D_2(\theta) - D_1(\theta)\right]$
GH[3]	$\frac{\theta-1}{\theta}$	$12\int_0^1 \left[1 + B(u,\theta)\right]^{-2} du - 3$
Joe[3]	$1 + \frac{2}{2-\theta}\left[\Psi(2) - \Psi\left(\frac{2}{\theta}+1\right)\right]$	No closed form

[1]See Kumar (2010), Theorem 1.
[2]See Klein et al. (2013), Subsection 24.2.4.
[3]See Joe (2014), Chapter 4.

The following special functions present in Table 3.1 are:

- $\mathrm{dl}(x) = \int_1^x \log(a)(1-a)^{-1} da$ is the dilogarithm function;
- ${}_3F_2\left(a_1, a_2, a_3, a_4, a_5; a_6\right) = \sum_{l=0}^{\infty} \frac{\Gamma(a_1+l)\Gamma(a_2+l)\Gamma(a_3+l)\Gamma(a_4)\Gamma(a_5)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)\Gamma(a_4+l)\Gamma(a_5+l)} \frac{a_6^l}{l!}$ is a hypergeometric function;
- $D_k(x) = kx^{-k}\int_0^x a^k \left[\exp(a) - 1\right]^{-1} da$, with $k \in \mathbb{N}$, is the Dèbye function;
- $B(b,c) = \int_0^1 a^{b-1}(1-a)^{c-1} da$, with $b, c \in \mathbb{R}^+$, is the Beta function;
- $\Psi(x) = (d/dx)\log\left[\Gamma(x)\right]$ is the digamma function.

According to Joe (2014), tail dependence is a way to measure the amount of dependence in the joint lower or upper tails of a multivariate distribution function. In bivariate distributions, tail dependence is related to the amount of dependence in the lower or upper quadrant tails. For copula models, both tails will be in the square $[0,1]^2$. Tail dependence coefficients are also invariant to increasing transformations.

Definition 3.9 (Durante and Sempi, 2015, p. 73)**.** *Let (X_1, X_2) be a continuous random pair with marginal distribution functions F_1 and F_2. The lower and upper tail dependence coefficients χ_L and χ_U are defined by*

$$\chi_L = \lim_{x \to 0^+} \mathbb{P}\left[X_2 \leqslant F_2^{-1}(x) | X_1 \leqslant F_1^{-1}(x)\right], \tag{3.22}$$

$$\chi_U = \lim_{x \to 1^-} \mathbb{P}\left[X_2 > F_2^{-1}(x) | X_1 > F_1^{-1}(x)\right], \tag{3.23}$$

provided that both limits exist.

Thus, tail dependence coefficients provide an asymptotic approximation for the behavior of a distribution function in the tails. For Archimedean copulas, the expressions in (3.22) and (3.23) can be rewritten in terms

of the corresponding generator as (Nelsen, 2006, p. 215)

$$\chi_L = \lim_{w \to +\infty} \frac{\psi(2w)}{\psi(w)}, \chi_U = 2 - \lim_{w \to 0^+} \frac{1 - \psi(2w)}{1 - \psi(w)}.$$

Thereby, $\chi_L = \chi_U = 0$ for the AMH and Frank copulas; $\chi_L = 0$ and $\chi_U = 2 - 2^{1/\theta}$ for the GH and Joe copulas; and finally $\chi_L = 2^{-1/\theta}$ and $\chi_U = 0$ for the Clayton copula (when $\theta \geqslant 0$). As $\theta \to \infty$, the lower (upper) tail dependence for the Clayton copula (for the GH and Joe copulas) tends to 1. In the survival analysis context, the presence of lower tail dependence (as in the Clayton copula) implies a higher association between greater survival times, while the presence of upper tail dependence (as in the GH and Joe copulas) implies a stronger correlation between smaller survival times (Prenen et al., 2017).

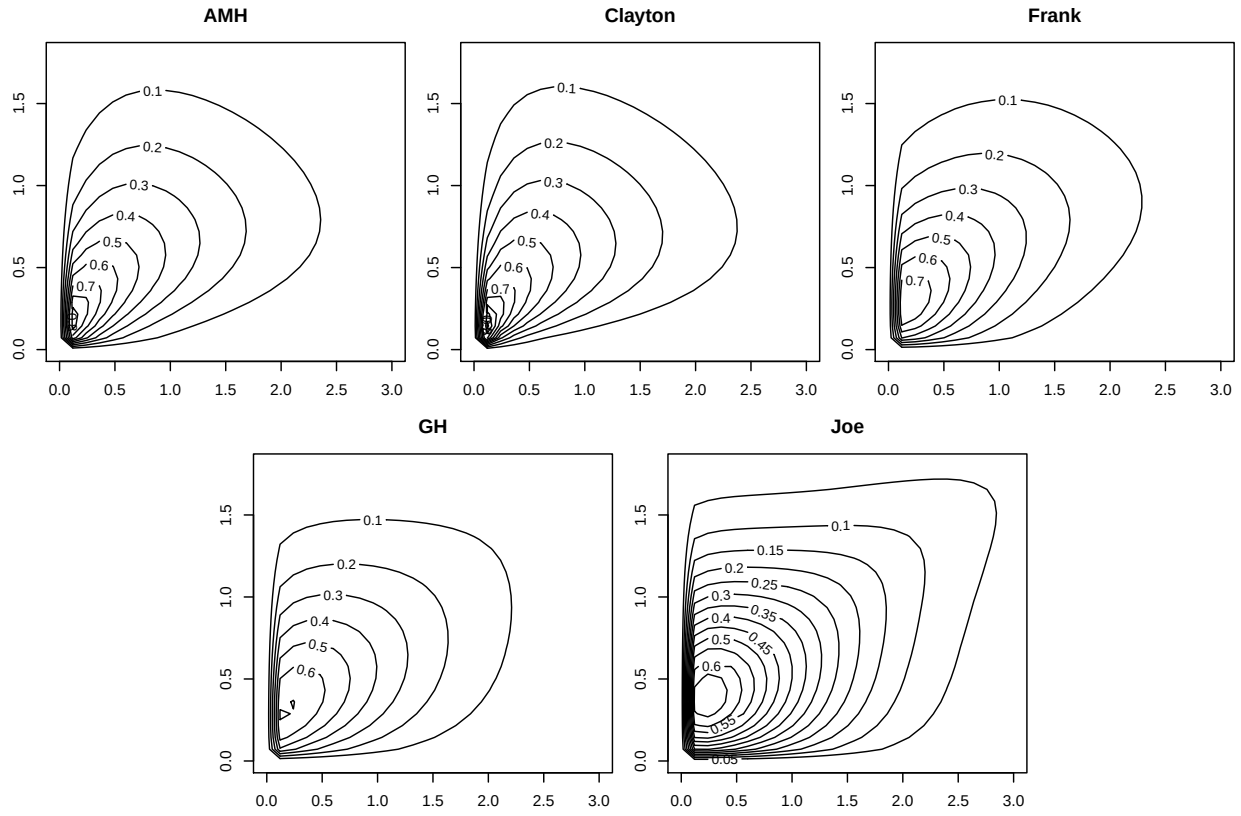


Figure 3.1: Contour plots on the density for AMH, Clayton, Frank, GH and Joe copulas, given $\tau = 0.25$ and Weibull margins with $\boldsymbol{\kappa}_1 = (\alpha_1, \lambda_1) = (1.2, 0.8)$ and $\boldsymbol{\kappa}_2 = (\alpha_2, \lambda_2) = (1.6, 0.2)$

CHAPTER 4

Proposed Survival Copula Models

The previous two chapters showed (i) parametric survival models and regression model classes of major concern for this work in a univariate framework, and (ii) copula theory and its applications using univariate survival functions as components of the Archimedean copula class. This chapter presents the survival copula likelihood functions for the developed models in this work, given a sample of observed times $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_d)$ and an array of design matrices $X = (X_1, \ldots, X_d)$, $j = 1, \ldots, d$. The parameters are specified by a collection $\mathcal{P} = \{\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}\}$ (for PH and PO regression model classes) or $\mathcal{P} = \{\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}^{(S)}, \boldsymbol{\beta}^{(L)}\}$ (for the YP model), with $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_d)$ being marginal survival model parameters from the same family of baseline distributions, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)$ being parameters from the same regression model class for all margins (similarly for the YP class).

For all possible combinations of baseline distribution and regression structure, the following likelihood functions are versions of (3.6) for the general case of an Archimedean copula in (3.8), and (3.14) for the particular case of a completely monotone generator for an Archimedean copula. In both cases, marginal survival and density functions are replaced according to the parametric survival model (one of Weibull, BP or PE) for the baseline distribution, and the regression model class (one of PH, PO or YP) considered. Since five Archimedean copulas (AMH, Clayton, Frank, GH and Joe) are addressed in this work, a total of 45 survival copula models are developed under the proposed framework. As seen in Section 1.3, Archimedean copula models with Weibull PH components are not exactly a novelty, but due to the use of marginal Weibull PH models as a benchmark, they are presented in the next sections. For the same reason, those copula models will also be used for data generation in the simulation study done in Chapter 5.

4.1 Survival Copula Models with Marginal Weibull Baseline

4.1.1 Weibull Proportional Hazards Copula Model

The survival copula likelihood function for a model with Weibull PH marginal components and parameters $\boldsymbol{\kappa}_j = (\alpha_j, \lambda_j)$ and $\boldsymbol{\beta}_j$, $j = 1, \ldots, d$, over all subjects i, $i = 1, \ldots, n$, is given by

$$L_{\overline{C}(\text{WPH})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) = \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \exp\left(-\lambda_j y_{i;j}^{\alpha_j} \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right) \right] \right\} \times \prod_{j=1}^{d} \left\{ \lambda_j \alpha_j y_i^{\alpha_j - 1} \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right) \exp\left[-\lambda_j y_{i;j}^{\alpha_j} \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right] \right\}^{\delta_{i;j}}. \tag{4.1}$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.1) can be rewritten as

$$L_{\overline{C}(\text{WPH})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\lambda_j \alpha_j y_i^{\alpha_j - 1} \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right) \exp\left(-\lambda_j y_{i;j}^{\alpha_j} \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)}{\psi_\theta^{(1)}\left(\psi_\theta^{-1}\left(\exp\left(-\lambda_j y_{i;j}^{\alpha_j} \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)\right)\right)} \right]^{\delta_{i;j}} \right\} \times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\exp\left(-\lambda_j y_{i;j}^{\alpha_j} \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right) \right] \right\}.$$

4.1.2 Weibull Proportional Odds Copula Model

Changing only the regression model class, the survival copula likelihood function for a model with Weibull PO marginal components, with parameters $\boldsymbol{\kappa}_j$ and $\boldsymbol{\beta}_j$, is given by

$$L_{\overline{C}(\text{WPO})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) = \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \frac{1}{\left(1 + \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)} \right] \right\} \times \prod_{j=1}^{d} \left\{ \frac{\lambda_j \alpha_j y_{i;j}^{\alpha_j - 1} \exp\left(\lambda_j y_{i;j}^{\alpha_j} + \boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)}{\left[1 + \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right]^2} \right\}^{\delta_{i;j}}. \tag{4.2}$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.2) can also be expressed as

$$L_{\overline{C}(\text{WPO})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\dfrac{\lambda_j \alpha_j y_{i;j}^{\alpha_j - 1} \exp\left(\lambda_j y_{i;j}^{\alpha_j} + \boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)}{\left[1 + \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right]^2}}{\psi_\theta^{(1)}\left(\psi_\theta^{-1}\left(\dfrac{1}{\left(1 + \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)}\right)\right)} \right]^{\delta_{i;j}} \right\} \times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\frac{1}{\left(1 + \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)} \right] \right\}.$$

4.1.3 Weibull Yang-Prentice Copula Model

Changing again the regression model class, the survival copula likelihood function for a model with Weibull YP marginal components, with parameters $\boldsymbol{\kappa}_j$, $\boldsymbol{\beta}_j^{(S)}$ and $\boldsymbol{\beta}_j^{(L)}$, is given by

$$
\begin{aligned}
&L_{\overline{C}(\mathrm{WYP})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}^{(S)}, \boldsymbol{\beta}^{(L)} | \boldsymbol{y}, X) = \\
&= \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} \left(y_{i;j}\right)^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \left(1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \right)^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right] \right\} \\
&\times \prod_{j=1}^{d} \left\{ \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \lambda_j \alpha_j y_{i;j}^{\alpha_j - 1}}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \left[1 - \exp\left(-\lambda_j y_{i;j}^{\alpha_j}\right)\right] + \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \exp\left(-\lambda_j y_{i;j}^{\alpha_j}\right)} \right\}^{\delta_{i;j}} \\
&\times \prod_{j=1}^{d} \left\{ \left[1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \right]^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right\} .
\end{aligned}
\tag{4.3}
$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.3) can also be expressed as

$$
\begin{aligned}
&L_{\overline{C}(\mathrm{WYP})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}^{(S)}, \boldsymbol{\beta}^{(L)} | \boldsymbol{y}, X) = \\
&= \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\dfrac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \lambda_j \alpha_j y_{i;j}^{\alpha_j - 1}}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \left[1 - \exp\left(-\lambda_j y_{i;j}^{\alpha_j}\right)\right] + \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \exp\left(-\lambda_j y_{i;j}^{\alpha_j}\right)}}{\left(1 + \dfrac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \right)^{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)}} \right]^{\delta_{i;j}} \right. \\
&\qquad \left. \times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\left(1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\lambda_j y_{i;j}^{\alpha_j}\right) - 1\right) \right)^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right] \right\} \right\} .
\end{aligned}
$$

4.2 Survival Copula Models with Marginal BP Baseline

4.2.1 Bernstein Polynomials Proportional Hazards Copula Model

Switching now the baseline, the survival copula likelihood function for a model with BP PH margins and parameters $\boldsymbol{\kappa}_j = \boldsymbol{\gamma}_j$ and $\boldsymbol{\beta}_j$, $j = 1, \ldots, d$, over all subjects i, $i = 1, \ldots, n$ and fixed degree m, is given by

$$L_{\overline{C}(\text{BPPH})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) = \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \exp\left(-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right) \right] \right\}$$
$$\times \prod_{j=1}^{d} \left\{ \boldsymbol{\gamma}_j' \boldsymbol{g}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right) \exp\left[-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right] \right\}^{\delta_{i;j}}. \tag{4.4}$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.4) can be rewritten as

$$L_{\overline{C}(\text{BPPH})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\boldsymbol{\gamma}_j' \boldsymbol{g}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right) \exp\left(-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)}{\psi_\theta^{(1)}\left(\psi_\theta^{-1}\left(\exp\left(-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)\right)\right)} \right]^{\delta_{i;j}} \right\}$$
$$\times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\exp\left(-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right) \right] \right\}.$$

4.2.2 Bernstein Polynomials Proportional Odds Copula Model

Again changing only the regression model class, the survival copula likelihood function for a model with BP PO marginal components, with parameters $\boldsymbol{\kappa}_j = \boldsymbol{\zeta}_j$ and $\boldsymbol{\beta}_j$, is given by

$$L_{\overline{C}(\text{BPPO})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) = \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \frac{1}{\left(1 + \boldsymbol{\zeta}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)} \right] \right\}$$
$$\times \prod_{j=1}^{d} \left\{ \frac{\boldsymbol{\zeta}_j' \boldsymbol{g}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)}{\left[1 + \boldsymbol{\zeta}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right]^2} \right\}^{\delta_{i;j}}. \tag{4.5}$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.5) can also be expressed as

$$L_{\overline{C}(\text{BPPO})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\dfrac{\boldsymbol{\zeta}_j' \boldsymbol{g}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)}{\left[1 + \boldsymbol{\zeta}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right]^2}}{\psi_\theta^{(1)}\left(\psi_\theta^{-1}\left(\dfrac{1}{\left(1 + \boldsymbol{\zeta}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)}\right)\right)} \right]^{\delta_{i;j}} \right\}$$
$$\times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\frac{1}{\left(1 + \boldsymbol{\zeta}_j' \boldsymbol{G}_m\left(y_{i;j}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)} \right] \right\}.$$

4.2.3 Bernstein Polynomials Yang-Prentice Copula Model

Switching the regression structure to the YP model, the survival copula likelihood function for a model with BP YP margins and parameters $\boldsymbol{\kappa}_j = \boldsymbol{\gamma}_j$, $\boldsymbol{\beta}_j^{(S)}$ and $\boldsymbol{\beta}_j^{(L)}$, is given by

$$
\begin{aligned}
&L_{\overline{C}(\text{BPYP})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}^{(S)}, \boldsymbol{\beta}^{(L)} | \boldsymbol{y}, X) = \\
&= \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \left(1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right) - 1\right) \right)^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right] \right\} \\
&\times \prod_{j=1}^{d} \left\{ \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \boldsymbol{\gamma}_j' \boldsymbol{g}_m\left(y_{i;j}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \left[1 - \exp\left(-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right)\right] + \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \exp\left(-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right)} \right\}^{\delta_{i;j}} \\
&\times \prod_{j=1}^{d} \left\{ \left[1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right) - 1\right) \right]^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right\}^{\delta_{i;j}} .
\end{aligned}
\tag{4.6}
$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.6) can also be expressed as

$$
\begin{aligned}
&L_{\overline{C}(\text{BPYP})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}^{(S)}, \boldsymbol{\beta}^{(L)} | \boldsymbol{y}, X) = \\
&= \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\dfrac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \boldsymbol{\gamma}_j' \boldsymbol{g}_m\left(y_{i;j}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \left[1 - \exp\left(-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right)\right] + \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \exp\left(-\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right)} \left(1 + \dfrac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right) - 1\right) \right)^{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)}}{\psi_\theta^{(1)}\left(\psi_\theta^{-1}\left(\left(1 + \dfrac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right) - 1\right) \right)^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right) \right)} \right]^{\delta_{i;j}} \right\} \\
&\times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\left(1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\boldsymbol{\gamma}_j' \boldsymbol{G}_m\left(y_{i;j}\right)\right) - 1\right) \right)^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right] \right\} .
\end{aligned}
$$

4.3 Survival Copula Models with Marginal PE Baseline

4.3.1 Piecewise Exponential Proportional Hazards Copula Model

Switching again the baseline distribution, the survival copula likelihood function for a model with PE PH margins and parameters $\boldsymbol{\kappa}_j = \boldsymbol{\lambda}_j$ and $\boldsymbol{\beta}_j$, $j = 1, \ldots, d$, over all subjects i, $i = 1, \ldots, n$, and fixing a $(d, p+1)$

time grid E with p intervals $E_{j;k} = (e_{j;k-1}, e_{j;k}]$ for each copula margin j, $k = 1, \ldots, p$, is given by

$$L_{\overline{C}(\text{PEPH})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X, E) = \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \exp\left(-\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right) \right] \right\}$$
$$\times \prod_{j=1}^{d} \left\{ \sum_{k=1}^{p} \lambda_{j;k} \mathbb{I}_{E_{j;k}}(y_{i;j}) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right) \exp\left[-\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right] \right\}^{\delta_{i;j}}. \quad (4.7)$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.7) can be rewritten as

$$L_{\overline{C}(\text{PEPH})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X, E) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\sum_{k=1}^{p} \lambda_{j;k} \mathbb{I}_{E_{j;k}}(y_{i;j}) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)}{\exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right) \psi_\theta^{(1)}\left(\psi_\theta^{-1}\left(\exp\left(-\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)\right)\right)} \right]^{\delta_{i;j}} \right\}$$
$$\times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\exp\left(-\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right) \right] \right\}.$$

4.3.2 Piecewise Exponential Proportional Odds Copula Model

Changing now the regression model class, the survival copula likelihood function for a model with PE PO marginal components, with parameters $\boldsymbol{\kappa}_j$ and $\boldsymbol{\beta}_j$, is given by

$$L_{\overline{C}(\text{PEPO})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X, E) = \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \frac{1}{\left(1 + \left(\exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)} \right] \right\}$$
$$\times \prod_{j=1}^{d} \left\{ \frac{\sum_{k=1}^{p} \lambda_{j;k} \mathbb{I}_{E_{j;k}}(y_{i;j}) \exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right) + \boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)}{\left[1 + \left(\exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right]^2} \right\}^{\delta_{i;j}}. \quad (4.8)$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.8) can also be expressed as

$$L_{\overline{C}(\text{PEPO})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}|\boldsymbol{y}, X, E) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\sum_{k=1}^{p} \lambda_{j;k} \mathbb{I}_{E_{j;k}}(y_{i;j}) \exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right) + \boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)}{\left[1 + \left(\exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right]^2} \right. \right.$$
$$\left. \left. \frac{}{\psi_\theta^{(1)}\left(\psi_\theta^{-1}\left(\frac{1}{\left(1 + \left(\exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)}\right)\right)} \right]^{\delta_{i;j}} \right\}$$
$$\times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\frac{1}{\left(1 + \left(\exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right) - 1\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j\right)\right)} \right] \right\}.$$

4.3.3 Piecewise Exponential Yang-Prentice Copula Model

Changing again the regression structure to the YP model, the survival copula likelihood function for a model with PE YP margins and parameters $\boldsymbol{\kappa}_j$, $\boldsymbol{\beta}_j^{(S)}$ and $\boldsymbol{\beta}_j^{(L)}$, is given by

$$
\begin{aligned}
&L_{\overline{C}(\text{PEYP})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}^{(S)}, \boldsymbol{\beta}^{(L)} | \boldsymbol{y}, X, E) = \\
&= \prod_{i=1}^{n} \frac{\partial^{d_i}}{\partial \prod_{j=1}^{d} (y_{i;j})^{\delta_{i;j}}} \psi_\theta \left\{ \psi_\theta^{-1} \left[\sum_{j=1}^{d} \left(1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right) - 1 \right) \right)^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right] \right\} \\
&\times \prod_{j=1}^{d} \left\{ \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \sum_{k=1}^{p} \lambda_{j;k} \mathbb{I}_{E_{j;k}}(y_{i;j})}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \left[1 - \exp\left(-\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right)\right] + \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \exp\left(-\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right)} \right\}^{\delta_{i;j}} \\
&\times \prod_{j=1}^{d} \left\{ \left[1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(\boldsymbol{\lambda}_j' \left(\boldsymbol{t}_j - \boldsymbol{e}_j\right)\right) - 1 \right) \right]^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right\}^{\delta_{i;j}} .
\end{aligned} \tag{4.9}
$$

If $\psi_\theta = \mathcal{LS}_{V|\theta}$ is a completely monotone generator, then (4.9) can also be expressed as

$$
\begin{aligned}
&L_{\overline{C}(\text{PEYP})}(\theta, \boldsymbol{\kappa}, \boldsymbol{\beta}^{(S)}, \boldsymbol{\beta}^{(L)} | \boldsymbol{y}, X, E) = \\
&= \prod_{i=1}^{n} \left\{ \prod_{j=1}^{d} \left[\frac{\dfrac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \sum_{k=1}^{p} \lambda_{j;k} \mathbb{I}_{E_{j;k}}(y_{i;j})}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right) \left[1 - \exp\left(-H_0\left(y_{i;j}\right)\right)\right] + \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right) \exp\left(-H_0\left(y_{i;j}\right)\right)} \left(1 + \dfrac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(H_0\left(y_{i;j}\right)\right) - 1 \right) \right)^{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} }{\psi_\theta^{(1)} \left(\psi_\theta^{-1} \left(\left(1 + \dfrac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(H_0\left(y_{i;j}\right)\right) - 1 \right) \right)^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right) \right)} \right]^{\delta_{i;j}} \right\} \\
&\times \psi_\theta^{(d_i)} \left\{ \sum_{j=1}^{d} \psi_\theta^{-1} \left[\left(1 + \frac{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)}{\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \left(\exp\left(H_0\left(y_{i;j}\right)\right) - 1 \right) \right)^{-\exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)} \right] \right\} .
\end{aligned}
$$

CHAPTER 5

Numerical Results

This chapter presents Monte Carlo simulation studies to evaluate the performance of bivariate Archimedean survival copula models taking into account $M = 1000$ replications of data sets. To generate the simulated data, the R `copula` package (Hofert et al., 2020) was used to obtain each marginal uniform realizations $u_{i;j}$, $i = 1, \ldots, n$ and $j = 1, 2$, for the five Archimedean copulas addressed in this work (AMH, Clayton, Frank, GH and Joe), according to the stochastic representation in (3.10). For each margin, the same design matrix was specified, with two covariates $X_{i,1} \sim \text{Bern}(0.5)$ and $X_{i,2} \sim \text{N}(0, 1)$ generated independently and identically distributed for all i. Assuming $u_{i;j} = S(t_{i;j})$, where $S(\cdot)$ is a survival function with inverse $S^{-1}(\cdot)$, each marginal survival time was generated as $t_{i;j} = S^{-1}(u_{i;j})$. Two baseline distributions were considered in the data generation process: the traditional Weibull model, and the Exponentiated Weibull (EW) model proposed by Mudholkar and Srivastava (1993). The EW distribution is a parametric model widely used in survival analysis due to its high flexibility for the hazard function behavior, allowing both unimodal and bathtub forms, which are not accommodated by the traditional Weibull model. The EW model possesses a vector of parameters $\boldsymbol{\kappa} = (\alpha, \lambda, \xi)$, where α and λ are the shape and scale parameters from the Weibull model, and ξ is the exponentiation parameter. If $\xi = 1$, then the Weibull model is obtained. Each baseline distribution used for generation of marginal survival times was combined with one of the three regression model classes (PH, PO, or YP) presented in Section 2.2. Thereby, each marginal survival time $t_{i;j}$ was generated from the following expressions in (5.1) for the YP class, obtaining the PH and PO as particular cases by taking $\boldsymbol{\beta}_j^{(S)} = \boldsymbol{\beta}_j^{(L)}$ or $\boldsymbol{\beta}_j^{(L)} = \mathbf{0}$ (respectively), and depending on the chosen baseline distribution (Weibull or EW).

$$t_{i;j} = \begin{cases} \left\{ \log\left[1 - q\left(u_{i;j}, \boldsymbol{x}_{i;j}, \boldsymbol{\beta}_j^{(S)}, \boldsymbol{\beta}_j^{(L)}\right)\right] \lambda_j^{-1} \right\}^{\left(\alpha_j^{-1}\right)}; & \text{(Weibull YP)} \\ \left\{ -\log\left[1 - \left(1 - \left(1 - q\left(u_{i;j}, \boldsymbol{x}_{i;j}, \boldsymbol{\beta}_j^{(S)}, \boldsymbol{\beta}_j^{(L)}\right)\right)^{-1}\right)^{\xi_j^{-1}}\right] \lambda_j^{-1} \right\}^{\left(\alpha_j^{-1}\right)}, & \text{(EW YP)} \end{cases} \tag{5.1}$$

with

$$q\left(u_{i;j}, \boldsymbol{x}_{i;j}, \boldsymbol{\beta}_j^{(S)}, \boldsymbol{\beta}_j^{(L)}\right) = \exp\left(\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)} - \boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(S)}\right)\left(1 - u_{i;j}^{-\exp\left(-\boldsymbol{x}_{i;j}\boldsymbol{\beta}_j^{(L)}\right)}\right).$$

Given a survival copula model, the baseline distribution and the associated regression model class are generated from the same family for both margins. The simulation scenarios are primarily defined by: (i) a fixed and same sample size ($n = 500$) for all margins, and (ii) three Kendall's τ correlation values ($\tau \in \{0.25, 0.5, 0.75\}$). In other words, there are three scenarios for study with the same sample size: S1, with $\tau = 0.25$, S2, with $\tau = 0.50$, and S3, with $\tau = 0.75$. These true values for τ were chosen in order to obtain results given different levels of dependence, and to retain comparability among all fitted copulas: for a same value of τ, θ can be quite different from one copula to another. The Kendall's τ choice as the measure of association for this work is justified by its easier computation if compared to Spearman's ρ for all copulas (as showed in 3.1), and the recurrent use of Kendall's τ in the literature, as seen in Prenen et al. (2017).

Although the baseline distribution and regression model class are from the same family for both margins, the chosen values for $\boldsymbol{\kappa}_j$ and $\boldsymbol{\beta}_j$ (for PH or PO), or $\boldsymbol{\kappa}_j$, $\boldsymbol{\beta}_j^{(S)}$, and $\boldsymbol{\beta}_j^{(L)}$ (for YP), $j = 1, 2$, are different. Regardless of copula or regression model class, for the Weibull model as the baseline generator, $\boldsymbol{\kappa}_1 = (\alpha_1, \lambda_1) = (1.2, 0.8)$ and $\boldsymbol{\kappa}_2 = (\alpha_2, \lambda_2) = (1.6, 1.2)$. Both parameter specifications yield increasing marginal hazard functions. If the generator process for the baseline is the EW model, $\boldsymbol{\kappa}_1 = (\alpha_1, \lambda_1, \xi_1) = (2.1, 0.5, 0.3)$, and $\boldsymbol{\kappa}_2 = (\alpha_2, \lambda_2, \xi_2) = (2.5, 0.6, 0.2)$. Those parameter specifications produce "bathtub"-shaped marginal hazard functions. Concerning the regression model classes, regardless of copula or baseline distribution, for both PH and PO models $\boldsymbol{\beta}_1 = (-0.7, 0.4)$ and $\boldsymbol{\beta}_2 = (-0.9, 0.6)$. For the YP model, $\boldsymbol{\beta}_1^{(S)} = (-0.7, 0.4)$, $\boldsymbol{\beta}_2^{(S)} = (-0.9, 0.6)$, $\boldsymbol{\beta}_1^{(L)} = (0.8, -0.6)$, and $\boldsymbol{\beta}_2^{(L)} = (1.0, -0.8)$. To introduce censoring, all generated times $t_{i;j}$, $i = 1, \ldots, n$ and $j = 1, 2$, were compared to a corresponding threshold value $a_{i;j}$ sampled from a continuous uniform distribution $\mathrm{U}(0, a_j)$. If $t_{i;j} \leqslant a_{i;j}$, then $y_{i;j} = t_{i;j}$ is a failure time. Otherwise, $y_{i;j} = a_{i;j}$ is a censored time. The values a_j were chosen in order to achieve a failure rate between 65% and 85% for each margin, given their corresponding set of parameters values, for all scenarios. The threshold choices were $a_1 = 6$, $a_2 = 4$ when the generated baseline is a Weibull model, and $a_1 = 4$, $a_2 = 3$ for a EW model.

The simulation study has three main goals: (i) to compare results for different survival copula models, given the same baseline function and regression model class, when fitting all five Archimedean copulas to a set of data generated from a given copula, (ii) given the correct fitting for the copula and regression structure, to compare results for different fitted baseline functions, and (iii) for a same combination of copula and baseline, evaluate the fitting of nested regression model classes when generating from a given one. Goal (i) is achieved by presenting results for regression parameter estimation and information criteria, while goal (ii) is reached by presenting also results for Kendall's τ estimation. To achieve goal (iii), an analysis through the Likelihood Ratio (LR) test is done, confronting pairs of nested models for each regression model class used for generation (supposed unknown for each test). Since fitted baseline functions from Weibull, BP and PE families have distinct specifications for $\boldsymbol{\kappa}_j$, which do not allow to compare them directly, their parameter estimation will not be addressed here. To evaluate MC estimates for the regression parameter set $\boldsymbol{\beta}_j$ (for PH and PO), or $\boldsymbol{\beta}_j^{(S)}, \boldsymbol{\beta}_j^{(L)}$ (for YP), and the Kendall's τ correlation as function of the original dependence parameter θ, the following statistics were computed over simulation results for all M data sets. If ν_l, $l = 1, \ldots, M$, is a parameter

of interest for inference with maximum likelihood estimate $\widehat{\nu}_l$ and associated standard error estimate $\text{se}(\widehat{\nu}_l)$ obtained from the hessian matrix, it is possible to compute:

- The Average Estimate (AE) of ν, given by the mean of point estimates $\widehat{\nu}_l$:

$$\text{AE}(\nu) = \frac{1}{M}\sum_{l=1}^{M}\widehat{\nu}_l; \tag{5.2}$$

- The Standard Deviation Estimate (SDE) of ν, given by the standard deviation of point estimates $\widehat{\nu}_l$:

$$\text{SDE}(\nu) = \left\{\frac{1}{M-1}\sum_{l=1}^{M}\left[\widehat{\nu}_l - AE\left(\nu\right)\right]^2\right\}^{1/2}; \tag{5.3}$$

- The Average Standard Error (ASE) of ν, given by the mean of standard error estimates $\text{se}(\widehat{\nu}_l)$:

$$\text{ASE}(\nu) = \frac{1}{M}\sum_{l=1}^{M}\text{se}\left(\widehat{\nu}_l\right); \tag{5.4}$$

- The Average Relative Bias (ARB) of ν, generally expressed by a percentage, given by the mean of relative biases computed over all estimates $\widehat{\nu}_l$, with respect to the true value ν:

$$\text{ARB}(\nu) = 100 \times \frac{1}{M}\sum_{l=1}^{M}\frac{(\widehat{\nu}_l - \nu)}{|\nu|} \tag{5.5}$$

- The Coverage Rate (CR) of ν, *i.e.*, the proportion of M data sets that provides a interval with a pre-specified confidence level (95%) that contains the true value ν.

On a well-fitted model, the AE is expected to be near the true parameter value, the SDE and ASE are expected to have values close to each other, the ARB is expected to be around 0, and the CR is expected to be near the pre-specified confidence level. If the ARB is close to 0, underestimating the true standard deviation of a given parameter (ASE $<$ SDE) leads to a CR lower than the confidence level. When overestimating the true standard deviation (ASE $>$ SDE), the CR is greater than the confidence level. If the data is marginally generated from a Weibull baseline function, under the correct fitting of both copula and regression structure, fitted Weibull models are expected to have great performance. Also, semiparametric fitted models (BP and PE) are expected to perform well due to its nonparametric appeal. However, for marginal generation from an EW baseline function, it is expected that semiparametric models perform better than the Weibull fitting due to their greater flexibility to capture non-monotonic behaviors potentially present in both hazard functions.

Another way to compare fitted models is computing an information criteria based on the log-likelihood $\ell = \ell(\boldsymbol{\nu})$ for a set of n observed values, given the p-dimensional vector $\widehat{\boldsymbol{\nu}}$ of estimates. A model is preferable if it has the lowest criteria value. For a frequentist approach, a method frequently used is the Akaike Information Criteria (AIC, Akaike, 1974). Comparisons for fitted models in this work will be done through the mean AIC and, given a copula model used for generation, using also the proportion of choice for each of the five

fitted copula models. Furthermore, given any pair of fitted survival copula models nested with respect to the regression model class, an analysis through the LR test is done to conclude if the augmented model (the one with more regression parameters) is significant. It is expected that fitted PH (or PO) models will be chosen due to their parsimony over YP model for PH (PO) marginally generated data, but also that fitted YP models perform significantly better than fitted PH or PO models for YP marginally generated data.

Defined all statistics, criteria and tests to be evaluated, simulation results are produced for all scenarios. Due to the high number of tables and figures, along with the detection of similar patterns, in the main text these results are limited to scenario S1 ($\tau = 0.25$), while the others can be accessed through the link `https://wrmfstat.shinyapps.io/CopRegEst/` (for regression parameter estimation) or seen in the Appendix, which also presents some instructions on how to visualize them online. For the regression parameter estimates presented here, results are restricted to the 1st copula margin and given generated data from the AMH model, since Kendall's τ for that copula varies in the interval $[(5 - 8\log(2))/3, 1/3] \approx [-0.1817, 0.3333]$ (the only one that does not cover the open unit interval). This way, AMH generated data used a value of τ truncated to the upper limit when necessary. The idea is to show that this change has little effect on regression parameter estimates from different fitted copula models, since the true τ value is "unknown". Thus, results for AMH generated data are comparable to those of other Archimedean copulas used for generation, which in general appoints to similar conclusions. Concerning the amount of time to fit all MC replicas, given a scenario, the BP model was about 5.1 times slower than the PE model, and about 8.2 times than the fitted Weibull model. This is justified by BP models presenting hessian matrices near singularity. In its turn, all five Archimedean copulas had similar amounts of fitting time, with AMH (the fastest) spending about 84% the time of GH (the slowest). Finally, the YP class fitting was about 1.7 times slower than the PH and PO ones.

Results for fitted survival copula models over simulated data are presented in the following sections, divided by the marginally generated baseline distribution (Weibull or EW). Each section contains MC estimates for regression parameters (including the average lower and upper bounds – ALB and AUB – for their estimated intervals), AIC for the fitted models, MC estimates for Kendall's τ correlation, and LR tests for nested regression model classes. Finally, an additional simulation study is done over a specific scenario ($\tau = 0.25$), and only for copula generated data with Weibull YP margins, to estimate marginal crossing survival times given a combination of fitted Archimedean copula (one of the five discussed in this work, and always fitting the correct copula), baseline distribution (one of Weibull, BP or PE), and the YP model class. To handle the solving of nonlinear equations involving survival functions of treated and control subjects in R, the command `uniroot` is used to find the corresponding root for each marginal crossing survival time (R Core Team, 2021).

5.1 Generated Copulas with Weibull Baseline

This section presents results for fitted survival copula models over generated data from Archimedean survival copulas with marginal Weibull baseline distribution, associated to a regression model class. Results are divided in the following subsections by Monte Carlo estimates for regression parameters, AIC for the fitted models, Monte Carlo estimates for Kendall's τ correlation, and LR tests for nested fitted models.

5.1.1 Regression Parameter Estimates

The Monte Carlo estimates on regression parameters for fitted survival copula models are showed from Tables 5.1 to 5.4, divided by fitted baseline distribution for each regression parameter set ($\boldsymbol{\beta}_j$, $\boldsymbol{\beta}_j^{(S)}$ or $\boldsymbol{\beta}_j^{(L)}$) from a given class, on the 1st copula margin ($j = 1$). For those results, comparisons are done among fitted models with different copulas, but always maintaining the same regression model class used for generation.

Table 5.1: MC statistics for 1st margin regression parameter estimates of fitted survival copula models over AMH Weibull PH generated data ($n = 500$; $\tau = 0.25$).

Parameter	Copula	Weibull PH Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7069	0.1048	0.1041	-0.9840	-0.9123	-0.5015	95.1807
	Clayton	-0.6970	0.1031	0.1052	0.4310	-0.8991	-0.4949	94.6000
	Frank	-0.7086	0.1059	0.1061	-1.2306	-0.9162	-0.5010	94.3888
	GH	-0.7027	0.1065	0.1069	-0.3916	-0.9114	-0.4941	93.7876
	Joe	-0.7020	0.1070	0.1080	-0.2873	-0.9116	-0.4924	93.3000
$\beta_{12} = 0.4$	AMH	0.4016	0.0536	0.0545	0.4000	0.2966	0.5066	94.1767
	Clayton	0.3964	0.0527	0.0552	-0.9103	0.2931	0.4997	93.1000
	Frank	0.4018	0.0541	0.0554	0.4469	0.2957	0.5079	93.7876
	GH	0.3985	0.0543	0.0558	-0.3736	0.2920	0.5050	93.0862
	Joe	0.3979	0.0546	0.0564	-0.5152	0.2909	0.5050	93.1000
Parameter	Copula	BP PH Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7089	0.1061	0.1055	-1.2776	-0.9169	-0.5010	94.9000
	Clayton	-0.6961	0.1043	0.1062	0.5600	-0.9005	-0.4916	94.8000
	Frank	-0.7129	0.1072	0.1078	-1.8472	-0.9230	-0.5029	94.0941
	GH	-0.7138	0.1075	0.1092	-1.9711	-0.9245	-0.5031	93.9940
	Joe	-0.7107	0.1079	0.1104	-1.5238	-0.9222	-0.4992	93.9880
$\beta_{12} = 0.4$	AMH	0.4026	0.0542	0.0553	0.6474	0.2963	0.5088	93.8000
	Clayton	0.3960	0.0533	0.0560	-0.9963	0.2915	0.5005	92.9000
	Frank	0.4044	0.0548	0.0563	1.0925	0.2970	0.5117	93.7938
	GH	0.4047	0.0549	0.0570	1.1748	0.2972	0.5122	92.6927
	Joe	0.4029	0.0551	0.0576	0.7237	0.2950	0.5108	92.7856
Parameter	Copula	PE PH Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7046	0.1059	0.1052	-0.6551	-0.9121	-0.4971	94.9648
	Clayton	-0.6919	0.1041	0.1056	1.1616	-0.8959	-0.4879	95.0000
	Frank	-0.7088	0.1069	0.1072	-1.2510	-0.9183	-0.4992	93.9819
	GH	-0.7099	0.1071	0.1083	-1.4116	-0.9197	-0.5001	94.1650
	Joe	-0.7064	0.1073	0.1095	-0.9206	-0.9168	-0.4961	93.9759
$\beta_{12} = 0.4$	AMH	0.4000	0.0541	0.0548	-0.0094	0.2939	0.5060	94.3605
	Clayton	0.3934	0.0532	0.0556	-1.6573	0.2891	0.4976	93.4000
	Frank	0.4022	0.0547	0.0558	0.5506	0.2951	0.5093	93.6810
	GH	0.4025	0.0546	0.0564	0.6236	0.2955	0.5095	93.3602
	Joe	0.4006	0.0547	0.0571	0.1481	0.2933	0.5079	93.0723

Taking the results in Table 5.1, when correctly fitting the PH class under a low correlation, the ARB is always lower (in magnitude) than 2%, and the CR is at most 0.03 away from the confidence level (set as 95%) for all regression parameters, even when fitting the wrong copula. Correctly fitted (AMH) copula models had,

in general, smaller ARB values and closer CR values to the confidence level. As expected, fitted semiparametric models (BP and PE) perform similar to (correctly) fitted Weibull models, but without imposing any parametric restriction for the (marginal) hazard rate function to obtain good regression parameter estimates.

Table 5.2: MC statistics for 1st margin regression parameter estimates of fitted survival copula models over AMH Weibull PO generated data ($n = 500$; $\tau = 0.25$).

Parameter	Copula	Weibull PO Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7074	0.1627	0.1596	-1.0597	-1.0263	-0.3885	95.4955
	Clayton	-0.7050	0.1621	0.1604	-0.7157	-1.0227	-0.3874	95.2906
	Frank	-0.7039	0.1637	0.1627	-0.5541	-1.0247	-0.3831	95.2953
	GH	-0.6988	0.1640	0.1663	0.1678	-1.0203	-0.3774	94.9900
	Joe	-0.6978	0.1643	0.1676	0.3118	-1.0199	-0.3758	94.7844
$\beta_{12} = 0.4$	AMH	0.3997	0.0819	0.0850	-0.0830	0.2392	0.5602	94.0941
	Clayton	0.3980	0.0816	0.0850	-0.4963	0.2382	0.5579	94.2886
	Frank	0.3968	0.0824	0.0862	-0.7885	0.2354	0.5583	94.3944
	GH	0.3936	0.0825	0.0881	-1.6097	0.2318	0.5554	93.1864
	Joe	0.3930	0.0827	0.0889	-1.7474	0.2309	0.5551	92.9789
Parameter	Copula	BP PO Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7071	0.1633	0.1604	-1.0134	-1.0272	-0.3870	95.1904
	Clayton	-0.7082	0.1629	0.1619	-1.1705	-1.0274	-0.3890	95.0902
	Frank	-0.7036	0.1643	0.1633	-0.5193	-1.0256	-0.3816	95.1000
	GH	-0.7011	0.1642	0.1669	-0.1517	-1.0229	-0.3792	94.8000
	Joe	-0.7009	0.1642	0.1695	-0.1339	-1.0227	-0.3792	94.2886
$\beta_{12} = 0.4$	AMH	0.3992	0.0822	0.0854	-0.2035	0.2380	0.5604	94.4890
	Clayton	0.4000	0.0820	0.0858	-0.0075	0.2393	0.5607	94.1884
	Frank	0.3967	0.0827	0.0863	-0.8250	0.2346	0.5588	94.5000
	GH	0.3954	0.0826	0.0879	-1.1611	0.2335	0.5572	93.2000
	Joe	0.3951	0.0826	0.0892	-1.2283	0.2332	0.5570	92.3848
Parameter	Copula	PE PO Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7019	0.1631	0.1589	-0.2761	-1.0215	-0.3823	95.1952
	Clayton	-0.7018	0.1625	0.1605	-0.2637	-1.0204	-0.3833	94.9648
	Frank	-0.6986	0.1640	0.1620	0.2017	-1.0201	-0.3771	95.4683
	GH	-0.6972	0.1635	0.1645	0.3991	-1.0176	-0.3768	94.8640
	Joe	-0.6961	0.1632	0.1678	0.5527	-1.0159	-0.3764	94.3548
$\beta_{12} = 0.4$	AMH	0.3959	0.0821	0.0847	-1.0276	0.2350	0.5568	94.4945
	Clayton	0.3965	0.0818	0.0851	-0.8827	0.2362	0.5567	94.2598
	Frank	0.3942	0.0826	0.0858	-1.4482	0.2324	0.5561	94.3605
	GH	0.3936	0.0822	0.0868	-1.5970	0.2326	0.5547	93.4542
	Joe	0.3926	0.0820	0.0881	-1.8446	0.2319	0.5534	92.7419

Looking now on Table 5.2, but now correctly fitting the PO regression structure, conclusions for ARB and CR values are similar to those obtained from Table 5.1. For results in Table 5.3, when correctly fitting the YP class, the ARB for short-term parameters is a bit larger than the ones in Tables 5.1 and 5.2, but always lower than 6% even when fitting the wrong copula, and lower than 3% when fitting the correct one. This was expected, since the YP class specifies more regression parameters for the same covariates. On the other hand, the CR is as close as the obtained for PH and PO models to the confidence level. When comparing the fitted

baseline distributions, BP models had in general lowest ARB and closer CR values to the confidence level.

Table 5.3: MC statistics for 1st margin short-term regression parameter estimates of fitted survival copula models over AMH Weibull YP generated data ($n = 500$; $\tau = 0.25$).

Parameter	Copula	Weibull YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(S)} = -0.7$	AMH	-0.7077	0.1542	0.1526	-1.0935	-1.0099	-0.4054	96.3783
	Clayton	-0.6939	0.1544	0.1545	0.8667	-0.9965	-0.3914	96.0765
	Frank	-0.7137	0.1543	0.1542	-1.9526	-1.0161	-0.4112	96.4895
	GH	-0.7387	0.1537	0.1550	-5.5346	-1.0400	-0.4375	94.8949
	Joe	-0.7281	0.1543	0.1565	-4.0175	-1.0306	-0.4256	94.7791
$\beta_{12}^{(S)} = 0.4$	AMH	0.3973	0.0917	0.0965	-0.6857	0.2176	0.5770	94.0644
	Clayton	0.3938	0.0911	0.0958	-1.5552	0.2152	0.5724	94.1650
	Frank	0.3946	0.0921	0.0978	-1.3604	0.2141	0.5750	93.9819
	GH	0.3900	0.0917	0.0996	-2.5020	0.2102	0.5698	93.5936
	Joe	0.3891	0.0920	0.1008	-2.7225	0.2088	0.5694	93.0723
Parameter	Copula	BP YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(S)} = -0.7$	AMH	-0.7047	0.1610	0.1609	-0.6751	-1.0202	-0.3893	94.6894
	Clayton	-0.7031	0.1612	0.1619	-0.4359	-1.0189	-0.3872	95.1000
	Frank	-0.7006	0.1616	0.1632	-0.0923	-1.0175	-0.3838	94.9000
	GH	-0.7000	0.1618	0.1656	0.0033	-1.0170	-0.3830	95.0853
	Joe	-0.6980	0.1619	0.1673	0.2905	-1.0152	-0.3807	94.5892
$\beta_{12}^{(S)} = 0.4$	AMH	0.3983	0.0920	0.0969	-0.4133	0.2180	0.5787	94.0882
	Clayton	0.3965	0.0919	0.0968	-0.8667	0.2165	0.5766	93.6000
	Frank	0.3956	0.0923	0.0976	-1.0976	0.2147	0.5765	94.4000
	GH	0.3929	0.0916	0.0980	-1.7673	0.2134	0.5724	93.6810
	Joe	0.3920	0.0918	0.1001	-1.9878	0.2120	0.5721	93.0862
Parameter	Copula	PE YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(S)} = -0.7$	AMH	-0.7142	0.1604	0.1594	-2.0219	-1.0285	-0.3998	95.0902
	Clayton	-0.7110	0.1604	0.1604	-1.5749	-1.0254	-0.3966	95.1807
	Frank	-0.7121	0.1610	0.1615	-1.7298	-1.0277	-0.3965	95.0853
	GH	-0.7101	0.1607	0.1640	-1.4432	-1.0251	-0.3951	95.0853
	Joe	-0.7071	0.1606	0.1652	-1.0206	-1.0219	-0.3924	94.8847
$\beta_{12}^{(S)} = 0.4$	AMH	0.3899	0.0914	0.0951	-2.5355	0.2107	0.5690	94.4890
	Clayton	0.3882	0.0912	0.0951	-2.9483	0.2095	0.5670	94.0763
	Frank	0.3876	0.0917	0.0961	-3.0900	0.2080	0.5673	94.5838
	GH	0.3878	0.0908	0.0969	-3.0581	0.2097	0.5658	93.8816
	Joe	0.3869	0.0909	0.0985	-3.2749	0.2087	0.5651	92.9789

Again for correctly fitted YP models, but now looking results on long-term parameters in Table 5.4, the ARB is always below 10% when fitting the correct copula, being higher for the first (dichotomous) covariate than for the second (continuous) one, regardless of the fitted baseline distribution. The CR values, however, are still near the confidence level, at most 0.03 away from it. The SDE and ASE estimates are also greater than those obtained for short-term parameters. These evidences implies more accurate estimates for short-term parameters than for their long-term counterpart. This is not a surprise: it is harder to estimate $\boldsymbol{\beta}_j^{(L)}$ than $\boldsymbol{\beta}_j^{(S)}$ since there are fewer subjects under risk the closer a follow-up is to its end.

Table 5.4: MC statistics for 1st margin long-term regression parameter estimates of fitted survival copula models over AMH Weibull YP generated data ($n = 500$; $\tau = 0.25$).

Parameter	Copula	Weibull YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(L)} = 0.8$	AMH	0.8454	0.3206	0.3317	5.6743	0.2171	1.4737	95.8753
	Clayton	0.8033	0.2984	0.3197	0.4179	0.2186	1.3881	93.7626
	Frank	0.8734	0.3346	0.3453	9.1696	0.2175	1.5292	97.1916
	GH	0.9268	0.3543	0.3693	15.8551	0.2325	1.6212	97.7978
	Joe	0.9124	0.3655	0.3767	14.0457	0.1961	1.6287	97.9920
$\beta_{12}^{(L)} = -0.6$	AMH	-0.5985	0.1231	0.1199	0.2430	-0.8397	-0.3574	95.1710
	Clayton	-0.5832	0.1190	0.1197	2.8063	-0.8164	-0.3499	94.2656
	Frank	-0.6063	0.1263	0.1218	-1.0567	-0.8538	-0.3588	95.9880
	GH	-0.6041	0.1280	0.1224	-0.6776	-0.8549	-0.3532	96.3964
	Joe	-0.6041	0.1288	0.1239	-0.6877	-0.8565	-0.3517	96.4859
Parameter	Copula	BP YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(L)} = 0.8$	AMH	0.8732	0.3456	0.5894	9.1463	0.1756	1.5304	95.3815
	Clayton	0.8440	0.3284	0.5070	5.5004	0.1889	1.4762	94.7948
	Frank	0.8636	0.3519	0.4083	7.9500	0.1683	1.5479	96.2963
	GH	0.8558	0.3507	0.4249	6.9739	0.1617	1.5363	96.0843
	Joe	0.8777	0.3634	0.6558	9.7078	0.1415	1.5660	96.2814
$\beta_{12}^{(L)} = -0.6$	AMH	-0.6073	0.1276	0.1256	-1.2173	-0.8574	-0.3573	95.5868
	Clayton	-0.5933	0.1248	0.1237	1.1244	-0.8379	-0.3485	95.4955
	Frank	-0.6111	0.1302	0.1281	-1.8558	-0.8664	-0.3559	95.9000
	GH	-0.5996	0.1305	0.1288	0.0696	-0.8554	-0.3438	96.0883
	Joe	-0.6027	0.1317	0.1301	-0.4453	-0.8607	-0.3444	96.4895
Parameter	Copula	PE YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(L)} = 0.8$	AMH	0.8664	0.3452	0.3615	8.2977	0.1899	1.5429	95.9920
	Clayton	0.8452	0.3268	0.3504	5.6472	0.2046	1.4858	95.7831
	Frank	0.8727	0.3556	0.3698	9.0874	0.1757	1.5697	96.5898
	GH	0.8601	0.3515	0.3650	7.5125	0.1711	1.5491	96.2889
	Joe	0.8637	0.3570	0.3707	7.9609	0.1641	1.5633	96.4895
$\beta_{12}^{(L)} = -0.6$	AMH	-0.5917	0.1249	0.1204	1.3806	-0.8366	-0.3469	96.0922
	Clayton	-0.5805	0.1218	0.1191	3.2541	-0.8192	-0.3417	94.9799
	Frank	-0.5958	0.1275	0.1226	0.6950	-0.8458	-0.3459	96.4895
	GH	-0.5869	0.1276	0.1234	2.1866	-0.8370	-0.3368	96.4895
	Joe	-0.5902	0.1290	0.1249	1.6344	-0.8430	-0.3374	96.1886

Figures 5.1 (for PH and PO model classes) and 5.2 (for YP) show boxplots for the relative bias on regression parameter estimates for fitted survival copula models, divided by baseline, over AMH Weibull generated data (again, marginally fitting the correct regression structure), given Scenario S1 ($n = 500$; $\tau = 0.25$). For the majority of regression parameters, boxplots for fitted models with the correct copula had, in general, smaller range or fewer outlier count, given a fitted baseline distribution, even on a scenario with weak dependence. Also, given a regression model class, fitted BP and PE models' boxplot ranges are similar to those obtained for (correctly) fitted Weibull models, another evidence of the great flexibility from these semiparametric models to fit marginal hazard functions without bind them to a parametric functional form.

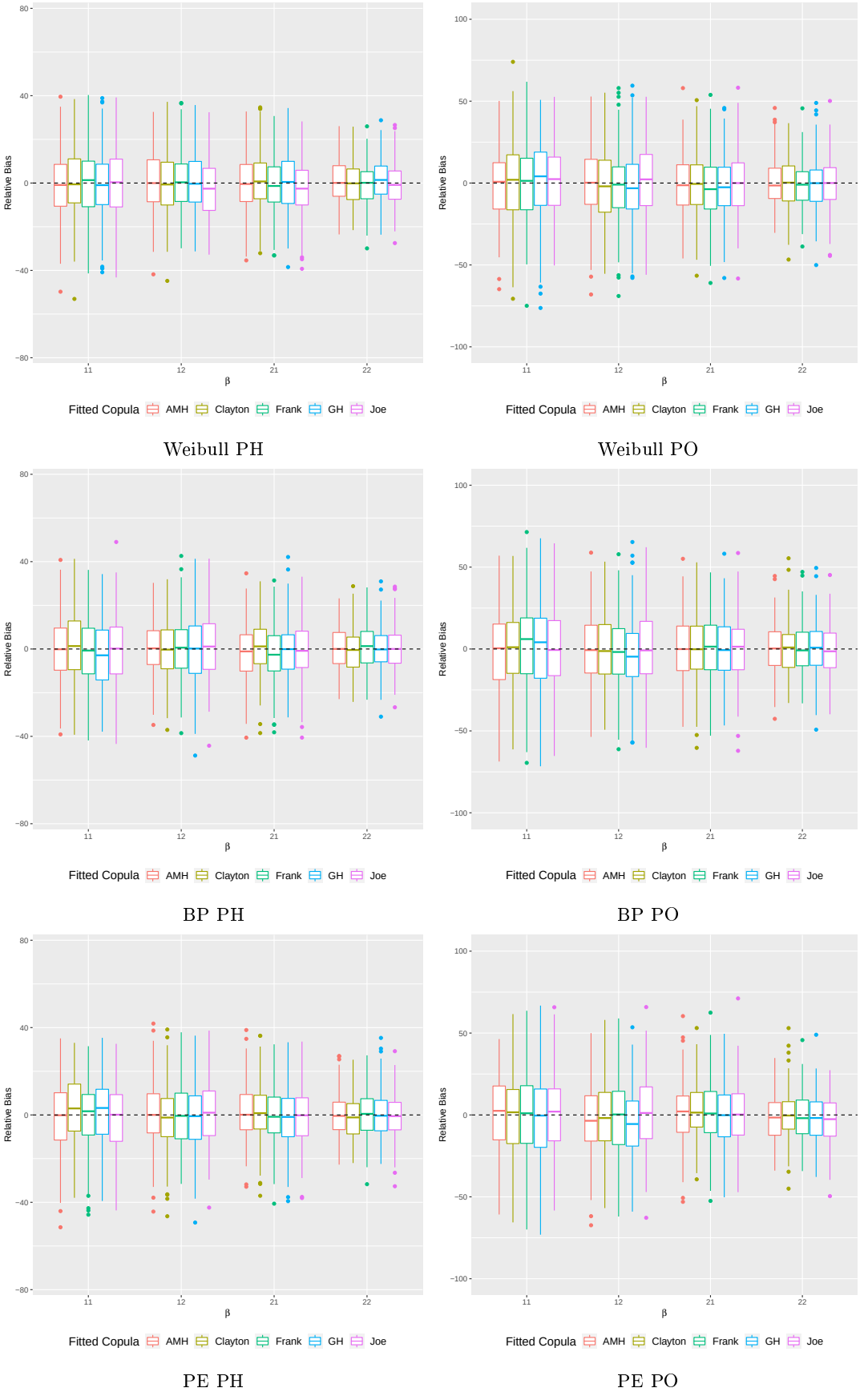
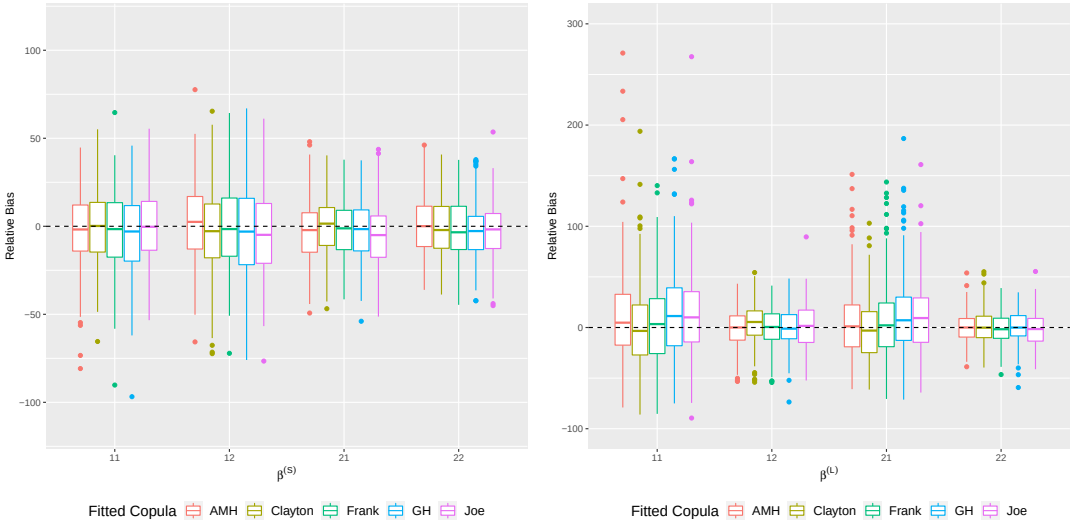
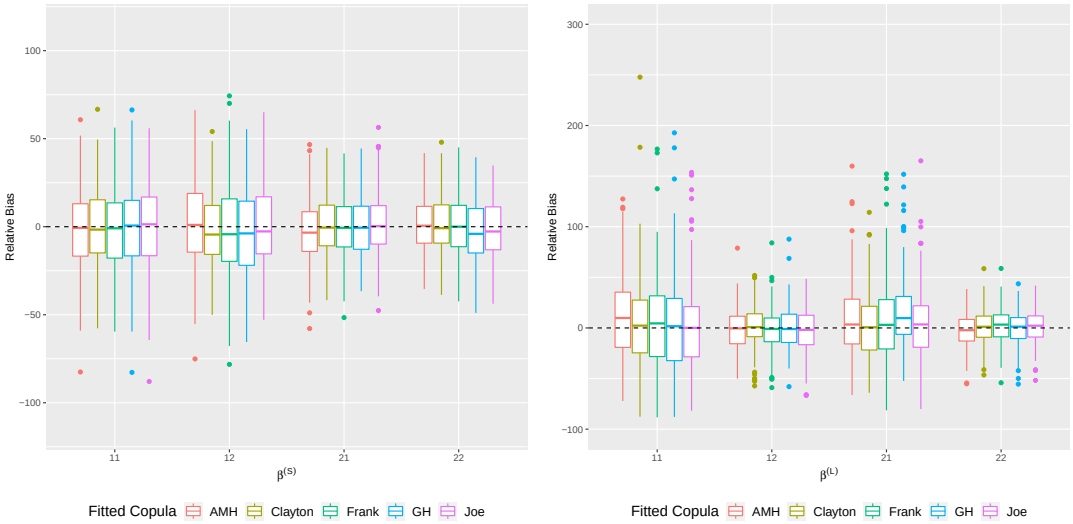


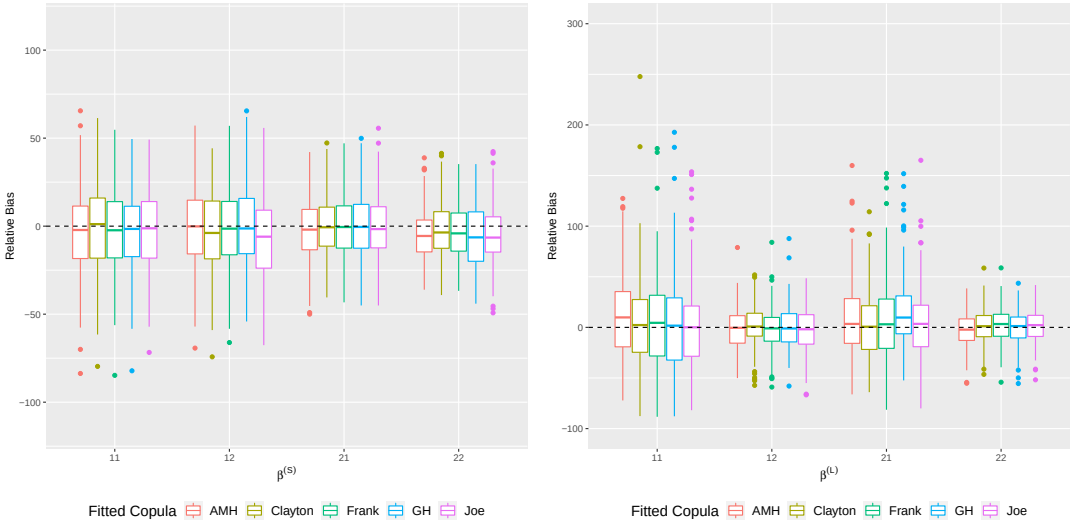
Figure 5.1: Relative bias on regression parameter estimates for fitted survival copula models over AMH Weibull (PH or PO) generated data ($n = 500$; $\tau = 0.25$).



Weibull YP (short) Weibull YP (long)



BP YP (short) BP YP (long)



PE YP (short) PE YP (long)

Figure 5.2: Relative bias on regression parameter estimates for fitted survival copula models over AMH Weibull YP generated data ($n = 500$; $\tau = 0.25$).

Even simulating with a high sample size, given generated data from an Archimedean copula model with marginal Weibull baseline distribution, fitted models with an incorrect (Archimedean) copula, among the five treated in this work (AMH, Clayton, Frank, GH and Joe), still show results close to the correct copula for regression parameters (regarding the ARB and CR) under a PH, PO or YP regression structure. Therefore, the choice of a copula for fitting has little impact on estimation of regression parameters. Also, given a correct copula fitting and a regression structure, the ARB and CR values for fitted models with a nonparametric baseline distribution (BP or PE) are similar to those obtained for (correctly) fitted Weibull models.

5.1.2 Akaike Information Criteria

Unlike conclusions obtained for regression parameter tables, results for the mean of the AIC values and their proportion of choice (by the smallest AIC, given each MC replica) appoint clearly to the correct copula model choice, when generating marginally from the Weibull baseline distribution, as seen from Tables 5.5 to 5.7 for PH, PO and YP regression structures, respectively. Looking only the correct copula fitted models, the highest proportions of choice are observed for the Joe copula, while the AMH and GH copulas (this last when fitted with the PE baseline) exhibit the lowest proportions, regardless of regression model class.

Table 5.5: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull PH margins ($n = 500$; $\tau = 0.25$).

True	Fitted	Weibull PH		BP PH		PE PH	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1631.46	67.90	1663.57	66.90	1663.27	66.90
	Clayton	1634.55	18.20	1666.40	19.50	1666.35	18.70
	Frank	1636.44	13.70	1668.51	13.40	1668.44	13.90
	GH	1655.72	0.20	1685.77	0.20	1682.51	0.50
	Joe	1669.42	0.00	1699.73	0.00	1696.41	0.00
Clayton	Clayton	1624.10	81.90	1655.94	82.50	1656.24	82.20
	AMH	1626.91	17.50	1658.52	16.90	1658.74	17.20
	Frank	1639.29	0.60	1671.08	0.60	1671.35	0.60
	GH	1658.35	0.00	1688.23	0.00	1685.50	0.00
	Joe	1672.79	0.00	1703.19	0.00	1700.91	0.00
Frank	Frank	1629.86	84.10	1661.78	83.30	1661.46	80.30
	AMH	1634.42	12.50	1666.31	11.90	1666.29	11.40
	Clayton	1643.45	0.90	1675.01	1.10	1674.83	0.90
	GH	1644.40	2.50	1674.18	3.70	1671.21	7.30
	Joe	1656.39	0.00	1685.65	0.00	1682.12	0.10
GH	GH	1602.46	82.10	1634.60	77.30	1635.05	65.00
	AMH	1636.56	0.00	1668.58	0.00	1669.10	0.00
	Clayton	1647.34	0.00	1678.96	0.10	1679.21	0.00
	Frank	1623.95	1.50	1655.94	1.70	1656.39	0.70
	Joe	1606.21	16.40	1637.76	20.90	1637.02	34.30
Joe	Joe	1567.79	89.60	1599.67	89.90	1600.63	95.70
	AMH	1637.18	0.00	1668.89	0.00	1669.37	0.00
	Clayton	1652.74	0.00	1683.90	0.00	1684.21	0.00
	Frank	1615.81	0.00	1647.55	0.00	1648.17	0.00
	GH	1573.57	10.40	1605.46	10.10	1608.51	4.30

For results in Table 5.5 for correctly fitted PH models, and taking also the correct copula fitting, the Joe copula presents the smallest mean AIC values, given each fitted baseline distribution. Inside each (correct) copula, fitted Weibull models exhibit the smallest mean AIC values, as expected. Fitted semiparametric models perform well and similarly to each other, with BP models being slightly better for the majority of copulas (it is only outperformed by the PE models when given the Frank copula).

Table 5.6: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull PO margins ($n = 500$; $\tau = 0.25$).

True	Fitted	Weibull PO		BP PO		PE PO	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1524.02	73.40	1553.17	71.70	1556.26	72.30
	Clayton	1527.58	16.20	1556.92	16.10	1559.77	15.00
	Frank	1530.21	10.30	1558.63	12.00	1561.82	12.20
	GH	1552.28	0.10	1576.42	0.20	1575.93	0.50
	Joe	1566.80	0.00	1591.85	0.00	1591.39	0.00
Clayton	Clayton	1514.42	84.40	1543.70	81.70	1546.57	81.80
	AMH	1517.85	15.40	1546.46	17.90	1549.57	17.80
	Frank	1532.87	0.20	1561.31	0.40	1563.96	0.40
	GH	1554.76	0.00	1579.05	0.00	1578.39	0.00
	Joe	1570.19	0.00	1596.16	0.00	1595.80	0.00
Frank	Frank	1525.10	82.70	1554.30	83.00	1557.26	79.50
	AMH	1529.32	14.80	1558.94	12.30	1562.23	11.50
	Clayton	1539.38	0.80	1568.95	0.70	1571.83	0.80
	GH	1542.13	1.70	1566.96	3.90	1566.61	8.10
	Joe	1554.97	0.00	1579.00	0.10	1578.15	0.10
GH	GH	1497.29	88.30	1526.84	77.80	1530.60	64.90
	AMH	1531.11	0.30	1561.09	0.20	1564.52	0.00
	Clayton	1542.80	0.00	1572.75	0.10	1575.92	0.00
	Frank	1518.31	2.50	1548.16	1.80	1551.88	0.50
	Joe	1502.21	8.90	1530.33	20.10	1532.28	34.60
Joe	Joe	1463.40	82.80	1492.63	90.90	1496.25	95.70
	AMH	1532.57	0.00	1562.43	0.00	1565.82	0.00
	Clayton	1549.23	0.00	1579.10	0.00	1582.01	0.00
	Frank	1510.95	0.00	1540.94	0.00	1544.67	0.00
	GH	1467.89	17.20	1498.54	9.10	1505.09	4.30

Taking results in Table 5.6 for correctly fitted PO models, along with the correct copula fitting, the Joe copula presents again the smallest mean AIC values, given each fitted baseline. Given each (correct) copula, once again fitted Weibull models exhibit the smallest mean AIC values. Fitted semiparametric models also perform well, but now BP baseline overcome PE for all five copulas. Similar results are observed in Table 5.7 for correctly fitted YP models, but for semiparametric models the BP baseline are slightly better than PE for the majority of copulas, only being outperformed when given the Clayton copula. Having more marginal regression parameters, YP models possess the lowest mean AIC given any combination of copula and fitted baseline. This was expected, since the YP regression structure is a more flexible model class that contains the other two (PH and PO) as particular cases.

Although the highest proportions of choice by the smallest AIC always point out to the correct copula choice, given any Archimedean copula for generation, it is possible to identify a similar pattern of fitted copulas with

Table 5.7: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull YP margins ($n = 500$; $\tau = 0.25$).

True	Fitted	Weibull YP		BP YP		PE YP	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1452.21	70.10	1483.53	70.50	1484.45	68.90
	Clayton	1456.01	17.50	1487.01	16.80	1487.78	18.00
	Frank	1457.82	12.30	1489.12	12.50	1489.85	12.70
	GH	1477.82	0.10	1506.95	0.20	1504.21	0.40
	Joe	1492.98	0.00	1522.31	0.00	1519.60	0.00
Clayton	Clayton	1443.29	83.10	1474.76	82.70	1475.45	83.60
	AMH	1446.29	16.70	1477.91	16.90	1478.38	16.10
	Frank	1461.58	0.20	1492.95	0.40	1493.41	0.30
	GH	1481.45	0.00	1510.66	0.00	1507.64	0.00
	Joe	1498.01	0.00	1527.91	0.00	1525.19	0.00
Frank	Frank	1454.63	84.90	1486.48	83.10	1487.11	79.20
	AMH	1459.53	12.30	1491.15	12.80	1491.76	12.40
	Clayton	1469.78	0.50	1500.81	0.50	1501.42	0.40
	GH	1469.61	2.30	1499.11	3.60	1496.36	8.00
	Joe	1482.36	0.00	1511.19	0.00	1507.76	0.00
GH	GH	1426.64	83.10	1458.16	79.00	1459.69	65.00
	AMH	1460.84	0.10	1492.32	0.10	1493.69	0.00
	Clayton	1472.77	0.00	1503.61	0.00	1504.59	0.00
	Frank	1448.32	1.40	1479.62	1.60	1481.05	0.40
	Joe	1430.85	15.40	1461.67	19.30	1461.39	34.60
Joe	Joe	1392.68	90.00	1424.21	89.50	1425.65	95.40
	AMH	1462.87	0.00	1494.04	0.00	1494.98	0.00
	Clayton	1479.56	0.00	1510.41	0.10	1510.97	0.00
	Frank	1441.53	0.00	1472.54	0.00	1473.74	0.00
	GH	1398.89	10.00	1430.24	10.40	1434.29	4.60

non-negligible proportions of choice, regardless of fitted baseline or regression structure. For example, when fitting generated data from an AMH copula, fitted models with the Clayton and Frank copulas always present proportions of choice about than 10% or higher. The same can be said for fitted AMH models over data generated from Clayton or Frank copulas, and so on. Therefore, it is possible to define two groups of copulas that, albeit not nested on themselves with respect to the copula function, seem to capture similar behaviors of dependence (to be checked by correlation estimation in the next subsection): the first one composed by Frank, AMH and Clayton copulas, and the second one by GH and Joe copulas.

5.1.3 Correlation Estimates

As well as observed AIC results, the correct copula choice provides the best Monte Carlo Kendall's τ estimates. For the sake of simplicity, comments on the dependence parameter are restricted to the ARB results, since computations for a Kendall's τ standard error estimate from its analogue for θ (either through a Delta Method or a nonparametric resampling technique) is extremely difficult for some copulas or computationally demanding. A special note should be taken on AMH models: as seen earlier, the Kendall's τ for the AMH copula is restricted to the interval $[-0.1817, 0.3333]$. To accommodate a stronger (positive) dependence, on

data generated from the AMH copula model, the supposed value for τ is truncated to the upper limit. That way, all other survival copula models fitted for AMH generated data will approximate well the truncated correlation value, but won't do the same for the supposed original τ value according to the copula chosen for fitting (treating the true copula as unknown). This explains the higher negative values for the ARB of fitted models for AMH generated data with $\tau = 0.5$ and $\tau = 0.75$ (see the Appendix for more details).

Table 5.8: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull PH margins ($n = 500$; $\tau = 0.25$)

True	Copula	Weibull PH		BP PH		PE PH	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.2501	0.0575	0.2507	0.2622	0.2518	0.7377
	Clayton	0.2357	-5.7147	0.2375	-4.9943	0.2394	-4.2569
	Frank	0.2385	-4.5856	0.2389	-4.4257	0.2415	-3.4193
	GH	0.1741	-30.3761	0.1875	-25.0138	0.2004	-19.8577
	Joe	0.1077	-56.9244	0.1223	-51.0694	0.1375	-44.9808
Clayton	Clayton	0.2513	0.5042	0.2516	0.6514	0.2538	1.5012
	AMH	0.2737	9.4982	0.2791	11.6256	0.2788	11.5064
	Frank	0.2308	-7.6685	0.2315	-7.4135	0.2339	-6.4577
	GH	0.1656	-33.7657	0.1796	-28.1539	0.1929	-22.8444
	Joe	0.0932	-62.7046	0.1059	-57.6590	0.1220	-51.2054
Frank	Frank	0.2504	0.1504	0.2504	0.1746	0.2528	1.1036
	AMH	0.2249	-10.0457	0.2253	-9.8906	0.2264	-9.4488
	Clayton	0.2177	-12.9197	0.2210	-11.6183	0.2227	-10.9247
	GH	0.1953	-21.8819	0.2068	-17.2841	0.2191	-12.3614
	Joe	0.1391	-44.3578	0.1557	-37.7189	0.1694	-32.2251
GH	GH	0.2487	-0.5239	0.2506	0.2495	0.2655	6.1938
	AMH	0.2130	-14.8199	0.2138	-14.4629	0.2147	-14.1087
	Clayton	0.2090	-16.3891	0.2133	-14.6809	0.2150	-13.9853
	Frank	0.2650	6.0163	0.2655	6.2063	0.2674	6.9612
	Joe	0.2068	-17.2847	0.2134	-14.6314	0.2300	-8.0025
Joe	Joe	0.2514	0.5567	0.2542	1.6760	0.2717	8.6645
	AMH	0.2007	-19.7268	0.2017	-19.3366	0.2024	-19.0564
	Clayton	0.1982	-20.7288	0.2037	-18.5174	0.2053	-17.8804
	Frank	0.2826	13.0336	0.2838	13.5001	0.2854	14.1603
	GH	0.2824	12.9736	0.2810	12.4060	0.2966	18.6229

Given Scenario S1 ($n = 500$; $\tau = 0.25$) from Tables 5.8 to 5.10, the least ARB was almost always observed for the correct copula choice (except for GH generated data when fitting the Frank copula with PE PO or PE YP for its margins). Moreover, given any fitted model with the correct copula, when plugging the Weibull model as fitted baseline distribution, it presented the lowest ARB among all fitted baseline models for correctly fitted PH and YP regression structures, while the BP model exhibited the lowest ARB for correctly fitted PO models. However, the BP baseline performs well for PH and YP models, as the Weibull baseline for PO models: their ARB values are always below 4%, regardless of the regression structure. Concerning the PE baseline, it also performs well, specially for the correct fitting of AMH, Clayton and Frank copulas (always below 2%), but not as much as Weibull and BP models for GH and Joe copulas, given any regression model class.

For correctly fitted survival copula models with the PH regression structure in Table 5.8, and given each one of the three fitted baseline distributions, the smallest ARB were observed for the AMH (Weibull, PE) and

Frank (BP) copulas. For all fitted (correct) AMH survival copula models, their ARB values were below 1%, while for the Joe copula it is observed only when fitting the Weibull baseline. On the other hand, for fitted models with the wrong copula, regardless of the baseline chosen for its margins, the ARB frequently exceeds 10%, specially when generating data from the Joe copula model (and fitting incorrectly any other Archimedean survival copula model).

Table 5.9: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull PO margins ($n = 500$; $\tau = 0.25$)

True	Copula	Weibull PO		BP PO		PE PO	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.2522	0.8760	0.2502	0.0823	0.2511	0.4248
	Clayton	0.2340	-6.4012	0.2326	-6.9744	0.2347	-6.1096
	Frank	0.2394	-4.2344	0.2425	-2.9897	0.2448	-2.0625
	GH	0.1665	-33.3865	0.1941	-22.3417	0.2079	-16.8443
	Joe	0.0976	-60.9664	0.1272	-49.1358	0.1440	-42.4011
Clayton	Clayton	0.2540	1.5988	0.2511	0.4568	0.2530	1.2178
	AMH	0.2826	13.0418	0.2826	13.0421	0.2824	12.9613
	Frank	0.2332	-6.7320	0.2374	-5.0269	0.2398	-4.0750
	GH	0.1597	-36.1176	0.1895	-24.2055	0.2035	-18.6034
	Joe	0.0841	-66.3741	0.1112	-55.5349	0.1287	-48.5108
Frank	Frank	0.2498	-0.0906	0.2508	0.3079	0.2529	1.1439
	AMH	0.2246	-10.1600	0.2228	-10.8646	0.2235	-10.5958
	Clayton	0.2132	-14.7223	0.2128	-14.8752	0.2147	-14.1165
	GH	0.1868	-25.2842	0.2098	-16.0651	0.2229	-10.8207
	Joe	0.1284	-48.6378	0.1597	-36.1350	0.1745	-30.2145
GH	GH	0.2411	-3.5522	0.2508	0.3010	0.2663	6.5337
	AMH	0.2127	-14.9300	0.2111	-15.5732	0.2116	-15.3635
	Clayton	0.2041	-18.3629	0.2031	-18.7502	0.2052	-17.9385
	Frank	0.2643	5.7298	0.2645	5.7817	0.2660	6.4176
	Joe	0.1968	-21.2627	0.2153	-13.8696	0.2331	-6.7525
Joe	Joe	0.2409	-3.6527	0.2542	1.6621	0.2730	9.2106
	AMH	0.1992	-20.3069	0.1974	-21.0261	0.1978	-20.8973
	Clayton	0.1919	-23.2354	0.1904	-23.8456	0.1923	-23.0789
	Frank	0.2807	12.2675	0.2806	12.2461	0.2814	12.5418
	GH	0.2739	9.5605	0.2781	11.2486	0.2944	17.7418

When correctly fitting a survival copula model with the PO regression structure in Table 5.9, the smallest ARB were observed for the AMH (PE), Frank (Weibull) and GH (BP) copulas. Again, for all fitted (correct) AMH survival copula models, their ARB values were below 1%, while for the Joe copula it is always above this percentage. Again, when fitting the wrong copula, regardless of the marginal baseline, the ARB often surpass 10%, now specially for data generated from Frank and Joe copula models. Finally, for correctly fitted survival copula models with the YP regression structure in Table 5.10, given each one of the three fitted baselines, the smallest ARB were always observed for the AMH copula (all of them below 1%), regardless of the fitted baseline, while for the Joe copula it is observed only when fitting the Weibull baseline. For fitted models with the wrong copula, the ARB frequently exceeds 10%, specially when generating from Frank and Joe copulas.

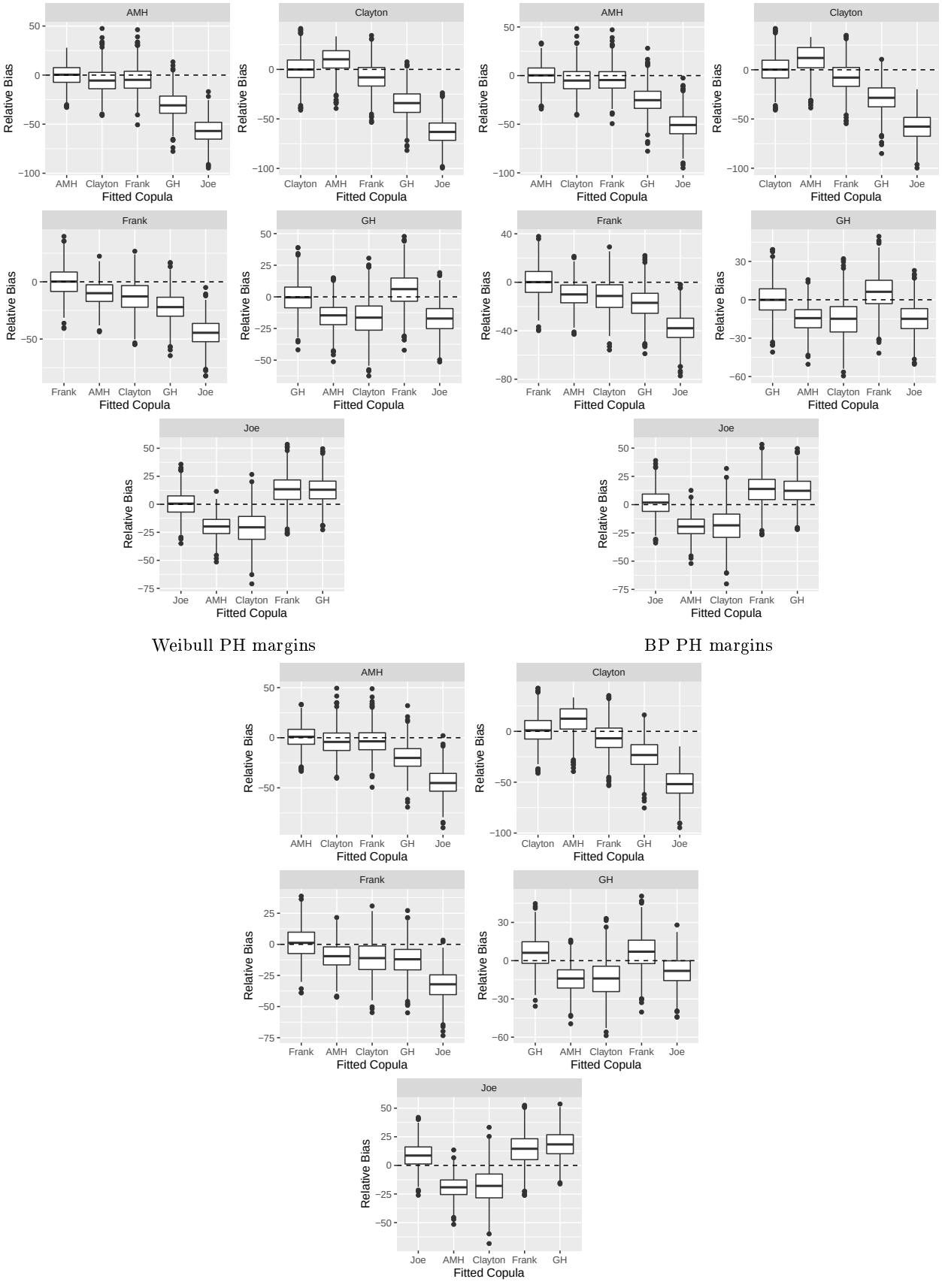
Figures 5.3 to 5.5 show boxplots for the relative bias on Kendall's τ estimation for all fitted survival copula models in Scenario S1 ($n = 500$; $\tau = 0.25$). Again, the correct copula choice provides the best results (*i.e.*,

more centered around 0 when compared to fitted models with a wrong copula), regardless of the baseline distribution or regression model class used for marginal fitting. Taking only the fitted models with the correct copula, given any regression structure, the best results are observed for the Weibull baseline, although BP models have similar performance for all copulas, as well as PE models (except for GH and Joe copulas).

Table 5.10: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull YP margins ($n = 500$; $\tau = 0.25$)

True	Copula	Weibull YP		BP YP		PE YP	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.2504	0.1433	0.2506	0.2271	0.2513	0.5034
	Clayton	0.2312	-7.5121	0.2338	-6.4906	0.2353	-5.8715
	Frank	0.2415	-3.3931	0.2431	-2.7794	0.2451	-1.9716
	GH	0.1798	-28.0668	0.1956	-21.7589	0.2091	-16.3777
	Joe	0.1114	-55.4558	0.1286	-48.5525	0.1454	-41.8507
Clayton	Clayton	0.2517	0.6780	0.2522	0.8913	0.2536	1.4278
	AMH	0.2855	14.1892	0.2859	14.3573	0.2830	13.1897
	Frank	0.2362	-5.5211	0.2386	-4.5782	0.2405	-3.7861
	GH	0.1742	-30.3022	0.1916	-23.3753	0.2054	-17.8321
	Joe	0.0974	-61.0555	0.1124	-55.0309	0.1301	-47.9483
Frank	Frank	0.2507	0.2944	0.2512	0.4942	0.2530	1.1973
	AMH	0.2226	-10.9621	0.2232	-10.7273	0.2238	-10.4709
	Clayton	0.2100	-15.9833	0.2150	-13.9823	0.2164	-13.4214
	GH	0.1987	-20.5014	0.2108	-15.6613	0.2238	-10.4984
	Joe	0.1427	-42.9139	0.1610	-35.5982	0.1757	-29.7065
GH	GH	0.2488	-0.4624	0.2511	0.4276	0.2663	6.5289
	AMH	0.2106	-15.7413	0.2117	-15.3085	0.2121	-15.1727
	Clayton	0.2008	-19.6602	0.2056	-17.7519	0.2069	-17.2430
	Frank	0.2631	5.2348	0.2648	5.9315	0.2660	6.3892
	Joe	0.2082	-16.7336	0.2158	-13.6746	0.2333	-6.6855
Joe	Joe	0.2516	0.6351	0.2546	1.8259	0.2732	9.2728
	AMH	0.1968	-21.2954	0.1979	-20.8251	0.1983	-20.6982
	Clayton	0.1874	-25.0559	0.1934	-22.6403	0.1943	-22.2730
	Frank	0.2783	11.3159	0.2811	12.4523	0.2817	12.6789
	GH	0.2803	12.1088	0.2784	11.3529	0.2945	17.8041

When simulating with a high sample size, even on a scenario with weak dependence, the choice of the correct Archimedean copula for fitting is crucial to ensure a suitable estimation of θ and consequently of the Kendall's τ correlation. This was expected, since all copulas considered for fitting are defined as functions of a single dependence parameter, besides being enough to produce, on average, better AIC results. Also, semiparametric models have similar performance to (correctly fitted) Weibull models for almost all fitted survival models with the correct copula, no matter which results are taken to comparison, be it regression parameters, an information criteria, or the correlation parameter. Concerning the identified groups of copulas earlier on mean AIC evaluation (Frank, AMH and Clayton composing the first group, and GH and Joe the second one), this is corroborated by ARB results on the correlation estimation, although the Frank copula also seems to compete with the correct fitting of a GH copula along with Joe.

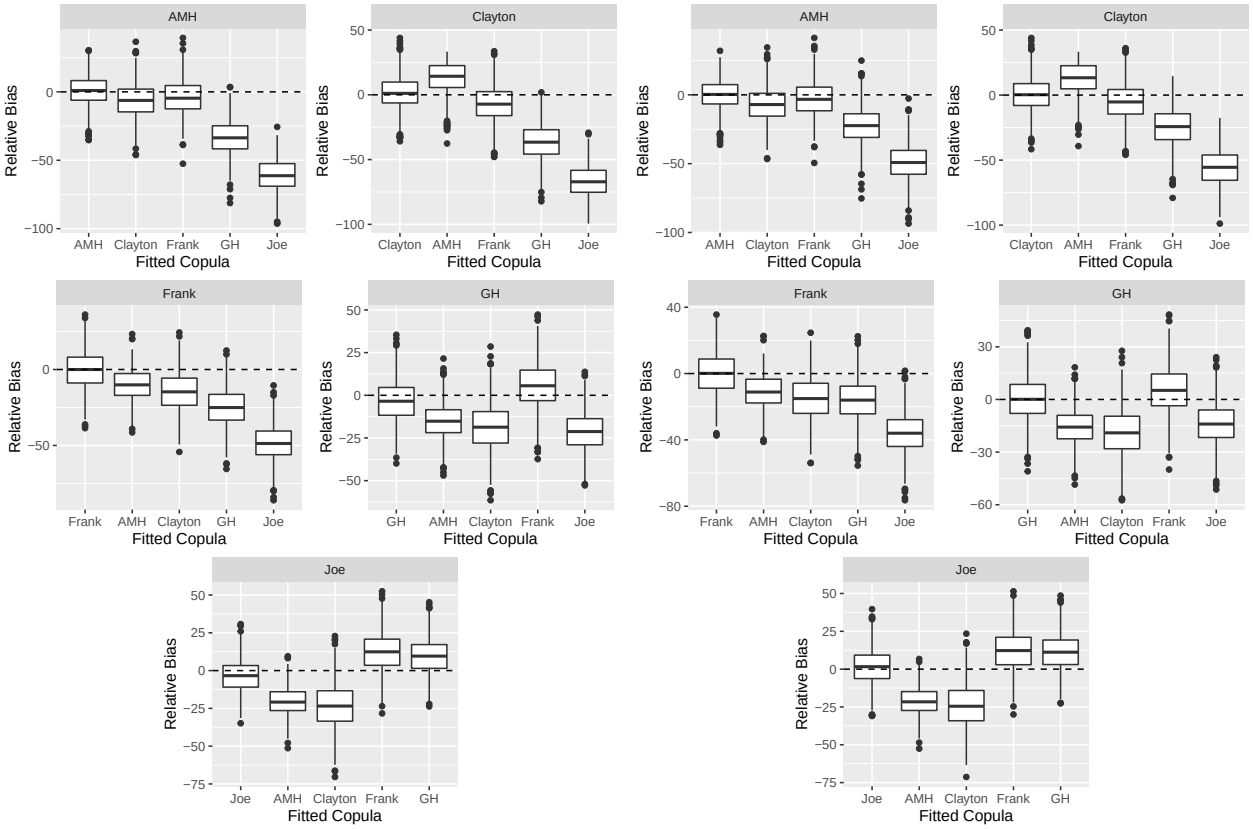


Weibull PH margins

BP PH margins

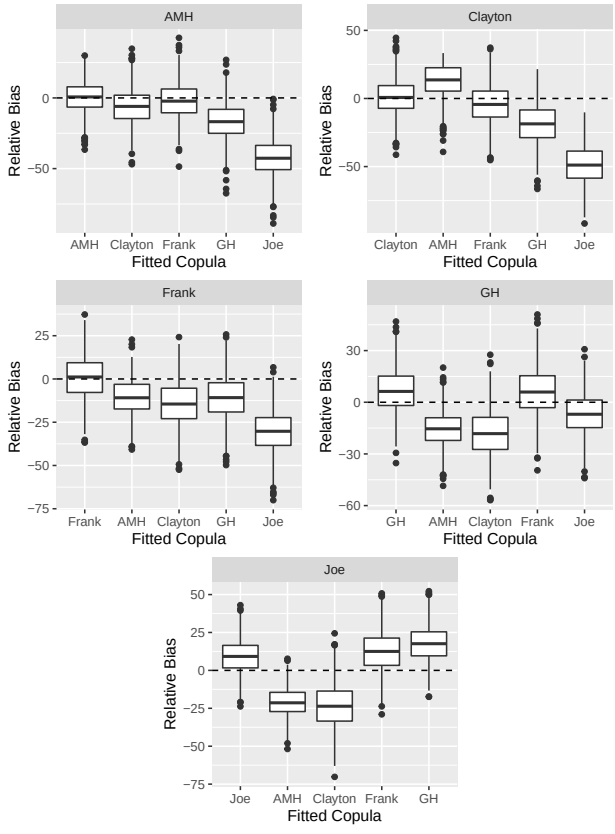
PE PH margins

Figure 5.3: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull PH margins ($n = 500$; $\tau = 0.25$)



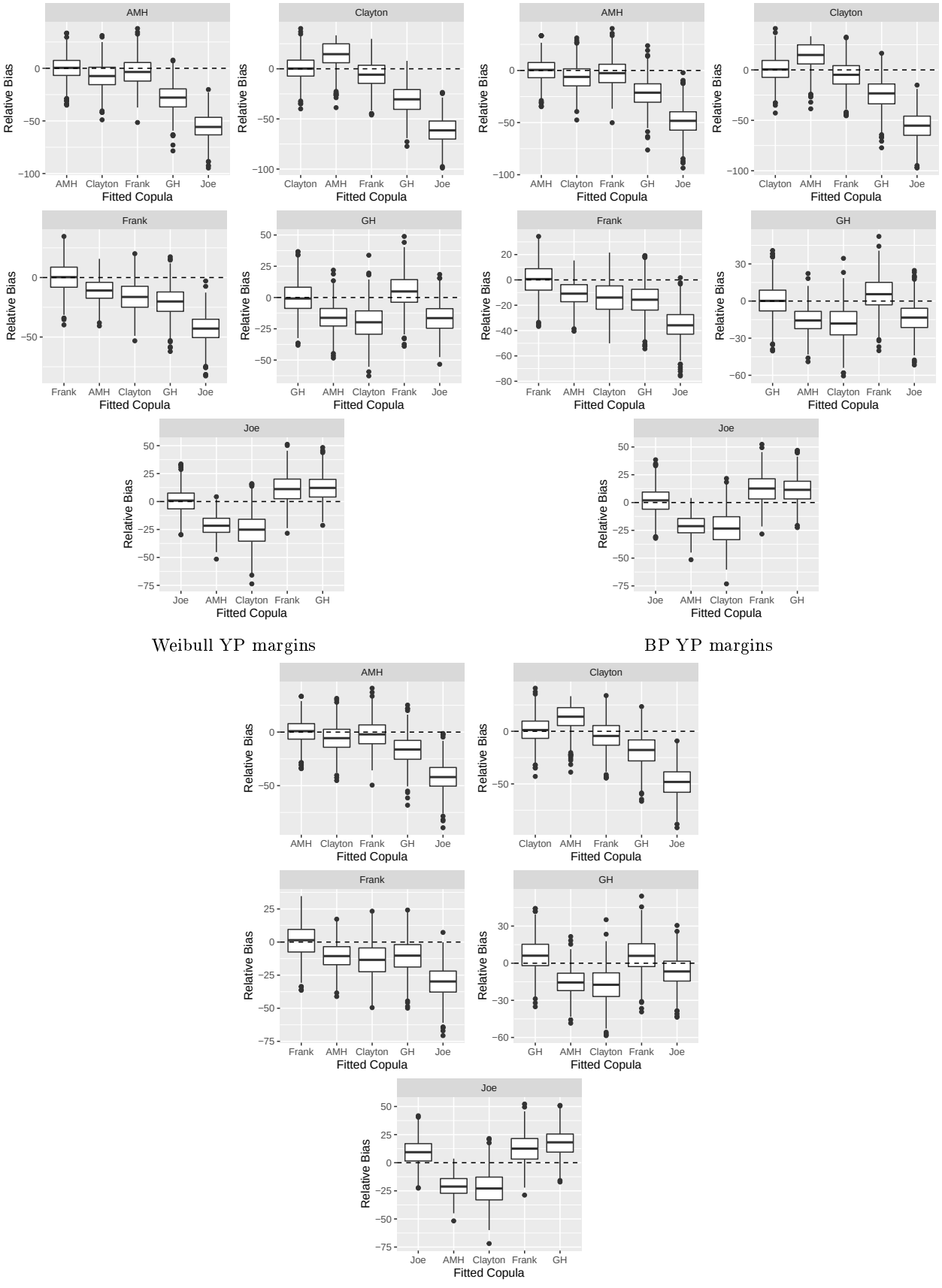
Weibull PO margins

BP PO margins



PE PO margins

Figure 5.4: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull PO margins ($n = 500$; $\tau = 0.25$)



Weibull YP margins

BP YP margins

PE YP margins

Figure 5.5: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull YP margins ($n = 500$; $\tau = 0.25$)

5.1.4 Likelihood Ratio Tests

The analysis presented below compares, through the Likelihood Ratio (LR) test, two nested models with respect to their regression structure (*i.e.*, PH *vs.* YP, and PO *vs.* YP), given each class used for generation (always fitting the correct copula). Note that there are no pairs of nested Archimedean copulas, or fitted baseline models, among those treated in this work. Since there are two covariates for each margin, keeping the same specification for fitted PH, PO and YP models, p-values for all LR statistics from tests comparing nested models from each MC replica will be obtained from a χ^2 distribution with $2 \times 2 = 4$ degrees of freedom (note that PH and PO classes have the same number of parameters). Under the null hypothesis, it is supposed that the additional regression parameters from YP model are not significant. If the LR statistic does not surpass the critical value under a significance of 5% (here, equal to $\chi^2(0.95, 4) \approx 9.4877$), the regression class with less parameters (more parsimonious model), is chosen. Otherwise, the YP model is selected as the best class. Results for the average statistics and p-values from LR tests are presented from Tables 5.11 to 5.13.

Table 5.11: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull PH margins ($n = 500; \tau = 0.25$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	4.1965	0.4846	4.2689	0.4787	4.1354	0.4907
	PO	48.0752	$<$ 0.0001	33.8761	0.0005	34.1804	0.0002
Clayton	PH	4.0989	0.4923	4.0398	0.4921	3.9649	0.4991
	PO	48.7359	$<$ 0.0001	34.4732	0.0003	34.6441	0.0002
Frank	PH	4.1500	0.4742	4.1365	0.4736	4.0266	0.4840
	PO	47.6290	$<$ 0.0001	34.1908	0.0003	34.6151	0.0002
GH	PH	4.0212	0.5008	4.0232	0.5000	3.8907	0.5124
	PO	47.3462	$<$ 0.0001	33.5009	0.0003	34.0694	0.0002
Joe	PH	4.1007	0.4867	4.1930	0.4755	4.0980	0.4858
	PO	47.2089	$<$ 0.0001	33.6742	0.0004	34.7567	0.0003

For Table 5.11, when generating from the PH regression model class, all LR tests accept fitted YP models against the (incorrect) PO regression structure, given any combination of fitted baseline and copula, as expected. Remember that the YP model is a generalization for the PH class, which are not nested within the PO structure. However, the same cannot be said for LR tests confronting fitted PH and YP models. Their results are always non-significant, leading to the choice of (correctly) fitted PH models, since they are more parsimonious (fewer regression parameters). Therefore, given a large sample size, introducing more regression parameters with a wider functional form to capture both short and long-term covariate effects (without increasing the number of original covariates) does not provide a significantly better fitting.

Looking now for Table 5.12, this time generating from the PO class, the converse is also true: all LR tests accept fitted YP models against the (incorrect) PH structure, given any combination of fitted baseline and copula, as expected, since the YP model is also a generalization for the PO class. However, the same cannot be said for LR tests confronting fitted PO and YP models. Their results lead towards the choice of fitted PO models, since (in this case) they are more parsimonious. Finally, for Table 5.13, when generating from the

Table 5.12: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull PO margins ($n = 500; \tau = 0.25$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	30.1272	0.0018	23.7872	0.0100	24.9197	0.0066
	PO	5.5396	0.3600	4.1265	0.4960	5.3370	0.3465
Clayton	PH	31.1509	0.0015	24.6702	0.0076	25.9595	0.0051
	PO	5.5430	0.3617	4.0845	0.4890	5.2688	0.3469
Frank	PH	29.9082	0.0017	23.6726	0.0096	24.7241	0.0066
	PO	5.4930	0.3664	4.1064	0.4922	5.2970	0.3465
GH	PH	29.0730	0.0026	23.2333	0.0096	24.0902	0.0063
	PO	5.5658	0.3659	4.3068	0.4800	5.4802	0.3429
Joe	PH	30.1505	0.0029	23.7688	0.0139	24.2719	0.0107
	PO	5.5344	0.3654	4.2056	0.4780	5.4340	0.3316

wider YP regression model class, all LR tests accept fitted YP models against the PH or PO structures, given any combination of fitted baseline and copula. This was also expected, since (nested) PH and PO models do not account for covariate short and long-term effects, but only during the whole time of follow-up.

Table 5.13: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull YP margins ($n = 500; \tau = 0.25$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	122.1934	< 0.0001	108.0926	< 0.0001	109.9563	< 0.0001
	PO	94.2805	< 0.0001	80.5264	< 0.0001	82.8430	< 0.0001
Clayton	PH	128.0063	< 0.0001	113.2562	< 0.0001	115.0452	< 0.0001
	PO	100.5764	< 0.0001	85.5648	< 0.0001	87.9450	< 0.0001
Frank	PH	122.1422	< 0.0001	107.9013	< 0.0001	109.4467	< 0.0001
	PO	96.4341	< 0.0001	82.5658	< 0.0001	84.6061	< 0.0001
GH	PH	120.1831	< 0.0001	106.2706	< 0.0001	107.3817	< 0.0001
	PO	93.6736	< 0.0001	80.5333	< 0.0001	82.1474	< 0.0001
Joe	PH	123.6797	< 0.0001	108.7011	< 0.0001	109.9063	< 0.0001
	PO	96.3288	< 0.0001	82.1735	< 0.0001	84.0249	< 0.0001

Tables 5.11 to 5.13 showed that the analysis through the LR test for nested regression model classes, when generating from an Archimedean survival copula with marginal Weibull baseline distribution, and regardless of the fitted copula or baseline models (among those considered in this work) is a useful tool to choose the regression structure for fitting if the one that generated the data is unknown.

5.2 Generated Copulas with EW Baseline

This section shows results for fitted survival copula models over generated data from Archimedean survival copulas with marginal EW baseline distribution, also associated to a regression model class. Results in the following subsections are divided in the same way as done for generated copula data with Weibull margins.

5.2.1 Regression Parameter Estimates

The Monte Carlo estimates on regression parameters for fitted survival copula models, when marginally generating from the EW distribution, are showed from Tables 5.14 to 5.17, divided by fitted baseline distribution for each regression parameter set from a given class (for $j = 1$). For those results, comparisons are again done among fitted models with different copulas, yet keeping the same regression model class used for generation.

Table 5.14: MC statistics for 1st margin regression parameter estimates of fitted survival copula models over AMH EW PH generated data ($n = 500$; $\tau = 0.25$)

Parameter	Copula	Weibull PH Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.6066	0.1015	0.0895	13.3455	-0.8055	-0.4077	88.9972
	Clayton	-0.5993	0.0995	0.0899	14.3829	-0.7943	-0.4044	85.6960
	Frank	-0.6095	0.1026	0.0917	12.9305	-0.8106	-0.4084	89.4268
	GH	-0.6141	0.1040	0.0940	12.2656	-0.8180	-0.4103	89.6552
	Joe	-0.6165	0.1045	0.0947	11.9343	-0.8213	-0.4116	89.8848
$\beta_{12} = 0.4$	AMH	0.3502	0.0519	0.0488	-12.4418	0.2486	0.4519	84.2618
	Clayton	0.3454	0.0507	0.0488	-13.6555	0.2460	0.4447	81.8646
	Frank	0.3494	0.0525	0.0500	-12.6491	0.2465	0.4524	84.7134
	GH	0.3515	0.0532	0.0518	-12.1180	0.2472	0.4559	85.9515
	Joe	0.3529	0.0535	0.0521	-11.7710	0.2480	0.4578	86.2996
Parameter	Copula	BP PH Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7100	0.1039	0.1048	-1.4239	-0.9137	-0.5062	93.8878
	Clayton	-0.6990	0.1023	0.1061	0.1419	-0.8994	-0.4986	93.5936
	Frank	-0.7105	0.1053	0.1070	-1.4981	-0.9168	-0.5042	93.7938
	GH	-0.7115	0.1071	0.1083	-1.6477	-0.9214	-0.5017	94.0798
	Joe	-0.7086	0.1073	0.1077	-1.2283	-0.9189	-0.4983	94.3590
$\beta_{12} = 0.4$	AMH	0.4025	0.0530	0.0553	0.6300	0.2986	0.5064	93.3868
	Clayton	0.3973	0.0522	0.0560	-0.6811	0.2950	0.4995	92.2923
	Frank	0.4016	0.0537	0.0566	0.4077	0.2963	0.5070	93.0931
	GH	0.4043	0.0547	0.0579	1.0772	0.2971	0.5115	92.7928
	Joe	0.4032	0.0548	0.0577	0.8078	0.2958	0.5107	92.9487
Parameter	Copula	PE PH Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7014	0.1037	0.1037	-0.1956	-0.9046	-0.4981	94.1767
	Clayton	-0.6875	0.1018	0.1048	1.7838	-0.8870	-0.4881	93.3801
	Frank	-0.7053	0.1048	0.1056	-0.7579	-0.9107	-0.4999	94.0704
	GH	-0.7057	0.1057	0.1062	-0.8144	-0.9129	-0.4985	94.3517
	Joe	-0.7032	0.1063	0.1066	-0.4513	-0.9115	-0.4948	94.6223
$\beta_{12} = 0.4$	AMH	0.3981	0.0529	0.0549	-0.4838	0.2943	0.5018	93.4739
	Clayton	0.3906	0.0520	0.0553	-2.3384	0.2888	0.4925	92.3771
	Frank	0.3993	0.0535	0.0559	-0.1666	0.2944	0.5042	93.2663
	GH	0.4012	0.0539	0.0566	0.2880	0.2954	0.5069	92.4262
	Joe	0.3998	0.0542	0.0571	-0.0478	0.2935	0.5061	92.5736

Taking the results in Table 5.14, when correctly fitting the PH class under a low correlation, the ARB is always lower (in magnitude) than 3%, and the CR is at most 0.03 away from the confidence level when

fitting a semiparametric model, for all regression parameters, even when fitting the wrong copula. However, for (incorrectly) fitted Weibull models, the ARB is always above 10%, while the CR is always below 90%, even with the correct copula. On the other hand, fitted BP and PE models with the correct (AMH) copula have produced, in general, smaller ARB values and closer CR values to 95%. Therefore, they perform better than fitted Weibull models. This was expected due to the high flexibility of these semiparametric models: they do not impose any parametric restriction for the (marginal) hazard rate function. It allows to capture non-monotonic behaviors for the (marginal) hazard function that cannot be wrapped by any parametric specification in the Weibull family, and therefore produce more accurate and better regression parameter estimates.

Table 5.15: MC statistics for 1st margin regression parameter estimates of fitted survival copula models over AMH EW PO generated data ($n = 500$; $\tau = 0.25$)

Parameter	Copula	Weibull PO Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7267	0.1615	0.1601	-3.8092	-1.0432	-0.4102	95.6958
	Clayton	-0.7231	0.1603	0.1585	-3.2944	-1.0373	-0.4089	95.6577
	Frank	-0.7103	0.1624	0.1628	-1.4767	-1.0287	-0.3920	95.4023
	GH	-0.7043	0.1640	0.1678	-0.6213	-1.0258	-0.3829	95.2685
	Joe	-0.7024	0.1645	0.1681	-0.3468	-1.0249	-0.3800	94.8914
$\beta_{12} = 0.4$	AMH	0.4151	0.0810	0.0839	3.7873	0.2565	0.5738	93.9742
	Clayton	0.4124	0.0804	0.0845	3.0914	0.2548	0.5700	93.4866
	Frank	0.4023	0.0817	0.0868	0.5804	0.2422	0.5624	93.6143
	GH	0.3970	0.0826	0.0894	-0.7601	0.2351	0.5588	92.9668
	Joe	0.3958	0.0829	0.0895	-1.0499	0.2334	0.5582	93.1034
Parameter	Copula	BP PO Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7255	0.1616	0.1610	-3.6441	-1.0423	-0.4088	94.8000
	Clayton	-0.7291	0.1614	0.1633	-4.1529	-1.0454	-0.4128	94.5000
	Frank	-0.7160	0.1625	0.1634	-2.2924	-1.0345	-0.3976	94.6000
	GH	-0.7126	0.1653	0.1683	-1.7988	-1.0366	-0.3885	94.5083
	Joe	-0.7101	0.1656	0.1683	-1.4413	-1.0347	-0.3855	94.3152
$\beta_{12} = 0.4$	AMH	0.4107	0.0814	0.0866	2.6638	0.2511	0.5702	93.1000
	Clayton	0.4133	0.0813	0.0873	3.3176	0.2539	0.5726	92.7000
	Frank	0.4038	0.0819	0.0875	0.9387	0.2433	0.5642	93.0000
	GH	0.4025	0.0833	0.0900	0.6137	0.2391	0.5658	92.0817
	Joe	0.4007	0.0835	0.0889	0.1824	0.2372	0.5643	93.1525
Parameter	Copula	PE PO Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11} = -0.7$	AMH	-0.7111	0.1612	0.1579	-1.5817	-1.0271	-0.3951	95.2764
	Clayton	-0.7108	0.1607	0.1605	-1.5469	-1.0258	-0.3959	95.0505
	Frank	-0.7055	0.1622	0.1609	-0.7889	-1.0233	-0.3877	94.9648
	GH	-0.7076	0.1635	0.1653	-1.0901	-1.0280	-0.3872	95.1469
	Joe	-0.7064	0.1640	0.1662	-0.9186	-1.0279	-0.3849	94.8849
$\beta_{12} = 0.4$	AMH	0.4012	0.0812	0.0849	0.3037	0.2421	0.5604	93.4673
	Clayton	0.4017	0.0809	0.0850	0.4323	0.2431	0.5603	93.6364
	Frank	0.3972	0.0817	0.0859	-0.7022	0.2371	0.5573	93.5549
	GH	0.3997	0.0823	0.0879	-0.0796	0.2384	0.5609	92.5926
	Joe	0.3988	0.0826	0.0881	-0.3080	0.2369	0.5606	92.9668

For results in Table 5.15, this time correctly fitting the PO structure, (incorrectly) fitted Weibull models

perform worse than, with respect to regression parameter estimation, fitted semiparametric models, although that difference is smaller than the observed given the PH class. Other than that, conclusions over the ARB and CR values are similar to those obtained for fitted semiparametric PH models in Table 5.14. Moreover, their magnitude are similar to those obtained for marginally generated data from the Weibull baseline.

Table 5.16: MC statistics for 1st margin short-term regression parameter estimates of fitted survival copula models over AMH EW YP generated data ($n = 500$; $\tau = 0.25$)

Parameter	Copula	Weibull YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(S)} = -0.7$	AMH	-0.9417	0.1450	0.1420	-34.5262	-1.2259	-0.6574	60.7692
	Clayton	-0.9342	0.1446	0.1425	-33.4615	-1.2177	-0.6508	63.4615
	Frank	-0.9400	0.1457	0.1427	-34.2802	-1.2255	-0.6545	61.5877
	GH	-0.9372	0.1473	0.1451	-33.8843	-1.2258	-0.6486	63.3205
	Joe	-0.9184	0.1481	0.1474	-31.2050	-1.2086	-0.6282	68.4547
$\beta_{12}^{(S)} = 0.4$	AMH	0.3796	0.0941	0.1101	-5.1048	0.1952	0.5640	90.1282
	Clayton	0.3837	0.0933	0.1078	-4.0625	0.2009	0.5666	90.7692
	Frank	0.3685	0.0956	0.1131	-7.8817	0.1811	0.5559	89.1165
	GH	0.3594	0.0960	0.1168	-10.1546	0.1712	0.5476	86.3578
	Joe	0.3588	0.0967	0.1173	-10.2903	0.1694	0.5483	86.5900
Parameter	Copula	BP YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(S)} = -0.7$	AMH	-0.7013	0.1591	0.1614	-0.1835	-1.0132	-0.3894	95.1904
	Clayton	-0.7022	0.1595	0.1628	-0.3190	-1.0148	-0.3896	94.8795
	Frank	-0.6912	0.1601	0.1643	1.2584	-1.0049	-0.3775	94.8000
	GH	-0.6949	0.1631	0.1683	0.7306	-1.0146	-0.3752	95.2381
	Joe	-0.6961	0.1633	0.1678	0.5590	-1.0162	-0.3760	95.6522
$\beta_{12}^{(S)} = 0.4$	AMH	0.4284	0.0923	0.1003	7.0937	0.2475	0.6093	92.2846
	Clayton	0.4260	0.0921	0.1010	6.5109	0.2456	0.6065	92.1687
	Frank	0.4228	0.0926	0.1016	5.6939	0.2413	0.6043	92.5000
	GH	0.4168	0.0940	0.1037	4.2096	0.2326	0.6010	92.9215
	Joe	0.4138	0.0942	0.1038	3.4473	0.2292	0.5984	92.5831
Parameter	Copula	PE YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(S)} = -0.7$	AMH	-0.7154	0.1589	0.1576	-2.1955	-1.0268	-0.4039	94.9648
	Clayton	-0.7125	0.1589	0.1584	-1.7868	-1.0240	-0.4010	94.7844
	Frank	-0.7111	0.1596	0.1601	-1.5865	-1.0239	-0.3983	94.8692
	GH	-0.7125	0.1611	0.1633	-1.7907	-1.0282	-0.3968	95.5128
	Joe	-0.7111	0.1617	0.1639	-1.5849	-1.0280	-0.3942	95.3668
$\beta_{12}^{(S)} = 0.4$	AMH	0.4020	0.0913	0.0961	0.5080	0.2230	0.5810	93.9577
	Clayton	0.4002	0.0911	0.0960	0.0597	0.2218	0.5787	94.2828
	Frank	0.3988	0.0916	0.0968	-0.3012	0.2193	0.5783	94.7686
	GH	0.3991	0.0921	0.0992	-0.2233	0.2185	0.5797	93.0769
	Joe	0.3977	0.0927	0.0997	-0.5820	0.2160	0.5794	92.9215

From Table 5.16, when correctly fitting the YP class under a nonparametric baseline, the ARB for short-term parameters is larger than the ones in Tables 5.14 and 5.15, but still always lower than 3% for fitted PE models and 8% for BP models, even when fitting the wrong copula. However, for (incorrectly) fitted Weibull models, ARB and CR results are poor: for the first covariate, the ARB is always higher than 30%, leading to

CR values below 70%. Again, for semiparametric models, fitting the correct copula has produced, in general, smaller ARB values and closer CR values to the confidence level, specially when fitting the PE baseline. When looking the results for long-term parameters in Table 5.17, conclusions are similar to those obtained for short-term parameters. Along with results in Table 5.16, for the wider YP class, fitted semiparametric models (BP and PE) perform better than (incorrectly) fitted Weibull models, given copula generated data with marginal EW baseline distribution. Figures 5.6 (for PH and PO model classes) and 5.7 (for YP) show the boxplots for the relative bias on regression parameter estimates for fitted survival copula models, divided by baseline, over AMH EW generated data, given Scenario S1 ($n = 500$; $\tau = 0.25$).

Table 5.17: MC statistics for 1st margin long-term regression parameter estimates of fitted survival copula models over AMH EW YP generated data ($n = 500$; $\tau = 0.25$)

Parameter	Copula	Weibull YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(L)} = 0.8$	AMH	1.5848	1.4085	1.2081	98.0972	-1.1837	4.3373	88.5751
	Clayton	1.4957	0.5825	1.2353	86.9635	0.3539	2.6375	88.4615
	Frank	1.7089	1.8974	1.5321	113.6081	-2.0099	5.4277	90.3969
	GH	1.8217	2.6840	1.7635	127.7085	-3.4389	7.0822	89.1892
	Joe	1.7977	2.6400	1.8409	124.7111	-3.3923	6.9564	91.8159
$\beta_{12}^{(L)} = -0.6$	AMH	-0.6478	0.1324	0.1224	-7.9588	-0.9073	-0.3882	95.0000
	Clayton	-0.6401	0.1286	0.1174	-6.6816	-0.8921	-0.3881	95.6410
	Frank	-0.6428	0.1393	0.1276	-7.1338	-0.9159	-0.3698	95.7746
	GH	-0.6269	0.1426	0.1329	-4.4868	-0.9063	-0.3475	95.8816
	Joe	-0.6221	0.1434	0.1348	-3.6821	-0.9031	-0.3411	95.9132
Parameter	Copula	BP YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(L)} = 0.8$	AMH	0.8385	0.3274	0.4899	4.8121	0.1860	1.4694	95.1856
	Clayton	0.8122	0.3103	0.4292	1.5309	0.1956	1.4120	94.0704
	Frank	0.8570	0.3364	0.6558	7.1276	0.1859	1.5045	95.3954
	GH	0.8562	0.4646	0.5860	7.0267	-0.0543	1.7667	95.7529
	Joe	0.8594	0.5793	0.5502	7.4200	-0.2763	1.9946	95.6466
$\beta_{12}^{(L)} = -0.6$	AMH	-0.6398	0.1257	0.1261	-6.6401	-0.8865	-0.3938	93.6810
	Clayton	-0.6227	0.1226	0.1232	-3.7775	-0.8629	-0.3825	94.6787
	Frank	-0.6432	0.1291	0.1295	-7.1932	-0.8961	-0.3902	93.5000
	GH	-0.6325	0.1304	0.1310	-5.4141	-0.8881	-0.3769	94.0798
	Joe	-0.6325	0.1307	0.1314	-5.4090	-0.8889	-0.3766	94.1101
Parameter	Copula	PE YP Fitting						
		AE	SDE	ASE	ARB (%)	ALB	AUB	CR (%)
$\beta_{11}^{(L)} = 0.8$	AMH	0.8352	0.3320	0.3398	4.4056	0.1845	1.4860	96.5760
	Clayton	0.8138	0.3130	0.3264	1.7189	0.2003	1.4272	95.7874
	Frank	0.8446	0.3448	0.3545	5.5735	0.1687	1.5204	97.1831
	GH	0.8426	0.3412	0.3376	5.3285	0.1740	1.5113	97.0513
	Joe	0.8501	0.3460	0.3431	6.2629	0.1720	1.5283	97.0399
$\beta_{12}^{(L)} = -0.6$	AMH	-0.5922	0.1223	0.1172	1.2967	-0.8319	-0.3526	96.3746
	Clayton	-0.5795	0.1188	0.1153	3.4173	-0.8124	-0.3466	95.5868
	Frank	-0.5955	0.1253	0.1195	0.7452	-0.8411	-0.3500	96.3783
	GH	-0.5908	0.1261	0.1225	1.5272	-0.8381	-0.3436	95.6410
	Joe	-0.5946	0.1273	0.1230	0.8981	-0.8440	-0.3452	95.7529

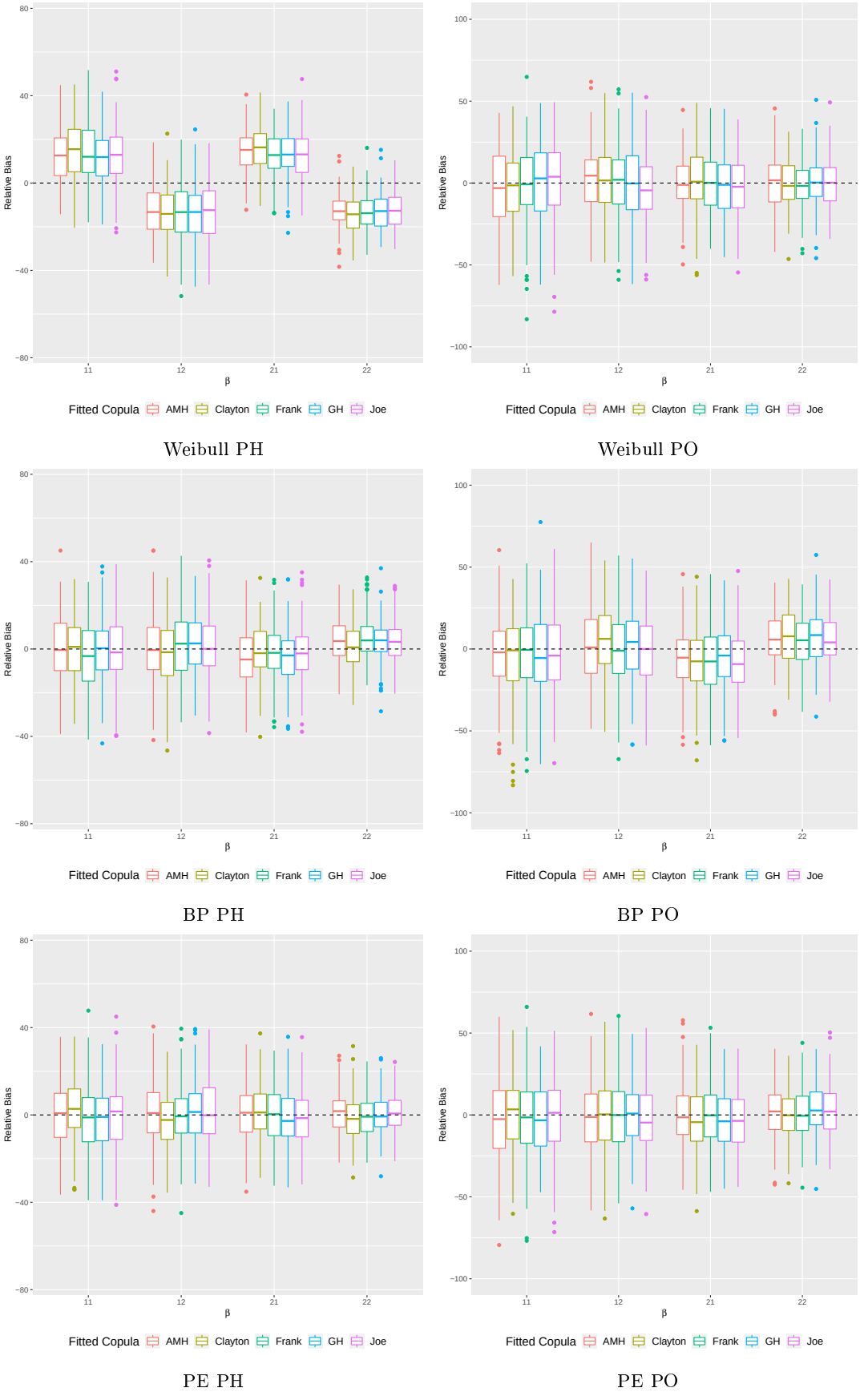
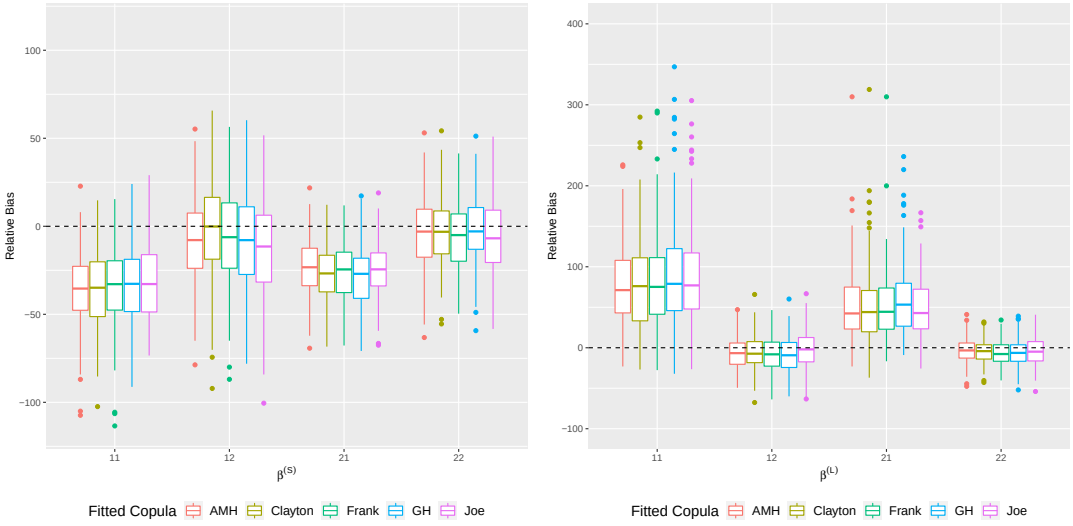
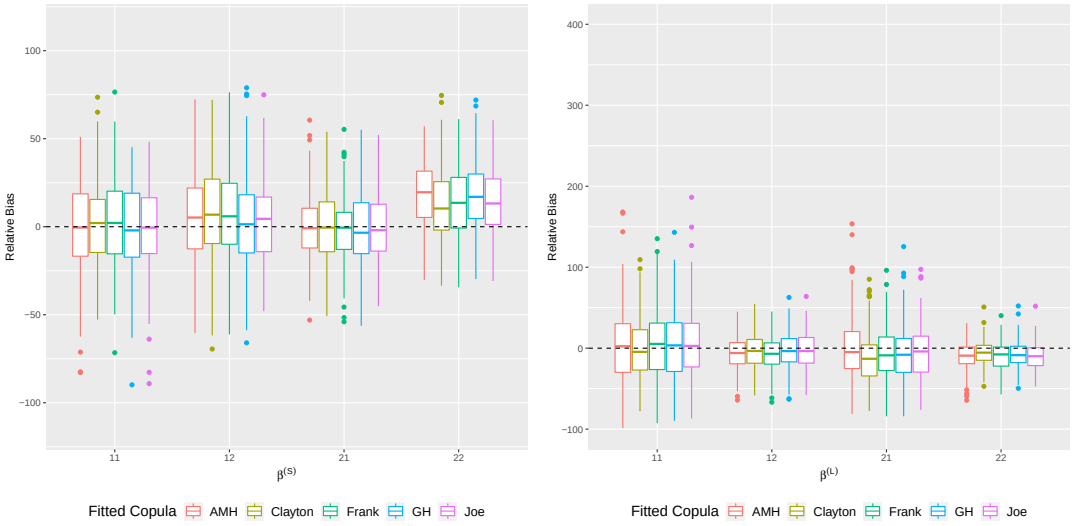


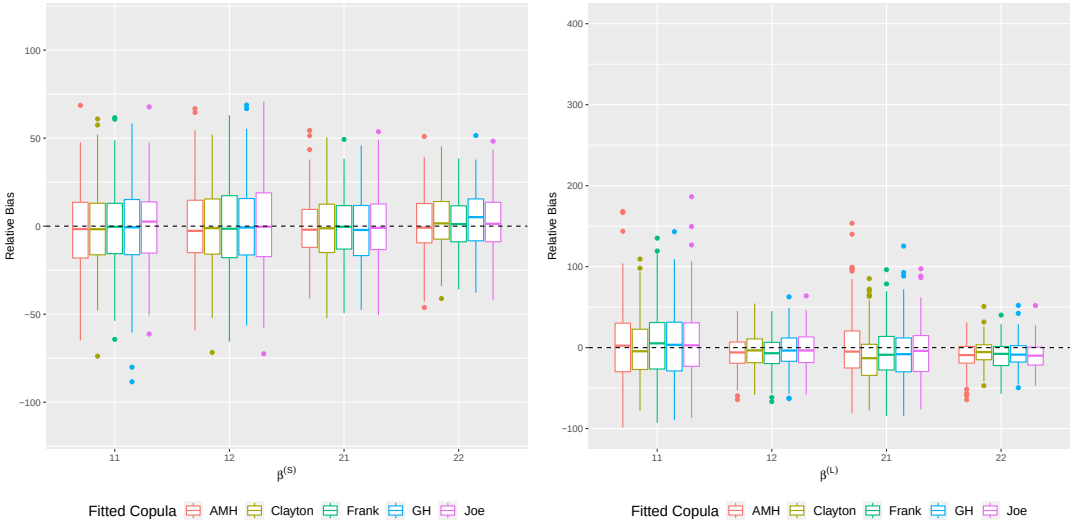
Figure 5.6: Relative bias on regression parameter estimates for fitted survival copula models over AMH EW (PH or PO) generated data ($n = 500$; $\tau = 0.25$).



Weibull YP (short) Weibull YP (long)



BP YP (short) BP YP (long)



PE YP (short) PE YP (long)

Figure 5.7: Relative bias on regression parameter estimates for fitted survival copula models over AMH EW YP generated data ($n = 500$; $\tau = 0.25$).

For fitted semiparametric models, the boxplots for the majority of regression parameters with the correct copula fitting had, in general, smaller range or fewer outlier count. Fitted PE models seem to have slightly better results for relative bias on regression parameter estimates than BP models, specially for the PO class. On the other hand, the boxplots from (incorrectly) fitted Weibull models, given the PH and YP classes, are far from being centered to the null value, even for the correct copula fitting.

Given generated data from an Archimedean copula model with marginal EW baseline distribution, again fitted models with an incorrect copula show results close to the correct one for regression parameters (regarding the ARB and CR) under a PH, PO or YP structure, when marginally fitting a semiparametric baseline distribution. In that case, the choice of a copula for fitting has little impact on estimation of regression parameters. However, choosing a wrong parametric specification for the baseline distribution can lead to poor estimation on regression parameters. This was expected, since the hazard function for the Weibull model cannot accommodate non-monotone forms, as those that can arise from marginally generated EW models.

5.2.2 Akaike Information Criteria

For copula generated data with marginal EW baseline distribution, results on the AIC of fitted BP and PE

Table 5.18: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW PH margins ($n = 500$; $\tau = 0.25$)

True	Fitted	Weibull PH		BP PH		PE PH	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	905.73	34.01	909.44	67.30	859.54	68.60
	Clayton	904.69	62.04	912.33	22.30	862.73	19.80
	Frank	913.69	3.95	915.33	10.40	865.25	11.60
	GH	941.99	0.00	949.11	0.00	897.45	0.00
	Joe	952.82	0.00	959.67	0.00	910.99	0.00
Clayton	Clayton	898.76	97.02	899.48	86.10	849.32	83.40
	AMH	902.90	2.72	903.05	13.40	852.46	16.00
	Frank	917.27	0.26	916.67	0.50	866.17	0.60
	GH	945.09	0.00	951.43	0.00	898.52	0.00
	Joe	955.16	0.00	961.45	0.00	911.71	0.00
Frank	Frank	903.39	65.32	906.36	83.40	856.10	85.70
	AMH	905.55	26.10	910.52	14.90	860.55	12.90
	Clayton	908.98	7.82	920.23	1.00	869.80	0.90
	GH	928.44	0.76	936.19	0.70	884.28	0.50
	Joe	940.57	0.00	947.78	0.00	897.85	0.00
GH	GH	873.88	90.75	879.71	71.90	835.03	70.90
	AMH	905.11	0.13	913.51	0.20	864.46	0.10
	Clayton	910.95	0.25	925.06	0.10	875.56	0.00
	Frank	892.74	8.24	900.66	27.60	850.48	28.40
	Joe	882.09	0.63	888.56	0.20	843.11	0.60
Joe	Joe	833.34	29.69	841.11	19.10	796.33	37.30
	AMH	901.10	0.00	909.95	0.00	860.09	0.00
	Clayton	910.85	0.00	926.90	0.00	875.94	0.00
	Frank	877.35	0.13	887.86	22.30	836.71	22.30
	GH	831.08	70.18	838.38	58.60	795.97	40.40

models corroborate to choose the correct copula, as seen from Tables 5.18 to 5.20, except when generating from the Joe copula. In that case, fitted Frank and GH copulas have non-negligible or even higher proportions of choice, while fitted GH copula models also present smaller mean AIC values, even under the YP class. When taking other fitted models with the correct copula, the highest proportions of choice are observed for the Clayton and Frank copulas. On the other hand, for fitted models with the (incorrect) Weibull baseline, AIC results induce to the choice of Clayton copula instead of AMH when generating from the last one, although leading to the correct copula fitting for Clayton, Frank or GH generated data.

For results in Table 5.18 for correctly fitted PH models, and taking also the correct copula fitting, the Joe copula presents the smallest mean AIC values (narrowly losing to fitted GH models over Joe generated data), given each fitted nonparametric baseline distribution. This conclusion is the same for correctly fitted PO models in Table 5.19. Inside each (correct) copula with marginal PH class, fitted PE models exhibit by far the smallest mean AIC values, while fitted Weibull models perform slightly better than BP ones. Changing to the marginal PO class, once again fitted PE models exhibit the smallest mean AIC values, but now fitted BP models clearly perform better than Weibull ones.

Table 5.19: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW PO margins ($n = 500$; $\tau = 0.25$)

True	Fitted	Weibull PO		BP PO		PE PO	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	815.06	18.47	792.56	71.70	755.36	70.50
	Clayton	814.24	79.87	796.03	19.60	758.29	18.10
	Frank	827.00	1.66	799.35	8.70	761.37	11.40
	GH	857.86	0.00	835.71	0.00	796.08	0.00
	Joe	868.58	0.00	847.99	0.00	810.98	0.00
Clayton	Clayton	808.05	99.61	782.04	86.30	744.41	85.90
	AMH	813.07	0.26	785.62	13.30	748.29	13.70
	Frank	831.17	0.13	802.07	0.40	763.53	0.40
	GH	861.15	0.00	839.67	0.00	798.32	0.00
	Joe	870.79	0.00	851.03	0.00	813.27	0.00
Frank	Frank	815.65	51.58	792.98	83.30	755.35	86.40
	AMH	816.49	28.88	797.16	15.40	760.11	12.70
	Clayton	818.47	19.04	807.79	0.80	770.21	0.60
	GH	843.35	0.50	824.00	0.50	784.32	0.30
	Joe	855.62	0.00	836.66	0.00	798.59	0.00
GH	GH	785.44	90.11	767.94	70.80	734.21	69.90
	AMH	815.76	0.25	800.33	0.30	764.36	0.10
	Clayton	820.55	0.38	813.15	0.00	776.89	0.00
	Frank	804.59	9.13	787.22	28.50	751.25	28.70
	Joe	794.75	0.13	777.23	0.40	743.13	1.30
Joe	Joe	745.21	17.61	730.38	20.70	695.93	36.60
	AMH	812.18	0.00	797.71	0.00	759.91	0.00
	Clayton	821.26	0.00	816.21	0.00	777.52	0.00
	Frank	789.53	0.13	775.02	22.30	737.14	22.30
	GH	741.49	82.26	727.71	57.00	695.62	41.10

Similar results to those obtained for the PH class are observed in Table 5.20 for correctly fitted YP models, but now the gap between Weibull and BP mean AIC is a bit larger. Having more marginal regression

parameters, fitted YP models possess again the lowest mean AIC when compared to fitted PH and PO classes, given any combination of copula and baseline distribution, when fitting the correct regression structure.

Table 5.20: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW YP margins ($n = 500$; $\tau = 0.25$)

True	Fitted	Weibull YP		BP YP		PE YP	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	747.29	25.22	760.94	71.10	720.23	70.50
	Clayton	745.76	73.63	764.42	19.50	723.36	18.50
	Frank	757.80	1.15	767.78	9.40	726.24	11.00
	GH	787.88	0.00	806.15	0.00	762.19	0.00
	Joe	800.04	0.00	817.86	0.00	777.21	0.00
Clayton	Clayton	738.31	98.70	749.94	85.40	709.60	85.60
	AMH	742.98	1.30	753.99	14.40	713.19	14.20
	Frank	761.62	0.00	770.33	0.20	729.10	0.20
	GH	790.49	0.00	808.30	0.00	763.97	0.00
	Joe	802.17	0.00	819.70	0.00	778.96	0.00
Frank	Frank	746.74	57.25	762.06	83.50	721.45	85.50
	AMH	747.55	31.15	766.39	15.50	726.34	13.70
	Clayton	751.49	11.10	776.96	0.60	735.91	0.50
	GH	771.96	0.50	793.08	0.40	750.36	0.30
	Joe	785.56	0.00	805.68	0.00	764.70	0.00
GH	GH	717.57	91.40	735.00	71.60	698.76	70.20
	AMH	748.38	0.13	768.72	0.20	729.69	0.10
	Clayton	754.07	0.38	781.10	0.00	741.32	0.00
	Frank	737.14	7.08	755.41	28.10	716.00	28.70
	Joe	726.33	1.01	744.60	0.10	707.78	1.00
Joe	Joe	678.42	31.23	698.14	18.10	661.12	35.50
	AMH	745.59	0.00	766.69	0.00	725.67	0.20
	Clayton	756.43	0.00	784.82	0.00	742.78	0.00
	Frank	723.87	0.13	743.68	22.50	702.14	22.20
	GH	676.54	68.64	695.33	59.40	660.71	42.10

Although the majority of highest proportions of choice by the smallest AIC point out, in general, to the correct copula choice, it is possible to identify the same pattern (verified earlier for generated data with Weibull baseline) of fitted copulas with non-negligible proportions of choice, regardless of fitted baseline or regression structure. Therefore, the same two copula groups (one involving Frank, AMH and Clayton, and the other composed by GH and Joe) that seem to capture similar behaviors of dependence are once again defined.

5.2.3 Correlation Estimates

As well as observed AIC results for generated data with marginal EW baseline distribution, the correct copula choice yields, in general, the best MC Kendall's τ estimates, as seen from Tables 5.21 to 5.23, except when generating from GH or Joe copulas. For those two cases, fitted models with the correct copula present non-negligible (and negative) ARB values, even when fitting a nonparametric baseline, regardless of the generated regression model class. This possibly evidences a difficulty on identifying the dependence parameter over more general behaviors for the marginal hazard function, with respect to some copulas.

Table 5.21: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW PH margins ($n = 500$; $\tau = 0.25$)

True	Copula	Weibull PH		BP PH		PE PH	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.2692	7.6685	0.2450	-1.9970	0.2506	0.2438
	Clayton	0.2720	8.7965	0.2311	-7.5745	0.2386	-4.5715
	Frank	0.2269	-9.2588	0.2253	-9.8928	0.2370	-5.1902
	GH	0.1151	-53.9482	0.0955	-61.8159	0.1353	-45.8630
	Joe	0.0571	-77.1447	0.0430	-82.8188	0.0663	-73.4804
Clayton	Clayton	0.2772	10.8926	0.2446	-2.1504	0.2524	0.9446
	AMH	0.2788	11.5088	0.2730	9.2034	0.2793	11.7282
	Frank	0.2164	-13.4423	0.2178	-12.8868	0.2298	-8.0825
	GH	0.1034	-58.6357	0.0865	-65.4156	0.1268	-49.2837
	Joe	0.0467	-81.3315	0.0345	-86.1940	0.0539	-78.4380
Frank	Frank	0.2459	-1.6421	0.2380	-4.8003	0.2487	-0.5051
	AMH	0.2461	-1.5459	0.2212	-11.5032	0.2255	-9.7840
	Clayton	0.2674	6.9786	0.2165	-13.4000	0.2230	-10.7841
	GH	0.1414	-43.4516	0.1176	-52.9479	0.1563	-37.4759
	Joe	0.0835	-66.5869	0.0653	-73.8774	0.0950	-62.0043
GH	GH	0.2013	-19.4865	0.1750	-30.0019	0.2143	-14.2728
	AMH	0.2356	-5.7770	0.2098	-16.0892	0.2141	-14.3517
	Clayton	0.2662	6.4946	0.2087	-16.5322	0.2153	-13.8865
	Frank	0.2673	6.9242	0.2516	0.6332	0.2645	5.8170
	Joe	0.1469	-41.2225	0.1234	-50.6293	0.1637	-34.5352
Joe	Joe	0.1901	-23.9626	0.1623	-35.0645	0.2085	-16.6136
	AMH	0.2232	-10.7372	0.1979	-20.8271	0.2015	-19.3810
	Clayton	0.2658	6.3373	0.1993	-20.2615	0.2058	-17.6903
	Frank	0.2907	16.2822	0.2688	7.5022	0.2829	13.1531
	GH	0.2403	-3.8876	0.2104	-15.8259	0.2503	0.1352

Given any fitted model with the correct copula, among AMH, Clayton, and Frank, the lowest ARB values were observed when fitting the PE model as the baseline distribution (always below 1% in magnitude), followed by the BP and Weibull models, regardless of the regression structure. Although not presenting good ARB values for correctly fitted models with GH or Joe copulas (always above 10%), the PE baseline still performs better than (incorrect) Weibull and BP models (in that order). Comparing only the semiparametric models for those cases, using the BP baseline produces an ARB about twice the obtained when using the PE model (a bit more under the PH regression structure, and less under PO or YP classes).

Table 5.22: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW PO margins ($n = 500$; $\tau = 0.25$)

True	Copula	Weibull PO		BP PO		PE PO	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.2737	9.4981	0.2467	-1.3137	0.2496	-0.1653
	Clayton	0.2711	8.4380	0.2286	-8.5671	0.2347	-6.1307
	Frank	0.2227	-10.9094	0.2309	-7.6308	0.2401	-3.9497
	GH	0.1057	-57.7121	0.1044	-58.2329	0.1395	-44.2022
	Joe	0.0495	-80.1828	0.0463	-81.4680	0.0675	-73.0074
Clayton	Clayton	0.2764	10.5568	0.2453	-1.8737	0.2516	0.6336
	AMH	0.2865	14.6094	0.2775	10.9822	0.2802	12.0855
	Frank	0.2122	-15.1248	0.2248	-10.0944	0.2347	-6.1226
	GH	0.0944	-62.2345	0.0961	-61.5624	0.1323	-47.0846
	Joe	0.0402	-83.9202	0.0372	-85.1124	0.0552	-77.9196
Frank	Frank	0.2418	-3.2838	0.2407	-3.7264	0.2485	-0.6146
	AMH	0.2484	-0.6586	0.2204	-11.8350	0.2225	-10.9926
	Clayton	0.2651	6.0267	0.2107	-15.7333	0.2157	-13.7379
	GH	0.1318	-47.2975	0.1247	-50.1255	0.1576	-36.9556
	Joe	0.0745	-70.2033	0.0692	-72.3315	0.0955	-61.8100
GH	GH	0.1909	-23.6432	0.1820	-27.2177	0.2132	-14.7030
	AMH	0.2368	-5.2942	0.2085	-16.6036	0.2104	-15.8425
	Clayton	0.2623	4.9357	0.2010	-19.5860	0.2061	-17.5774
	Frank	0.2621	4.8592	0.2538	1.5231	0.2621	4.8452
	Joe	0.1358	-45.6651	0.1297	-48.1014	0.1630	-34.8036
Joe	Joe	0.1781	-28.7560	0.1698	-32.0668	0.2068	-17.2698
	AMH	0.2242	-10.3358	0.1959	-21.6590	0.1970	-21.1959
	Clayton	0.2608	4.3090	0.1893	-24.2804	0.1938	-22.4926
	Frank	0.2856	14.2578	0.2705	8.2015	0.2786	11.4376
	GH	0.2296	-8.1666	0.2167	-13.3032	0.2473	-1.0807

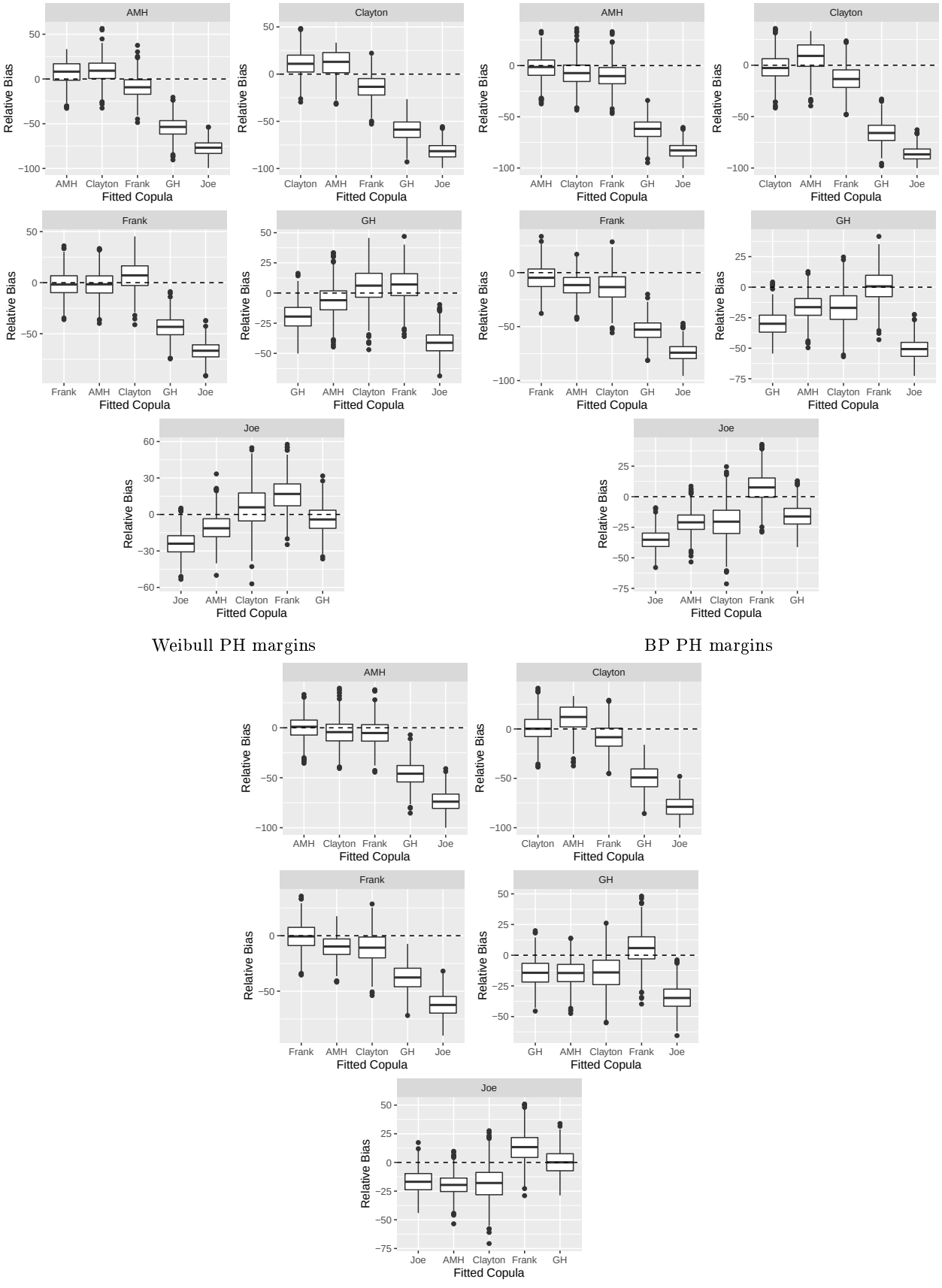
Figures 5.8 to 5.10 show boxplots for the relative bias on Kendall's τ estimation for all fitted survival copula models in Scenario S1 ($n = 500$; $\tau = 0.25$) over generated copula data with EW baseline for each margin. Given any regression model class, results for their associated boxplots confirms those obtained for point estimation and ARB from Tables 5.21 to 5.23: fitted Weibull and semiparametric models (specially these last ones) with the correct copula have boxplots well-centered in the null line for generated data from AMH, Clayton and Frank copulas. As expected, the same cannot be said for any fitted baseline model when correctly fitting the GH and Joe copulas: for fitted Weibull and BP models, their corresponding boxplots barely touch the null line, while for fitted PE models only a small part from their interquartile range is around it. On the other hand, given any copula for generation, the boxplot dispersion for the relative bias of each fitted copula model is similar, in general, across all five Archimedean copula models, even if choosing a wrong one.

From Tables 5.21 to 5.23 and also from Figures 5.8 to 5.10, it is possible to conclude that the choice of a correct Archimedean copula is necessary to ensure an appropriate estimation of the Kendall's τ correlation if marginal survival times were generated by a more general process, but it is not sufficient depending on the true copula function. However, when looking to mean AIC values and proportions of choice for correctly fitted GH and Joe copulas from tables 5.18 to 5.20 it can be said that, if τ estimates alone are not the best, those models are better fitted if compared to almost all wrong copula choices, regardless of the fitted baseline distribution.

Finally, fitted PE models performed exceptionally better than the (also semiparametric) BP and the (incorrect) Weibull models given any correct copula fitting, over generated data with EW margins. This contrasts with results obtained when generating marginally with the Weibull baseline, for which BP models had better fitting than PE ones for the majority of copula and regression model class combinations (although still being outperformed to fitted Weibull models in general). That said, there is no immediate response for which semiparametric model is better when the generator process of copula data and marginal baseline distribution are unknown, but both BP and PE have proven to be useful.

Table 5.23: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW YP margins ($n = 500$; $\tau = 0.25$)

True	Copula	Weibull YP		BP YP		PE YP	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.2681	7.2207	0.2480	-0.8038	0.2502	0.0607
	Clayton	0.2606	4.2588	0.2310	-7.6188	0.2356	-5.7568
	Frank	0.2235	-10.6162	0.2322	-7.1067	0.2410	-3.6106
	GH	0.1209	-51.6409	0.1021	-59.1694	0.1394	-44.2551
	Joe	0.0596	-76.1624	0.0446	-82.1799	0.0669	-73.2306
Clayton	Clayton	0.2672	6.8785	0.2476	-0.9431	0.2521	0.8494
	AMH	0.2980	19.2168	0.2813	12.5057	0.2815	12.5824
	Frank	0.2141	-14.3685	0.2266	-9.3537	0.2359	-5.6588
	GH	0.1099	-56.0272	0.0943	-62.2957	0.1326	-46.9619
	Joe	0.0484	-80.6363	0.0354	-85.8384	0.0545	-78.2067
Frank	Frank	0.2402	-3.9398	0.2416	-3.3708	0.2489	-0.4216
	AMH	0.2395	-4.2144	0.2212	-11.5299	0.2230	-10.8022
	Clayton	0.2525	1.0081	0.2134	-14.6327	0.2174	-13.0395
	GH	0.1458	-41.6675	0.1226	-50.9588	0.1572	-37.1380
	Joe	0.0865	-65.3876	0.0675	-73.0175	0.0946	-62.1548
GH	GH	0.2006	-19.7675	0.1800	-27.9929	0.2125	-15.0124
	AMH	0.2267	-9.3075	0.2098	-16.0970	0.2112	-15.5357
	Clayton	0.2481	-0.7549	0.2044	-18.2425	0.2079	-16.8314
	Frank	0.2571	2.8342	0.2552	2.0885	0.2627	5.0931
	Joe	0.1477	-40.9160	0.1274	-49.0377	0.1618	-35.2800
Joe	Joe	0.1899	-24.0296	0.1673	-33.0780	0.2061	-17.5799
	AMH	0.2148	-14.0708	0.1970	-21.1825	0.1977	-20.9172
	Clayton	0.2452	-1.9213	0.1938	-22.4956	0.1961	-21.5458
	Frank	0.2780	11.2183	0.2720	8.8198	0.2794	11.7567
	GH	0.2370	-5.1906	0.2150	-13.9848	0.2468	-1.2794

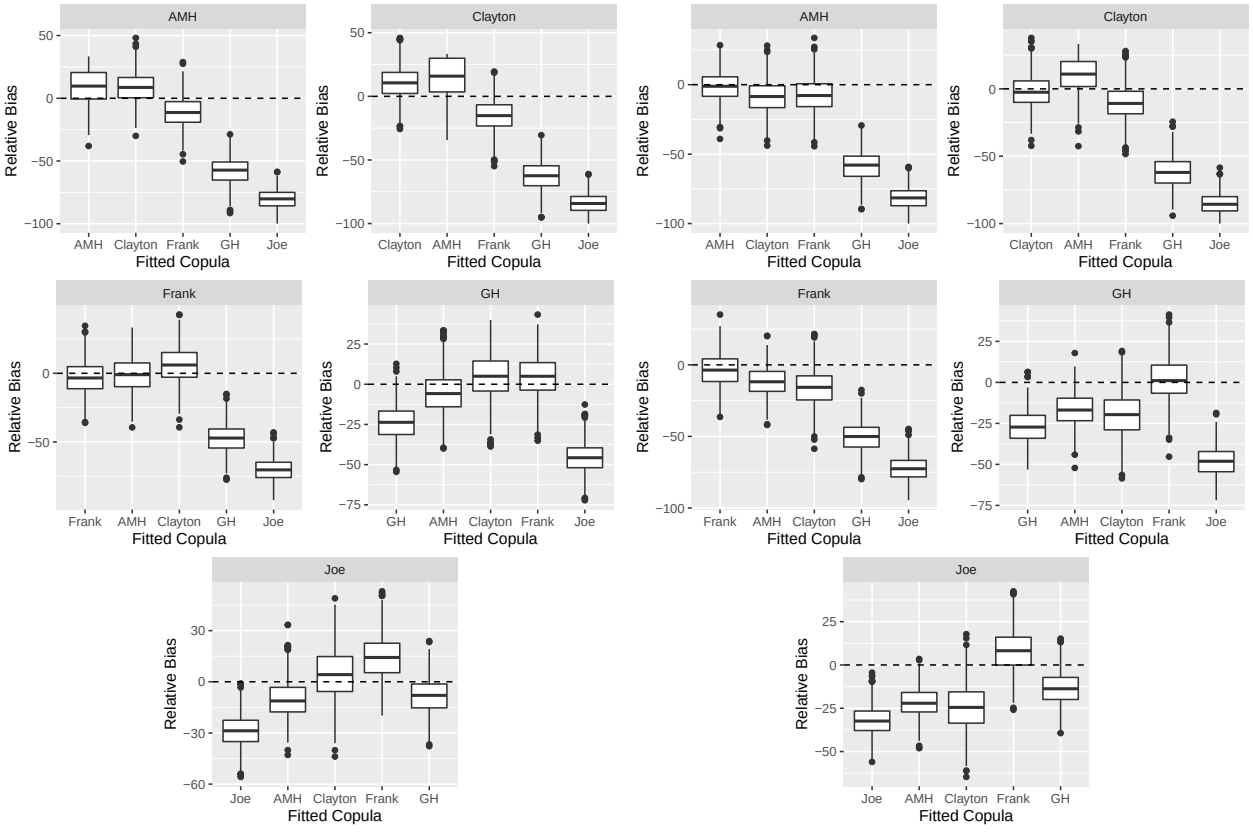


Weibull PH margins

BP PH margins

PE PH margins

Figure 5.8: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW PH margins ($n = 500$; $\tau = 0.25$)

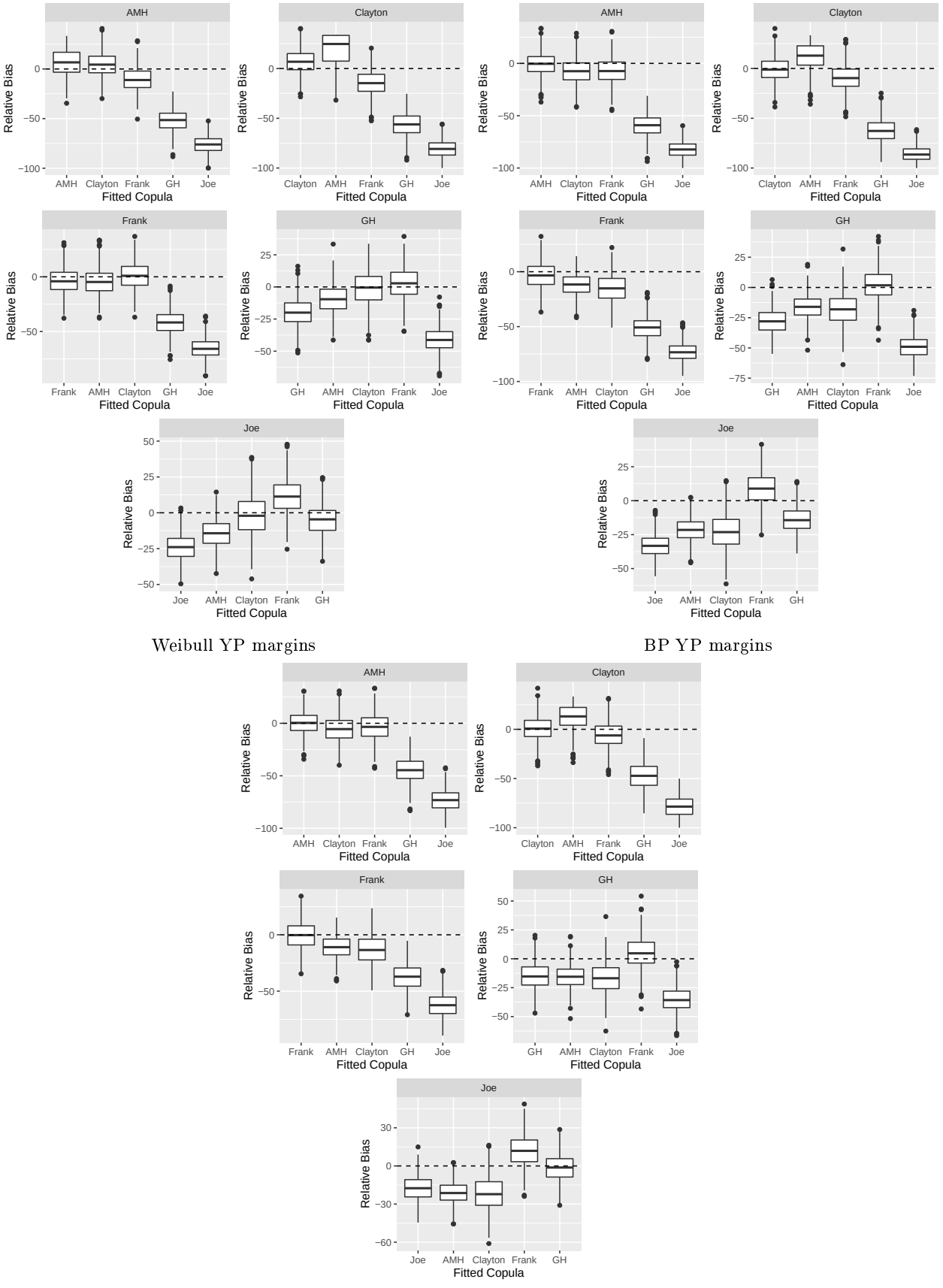


Weibull PO margins

BP PO margins

PE PO margins

Figure 5.9: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW PO margins ($n = 500$; $\tau = 0.25$)



Weibull YP margins

BP YP margins

PE YP margins

Figure 5.10: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW YP margins ($n = 500$; $\tau = 0.25$)

5.2.4 Likelihood Ratio Tests

For copula generated data with EW margins, the analysis through LR tests will not account for fitted Weibull models: for the purpose of this work, evaluation of nested models is done with respect to the regression model class. Therefore, it is restricted to fitted semiparametric models. Again, p-values for all LR statistics will be obtained from a $\chi^2_{(4)}$. Tables 5.24 to 5.26 present results for average statistics and p-values from LR tests on fitted BP and PE models over generated data with the EW baseline (under the correct copula fitting).

Table 5.24: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW PH margins ($n = 500; \tau = 0.25$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	6.1175	0.3425	4.2940	0.4717
	PO	26.0845	0.0172	30.8931	0.0009
Clayton	PH	5.6297	0.3771	4.0687	0.4891
	PO	26.4718	0.0124	31.3756	0.0008
Frank	PH	5.9444	0.3527	4.1757	0.4726
	PO	25.9596	0.0192	31.2596	0.0007
GH	PH	5.4472	0.3895	4.1775	0.4853
	PO	24.9161	0.0176	29.9006	0.0013
Joe	PH	5.6756	0.3807	4.4143	0.4622
	PO	23.1901	0.0283	28.8543	0.0023

On Table 5.24, when generating from the PH class, all LR tests accept fitted YP models against the (incorrect) PO class, for any combination of fitted nonparametric baseline and copula, as expected. The same cannot be said for LR tests confronting fitted PH and YP models, whose results are always non-significant, leading to the choice of (correctly and more parsimonious) fitted PH models. Thus, even on a more general baseline distribution for marginal survival times and given a large sample size, introducing more regression parameters to capture short and long-term covariate effects does not provide a significantly better fitting.

Table 5.25: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW PO margins ($n = 500; \tau = 0.25$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	27.9217	0.0071	26.7120	0.0051
	PO	1.0745	0.7665	4.4347	0.4495
Clayton	PH	28.9169	0.0040	27.9565	0.0038
	PO	0.9686	0.7644	4.4032	0.4478
Frank	PH	28.0047	0.0062	26.6746	0.0043
	PO	1.1426	0.7550	4.3725	0.4536
GH	PH	28.0819	0.0072	26.9723	0.0045
	PO	1.7151	0.7208	4.6114	0.4355
Joe	PH	30.1212	0.0061	28.6870	0.0051
	PO	1.7149	0.7188	4.6527	0.4233

Looking now for Table 5.25, this time generating from the PO class, the converse is also true: all LR tests accept fitted YP models against the (incorrect) PH regression structure, given any combination of fitted nonparametric baseline and copula. However, the same cannot be said for LR tests confronting fitted PO and YP models. Their results lead towards the choice of PO class, since its associated models are more parsimonious. Finally, for Table 5.26, when generating from the wider YP class, all LR tests accept fitted YP models against the PH or PO regression structures, given any combination of fitted nonparametric baseline and copula function, as occurred for marginally generated data from the Weibull baseline.

Table 5.26: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW YP margins ($n = 500; \tau = 0.25$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	118.2559	< 0.0001	111.9872	< 0.0001
	PO	86.1503	< 0.0001	83.1002	< 0.0001
Clayton	PH	124.1241	< 0.0001	117.2673	< 0.0001
	PO	91.4135	< 0.0001	87.9554	< 0.0001
Frank	PH	118.5429	< 0.0001	111.6295	< 0.0001
	PO	88.7532	< 0.0001	84.9454	< 0.0001
GH	PH	117.8587	< 0.0001	111.5238	< 0.0001
	PO	87.7842	< 0.0001	83.9300	< 0.0001
Joe	PH	120.8868	< 0.0001	114.6986	< 0.0001
	PO	90.1515	< 0.0001	86.4063	< 0.0001

Tables 5.24 to 5.26 showed that the analysis through the LR test for nested regression model classes, also when generating from an Archimedean survival copula with marginal EW baseline distribution (regardless of the fitted copula function or nonparametric baseline model) is a useful tool to choose the regression structure for fitting if the one that generated the data is unknown, even on a more general behavior for the true marginal baseline distribution.

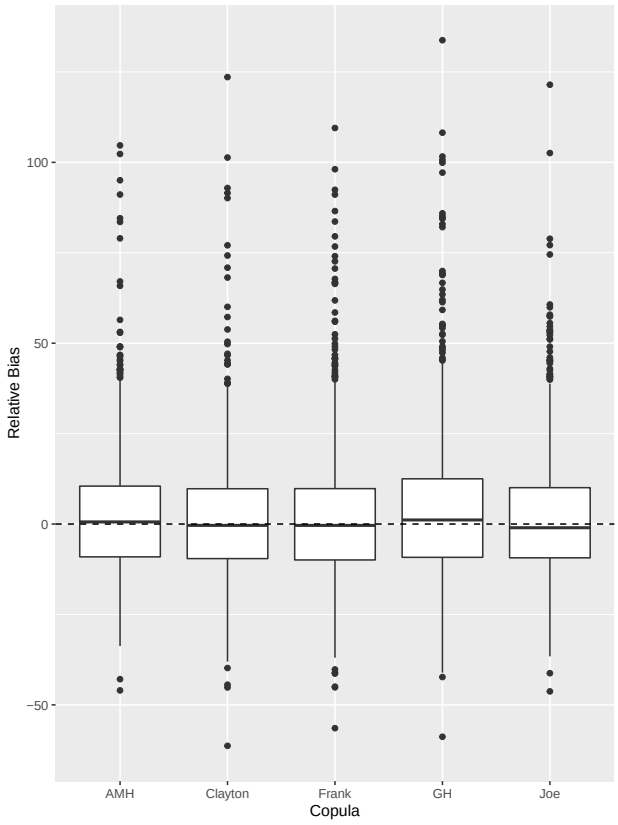
5.3 Crossing Time Estimation

Both PH and PO classes allows a feasible interpretation on regression parameter estimates if their corresponding assumptions are valid. However, neither of them account for the situation where, given two levels of a covariate (*e.g.*, the indicator for a treatment), the associated survival functions cross each other. It might be of interest to estimate such crossing survival time t^*. This is possible for fitted models with the YP class, but it is not straightforward to find an interval estimate for t^*, since the standard error of $\widehat{t^*}$ has no closed form expression (Demarqui and Mayrink, 2021). The usual solution is to implement a numerical procedure to find the root that solves the nonlinear equation $S_C(t^*) - S_T(t^*) = 0$, where $S_C(\cdot)$ and $S_T(\cdot)$ are the survival functions given control and treatment values, respectively. This can be done through nonparametric bootstrap, generating a set of new samples from the original data and fitting the same model for each bootstrap sample to obtain the associated parameter estimates and the quantities $\widehat{S}_C(t^*)$ and $\widehat{S}_T(t^*)$. To search the root for each marginal crossing survival time, the R command `uniroot` (see Brent, 1973 for more details) will be used.

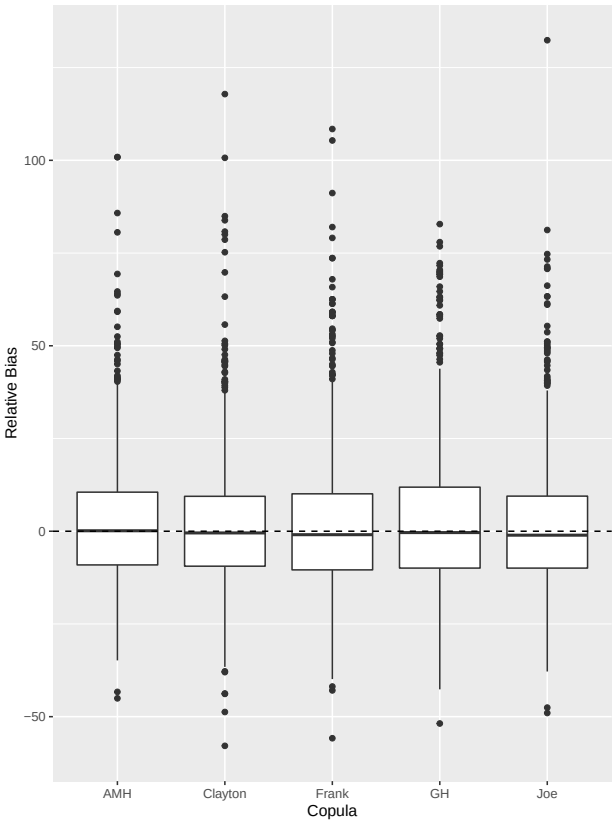
Table 5.27: MC statistics for marginal crossing time estimates of fitted survival copula models over generated copulas with Weibull YP margins ($n = 500$; $\tau = 0.25$)

Quantity	Copula	Weibull YP Fitting				
		AE	ARB (%)	ALB	AUB	CR (%)
t_1^*	AMH	2.2435	2.2842	1.5853	3.2383	95.4
	Clayton	2.2234	1.3651	1.5865	3.1295	95.2
	Frank	2.2268	1.5205	1.5680	3.2149	95.2
	GH	2.2649	3.2590	1.5835	3.3060	94.4
	Joe	2.2318	1.7471	1.5932	3.2586	95.3
t_2^*	AMH	1.5041	0.6694	1.2589	1.8630	95.0
	Clayton	1.4964	0.1520	1.2602	1.8160	94.7
	Frank	1.5108	1.1208	1.2626	1.8814	95.6
	GH	1.4963	0.1450	1.2460	1.8615	93.2
	Joe	1.5034	0.6257	1.2631	1.8663	94.7
Quantity	Copula	BP YP Fitting				
		AE	ARB (%)	ALB	AUB	CR (%)
t_1^*	AMH	2.2392	2.0862	1.5573	3.2609	96.0
	Clayton	2.2237	1.3778	1.5647	3.1783	95.1
	Frank	2.2275	1.5521	1.5468	3.2470	94.2
	GH	2.2444	2.3242	1.5516	3.3123	94.7
	Joe	2.2212	1.2645	1.5752	3.2832	96.1
t_2^*	AMH	1.5011	0.4668	1.2506	1.9020	95.1
	Clayton	1.4946	0.0342	1.2530	1.8560	94.7
	Frank	1.5118	1.1868	1.2554	1.9264	95.1
	GH	1.4944	0.0183	1.2391	1.8914	94.5
	Joe	1.5009	0.4540	1.2560	1.9069	94.7
Quantity	Copula	PE YP Fitting				
		AE	ARB (%)	ALB	AUB	CR (%)
t_1^*	AMH	2.2432	2.2674	1.5826	3.2278	96.3
	Clayton	2.2284	1.5944	1.5880	3.1294	94.9
	Frank	2.2407	2.1555	1.5835	3.2242	94.3
	GH	2.2655	3.2843	1.5871	3.3091	94.8
	Joe	2.2322	1.7665	1.6068	3.2489	95.5
t_2^*	AMH	1.5051	0.7345	1.2596	1.8560	94.3
	Clayton	1.4995	0.3626	1.2627	1.8138	94.3
	Frank	1.5164	1.4968	1.2653	1.8810	95.0
	GH	1.4955	0.0973	1.2456	1.8481	94.1
	Joe	1.5006	0.4381	1.2602	1.8531	95.0

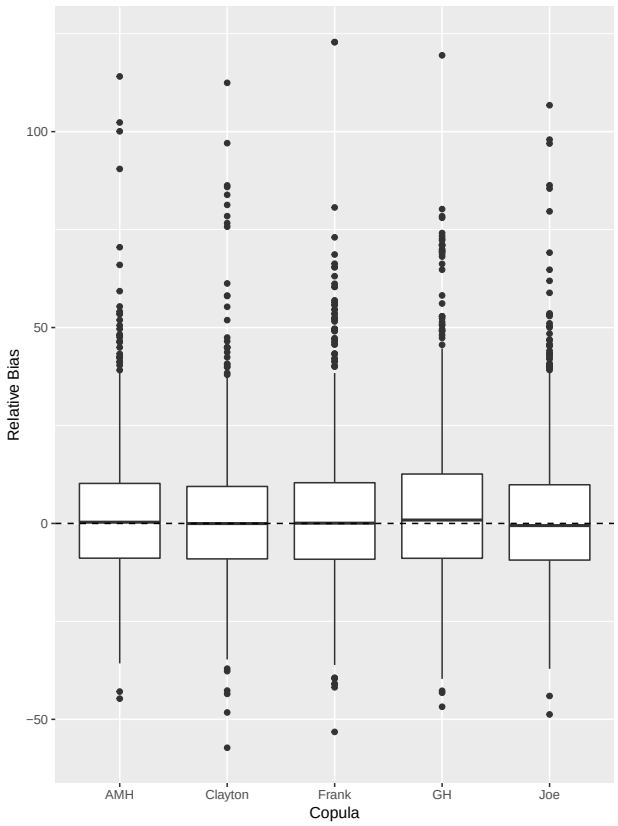
Due to the use of a resampling method for each MC replica, the estimation of crossing survival times is far more computationally intensive. Thus, the evaluation of marginal crossing time estimates through simulation for the survival copula models proposed here is restricted to a single correlation value, given copula generated data with Weibull YP margins and fitting only the correct copula. Therefore, consider a new MC simulation study with $M = 1000$ replications of copula data sets with Weibull YP margins and $n = 500$, using the same covariates and values for the baseline and regression parameters as before, and varying only the true Archimedean copula, but always with a fixed θ value such that $\tau = 0.25$. To estimate marginal crossing survival times associated to the treatment effect (dichotomous covariate), take two new subjects, control and treated, with covariate values $\boldsymbol{x}^*_{C;j} = (0, 0)$ and $\boldsymbol{x}^*_{T;j} = (1, 0)$, respectively, for $j = 1, 2$.



Weibull YP margins

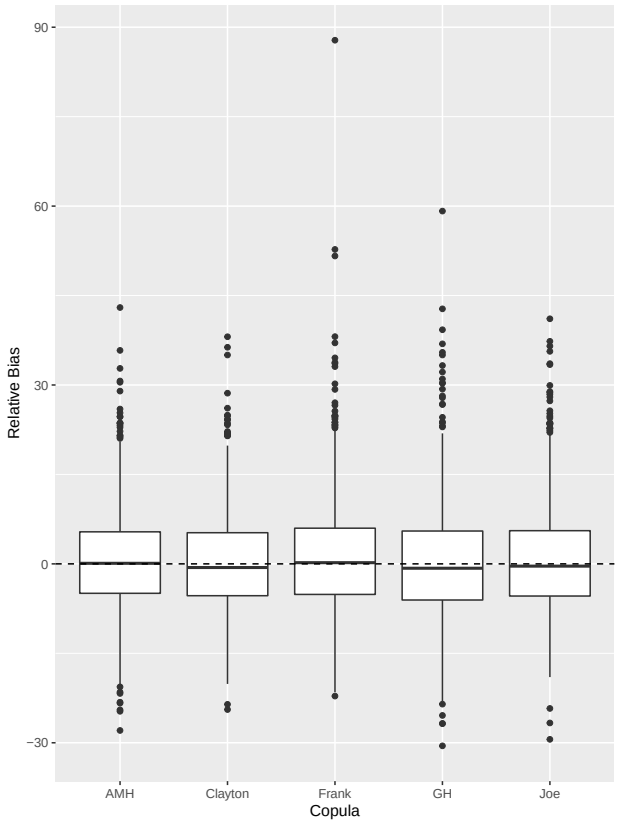


BP YP margins

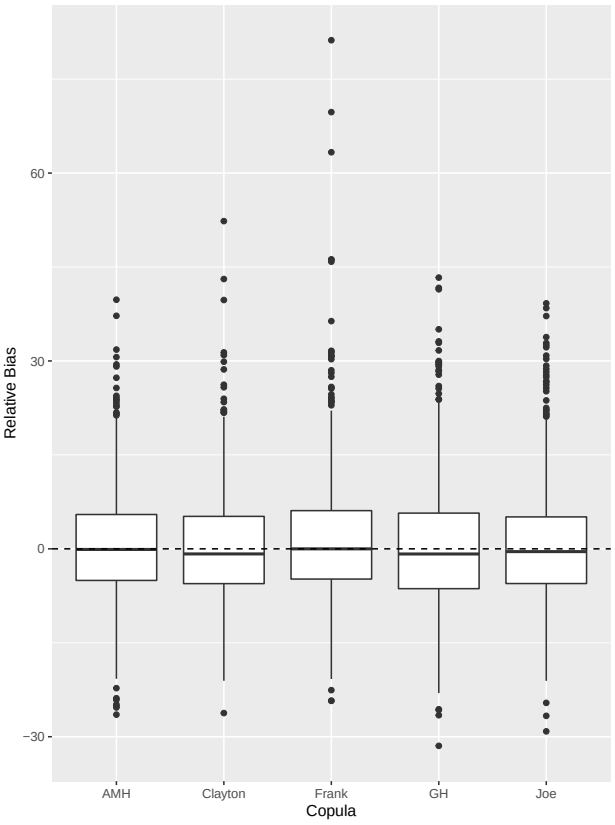


PE YP margins

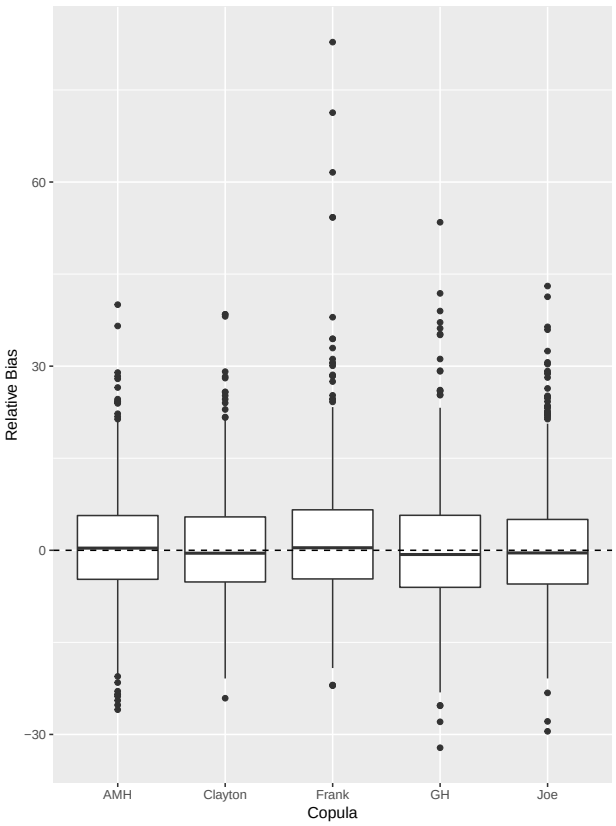
Figure 5.11: Crossing time estimates on the 1st margin for fitted survival copula models over data generated from the correct copula model with Weibull YP margins ($n = 500$; $\tau = 0.25$)



Weibull YP margins



BP YP margins



PE YP margins

Figure 5.12: Crossing time estimates on the 2nd margin for fitted survival copula models over data generated from the correct copula model with Weibull YP margins ($n = 500$; $\tau = 0.25$)

Note that the continuous covariate is set constant (here, equal to a reference level) for both subjects. To infer on marginal crossing survival times, a nonparametric bootstrap is applied over each MC replica, using a total of 1000 bootstrap samples to obtain their associated point and interval estimates (using the corresponding percentiles to a confidence level of 95%). Table 5.27 presents the MC results on the estimation of crossing survival times, divided for each copula and baseline distribution. As expected, the ARB was always lower than 4% for the first marginal crossing survival time (2% for the second one), and the CR is at most 0.02 (0.01) away from the confidence level of 95%, for fitted survival copula models with Weibull YP margins. However, the same can be said from the estimation for fitted BP and PE models, which can still overcome the Weibull model as seen for the GH copula. Thus, such semiparametric models obtain estimates as good as those from the correctly fitted Weibull models even for marginal crossing survival times, but without imposing any parametric functional form for the hazard function. Concerning the copula itself, changing only its expression (maintaining the true correlation and other unrelated parameters and quantities) has little effect over the estimated marginal crossing survival times. Also, none of them is far better or worse than another with respect to the ARB or CR values. Figures 5.11 and 5.12 corroborate the above conclusions from Table 5.27. For the relative bias of crossing survival time estimates, their corresponding boxplots are all well-centered around 0, regardless of fitted copula or baseline distribution. Concerning the margins, boxplot dispersion is greater the for 1st margin, which also seems to have more (positive) outlier counts. This result confirms the lower ARB and more accurate CR values observed for the 2nd margin in Table 5.27.

CHAPTER 6

A Real Data Application

This work addresses the study of a manually curated data collection of patients with ovarian cancer, described by Ganzfried et al. (2013). Their resource provides data for a total of 23 distinct studies, but a single one (TCGA), with a total of $n = 508$ subjects (after removing 49 of them with missing information on the tumor stage or treatment indicators, or null values for survival times), is considered as an application to fit the survival copula models proposed in Chapter 4. Each subject i, $i = 1, \ldots, n$, has 2 observed times, the first one representing a nonterminal event time (in this study, the time-tumor progression) T_i or a random, independent censored time A_i, and the second, a terminal event time (here, the overall survival of a subject) T_i^* or a censored time A_i (the censoring mechanism is always the same). This is a semi-competing risks study: a subject may experience two potential events: the nonterminal event (tumor progression) or the terminal event (death). The nonterminal event may not be observable due to an earlier occurrence of the terminal event. On the other hand, an occurrence of the nonterminal event does not change the observational condition for the terminal event. Therefore, the terminal event is a competing risk for the non-terminal event, but the reciprocal is not true (Wu et al., 2020). In order to fit a survival copula model, it is necessary to define what are the survival times and censoring statuses for each subject i at each copula margin j, $j = 1, 2$:

- The survival time for the 1st copula margin is $Y_{i1} = \min(T_i, T_i^*, A_i)$;
- The censoring status for the 1st copula margin is $\delta_{i1} = \mathbb{I}(Y_{i1} = T_i)$;
- The survival time for the 2nd copula margin is $Y_{i2} = \min(T_i^*, A_i)$;
- The censoring status for the 2nd copula margin is $\delta_{i2} = \mathbb{I}(Y_{i2} = T_i^*)$.

The failure rates were about 53,15% for the first margin and 50,39% for the second one. Defined the survival times and censoring statuses for each margin j, a set of covariates must be chosen for the regression

structure specification. From the TCGA study, two covariates are specified: *CXCL12*, a biomarker for the gene expression of a ovarian cancer (continuous, with concentration values standardized to have zero mean and unit variance), and *pltx*, an indicator for a platinum-based treatment (dichotomous, the reference level is the subject who did not receive this treatment). Note that this covariate specification is similar to the one defined for Monte Carlo simulations in Chapter 5.

First, survival copula models accounting for all possible combinations of bivariate Archimedean copulas (one of AMH, Clayton, Frank, GH or Joe copulas), baseline distributions (one among Weibull, BP or PE models), and regression model classes (one of PH, PO or YP models) are fitted in order to obtain their AIC values, which are shown in Table 6.1. Thus, results on the regression parameter and Kendall's τ estimates will be presented for the "best" combination (given the AIC) of copula, baseline and regression structures.

Table 6.1: AIC values for fitted survival copula models on cancer ovarian data

Copula	PH			PO			YP		
	Weibull	BP	PE	Weibull	BP	PE	Weibull	BP	PE
AMH	8621.72	8488.80	8492.15	–	8485.22	8499.18	–	8493.96	8482.57
Clayton	–	8489.49	8492.82	–	8485.99	8499.25	–	8494.60	8482.56
Frank	–	8482.81	8485.61	–	8477.73	8492.78	–	8487.71	8473.49
GH	8602.52	8475.95	8479.39	–	8471.22	8485.91	–	8480.06	**8467.78**
Joe	–	8484.92	8487.88	–	8480.05	8494.15	–	8489.00	8476.54

For almost all combinations specifying the Weibull model for the baseline distribution, the survival copula fitting fails. Even when the log-likelihood is successfully maximized, the AIC values are higher when compared to all other fitted models with a semiparametric baseline distribution. This is an evidence of a generator process with non-monotonic hazard function for survival times at each margin. All copulas are close to each other concerning the AIC criterion, but fitted models with the GH copula have the lowest AIC values, regardless of (semiparametric) baseline distribution or regression model class. Given all fitted models with the GH copula, the GH PE YP has the lowest AIC values compared to all other marginal specifications for baseline and regression models.

Table 6.2: LR tests for nested models against the GH PE YP model on cancer ovarian data

Fitted Model	Log-lik.	LR stat.	DF	P-value
GH PE YP	-4224.89	–	–	–
GH PE PH	-4234.70	19.62	4	5.94×10^{-4}
GH PE PO	-4237.96	26.14	4	2.97×10^{-5}

Now, when taking the Likelihood Ratio (LR) test statistic for the GH PE YP model against its nested models (those last ones as the null hypothesis for each test) with the same copula and baseline specification (*i.e.*, the GH PE PH and GH PE PO models) in Table 6.2, the LR statistic is significant at the level of 5% (1 minus the confidence level) for both tests. Thus, there is no evidence to not reject the GH PE YP model instead of any nested model more parsimonious with respect to the regression model class. Hence, results on the regression parameter coefficients (point estimate, standard error, lower and upper limits of the 95% confidence interval, Z-statistics and p-values) for the GH PE YP model are presented in Table 6.3.

Table 6.3: Regression parameter results for the GH PE YP model on cancer ovarian data

Margin	Covariate	Coef.	Estimate	SE	Lower	Upper	Z-stat.	P-value
1st	*CXCL12*	$\widehat{\beta}_{11}^{(S)}$	0.0900	0.0620	-0.0316	0.2116	1.4505	0.1469
	CXCL12	$\widehat{\beta}_{11}^{(L)}$	1.8903	0.7537	0.4131	3.3675	2.5081	0.0121
	pltx	$\widehat{\beta}_{12}^{(S)}$	-1.7650	0.3898	-2.5289	-1.0011	-4.5286	5.94×10^{-6}
	pltx	$\widehat{\beta}_{12}^{(L)}$	8.1402	14.1470	-19.5874	35.8678	0.5754	0.5650
2nd	*CXCL12*	$\widehat{\beta}_{21}^{(S)}$	0.2372	0.1087	0.0241	0.4503	2.1817	0.0291
	CXCL12	$\widehat{\beta}_{21}^{(L)}$	-0.1160	0.1414	-0.3932	0.1612	-0.8205	0.4119
	pltx	$\widehat{\beta}_{22}^{(S)}$	-1.2938	0.4460	-2.1680	-0.4196	-2.9007	0.0037
	pltx	$\widehat{\beta}_{22}^{(L)}$	-0.4670	0.6098	-1.6621	0.7281	-0.7659	0.4437

Given the GH PE YP model in Table 6.3, except for $\widehat{\beta}_{11}^{(S)}$, all other short-term regression coefficients were significant at the level of 5%. However, the long-term counterpart of $\widehat{\beta}_{11}^{(S)}$, $\widehat{\beta}_{11}^{(L)}$, has significance. When interpreting the significant regression coefficients for the GH PE YP model, it can be said that the ratio of hazard rates between a treated subject and a control for tumor progression (nonterminal event) is $\exp(-1.7650) \approx 0.1712$ (or 17.12%), *i.e.*, the treatment (*pltx*) reduces the hazard rate in 82.88% for a tumor progression. Also, the ratio of hazard rates between a treated subject and a control for death (terminal event) is $\exp(-1.2938) \approx 0.2742$ (or 27.42%). In other words, the treatment reduces the hazard rate in 72.58% for a death. On the other hand, each level gained for the *CXCL12* biomarker increases the hazard rate in $\exp(1.8903) \approx 6.6214$ times in the long-term for a tumor progression, and in $\exp(0.2372) \approx 1.2677$ times in the short-term for a death. Therefore, the platinum treatment reduces the hazard for both events in the short-term, but does not have significant influence in the long-term. However, greater levels of *CXCL12* increases the hazard for both events, but in distinct moments of the follow-up.

Concerning the dependence estimation, for the fitted GH PE YP survival copula model, the estimated Kendall's correlation is $\widehat{\tau} = 0.3332$, with confidence interval $I_{\widehat{\tau}} = [0.1576, 0.4483]$. This significant correlation gives an estimated upper tail dependence $\widehat{\chi}_U = 2 - 2^{1/0.3332} \approx -6.0067$. In other words, marginal survival times have moderate overall dependence given a same subject, but they also have a moderate upper tail dependence, thus implying on a mild correlation across smaller survival times. For greater marginal survival times, there is no lower tail dependence, since it is always null for GH copulas.

CHAPTER 7

Final Remarks and Future Research

The present work proposed a new theoretical framework for semiparametric modeling on clustered survival data through the introduction of copulas under independent random censoring. Each margin of a copula model was given by a survival time from a cluster under study, modeled by a combination of two components: a survival model for a baseline function, and a regression model class to accommodate explanatory variables. Hence, a wider and flexible survival copula model class was proposed, with programmed functions for generation, fitting and methods for inference on the `copSurv` R package.

Three models were considered to fit the baseline distribution: the parametric Weibull model, vastly used in the literature; the nonparametric Bernstein Polynomials (BP) model, which uses a polynomial approximation for the baseline function, and the nonparametric Piecewise Exponential (PE) model, which approximates the baseline function through constant hazard functions for each interval from a partition on the time axis.

For the regression structure, the traditional Proportional Hazards (PH); the Proportional Odds (PO) and the Yang-Prentice (YP) models were addressed. The first two assume a proportionality relationship for survival times of different subjects conditional on the baseline functional form. For PH models, the proportionality is in the hazard function, while for PO models it is in the odds function. The YP model was originally proposed to accommodate crossing survival curves, but it also contains both PH and PO as particular cases.

Although many copulas developed in the literature can be used for a survival copula model, this work focused on five models from the Archimedean copula class: the Ali-Mikhail-Haq (AMH); Clayton; Frank; Gumbel-Hougaard (GH) and Joe copula models, obtaining results from simulated data for all of them. Archimedean copulas have been widely studied and enjoys from well-established properties that fulfill an important role in the (survival) copula likelihood function evaluation, and even for dependence measures.

It should be noted that some combinations arisen from the models above for a joint survival copula modeling

are a novelty in the literature, in particular those involving the BP model for the baseline distribution or the PO and YP regression model classes. To evaluate and compare all proposed (bivariate) survival copula models, an extensive simulation study was realized. All simulated data was generated from a bivariate survival copula model with two covariates for each margin. The generation could vary on the copula used (one of the five Archimedean copulas aforementioned); on the survival model for the baseline distribution on both margins (from Weibull or EW models), or on the regression model class (PH, PO or YP). For a fixed sample size $n = 500$, simulation scenarios also were divided by three chosen values for the true Kendall's τ correlation ($\tau \in \{0.25, 0.5, 0.75\}$). In its turn, the proposed survival copula models for fitting over simulated data were always a combination of a copula function (again, one of the five Archimedean copulas), a baseline distribution (one of Weibull, BP or PE models), and a regression model class (one of PH, PO or YP classes). For parameter estimation and choice through the mean AIC, fitted models were compared by exchanging the copula function or baseline distribution used for fitting, while for an analysis through the LR test they were compared by swapping the regression model class. Finally, all proposed survival copula models for fitting were applied to a set of bivariate real data, in order to choose the best one through the AIC combined to an LR test analysis, and also present its results on regression parameter and Kendall's τ correlation estimates.

One of the main goals of this work was to compare results on regression parameter estimates, specially in terms of the Average Relative Bias (ARB) and Coverage Rate (CR), and mean AIC, among each Archimedean copula used for fitting, given a fixed copula for generation of the simulated data, and verify if a correct copula fitting has suitable results under a high sample size for three distinct levels of dependence. Illustrating for the AMH copula in the main text and presenting results when generating from other copulas in the Appendix, fitting the correct copula has generally produced good estimates (when looking for the ARB and CR), although for many cases a wrong fitting of the copula model can still have similar performance, even for a scenario with higher correlation. In general, fitting a wrong copula does not lead to a severe loss of performance for the regression parameter estimation, but this in fact occurs for the computed AIC and τ estimates, regardless of the regression model class.

Given only fitted models with the correct copula and fixed the regression model class, another main goal of this work was to compare results for three baseline distributions, two of them offering a nonparametric appeal. As expected, when generating copula data from the Weibull baseline, fitting the same distribution provided the best results, although fitted models with a nonparametric baseline (BP or PE) follow closely in terms of ARB. On the other hand, when generating copula data from the EW baseline, PE and BP models perform better than Weibull ones for the majority of copulas. This was expected due to their nonparametric nature: their number of baseline parameters depends on the sample size, thus fitting well when that size is high, including when the true marginal hazard functions have a non-monotonic behavior.

Finally, given the correct copula fitting and now exchanging the regression model class, the last goal involved an analysis through the LR test to check if the YP class is preferable or no over the more parsimonious (and nested) PH and PO classes. Given one of the smaller regression structures for marginal survival data generation, the YP class fits significantly better if tested against the other not used for generation, but it does not if tested against the true one, regardless of the true or fitted baseline distribution and copula function. On the other

hand, if the YP class is also part of the generator process for marginal survival data, LR tests always accept it against both PH and PO classes.

Based on the conclusions set out above, the next steps for a future work include:

- Obtain results for fitted survival copula models under a Bayesian approach of inference, using appropriate criteria and comparing with results already presented under the frequentist approach in this work;
- Study, through a sensitivity analysis, more conservative choices for the polynomial degree in BP models when marginally fitted on a survival copula model. Since copulas themselves introduce at least one parameter for dependence and multiplies each marginal parameterization by the copula dimension, it is reasonable to suppose that smaller degrees for each margin can ensure a more accurate estimation;
- Implement, to survival copula models, the usual residual analysis for survival data, in order to find potentially outlier or influential subjects, and redefine the correct covariate specification, potentially including the use of non-linear or time-dependent covariates;
- Introduce, among all margins from a survival copula model, distinct choices on the baseline distribution (for non-nested parametric survival models), number of baseline parameters (for BP and PE nonparametric models), regression model class (allowing parsimonious choices for some, but not necessarily all, copula margins), and number of covariates for estimation of regression parameters;
- Extend marginal survival models to also incorporate cure fraction estimation when the marginal survival function seems to stop at a positive lower limit s_j, after a time t_j for each copula margin $j = 1, \ldots, d$ (see chapter 6 of Klein et al., 2013 for an insight on cure fraction models);
- Extend the developed combinations of baseline and regression structures in this work for a joint frailty-copula modeling, when the heterogeneity can come from a known source other than the subjects themselves, such as interviews from a small number of distinct studies (see Wu et al., 2020 for an example);
- Provide in the CRAN repository the `copSurv` R package, already in development, for all survival copula models treated in this work and their associated methods for inference.

Bibliography

Aalen, O. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6:701–726, 1978.

Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19: 716–723, 1974.

Ali, M. M., Mikhail, N. N., and Haq, M. S. A class of bivariate distributions including the Bivariate Logistic. *Journal of Multivariate Analysis*, 8:405–412, 1978.

Andersen, P. K. and Keiding, N. *Encyclopedia of Biostatistics 6: Survival Analysis*. Wiley, 1998.

Barry, D. and Hartingan, J. A. Product partition model for change point problems. *The Annals of Statistics*, 20:260–279, 1992.

Bartle, R. G. and Sherbert, D. R. *Introduction to Real Analysis*. John Wiley & Sons, 4$^{\text{th}}$ edition, 2011.

Bennett, S. Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2:273–277, 1983.

Bernstein, S. N. Démonstration du théorème de weierstrass fondée sur le calcul des probabilités. *Communications of the Kahrkov Mathematical Society*, 13:1–2, 1913.

Bernstein, S. N. Sur les fonctions absolument monotones. *Acta Mathematica*, 52:1–66, 1928.

Box, G. E. P. and Cox, D. R. An analysis of transformations. *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 26:211–252, 1964.

Brent, R. P. *Algorithms for Minimization Without Derivatives*. Prentice-Hall, 1973.

Carnicer, J. M. and Peña, J. M. Shape preserving representations and optimality of the Bernstein basis. *Advances in Computational Mathematics*, 1:173–196, 1993.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Bru-baker, M., Guo, J., Li, P., and Riddell, A. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76: 1–32, 2017.

Chang, I. S., Hsiung, C. A., Wu, Y. J., and Yang, C. C. Bayesian survival analysis using Bernstein polynomials. *Scandinavian Journal of Statistics*, 32:447–466, 2005.

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

Collett, D. *Modelling Survival Data in Medical Research*. CRC Press, 3rd edition, 2015.

Cooray, K. Strictly Archimedean copulas with complete association for multivariate dependence based on the Clayton family. *Dependence Modeling*, 6:1–18, 2018.

Cox, D. R. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 34:187–220, 1972.

Demarqui, F. N. and Mayrink, V. D. Yang and Prentice model with piecewise exponential baseline distribution for modeling lifetime data with crossing survival curves. *Brazilian Journal of Probability and Statistics*, 35: 172–186, 2021.

Demarqui, F. N., Dey, D. K., Loschi, R. H., and Colosimo, E. A. Fully semiparametric Bayesian approach for modeling survival data with cure fraction. *Biometrical Journal*, 56:198–218, 2014.

Duchateau, L. and Janssen, P. *The Frailty Model*. Springer New York, 2007.

Durante, F. and Sempi, C. *Principles of Copula Theory*. Taylor & Francis, 2015.

Emura, T. and Michimae, H. *A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae*, volume 24. 2017.

Farouki, R. T. The Bernstein polynomial basis: A centennial retrospective. *Computer Aided Geometric Design*, 29:379–419, 2012.

Farouki, R. T. and Rajan, V. On the numerical condition of polynomials in Bernstein form. *Computer Aided Geometric Design*, 4:191–216, 1987.

Fleming, T. R. and Harrington, D. P. *Counting Processes and Survival Analysis*. Wiley, 1991.

Flores, M. U., Artero, E. D. A., Durante, F., and Sánchez, J. F. *Copulas and Dependence Models with Applications*. Springer, 2017.

Frank, M. J. On the simultaneous associativity ot F(x, y) and x + y- F(x, y). *Aequationes Mathematicae*, 19: 194–226, 1979.

Frank, M. J. The solution of a problem of alsina, and its generalization. *Aequationes Mathematicae*, 73:37–38, 1981.

Friedman, M. Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, 10: 101–113, 1982.

Féron, R. Sur les tableaux de corrélation dont les marges sont données, cas de l'espace à trois dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 5:3–12, 1956.

Ganzfried, B. F., Riester, M., Haibe-Kains, B., Risch, T., Tyekucheva, S., Jazic, I., Wang, X. V., Ahmadifar, M., Birrer, M. J., Parmigiani, G., Huttenhower, C., and Waldron, L. curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome. https://doi.org/10.1093/database/bat013, 2013. [Database (Oxford); Online; accessed 06-December-2021].

Goethals, K., Janssen, P., and Duchateau, L. Frailties and copulas, not two of a kind. *Risk and Decision Analysis*, 3:247–253, 2012.

Gumbel, E. J. Distributions des valeurs extrêmes en plusiers dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, 9:171–173, 1960.

Hacking, I. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge University Press, 2006.

Hoeffding, W. Massstabinvariante korrelationstheorie. *Schriften des Mathematischen Seminars und des Instituts für Angewandte Mathematik der Universität Berlin*, 5:181–233, 1940.

Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. *Elements of Copula Modeling with R*. Springer, 2018.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. *copula: Multivariate Dependence with Copulas*, 2020. URL https://CRAN.R-project.org/package=copula. R package version 0.999-20.

Hoffman, M. D. and Gelman, A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15:1593–1623, 2014.

Hougaard, P. A class of multivariate failure time distributions. *Biometrika*, 21:671–678, 1986.

Hu, T., Zhou, Q., and Sun, J. Regression analysis of bivariate current status data under the proportional hazards model. *The Canadian Journal of Statistics*, 45:410–424, 2017.

Ibrahim, J. G., Chen, M. H., and Sinha, D. *Bayesian Survival Analysis*. Springer Science & Business Media, 2001.

Joe, H. Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis*, 46:262–282, 1993.

Joe, H. *Multivariate Models and Dependence Concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 1997.

Joe, H. *Dependence Modeling with Copulas*, volume 134 of *Monographs on Statistics and Applied Probability*. CRC Press, 2014.

Kalbfleisch, J. D. and Prentice, R. L. Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60:267–278, 1973.

Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure Time Data*. Wiley, 2$^{\text{th}}$ edition, 2002.

Kaplan, L., E and Meier, P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

Klein, J. P. and Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer New York, 2$^{\text{th}}$ edition, 2005.

Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G., and Scheike, T. H. *Handbook of Survival Analysis*. CRC Press, 2013.

Knaus, J. *snowfall: Easier cluster computing (based on snow).*, 2015. URL https://CRAN.R-project.org/package=snowfall. R package version 1.84-6.1.

Kumar, P. Probability Distributions and Estimation of Ali-Mikhail-Haq Copula. *Applied Mathematical Sciences*, 4:657–666, 2010.

Lawless, J. F. *Statistical Models and Methods for Lifetime Data*. Wiley, 2$^{\text{th}}$ edition, 2011.

Lorentz, G. G. *Bernstein Polynomials*. American Mathematical Society, 1986.

Louzada, F., Suzuki, A. K., and Cancho, V. G. The FGM Long-Term Bivariate Survival Copula Model: Modeling, Bayesian Estimation, and Case Influence Diagnostics. *Communications in Statistics - Theory and Methods*, 42:673–691, 2013.

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

Marshall, A. W. and Olkin, I. Families of multivariate distributions. *Journal of the American Statistical Association*, 83:834–841, 1988.

Mello e Silva, J. F. D. Piecewise exponential model for long-term survival data. Master's thesis, Departamento de Estatística – ICEx – UFMG, 2016.

Mudholkar, G. S. and Srivastava, D. K. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42:299–302, 1993.

Nelsen, R. *An Introduction to Copulas*. Springer, 2$^{\text{th}}$ edition, 2006.

Nocedal, J. and Wright, S. *Numerical Optimization*. Springer New York, 2006.

Nolan, J. P. *Univariate Stable Distributions*. Springer, 2020.

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

Osman, M. and Ghosh, S. K. Nonparametric regression models for right-censored data using Bernstein polynomials. *Computational Statistics & Data Analysis*, 56:559–573, 2012.

Panaro, R. V. spsurv: An R package for semiparametric survival analysis. Master's thesis, Departamento de Estatística – ICEx – UFMG, 2020.

Prenen, L., Braekers, R., and Duchateau, L. Extending the Archimedean copula methodology to model multivariate survival data grouped in clusters of variable size. *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 79:483–505, 2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

Rickert, J. Survival Analysis with R. https://rviews.rstudio.com/2017/09/25/survival-analysis-with-r/, 2017. [Online; accessed 03-March-2020].

Schneider, S. *Modelos para dados de sobrevivência multivariados com censura informativa*. PhD thesis, Departamento de Estatística – ICEx – UFMG, 2017.

Schneider, S., Demarqui, F. N., Colosimo, E. A., and Mayrink, V. D. An approach to model clustered survival data with dependent censoring. *Biometrical Journal*, 62:157–174, 2020.

Schweizer, B. *Advances in Probability Distributions with Given Marginals*, chapter Thirty years of copulas, pages 13–50. Kluwer Academic Publishers, 1991.

Schweizer, B. and Wolff, E. F. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9:879–885, 1981.

Shih, J. H. and Louis, T. A. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51:1384–1399, 1995.

Sklar, A. Fonctions de répartition á n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8:229–231, 1959.

Stan Development Team. *RStan: the R interface to Stan*, 2020a. URL http://mc-stan.org/. R package version 2.19.3.

Stan Development Team. Stan. https://mc-stan.org/, 2020b. [Online; accessed 30-March-2020].

Therneau, T. M. *A Package for Survival Analysis in R*, 2021. URL https://CRAN.R-project.org/package=survival. R package version 3.2-13.

Weibull, W. A statistical distribution of wide applicability. *Journal of Applied Mechanics*, 18:293–297, 1951.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.

Wu, B., Michimae, H., and Emura, T. Meta-analysis of individual patient data with semi-competing risks under the Weibull joint frailty–copula model. *Computational Statistics*, 35:1525–1552, 2020.

Xu, D., Zhao, S., Hu, T., Yu, M., and Sun, J. Regression analysis of informative current status data with the semiparametric linear transformation model. *Journal of Applied Statistics*, 46:187–202, 2019.

Yang, S. and Prentice, R. Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. *Biometrika*, 92:1–17, 2005.

Appendix

APPENDIX A

Additional Results for Generated Copulas with Weibull Baseline

A.1 Akaike Information Criteria

Results in this appendix for the AIC consist of tables on two statistics, the mean AIC and the proportion of choice by the least AIC, for each fitted copula model (among AMH, Clayton, Frank, GH and Joe), given a specification for the baseline distribution (Weibull, BP or PE) and regression model class (PH, PO or YP), on Scenarios S2 ($n = 500, \tau = 0.5$) and S3 ($n = 500, \tau = 0.75$).

A.2 Correlation Estimates

Results for the Kendall's τ correlation estimates in this appendix comprehend tables of two statistics (AE and ARB), and boxplots of the relative biases, for all Archimedean copulas addressed in this work (AMH, Clayton, Frank, GH and Joe), given each combination of fitted baseline distribution (Weibull, BP or PE) and regression model class (PH, PO or YP), on Scenarios S2 and S3.

A.3 Likelihood Ratio Tests

Finally, results for the LR tests in this appendix comprehend tables of means for the LR statistic and its corresponding p-value for nested fitted models with respect to the regression model class, given each one of the five Archimedean copulas discussed in this work (AMH, Clayton, Frank, GH and Joe), combined with a fitted baseline distribution (Weibull, BP or PE), on Scenarios S2 and S3, given a regression model class (PH, PO or YP) used for marginal data generation.

Table A.1: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull PH margins ($n = 500$; $\tau = 0.5$)

True	Fitted	Weibull PH		BP PH		PE PH	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1580.73	16.80	1610.81	17.30	1611.73	14.90
	Clayton	1578.36	82.70	1610.46	82.20	1610.66	84.60
	Frank	1602.40	0.50	1634.17	0.50	1633.98	0.50
	GH	1633.77	0.00	1661.41	0.00	1657.19	0.00
	Joe	1659.58	0.00	1688.05	0.00	1684.02	0.00
Clayton	Clayton	1440.89	100.00	1472.84	100.00	1473.21	99.40
	AMH	1474.12	0.00	1509.16	0.00	1507.53	0.30
	Frank	1484.36	0.00	1515.23	0.00	1515.26	0.30
	GH	1551.93	0.00	1569.36	0.00	1564.09	0.00
	Joe	1606.30	0.00	1625.25	0.00	1619.73	0.00
Frank	Frank	1449.39	99.50	1481.15	99.80	1481.40	99.40
	AMH	1515.36	0.00	1545.28	0.00	1546.92	0.00
	Clayton	1500.82	0.00	1527.50	0.00	1527.75	0.00
	GH	1498.98	0.50	1518.92	0.20	1514.02	0.60
	Joe	1541.71	0.00	1556.16	0.00	1550.64	0.00
GH	GH	1382.26	97.30	1414.21	95.20	1416.06	90.60
	AMH	1523.14	0.00	1553.47	0.00	1554.91	0.00
	Clayton	1509.20	0.00	1535.12	0.00	1535.86	0.00
	Frank	1433.55	0.10	1464.37	0.10	1465.57	0.00
	Joe	1396.53	2.60	1426.23	4.70	1425.56	9.40
Joe	Joe	1300.54	98.30	1332.84	97.90	1336.24	98.70
	AMH	1537.29	0.00	1567.93	0.00	1568.54	0.00
	Clayton	1534.40	0.00	1556.46	0.00	1556.31	0.00
	Frank	1409.46	0.00	1438.03	0.00	1439.33	0.00
	GH	1320.72	1.70	1350.37	2.10	1357.67	1.30

Table A.2: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull PO margins ($n = 500$; $\tau = 0.5$)

True	Fitted	Weibull PO		BP PO		PE PO	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1464.57	15.20	1492.62	19.30	1497.99	14.70
	Clayton	1464.04	84.50	1493.02	80.10	1496.40	84.60
	Frank	1493.44	0.30	1520.64	0.60	1523.32	0.70
	GH	1529.47	0.00	1548.50	0.00	1546.55	0.00
	Joe	1557.16	0.00	1578.94	0.00	1577.23	0.00
Clayton	Clayton	1316.40	99.70	1345.43	100.00	1348.84	99.50
	AMH	1358.76	0.00	1384.25	0.00	1387.46	0.20
	Frank	1370.03	0.30	1393.86	0.00	1396.33	0.30
	GH	1447.01	0.00	1448.85	0.00	1445.53	0.00
	Joe	1505.85	0.00	1511.74	0.00	1508.18	0.00
Frank	Frank	1337.94	99.70	1367.22	99.80	1369.99	99.30
	AMH	1403.92	0.00	1433.04	0.00	1435.94	0.00
	Clayton	1393.15	0.20	1417.39	0.00	1420.26	0.00
	GH	1395.79	0.10	1404.98	0.20	1401.86	0.70
	Joe	1442.43	0.00	1443.49	0.00	1439.96	0.00
GH	GH	1270.66	99.40	1300.06	96.40	1304.90	90.10
	AMH	1411.40	0.00	1441.11	0.00	1444.70	0.00
	Clayton	1399.72	0.00	1425.60	0.00	1429.07	0.00
	Frank	1320.09	0.10	1350.00	0.00	1354.43	0.00
	Joe	1289.63	0.50	1313.10	3.60	1314.83	9.90
Joe	Joe	1191.99	97.20	1221.57	98.10	1227.66	99.20
	AMH	1428.72	0.00	1458.55	0.00	1461.70	0.00
	Clayton	1430.90	0.00	1453.68	0.00	1456.20	0.00
	Frank	1297.86	0.00	1327.18	0.00	1331.83	0.00
	GH	1209.46	2.80	1240.02	1.90	1250.92	0.80

Table A.3: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull YP margins ($n = 500$; $\tau = 0.5$)

True	Fitted	Weibull YP		BP YP		PE YP	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1391.61	16.70	1423.89	17.80	1424.75	14.10
	Clayton	1390.87	82.80	1422.23	81.70	1423.21	85.20
	Frank	1419.77	0.50	1450.42	0.50	1451.15	0.70
	GH	1452.44	0.00	1478.13	0.00	1474.24	0.00
	Joe	1481.63	0.00	1509.04	0.00	1505.30	0.00
Clayton	Clayton	1242.58	99.30	1273.94	100.00	1274.69	99.80
	AMH	1282.07	0.50	1313.24	0.00	1316.08	0.00
	Frank	1295.03	0.20	1324.00	0.00	1323.97	0.10
	GH	1364.99	0.00	1377.96	0.00	1372.05	0.10
	Joe	1427.29	0.00	1441.95	0.00	1435.87	0.00
Frank	Frank	1267.05	99.70	1298.67	99.80	1299.26	99.10
	AMH	1334.51	0.00	1364.32	0.00	1366.27	0.00
	Clayton	1324.07	0.20	1346.56	0.00	1347.20	0.20
	GH	1317.05	0.10	1335.68	0.20	1330.28	0.70
	Joe	1362.76	0.00	1373.77	0.00	1368.41	0.00
GH	GH	1199.30	98.30	1230.49	96.00	1233.67	90.60
	AMH	1340.76	0.00	1370.90	0.00	1373.07	0.00
	Clayton	1330.14	0.00	1353.64	0.00	1355.73	0.00
	Frank	1251.14	0.10	1280.34	0.00	1283.14	0.00
	Joe	1215.02	1.60	1243.68	4.00	1243.62	9.40
Joe	Joe	1119.84	98.90	1151.48	98.10	1155.89	99.60
	AMH	1359.21	0.00	1388.63	0.00	1389.82	0.00
	Clayton	1361.06	0.00	1379.26	0.00	1380.24	0.00
	Frank	1230.67	0.00	1256.75	0.00	1259.48	0.00
	GH	1141.21	1.10	1169.66	1.90	1179.27	0.40

Table A.4: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull PH margins ($n = 500$; $\tau = 0.75$)

True	Fitted	Weibull PH		BP PH		PE PH	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1580.59	16.90	1610.81	17.20	1611.60	14.60
	Clayton	1578.43	82.60	1610.49	82.30	1610.61	84.90
	Frank	1602.36	0.50	1634.17	0.50	1634.08	0.50
	GH	1633.82	0.00	1661.41	0.00	1657.24	0.00
	Joe	1659.54	0.00	1687.99	0.00	1684.08	0.00
Clayton	Clayton	1062.09	99.90	1093.63	100.00	1094.16	99.90
	AMH	1350.25	0.00	1385.33	0.00	1386.93	0.00
	Frank	1133.70	0.10	1161.21	0.00	1160.83	0.10
	GH	1310.52	0.00	1293.16	0.00	1287.85	0.00
	Joe	1417.59	0.00	1398.52	0.00	1393.01	0.00
Frank	Frank	1058.83	99.80	1090.50	99.90	1090.44	99.40
	AMH	1389.72	0.00	1417.09	0.00	1418.62	0.00
	Clayton	1175.85	0.10	1183.91	0.10	1184.98	0.40
	GH	1194.37	0.10	1186.40	0.00	1181.52	0.20
	Joe	1286.19	0.00	1261.93	0.00	1256.92	0.00
GH	GH	941.27	99.50	973.50	99.50	980.29	99.10
	AMH	1384.48	0.00	1417.03	0.00	1417.25	0.00
	Clayton	1173.91	0.00	1180.39	0.00	1181.78	0.00
	Frank	1032.93	0.00	1061.15	0.00	1062.26	0.00
	Joe	971.95	0.50	999.23	0.50	1005.02	0.90
Joe	Joe	809.76	99.60	842.04	99.70	860.19	99.70
	AMH	1416.84	0.00	1443.97	0.00	1446.40	0.00
	Clayton	1248.05	0.00	1224.22	0.00	1225.39	0.00
	Frank	978.85	0.00	1002.31	0.00	1004.28	0.00
	GH	856.15	0.40	879.14	0.30	896.76	0.30

Table A.5: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull PO margins ($n = 500$; $\tau = 0.75$)

True	Fitted	Weibull PO		BP PO		PE PO	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1464.57	15.50	1492.58	19.30	1498.00	14.70
	Clayton	1464.12	84.10	1493.15	80.10	1496.29	84.60
	Frank	1493.22	0.40	1520.55	0.60	1523.35	0.70
	GH	1529.50	0.00	1548.43	0.00	1546.73	0.00
	Joe	1557.10	0.00	1578.96	0.00	1577.24	0.00
Clayton	Clayton	909.56	99.80	938.30	100.00	941.81	99.80
	AMH	1221.73	0.00	1256.84	0.00	1260.65	0.00
	Frank	999.32	0.20	1015.24	0.00	1017.05	0.20
	GH	1200.94	0.00	1150.41	0.00	1148.54	0.00
	Joe	1319.09	0.00	1266.40	0.00	1264.80	0.00
Frank	Frank	928.25	99.80	957.15	100.00	959.57	99.50
	AMH	1270.17	0.00	1292.09	0.00	1298.31	0.00
	Clayton	1059.34	0.20	1059.44	0.00	1064.44	0.20
	GH	1083.06	0.00	1051.66	0.00	1049.65	0.30
	Joe	1185.71	0.00	1129.92	0.00	1129.21	0.00
GH	GH	811.62	100.00	840.87	99.80	851.73	99.40
	AMH	1263.97	0.00	1292.42	0.00	1297.80	0.00
	Clayton	1050.73	0.00	1059.62	0.00	1066.05	0.00
	Frank	901.04	0.00	927.14	0.00	932.87	0.00
	Joe	853.18	0.00	868.81	0.20	878.50	0.60
Joe	Joe	685.75	99.90	714.74	99.70	737.42	99.50
	AMH	1301.47	0.00	1324.75	0.00	1330.48	0.00
	Clayton	1147.59	0.00	1122.85	0.00	1128.83	0.00
	Frank	850.80	0.00	876.04	0.00	882.98	0.00
	GH	730.94	0.10	753.78	0.30	777.02	0.50

Table A.6: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with Weibull YP margins ($n = 500$; $\tau = 0.75$)

True	Fitted	Weibull YP		BP YP		PE YP	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	1391.65	16.60	1423.93	17.50	1424.75	14.10
	Clayton	1390.94	82.90	1422.14	82.00	1423.30	85.40
	Frank	1419.55	0.50	1450.43	0.50	1451.17	0.50
	GH	1452.49	0.00	1478.13	0.00	1474.13	0.00
	Joe	1481.86	0.00	1509.04	0.00	1505.15	0.00
Clayton	Clayton	830.73	99.90	861.71	100.00	863.22	99.80
	AMH	1153.41	0.00	1184.42	0.00	1186.54	0.00
	Frank	918.79	0.10	942.14	0.00	941.79	0.20
	GH	1103.06	0.00	1073.34	0.00	1069.41	0.00
	Joe	1228.08	0.00	1190.76	0.00	1187.42	0.00
Frank	Frank	855.11	100.00	886.63	99.90	886.84	99.70
	AMH	1196.81	0.00	1223.97	0.00	1224.00	0.00
	Clayton	986.97	0.00	980.67	0.00	983.04	0.20
	GH	992.36	0.00	978.60	0.10	974.68	0.10
	Joe	1093.19	0.00	1056.09	0.00	1053.50	0.00
GH	GH	737.48	100.00	768.85	99.70	777.60	99.50
	AMH	1192.60	0.00	1220.47	0.00	1225.71	0.00
	Clayton	978.45	0.00	980.49	0.00	984.59	0.00
	Frank	831.72	0.00	854.81	0.00	858.42	0.00
	Joe	773.32	0.00	797.05	0.30	804.95	0.50
Joe	Joe	611.89	99.90	643.56	99.80	664.46	99.60
	AMH	1229.83	0.00	1253.97	0.00	1256.33	0.00
	Clayton	1070.57	0.00	1033.88	0.00	1037.84	0.00
	Frank	785.61	0.00	804.59	0.00	809.09	0.00
	GH	663.13	0.10	681.19	0.20	702.33	0.40

Table A.7: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull PH margins ($n = 500$; $\tau = 0.5$)

True	Copula	Weibull PH		BP PH		PE PH	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3220	-3.4062	0.3282	-1.5478	0.3276	-1.7089
	Clayton	0.3362	0.8722	0.3368	1.0382	0.3389	1.6829
	Frank	0.3112	-6.6528	0.3127	-6.1951	0.3158	-5.2609
	GH	0.2332	-30.0266	0.2564	-23.0925	0.2677	-19.6754
	Joe	0.1428	-57.1482	0.1661	-50.1684	0.1800	-46.0020
Clayton	Clayton	0.5015	0.3036	0.5026	0.5144	0.5048	0.9697
	AMH	0.3331	-33.3833	0.3333	-33.3352	0.3333	-33.3358
	Frank	0.4709	-5.8137	0.4758	-4.8308	0.4796	-4.0829
	GH	0.3648	-27.0300	0.4067	-18.6528	0.4130	-17.3993
	Joe	0.2492	-50.1590	0.2973	-40.5371	0.3043	-39.1408
Frank	Frank	0.5010	0.1903	0.5013	0.2678	0.5047	0.9433
	AMH	0.3180	-36.3973	0.3205	-35.8930	0.3205	-35.9086
	Clayton	0.4575	-8.4912	0.4717	-5.6685	0.4735	-5.3081
	GH	0.4135	-17.3090	0.4401	-11.9863	0.4465	-10.7064
	Joe	0.3177	-36.4568	0.3608	-27.8485	0.3660	-26.8084
GH	GH	0.4991	-0.1771	0.5013	0.2568	0.5152	3.0417
	AMH	0.3172	-36.5664	0.3205	-35.9061	0.3205	-35.9054
	Clayton	0.4548	-9.0452	0.4744	-5.1241	0.4757	-4.8610
	Frank	0.5222	4.4368	0.5249	4.9847	0.5277	5.5367
	Joe	0.4328	-13.4483	0.4459	-10.8135	0.4582	-8.3633
Joe	Joe	0.5011	0.2251	0.5038	0.7587	0.5168	3.3671
	AMH	0.2980	-40.3908	0.3015	-39.6924	0.3021	-39.5850
	Clayton	0.4364	-12.7182	0.4718	-5.6342	0.4731	-5.3778
	Frank	0.5488	9.7649	0.5542	10.8388	0.5567	11.3301
	GH	0.5439	8.7749	0.5395	7.9063	0.5549	10.9711

Table A.8: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull PO margins ($n = 500$; $\tau = 0.5$)

True	Copula	Weibull PO		BP PO		PE PO	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3260	-2.2133	0.3280	-1.6142	0.3277	-1.6795
	Clayton	0.3386	1.5695	0.3361	0.8199	0.3382	1.4541
	Frank	0.3136	-5.9237	0.3207	-3.7856	0.3238	-2.8488
	GH	0.2265	-32.0564	0.2700	-18.9882	0.2817	-15.4985
	Joe	0.1312	-60.6462	0.1748	-47.5518	0.1898	-43.0519
Clayton	Clayton	0.5022	0.4429	0.5025	0.4989	0.5039	0.7865
	AMH	0.3331	-33.3767	0.3333	-33.3343	0.3333	-33.3367
	Frank	0.4741	-5.1774	0.4877	-2.4636	0.4913	-1.7413
	GH	0.3576	-28.4744	0.4270	-14.5947	0.4315	-13.6969
	Joe	0.2355	-52.9053	0.3150	-36.9930	0.3201	-35.9767
Frank	Frank	0.4997	-0.0627	0.5016	0.3111	0.5048	0.9645
	AMH	0.3171	-36.5883	0.3191	-36.1737	0.3197	-36.0676
	Clayton	0.4438	-11.2458	0.4540	-9.1993	0.4555	-8.9088
	GH	0.4036	-19.2762	0.4455	-10.8943	0.4514	-9.7204
	Joe	0.3038	-39.2328	0.3716	-25.6797	0.3750	-25.0064
GH	GH	0.4922	-1.5584	0.5014	0.2753	0.5157	3.1428
	AMH	0.3170	-36.6061	0.3186	-36.2735	0.3195	-36.1094
	Clayton	0.4415	-11.7094	0.4533	-9.3312	0.4540	-9.2068
	Frank	0.5207	4.1334	0.5252	5.0352	0.5269	5.3886
	Joe	0.4230	-15.3910	0.4514	-9.7166	0.4638	-7.2416
Joe	Joe	0.4925	-1.4927	0.5042	0.8323	0.5180	3.5968
	AMH	0.2961	-40.7871	0.2979	-40.4198	0.2986	-40.2799
	Clayton	0.4171	-16.5855	0.4416	-11.6760	0.4421	-11.5736
	Frank	0.5457	9.1309	0.5512	10.2497	0.5524	10.4741
	GH	0.5356	7.1229	0.5335	6.7053	0.5496	9.9213

Table A.9: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull YP margins ($n = 500$; $\tau = 0.5$)

True	Copula	Weibull YP		BP YP		PE YP	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3291	-1.2604	0.3286	-1.4134	0.3265	-2.0618
	Clayton	0.3364	0.9205	0.3372	1.1611	0.3385	1.5572
	Frank	0.3172	-4.8475	0.3219	-3.4215	0.3246	-2.6318
	GH	0.2435	-26.9446	0.2730	-18.1096	0.2839	-14.8252
	Joe	0.1483	-55.5071	0.1774	-46.7665	0.1919	-42.4151
Clayton	Clayton	0.5020	0.3954	0.5040	0.8008	0.5049	0.9776
	AMH	0.3333	-33.3356	0.3333	-33.3365	0.3333	-33.3396
	Frank	0.4793	-4.1396	0.4902	-1.9665	0.4933	-1.3437
	GH	0.3796	-24.0859	0.4321	-13.5711	0.4357	-12.8521
	Joe	0.2589	-48.2225	0.3211	-35.7896	0.3248	-35.0371
Frank	Frank	0.5012	0.2377	0.5020	0.4077	0.5049	0.9743
	AMH	0.3167	-36.6525	0.3206	-35.8744	0.3200	-35.9962
	Clayton	0.4444	-11.1239	0.4609	-7.8243	0.4618	-7.6478
	GH	0.4198	-16.0481	0.4474	-10.5108	0.4527	-9.4699
	Joe	0.3258	-34.8346	0.3753	-24.9443	0.3776	-24.4769
GH	GH	0.4997	-0.0686	0.5020	0.3980	0.5160	3.2085
	AMH	0.3175	-36.4954	0.3208	-35.8440	0.3200	-35.9918
	Clayton	0.4424	-11.5208	0.4610	-7.7972	0.4610	-7.8097
	Frank	0.5198	3.9543	0.5263	5.2640	0.5279	5.5895
	Joe	0.4360	-12.8032	0.4527	-9.4686	0.4647	-7.0535
Joe	Joe	0.5013	0.2676	0.5045	0.8998	0.5179	3.5893
	AMH	0.2948	-41.0462	0.2998	-40.0415	0.3001	-39.9724
	Clayton	0.4195	-16.1001	0.4555	-8.9031	0.4551	-8.9761
	Frank	0.5430	8.6092	0.5524	10.4785	0.5535	10.6905
	GH	0.5402	8.0495	0.5340	6.8086	0.5496	9.9168

Table A.10: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull PH margins ($n = 500$; $\tau = 0.75$)

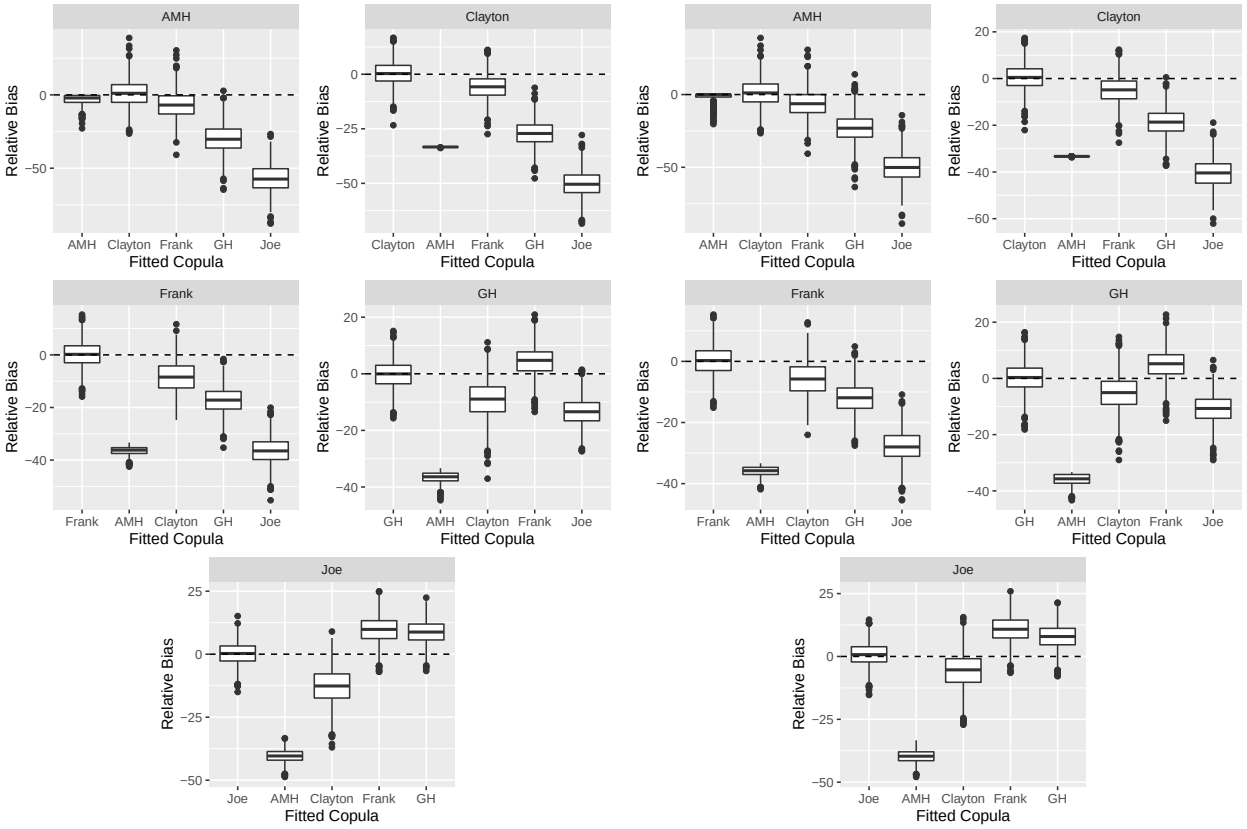
True	Copula	Weibull PH		BP PH		PE PH	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3221	-3.3792	0.3282	-1.5478	0.3277	-1.7030
	Clayton	0.3362	0.8537	0.3368	1.0377	0.3390	1.6855
	Frank	0.3110	-6.7011	0.3127	-6.1951	0.3157	-5.2786
	GH	0.2333	-30.0249	0.2564	-23.0861	0.2677	-19.6924
	Joe	0.1428	-57.1554	0.1661	-50.1810	0.1801	-45.9683
Clayton	Clayton	0.7507	0.0904	0.7524	0.3205	0.7543	0.5694
	AMH	0.3333	-55.5556	0.3333	-55.5556	0.3333	-55.5556
	Frank	0.7275	-2.9980	0.7345	-2.0621	0.7375	-1.6648
	GH	0.5764	-23.1459	0.6306	-15.9143	0.6341	-15.4594
	Joe	0.4369	-41.7473	0.5075	-32.3322	0.5117	-31.7711
Frank	Frank	0.7508	0.1123	0.7517	0.2258	0.7547	0.6202
	AMH	0.3310	-55.8678	0.3322	-55.7126	0.3321	-55.7233
	Clayton	0.7279	-2.9435	0.7381	-1.5885	0.7392	-1.4341
	GH	0.6417	-14.4442	0.6710	-10.5390	0.6735	-10.2023
	Joe	0.5222	-30.3682	0.5747	-23.3741	0.5778	-22.9646
GH	GH	0.7500	0.0026	0.7517	0.2233	0.7576	1.0197
	AMH	0.3324	-55.6829	0.3331	-55.5851	0.3330	-55.5933
	Clayton	0.7331	-2.2491	0.7502	0.0232	0.7517	0.2226
	Frank	0.7691	2.5463	0.7718	2.9078	0.7752	3.3637
	Joe	0.6850	-8.6639	0.6967	-7.1132	0.6998	-6.6908
Joe	Joe	0.7511	0.1532	0.7527	0.3660	0.7528	0.3691
	AMH	0.3275	-56.3328	0.3296	-56.0485	0.3296	-56.0533
	Clayton	0.7241	-3.4482	0.7658	2.1090	0.7674	2.3145
	Frank	0.7978	6.3718	0.8002	6.6925	0.8031	7.0832
	GH	0.7862	4.8273	0.7839	4.5156	0.7892	5.2226

Table A.11: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull PO margins ($n = 500$; $\tau = 0.75$)

True	Copula	Weibull PO		BP PO		PE PO	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3260	-2.2133	0.3279	-1.6163	0.3277	-1.6812
	Clayton	0.3385	1.5647	0.3361	0.8395	0.3382	1.4686
	Frank	0.3136	-5.9252	0.3207	-3.7802	0.3238	-2.8499
	GH	0.2265	-32.0619	0.2700	-18.9882	0.2816	-15.5298
	Joe	0.1312	-60.6543	0.1748	-47.5524	0.1899	-43.0420
Clayton	Clayton	0.7488	-0.1603	0.7527	0.3605	0.7541	0.5402
	AMH	0.3333	-55.5560	0.3333	-55.5556	0.3333	-55.5556
	Frank	0.7316	-2.4596	0.7458	-0.5646	0.7487	-0.1763
	GH	0.5703	-23.9658	0.6537	-12.8340	0.6534	-12.8815
	Joe	0.4231	-43.5931	0.5343	-28.7640	0.5334	-28.8789
Frank	Frank	0.7496	-0.0499	0.7518	0.2371	0.7547	0.6247
	AMH	0.3310	-55.8714	0.3327	-55.6359	0.3325	-55.6650
	Clayton	0.7087	-5.5005	0.7173	-4.3655	0.7181	-4.2475
	GH	0.6344	-15.4142	0.6790	-9.4699	0.6792	-9.4355
	Joe	0.5097	-32.0430	0.5908	-21.2294	0.5895	-21.4028
GH	GH	0.7461	-0.5147	0.7518	0.2385	0.7577	1.0267
	AMH	0.3324	-55.6784	0.3332	-55.5674	0.3332	-55.5779
	Clayton	0.7146	-4.7150	0.7277	-2.9725	0.7274	-3.0138
	Frank	0.7667	2.2306	0.7724	2.9835	0.7751	3.3462
	Joe	0.6802	-9.3046	0.7047	-6.0403	0.7061	-5.8487
Joe	Joe	0.7473	-0.3654	0.7530	0.4022	0.7525	0.3363
	AMH	0.3270	-56.4040	0.3303	-55.9561	0.3302	-55.9748
	Clayton	0.6906	-7.9159	0.7292	-2.7677	0.7288	-2.8201
	Frank	0.7935	5.8018	0.7964	6.1846	0.7984	6.4582
	GH	0.7798	3.9743	0.7777	3.6996	0.7831	4.4197

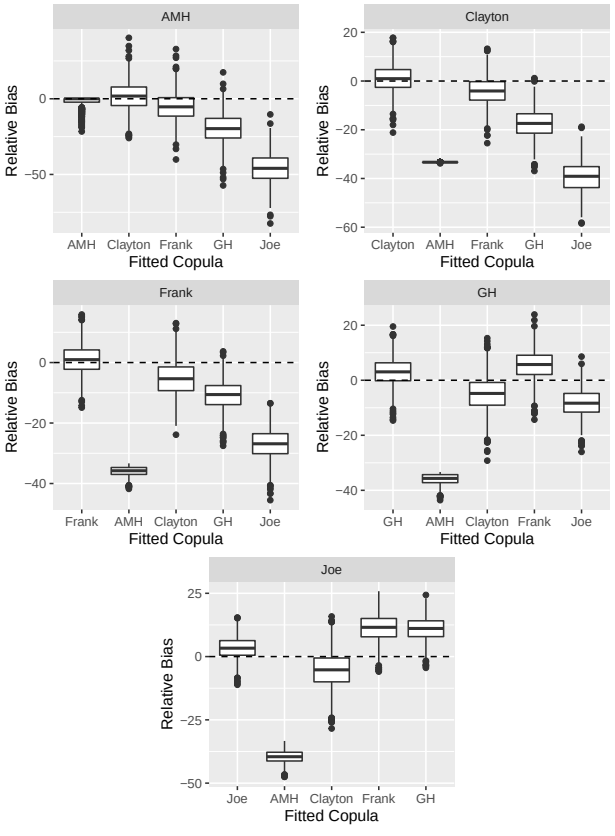
Table A.12: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with Weibull YP margins ($n = 500$; $\tau = 0.75$)

True	Copula	Weibull YP		BP YP		PE YP	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3291	-1.2631	0.3286	-1.4159	0.3265	-2.0618
	Clayton	0.3364	0.9127	0.3372	1.1471	0.3385	1.5550
	Frank	0.3172	-4.8445	0.3219	-3.4248	0.3246	-2.6302
	GH	0.2435	-26.9506	0.2730	-18.1096	0.2839	-14.8343
	Joe	0.1484	-55.4932	0.1774	-46.7665	0.1919	-42.4175
Clayton	Clayton	0.7515	0.2021	0.7539	0.5215	0.7547	0.6242
	AMH	0.3333	-55.5556	0.3333	-55.5556	0.3333	-55.5556
	Frank	0.7358	-1.8996	0.7476	-0.3164	0.7503	0.0342
	GH	0.5962	-20.5039	0.6587	-12.1719	0.6584	-12.2122
	Joe	0.4531	-39.5873	0.5400	-27.9963	0.5396	-28.0487
Frank	Frank	0.7513	0.1719	0.7521	0.2767	0.7548	0.6436
	AMH	0.3317	-55.7770	0.3330	-55.5993	0.3327	-55.6459
	Clayton	0.7145	-4.7300	0.7232	-3.5716	0.7236	-3.5190
	GH	0.6520	-13.0664	0.6809	-9.2197	0.6807	-9.2390
	Joe	0.5372	-28.3751	0.5937	-20.8409	0.5924	-21.0181
GH	GH	0.7504	0.0583	0.7523	0.3058	0.7579	1.0500
	AMH	0.3330	-55.5949	0.3333	-55.5631	0.3332	-55.5713
	Clayton	0.7205	-3.9280	0.7324	-2.3480	0.7320	-2.3987
	Frank	0.7668	2.2443	0.7731	3.0739	0.7758	3.4362
	Joe	0.6886	-8.1820	0.7057	-5.9092	0.7070	-5.7396
Joe	Joe	0.7515	0.2026	0.7533	0.4373	0.7526	0.3419
	AMH	0.3277	-56.3037	0.3316	-55.7852	0.3310	-55.8704
	Clayton	0.7049	-6.0130	0.7417	-1.1064	0.7411	-1.1917
	Frank	0.7927	5.6965	0.7964	6.1887	0.7987	6.4986
	GH	0.7827	4.3545	0.7788	3.8416	0.7841	4.5458



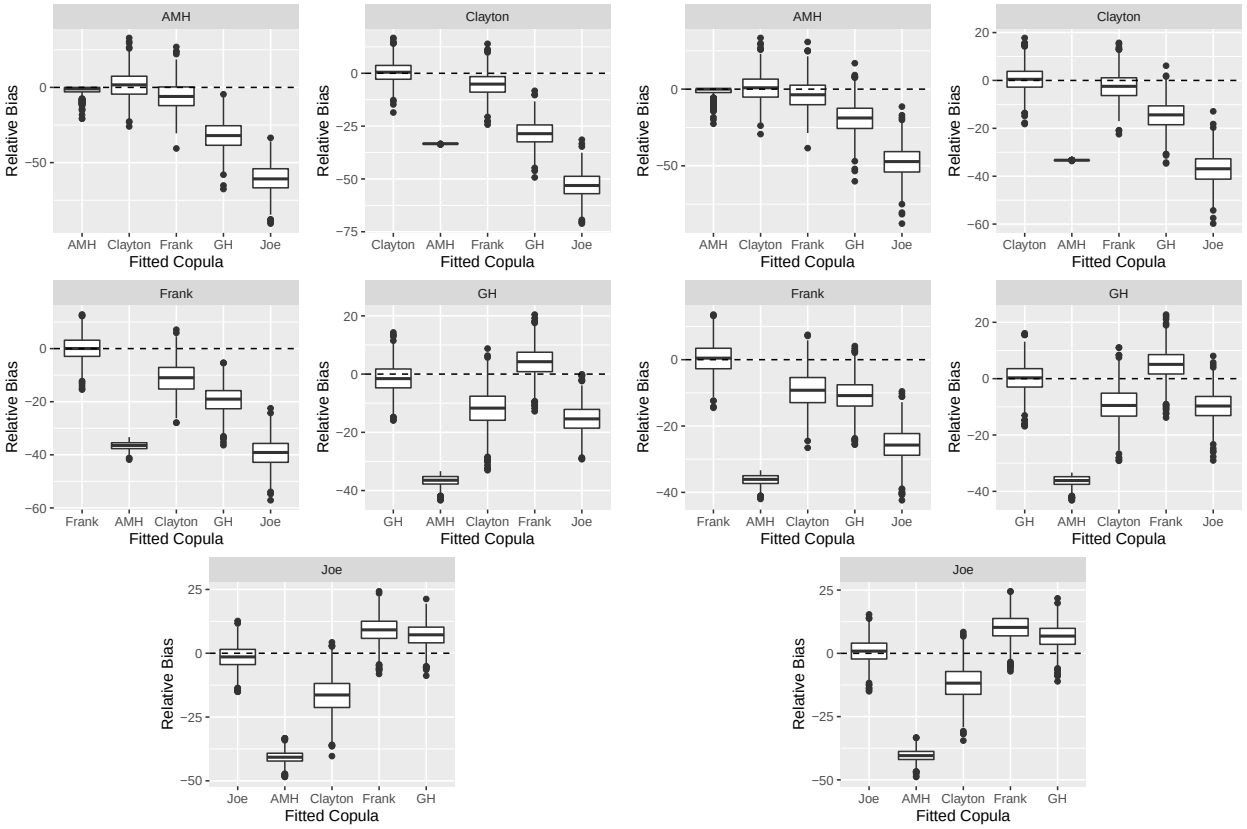
Weibull PH margins

BP PH margins



PE PH margins

Figure A.1: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull PH margins ($n = 500$; $\tau = 0.5$)

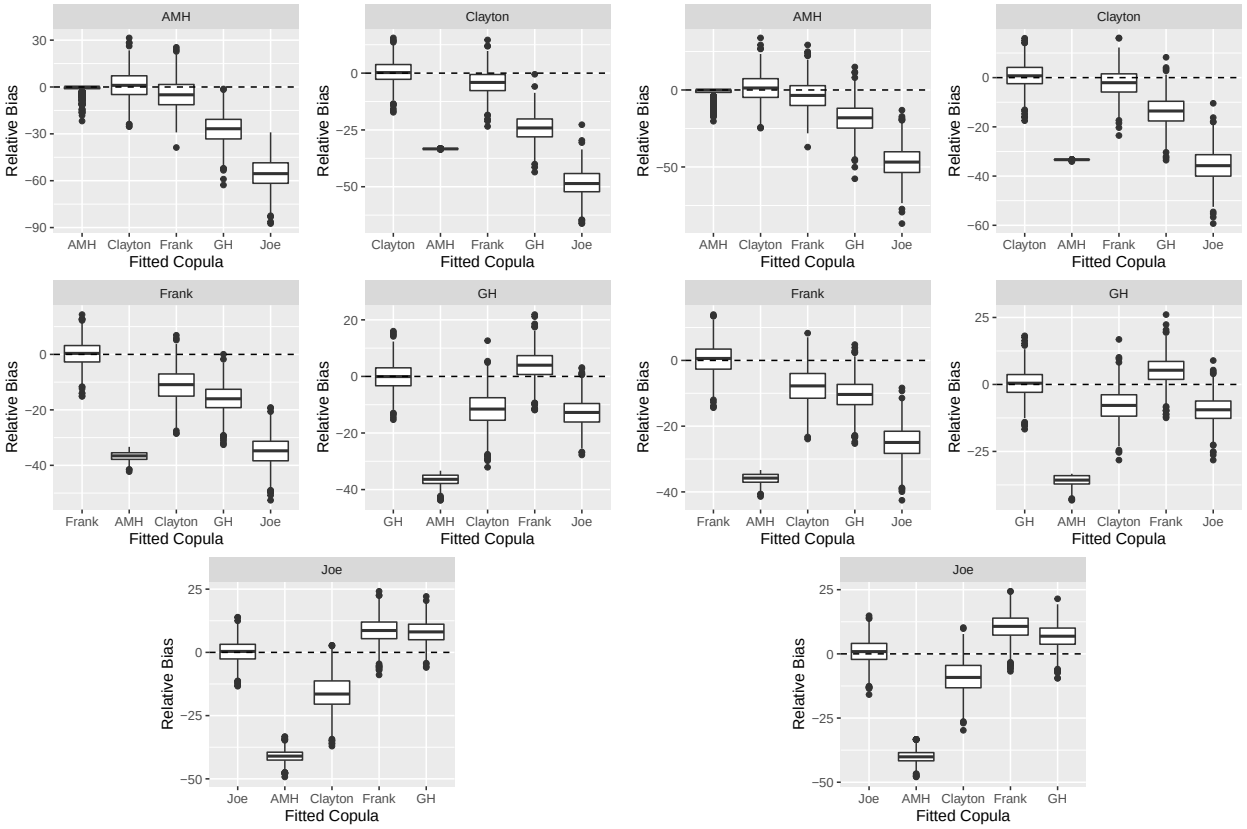


Weibull PO margins

BP PO margins

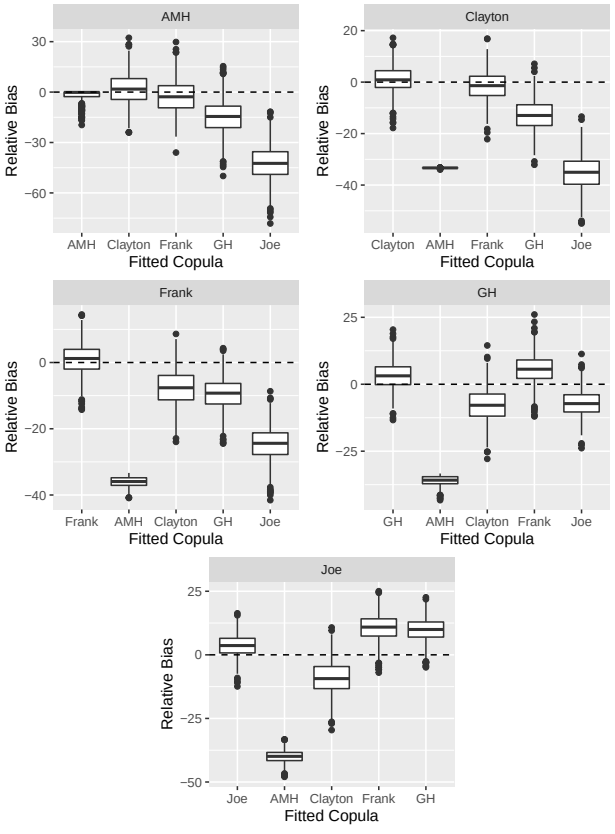
PE PO margins

Figure A.2: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull PO margins ($n = 500$; $\tau = 0.5$)



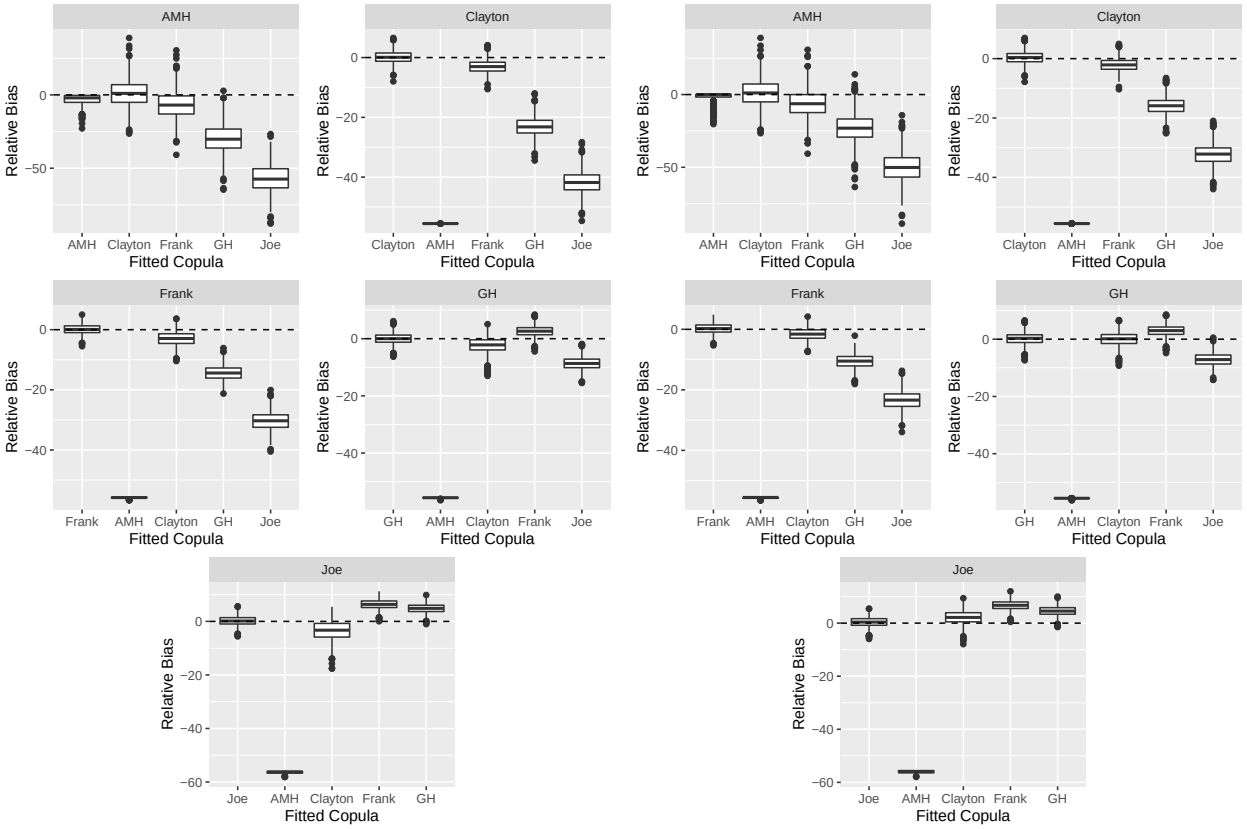
Weibull YP margins

BP YP margins



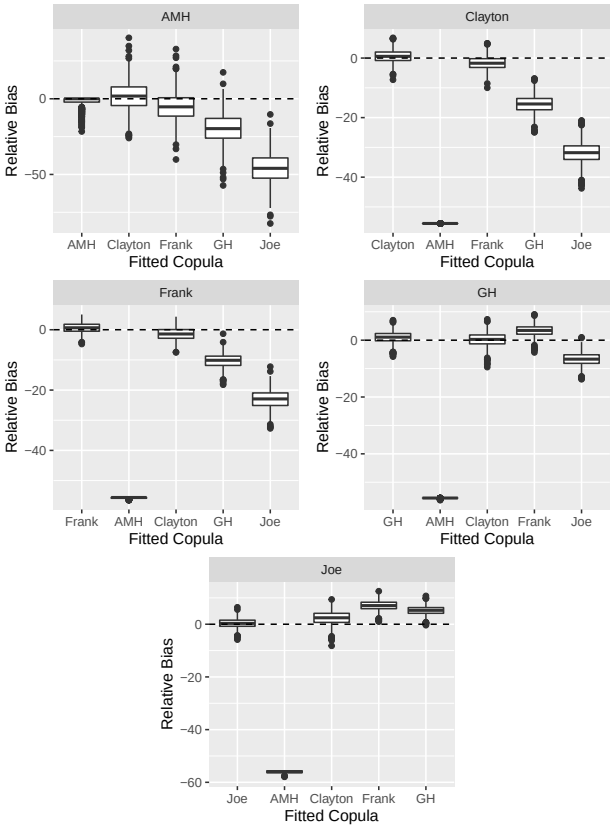
PE YP margins

Figure A.3: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull YP margins ($n = 500$; $\tau = 0.5$)



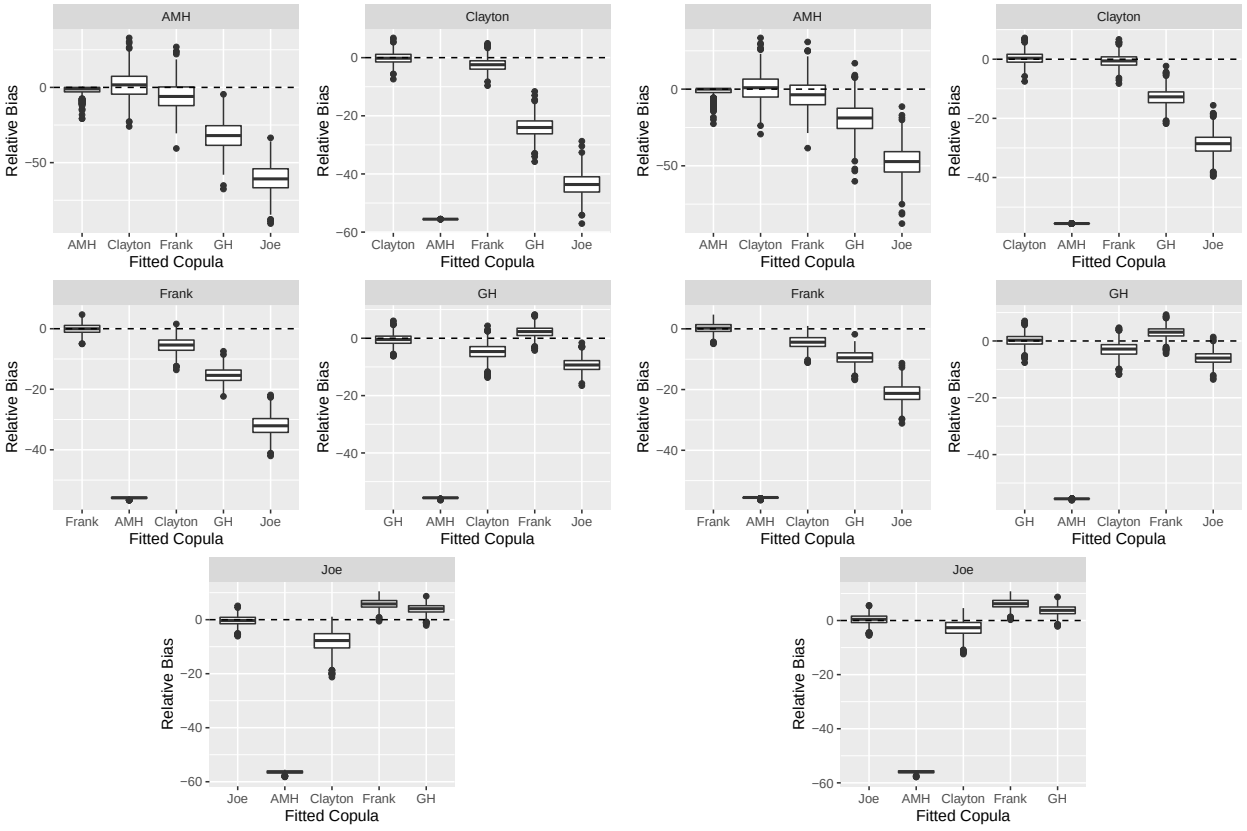
Weibull PH margins

BP PH margins



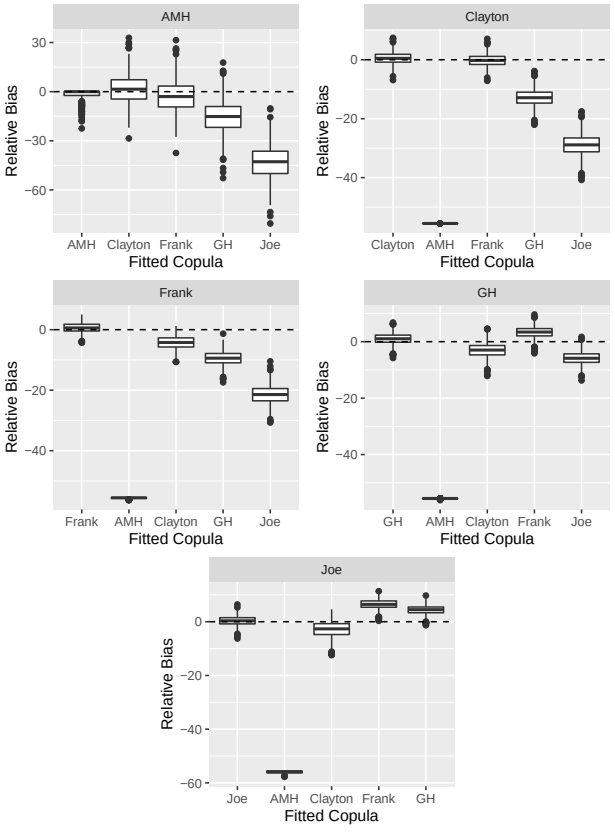
PE PH margins

Figure A.4: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull PH margins ($n = 500$; $\tau = 0.75$)



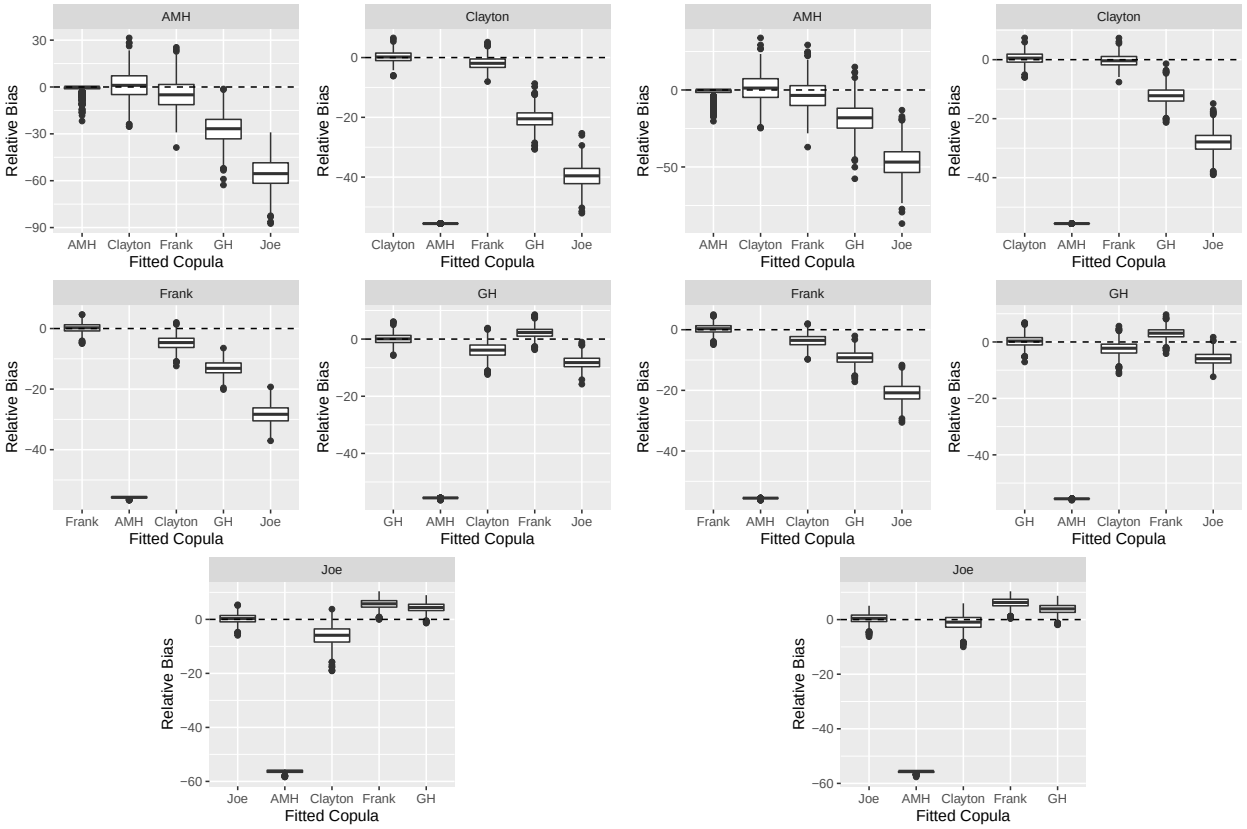
Weibull PO margins

BP PO margins



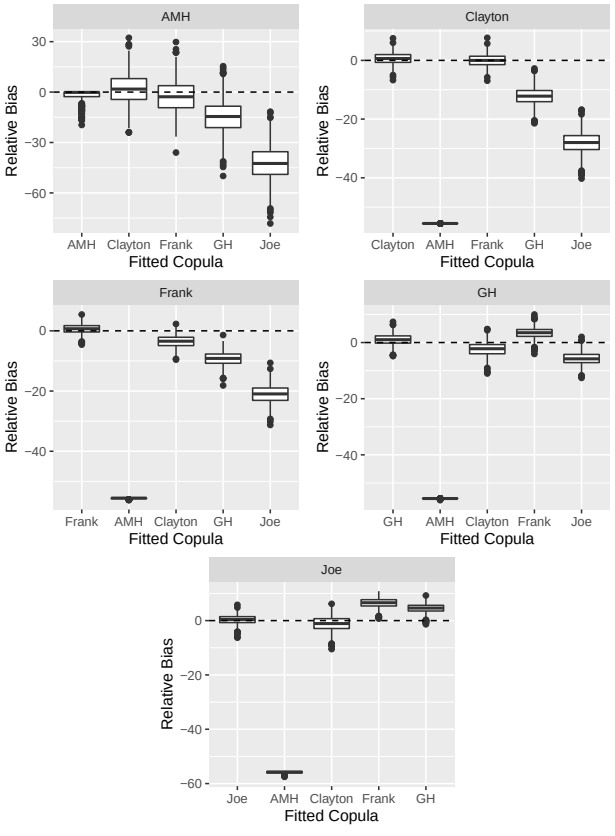
PE PO margins

Figure A.5: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull PO margins ($n = 500$; $\tau = 0.75$)



Weibull YP margins

BP YP margins



PE YP margins

Figure A.6: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with Weibull YP margins ($n = 500$; $\tau = 0.75$)

Table A.13: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull PH margins ($n = 500; \tau = 0.5$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	4.0786	0.4915	4.2633	0.4733	3.9422	0.5066
	PO	53.6566	< 0.0001	37.5120	0.0002	35.9249	0.0001
Clayton	PH	4.1552	0.4866	4.0848	0.4888	3.9933	0.4998
	PO	52.1819	< 0.0001	37.0372	0.0002	37.1081	0.0001
Frank	PH	4.1845	0.4789	4.1552	0.4793	4.0751	0.4859
	PO	46.9079	< 0.0001	33.7010	0.0003	33.8977	0.0002
GH	PH	4.0814	0.4958	4.1157	0.4876	3.9735	0.5023
	PO	46.9448	< 0.0001	33.8404	0.0005	34.3718	0.0003
Joe	PH	4.1289	0.4872	4.3066	0.4701	4.1924	0.4786
	PO	47.5744	< 0.0001	34.0852	0.0004	35.2987	0.0002

Table A.14: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull PO margins ($n = 500; \tau = 0.5$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	31.4730	0.0012	23.6811	0.0104	24.5327	0.0084
	PO	5.8156	0.3537	4.1720	0.4915	5.4295	0.3392
Clayton	PH	30.5809	0.0013	23.9132	0.0083	25.3416	0.0051
	PO	5.6386	0.3562	4.0893	0.4942	5.3627	0.3428
Frank	PH	29.4902	0.0016	23.2168	0.0106	24.3390	0.0074
	PO	5.5510	0.3632	4.1485	0.4870	5.3165	0.3471
GH	PH	29.4416	0.0026	23.2887	0.0106	23.9788	0.0082
	PO	5.6593	0.3622	4.3634	0.4732	5.5595	0.3342
Joe	PH	32.2492	0.0015	25.1835	0.0103	25.4390	0.0084
	PO	5.5502	0.3721	4.2044	0.4806	5.4813	0.3285

Table A.15: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull YP margins ($n = 500; \tau = 0.5$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	129.6494	< 0.0001	112.7223	< 0.0001	114.8362	< 0.0001
	PO	101.2200	< 0.0001	86.4514	< 0.0001	88.9593	< 0.0001
Clayton	PH	130.4050	< 0.0001	115.8957	< 0.0001	117.6691	< 0.0001
	PO	113.6766	< 0.0001	97.1271	< 0.0001	99.9802	< 0.0001
Frank	PH	119.8104	< 0.0001	105.9319	< 0.0001	107.5134	< 0.0001
	PO	94.5675	< 0.0001	81.2991	< 0.0001	83.4815	< 0.0001
GH	PH	119.5611	< 0.0001	105.2201	< 0.0001	106.0989	< 0.0001
	PO	92.3302	< 0.0001	78.9427	< 0.0001	80.5208	< 0.0001
Joe	PH	127.9855	< 0.0001	111.9543	< 0.0001	113.1972	< 0.0001
	PO	96.6103	< 0.0001	81.6008	< 0.0001	83.4350	< 0.0001

Table A.16: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull PH margins ($n = 500; \tau = 0.75$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	4.0679	0.4927	4.2633	0.4733	3.9458	0.5062
	PO	53.6867	< 0.0001	37.5177	0.0002	35.8834	0.0001
Clayton	PH	4.2530	0.4732	4.3030	0.4674	4.2003	0.4789
	PO	59.3317	< 0.0001	38.8624	0.0002	38.8464	0.0001
Frank	PH	4.1085	0.4952	4.1176	0.4912	4.0613	0.4965
	PO	46.6895	< 0.0001	31.9956	0.0006	32.0962	0.0005
GH	PH	4.0969	0.4947	4.2173	0.4808	4.2422	0.4814
	PO	48.3851	< 0.0001	34.5418	0.0004	35.4898	0.0002
Joe	PH	4.1397	0.4869	4.2477	0.4767	4.4248	0.4602
	PO	48.1058	0.0001	34.5065	0.0003	36.0928	0.0003

Table A.17: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull PO margins ($n = 500; \tau = 0.75$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	31.4746	0.0012	23.6687	0.0104	24.5679	0.0084
	PO	5.8156	0.3537	4.1670	0.4921	5.4246	0.3397
Clayton	PH	34.1277	0.0006	25.1576	0.0072	27.3388	0.0037
	PO	5.8548	0.3422	4.2676	0.4816	5.5711	0.3310
Frank	PH	31.9510	0.0010	24.4964	0.0079	25.6515	0.0051
	PO	5.6419	0.3590	4.2079	0.4825	5.4412	0.3370
GH	PH	33.2654	0.0008	25.3660	0.0093	25.8902	0.0070
	PO	5.9434	0.3465	4.3894	0.4705	5.8142	0.3159
Joe	PH	37.9297	0.0005	27.9655	0.0051	28.1033	0.0034
	PO	6.0286	0.3396	4.0225	0.4971	5.5462	0.3283

Table A.18: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with Weibull YP margins ($n = 500; \tau = 0.75$)

Copula	Class	Weibull Fitting		BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value	LR stat.	P-value
AMH	PH	129.5452	< 0.0001	112.6162	< 0.0001	114.8362	< 0.0001
	PO	101.2151	< 0.0001	86.4675	< 0.0001	88.8835	< 0.0001
Clayton	PH	156.1917	< 0.0001	138.7861	< 0.0001	141.6539	< 0.0001
	PO	136.5450	< 0.0001	118.2574	< 0.0001	121.5292	< 0.0001
Frank	PH	128.1955	< 0.0001	113.4838	< 0.0001	115.5439	< 0.0001
	PO	95.3431	< 0.0001	82.5949	< 0.0001	84.7231	< 0.0001
GH	PH	129.0373	< 0.0001	112.5214	< 0.0001	112.5278	< 0.0001
	PO	94.5899	< 0.0001	80.8641	< 0.0001	82.1328	< 0.0001
Joe	PH	141.7503	< 0.0001	122.4834	< 0.0001	122.2895	< 0.0001
	PO	96.1089	< 0.0001	81.3977	< 0.0001	83.1340	< 0.0001

APPENDIX B

Additional Results for Generated Copulas with EW Baseline

B.1 Akaike Information Criteria

Results in this appendix for the AIC consist of tables on two statistics, the mean AIC and the proportion of choice by the least AIC, for each fitted copula model (among AMH, Clayton, Frank, GH and Joe), given a specification for the baseline distribution (Weibull, BP or PE) and regression model class (PH, PO or YP), on Scenarios S2 ($n = 500, \tau = 0.5$) and S3 ($n = 500, \tau = 0.75$).

B.2 Correlation Estimates

Results for the Kendall's τ correlation estimates in this appendix comprehend tables of two statistics (AE and ARB), and boxplots of the relative biases, for all Archimedean copulas addressed in this work (AMH, Clayton, Frank, GH and Joe), given each combination of fitted baseline distribution (Weibull, BP or PE) and regression model class (PH, PO or YP), on Scenarios S2 and S3.

B.3 Likelihood Ratio Tests

Finally, results for the LR tests in this appendix comprehend tables of means for the LR statistic and its corresponding p-value for nested fitted models with respect to the regression model class, given each one of the five Archimedean copulas discussed in this work (AMH, Clayton, Frank, GH and Joe), combined with a fitted baseline distribution (BP or PE), on Scenarios S2 and S3, given a regression model class (PH, PO or YP) used for marginal data generation.

Table B.1: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW PH margins ($n = 500$; $\tau = 0.5$)

True	Fitted	Weibull PH		BP PH		PE PH	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	842.56	0.64	850.72	16.70	802.05	15.00
	Clayton	850.63	99.11	852.00	82.90	802.23	84.30
	Frank	881.37	0.25	878.98	0.40	827.68	0.70
	GH	926.95	0.00	934.18	0.00	877.91	0.00
	Joe	946.44	0.00	954.14	0.00	903.66	0.00
Clayton	Clayton	702.77	99.87	701.82	99.90	651.68	99.50
	AMH	803.74	0.00	735.21	0.00	690.01	0.30
	Frank	761.95	0.13	750.58	0.10	698.17	0.20
	GH	858.61	0.00	863.57	0.00	796.70	0.00
	Joe	904.09	0.00	913.12	0.00	855.55	0.00
Frank	Frank	707.85	87.98	710.45	99.70	660.28	99.60
	AMH	772.43	0.00	775.52	0.00	729.26	0.00
	Clayton	724.39	12.02	757.29	0.30	708.41	0.40
	GH	794.01	0.00	803.56	0.00	743.56	0.00
	Joe	840.97	0.00	850.66	0.00	793.14	0.00
GH	GH	639.04	97.84	648.97	76.60	601.30	76.70
	AMH	773.86	0.00	783.19	0.00	737.88	0.00
	Clayton	728.58	0.00	764.18	0.00	715.86	0.00
	Frank	678.92	2.16	691.74	23.40	640.59	22.90
	Joe	669.14	0.00	679.74	0.00	626.96	0.40
Joe	Joe	551.53	22.25	570.15	19.20	517.03	56.60
	AMH	788.05	0.00	794.36	0.00	746.84	0.00
	Clayton	736.63	0.00	781.67	0.00	731.61	0.00
	Frank	636.36	0.00	656.86	23.60	604.01	23.60
	GH	545.64	77.75	564.73	57.20	521.68	19.80

Table B.2: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW PO margins ($n = 500$; $\tau = 0.5$)

True	Fitted	Weibull PO		BP PO		PE PO	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	734.86	0.00	726.86	17.60	691.26	12.30
	Clayton	755.49	98.98	727.74	82.20	690.47	87.10
	Frank	793.76	1.02	759.09	0.20	720.15	0.60
	GH	843.30	0.00	818.43	0.00	774.23	0.00
	Joe	862.69	0.00	841.86	0.00	803.00	0.00
Clayton	Clayton	597.16	99.87	567.32	100.00	529.63	99.10
	AMH	706.61	0.00	606.02	0.00	582.36	0.00
	Frank	670.28	0.13	623.11	0.00	582.12	0.90
	GH	775.24	0.00	743.61	0.00	688.42	0.00
	Joe	822.36	0.00	800.38	0.00	753.79	0.00
Frank	Frank	611.48	72.51	588.85	99.90	550.81	99.90
	AMH	677.36	0.00	655.69	0.00	620.67	0.00
	Clayton	621.10	27.49	638.64	0.10	601.85	0.10
	GH	707.12	0.00	685.44	0.00	636.92	0.00
	Joe	757.64	0.00	737.07	0.00	690.16	0.00
GH	GH	542.96	96.82	529.58	75.90	492.91	76.90
	AMH	695.28	0.00	664.12	0.00	631.09	0.00
	Clayton	627.88	0.00	647.41	0.00	612.13	0.20
	Frank	581.73	3.05	569.02	24.10	532.17	22.60
	Joe	578.46	0.13	562.97	0.00	520.31	0.30
Joe	Joe	456.97	7.33	453.20	23.40	411.62	54.10
	AMH	684.34	0.00	677.61	0.00	642.47	0.10
	Clayton	641.04	0.00	670.92	0.00	634.22	0.10
	Frank	538.36	0.00	535.79	23.60	498.61	23.40
	GH	445.47	92.67	448.68	53.00	416.65	22.30

Table B.3: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW YP margins ($n = 500$; $\tau = 0.5$)

True	Fitted	Weibull YP		BP YP		PE YP	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	685.86	0.76	695.67	18.50	650.93	14.30
	Clayton	684.72	99.11	694.66	81.30	653.92	85.40
	Frank	722.60	0.13	726.71	0.20	684.22	0.30
	GH	769.73	0.00	788.14	0.00	739.25	0.00
	Joe	792.61	0.00	811.19	0.00	768.72	0.00
Clayton	Clayton	523.68	99.87	532.18	100.00	492.09	99.70
	AMH	566.81	0.13	572.94	0.00	532.07	0.20
	Frank	594.39	0.00	589.46	0.00	545.28	0.10
	GH	693.66	0.00	712.85	0.00	653.01	0.00
	Joe	747.08	0.00	770.06	0.00	720.27	0.00
Frank	Frank	540.38	87.08	556.35	99.60	515.92	99.70
	AMH	590.93	0.00	624.04	0.00	586.54	0.00
	Clayton	556.60	12.92	604.32	0.40	564.38	0.30
	GH	624.90	0.00	654.00	0.00	602.32	0.00
	Joe	678.84	0.00	705.78	0.00	656.20	0.00
GH	GH	473.05	98.98	496.04	75.70	456.36	77.10
	AMH	599.54	0.00	631.57	0.00	594.39	0.00
	Clayton	566.23	0.00	611.94	0.00	574.01	0.00
	Frank	517.60	0.76	535.43	24.30	495.89	22.60
	Joe	505.78	0.25	529.91	0.00	484.93	0.30
Joe	Joe	386.90	27.49	419.35	20.60	375.56	53.20
	AMH	612.85	0.00	645.73	0.00	607.06	0.00
	Clayton	580.97	0.00	633.78	0.00	594.59	0.00
	Frank	477.73	0.00	501.88	23.60	462.29	23.60
	GH	381.78	72.51	413.62	55.80	379.89	23.20

Table B.4: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW PH margins ($n = 500$; $\tau = 0.75$)

True	Fitted	Weibull PH		BP PH		PE PH	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	842.56	0.64	850.86	16.50	801.92	14.90
	Clayton	850.93	98.85	851.91	82.90	802.18	84.60
	Frank	881.43	0.51	878.89	0.60	827.96	0.50
	GH	927.17	0.00	934.18	0.00	878.21	0.00
	Joe	946.53	0.00	954.14	0.00	903.62	0.00
Clayton	Clayton	303.09	100.00	291.48	100.00	241.33	99.70
	AMH	614.28	0.00	594.49	0.00	555.18	0.00
	Frank	402.83	0.00	369.01	0.00	315.15	0.30
	GH	638.26	0.00	632.67	0.00	551.67	0.00
	Joe	740.24	0.00	741.07	0.00	668.82	0.00
Frank	Frank	287.89	97.92	285.51	100.00	233.54	99.90
	AMH	619.94	0.00	627.32	0.00	586.77	0.00
	Clayton	329.95	2.08	377.84	0.00	329.94	0.10
	GH	496.40	0.00	500.74	0.00	435.10	0.00
	Joe	602.79	0.00	601.33	0.00	535.25	0.00
GH	GH	169.28	99.62	193.98	79.40	133.32	79.50
	AMH	618.75	0.00	625.37	0.00	585.38	0.00
	Clayton	326.77	0.00	374.75	0.00	323.05	0.10
	Frank	242.81	0.25	261.93	20.50	195.93	20.40
	Joe	232.78	0.13	257.48	0.10	181.85	0.00
Joe	Joe	33.27	33.59	81.57	23.00	11.32	63.00
	AMH	651.64	0.00	654.62	0.00	610.17	0.00
	Clayton	353.38	0.00	411.45	0.00	357.59	0.30
	Frank	157.93	0.00	192.39	23.20	121.21	22.90
	GH	27.53	66.41	74.34	53.80	24.13	13.80

Table B.5: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW PO margins ($n = 500$; $\tau = 0.75$)

True	Fitted	Weibull PO		BP PO		PE PO	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	734.86	0.00	726.86	17.70	691.13	12.20
	Clayton	755.84	99.75	727.86	82.10	690.55	87.40
	Frank	793.77	0.25	759.20	0.20	720.44	0.40
	GH	843.54	0.00	818.46	0.00	774.13	0.00
	Joe	862.72	0.00	841.86	0.00	802.98	0.00
Clayton	Clayton	162.89	99.87	124.73	100.00	87.62	99.70
	AMH	544.66	0.00	457.87	0.00	449.90	0.00
	Frank	287.96	0.13	213.85	0.00	170.12	0.30
	GH	548.95	0.00	497.96	0.00	425.81	0.00
	Joe	658.42	0.00	619.97	0.00	555.68	0.00
Frank	Frank	166.10	93.37	139.96	99.80	101.36	99.70
	AMH	500.44	0.00	495.77	0.00	474.38	0.00
	Clayton	200.65	6.63	239.27	0.20	205.45	0.30
	GH	398.96	0.00	365.54	0.00	309.41	0.00
	Joe	516.11	0.00	476.02	0.00	418.25	0.00
GH	GH	51.21	99.87	50.77	79.40	3.95	79.60
	AMH	563.20	0.00	495.28	0.00	482.83	0.00
	Clayton	205.34	0.00	240.06	0.00	206.00	0.00
	Frank	123.35	0.13	111.70	20.60	64.96	20.30
	Joe	130.48	0.00	121.87	0.00	59.17	0.10
Joe	Joe	-80.26	7.16	-59.29	23.20	-110.98	59.90
	AMH	542.06	0.00	525.90	0.00	498.44	0.00
	Clayton	246.09	0.00	294.65	0.00	260.46	0.30
	Frank	32.09	0.00	43.53	23.20	-2.59	22.90
	GH	-99.85	92.84	-66.94	53.60	-100.52	16.90

Table B.6: Mean AIC and choice proportion for all fitted survival copula models over generated data from each copula with EW YP margins ($n = 500$; $\tau = 0.75$)

True	Fitted	Weibull YP		BP YP		PE YP	
		AIC	Choice (%)	AIC	Choice (%)	AIC	Choice (%)
AMH	AMH	685.84	0.76	695.62	18.50	650.93	14.50
	Clayton	684.58	98.98	694.66	81.30	653.89	84.90
	Frank	722.54	0.25	726.67	0.20	684.30	0.60
	GH	769.89	0.00	788.48	0.00	739.49	0.00
	Joe	792.61	0.00	811.19	0.00	768.76	0.00
Clayton	Clayton	84.58	99.62	84.45	99.90	45.15	99.80
	AMH	419.09	0.00	426.76	0.00	389.07	0.00
	Frank	205.33	0.38	176.61	0.10	130.66	0.20
	GH	446.87	0.00	467.12	0.00	388.71	0.00
	Joe	566.64	0.00	590.34	0.00	520.84	0.00
Frank	Frank	90.70	98.06	103.51	100.00	63.31	100.00
	AMH	434.66	0.00	465.30	0.00	424.96	0.00
	Clayton	136.26	1.94	194.74	0.00	158.68	0.00
	GH	296.89	0.00	332.11	0.00	272.12	0.00
	Joe	420.13	0.00	443.50	0.00	381.88	0.00
GH	GH	-24.73	99.87	15.25	79.30	-34.76	79.70
	AMH	437.08	0.00	461.79	0.00	428.65	0.00
	Clayton	144.52	0.00	194.57	0.00	158.87	0.10
	Frank	59.44	0.00	72.91	20.70	24.73	20.20
	Joe	47.32	0.13	87.72	0.00	21.86	0.00
Joe	Joe	-159.69	31.12	-94.18	18.50	-149.06	56.30
	AMH	458.61	0.00	492.42	0.00	453.93	0.00
	Clayton	183.45	0.00	237.63	0.00	202.13	0.00
	Frank	-25.67	0.00	5.68	23.20	-42.27	23.20
	GH	-166.78	68.88	-104.06	58.30	-140.43	20.50

Table B.7: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW PH margins ($n = 500$; $\tau = 0.5$)

True	Copula	Weibull PH		BP PH		PE PH	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3221	-3.3813	0.3273	-1.8142	0.3283	-1.5075
	Clayton	0.3594	7.8141	0.3274	-1.7819	0.3369	1.0617
	Frank	0.2908	-12.7635	0.2946	-11.6201	0.3099	-7.0195
	GH	0.1583	-52.5150	0.1371	-58.8716	0.1919	-42.4344
	Joe	0.0782	-76.5306	0.0606	-81.8244	0.0952	-71.4277
Clayton	Clayton	0.5101	2.0300	0.4912	-1.7667	0.5018	0.3514
	AMH	0.3330	-33.4079	0.3333	-33.3345	0.3333	-33.3346
	Frank	0.4429	-11.4296	0.4539	-9.2264	0.4721	-5.5748
	GH	0.2750	-45.0039	0.2481	-50.3802	0.3230	-35.4001
	Joe	0.1582	-68.3677	0.1334	-73.3255	0.1956	-60.8703
Frank	Frank	0.4902	-1.9591	0.4834	-3.3172	0.4989	-0.2294
	AMH	0.3284	-34.3240	0.3197	-36.0549	0.3211	-35.7877
	Clayton	0.5097	1.9375	0.4662	-6.7536	0.4745	-5.0923
	GH	0.3397	-32.0500	0.3010	-39.8037	0.3636	-27.2833
	Joe	0.2285	-54.2924	0.1987	-60.2656	0.2615	-47.7025
GH	GH	0.4473	-10.5415	0.4072	-18.5656	0.4632	-7.3610
	AMH	0.3287	-34.2597	0.3201	-35.9727	0.3216	-35.6778
	Clayton	0.5179	3.5857	0.4694	-6.1148	0.4792	-4.1502
	Frank	0.5199	3.9782	0.5033	0.6619	0.5254	5.0701
	Joe	0.3560	-28.8078	0.3174	-36.5214	0.3855	-22.8934
Joe	Joe	0.4368	-12.6338	0.3898	-22.0366	0.4563	-8.7415
	AMH	0.3214	-35.7296	0.3010	-39.7963	0.3028	-39.4341
	Clayton	0.5250	5.0068	0.4713	-5.7466	0.4801	-3.9762
	Frank	0.5557	11.1423	0.5327	6.5444	0.5572	11.4379
	GH	0.5091	1.8175	0.4643	-7.1373	0.5137	2.7330

Table B.8: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW PO margins ($n = 500$; $\tau = 0.5$)

True	Copula	Weibull PO		BP PO		PE PO	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3277	-1.7005	0.3276	-1.7089	0.3279	-1.6237
	Clayton	0.3561	6.8446	0.3288	-1.3677	0.3356	0.6947
	Frank	0.2855	-14.3436	0.3046	-8.6209	0.3170	-4.8964
	GH	0.1469	-55.9160	0.1512	-54.6419	0.1999	-40.0417
	Joe	0.0683	-79.5094	0.0656	-80.3158	0.0976	-70.7279
Clayton	Clayton	0.5028	0.5610	0.4927	-1.4592	0.5002	0.0363
	AMH	0.3333	-33.3333	0.3333	-33.3344	0.3333	-33.3340
	Frank	0.4379	-12.4202	0.4675	-6.4930	0.4828	-3.4345
	GH	0.2618	-47.6453	0.2677	-46.4700	0.3346	-33.0834
	Joe	0.1435	-71.3062	0.1440	-71.2094	0.2014	-59.7295
Frank	Frank	0.4859	-2.8244	0.4874	-2.5122	0.4986	-0.2830
	AMH	0.3313	-33.7380	0.3194	-36.1217	0.3214	-35.7201
	Clayton	0.4935	-1.3085	0.4513	-9.7494	0.4570	-8.5962
	GH	0.3273	-34.5401	0.3141	-37.1847	0.3664	-26.7288
	Joe	0.2130	-57.3960	0.2096	-58.0860	0.2644	-47.1251
GH	GH	0.4362	-12.7570	0.4189	-16.2286	0.4621	-7.5747
	AMH	0.3308	-33.8352	0.3185	-36.2935	0.3209	-35.8154
	Clayton	0.4967	-0.6532	0.4509	-9.8115	0.4567	-8.6664
	Frank	0.5139	2.7886	0.5093	1.8588	0.5238	4.7651
	Joe	0.3410	-31.7927	0.3304	-33.9206	0.3867	-22.6625
Joe	Joe	0.4241	-15.1818	0.4024	-19.5212	0.4545	-9.0913
	AMH	0.3250	-34.9994	0.2990	-40.2011	0.3001	-39.9732
	Clayton	0.4960	-0.7913	0.4440	-11.2092	0.4481	-10.3868
	Frank	0.5505	10.1061	0.5379	7.5739	0.5524	10.4874
	GH	0.4991	-0.1808	0.4723	-5.5423	0.5079	1.5737

Table B.9: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW YP margins ($n = 500$; $\tau = 0.5$)

True	Copula	Weibull YP		BP YP		PE YP	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3330	-0.0954	0.3279	-1.6380	0.3279	-1.6435
	Clayton	0.3465	3.9355	0.3317	-0.5027	0.3368	1.0399
	Frank	0.2876	-13.7059	0.3068	-7.9671	0.3187	-4.3908
	GH	0.1672	-49.8368	0.1488	-55.3711	0.2006	-39.8056
	Joe	0.0821	-75.3779	0.0637	-80.8966	0.0971	-70.8818
Clayton	Clayton	0.4951	-0.9845	0.4964	-0.7135	0.5016	0.3173
	AMH	0.3333	-33.3335	0.3333	-33.3357	0.3333	-33.3356
	Frank	0.4413	-11.7491	0.4708	-5.8330	0.4853	-2.9483
	GH	0.2902	-41.9507	0.2656	-46.8750	0.3369	-32.6190
	Joe	0.1671	-66.5758	0.1412	-71.7619	0.2019	-59.6159
Frank	Frank	0.4840	-3.1984	0.4889	-2.2195	0.4993	-0.1305
	AMH	0.3325	-33.4963	0.3210	-35.8042	0.3220	-35.6049
	Clayton	0.4855	-2.9045	0.4606	-7.8819	0.4641	-7.1763
	GH	0.3502	-29.9540	0.3115	-37.7095	0.3662	-26.7697
	Joe	0.2383	-52.3378	0.2068	-58.6479	0.2643	-47.1435
GH	GH	0.4464	-10.7109	0.4170	-16.6046	0.4621	-7.5788
	AMH	0.3326	-33.4714	0.3208	-35.8474	0.3217	-35.6558
	Clayton	0.4876	-2.4710	0.4617	-7.6549	0.4643	-7.1484
	Frank	0.5080	1.5981	0.5116	2.3242	0.5254	5.0764
	Joe	0.3590	-28.2099	0.3276	-34.4782	0.3866	-22.6761
Joe	Joe	0.4374	-12.5158	0.4000	-19.9959	0.4542	-9.1671
	AMH	0.3254	-34.9201	0.3007	-39.8611	0.3014	-39.7111
	Clayton	0.4889	-2.2172	0.4614	-7.7300	0.4621	-7.5874
	Frank	0.5420	8.3904	0.5402	8.0402	0.5542	10.8416
	GH	0.5041	0.8269	0.4713	-5.7431	0.5079	1.5827

Table B.10: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW PH margins ($n = 500$; $\tau = 0.75$)

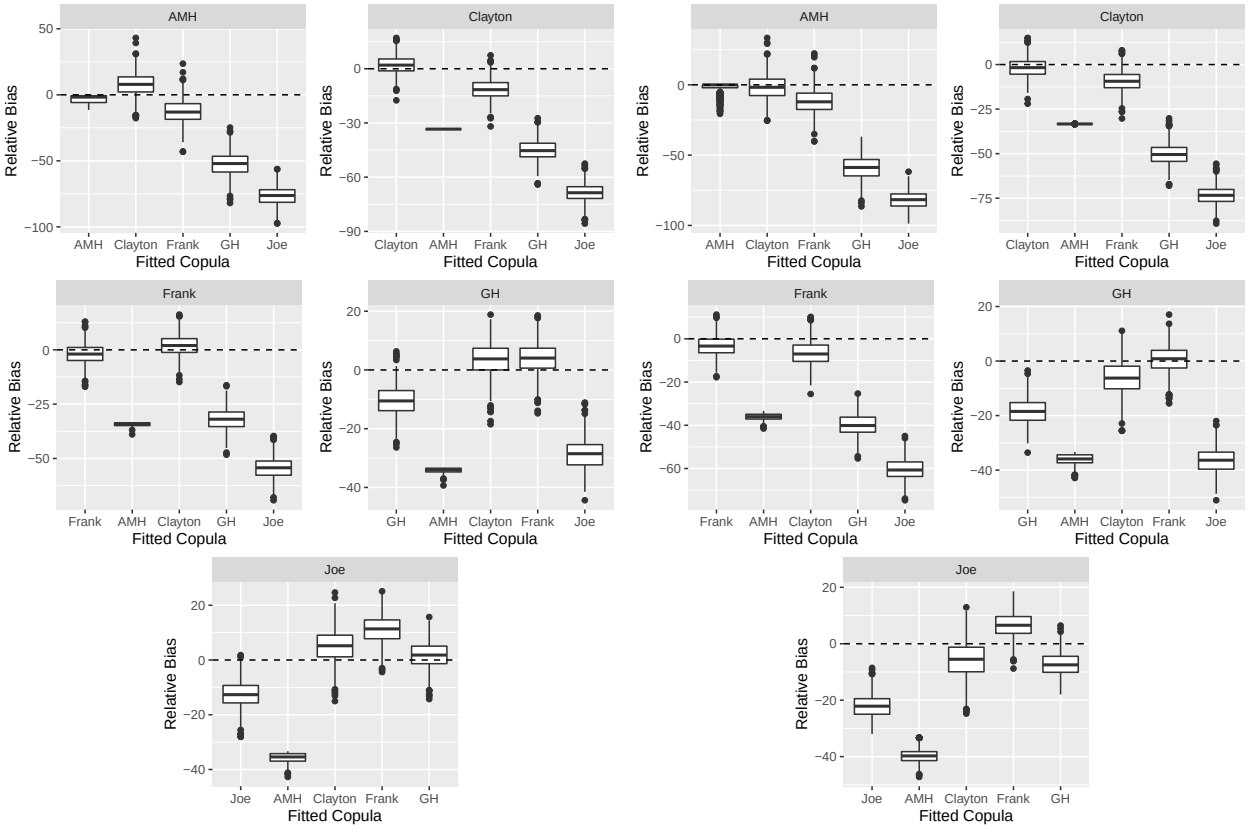
True	Copula	Weibull PH		BP PH		PE PH	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3221	-3.3813	0.3273	-1.8053	0.3283	-1.4998
	Clayton	0.3593	7.7998	0.3273	-1.8054	0.3368	1.0476
	Frank	0.2908	-12.7639	0.2946	-11.6259	0.3098	-7.0459
	GH	0.1583	-52.5037	0.1371	-58.8716	0.1918	-42.4585
	Joe	0.0782	-76.5351	0.0606	-81.8244	0.0952	-71.4451
Clayton	Clayton	0.7334	-2.2172	0.7420	-1.0648	0.7506	0.0863
	AMH	0.3333	-55.5556	0.3333	-55.5556	0.3333	-55.5556
	Frank	0.7006	-6.5810	0.7188	-4.1590	0.7310	-2.5295
	GH	0.4854	-35.2798	0.4522	-39.7042	0.5335	-28.8725
	Joe	0.3291	-56.1143	0.3026	-59.6582	0.3854	-48.6103
Frank	Frank	0.7397	-1.3668	0.7381	-1.5877	0.7496	-0.0594
	AMH	0.3324	-55.6797	0.3321	-55.7184	0.3321	-55.7138
	Clayton	0.7409	-1.2134	0.7298	-2.6994	0.7394	-1.4112
	GH	0.5712	-23.8451	0.5259	-29.8808	0.5858	-21.8877
	Joe	0.4221	-43.7227	0.3893	-48.0877	0.4595	-38.7304
GH	GH	0.7154	-4.6170	0.6725	-10.3275	0.7214	-3.8096
	AMH	0.3329	-55.6159	0.3331	-55.5857	0.3331	-55.5914
	Clayton	0.7511	0.1438	0.7379	-1.6089	0.7553	0.7106
	Frank	0.7611	1.4755	0.7480	-0.2660	0.7729	3.0595
	Joe	0.6261	-16.5184	0.5721	-23.7207	0.6428	-14.2867
Joe	Joe	0.7120	-5.0613	0.6523	-13.0299	0.7079	-5.6068
	AMH	0.3315	-55.7970	0.3294	-56.0785	0.3298	-56.0309
	Clayton	0.7656	2.0832	0.7554	0.7141	0.7759	3.4570
	Frank	0.7955	6.0650	0.7767	3.5554	0.8027	7.0242
	GH	0.7692	2.5605	0.7258	-3.2215	0.7618	1.5691

Table B.11: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW PO margins ($n = 500$; $\tau = 0.75$)

True	Copula	Weibull PO		BP PO		PE PO	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3277	-1.7005	0.3276	-1.7089	0.3279	-1.6254
	Clayton	0.3562	6.8618	0.3288	-1.3718	0.3358	0.7301
	Frank	0.2855	-14.3553	0.3046	-8.6308	0.3172	-4.8510
	GH	0.1469	-55.9233	0.1511	-54.6557	0.1998	-40.0464
	Joe	0.0683	-79.5077	0.0656	-80.3158	0.0976	-70.7147
Clayton	Clayton	0.7273	-3.0267	0.7436	-0.8499	0.7500	0.0059
	AMH	0.3333	-55.5556	0.3333	-55.5556	0.3333	-55.5556
	Frank	0.7017	-6.4432	0.7309	-2.5486	0.7413	-1.1569
	GH	0.4733	-36.8897	0.4738	-36.8210	0.5465	-27.1354
	Joe	0.3110	-58.5379	0.3196	-57.3883	0.3961	-47.1868
Frank	Frank	0.7395	-1.4031	0.7413	-1.1645	0.7495	-0.0647
	AMH	0.3330	-55.6035	0.3328	-55.6254	0.3327	-55.6415
	Clayton	0.7244	-3.4182	0.7130	-4.9328	0.7192	-4.1033
	GH	0.5612	-25.1724	0.5411	-27.8595	0.5908	-21.2210
	Joe	0.4046	-46.0471	0.4045	-46.0689	0.4659	-37.8841
GH	GH	0.7097	-5.3715	0.6837	-8.8412	0.7207	-3.9075
	AMH	0.3332	-55.5722	0.3333	-55.5649	0.3332	-55.5727
	Clayton	0.7301	-2.6486	0.7216	-3.7830	0.7311	-2.5208
	Frank	0.7582	1.0998	0.7548	0.6393	0.7722	2.9653
	Joe	0.6163	-17.8222	0.5871	-21.7265	0.6455	-13.9398
Joe	Joe	0.7068	-5.7593	0.6651	-11.3244	0.7065	-5.7990
	AMH	0.3327	-55.6392	0.3307	-55.9093	0.3311	-55.8543
	Clayton	0.7350	-2.0020	0.7278	-2.9556	0.7384	-1.5432
	Frank	0.7937	5.8288	0.7822	4.2868	0.7983	6.4384
	GH	0.7645	1.9371	0.7332	-2.2438	0.7568	0.9121

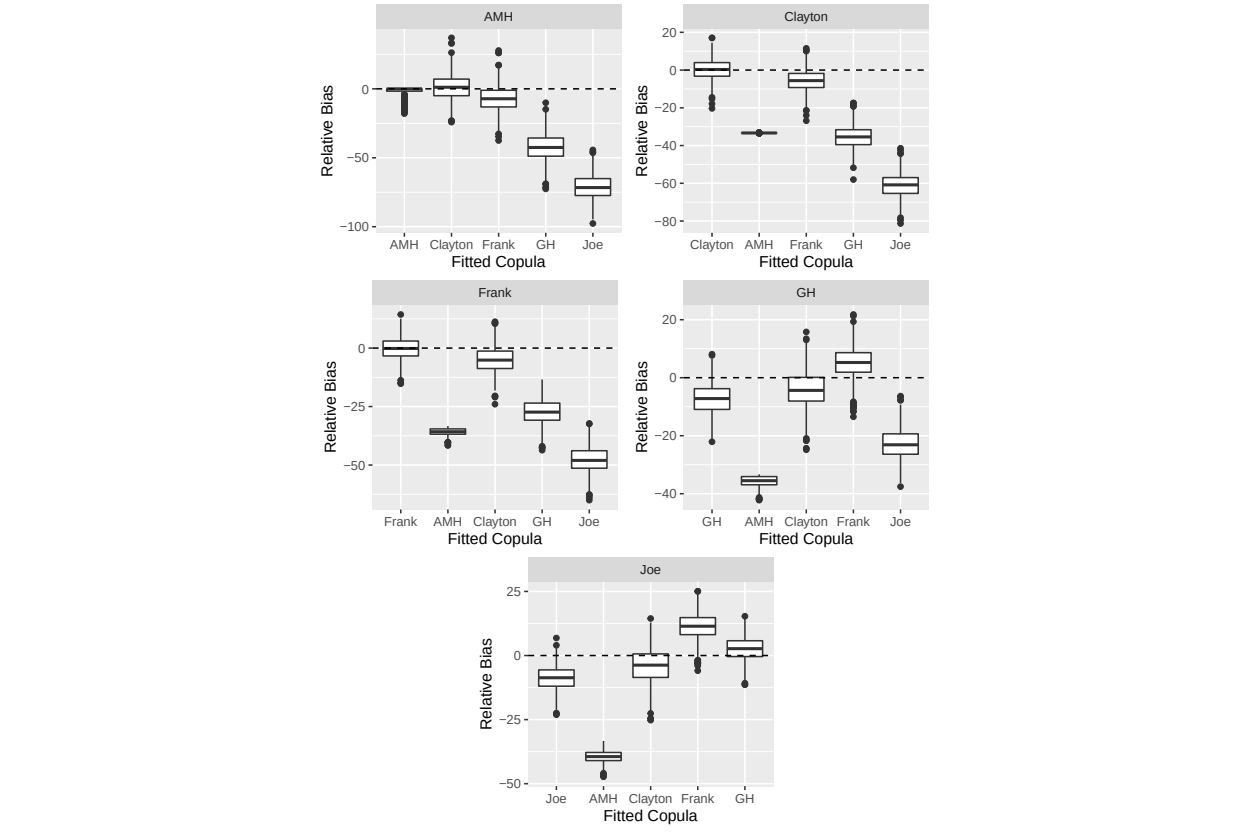
Table B.12: MC statistics for Kendall's tau estimates of fitted survival copula models over generated data from each copula model with EW YP margins ($n = 500$; $\tau = 0.75$)

True	Copula	Weibull YP		BP YP		PE YP	
		AE	ARB (%)	AE	ARB (%)	AE	ARB (%)
AMH	AMH	0.3330	-0.0956	0.3279	-1.6380	0.3279	-1.6435
	Clayton	0.3466	3.9656	0.3317	-0.5027	0.3367	1.0085
	Frank	0.2877	-13.6999	0.3067	-7.9805	0.3187	-4.3913
	GH	0.1671	-49.8642	0.1488	-55.3663	0.2008	-39.7735
	Joe	0.0821	-75.3779	0.0637	-80.8966	0.0971	-70.8695
Clayton	Clayton	0.7279	-2.9459	0.7464	-0.4780	0.7511	0.1439
	AMH	0.3333	-55.5556	0.3333	-55.5556	0.3333	-55.5556
	Frank	0.7049	-6.0117	0.7333	-2.2292	0.7430	-0.9386
	GH	0.5087	-32.1729	0.4727	-36.9697	0.5494	-26.7411
	Joe	0.3471	-53.7170	0.3178	-57.6248	0.3988	-46.8307
Frank	Frank	0.7399	-1.3442	0.7426	-0.9854	0.7500	0.0003
	AMH	0.3333	-55.5591	0.3330	-55.6007	0.3330	-55.5953
	Clayton	0.7257	-3.2364	0.7212	-3.8377	0.7252	-3.3036
	GH	0.5886	-21.5168	0.5397	-28.0397	0.5917	-21.1109
	Joe	0.4413	-41.1621	0.4031	-46.2588	0.4670	-37.7399
GH	GH	0.7154	-4.6153	0.6829	-8.9495	0.7211	-3.8553
	AMH	0.3333	-55.5563	0.3333	-55.5643	0.3333	-55.5623
	Clayton	0.7301	-2.6478	0.7295	-2.7376	0.7363	-1.8215
	Frank	0.7558	0.7778	0.7570	0.9290	0.7735	3.1286
	Joe	0.6303	-15.9590	0.5862	-21.8365	0.6467	-13.7793
Joe	Joe	0.7140	-4.7942	0.6640	-11.4639	0.7070	-5.7289
	AMH	0.3332	-55.5684	0.3318	-55.7664	0.3321	-55.7244
	Clayton	0.7411	-1.1876	0.7431	-0.9180	0.7508	0.1130
	Frank	0.7907	5.4267	0.7833	4.4432	0.7991	6.5496
	GH	0.7677	2.3564	0.7331	-2.2577	0.7578	1.0396



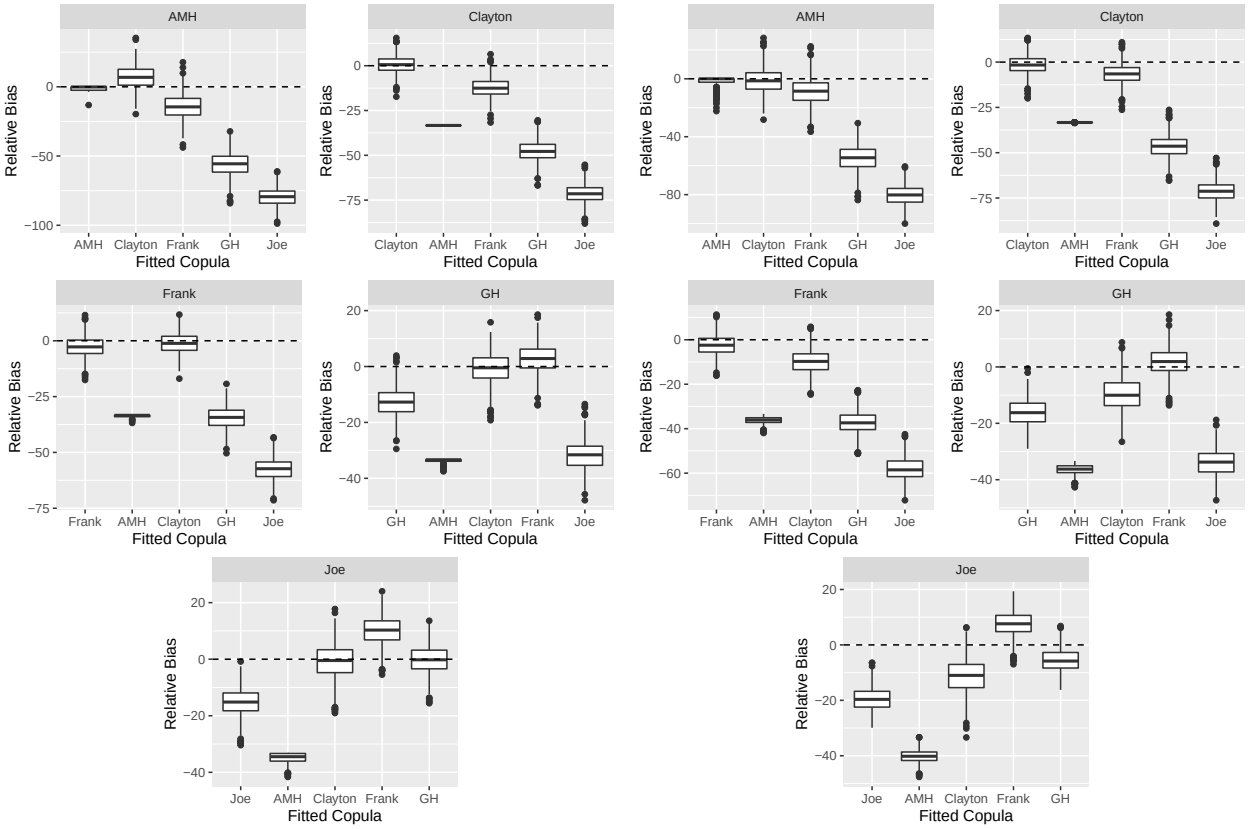
Weibull PH margins

BP PH margins



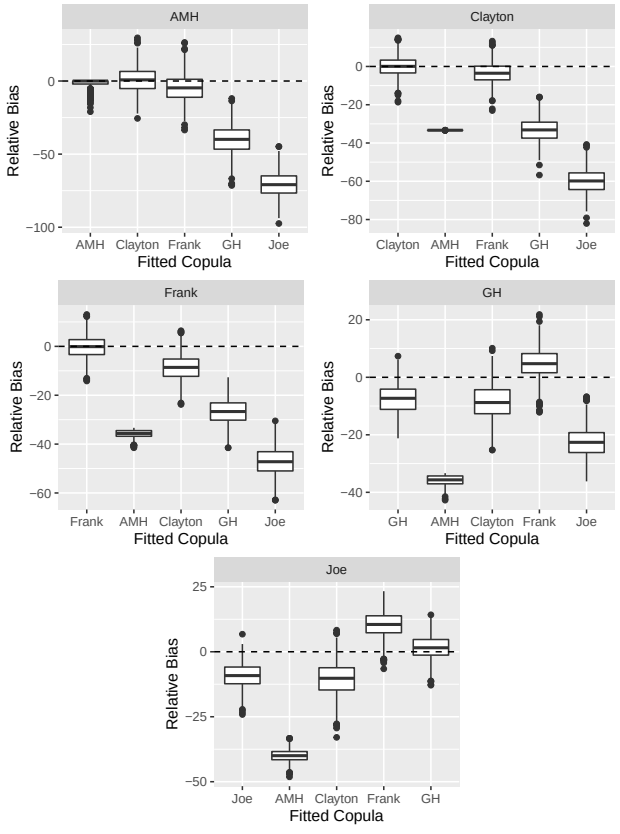
PE PH margins

Figure B.1: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW PH margins ($n = 500$; $\tau = 0.5$)



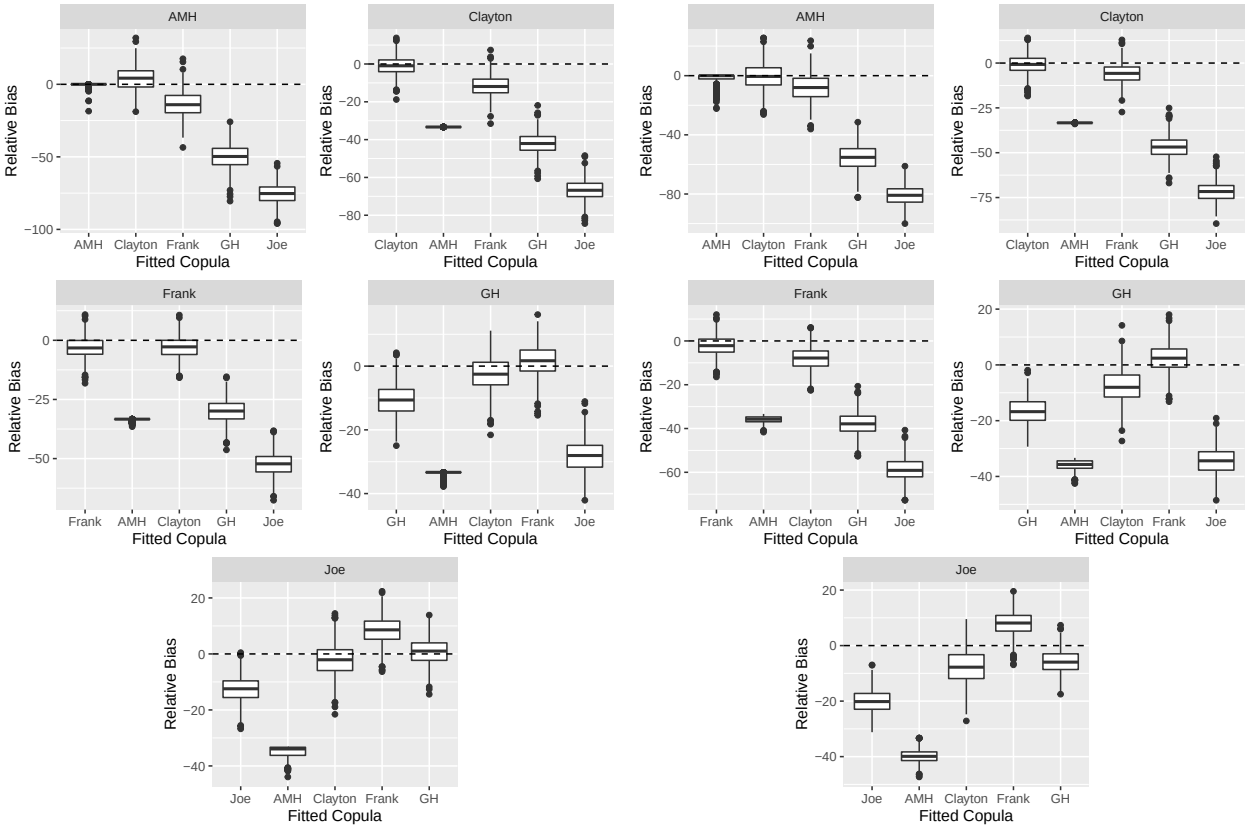
Weibull PO margins

BP PO margins



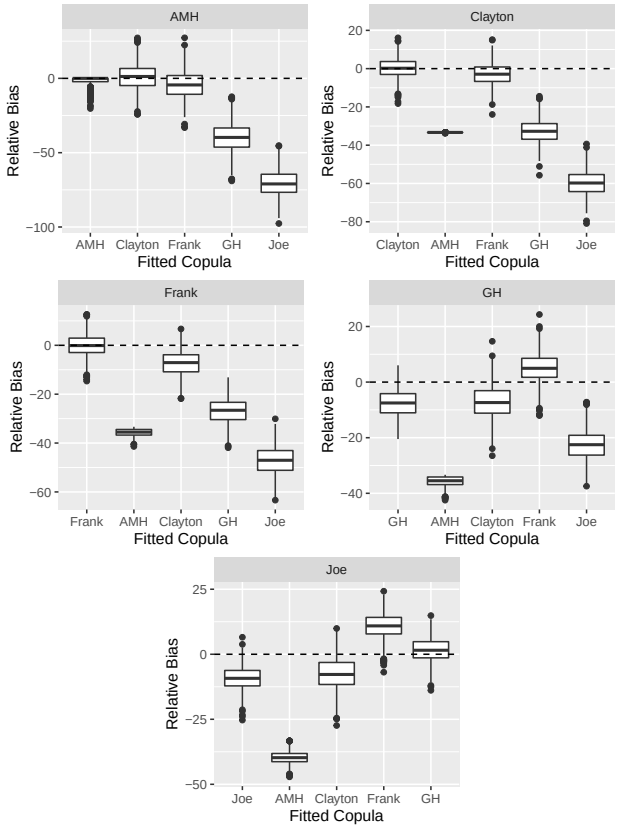
PE PO margins

Figure B.2: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW PO margins ($n = 500$; $\tau = 0.5$)



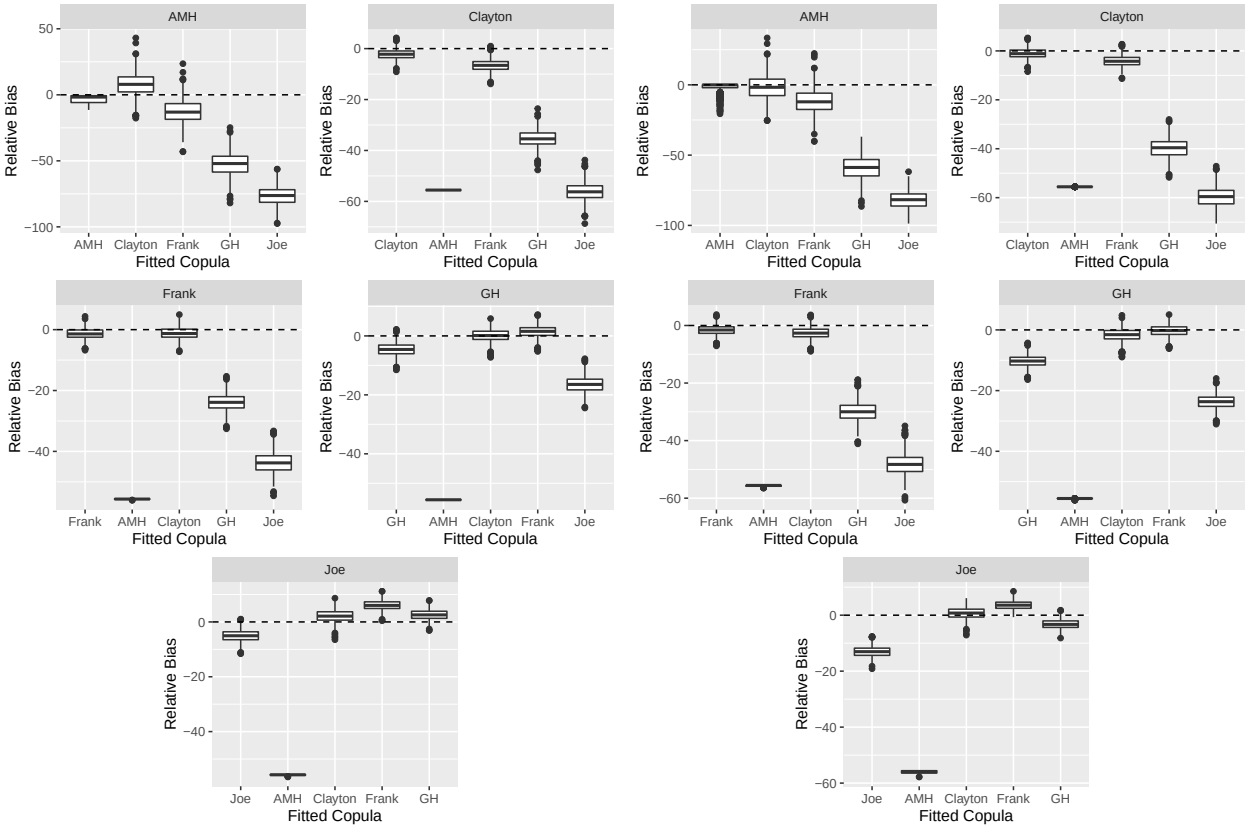
Weibull YP margins

BP YP margins



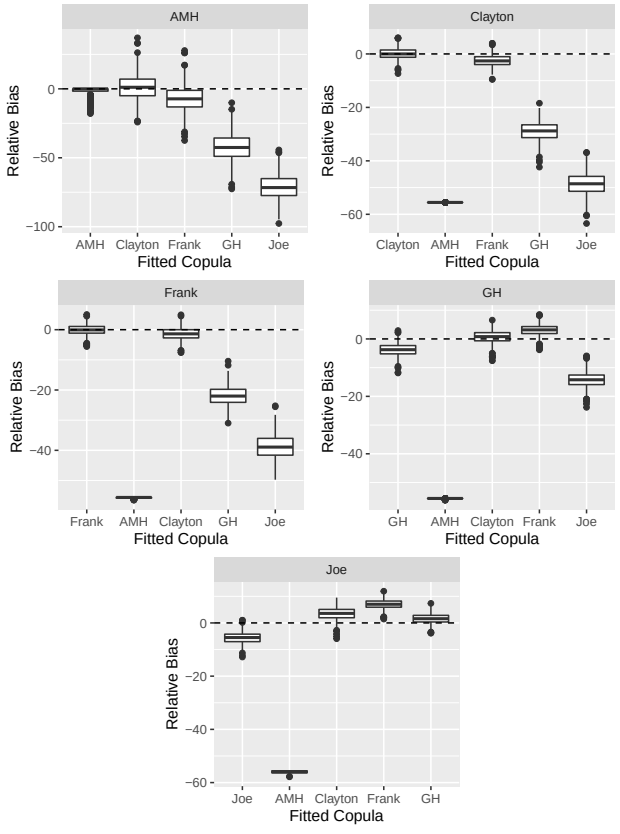
PE YP margins

Figure B.3: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW YP margins ($n = 500$; $\tau = 0.5$)



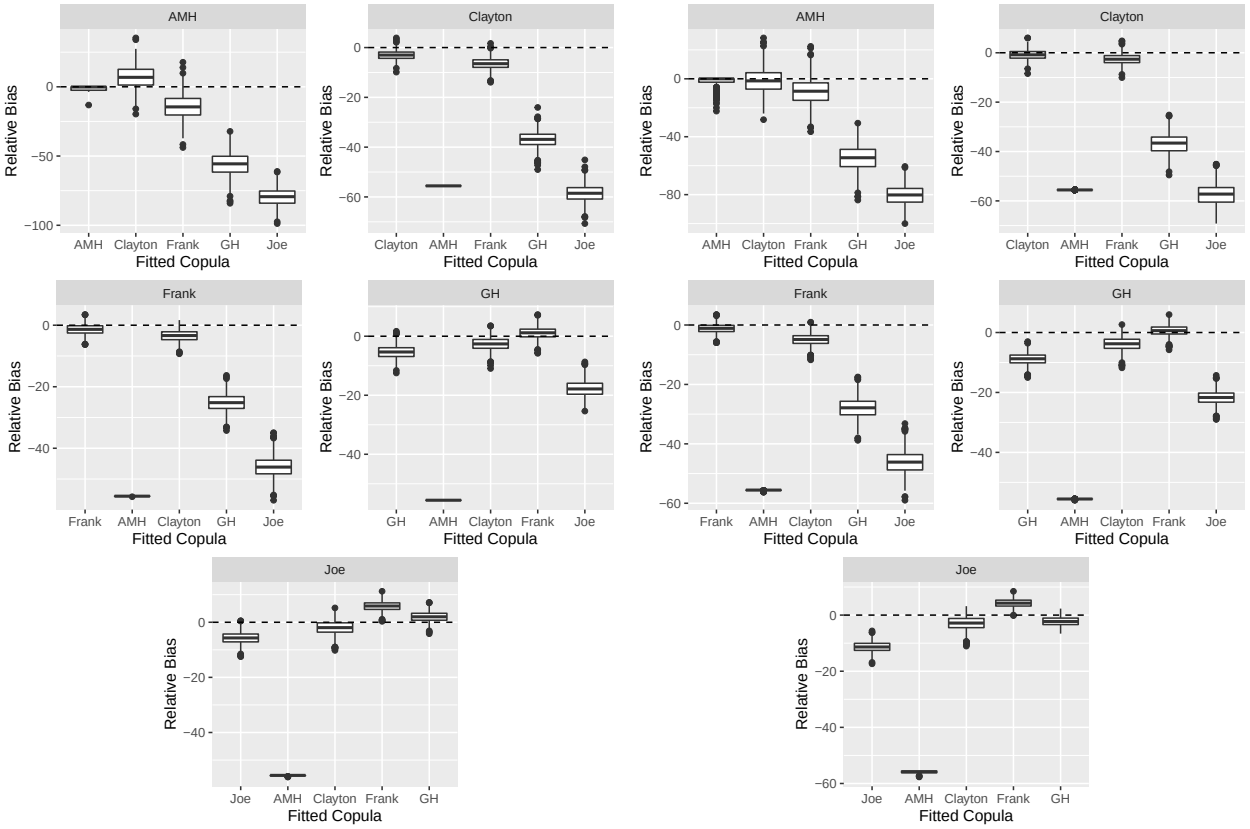
Weibull PH margins

BP PH margins



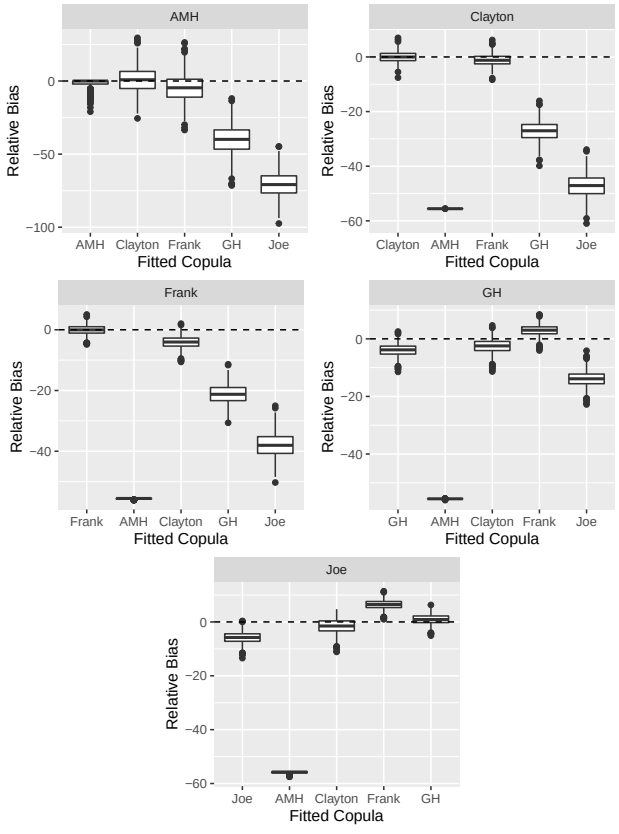
PE PH margins

Figure B.4: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW PH margins ($n = 500$; $\tau = 0.75$)



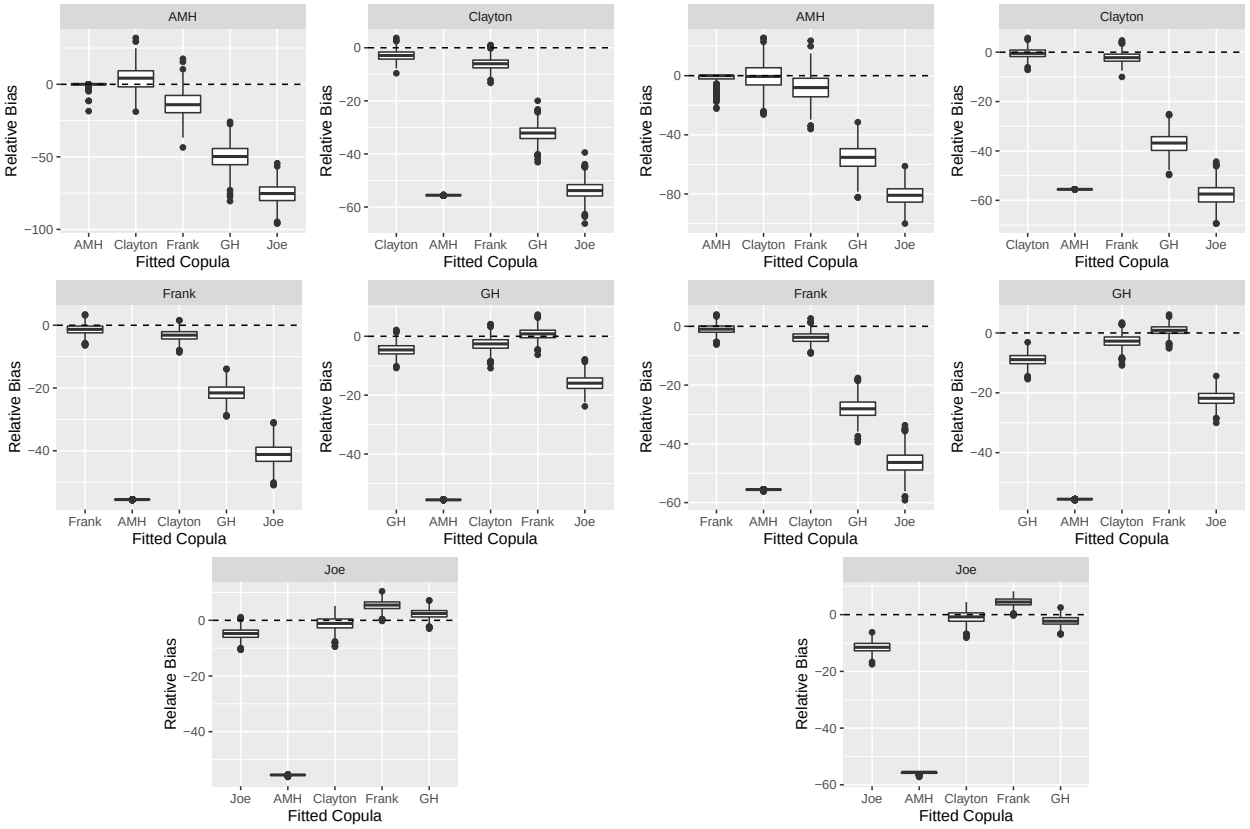
Weibull PO margins

BP PO margins



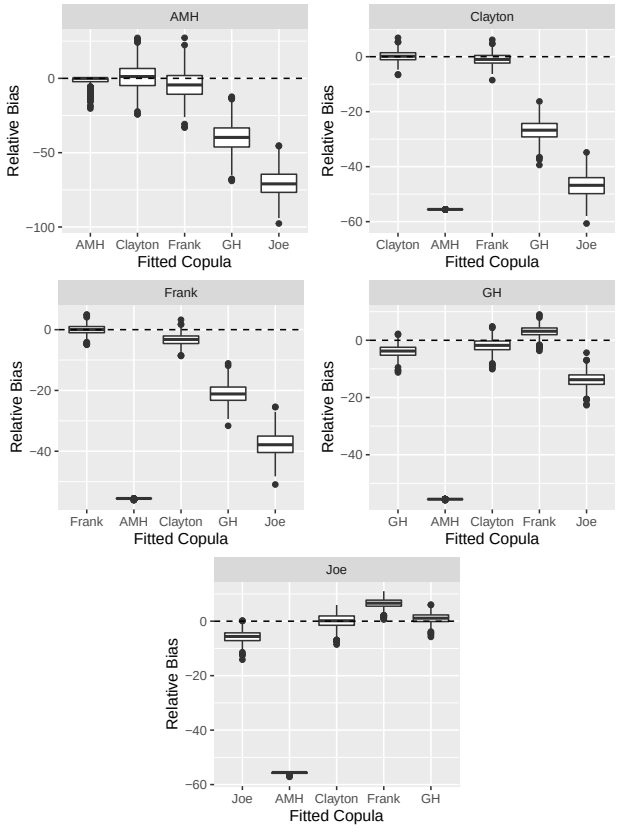
PE PO margins

Figure B.5: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW PO margins ($n = 500$; $\tau = 0.75$)



Weibull YP margins

BP YP margins



PE YP margins

Figure B.6: Kendall's τ estimates for fitted survival copula models over data generated from each copula model with EW YP margins ($n = 500$; $\tau = 0.75$)

Table B.13: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW PH margins ($n = 500; \tau = 0.5$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	6.0366	0.3392	4.1335	0.4861
	PO	29.3335	0.0135	33.2383	0.0007
Clayton	PH	5.5854	0.3775	4.1538	0.4848
	PO	29.5332	0.0093	34.1537	0.0005
Frank	PH	6.0102	0.3560	4.1935	0.4756
	PO	26.0691	0.0186	30.9635	0.0008
GH	PH	5.9059	0.3522	4.2574	0.4698
	PO	24.5219	0.0180	29.7684	0.0014
Joe	PH	6.8017	0.3171	4.7982	0.4322
	PO	20.8203	0.0519	28.3163	0.0019

Table B.14: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW PO margins ($n = 500; \tau = 0.5$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	27.2364	0.0085	26.6160	0.0060
	PO	0.9229	0.7797	4.4759	0.4503
Clayton	PH	27.4125	0.0049	27.1481	0.0034
	PO	0.8394	0.7759	4.3992	0.4492
Frank	PH	27.2115	0.0077	25.9299	0.0054
	PO	1.2819	0.7457	4.4113	0.4541
GH	PH	30.0100	0.0055	27.8126	0.0044
	PO	1.7602	0.7161	4.6868	0.4252
Joe	PH	35.2895	0.0017	31.3643	0.0022
	PO	1.8001	0.7091	4.7699	0.4072

Table B.15: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW YP margins ($n = 500; \tau = 0.5$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	123.5398	< 0.0001	116.1681	< 0.0001
	PO	92.1303	< 0.0001	88.9388	< 0.0001
Clayton	PH	125.9010	< 0.0001	119.2945	< 0.0001
	PO	103.1754	< 0.0001	99.9462	< 0.0001
Frank	PH	115.6478	< 0.0001	109.0777	< 0.0001
	PO	87.0588	< 0.0001	83.3261	< 0.0001
GH	PH	117.8532	< 0.0001	110.4258	< 0.0001
	PO	86.7865	< 0.0001	82.2861	< 0.0001
Joe	PH	126.4209	< 0.0001	118.7282	< 0.0001
	PO	92.0491	< 0.0001	87.0614	< 0.0001

Table B.16: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW PH margins ($n = 500; \tau = 0.75$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	6.0405	0.3390	4.1592	0.4848
	PO	29.2964	0.0134	33.3045	0.0007
Clayton	PH	5.7289	0.3594	4.3249	0.4679
	PO	30.6600	0.0108	35.6394	0.0011
Frank	PH	6.5995	0.3241	4.1748	0.4871
	PO	23.7234	0.0295	29.2761	0.0019
GH	PH	7.6009	0.2631	4.7821	0.4350
	PO	23.8162	0.0230	29.6960	0.0016
Joe	PH	10.5771	0.1569	5.8501	0.3600
	PO	18.5547	0.0638	27.0367	0.0030

Table B.17: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW PO margins ($n = 500; \tau = 0.75$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	27.2287	0.0086	26.5805	0.0060
	PO	0.9229	0.7797	4.4834	0.4501
Clayton	PH	28.6215	0.0060	29.7224	0.0025
	PO	0.9756	0.7666	4.5895	0.4332
Frank	PH	29.0585	0.0044	27.0385	0.0035
	PO	1.4139	0.7319	4.5591	0.4366
GH	PH	36.4658	0.0021	31.9641	0.0023
	PO	1.4755	0.7189	4.9458	0.4063
Joe	PH	46.1807	0.0001	38.6782	0.0005
	PO	1.7272	0.7056	5.0648	0.4022

Table B.18: LR tests for PH and PO classes against YP class, given fitted survival copula models over generated data from each copula model with EW YP margins ($n = 500; \tau = 0.75$)

Copula	Class	BP Fitting		PE Fitting	
		LR stat.	P-value	LR stat.	P-value
AMH	PH	123.5554	< 0.0001	116.2059	< 0.0001
	PO	92.0240	< 0.0001	88.9409	< 0.0001
Clayton	PH	147.7122	< 0.0001	143.4252	< 0.0001
	PO	123.3079	< 0.0001	121.3521	< 0.0001
Frank	PH	122.9050	< 0.0001	116.6629	< 0.0001
	PO	87.8778	< 0.0001	84.4777	< 0.0001
GH	PH	129.2697	< 0.0001	120.1143	< 0.0001
	PO	89.3218	< 0.0001	85.0889	< 0.0001
Joe	PH	143.1140	< 0.0001	133.6036	< 0.0001
	PO	94.9875	< 0.0001	89.4466	< 0.0001

APPENDIX C

The copSurv package

This chapter introduces the R package `copSurv`, with implemented routines for fitting and analyzing survival copula models from the proposed class in this work. The package was created by integrating R functions with the Stan platform (Carpenter et al., 2017), which is strongly based in the C++ programming language, through the `rstan` R package (Stan Development Team, 2020a). This integration was done in order to obtain faster estimation results from the log-likelihood maximization (compared to an implementation using only the R language), and with lesser computational effort. In addition, the `copSurv` package imports specific routines from other R packages to support its internal computations, such as the `copula` package (Hofert et al., 2020) to generate from all Archimedean copulas and compute the Kendall's τ correlation from the original θ estimates; the `survival` package (Therneau, 2021) to borrow functions for clustering subject information in the regression structure and to analyze marginal survival functions; the `snowfall` package (Knaus, 2015) to parallelize all MC computations through as many CPU cores as possible; and the `tidyverse` set of packages (Wickham et al., 2019) to enhance data storage and manipulation. The `copSurv` package can be installed through the link `https://github.com/wrmfstat/copSurv`. Then, open R and put the commands below:

```
install.packages("devtools")
devtools::install\_github("wrmfstat/copSurv")
library(copSurv)
```

The function `survcop` is the main routine from the `copSurv` package for fitting any particular survival copula model belonging to the proposed class in this work. The `formula` argument in `survcop` takes the same structure for variable specification from the `survival` package to provide a more familiar environment to the user. The data to be fitted is passed to the scope of `survcop` through a `data.frame` object class. Additional arguments can be passed directly to Stan in order to apply a maximum likelihood method with

`rstan::optimizing`, or a Markov Chain Monte Carlo method through `rstan::sampling` (to be developed), inside the `survcop` routine. As seen in Chapter 2, the polynomial degree for a BP model fitting (*i.e.*, the highest basis order), or the number of intervals for a PE model, can be chosen in the `m` argument by the user, but it must be an integer positive. If not specified by the user, `m` is set as the smallest integer greater than $n^{0.4}$, where n is the sample size, following Osman and Ghosh (2012), and at most equal to 15. Setting a degree value too high for a BP model can cause numerical instability and even to a failure in maximizing the log-likelihood function. As with the number of baseline parameters m for a semiparametric model, some arguments are optional to provide by the user, such as the vector of maximum times `time_max` and the list of time grids `rho` for each copula margin, but they are computed internally if not. For any object being an output from the `survcop` function, the class `scm` was defined to extend some S3 methods for inference and analysis (including AIC computations and LR tests). The arguments for the `survcop` function are set as follows:

```
survcop(formula, data, d=2, approach=c("ML", "Bayes"),
        copula=c("AMH", "Clayton", "Frank", "GH", "Joe"), baseline=c("W", "BP", "PE"),
        time_max=NULL, m=NULL, rho=NULL, survreg=c("PH", "PO", "YP"),
        hessian=TRUE, init=NULL,
        hp=list(h1_psi=0, h2_psi=4, h1_phi=0, h2_phi=4, h1_gamma=0, h2_gamma=2), ...)
```

- `formula`: a survival copula object of class `formula`.
- `data`: an optional data frame containing all variables. If not specified, they are taken from `formula`.
- `d`: dimension of the observed multivariate data (default is `d=2`).
- `approach`: the inference approach to be used (`ML`: maximum likelihood (frequentist); `Bayes`: Bayesian).
- `copula`: the Archimedean copula to be fitted.
- `baseline`: the baseline distribution for each copula margin.
- `time_max`: the vector of maximum times for each copula margin.
- `m`: the degree for the BP basis, or the number of intervals on the PE model for each copula margin.
- `rho`: a list with the time grids for the PE model.
- `survreg`: regression model class for each copula margin.
- `hessian`: logical. If `TRUE` (default), a hessian matrix is returned when `approach="ML"`.
- `init`: initial value when `approach="ML"`.
- `h_ps`: a list with hyperparameters when `approach="Bayes"`. If not specified, default values are used.
- `...` : arguments passed to either `rstan::optimizing` or `rstan::sampling`.

To fit a survival copula model with the `survcop` function, consider the following code for the "best" model fitted to the Ganzfried et al. (2013) ovarian cancer data in Chapter 6:

```
> fitGHPEYP = survcop(formula = Surv(time, status) ~ cxcl12 + pltx + margin + cluster(id),
                      data=data, approach="ML", copula="GH", baseline="PE", survreg="YP")
```

Applying some proper S3 methods to the fitted model object (class `scm`), the following results are obtained:

```
> coef(gpeyp)
psi[1,1] psi[2,1] psi[1,2] psi[2,2] phi[1,1] phi[2,1] phi[1,2] phi[2,2]    theta
  0.0900  -1.7650   0.2372  -1.2938   1.8903   8.1402  -0.1160  -0.4670   1.4998
> se(gpeyp)
psi[1,1] psi[2,1] psi[1,2] psi[2,2] phi[1,1] phi[2,1] phi[1,2] phi[2,2]    theta
  0.0620   0.3898   0.1087   0.4460   0.7537  14.1470   0.1414   0.6098   0.1595
> confint(gpeyp)
             2.5%   97.5%
psi[1,1]  -0.0316  0.2116
psi[2,1]  -2.5289 -1.0011
psi[1,2]   0.0241  0.4503
psi[2,2]  -2.1680 -0.4196
phi[1,1]   0.4131  3.3675
phi[2,1] -19.5874 35.8678
phi[1,2]  -0.3932  0.1612
phi[2,2]  -1.6621  0.7281
theta      1.1871  1.8124
> AIC(gpeyp)
[1] 8467.78
> anova(gpeph, gpeyp)
Copula model: GH
Baseline distribution: PE
---
Model 1 ( PH ):  Surv(time, status) ~ cxcl12 + pltx + margin + cluster(id)
Model 2 ( YP ):  Surv(time, status) ~ cxcl12 + pltx + margin + cluster(id)
---
           loglik    LR df      P-value
Model 1: -4234.70 19.62  4 0.000594 ***
Model 2: -4224.89     -  -        -
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

For the programmed code that yielded the results above, $\boldsymbol{\beta}^{(S)}$ and $\boldsymbol{\beta}^{(L)}$ are denoted by `psi` and `phi`, respectively, following the notation of Demarqui and Mayrink (2021) for the YP model; `psi[2,1]` denotes the $\boldsymbol{\beta}_{12}^{(S)}$ regression term on table 6.3, *i.e.*, the order of regression parameters and copula margins were exchanged to ensure more general matrix operations in Stan. Finally, `gpeph` denotes a fitted GH PE PH model object.

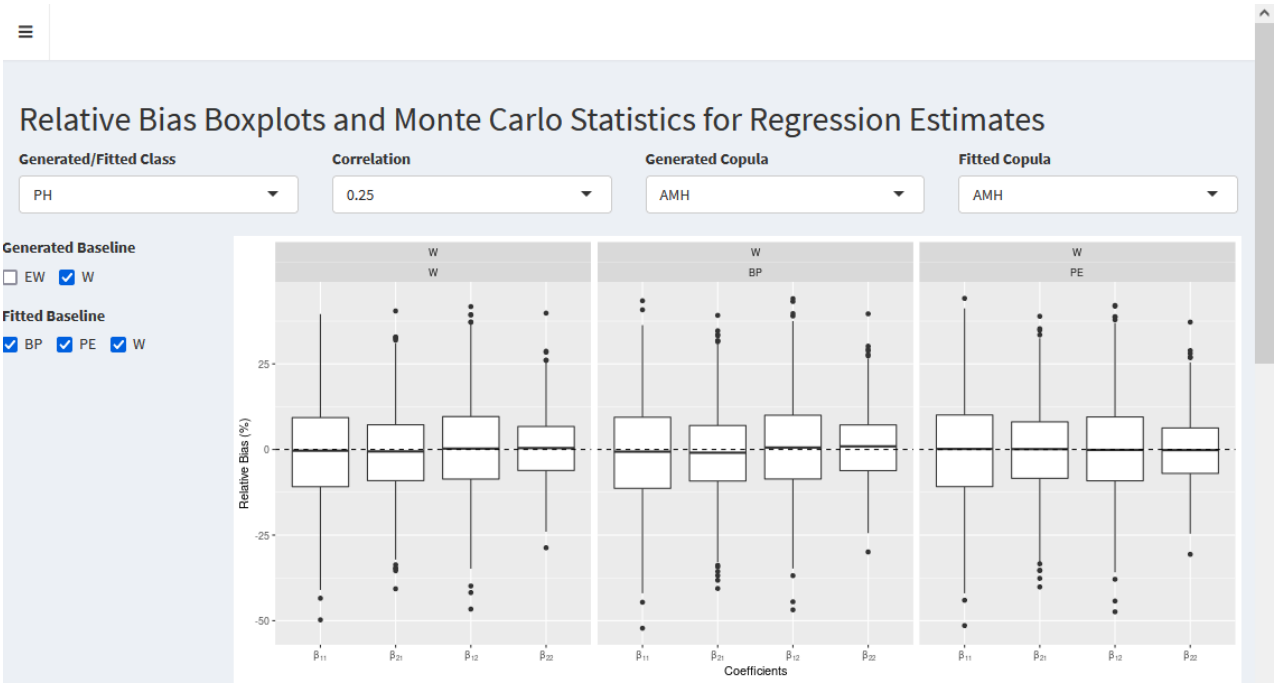
APPENDIX D

Visualization of Simulation Results in Shiny

In the pages that follows, Shiny tabs for some simulation results, given specific scenarios for data generation and model fitting, are presented along with instructions on how to visualize other possible scenarios.

Starting from regression parameter estimates, boxplots for the relative bias and a table summarizing the MC statistics are shown in Figure D.1. From the bars just below the title, the user can select one of the three regression model classes (PH, PO or YP, being always the same for generation and fitting), three Kendall's τ correlation values (0.25, 0.50 or 0.75), five Archimedean copulas used for generation (AMH, Clayton, Frank, GH or Joe), and the same options for the fitted copula (including scenarios with distinct copulas for generation and fitting). By default, all bars set their corresponding first option. On the left of the boxplots, the user can mark one of (or both) the two generated baseline models (EW or Weibull, with the last denoted by W and set the default), and one of the three fitted baseline models (Weibull, again denoted by W, BP or PE, with all of them marked by default).

For the boxplots and MC statistics for AIC values in Figure D.2, the scheme is pretty much similar, except that now is possible to visualize all fitted copula models (given one for generation) in the same plot (now there is one measure instead of four, or eight, as seen for regression parameter estimates). Thus, the bar for select of the fitted copula is replaced by a new mark section, which selects by default all five Archimedean copulas. Since it is of interest to evaluate the proportion of choice for each fitted copula given the generated one, a single fitted baseline is marked by default. For correlation estimates in Figure D.3, the same idea is applied, but now all fitted baseline models are selected by default. Finally, for the LR results on Figure D.4, the generated and fitted regression model classes can be different instead of the copula model (which is always the same). Thus, an additional mark section appear for select of the smaller class (PH or PO) to be tested against the YP model (both are marked by default).



Boxplots

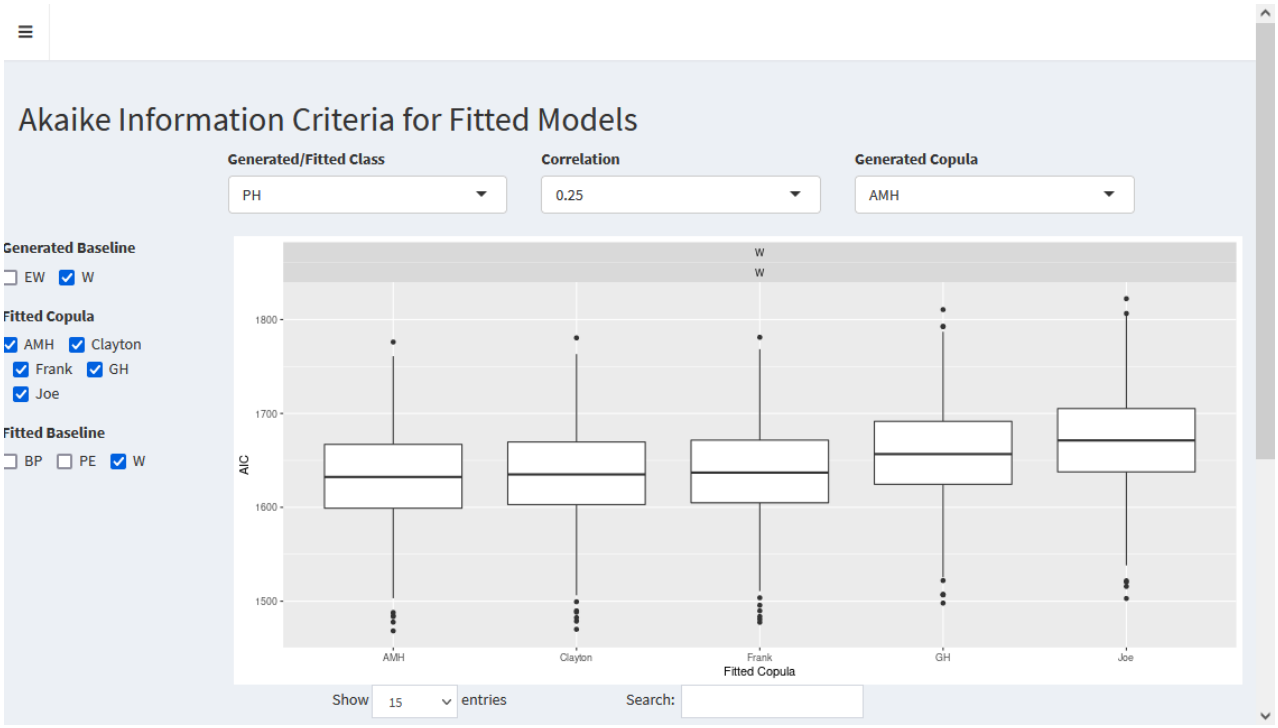
Show 24 entries Search:

G.Copula	G.Baseline	F.Copula	F.Baseline	Parameter	Real	AE	SDE	ASE	ARB	ALB	AUB	CR
AMH	W	AMH	W	β_{11}	-0.7	-0.7069	0.1041	0.1048	-0.9840	-0.9123	-0.5015	95.1807
AMH	W	AMH	W	β_{21}	0.4	0.4016	0.0545	0.0536	0.4000	0.2966	0.5066	94.1767
AMH	W	AMH	W	β_{12}	-0.9	-0.9082	0.1088	0.1081	-0.9098	-1.1200	-0.6964	94.5783
AMH	W	AMH	W	β_{22}	0.6	0.6031	0.0572	0.0568	0.5238	0.4918	0.7145	95.1807
AMH	W	AMH	BP	β_{11}	-0.7	-0.7089	0.1055	0.1061	-1.2776	-0.9169	-0.5010	94.9000
AMH	W	AMH	BP	β_{21}	0.4	0.4026	0.0553	0.0542	0.6474	0.2963	0.5088	93.8000
AMH	W	AMH	BP	β_{12}	-0.9	-0.9110	0.1111	0.1099	-1.2215	-1.1264	-0.6956	94.8000
AMH	W	AMH	BP	β_{22}	0.6	0.6049	0.0594	0.0579	0.8216	0.4914	0.7184	94.8000
AMH	W	AMH	PE	β_{11}	-0.7	-0.7046	0.1052	0.1059	-0.6551	-0.9121	-0.4971	94.9648
AMH	W	AMH	PE	β_{21}	0.4	0.4000	0.0548	0.0541	-0.0094	0.2939	0.5060	94.3605
AMH	W	AMH	PE	β_{12}	-0.9	-0.9036	0.1102	0.1096	-0.3975	-1.1184	-0.6888	95.1662
AMH	W	AMH	PE	β_{22}	0.6	0.5993	0.0583	0.0577	-0.1152	0.4862	0.7124	94.8640
G.Copula	G.Baseline	F.Copula	F.Baseline	Parameter	Real	AE	SDE	ASE	ARB	ALB	AUB	CR

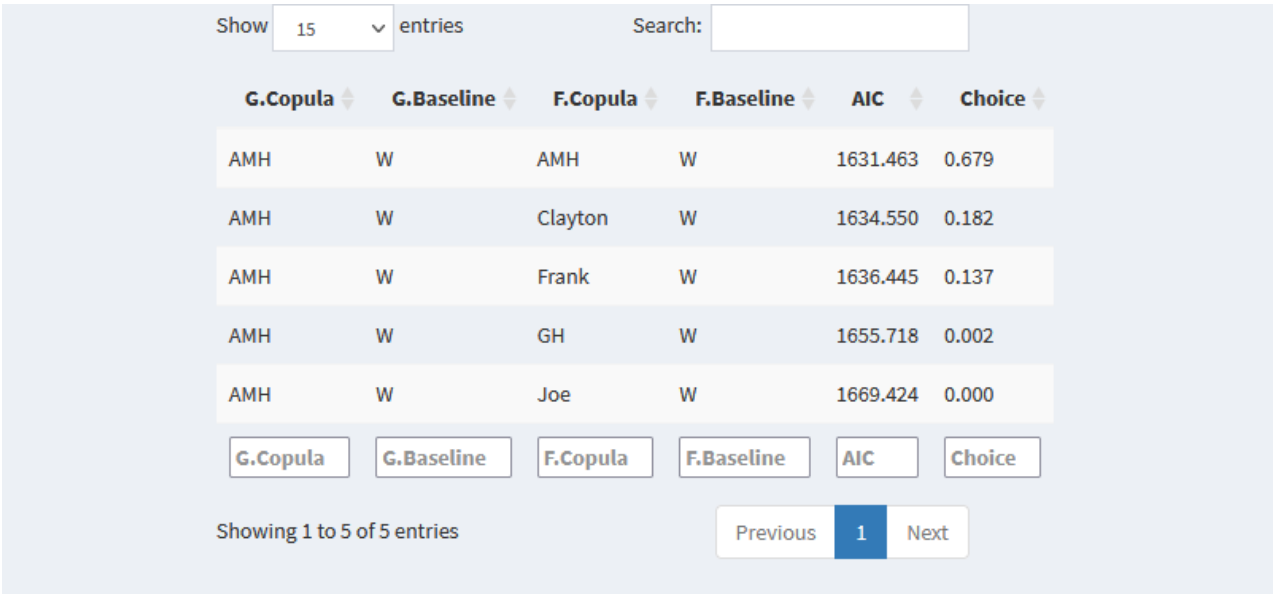
Showing 1 to 12 of 12 entries Previous 1 Next

MC Statistics

Figure D.1: Example of Shiny screen for regression parameter estimates

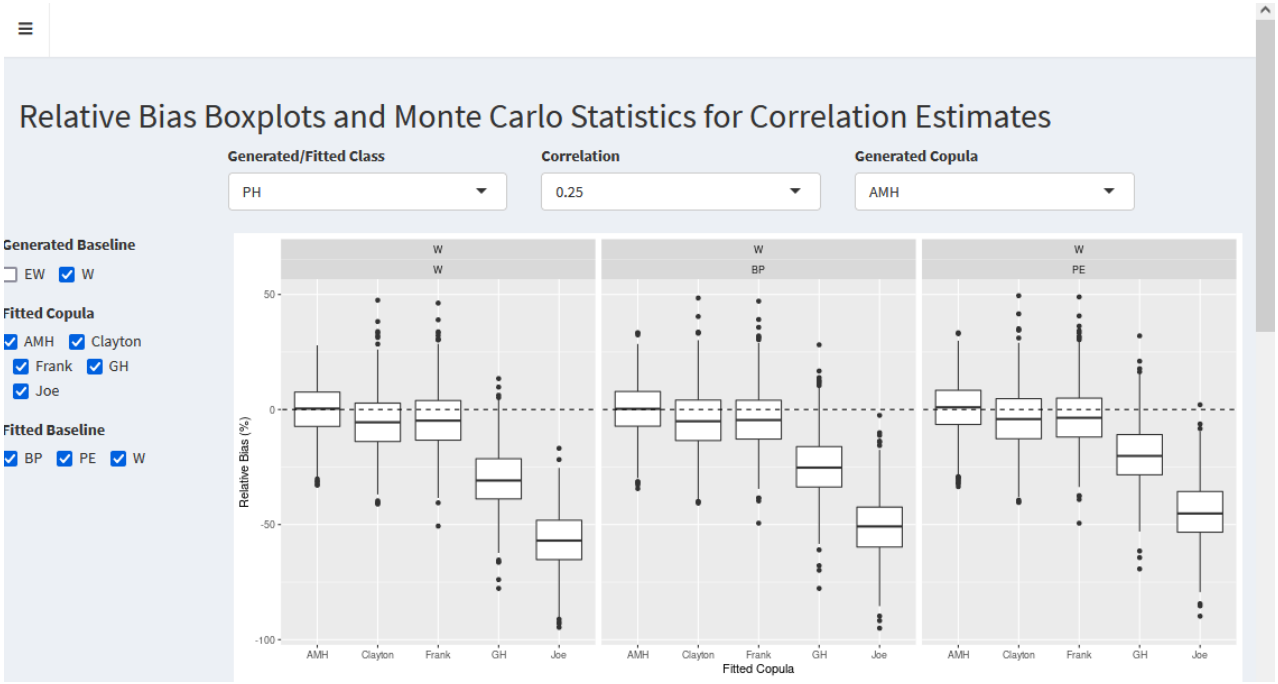


Boxplots

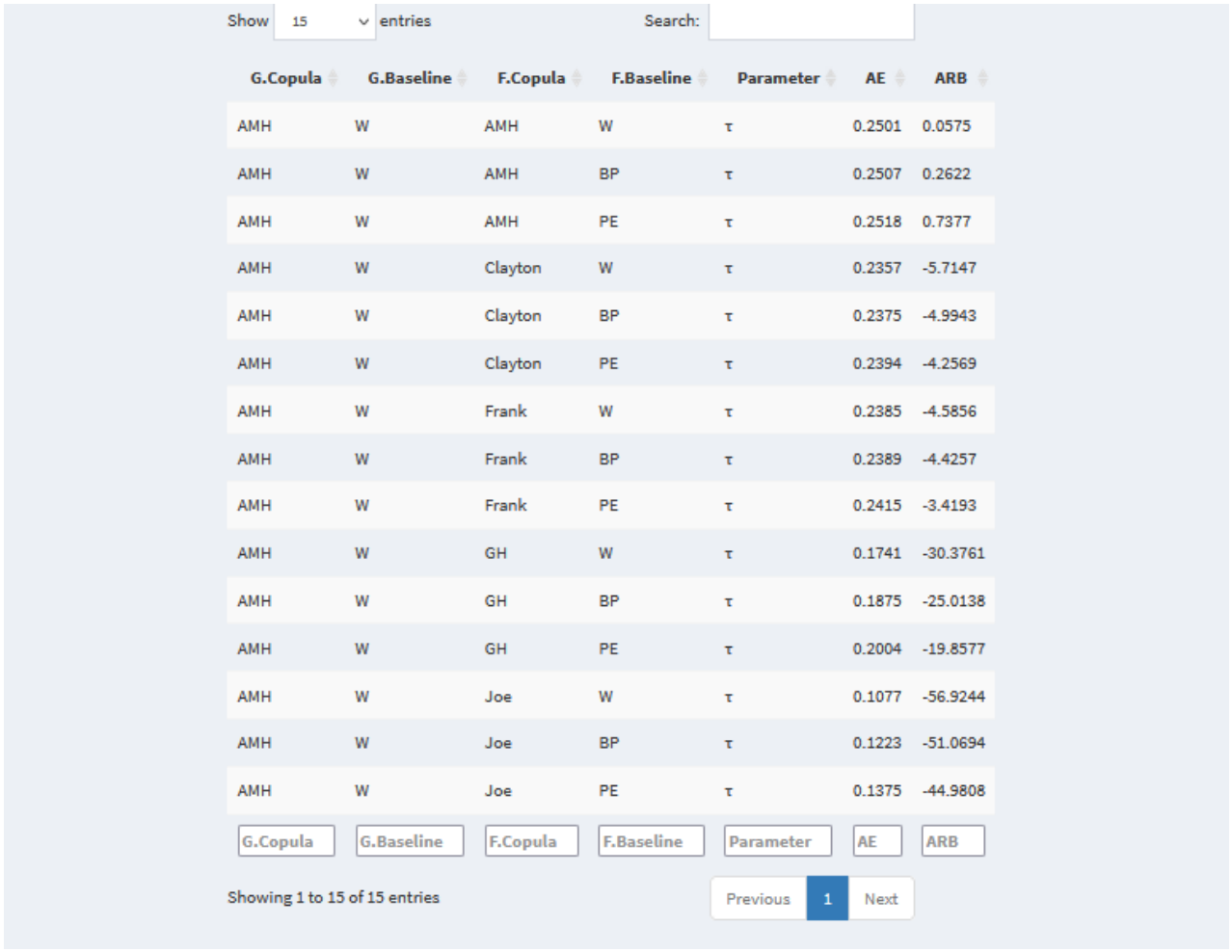


MC Statistics

Figure D.2: Example of Shiny screen for AIC values



Boxplots



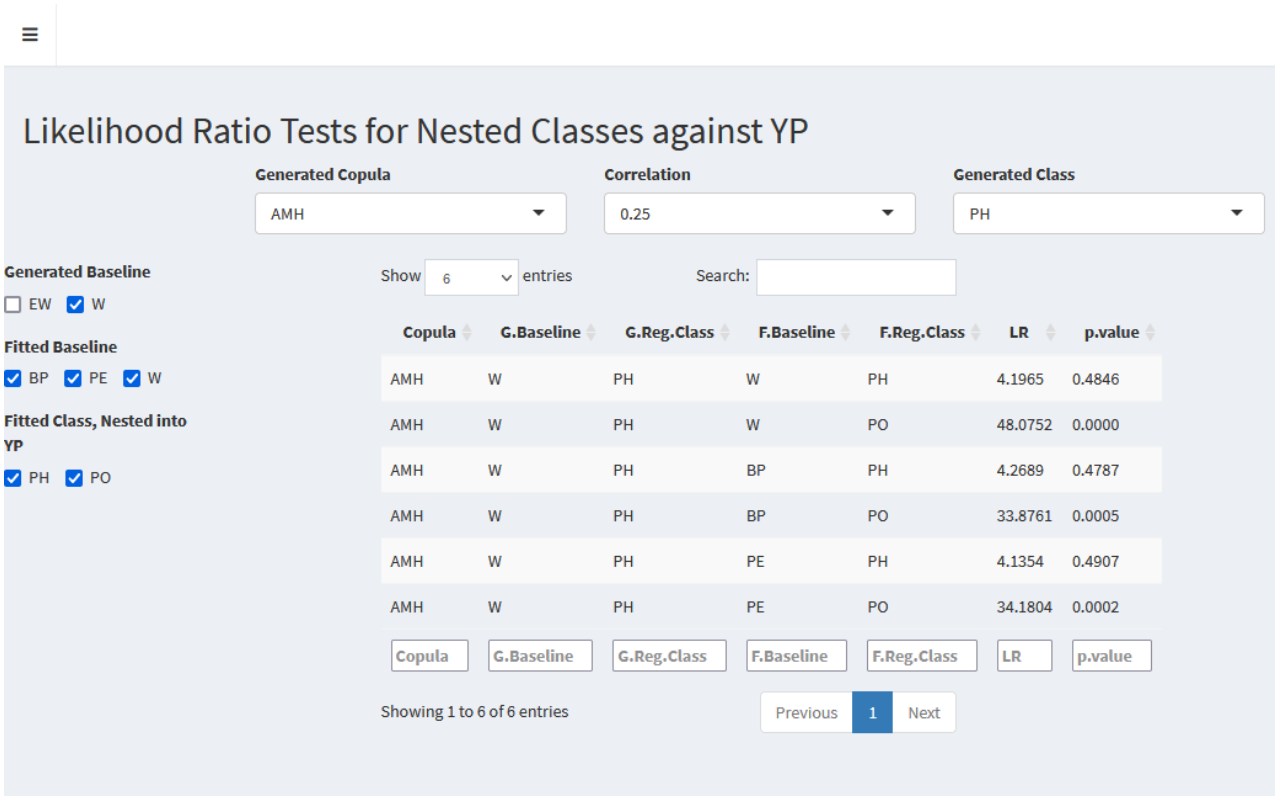
Show 15 entries Search:

G.Copula	G.Baseline	F.Copula	F.Baseline	Parameter	AE	ARB
AMH	W	AMH	W	τ	0.2501	0.0575
AMH	W	AMH	BP	τ	0.2507	0.2622
AMH	W	AMH	PE	τ	0.2518	0.7377
AMH	W	Clayton	W	τ	0.2357	-5.7147
AMH	W	Clayton	BP	τ	0.2375	-4.9943
AMH	W	Clayton	PE	τ	0.2394	-4.2569
AMH	W	Frank	W	τ	0.2385	-4.5856
AMH	W	Frank	BP	τ	0.2389	-4.4257
AMH	W	Frank	PE	τ	0.2415	-3.4193
AMH	W	GH	W	τ	0.1741	-30.3761
AMH	W	GH	BP	τ	0.1875	-25.0138
AMH	W	GH	PE	τ	0.2004	-19.8577
AMH	W	Joe	W	τ	0.1077	-56.9244
AMH	W	Joe	BP	τ	0.1223	-51.0694
AMH	W	Joe	PE	τ	0.1375	-44.9808

Showing 1 to 15 of 15 entries Previous 1 Next

MC Statistics

Figure D.3: Example of Shiny screen for correlation estimates



MC Statistics

Figure D.4: Example of Shiny screen for LR tests