ERGO Center of Excellence in Insurance

Eine Einrichtung der TUM gefördert von der ERGO Group

A Dirichlet Process Mixture model for the analysis of competing risks

Dr. Francesco Ungolo

Technische Universität München

Insurance Data Science Conference

18th June 2021

Agenda

Introduction

- **Aim**: estimate the joint distribution of the time to M competing events $(T_1, \ldots, T_M)$. However, the researcher can only observe $T = \min(T_1, \ldots, T_M)$, and the corresponding *cause of decrement* C;

Introduction

- **Aim**: estimate the joint distribution of the time to M competing events $(T_1, \ldots, T_M)$. However, the researcher can only observe $T = \min(T_1, \ldots, T_M)$, and the corresponding *cause of decrement* C;

- **Problem**: given the available data this joint distribution cannot be identified;

Introduction

- **Aim**: estimate the joint distribution of the time to M competing events $(T_1, \ldots, T_M)$. However, the researcher can only observe $T = \min(T_1, \ldots, T_M)$, and the corresponding *cause of decrement* C;

- **Problem**: given the available data this joint distribution cannot be identified;

$\Rightarrow$ Further point identifying assumptions are thus needed, for example:

- $T_1, \ldots, T_M$ are pairwise independent;
- Specify a copula model with known dependence parameter;
- Subdistribution approach.

Model (1)

For $c = 1, \ldots, M$, for each unit we have the following data generating process:

$$\log T_{c,i} = x_{c,i}\beta_c^T + \theta_{c,i} + \varepsilon_{c,i}; \qquad \varepsilon_{c,i} \sim N\left(0, \sigma_c^2\right) \tag{1}$$

$$\theta_i = (\theta_{1,i} \ldots, \theta_{M,i}) \sim P; \tag{2}$$

$$P \sim DP(\phi, MVN(m_\theta, \Sigma_\theta)) \tag{3}$$

Model (1)

For $c = 1, \ldots, M$, for each unit we have the following data generating process:

$$\log T_{c,i} = x_{c,i}\beta_c^T + \theta_{c,i} + \varepsilon_{c,i}; \qquad \varepsilon_{c,i} \sim N\left(0, \sigma_c^2\right) \tag{4}$$

$$\theta_i = \left(\theta_{1,i} \ldots, \theta_{M,i}\right) \sim P; \tag{5}$$

$$P \sim DP\left(\phi, MVN\left(m_\theta, \Sigma_\theta\right)\right) \tag{6}$$

This specification induces the following Dirichlet Process Mixture model for the joint density of $(T_1, \ldots, T_M)$:

$$\begin{aligned} & f\left(t_1, \ldots, t_M \mid x; \beta_1, \ldots, \beta_M, \sigma_1, \ldots, \sigma_M\right) \\ &= \int_\theta f\left(t_1 \mid x_1; \beta_1, \sigma_1, \theta_1\right) \cdots f\left(t_M \mid x_M; \beta_M, \sigma_M, \theta_M\right) dP(\theta) \\ &= \sum_{k=1}^{\infty} \pi_k f\left(t_1 \mid x_1; \beta_1, \sigma_1, \theta_{1,k}\right) \cdots f\left(t_M \mid x_M; \beta_M, \sigma_M, \theta_{M,k}\right) \end{aligned} \tag{7}$$

Model (2)

Furthermore, to complete the full Bayesian specification:

$$\phi \sim \text{Ga}(1, 1); \tag{8}$$

$$m_\theta \sim MVN(\Lambda_3, \Lambda_4); \tag{9}$$

$$\Sigma_\theta \sim \text{Inv-Wish}(\Lambda_5, \Lambda_6); \tag{10}$$

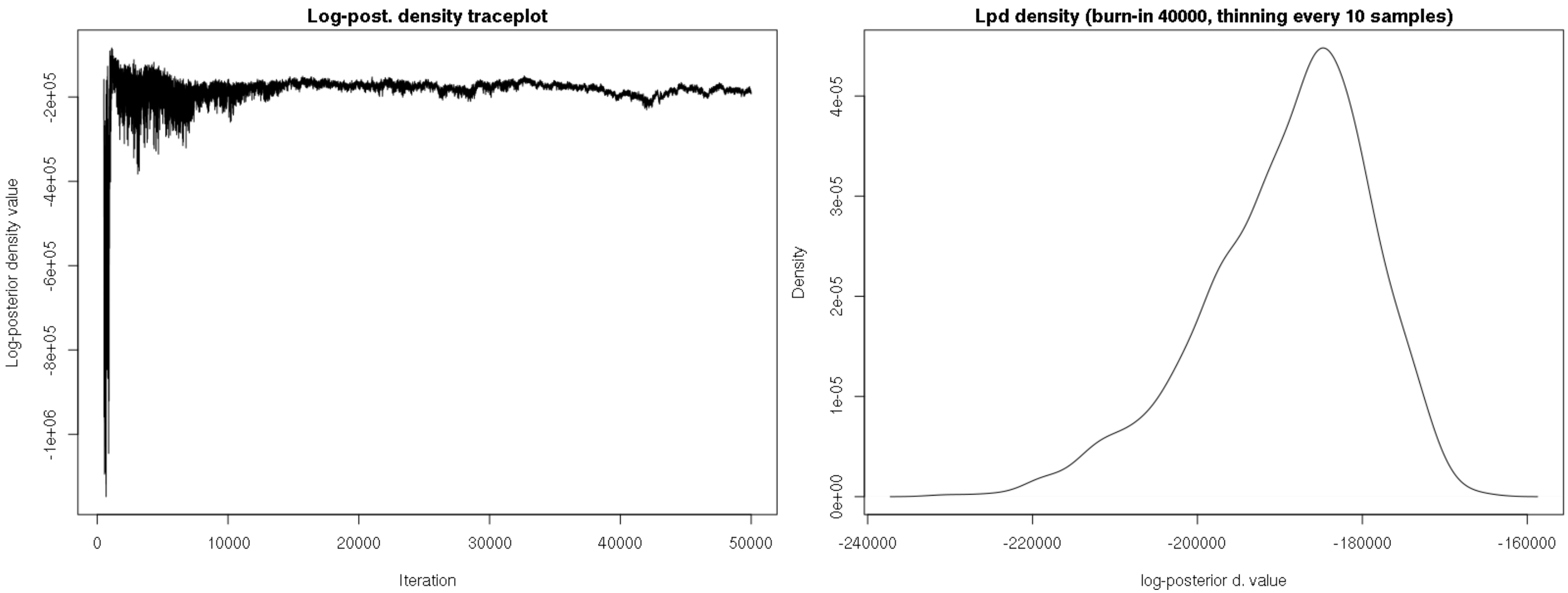
$$\beta_c \sim N(0, 10); \tag{11}$$

$$\sigma_c^2 \sim \text{Inv-Gamma}(\Lambda_{c,1}, \Lambda_{c,2}). \tag{12}$$

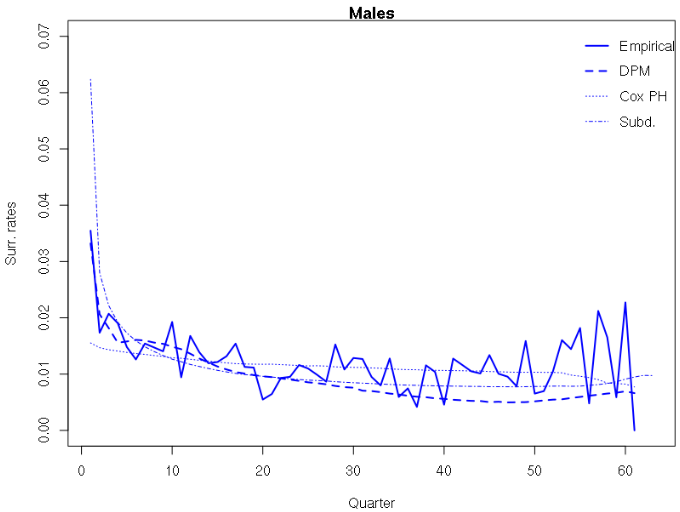
Dataset

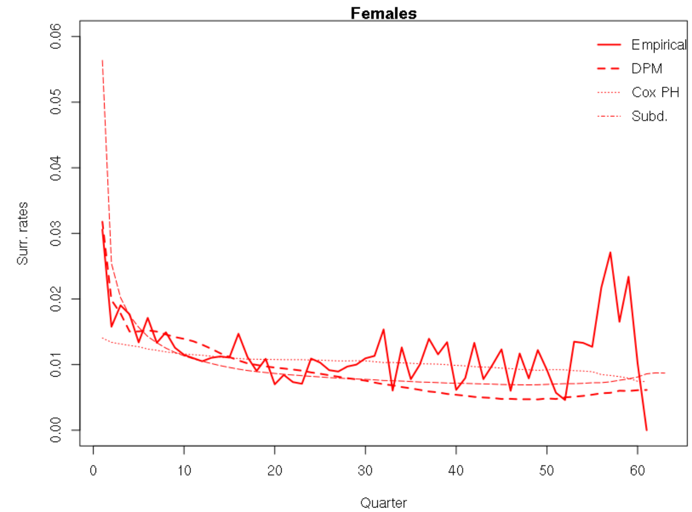
- 29,317 Whole life insurance policies (75% training set, 25% test set), from the R package `CASdatasets` (Dutang and Charpentier (2020));
- Observation period: 1st January 1995 - 31th December 2008;
- Causes of decrement ($M = 3$):
  - Surrending ($C = 1$);
  - Death ($C = 2$);
  - Other ($C = 3$);
- Covariates:
  - Age group;
  - Annual premium;
  - Payment frequency;
  - Smoking status;
  - Gender;
  - ...

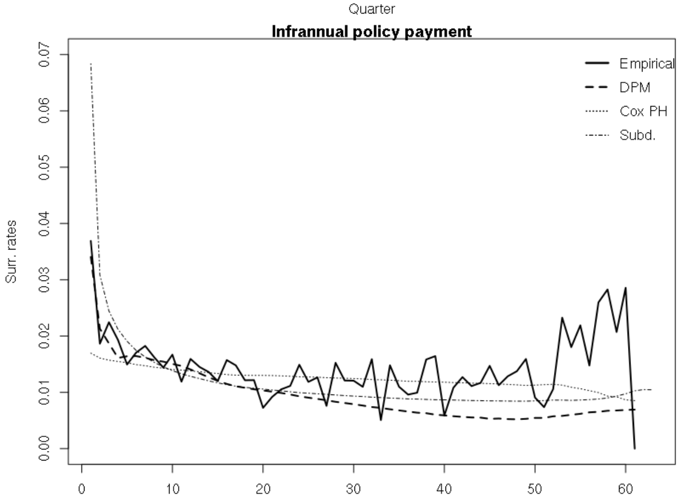
MCMC Convergence

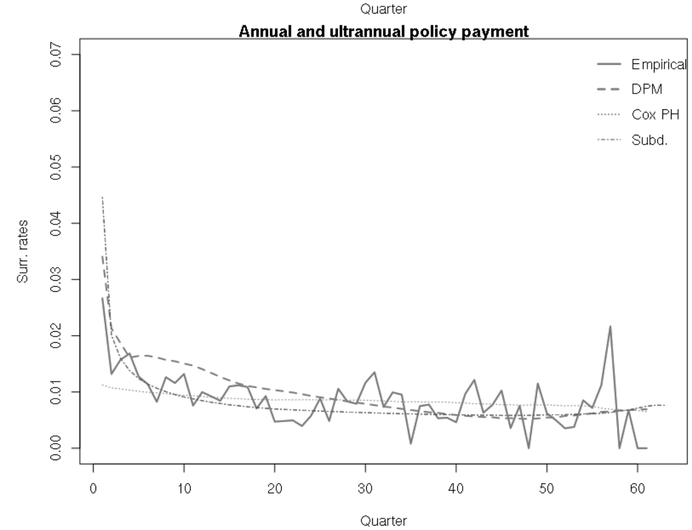


Prediction of surrending rates



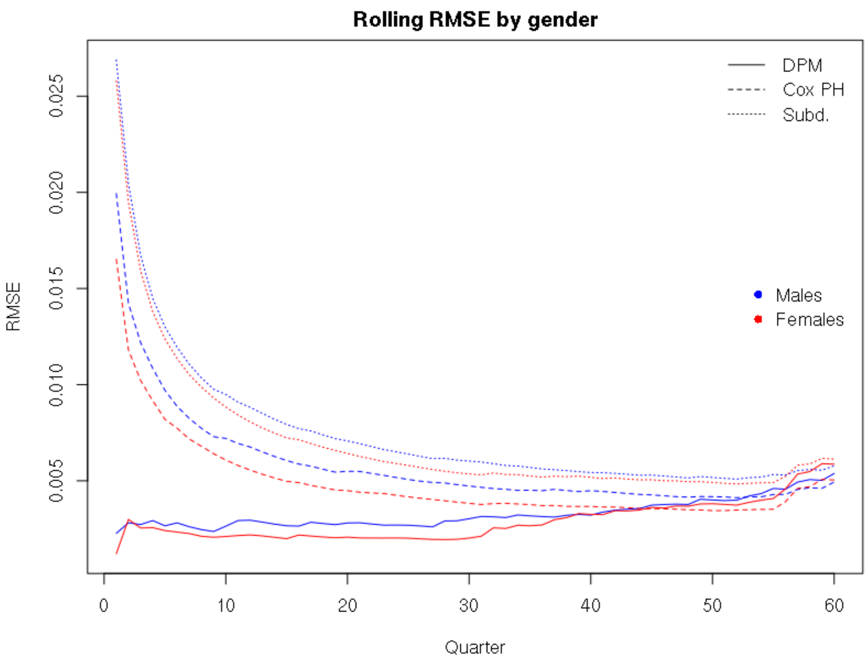


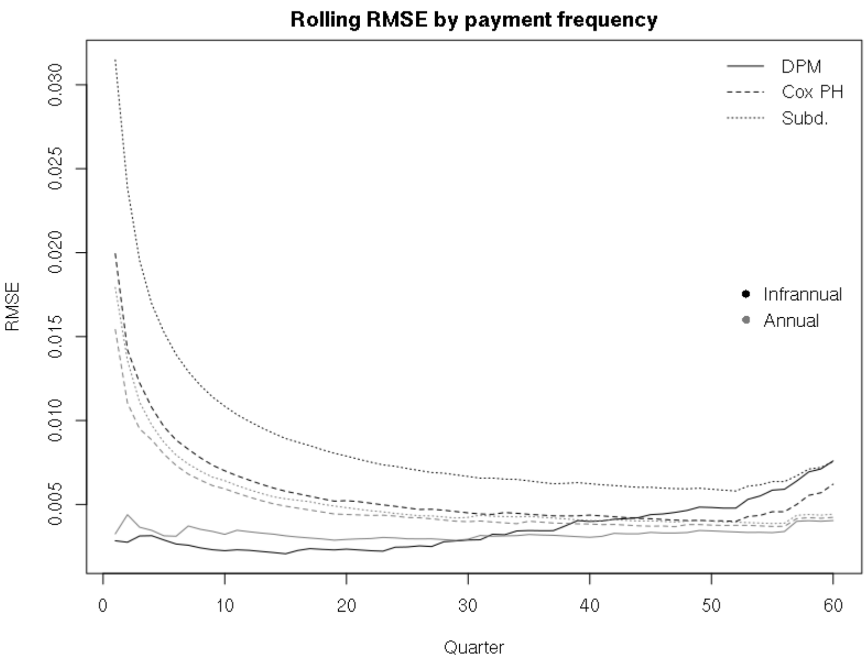




Prediction of surrending rates (RMSE)

$$\text{RMSE}_Q = \sqrt{\frac{\sum_{q=1}^{Q} \left(r_q^M - r_q^E\right)^2}{Q}} \tag{13}$$





References

- C. Dutang and A. Charpentier (2020), *CASdatasets R Package*;
- F. Ungolo and E.R. van den Heuvel (2021), *A Dirichlet Process Mixture model for the analysis of competing risks*, Working paper;

ERGO Center of Excellence in Insurance

Eine Einrichtung der TUM gefördert von der ERGO Group

Thank you for your attention!