Flexible nonparametric estimation of the cross ratio function for bivariate right-censored data using penalized tensor product splines

Miguel Beynaerts, Steven Abrams, Niel Hens, Anneleen Verhasselt

VUB meeting – 15 September 2026

A very brief introduction to association measures

Consider a pair of possibly correlated random variables (event times) (T_1, T_2) with IID realizations (t_1, t_2). Often, dependence between two random variables is measured using global measures

- Pearson's correlation r
- Spearman's ρ
- Kendall's τ
- ...

These measures are not able to capture local dependence (ie. for a given $T_1 = t_1$, what is the strength of association between T_1 and T_2?)

Some examples:

- Disease incidence with familial histories (Clayton, 1978)
- Time to appendectomy in Australian twins (See, e.g., Hu et al., 2011; Hu et al., 2019)

Introduction to the cross ratio function

The cross ratio function (CRF) is a measure of **time-dependent (or local) association**

$$\theta(t_1, t_2) = \frac{\lambda_1(t_1 \mid T_2 = t_2)}{\lambda_1(t_1 \mid T_2 > t_2)} = \frac{\lambda_2(t_2 \mid T_1 = t_1)}{\lambda_2(t_2 \mid T_1 > t_1)},$$

where t_1 and t_2 are realizations of T_1 and T_2, and λ_i the conditional hazard function of T_i given $T_{j \neq i}$ $(i, j = 1, 2)$. As a ratio between conditional hazards, it measures the risk between an exposure and non-exposure group.

Some properties of θ:

- Symmetric wrt switching the roles of T_1 and T_2.
- Invariant to transformations of the marginal distributions.
- Can be expressed in terms of the survival function and its derivatives.

What has been done already...

Various approaches have been proposed to estimate $\theta(t_1, t_2)$:

- Copula-based approaches: (empirical) copula determines $S(t_1, t_2)$
  - Parametric copula and margins
  - Smoothing of conditional hazard (Abrams et al., 2020)
  - Smoothing of (conditional) copula (Abrams et al., 2024; Sercik et al., 2026)
- Semiparametric pseudo-likelihood approach (Oakes, 1986)
- Parametric models for right-censored data
  - Piecewise exponential CRF estimation (Nan et al., 2006)
  - Polynomial surface (Hu et al., 2011)
  - Estimation of the CRF conditional on covariates (Hu et al., 2014)

Our focus: Extend the work by Hu et al. (2011) to flexible spline-based estimation of the CRF

Pseudo-partial likelihood for bivariate censored data

In applications with independent right censoring, we observe $X_j = \min(T_j, C_j)$ and a censoring indicator $\Delta_j = I(T_j \leq C_j)$ for $j = 1, 2$ where (C_1, C_2) are bivariate censoring times.

Hu et al. (2011) proposed a pseudo-partial likelihood by considering T_1 as a covariate in a Cox PH model. Then the ratio between $\{j \mid T_{1j} = t_1\}$ and $\{j \mid T_{1j} > t_1\}$ is the hazard ratio of T_2.

The CRF is symmetric with respect to switching the roles of T_1 and T_2, so a second likelihood contribution is needed to guarantee a symmetric estimator.

A tensor product spline approach

Let $\boldsymbol{\beta}$ be the parameter vector of interest. The implied pseudo-partial log-likelihood function

$$l(\boldsymbol{\beta}) = \sum_i \left\{ l_i^{(1)}(\boldsymbol{\beta}) + l_i^{(2)}(\boldsymbol{\beta}) \right\}$$

is maximized to obtain the Maximum Likelihood (ML) estimates $\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} l(\boldsymbol{\beta})$.

Approach of Hu et al. (2011)

$$\log \theta(t_1, t_2) = \sum_{r+s \leq 3} \beta_{rs} t_1^r t_2^s$$

Our approach:

$$\log \theta(t_1, t_2) = \sum_{r=1}^{r'} \sum_{s=1}^{s'} \beta_{rs} B_{1r}(t_1) B_{2s}(t_2)$$

where B_{1r} and B_{2s} are B-spline basis functions. Note that since the CRF is positive, a log transformation allows for unconstrained optimization.

Penalized pseudo-partial likelihood

Flexible estimation of the (log) CRF requires a large parameter space. Identifiability might not be guaranteed. We introduce penalization to the likelihood

$$l_{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - \frac{\lambda_1}{2}\boldsymbol{\beta}^T \boldsymbol{S}_1 \boldsymbol{\beta} - \frac{\lambda_2}{2}\boldsymbol{\beta}^T \boldsymbol{S}_2 \boldsymbol{\beta}$$

with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ a vector of smoothing parameters to be chosen in a data-driven way and $\boldsymbol{S}_i$ $(i = 1, 2)$ the corresponding positive semidefinite penalty matrices.

Possible penalties:

- Discrete (d-order difference) penalties: P-splines (Eilers & Marx, 1996)
- Integrated squared derivative penalties: O-splines (O'Sullivan, 1986)

Selecting smoothing parameters $\boldsymbol{\lambda}$

Bayesian motivation (Wahba, 1983):

$$\boldsymbol{\beta} \sim N\left(\mathbf{0}, \left(\sum_k \lambda_k \boldsymbol{S_k}\right)^{-1}\right)$$

This leads to the Laplace approximation of the (penalized) log-likelihood

$$\mathcal{V}(\boldsymbol{\lambda}) = l(\boldsymbol{\beta}) - \frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{S_\lambda} \boldsymbol{\beta} + \frac{1}{2}\log|S_{\boldsymbol{\lambda}}|^{+} - \frac{1}{2}\log|\mathcal{H}_{\boldsymbol{\lambda}}| + \frac{\mathcal{M}_p}{2}\log(2\pi),$$

where $\boldsymbol{S_\lambda} = \sum_i \lambda_i \boldsymbol{S}_i$, $|\cdot|^{+}$ the product of all positive eigenvalues, $\mathcal{H}_{\boldsymbol{\lambda}}$ the negative Hessian of the penalized log-likelihood and $\mathcal{M}_p$ the number of zero eigenvalues of $S_{\boldsymbol{\lambda}}$ when all λ_i are strictly positive.

The generalized Fellner-Schall method

Wood and Fasiolo (2017) proposed a simple iterative procedure to maximize the Laplace approximated (penalized pseudo-partial) log-likelihood function

$$\lambda_i^* = \frac{\mathrm{tr}\left(\boldsymbol{S}_{\boldsymbol{\lambda}}^{-}\boldsymbol{S}_i\right) - \mathrm{tr}\left(\mathcal{H}_{\boldsymbol{\lambda}}^{-1}\boldsymbol{S}_i\right)}{\hat{\beta}^T\boldsymbol{S}_{\boldsymbol{\lambda}}\hat{\beta}}\lambda_i.$$

Procedure:

1. Choose initial $\boldsymbol{\lambda}_{\mathbf{(0)}}$
2. Maximize penalized likelihood to obtain $\hat{\beta}|\boldsymbol{\lambda}_{\mathbf{(0)}}$
3. Update $\boldsymbol{\lambda}_{\mathbf{(0)}}$ to $\boldsymbol{\lambda}_{\mathbf{(1)}}$ using $\hat{\beta}|\boldsymbol{\lambda}_{\mathbf{(0)}}$
4. Re-estimate $\hat{\beta}|\boldsymbol{\lambda}_{\mathbf{(1)}}$
5. Check convergence criteria
6. Repeat steps 2–5 until convergence

Simulation study

We fit P- and B-splines with dimension 17×17 to samples of size 500, 750 and 1000 generated from the bivariate survival function

$$S(t_1, t_2) = \exp\left\{-at_1^b - at_2^b + h(at_1^b + at_2^b) - h(0)\right\}.$$

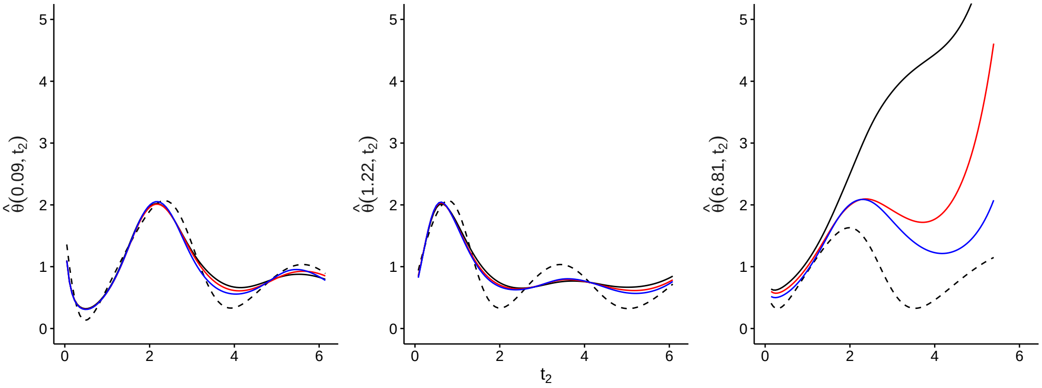
This implies the flexible CRF

$$\theta(t_1, t_2) = 1 + \frac{h''(at_1^b + at_2^b)}{\left(1 - h'(at_1^b + at_2^b)\right)^2}.$$

where $0 < a < 1$, $0 < b < 1$ and $h(\cdot)$ a suitable trigonometric transformation function to control the flexibility (of the survival function). In this example, we have $a = b = 0.7$ and

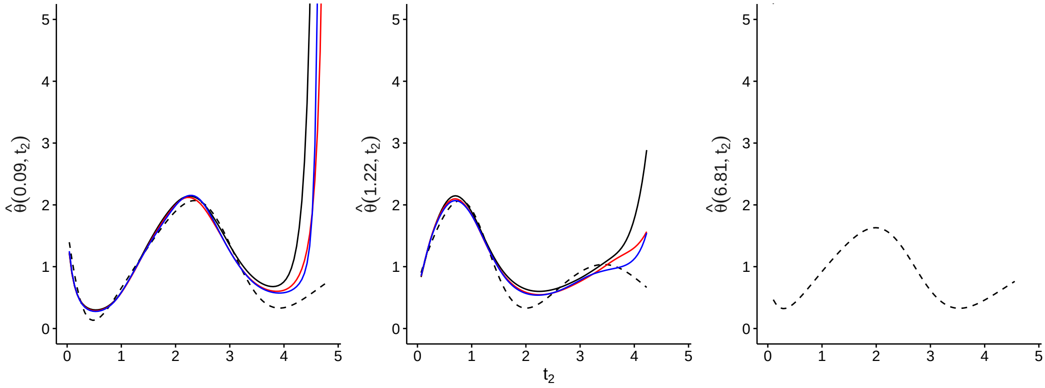
$$h(z) = \frac{1}{2}\{-0.35\cos(z) + 0.08\sin(3z) - 0.04\cos(5z)\}.$$

Selected results: fit (35% censoring)



P-spline fit with 35% censoring for $n = 500, 750, 1000$

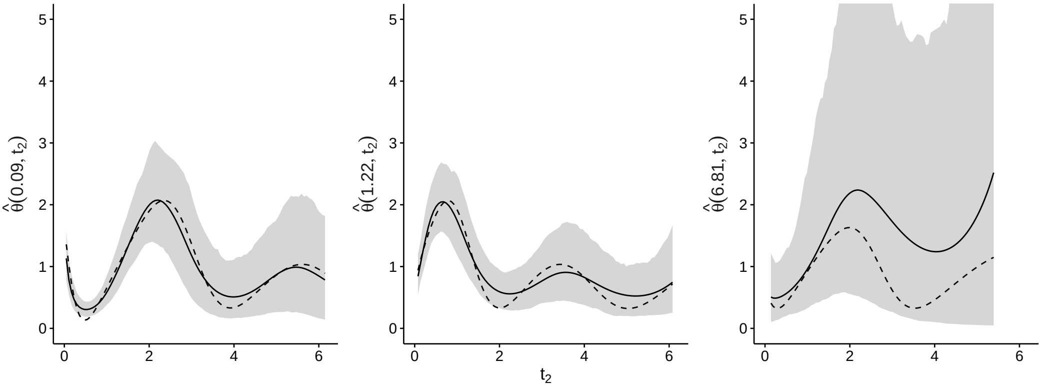
Selected results: fit (60% censoring)



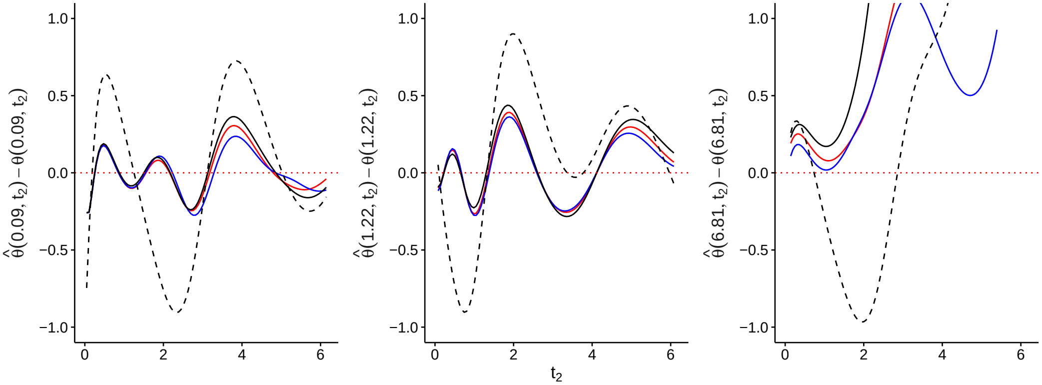
P-spline fit with 60% censoring for $n = 500, 750, 1000$

A note on knot selection...

In practice, one should take advantage of penalization by using many more knots!

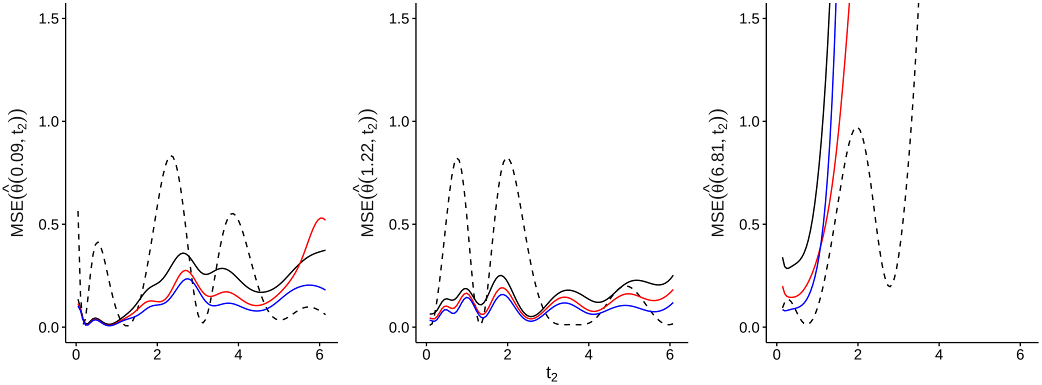


Selected results: bias (35% censoring)



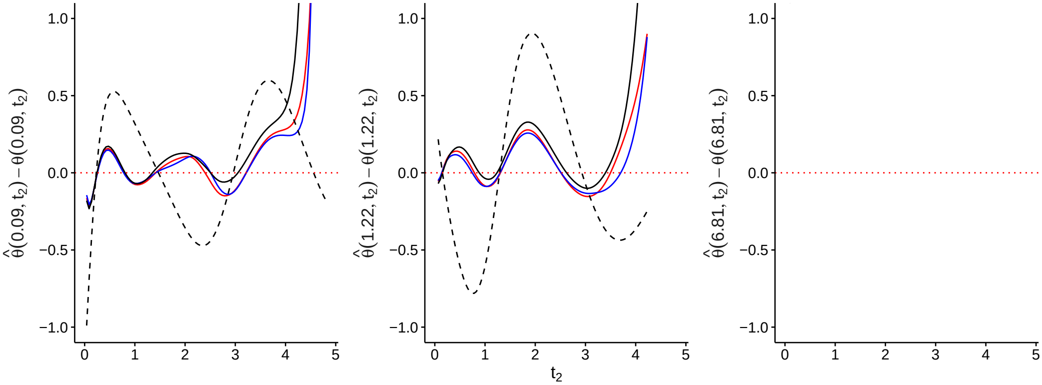
Bias of the P-spline model compared to parametric model from Hu et al. (2011) with 30% censoring for $n = 500, 750, 1000$

Selected results: MSE (35% censoring)



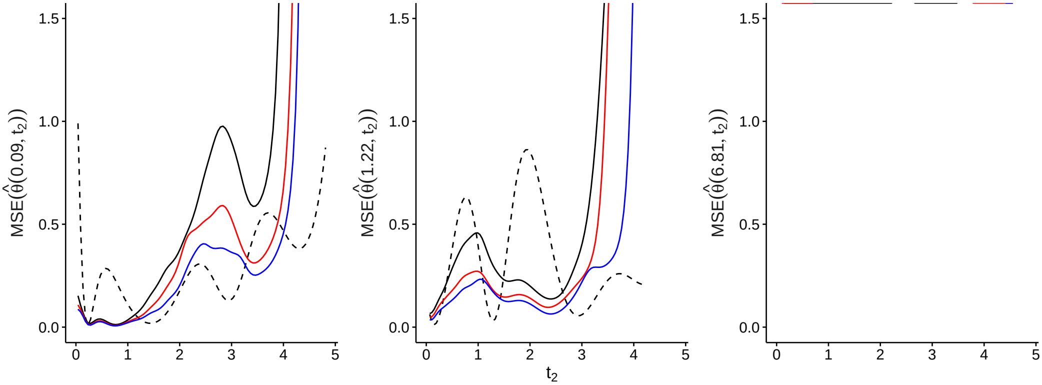
Pointwise MSE of the P-spline model compared to parametric model from Hu et al. (2011) with 35% censoring for $n = 500, 750, 1000$

Selected results: bias (60% censoring)



Bias of the P-spline model compared to parametric model from Hu et al. (2011) with 60% censoring for $n = 500, 750, 1000$

Selected results: MSE (60% censoring)



Pointwise MSE of the P-spline model compared to parametric model from Hu et al. (2011) with 60% censoring for $n = 500, 750, 1000$

Real-life data example: CCASAnet

Caribbean, Central and South America Network for HIV research (CCASAnet)

6691 HIV-positive patients in 5 different countries are treated with antiretroviral therapy (Eden et al., 2022)

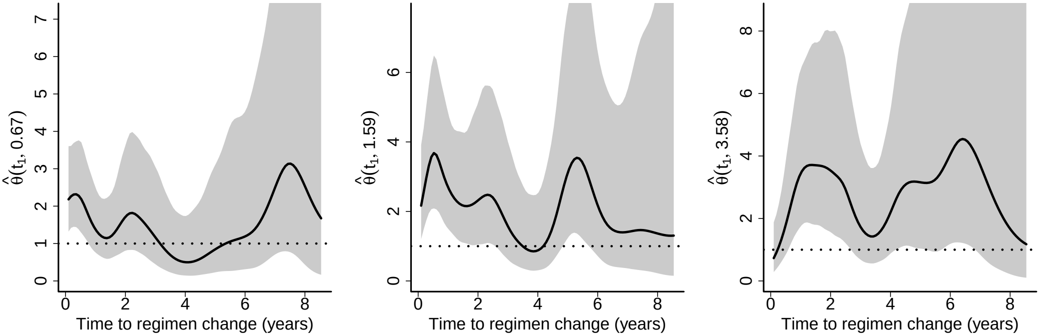
Two event times of interest (under bivariate right censoring):

- t_1: time to regimen change (years)
- t_2: time to viral failure (years)

Goal: We investigate the time-varying dependence between the two event times in countries A and B using our novel estimator with bivariate B-splines of dimension 17×17.

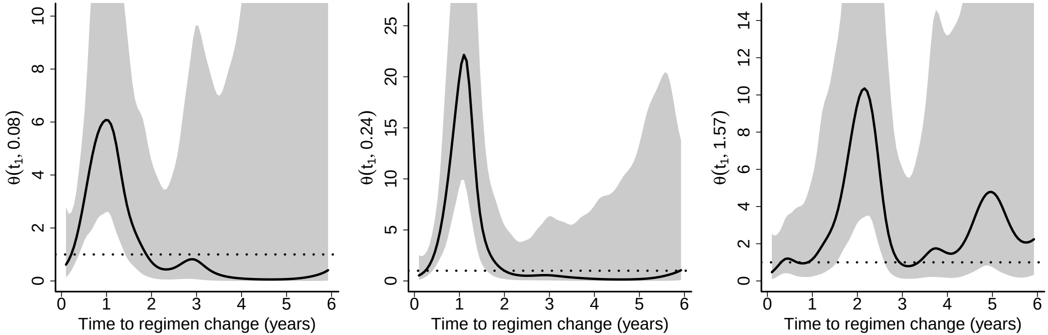
Real-life data example: CCASAnet

Country A



Real-life data example: CCASAnet

Country B



Future work...

- Theoretical framework: rate of convergence and asymptotic normality.
- Goodness-of-fit.
- Model extensions.
  - including covariates.
  - (semi-)competing risks with applications in surrogate marker identification.
- Taking advantage of censoring for computational gains.
- Alternatives ways to approximate infinite parameter space (e.g. sieves)?

Acknowledgements

SA, AV: FWO grant G011022N

MB, SA, AV, NH: BOF funding

MB, NH: Special Research Fund through the Methusalem project BOF22M01

We gratefuly acknowledge the VSC (Flemish Supercomputer center)