The Power Variance Function Copula Model in Bivariate Survival Analysis: An application to Twin Data

Jose (Pepe) Romeo

University of Santiago, Chile

jose.romeo@usach.cl

Joint work with:

Renate Meyer - University of Auckland, New Zealand
Diego Gallardo - University of Sao Paulo, Brazil

Workshop on Flexible Models for Longitudinal and Survival Data with Applications in Biostatistics

Coventry – UK, July 2015

Outline

- Introduction, some examples of multivariate lifetimes
- Modeling strategies
- Copulas
- The PVF copula model
  - Definition, particular and limiting cases and dependence properties
  - Simulating
  - Estimation
- Simulation study
- Application
- Final comments and future work

Introduction

- Survival analysis deals with time-to-event data, e.g. time to death or onset of disease.
- In multivariate survival analysis there may be a natural association because individuals share biological and/or environmental conditions.

  Examples:

Introduction

- Survival analysis deals with time-to-event data, e.g. time to death or onset of disease.
- In multivariate survival analysis there may be a natural association because individuals share biological and/or environmental conditions.

  Examples:
  - Groups of individuals under to similar environmental conditions
  - Time to relapse of a disease and time of death, for a subject
  - Lifetimes of pairs of human organs (e.g. kidneys, eyes)
  - Recurrent events: asthma attacks for a subject

Introduction

- Survival analysis deals with time-to-event data, e.g. time to death or onset of disease.
- In multivariate survival analysis there may be a natural association because individuals share biological and/or environmental conditions.

  Examples:
  - Groups of individuals under to similar environmental conditions
  - Time to relapse of a disease and time of death, for a subject
  - Lifetimes of pairs of human organs (e.g. kidneys, eyes)
  - Recurrent events: asthma attacks for a subject
  - Clustered failure times such as failure times of twins: to investigate whether the strength of dependence within twin pairs as to the risk of the onset of various diseases is different for MZ and DZ twins (time to appendectomy)

Introduction

- Survival analysis deals with time-to-event data, e.g. time to death or onset of disease.
- In multivariate survival analysis there may be a natural association because individuals share biological and/or environmental conditions.

  Examples:
  - Groups of individuals under to similar environmental conditions
  - Time to relapse of a disease and time of death, for a subject
  - Lifetimes of pairs of human organs (e.g. kidneys, eyes)
  - Recurrent events: asthma attacks for a subject
  - Clustered failure times such as failure times of twins: to investigate whether the strength of dependence within twin pairs as to the risk of the onset of various diseases is different for MZ and DZ twins (time to appendectomy)

- ➪ The assumption of independence among lifetimes can be unrealistic.
- ➪ It is of interest to estimate and quantify the dependence among the lifetimes and the effects of covariates under the dependence structure.

Modeling strategies

- Shared Frailty Models (random effect models)
  Clayton (1978), Hougaard (2000), Therneau & Grambsch (2000)

  Let T_{ij} denote the failure time i-th subject in the j-th cluster, the conditional hazard function is given by

  $$h(t|w_j, x_i) = h_0(t)\exp(\beta' x_i)w_j, \quad \text{where e.g., } W_j \overset{iid}{\sim} \text{Gamma, Log Normal, } \ldots$$

  Conditional on the frailty term w, the lifetimes are assumed independent, $S(t_1, \ldots, t_p|w) = S(t_1|w) \cdots S(t_p|w)$

Modeling strategies

- Shared Frailty Models (random effect models)
  Clayton (1978), Hougaard (2000), Therneau & Grambsch (2000)

  Let T_{ij} denote the failure time i-th subject in the j-th cluster, the conditional hazard function is given by

  $$h(t|w_j, x_i) = h_0(t)\exp(\beta' x_i)w_j, \quad \text{where e.g., } W_j \overset{iid}{\sim} \text{Gamma, Log Normal, } \ldots$$

  Conditional on the frailty term w, the lifetimes are assumed independent, $S(t_1, \ldots, t_p|w) = S(t_1|w) \cdots S(t_p|w)$

- Copulas Models (marginal models)
  Shih & Louis (1995), Duchateau & Janssen (2008), Wienke (2010)

  $$F(t_1, \ldots, t_p) = C_\alpha\left(F_1(t_1), \ldots, F_p(t_p)\right)$$

Copulas

Schweizer & Sklar (1983), Joe (1997), Owzar & Sen (2003), Nelsen (2006)

Let $T_1, \ldots, T_p$ r.v.'s with $T_j \sim F_j$, $j = 1, \ldots, p$, and joint cdf

$$F(t_1, \ldots, t_p) = \Pr\{T_1 \leq t_1, \ldots, T_p \leq t_p\}$$

Copulas

Schweizer & Sklar (1983), Joe (1997), Owzar & Sen (2003), Nelsen (2006)

Let $T_1, \ldots, T_p$ r.v.'s with $T_j \sim F_j$, $j = 1, \ldots, p$, and joint cdf

$$F(t_1, \ldots, t_p) = \Pr\{T_1 \leq t_1, \ldots, T_p \leq t_p\}$$

Therefore, considering $U_j = F_j(T_j) \sim \text{Uniform}(0, 1)$,

$$\begin{aligned} F(t_1, \ldots, t_p) &= \Pr\{U_1 \leq F_1(t_1), \ldots, U_p \leq F_p(t_p)\} \\ &= C_\alpha(F_1(t_1), \ldots, F_p(t_p)), \end{aligned} \tag{1}$$

is called the Copula of the vector $(T_1, \ldots, T_p)$, it is a multivariate cdf defined on $[0, 1]^p$ with uniformly distributed marginals and $\alpha \in \mathcal{A}$ is a dependence parameter.

Copulas

Schweizer & Sklar (1983), Joe (1997), Owzar & Sen (2003), Nelsen (2006)

Let $T_1, \ldots, T_p$ r.v.'s with $T_j \sim F_j$, $j = 1, \ldots, p$, and joint cdf

$$F(t_1, \ldots, t_p) = \Pr\{T_1 \leq t_1, \ldots, T_p \leq t_p\}$$

Therefore, considering $U_j = F_j(T_j) \sim \text{Uniform}(0, 1)$,

$$\begin{aligned} F(t_1, \ldots, t_p) &= \Pr\{U_1 \leq F_1(t_1), \ldots, U_p \leq F_p(t_p)\} \\ &= C_\alpha(F_1(t_1), \ldots, F_p(t_p)), \end{aligned} \tag{1}$$

is called the Copula of the vector $(T_1, \ldots, T_p)$, it is a multivariate cdf defined on $[0, 1]^p$ with uniformly distributed marginals and $\alpha \in \mathcal{A}$ is a dependence parameter.

From (1) the joint pdf is given by

$$f(t_1, \ldots, t_p) = c_\alpha(F_1(t_1), \ldots, F_p(t_p)) \prod_{j=1}^{p} f_j(t_j)$$

Archimedean Copulas

Genest & MacKay (1986), Joe (1997), Nelsen (2006).

A copula C_α is called **Archimedean** if there exists a convex function $\varphi_\alpha^{-1} : [0, \infty] \mapsto [0, 1]$, so that the copula C_α can be written as

$$C_\alpha(u_1, u_2) = \varphi_\alpha^{-1}(\varphi_\alpha(u_1) + \varphi_\alpha(u_2)),$$

for all $(u_1, u_2) \in [0, 1]^2$ and $\alpha \in \mathcal{A}$.

Archimedean Copulas

Genest & MacKay (1986), Joe (1997), Nelsen (2006).

A copula C_α is called **Archimedean** if there exists a convex function $\varphi_\alpha^{-1} : [0, \infty] \mapsto [0, 1]$, so that the copula C_α can be written as

$$C_\alpha(u_1, u_2) = \varphi_\alpha^{-1}(\varphi_\alpha(u_1) + \varphi_\alpha(u_2)),$$

for all $(u_1, u_2) \in [0, 1]^2$ and $\alpha \in \mathcal{A}$.

- when φ^{-1} corresponds to the Laplace transform of W in a **multiplicative frailty model**

  $$h(t|w) = h_0(t)w,$$

  copula models induced by frailty models are Archimedean copulas (Oakes, 1989), e.g:

Archimedean Copulas

Genest & MacKay (1986), Joe (1997), Nelsen (2006).

A copula C_α is called **Archimedean** if there exists a convex function $\varphi_\alpha^{-1} : [0, \infty] \mapsto [0, 1]$, so that the copula C_α can be written as

$$C_\alpha(u_1, u_2) = \varphi_\alpha^{-1}(\varphi_\alpha(u_1) + \varphi_\alpha(u_2)),$$

for all $(u_1, u_2) \in [0, 1]^2$ and $\alpha \in \mathcal{A}$.

- when φ^{-1} corresponds to the Laplace transform of W in a **multiplicative frailty model**

  $$h(t|w) = h_0(t)w,$$

  copula models induced by frailty models are Archimedean copulas (Oakes, 1989), e.g:

  - $W \sim$ Gamma then $(U_1, U_2) \sim$ Clayton copula
  - $W \sim$ Positive Stable then $(U_1, U_2) \sim$ Gumbel copula
  - $W \sim$ Inverse Gaussian then $(U_1, U_2) \sim$ Inverse Gaussian copula

Kendall's tau

- Let (T_1, T_2) a continuous r.v. with copula C_α, then the Kendall's tau coefficient for (T_1, T_2) is given by

$$\tau_\alpha = 4 \iint_{[0,1]^2} C_\alpha(u_1, u_2) dC_\alpha(u_1, u_2) - 1$$

Kendall's tau

- Let (T_1, T_2) a continuous r.v. with copula C_α, then the Kendall's tau coefficient for (T_1, T_2) is given by

$$\tau_\alpha = 4 \iint_{[0,1]^2} C_\alpha(u_1, u_2) dC_\alpha(u_1, u_2) - 1$$

- For Archimedean copulas, Kendall's tau is written as

$$\tau_\alpha = 4 \int_0^1 \frac{\varphi_\alpha(t)}{\varphi'_\alpha(t)} dt + 1$$

The PVF (Power variance function) distribution

$W \sim \text{PVF}(\alpha, \delta, \theta)$ for $\alpha \in (0, 1)$, $\delta > 0$, and $\theta \geq 0$ if for $w > 0$ the pdf is given by

The PVF (Power variance function) distribution

$W \sim \text{PVF}(\alpha, \delta, \theta)$ for $\alpha \in (0,1)$, $\delta > 0$, and $\theta \geq 0$ if for $w > 0$ the pdf is given by

$$f(w|\alpha,\delta,\theta) = -\frac{1}{\pi w}\exp\left(-\theta w + \frac{\delta\theta^{\alpha}}{\alpha}\right)\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha+1)}{\Gamma(k+1)}\left(\frac{-w^{-\alpha}\delta}{\alpha}\right)^{k}\sin(\alpha k\pi)$$

If $\theta > 0$ all moments exist, $\text{E}[W] = \delta\theta^{\alpha-1}$ and $\text{Var}[W] = (1-\alpha)\delta\theta^{\alpha-2}$.

The PVF (Power variance function) distribution

$W \sim \text{PVF}(\alpha, \delta, \theta)$ for $\alpha \in (0,1)$, $\delta > 0$, and $\theta \geq 0$ if for $w > 0$ the pdf is given by

$$f(w|\alpha, \delta, \theta) = -\frac{1}{\pi w} \exp\left(-\theta w + \frac{\delta \theta^{\alpha}}{\alpha}\right) \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{\Gamma(k+1)} \left(\frac{-w^{-\alpha}\delta}{\alpha}\right)^{k} \sin(\alpha k \pi)$$

If $\theta > 0$ all moments exist, $\text{E}[W] = \delta\theta^{\alpha-1}$ and $\text{Var}[W] = (1-\alpha)\delta\theta^{\alpha-2}$.

Under the reparametrization $\delta = \eta^{1-\alpha}$ and $\theta = \eta$, $W \sim \text{PVF}(\alpha, \eta)$, with $\alpha \in (0,1)$ and $\eta \geq 0$ and the Laplace transform is given by

$$\text{L}_W(s) = \exp\left\{-\frac{1}{\alpha}\left[\eta^{1-\alpha}(\eta + s)^{\alpha} - \eta\right]\right\}$$

Note that $\text{E}[W] = 1$, $\text{Var}[W] = (1-\alpha)\eta^{-1}$

The PVF copula

The Laplace transform of the PVF distribution

$$\mathsf{L}_W(s) = \exp\left\{-\frac{1}{\alpha}\left[\eta^{1-\alpha}(\eta+s)^{\alpha}-\eta\right]\right\}$$

The PVF copula

The Laplace transform of the PVF distribution

$$L_W(s) = \exp\left\{-\frac{1}{\alpha}\left[\eta^{1-\alpha}(\eta+s)^{\alpha}-\eta\right]\right\} = \varphi^{-1}_{\alpha,\eta}(s)$$

Following Oakes (1989), the Archimedean copula induced by PVF frailty distribution is

$$\begin{aligned} C_{\alpha,\eta}(u_1,u_2) &= \varphi^{-1}_{\alpha,\eta}\left(\varphi_{\alpha,\eta}(u_1)+\varphi_{\alpha,\eta}(u_2)\right) \\ &= \exp\left\{-\frac{1}{\alpha}\left[\eta^{1-\alpha}\left(g(u_1)^{\frac{1}{\alpha}}+g(u_2)^{\frac{1}{\alpha}}-\eta\right)^{\alpha}-\eta\right]\right\} \end{aligned}$$

where $g(u) = \eta^{\alpha} - \alpha\eta^{\alpha-1}\log u$.

The PVF copula

The Laplace transform of the PVF distribution

$$\mathsf{L}_W(s) = \exp\left\{-\frac{1}{\alpha}\left[\eta^{1-\alpha}(\eta+s)^{\alpha}-\eta\right]\right\} = \varphi^{-1}_{\alpha,\eta}(s)$$

Following Oakes (1989), the Archimedean copula induced by PVF frailty distribution is

$$\begin{aligned} C_{\alpha,\eta}(u_1,u_2) &= \varphi^{-1}_{\alpha,\eta}\left(\varphi_{\alpha,\eta}(u_1)+\varphi_{\alpha,\eta}(u_2)\right) \\ &= \exp\left\{-\frac{1}{\alpha}\left[\eta^{1-\alpha}\left(g(u_1)^{\frac{1}{\alpha}}+g(u_2)^{\frac{1}{\alpha}}-\eta\right)^{\alpha}-\eta\right]\right\} \end{aligned}$$

where $g(u) = \eta^{\alpha} - \alpha\eta^{\alpha-1}\log u$.

Note that

- if $\alpha \to 0$, $(U_1, U_2) \sim$ Clayton copula(η)
- if $\eta = 0$, $(U_1, U_2) \sim$ Gumbel copula(α)
- if $\alpha = 0.5$, $(U_1, U_2) \sim$ Inverse Gaussian copula(η)

The PVF copula - Dependence properties

Kendall's tau coefficient:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt$$

The PVF copula - Dependence properties

Kendall's tau coefficient:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \in (0, 1)$$

The PVF copula - Dependence properties

Kendall's tau coefficient:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \ \in \ (0,1)$$

- if $\alpha \to 1$, $\forall\ \eta \in R^+$ then $\tau \to 0$
- if $\eta \to \infty^+$, $\forall\ \alpha \in (0,1)$ then $\tau \to 0$

The PVF copula - Dependence properties

Kendall's tau coefficient:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \ \in \ (0,1)$$

- if $\alpha \to 1$, $\forall\ \eta \in R^+$ then $\tau \to 0$
- if $\eta \to \infty^+$, $\forall\ \alpha \in (0,1)$ then $\tau \to 0$
- if $\alpha \to 0$ and $\eta \to 0$ then $\tau \to 1$

The PVF copula - Dependence properties

Kendall's tau coefficient:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \ \in \ (0,1)$$

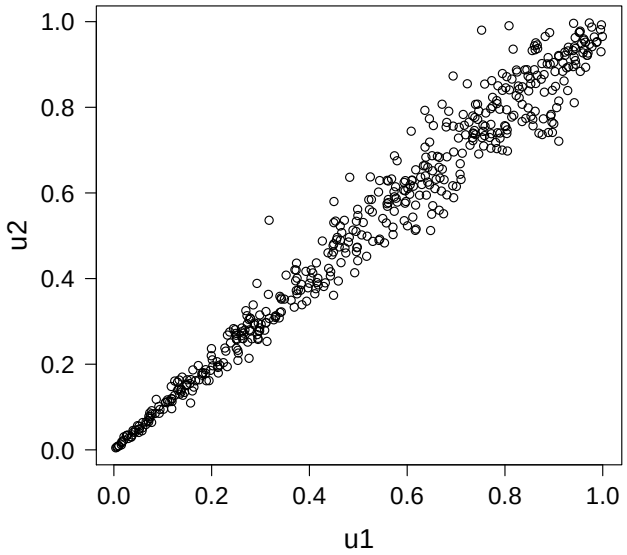
- if $\alpha \to 1$, $\forall\ \eta \in R^+$ then $\tau \to 0$
- if $\eta \to \infty^+$, $\forall\ \alpha \in (0,1)$ then $\tau \to 0$
- if $\alpha \to 0$ and $\eta \to 0$ then $\tau \to 1$

Copula	LTD_C	UTD_C	$\chi(v)$
PVF	0	0	$1-(1-\alpha)(\alpha \log v - \eta)^{-1}$
Clayton	$2^{-1/\eta}$	0	$\eta^{-1}+1$
Gumbel	0	$2-2^{1/\alpha}$	$1-(1-\alpha)(\alpha \log v)^{-1}$
Inverse Gaussian	0	0	$1-(\log v - 2\eta)^{-1}$

Table: Lower- and Upper-Tail Dependence and Cross-ratio function $\chi(v)$

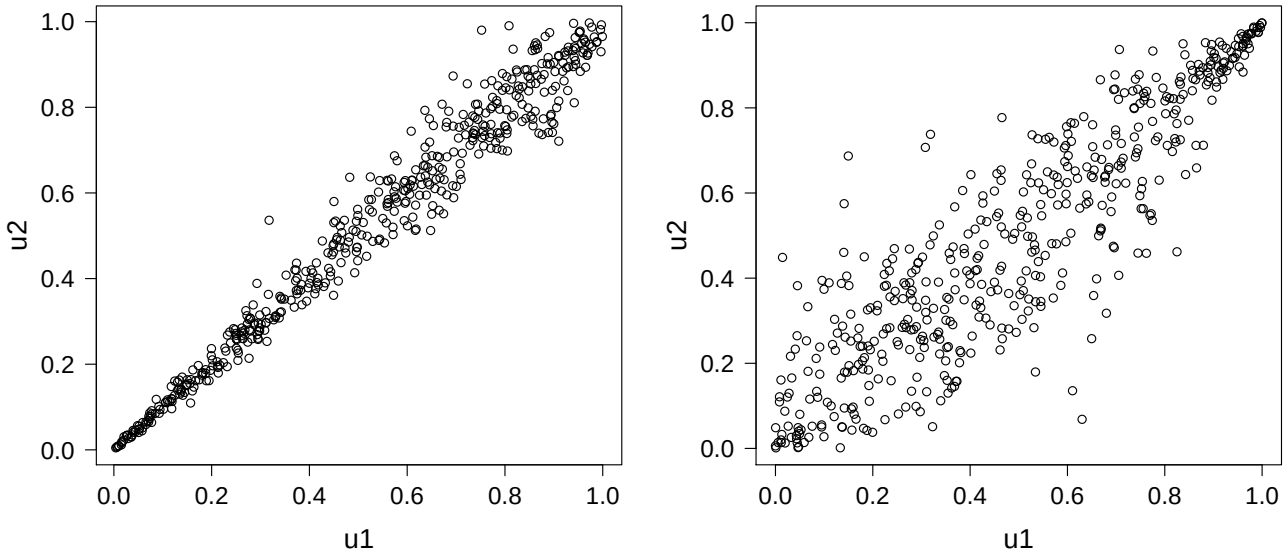
The PVF copula - Simulated data

Scatterplots of 500 samples from PVF copula $\tau = \{0.9, 0.7, 0.5, 0.2\}$



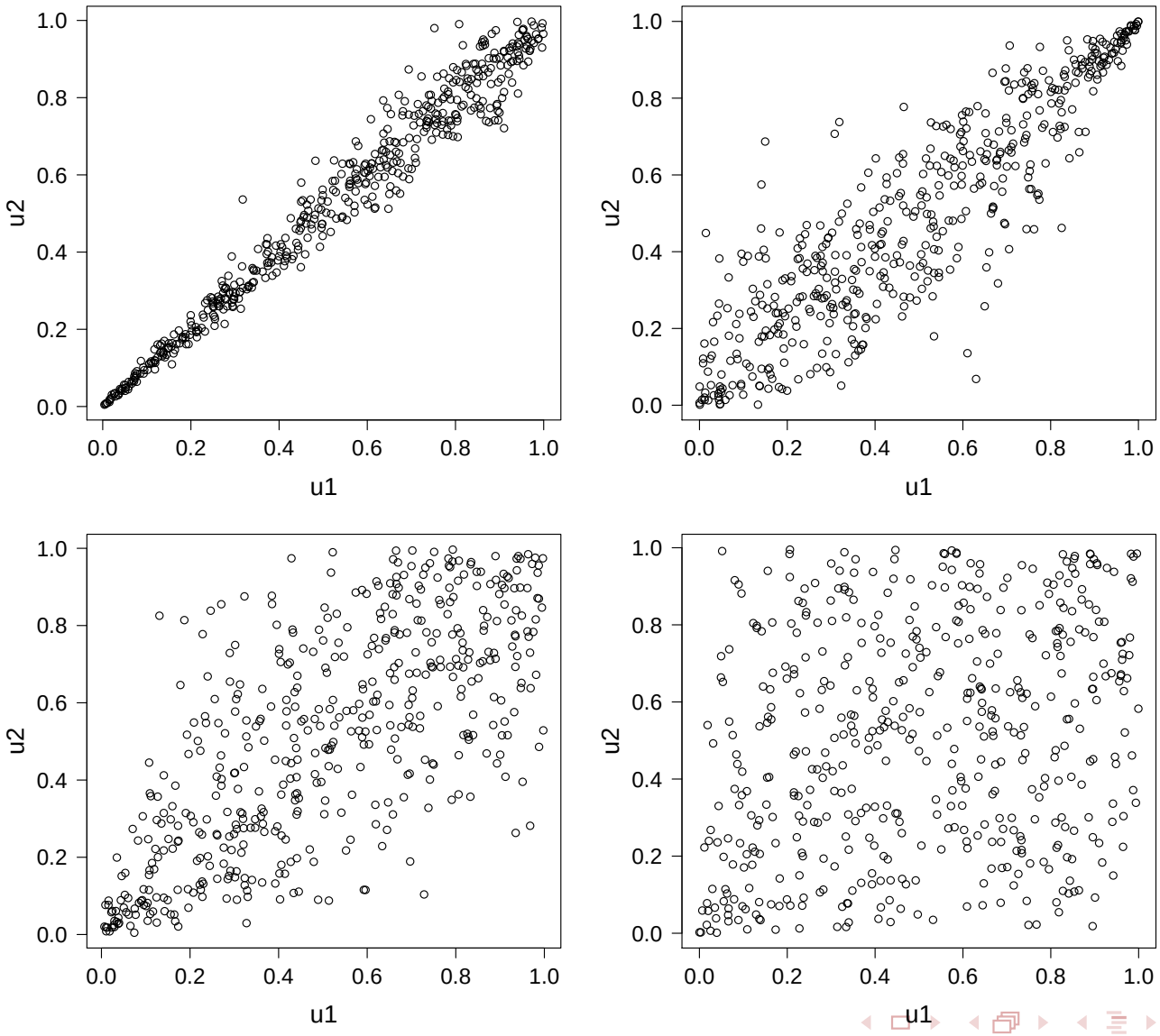
The PVF copula - Simulated data

Scatterplots of 500 samples from PVF copula $\tau = \{0.9, 0.7, 0.5, 0.2\}$



The PVF copula - Simulated data

Scatterplots of 500 samples from PVF copula $\tau = \{0.9, 0.7, 0.5, 0.2\}$



Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$

Algorithm: The conditional sampling for copulas, see e.g. Nelsen (2006):

Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$

Algorithm: The conditional sampling for copulas, see e.g. Nelsen (2006):

(*i*) simulate $u_2 \sim \text{Uniform}(0,1)$ and $q \sim \text{Uniform}(0,1)$

(*ii*) compute $F_{U_1|U_2}(u_1|u_2) := \frac{\partial}{\partial u_2} C(u_1, u_2)\Big|_{u_2}$ and set $u_1 = F^{-1}_{U_1|U_2}(q|u_2)$

(*iii*) return (u_1, u_2)

Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$

Algorithm: The conditional sampling for copulas, see e.g. Nelsen (2006):

(*i*) simulate $u_2 \sim \text{Uniform}(0,1)$ and $q \sim \text{Uniform}(0,1)$

(*ii*) compute $F_{U_1|U_2}(u_1|u_2) := \frac{\partial}{\partial u_2} C(u_1, u_2)\Big|_{u_2}$ and set $u_1 = F^{-1}_{U_1|U_2}(q|u_2)$

(*iii*) return (u_1, u_2)

$F^{-1}_{U_1|U_2}(q|u_2) = u_1$ cannot be solve explicitly in a close form, we solve it numerically by finding the root of $F_{U_1|U_2}(u_1|u_2) - q = 0$, i.e. $\frac{\partial}{\partial u_2} C(u_1, u_2) - q = 0$,

Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$

Algorithm: The conditional sampling for copulas, see e.g. Nelsen (2006):

(*i*) simulate $u_2 \sim \text{Uniform}(0,1)$ and $q \sim \text{Uniform}(0,1)$

(*ii*) compute $F_{U_1|U_2}(u_1|u_2) := \frac{\partial}{\partial u_2} C(u_1, u_2)\Big|_{u_2}$ and set $u_1 = F^{-1}_{U_1|U_2}(q|u_2)$

(*iii*) return (u_1, u_2)

$F^{-1}_{U_1|U_2}(q|u_2) = u_1$ cannot be solve explicitly in a close form, we solve it numerically by finding the root of $F_{U_1|U_2}(u_1|u_2) - q = 0$, i.e. $\frac{\partial}{\partial u_2} C(u_1, u_2) - q = 0$,

(*ii*) compute $F_{U_1|U_2}(u_1|u_2) = \frac{\partial}{\partial u_2} C(u_1, u_2)\Big|_{u_2}$ and set $u_1 = F^{-1}_{U_1|U_2}(q|u_2)$

(*ii.a*) set an initial value $u_1^{(0)} = u_0 \in (0,1)$

(*ii.b*) $u_1^{(j+1)} = u_1^{(j)} - (\frac{\partial}{\partial u_2} C(u_1^{(j)}, u_2) - q)$

(*ii.c*) if $|u_1^{(j+1)} - u_1^{(j)}| < \epsilon$ then return $u_1 = u_1^{(j+1)}$

Joint Bayesian estimation procedure

(Romeo et al., 2006) Let $(T_1, T_2) \sim (S_{1\theta_1}, S_{2\theta_2}), (f_{1\theta_1}, f_{2\theta_2})$. For $i = 1, \ldots, n$ suppose that (T_{i1}, T_{i2}) and the censoring times (C_{i1}, C_{i2}) are independent. The observed quantities are $Z_{ij} = \min\{T_{ij}, C_{ij}\}$ and $\delta_{ij} = I[Z_{ij} = T_{ij}]$, $j = 1, 2$.

The likelihood function for $(\alpha, \eta, \theta_1, \theta_2)$ is given by

$$
\begin{aligned}
L(\alpha, \eta, \theta_1, \theta_2 \mid \mathbf{z}_1, \boldsymbol{\delta}_1, \mathbf{z}_2, \boldsymbol{\delta}_2) = \prod_{i=1}^{n} & \left(c_{\alpha,\eta}(S_{1\theta_1}(z_{i1}), S_{2\theta_2}(z_{i2}))f_{1\theta_1}(z_{i1})f_{2\theta_2}(z_{i2})\right)^{\delta_{i1}\delta_{i2}} \\
& \cdot \left(\frac{\partial C_{\alpha,\eta}(S_{1\theta_1}(z_{i1}), S_{2\theta_2}(z_{i2}))}{\partial S_{1\theta_1}(z_{i1})} \cdot (-f_{1\theta_1}(z_{i1}))\right)^{\delta_{i1}(1-\delta_{i2})} \\
& \cdot \left(\frac{\partial C_{\alpha,\eta}(S_{1\theta_1}(z_{i1}), S_{2\theta_2}(z_{i2}))}{\partial S_{2\theta_2}(z_{i2})} \cdot (-f_{2\theta_2}(z_{i2}))\right)^{(1-\delta_{i1})\delta_{i2}} \\
& \cdot C_{\alpha,\eta}(S_{1\theta_1}(z_{i1}), S_{2\theta_2}(z_{i2}))^{(1-\delta_{i1})(1-\delta_{i2})}
\end{aligned}
$$

The posterior distribution can be written as

$$\pi(\alpha, \eta, \theta_1, \theta_2 \mid \mathbf{z}_1, \boldsymbol{\delta}_1, \mathbf{z}_2, \boldsymbol{\delta}_2) \propto L(\alpha, \eta, \theta_1, \theta_2 \mid \mathbf{z}_1, \boldsymbol{\delta}_1, \mathbf{z}_2, \boldsymbol{\delta}_2)\pi(\alpha)\pi(\eta)\pi(\theta_1)\pi(\theta_2)$$

Joint Bayesian estimation procedure

(Romeo et al., 2006) Let $(T_1, T_2) \sim (S_{1\theta_1}, S_{2\theta_2}), (f_{1\theta_1}, f_{2\theta_2})$. For $i = 1, \ldots, n$ suppose that (T_{i1}, T_{i2}) and the censoring times (C_{i1}, C_{i2}) are independent. The observed quantities are $Z_{ij} = \min\{T_{ij}, C_{ij}\}$ and $\delta_{ij} = I[Z_{ij} = T_{ij}]$, $j = 1, 2$.

The likelihood function for $(\alpha, \eta, \theta_1, \theta_2)$ is given by

$$
\begin{aligned}
L(\alpha, \eta, \theta_1, \theta_2 \mid \mathbf{z}_1, \boldsymbol{\delta}_1, \mathbf{z}_2, \boldsymbol{\delta}_2) = \prod_{i=1}^{n} & \left(c_{\alpha,\eta}(S_{1\theta_1}(z_{i1}), S_{2\theta_2}(z_{i2}))f_{1\theta_1}(z_{i1})f_{2\theta_2}(z_{i2})\right)^{\delta_{i1}\delta_{i2}} \\
& \cdot \left(\frac{\partial C_{\alpha,\eta}(S_{1\theta_1}(z_{i1}), S_{2\theta_2}(z_{i2}))}{\partial S_{1\theta_1}(z_{i1})} \cdot (-f_{1\theta_1}(z_{i1}))\right)^{\delta_{i1}(1-\delta_{i2})} \\
& \cdot \left(\frac{\partial C_{\alpha,\eta}(S_{1\theta_1}(z_{i1}), S_{2\theta_2}(z_{i2}))}{\partial S_{2\theta_2}(z_{i2})} \cdot (-f_{2\theta_2}(z_{i2}))\right)^{(1-\delta_{i1})\delta_{i2}} \\
& \cdot C_{\alpha,\eta}(S_{1\theta_1}(z_{i1}), S_{2\theta_2}(z_{i2}))^{(1-\delta_{i1})(1-\delta_{i2})}
\end{aligned}
$$

The posterior distribution can be written as

$$\pi(\alpha, \eta, \theta_1, \theta_2 \mid \mathbf{z}_1, \boldsymbol{\delta}_1, \mathbf{z}_2, \boldsymbol{\delta}_2) \propto L(\alpha, \eta, \theta_1, \theta_2 \mid \mathbf{z}_1, \boldsymbol{\delta}_1, \mathbf{z}_2, \boldsymbol{\delta}_2)\pi(\alpha)\pi(\eta)\pi(\theta_1)\pi(\theta_2)$$

Posterior computations implemented in `SAS - Proc MCMC`

Simulation study I - Small sample properties

We simulate $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$ under the following scheme:

- Three sample size: $n = \{50, 100, 200\}$
- Three level of association: $\tau = \{0.33, 0.50, 0.70\}$
- Three censoring percentages: $pc = \{5\%, 20\%, 50\%\}$
- 500 simulations (replicates)

Simulation study I - Small sample properties

We simulate $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$ under the following scheme:

- Three sample size: $n = \{50, 100, 200\}$
- Three level of association: $\tau = \{0.33, 0.50, 0.70\}$
- Three censoring percentages: $pc = \{5\%, 20\%, 50\%\}$
- 500 simulations (replicates)

		$n = 50$			$n = 100$			$n = 200$		
	true	estimate	bias	MSE	estimate	bias	MSE	estimate	bias	MSE
α	0.06	0.404	0.344	0.1182	0.269	0.209	0.0437	0.183	0.123	0.0151
η	0.90	0.399	-0.501	0.2632	0.532	-0.368	0.1380	0.662	-0.238	0.0603
τ	0.33	**0.297**	-0.034	0.0057	**0.318**	-0.013	0.0022	**0.326**	-0.005	0.0010
α	0.36	0.403	0.039	0.0090	0.377	0.013	0.0060	0.368	0.004	0.0023
η	0.10	0.080	-0.020	0.0051	0.093	-0.007	0.0027	0.097	-0.003	0.0012
τ	0.50	0.463	-0.037	0.0037	0.483	-0.017	0.0016	0.491	-0.009	0.0006
α	0.28	0.282	0.002	0.0018	0.277	-0.003	0.0009	0.286	0.006	0.0005
η	0.01	0.010	0.000	0.0001	0.011	0.001	0.0000	0.010	0.000	0.0000
τ	0.70	0.686	-0.015	0.0013	0.696	-0.005	0.0006	0.693	-0.008	0.0003

Table: Mean of Bayesian posterior median, bias and MSE for parameters of the PVF copula model $(pc = 20\%)$

Simulation study II - Comparison of different copula models

- We simulate $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$ under the same previous setup
- We fit the PVF copula model and Clayton, Gumbel and Inverse Gaussian
- Goodness-of-fit is compared using standard model selection criteria

Simulation study II - Comparison of different copula models

- We simulate $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$ under the same previous setup
- We fit the PVF copula model and Clayton, Gumbel and Inverse Gaussian
- Goodness-of-fit is compared using standard model selection criteria

	Model	AIC	BIC	DIC	LPML	τ-mean	τ-sd
$n = 50$	PVF	0.102	0.026	**0.480**	**0.462**	0.489	0.084
	Clayton	0.394	0.414	0.208	0.190	0.453	0.099
	Gumbel	0.050	0.068	0.002	0.004	0.414	0.080
	InvGaussian	**0.454**	**0.492**	0.310	0.344	0.449	0.051
$n = 100$	PVF	0.298	0.096	**0.602**	**0.580**	0.491	0.057
	Clayton	0.204	0.266	0.108	0.102	0.456	0.067
	Gumbel	0.022	0.030	0.008	0.012	0.417	0.056
	InvGaussian	**0.476**	**0.608**	0.282	0.306	0.445	0.032
$n = 200$	PVF	**0.618**	0.288	**0.786**	**0.776**	0.495	0.039
	Clayton	0.042	0.090	0.020	0.018	0.454	0.047
	Gumbel	0.000	0.000	0.000	0.002	0.422	0.038
	InvGaussian	0.340	**0.622**	0.194	0.204	0.441	0.021

Table: Comparison of different copula models simulating from $\text{PVF}(0.36, 0.10)$ copula through the **proportion of times a certain model was chosen** ($\tau = 0.50$, $pc = 20\%$)

Application - Australian NH&MRC Twin data

➪ To investigate whether the strength of dependence within twin pairs as to the risk of the onset of various diseases is different for MZ and DZ twins.

Application - Australian NH&MRC Twin data

➪ To investigate whether the strength of dependence within twin pairs as to the risk of the onset of various diseases is different for MZ and DZ twins.

- A questionnaire was applied for collecting the information, see Duffy et al. (1990).

Application - Australian NH&MRC Twin data

➪ To investigate whether the strength of dependence within twin pairs as to the risk of the onset of various diseases is different for MZ and DZ twins.

- A questionnaire was applied for collecting the information, see Duffy et al. (1990).
- We analyze **time to appendectomy** for adult twins comprised of 1798 **MZ** (1231 female and 567 male pairs) and 1098 **DZ** twins (748 female and 350 male pairs)

Application - Australian NH&MRC Twin data

➪ To investigate whether the strength of dependence within twin pairs as to the risk of the onset of various diseases is different for MZ and DZ twins.

- A questionnaire was applied for collecting the information, see Duffy et al. (1990).
- We analyze **time to appendectomy** for adult twins comprised of 1798 **MZ** (1231 female and 567 male pairs) and 1098 **DZ** twins (748 female and 350 male pairs)
- Subjects not undergoing appendectomy prior to survey were considered censored failure times (approximately 73% for each member of the twins in both types of zygotes)

Application - Australian NH&MRC Twin data

➪ To investigate whether the strength of dependence within twin pairs as to the risk of the onset of various diseases is different for MZ and DZ twins.

- A questionnaire was applied for collecting the information, see Duffy et al. (1990).
- We analyze **time to appendectomy** for adult twins comprised of 1798 **MZ** (1231 female and 567 male pairs) and 1098 **DZ** twins (748 female and 350 male pairs)
- Subjects not undergoing appendectomy prior to survey were considered censored failure times (approximately 73% for each member of the twins in both types of zygotes)
- Since any potential effect of a **shared environment** could be similar for MZ and DZ twins, a stronger dependence in the risks for appendicitis between MZ twin pair members would be indicative of a **genetic effect** and evidence of **heredity** in the onset of appendicitis

Application - Australian NH&MRC Twin data

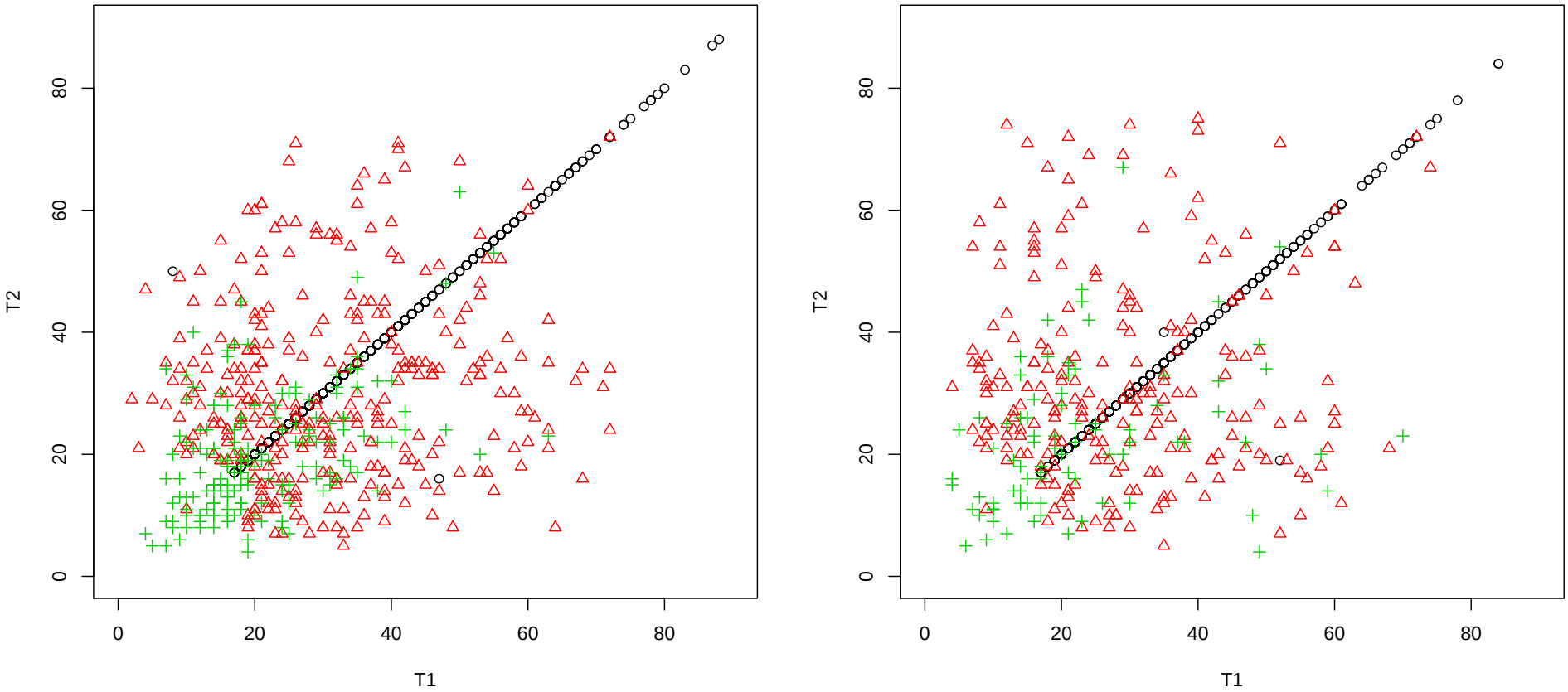


Figure: Scatterplot of MZ (left) and DZ (right) twin data. The data points correspond to non censored times (+), one censored time (△) and both times are censored (○)

Application - Australian NH&MRC Twin data

Modelling

- Marginal distributions: $T_j \sim$ Piecewise Exponential(λ_{lj}).

$$h_j(t) = \lambda_{lj}, \text{ for } t \in [a_l, a_{l+1}),\ l = 1, 2, \ldots, L,$$

and prior distribution $\lambda_{lj} \sim \text{Gamma}(0.001, 0.001),\ j = 1, 2.$

Application - Australian NH&MRC Twin data

Modelling

- Marginal distributions: $T_j \sim$ Piecewise Exponential(λ_{lj}).

$$h_j(t) = \lambda_{lj}, \text{ for } t \in [a_l, a_{l+1}),\ l = 1, 2, \ldots, L,$$

and prior distribution $\lambda_{lj} \sim \text{Gamma}(0.001, 0.001),\ j = 1, 2.$

- Priors for dependence parameters on copula models:
  - PVF $\alpha \sim \text{Beta}(1, 1)$, $\eta \sim \text{Gamma}(0.01, 0.01)$,
  - Clayton $\eta \sim \text{Gamma}(0.01, 0.01)$
  - Gumbel $\alpha \sim \text{Beta}(1, 1)$
  - Inverse Gaussian $\eta \sim \text{Gamma}(0.01, 0.01)$

Application - Australian NH&MRC Twin data

Modelling

- Marginal distributions: $T_j \sim$ Piecewise Exponential(λ_{lj}).

$$h_j(t) = \lambda_{lj}, \text{ for } t \in [a_l, a_{l+1}),\ l = 1, 2, \ldots, L,$$

  and prior distribution $\lambda_{lj} \sim \text{Gamma}(0.001, 0.001),\ j = 1, 2.$

- Priors for dependence parameters on copula models:
  - PVF $\alpha \sim \text{Beta}(1, 1)$, $\eta \sim \text{Gamma}(0.01, 0.01)$,
  - Clayton $\eta \sim \text{Gamma}(0.01, 0.01)$
  - Gumbel $\alpha \sim \text{Beta}(1, 1)$
  - Inverse Gaussian $\eta \sim \text{Gamma}(0.01, 0.01)$

- ... but, how we can include the gender and type of zygosity in the model?

Application - Australian NH&MRC Twin data

Modelling

- Marginal distributions: $T_j \sim$ Piecewise Exponential(λ_{lj}).

$$h_j(t) = \lambda_{lj}, \text{ for } t \in [a_l, a_{l+1}),\ l = 1, 2, \ldots, L,$$

and prior distribution $\lambda_{lj} \sim \text{Gamma}(0.001, 0.001),\ j = 1, 2.$

- Priors for dependence parameters on copula models:
  - PVF $\alpha \sim \text{Beta}(1, 1)$, $\eta \sim \text{Gamma}(0.01, 0.01)$,
  - Clayton $\eta \sim \text{Gamma}(0.01, 0.01)$
  - Gumbel $\alpha \sim \text{Beta}(1, 1)$
  - Inverse Gaussian $\eta \sim \text{Gamma}(0.01, 0.01)$

- ... but, how we can include the gender and type of zygosity in the model?

➪ Instead of the conventional approach of analysing MZ and DZ separately, we allow the **association parameter to depend on covariates**, i.e., including the type of zygosity as a dichotomous covariate as well as the sex of the twins.

Application - Australian NH&MRC Twin data

Modelling

- Specifically, we include binary covariates x_1, type of zygosity and x_2, sex of the twins, through the dependence parameter:
  - $logit(\alpha(x_1, x_2)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$, where $logit(u) = \log(u/(1-u))$ with $u \in (0,1)$ for α in PVF(α, η) and Gumbel(α)
  - $\log(\eta(x_1, x_2)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$, for Clayton$(\eta)$

Application - Australian NH&MRC Twin data

Modelling

- Specifically, we include binary covariates x_1, type of zygosity and x_2, sex of the twins, through the dependence parameter:
  - $logit(\alpha(x_1, x_2)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$, where $logit(u) = \log(u/(1-u))$ with $u \in (0,1)$ for α in PVF(α, η) and Gumbel(α)
  - $\log(\eta(x_1, x_2)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$, for Clayton$(\eta)$
- The prior distributions for γ_0, γ_1 and γ_2 are chosen as Normal$(0, 0.0001)$

Application - Australian NH&MRC Twin data

Modelling

- Specifically, we include binary covariates x_1, type of zygosity and x_2, sex of the twins, through the dependence parameter:
  - $logit(\alpha(x_1, x_2)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$, where $logit(u) = \log(u/(1-u))$ with $u \in (0,1)$ for α in PVF(α, η) and Gumbel(α)
  - $\log(\eta(x_1, x_2)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$, for Clayton$(\eta)$
- The prior distributions for γ_0, γ_1 and γ_2 are chosen as Normal$(0, 0.0001)$

Results

Model	AIC	BIC	DIC	p_D	LPML
PVF-PWE	15122.71	15254.07	15100.77	22.06	-7551.83
Clayton-PWE	15147.71	15273.10	15127.03	21.33	-7564.39
Gumbel-PWE	15122.61	15248.00	15101.74	21.13	-7551.68

Table: Model selection criteria for copula models, Australian twin data

Application - Australian NH&MRC Twin data

Modelling

- Specifically, we include binary covariates x_1, type of zygosity and x_2, sex of the twins, through the dependence parameter:
  - $logit(\alpha(x_1, x_2)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$, where $logit(u) = \log(u/(1-u))$ with $u \in (0,1)$ for α in PVF(α, η) and Gumbel(α)
  - $\log(\eta(x_1, x_2)) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2$, for Clayton$(\eta)$
- The prior distributions for γ_0, γ_1 and γ_2 are chosen as Normal$(0, 0.0001)$

Results

Model	AIC	BIC	DIC	p_D	LPML
PVF-PWE	15122.71	15254.07	15100.77	22.06	-7551.83
Clayton-PWE	15147.71	15273.10	15127.03	21.33	-7564.39
Gumbel-PWE	15122.61	15248.00	15101.74	21.13	-7551.68

Table: Model selection criteria for copula models, Australian twin data

➪ Note that for PVF, the posterior estimate of η is (0.007 ± 0.014)

Application - Australian NH&MRC Twin data

Results

Parameter	Mean	Median	SD	HPD
$\tau_{MZ-Female}$	0.229	0.230	0.024	0.183, 0.276
$\tau_{MZ-Male}$	0.143	0.143	0.031	0.079, 0.197
$\tau_{DZ-Female}$	0.141	0.141	0.025	0.094, 0.193
$\tau_{DZ-Male}$	0.085	0.083	0.026	0.040, 0.137

Table: Posterior Kendall's τ, Gumbel copula model, Australian twin data

Final comments and future work

- Models based on copulas are considerable flexible respect the marginal distributions and the dependence structure
- PVF copula and particular models
- Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$
- From simulation study: better performance of the PVF copula model for $n > 100$ and $\tau > 0.33$

Final comments and future work

- Models based on copulas are considerable flexible respect the marginal distributions and the dependence structure
- PVF copula and particular models
- Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$
- From simulation study: better performance of the PVF copula model for $n > 100$ and $\tau > 0.33$

- Application in others areas (finance)

Final comments and future work

- Models based on copulas are considerable flexible respect the marginal distributions and the dependence structure
- PVF copula and particular models
- Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$
- From simulation study: better performance of the PVF copula model for $n > 100$ and $\tau > 0.33$

- Application in others areas (finance)
- To implement the case of unbalanced lifetime data (Meyer & Romeo, 2015)

Final comments and future work

- Models based on copulas are considerable flexible respect the marginal distributions and the dependence structure
- PVF copula and particular models
- Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$
- From simulation study: better performance of the PVF copula model for $n > 100$ and $\tau > 0.33$

- Application in others areas (finance)
- To implement the case of unbalanced lifetime data (Meyer & Romeo, 2015)
- Testing precise or sharp hypothesis for PVF copula model:
  $H_0 : \alpha \to 0$ (Clayton),
  $H_0 : \eta = 0$ (Gumbel)
  $H_0 : \alpha = 0.5$ (Inverse Gaussian)

Final comments and future work

- Models based on copulas are considerable flexible respect the marginal distributions and the dependence structure
- PVF copula and particular models
- Simulation of $(u_1, u_2) \sim \text{PVF}(\alpha, \eta)$
- From simulation study: better performance of the PVF copula model for $n > 100$ and $\tau > 0.33$

- Application in others areas (finance)
- To implement the case of unbalanced lifetime data (Meyer & Romeo, 2015)
- Testing precise or sharp hypothesis for PVF copula model:
  $H_0 : \alpha \to 0$ (Clayton),
  $H_0 : \eta = 0$ (Gumbel)
  $H_0 : \alpha = 0.5$ (Inverse Gaussian)
- Temporal dependence: $\tau(t)$

Main References

- Duchateau, L. and Janssen, P. (2008). *The Frailty Model*, Springer, NY
- Duffy, D.L., Martin, N.G. and Mathews, J.D. (1990). Appendectomy in Australian twins. *American Journal of Human Genetics*, **47**(3), 590–592.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer, NY
- Mai, J. and Scherer, M. (2012). *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications.* Imperial College, Boca Raton.
- Meyer, R. and Romeo, J.S. (2015). Bayesian semi-parametric analysis of recurrent failure time data using copulas. *Biometrical Journal*, in press.
- Nelsen, R.B. (2006). *An Introduction to Copulas,* 2nd edition. Springer, NY.
- Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, **84**, 487-493.
- Romeo, J.S., Tanaka, N.I. and Pedroso de Lima, A.C. (2006). Bivariate survival modeling: A Bayesian approach based on Copulas. *Lifetime Data Analysis*, **12**, 205-222.
- Wienke, A. (2010). *Frailty Models in Survival Analysis.* Chapman and Hall/CRC, Boca Raton.