JOINT AND SURVIVOR ANNUITY VALUATION WITH A BIVARIATE REINFORCED URN PROCESS

A PREPRINT

Luis Souto Arias*
Centrum Wiskunde & Informatica
The Netherlands
Luis.Souto.Arias@cwi.nl

Pasquale Cirillo
Delft University of Technology
The Netherlands
P.Cirillo@tudelft.nl

July 27, 2020

ABSTRACT

Using a Bivariate Reinforced Urn Process (B-RUP), a novel way of modeling the dependence of coupled lifetimes is introduced, with application to the pricing of joint and survivor annuities.
In line with the machine learning paradigm, the model is able to improve its performances over time, but it also allows for the use of a priori information, like for example experts' judgements, to complement the empirical data.
Using a well-known Canadian dataset, the performances of the B-RUP are studied and compared with the existing literature.

Keywords Annuity · Bivariate Survival Function · Reinforced Urn Process · Bayesian Nonparametrics

1 Introduction

We propose a novel nonparametric approach to the modeling of joint and survivor annuities, particularly useful in the case of right-censored observations. We build on previous results by Bulla et al. [2007], Muliere et al. [2000], Walker and Muliere [1997]. The goal is to propose an alternative to copulas and other parametric constructions already used for the modeling of coupled lifetimes in insurance, for example in Frees et al. [1996], Carriere [2000] and Luciano et al. [2008] to cite a few important ones.

The basic ingredients of our approach are Pólya urns [Mahmoud, 2008], probabilistic objects with the ability of intuitively representing the idea of learning via a reinforcement mechanism, so that the model here presented can actually be seen as an unusual machine learning algorithm [Murphy, 2012].

Differently from the great majority of machine learning algorithms, the Bivariate Reinforced Urn Process (B-RUP) here discussed belongs to a particular class of models [Muliere et al., 2000, Walker and Muliere, 1997] with the extremely useful ability of combining some a priori knowledge–possibly referring to experts' judgements–with the information coming from actual data, perfectly in line with the Bayesian paradigm. This possibility allows for the incorporation of trends, tail events or other aspects that can be rarely observed in a dataset–hence invisible to standard machine learning approaches, yet possible and with dramatic consequences [Taleb, 2007]. For instance, if an expert thinks that their data under-represent a given phenomenon, like for example some unusual lifetime combinations in the modeling of survivor annuities, they could solve the problem by eliciting an a priori assigning a higher mass to those combinations, to oblige their model to always take into account their possibility, at least remotely, even if rarely (or not) present in the data. In a sense, a clever use of the priors can thus be an elegant way of dealing with sampling and historical biases [Derbyshire, 2017, Shackle, 1955].

Clearly, nothing guarantees the ability of eliciting a sound and reliable a priori: experts could naturally be wrong. The answer to such a relevant observation–we shall see–is that the bivariate urn model learns over time, at every interaction with actual data. A sufficient amount of data can therefore compensate for judgements one does not trust completely,

*Corresponding author.

for unrealistic beliefs, or for no belief at all. Furthermore, as recently observed [Cirillo et al., 2013, Cheng and Cirillo, 2018], reinforced urn processes are able to learn hidden patterns in the data, discovering previously ignored features. In any case, as it shall be clearer later, the elicitation of an a priori is not compulsory–it is only a relevant plus–and a B-RUP can also be used in an "objective" empirically-based way.

Another interesting feature of reinforced urn processes, when seen through statistical learning glasses, is that they are somehow immune to the "black box" argument [Knight, 2017] commonly used to criticize machine/deep learning. In fact, differently from other approaches, a B-RUP can be controlled and studied in its probabilistic details, controlling its features and allowing for fine tuning.

The paper is organized as follows. In Section 2 we briefly recall joint and last-survivor annuities, while in Section 3 we shortly deal with the problem of right-censoring. In Section 4, we describe the B-RUP model, focusing on its ability of dealing with censored observations. In Section 5, we discuss our results, using artificial and real data, and providing suggestions on how to use the B-RUP in practice, in terms of prior elicitation and parameter tuning. The real data are the same used by Frees et al. [1996], that we thank for sharing them with us, and that will be a benchmark for us. Interestingly, as we shall see, the B-RUP is able to obtain the same results of Frees et al. [1996], while guaranteeing more flexibility. Finally, Section 6 concludes the work and looks at future research.

2 Joint and survivor annuities

An annuity is a financial contract between two counterparties: an individual, called the annuitant, and a financial institution, like an insurance company. In such a contract, the former pays a given amount of money (in full or periodically) to the latter and, in return, they receive periodic payments, usually starting from a given time point in the future, and commonly until death. The goal of annuities is therefore to provide the annuitant with a steady stream of income, typically during retirement, and to offer insurance against lifetime uncertainty. It goes without saying that annuities are strictly connected to private pensions and other similar products [Sheshinski, 2007].

A simple and quite popular extension of the single annuitant case we have just described is given by annuities involving more annuitants, most of the times just two. These are often the members of a married couple or of a partnership, but parent-child and relative-relative combinations are also common [Frees et al., 1996, Brown and Poterba, 2000]. An example is offered by *joint and last-survivor* annuities, in which the insurance company pays as long as at least one of the two annuitants is alive. In this kind of contracts, the death of one of the annuitants may have an impact on the stream of payments that the survivor receives. For instance, for $p \in [0, 1]$, in a $100p\%$ (e.g. 50%) joint and survivor annuity contract, payments are made in full while both annuitants are alive, but if one of the two dies, only $100p\%$ (e.g. 50%) of the original amount is paid to the survivor [Sheshinski, 2007, Winklevoss, 1977]. A popular version is the contingent joint and survivor annuity, which pays the full benefit as long as the first annuitant (plan member) is alive, and then changes the benefit to a portion which is usually one-half or two-thirds.

From an actuarial point of view, to evaluate a joint and survivor annuity, one needs the marginal and joint survival functions of both annuitants. Let X and Y be two random variables representing the lifetimes of two annuitants in a joint and survivor annuity. Assume that today the age (say in years) of the first annuitant is observed to be x, while for the second we have y. The joint probability of survival for another k years, given the current ages, is

$$S_{XY}^{(x,y)}(k,k) := \mathbb{P}(X > x + k, Y > y + k | X > x, Y > y), \tag{1}$$

where $k = 0, 1, ..., K$.

We can then look at the probability that at least one of the two annuitants survives another k years, i.e.

$$\mathrm{P}_{LS}^{(x,y)}(k) := \mathbb{P}[(X > x + k) \cup (Y > y + k) | X > x, Y > y] = S_X^{(x,y)}(k) + S_Y^{(x,y)}(k) - S_{XY}^{(x,y)}(k,k), \tag{2}$$

where the subscript LS is the acronym of "last survivor", with $S_Y^{(x,y)}(k) = \mathbb{P}(Y > y + k | X > x, Y > y)$ indicating the marginal survival probability of the second annuitant given the present situation of the couple, and similarly $S_X^{(x,y)}(k)$ for the first annuitant. It is important to stress that, in Equation (2), both marginal probabilities are conditioned on the ages of both annuitants. This is because, in the general situation where the lifetimes X and Y share some kind of dependence, we have that $S_X^{(x,y)}(k) \neq S_X^{(x,0)}(k)$ [Youn et al., 2002].

If we assume a constant interest rate r, the basic one-unit pricing formula for a 100% joint and last-survivor annuity is given by

$$C(x, y) = \sum_{k=0}^{\infty} \frac{\mathrm{P}_{LS}^{(x,y)}(k)}{(1 + r)^k}. \tag{3}$$

In the $100p\%$ situation, with $0 < p < 1$, this formula is easily adjusted [Winklevoss, 1977].

Equations (2) and (3) imply that, in order to evaluate joint and survivor annuities, one needs to specify the dependence between the lifetimes of the annuitants. If we assume that their lifetimes are independent, then $S_{XY}^{(x,y)}(k,k) = S_X^{(x)}(k)S_Y^{(y)}(k)$, and the estimation of the marginal probabilities of survival is all that is needed. Since the estimation of univariate survival functions is a widely studied and well-known topic (e.g. Cox and Oakes [1984]), pricing annuities under independence makes both modelling and computations considerably easier [Winklevoss, 1977], but not necessarily realistic.

That of independence is indeed always a strong assumption. As Frees et al. [1996], Carriere [2000], Luciano et al. [2008], Sanders and Melenberd [2016] and others have shown, the lifetimes of several couples in annuity contracts present a non-negligible positive dependence. Moreover, some studies suggest that annuitants tend to have higher individual survival probabilities with respect to people who do not buy annuities [Mitchell et al., 1999]. Such a result opens to the possibility that also the dependence structure among couples that buy annuities, and couples that do not, differs. Although there is no conclusive study about differences in the coupled lifetimes of annuitants versus non-annuitants, Sanders and Melenberd [2016] observe that the magnitude of the positive dependence for a random subset of $50,000$ couples–not necessarily annuitants–sampled from the whole Dutch population is significantly smaller than what Frees et al. [1996] and Luciano et al. [2008] find for about 15000 couples of annuitants in Canada[2].

Regardless of its magnitude, there are several reasons that explain, at least qualitatively, why the dependence in the lifetimes of couples (especially married couples and partnerships) is positive. Some of them are rather intuitive, like the possibility of both individuals dying at the same time because of an accident or a contagious disease, or the common habits they might share as a consequence of living together, like the food they eat or the environment they live in. Others are more linked to emotional, psychological and medical aspects, like the "broken heart syndrome", a well-known phenomenon in medicine, in which the sudden passing of one spouse greatly increases (at least temporarily) the probability of death of the surviving one because of bereavement and sadness [Jagger and Sutton, 1991].

A question that naturally arises, when we consider the several possibilities for the modelling of joint mortality, is how these affect the annuity price in Equation (3). This question is well-addressed in Frees et al. [1996], where they compute the ratio between the annuity price under positive dependence and the price under independence (we shall call it annuity ration from now on). To model dependence, they make use of copula models with Gompertz and Weibull marginals, and they find out that the annuity ratio varies between approximately 0.95 and 1.05, depending on the initial ages of the annuitants at contract initiation. This means that, if one assumes independence to simplify modelling, they may end up with the underestimation or the overestimation of the annuity price by as much as a 5%. Compatible results are more recently obtained by Sanders and Melenberd [2016], even if the error they find is smaller than 5% on average, in absolute terms, for different types of contracts. Furthermore, the same authors underline that the error can be particularly larger and relevant for specific types of annuities, in which the joint survival distribution plays a major role, like for instance joint annuities.

3 Right-censoring

In survival analysis, right-censoring is a well-known problem one has to deal with in estimation [Kaplan and Meier, 1958].

When an observation is right-censored, it means that we only have reliable information until a certain point in time, and that, from that point onwards, we actually lose precision. In other words, we end up in a situation in which the quantity of interest is at least equal to x, but we do not known if the real value is actually $x+1$, $x+2$ or $x+n$, because, for several reasons, our observation is censored. For example we may know that an individual reached an age of 60 years, but then the observation is stopped, because this person exits our sample and we lose track of the actual age of death (is it 61 or 78?), a fundamental information for us.

The concept of censoring can also be defined on the left-side of a random variable, but it is less relevant here. For more details we refer for instance to Klein and Moeschberger [2003].

Mathematically, if X is our age variable, when right censoring occurs we only observe the minimum between X and the random censoring time T^X, that is, we observe $X^* = \min(X, T^X)$. The indicator δ_X is used to indicate if a given observation is censored ($\delta_X = 0$) or not ($\delta_X = 1$), so that $E[\delta_X]$ represents the probability of no censoring. Every potentially censored observation is therefore represented by the couple (X^*, δ_X). A common assumption [Gribkova and Lopez, 2015] is that the support of the distribution of X is included in the support of the censoring variable T^X.

[2]Naturally an analysis taking into consideration the differences in the two countries and the related markets would be necessary to clean the effects, but the idea of some selection mechanism does not seem ludicrous.

While univariate censoring has been extensively studied in the literature, e.g. in Kaplan and Meier [1958], Elandt-Johnson and Johnson [1980] or Cox and Oakes [1984], far less progress has been made in the bivariate case. One of the first works to study this problem is that of Dabrowska [1988], in which the Kaplan-Meier estimator is extended to the two-dimensional framework. However, this estimator is known to produce negative masses when a large amount of censoring is present [Pruitt, 1991]. In Wang and Wells [1997] or Gribkova and Lopez [2015], the censoring conditions are relaxed to obtain better estimators. For example, in the context of coupled lifetimes, Gribkova and Lopez [2015] assume that the age difference between individuals is known and related to the censoring variables; in Lin [1993], conversely, the censoring variable is taken to be the same for both lifetimes. Other recent articles on nonparametric estimators under bivariate censoring are those of Lopez [2012] and Shen and Yan [2008]. In all these works, censoring is assumed to be independent from the variables of interest, an assumption we will also make in our construction.

Similarly to the univariate case, what one observes in the presence of bivariate censoring is the quartet $(X^*, \delta_X, Y^*, \delta_Y)$, where X^*, δ_X, Y^* and δ_Y are defined as before. Once again, (X, Y) are assumed independent from (T^X, T^Y).

4 The Bivariate Reinforced Urn Process

The Bivariate Reinforced Urn Process (B-RUP) is our proposal for the modeling of joint and survivor annuities. It is a model originally introduced by Bulla [2005], Bulla et al. [2007], and here extended and adapted to the annuity framework.

We will show that this model is not only able to replicate the results one can obtain with other existing methodologies, like copulas, but it also allows for a more flexible modeling, it improves its performances over time, and it allows for the exploitation of experts' judgements.

To describe the model, we start by defining its main building blocks: the beta-Stacy process [Walker and Muliere, 1997] and the the Reinforced Urn Process (RUP) [Muliere et al., 2000]. We then show how several RUPs can be easily combined into a one-factor construction to give birth to a very flexible and powerful model, the B-RUP.

4.1 Reinforced Urn Processes and beta-Stacy processes

The Reinforced Urn Process was first described in Walker and Muliere [1997], where the results of Blackwell and MacQueen [1973] and Ferguson [1973] are extended to right-censored data, thus defining the so-called beta-Stacy process, a neutral-to-the-right [Doksum, 1974] generalization of the Dirichlet process [Ferguson, 1973].

The RUP is a combinatorial stochastic process, and it can be seen as a reinforced random walk over a state space of urns. Depending on how its parameters are specified, it can generate a large number of interesting models. Essential references on the topic are Muliere et al. [2000, 2003] and Fortini and Petrone [2012]. Examples of applications can be found for example in Peluso et al. [2015] and Cheng and Cirillo [2018].

In this paper we specify a reinforced urn process able to generate a discrete beta-Stacy process, a particular random distribution over the space of discrete distributions.

Definition 4.1 (Walker and Muliere [1997])**.** A random distribution function F is a discrete beta-Stacy process with jumps at $j \in \mathbb{N}$ and parameters $\{\beta_j, \omega_j \in \mathbb{R}^+, j \in \mathbb{N}\}$, if there exist mutually independent random variables $\{V_j\}_{j\in\mathbb{N}}$, each beta distributed with parameters (β_j, ω_j), such that the random mass assigned by F to $\{j\}$, written $F(\{j\})$, is given by $V_j \prod_{i<j}(1 - V_i)$.

Following Definition 4.1, we introduce couples $\{\beta_j, \omega_j\} \in \mathbb{R}^+ \times \mathbb{R}^+$, with $j \in \mathbb{N}$, such that $\beta_j, \omega_j \geq 0$, $\beta_j + \omega_j > 0$, and $\lim_{n\to\infty} \prod_{j=0}^{n} \frac{\omega_j}{\beta_j+\omega_j} = 0$. Next, given a discrete beta-Stacy process F with parameters $\{\beta_j, \omega_j \in \mathbb{R}^+, j \in \mathbb{N}\}$, and a sample $(\boldsymbol{X}_n^*, \boldsymbol{\delta}_n)$ of exchangeable[3] and potentially censored observations, with $\boldsymbol{X}_n^* = \{X_n^*, n \geq 1\}$, the sequence $\boldsymbol{X}_n = \{X_n, n \geq 1\}$ is a RUP if

$$\hat{S}_X(x) = \mathbb{P}(X_{n+1} > x | \boldsymbol{X}_n^* = \boldsymbol{x}_n^*, \boldsymbol{\delta}_n = \boldsymbol{d}_n) = \prod_{j=0}^{x} \left[1 - \frac{\beta_j + m_j^*(\boldsymbol{x}_n^*, \boldsymbol{d}_n)}{\beta_j + \omega_j + s_j(\boldsymbol{x}_n^*)}\right], \tag{4}$$

where $m_j^*(\boldsymbol{x}_n, \boldsymbol{d}_n) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i=j, d_i=1\}}$ is the number of exact observations at $x = j$ and $s_j(\boldsymbol{x}_n) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i\geq j\}}$ is the number of observations at $x \geq j$. As proved in Walker and Muliere [1997], the condition $\lim_{n\to\infty} \prod_{j=0}^{n} \frac{\omega_j}{\beta_j+\omega_j} = 0$ is necessary to ensure that the trajectories of the discrete beta-Stacy process are indeed random probability distributions.

[3] Exchangeability is a common assumption in Bayesian statistics [de Finetti, 2017], and it represents a relaxation of the stronger hypothesis of independence. A sequence of random variables is exchangeable, if their joint distribution is immune to permutations in the order of appearance of the variables.

Notice that, by defining $\beta_j^* = \beta_j + m_j^*(\boldsymbol{x}_n, \boldsymbol{d}_n)$ and $\omega_j^* = \omega_j + s_j(\boldsymbol{x}_n) - m_j^*(\boldsymbol{x}_n, \boldsymbol{d}_n)$, we can obtain a new beta-Stacy process F^* with parameters $\{\beta_j^*, \omega_j^* \in \mathbb{R}^+, j \in \mathbb{N}\}$. The beta-Stacy process is thus conjugate to right-censored data, a property somehow expected being the beta-Stacy process a particular case of a neutral-to-the-right process [Doksum, 1974]. The importance of conjugacy is to be stressed, as it allows for continuous updating of the process over time, every time new observations become available. In other words, the posterior one can obtain as the result of a first cycle of Bayesian updates can then represent the a priori of a new cycle, without the necessity of restarting everything from scratch [Muliere et al., 2000, Cirillo et al., 2013].

In the introduction we mentioned that the RUP is linked to Pólya urns, a feature that makes a complex mathematical tool like the beta-Stacy process very intuitive to explain and to work with. Following Muliere et al. [2000], we can easily show that Equation (4) can be obtained via a sequence of two-color Pólya urns, in the case of no censoring.

Assume we have $M+1$ Pólya urns [Mahmoud, 2008], where the j-th urn U_j, $j = 0, 1, ..., M$, initially contains $\omega_j > 0$ white balls and $\beta_j > 0$ blue balls. The only exception is urn U_0, which only possesses white balls. This starting urn composition is nothing more than a way of eliciting an a priori [Muliere et al., 2000]. We start by drawing a ball from U_0. The only color we can sample is white. We put back the sampled ball, add an extra white ball[4] and move to sample urn U_1. If in U_1 the sampled ball is again white, we put it back, add an extra ball of the same color, and go sampling urn U_2. If the ball sampled from U_2 is also white, then we add a white ball to U_2 and draw from U_3, and so on. Every sampled white ball makes us move one step further in the sequence of urns. But if the sampled ball is blue, say in urn U_2, we add an extra blue ball to U_2, set $X_1 = 2$, and we start drawing again from U_0. If the next blue ball is drawn, say, at U_{20}, then we set $X_2 = 20$, and start again from U_0. This sampling scheme defines a reinforced random walk on the state space of the $M+1$ urns: every extraction of a blue ball determines a cycle for the process $\{X_n\}$, which is nothing more than a RUP. Note in fact that, after n samplings of X, the probability distribution of X_{n+1} will indeed be given by Equation (4).

The urn construction just analyzed generates a RUP without right censoring. The possibility of censored observations can, however, be easily introduced: in case of a right-censored value j, under the k-th cycle[5] of the RUP, we add one white ball to all urns up to U_j (included), but no blue ball in any of the urns, then we restart from U_0.

In every cycle, the RUP learns and keeps memory of what happened before, thanks to the Pólya reinforcement, combining the information from samplings with the initial compositions of the urns. In Muliere et al. [2000], the properties of this urn construction, including the conditions for recurrence, are studied in detail.

The way in which actual observations are translated into urn samplings is rather intuitive. Imagine we observe the lifetime of an individual, whose death was at the age of 83. If every urn represents a year, it means that for 82 urns, starting from U_0 (birth) we have sampled a white balls, and only in U_{83} a blue one, which resets the RUP to 0. Each individual lifetime thus represents a cycle of the RUP, and we assume individuals to be exchangeable. The initial composition of all the urns, before the first individual is observed, and their lifetime transformed into samplings, just represents our a priori about lifetimes. The more individuals we observe, the more urns are updated, and the more we learn about the empirical distribution of lifetimes.

One of the characteristics of the beta-Stacy process, inherited from the Dirichlet process, is that its trajectories are centered around a certain probability distribution $G(\cdot)$, that is $\mathbb{E}[F(\{j\})] = G(\{j\})$, which—in Bayesian terms—plays the role of the prior. As shown in Walker and Muliere [1997], a necessary condition for this is that

$$\frac{\beta_j}{\beta_j + \omega_j} = \frac{G(j) - G(j-1)}{1 - G(j-1)}, \qquad j \in \mathbb{N} \tag{5}$$

where $G(j) = \mathbb{P}_G(X \leq j)$ is the probability that X is at most j under the centering measure.

The prior distribution G can also be used to define the initial compositions in a RUP. Following Walker and Muliere [1997] and Muliere et al. [2000], it is sufficient to set

$$\beta_j = c_j G(\{j\}), \qquad \omega_j = c_j(1 - G(j)), \qquad c_j \in \mathbb{R}^+, j \in \mathbb{N}, \tag{6}$$

with c_j denoting the so-called strength of belief in the prior knowledge, and $G(\{j\}) = \mathbb{P}_G(X = j)$. Notice that, given Equation (6), the necessary conditions for the couples $\{\beta_j, \omega_j\}$ to properly define a beta-Stacy process are automatically verified, if $G(\cdot)$ is a well-defined probability distribution.

[4]This is the reinforcement mechanism [Mahmoud, 2008] and it can be generalized to any fixed number $r > 0$ of extra balls. A high reinforcement means that the probability of picking a particular color greatly increases every time we sample it, while the opposite occurs for small values of r. See Cirillo et al. [2010], Peluso et al. [2015] and Cheng and Cirillo [2018] for a discussion about the impact of r in RUPs.

[5]Every cycle starts with U_0 and ends with the extraction of a blue ball in one of the subsequent urns. Given the condition $\lim_{n\to\infty} \prod_{j=0}^{n} \frac{\omega_j}{\beta_j + \omega_j} = 0$, it is easy to verify that the RUP is recurrent. More details, also for the case of censoring, in Muliere et al. [2000], Cirillo et al. [2013].

By giving high values to the strength of belief c_j, larger amounts of data are required in order for the posterior distribution, obtained via sampling and given in Equation (4), to move away from the a priori. On the contrary, when $c_j \to 0$, Equation (4) tends to the Kaplan-Meier (KM) estimator of Kaplan and Meier [1958].

4.2 Building dependence

We can extend Equation (4) to two dimensions, using a one-factor construction initially discussed in Bulla [2005], Bulla et al. [2007], to define the Bivariate Reinforced Urn Process (B-RUP).

Assume we observe n couples of possibly censored lifetime data of the form $((\boldsymbol{X}_n^*, \boldsymbol{\delta}_n), (\boldsymbol{Y}_n^*, \boldsymbol{\epsilon}_n))$, where $\boldsymbol{X}_n^*$ and $\boldsymbol{Y}_n^*$ are n-dimensional vectors of observations, and where $\boldsymbol{\delta}_n$ and $\boldsymbol{\epsilon}_n$ are the corresponding vectors of indicators for right censoring. A simple and flexible way of modelling the dependence between the observed lifetimes $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ is to consider three independent processes: one in common between $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$, and two idiosyncratic ones, assembled into a one-factor construction.

Formally, let $\boldsymbol{A}_n = \{A_i, 1 \leq i \leq n\}$, $\boldsymbol{B}_n = \{B_i, 1 \leq i \leq n\}$ and $\boldsymbol{C}_n = \{C_i, 1 \leq i \leq n\}$ be three independent reinforced urn processes, with respectively parameters (β_j^A, ω_j^A), (β_j^B, ω_j^B) and (β_j^C, ω_j^C) for $j \in \mathbb{N}$. One can easily create a bivariate model of the form

$$\begin{aligned} X_i &= A_i + B_i \\ Y_i &= A_i + C_i, \quad 1 \leq i \leq n. \end{aligned} \tag{7}$$

The lifetime of X_i and Y_i is thus composed of a common quantity A_i and two personal elements, B_i and C_i. In this way, the dependence between X_i and Y_i relies entirely on A_i, and thus, conditioned on A_i, X_i and Y_i are independent.

A straightforward calculation yields

$$\text{Cov}(X_{n+1}, Y_{n+1} | \boldsymbol{A}_n, \boldsymbol{B}_n, \boldsymbol{C}_n) = \text{Var}(A_{n+1} | \boldsymbol{A}_n) > 0, \quad n \geq 1, \tag{8}$$

thus implying that the B-RUP construction can only model positive dependence. For the application at hand, that is the joint modelling of lifetimes of coupled annuitants, this is not a problem, since previous studies have shown a strictly positive dependence [Frees et al., 1996]. However, it is important to stress that the B-RUP model should not be used when negative dependence is also possible.

In Bulla [2005], the sequence $\{(X_n, Y_n), n \geq 1\}$ is shown to be exchangeable (given that the RUPS $\{A_n\}$, $\{B_n\}$ and $\{C_n\}$ are exchangeable by construction), and therefore, by the de Finetti representation theorem [de Finetti, 2017], there exists a joint random distribution function F_{XY} conditionally on which the elements of $(\boldsymbol{X}_n, \boldsymbol{Y}_n)$ are independent and identically distributed according to F_{XY}. The properties of F_{XY} have been studied in detail in Bulla et al. [2007].

Let F_X and F_Y be the marginal distributions of X and Y, respectively. Clearly, we have

$$\begin{aligned} F_X &= F_A \times F_B, \\ F_Y &= F_A \times F_C, \end{aligned}$$

so that both F_X and F_Y are convolutions of beta-Stacy processes. Furthermore, if P is the probability function corresponding to F, one has

$$P_{XY}(x, y) = \sum_{a=0}^{\min(x,y)} P_A(a) P_B(x-a) P_C(y-a), \qquad \forall x, y \in \mathbb{N}_0^2. \tag{9}$$

Although the joint posterior distribution of Equation (9) can in principle be computed analytically[6] [Bulla et al., 2007], its complexity grows quickly with the number of observations, making it already unfeasible for a relatively small sample. A convenient way of by-passing the problem is therefore to use a Markov Chain Monte Carlo (MCMC) approach, also known as Gibbs sampler, to obtain the joint probability distribution from the posterior distributions of $\boldsymbol{A}_n$, $\boldsymbol{B}_n$ and $\boldsymbol{C}_n$, using the one-factor construction.

The MCMC method consists of the following steps:

1. Generate a realization of $\boldsymbol{A}_n^{[k]}$, with the superscript referring to the iteration number, at the k-th iteration conditioned on the available data of the previous iteration. That is, sample $\boldsymbol{A}_n^{[k]}$ from its conditional distribution,

[6] Also observe that the conjugacy of the constituent RUPs $\{A_n\}$, $\{B_n\}$ and $\{C_n\}$ allows for the continuous updating of the B-RUP as well.

i.e.

$$\begin{aligned}\mathbb{P}[A_n^{[k]} = a_n | \boldsymbol{A}_{n-1}^{[k-1]}, \boldsymbol{X}_n^*, \boldsymbol{\delta}_n, \boldsymbol{Y}_n^*, \boldsymbol{\epsilon}_n] \propto \\ \mathbb{P}[B_n = x_n - a_n, \delta_n = d_n | \boldsymbol{B}_{n-1} = \boldsymbol{X}_{n-1}^* - \boldsymbol{A}_{n-1}^{[k-1]}, \boldsymbol{\delta}_{n-1}] \\ \mathbb{P}[C_n = y_n - a_n, \epsilon_n = e_n | \boldsymbol{C}_{n-1} = \boldsymbol{Y}_{n-1}^* - \boldsymbol{A}_{n-1}^{[k-1]}, \boldsymbol{\epsilon}_{n-1}] \\ \mathbb{P}[A_n^{[k]} = a_n | \boldsymbol{A}_{n-1}^{[k-1]}],\end{aligned} \tag{10}$$

where, using the fact that $\mathbb{P}[X \geq k] - \mathbb{P}[X > k] = \mathbb{P}[X = k]$, one has

$$\begin{aligned}\mathbb{P}[B_n = x_n - a_n, \delta_n = d_n | \boldsymbol{B}_{n-1} = \boldsymbol{X}_{n-1}^* - \boldsymbol{A}_{n-1}^{[k-1]}, \boldsymbol{\delta}_{n-1}] = \\ \mathbb{P}[B_n \geq x_n - a_n | \boldsymbol{B}_{n-1} = \boldsymbol{X}_{n-1}^* - \boldsymbol{A}_{n-1}^{[k-1]}, \boldsymbol{\delta}_{n-1}] - \\ \mathbb{1}_{\{\delta_n = 1\}} \mathbb{P}[B_n > x_n - a_n | \boldsymbol{B}_{n-1} = \boldsymbol{X}_{n-1}^* - \boldsymbol{A}_{n-1}^{[k-1]}, \boldsymbol{\delta}_{n-1}],\end{aligned}$$

and with the conditional distribution of C_n defined in an analogous way.

Since $\boldsymbol{A}_n^{[k]}$ is an exchangeable process, Equation (10) also applies for any element of $\boldsymbol{A}_n^{[k]}$ by simply changing $\boldsymbol{A}_{n-1}^{[k-1]}$ with $\boldsymbol{A}_{-i}^{[k-1]} = \{A_n^{[k]}, 1 \leq n < i, \ or \ A_n^{[k-1]}, n > i \geq 1\}$. With this in mind, start simulating $A_1^{[k]}$ using the values of $\boldsymbol{A}_{-1}^{[k-1]}$ and repeat this process until a new $\boldsymbol{A}_n^{[k]}$ has been obtained. Once this is done, compute $\boldsymbol{B}_n = \boldsymbol{X}_n^* - \boldsymbol{A}_n^{[k]}$ and $\boldsymbol{C}_n = \boldsymbol{Y}_n^* - \boldsymbol{A}_n^{[k]}$. Note that, defined in this way, right censoring is only applied to $\boldsymbol{B}_n$ and $\boldsymbol{C}_n$, and not[7] to $\boldsymbol{A}_n^{[k]}$.

2. Create a new combination (A_k, B_k, C_k) from the respective conditional marginal distributions, as per Equation (4).
3. Compute $X_k = A_k + B_k$ and $Y_k = A_k + C_k$. Set $k = k + 1$.
4. Repeat steps 1-3 until k reaches the maximum number of iterations N.

Notice that the previous algorithm requires, besides the (possibly) censored observations of (X, Y), an initial sample for A, that is, $\boldsymbol{A}_n^{[0]}$. Since A is not observed in practice, a sensible approach is to run the MCMC several times for different, reasonable samples $\boldsymbol{A}_n^{[0]}$.

In the next section, we present a detailed analysis of the results obtained by implementing this algorithm for artificial and real datasets in the framework of annuity valuation.

5 Empirical results

In this section we apply the B-RUP model to calculate the dependence of coupled lifetimes, in order to price joint and survivor annuities.

We first consider an analytical example, where we sample possibly censored observations from a known distribution. Knowing the original distribution will not only allow us to estimate the difference between the original distribution and the posterior computed through MCMC, but also to study how these differences affect the final price of the annuity.

In the second part of the section, the B-RUP is tested on a well-known Canadian dataset originally used by Frees et al. [1996], which we will also use as a benchmark.

5.1 Analytical example

Assume that X and Y are linked through the one-factor model of Equation (7). Assume that $A \sim \text{Poi}(25)$, $B \sim \text{Poi}(35)$ and $C \sim \text{Poi}(40)$. This implies that the marginals of X and Y are Poisson distributions with parameters 60 and 65, respectively; and that $\text{Cov}(X, Y) = \text{Var}(A) = 25$. The correlation between X and Y is therefore approximately 0.4.

Moreover, while not necessary true in general, we will make some assumptions about the dependence between T^X and T^Y, so that we can generate censored observations and better study the properties of our construction. We start from the naive observation that the two annuitants of a joint and survivor annuity enter the contract at the same time; say that X_0 is the age of the first annuitant at the date of the signature, while Y_0 is that of the second person. If we denote the

[7] Assuming that the joint lifetime component is not censored appears acceptable from an empirical point of view [Frees et al., 1996, Carriere, 2000].

observation period by Δ, it is clear that $T^X = X_0 + \Delta$ and $T^Y = Y_0 + \Delta$. Moreover, $X_0 = Y_0 + \epsilon$, where ϵ is the age difference between the annuitants, which can be negative if $Y_0 > X_0$.

Therefore, to obtain right-censored observations, we have to choose the distributions of T^X, Δ and ϵ. In what follows we take $T^X \sim \text{Poi}(50)$ ("Poi" stands for Poisson), $\Delta \sim \text{Poi}(2)$ and $\epsilon = \epsilon^* - \epsilon_0$, with $\epsilon^* \sim \text{Poi}(7)$ and $\epsilon_0 = 5$. We then sample $n = 10^4$ couples of observations, of which almost 90% turn out to be at least partially censored (i.e. at least one of the annuitants in the couple is right-censored).

Following the procedure defined in Section 4, the first step to use the B-RUP is to define prior distributions for the target variables according to Equation (6), as well as the strength of belief parameters. The priors we choose are $G_A = \text{Poi}(20)$, $G_B = \text{Poi}(20)$ and $G_C = \text{Poi}(20)$, respectively. For the strengths of belief, c^k with $k \in \{A, B, C\}$, we consider two different scenarios: one where the strengths of belief are very small, so that the posterior distribution is strongly affected by the incoming data, as if we did not trust our a priori; and another scenario where the strengths of belief are big enough to keep memory of the a priori—which we trust—and, in particular, to influence those areas in the data with less observations. As said, the a priori can indeed be used to complement the data with information about rarely observed events and trends. In Table 1 the values of the strengths of belief for both scenarios[8] are provided. In the following we will refer to them as the "low belief" and the "high belief" scenarios, or with the subscripts l and h, when referring to the results obtained with the B-RUP model. For example, when we write B-RUP$_l$, we refer to the B-RUP estimators in the low belief scenario. B-RUP$_h$ is defined analogously for the high belief scenario.

In order to initialize the Gibbs sampler, we also need to define an initial sample for A, which is not observable in practice. Therefore, this sample will be "artificial", and we need a reasonable way of generating it. Moreover, the resulting posterior distribution will be affected by this artificial sample, and thus in practice it is recommended to compute the posterior for several appropriate initial samples. In our case, we will sample observations from the same distribution that we chose for the prior of A, so that, for this particular example, we generate 10^4 samples from a $\text{Poi}(20)$.

	Low belief	High belief
c^A, c^B, c^C	10^{-6}	10^2

Table 1: Proposed strength of belief scenarios. Notice that, since we are considering a dataset of size 10^4, even under the high belief scenario, the prior will not be able to actually affect those areas where the observations are more concentrated, but it will definitely influence the areas with fewer data points.

In Figure 1 we show the recovered marginal distribution using the B-RUP via MCMC, as well as the KM estimators and the original distribution. Notice how, due to the lack of exact samples on the right tail of the distribution, the KM estimator is not properly defined on the whole support of the original distribution, but only up to the maximum value of the uncensored observations. The B-RUP model, on the contrary, is able to recover the right tail of the distribution, although the fitting is clearly worse in that part of the curve, because of the evident lack of observations. If available, a better a priori could here be used to improve tail fitting.

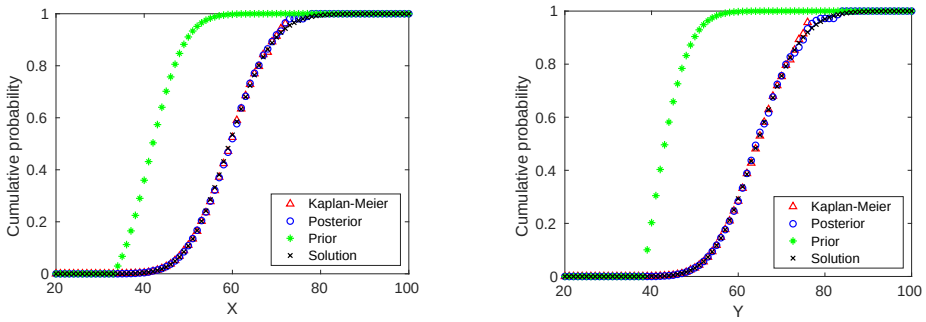
We also present the joint distribution in Figures 2a and 2b, for the low and high belief scenarios, respectively. Notice how, in Figure 2a, the fitting is worse in the upper-right corner than in the lower-left corner, in accordance with the right-censoring effect. For the high belief scenario, however, the prior distribution dominates in those areas of few observations and we recover a smoother surface, as per Figure 2b.

A more quantitative comparison can be found in Table 2, where we show the means, the variances and the correlation for both B-RUP scenarios, as well as for the KM estimator, the theoretical analytical solution, and the data. In this last case we provide the sample estimates under two different points of view: 1) correctly using only the truly uncensored observations ("Uncensored"), and 2) taking into consideration the entire dataset, ignoring the presence of censoring ("Whole"). While this second approach is not correct [Klein and Moeschberger, 2003], it can be heuristically useful to better understand the impact of right-censoring in the data. For example we can clearly see that not considering censoring can lead to a serious overestimation of the dependence in the data.

Always in Table 2, observe how the B-RUP is able to recover the correlation, under both scenarios[9], something not possible when using the KM approach. Moreover, observe how the B-RUP, in particular under the low strength of belief, better captures the variability of both X and Y.

[8] Note that we could go as far as to define different strengths of belief for each urn within a RUP [Muliere et al., 2000]. In this work, however, we will always assume a constant strength of belief for each RUP.

[9] Although both scenarios give reasonable results, the low belief scenario returns values closer to the actual ones, being more data-driven, as expected.



(a) Marginal of X.

(b) Marginal of Y.

Figure 1: Comparison of the marginal distributions of X and Y. In green we show the prior distribution, the Kaplan-Meier estimators in red, the B-RUP solution in blue and the original distribution in black. All distributions are conditioned on the minimum value of the uncensored observations. For this sample those were $X_{min} = 34$ and $Y_{min} = 39$.

As shown in Table 3, permutation tests performed on the means and the variances, under the null hypothesis of no difference between the obtained estimates and the true analytical values of Table 2, never reject the null for the B-RUP estimators under the low strength of belief scenario, at a standard 5% significance level. These permutation tests where performed simulating samples of size 10^3 from both the B-RUP (low and high strength of belief) and the KM marginals. For the mean, the test statistic used is the absolute difference between the means, while for the variance the Good's test [Good, 1994] is employed. Looking at Table 3, we can conclude that the B-RUP with low strength of belief even beats the KM, which is not able to correctly recover the variance of Y. If we also take into account that the KM approach cannot estimate the correlation, as already discussed for Table 2, the B-RUP$_l$ performance is even more appreciable.

Further, observe that, since the a priori is considerably different from the analytical solution, it is no surprise that, for the high strength of belief case, the B-RUP$_h$ performances worsen for the variances of both X and Y. As known, the variance is in fact particularly sensitive to discrepancies in the tails, and we already said that the B-RUP$_h$ puts a non-negligible mass on the right tail, which is never really updated by the data. Once again we want to stress that this is actually a sign of the power and flexibility of the B-RUP, in case of a credible a priori (clearly not the case in this very simple example).

	Uncensored	Whole	KM	B-RUP$_l$	B-RUP$_h$	Analytical
Mean(X)	52.877	50.684	59.811	59.811	59.341	60.005
Mean(Y)	56.690	53.088	64.691	64.691	64.227	65.007
Var(X)	35.498	39.524	53.486	55.755	50.886	59.905
Var(Y)	35.466	46.689	54.775	62.528	56.071	64.840
Corr(X,Y)	0.470	0.883	–	0.421	0.361	0.400

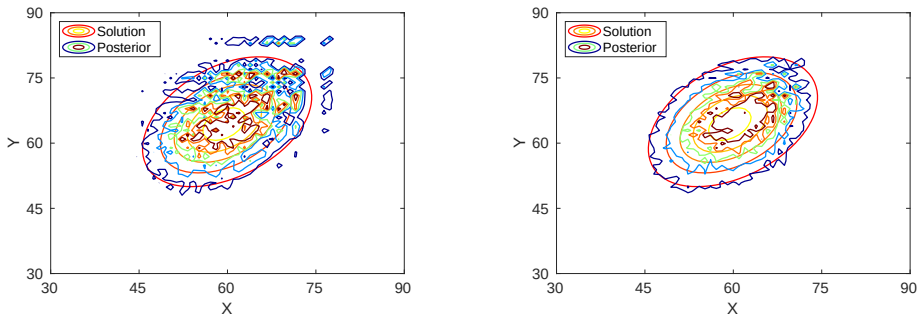
Table 2: Comparison of the means, the variances and the correlation of X and Y using the uncensored data, the whole dataset (wrongly assuming no censoring), the KM estimator, the B-RUP estimator and the original analytical solution, respectively.

Once we have the joint distribution of the pair (X, Y), we can price annuities using Equation (3). Following Frees et al. [1996], we are interested in the annuity ratio between the price when assuming dependence and the price when assuming independence. For the independent scenario we simply use the marginals obtained in Figure 1, while for the dependent scenario we use the joint distribution of Figure 2a. Moreover, according to Equation (3), once we know the joint distribution, the annuity price depends on three quantities: the ages of the two annuitants at the beginning of the contract, which we call entry/initial ages, and the interest rate. As in Frees et al. [1996], we first show the dependence of the annuity ratio for several initial ages with a fixed interest rate, and then we vary the interest rate and assume that both annuitants have the same age, so that we can still represent the results in a 3D plot.

In Figures 3a and 3b, we show the annuity ratio obtained with the B-RUP construction and the true (analytical) distribution for an interest rate of $r = 0.05$. Notice how, in this example, when considering dependence, the price can be up to 7% higher than in the independent scenario for some initial ages, especially when the age difference is large.

	KM	B-RUP$_l$	B-RUP$_h$
mean(X)	0.6612	0.4561	0.3254
mean(Y)	0.6256	0.1354	0.5106
var(X)	0.0881	0.1934	**0.0193**
var(Y)	**0.0239**	0.4859	**0.0217**

Table 3: Permutation test for the means and the variances of X and Y, using samples of 10^3 observations with 10^4 permutations. The sample are taken from the KM estimator and the B-RUP with low and high strength of belief. In bold we underline the situations for which the null hypothesis of no difference between the estimates and the true values (from the analytical model) is rejected, for a 5% significance level.

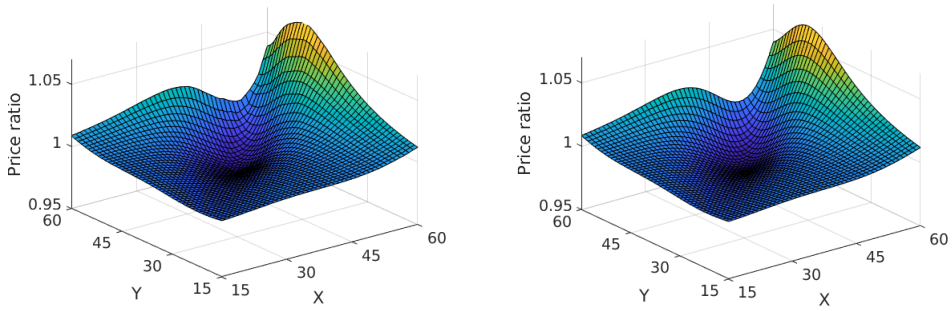


(a) Low strength of belief.

(b) High strength of belief.

Figure 2: Contour plot comparing the analytical distribution with the one obtained through the B-RUP model for the low (a) and high (b) belief scenarios. All distributions are conditioned on the minimum value of the uncensored observations: $X_{min} = 34$ and $Y_{min} = 39$.

For smaller age differences, conversely, the annuity ratio takes values below the unit. For low initial ages for both annuitants, the death event is still far away with high probability, and thus the annuity ratio barely moves from the unity in that area.



(a) B-RUP estimate.

(b) Original surface.

Figure 3: Ratio between the annuity values of the joint and independent survival models. The left plot (a) shows the surface obtained using the B-RUP construction with low strength of belief, while the right plot (b) shows the price ratio using the actual distribution (via the known analytics). Note how the ratio differs more between the two plots for high initial ages, since, because of the censored observations, the B-RUP fitting is worse in that area. The result is nevertheless quite positive.

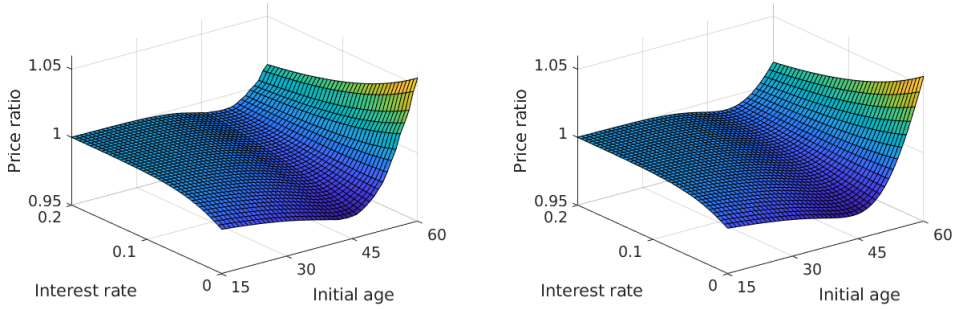
Looking at Figures 3a and 3b, it is precisely for high annuity ratios that the B-RUP estimators differ more from the analytical ones. This, again, is a consequence of the censored observations, that do not allow for a proper estimation due

to the lack of information about the right-tail of the distribution[10]. In Table 4 we present the absolute value differences between for several initial ages and a 5% interest rate. The results observed are clearly in line with what we have just said in terms of age differences. The performances of the B-RUP are definitely satisfactory.

Age FA \ Age SA	20	30	40	50	60
20	0.0003	0.0000	0.0001	0.0001	0.0002
30	0.0000	0.0005	0.0001	0.0001	0.0005
40	0.0001	0.0000	0.0009	0.0000	0.0008
50	0.0001	0.0003	0.0001	0.0013	0.0011
60	0.0005	0.0010	0.0024	0.0034	0.0002

Table 4: Absolute value differences in the annuity ratios obtained with the B-RUP model and the analytical distribution for several initial ages of the annuitants (Fist Annuitant - FA, Second Annuitant - SA) with an interest rate of 0.05. Given the small differences, we can say that the B-RUP is able to recover the "truth" from the data.

Finally, in Figures 4a and 4b, one can find the annuity ratio as a function of the interest rate for the same initial age for both annuitants. In accordance with the previous results, the annuity ratio is above one for large initial ages, and below one for low initial ages. Furthermore, this effect seems to be more pronounced for small interest rates, since the curves are steeper than for high interest rates. This tendency is naturally the same for both the B-RUP estimator and the original distribution.



(a) B-RUP estimate.

(b) Original surface.

Figure 4: Ratio between the annuity values of the joint and independent survival models. The left plot (a) shows the surface obtained using the B-RUP construction with low strength of belief, while the right plot (b) shows the price ratio using the true (analytical) distribution. A visual comparison suggests that the B-RUP definitely performs well.

5.2 Canadian dataset

Let us now consider a real-world application, using a well-known dataset, initially presented in Frees et al. [1996], and later also studied in the relevant works of Youn and Shemaykin [1999], Carriere [2000], Youn and Shemaykin [2001] and Luciano et al. [2008], among others.

The dataset[11] contains 14,497 contracts from a large Canadian insurer, and the period of observation runs from December 29, 1988, until December 31, 1993. To simplify the interpretation of the results, and also in line with other works [Frees et al., 1996, Luciano et al., 2008], we have removed same sex contracts and, for every couple, we have kept only one contract, leaving a total of 11,454 contracts. Finally, we have removed contracts were the annuitants age was less than 40 at the end of the observation period, obtaining a dataset of 11,421 male-female couples.

[10]Note that this should be seen as an indicator of the importance of choosing a proper prior distribution. Since the data barely contains any information about the right tail of the distribution, it is up to the expert to give a reasonable insight about tail events.

[11]We are grateful to Prof. Frees for sharing the data with us.

For each couple, the following information is available: dates of birth, dates of death or ages at the end of observation period, and date of contract initiation. Once the observation period ends, we no longer have any information about the individuals, and thus for those who lived longer than December 31, 1993, the lifetimes are right-censored.

Only a small portion of all the lifetimes in our data are actually uncensored, i.e. fully observed. In particular: 10.87% of the male lifetimes, 3.93% of the female lifetimes, and 1.71% of the joint (couple) lifetimes. The last 1.71% corresponds to 195 couples, of which 24 died within a one-day period, 59 within one month, and 92 within one year. Therefore, a considerable portion (47%) of the uncensored couple deaths are either likely due to an accident–which would explain both spouses' dying within a day–or to some other impactful reason like a highly contagious disease, or a strong broken heart syndrome [Jagger and Sutton, 1991]. In these situations, it is expected that the correlation between the annuitants' lifetimes is strongly positive, and it actually is. However, as we will see shortly, taking only into account the uncensored observations highly overestimates said dependence for all the other couples.

To model the data with the B-RUP, we follow the exact procedure as before. We start indeed by identifying the target variables, define their respective priors, and then apply the Gibbs sampler with predefined strengths of belief to merge the a priori knowledge with the data. The targets are, once again, A, B and C, and the priors we chose are $G_A = \text{Poi}(60)$, $G_B = \text{Poi}(15)$ and $G_C = \text{Poi}(15)$. With this choice, the marginals for X and Y are both Poisson distributions with parameter $\mu = 75$, and the correlation is approximately 0.8. This prior setting (from now on, the Poisson scenario) takes into account the sample moments and correlation one can estimate directly from the uncensored observations, and that are shown in the first column of Table 5. For the strengths of belief we will still use the scenarios in Table 1. As before, for initializing the Gibbs sampler we will generate observations from the prior we chose for A.

The decision of choosing distributions whose moments are in line with (a part of) the empirical data is to show a possible way of eliciting the a priori. Clearly, in case of experts' judgements or other sources of knowledge, one can also elicit completely subjective priors.

In order to analyze the impact of choosing a different a priori, we will also discuss the use of two alternative settings:

- $G_A = \text{Gomp}(50, 6)$, $G_B = \text{Gomp}(30, 2)$ and $G_C = \text{Gomp}(30, 3)$, where $\text{Gomp}(\mu, \sigma)$ denotes a Gompertz distribution with the following distribution function
$$\text{Gomp}(x; \mu, \sigma) = 1 - exp\left(e^{-\frac{\mu}{\sigma}}(1 - e^{\frac{x}{\sigma}})\right). \tag{11}$$
We will call this the Gompertz scenario.
- $G_A = \text{Uni}(85)$, $G_B = \text{Uni}(20)$ and $G_C = \text{Uni}(20)$, where $\text{Uni}(\mu)$ denotes a discrete uniform distribution with support $[0, \mu]$. This will be the uniform scenario.

In the following, when we write $\text{Poi}_{l(h)}$, we refer to the B-RUP estimator obtained with a Poisson prior and low (high) strength of belief. For the Gompertz and Uniform priors, we will use $\text{Gomp}_{l(h)}$ and $\text{Uni}_{l(h)}$, respectively.

For the Poisson scenario, Figure 5 shows the marginal distributions for both the low and high strengths of belief. We observe that the marginal distribution for the first annuitants (males) is nicely recovered. However, the differences are clearly bigger[12] for the marginal of Y–the second annuitant (females)–since censoring seems to affect more the right-tail.

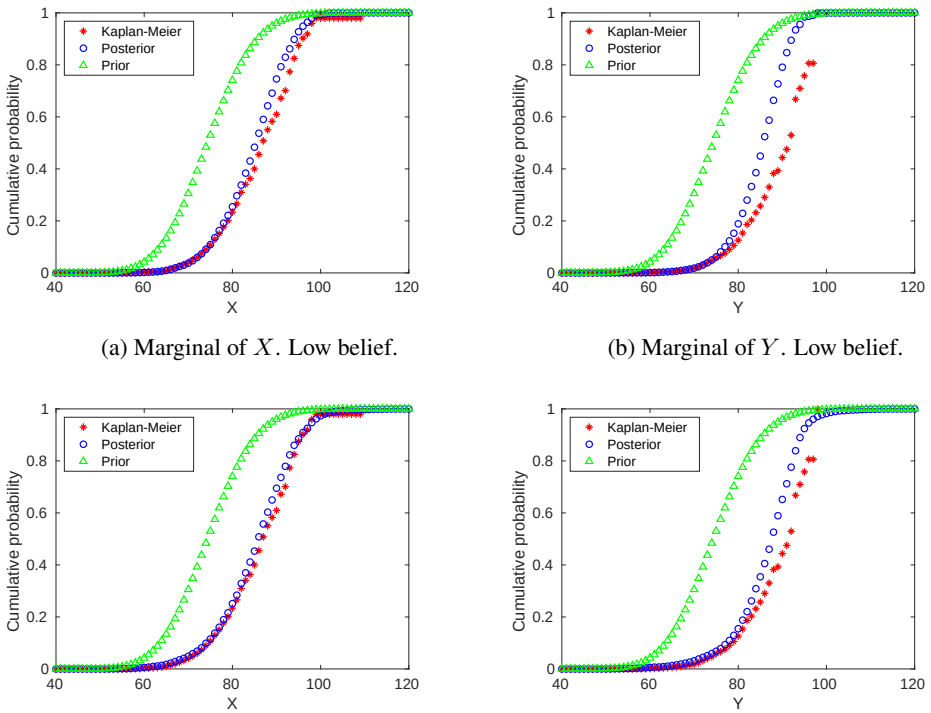
The lack of information about the right-tail of the distribution of female lifetimes is so large that the KM cumulative probability at the age of 97 is around 0.8, while for the age of 98 the cumulative distribution function (cdf) reaches its maximum value of 1. Taken literally, this would mean that there is a 20% probability for females of dying between 97 and 98 years, something we know is not true. This unrealistic behaviour could be due to the fact that, as per Table 5, female annuitants have a slightly longer life expectancy than males, while, from the initial ages at which they enter the contract, we see that on average [Luciano et al., 2008] females are three years younger than males, when the observation period starts. This combination of younger entry age plus higher life expectancy could be the reason why censoring affects more the distribution of females than that of males for this particular dataset.

Some additional numbers can give insight about the lack of data in the tails. Only 48 females have indeed survived until at least 90 years. Conversely, the number of males observed in the same range is 77. While this difference may not seem significant, notice that the life expectancy of males, according to the results in Table 5, is 3 years lower than that of women. Therefore one would expect to observe more women at high ages than men, whereas the opposite happens in our data. This could also be explained by the lower entry ages of females with respect to males.

The results obtained using the Gompertz scenario are in Figure 6, and they are very similar to the Poisson case. On the other hand, the results obtained with the Uniform scenario, shown in Figure 7, are significantly different, especially

[12] We observe that the differences are slightly smaller for the high strength of belief. This could be due to the fact that a strong a priori helps filling the gaps in the right tail of the marginals, which, because of censoring, are highly unrealistic.

for high strengths of belief–see Figures 7c and 7d–where the B-RUP and KM marginals have very little in common. Trivially, as one grows older the probability of demise increases, and therefore imposing a uniform distribution for the lifetime of an individual is highly unrealistic, so this mismatch was actually to be expected. In this sense, the B-RUP inherits the pros and the cons of the prior setting used, for sure under high strengths of belief. Naturally and conversely, for low strengths of belief, the B-RUP is able to capture both marginals, except for the areas with very few observations, so that the model is able to override a wrong prior given a sufficient amount observations.



(a) Marginal of X. Low belief.

(b) Marginal of Y. Low belief.

(c) Marginal of X. High belief.

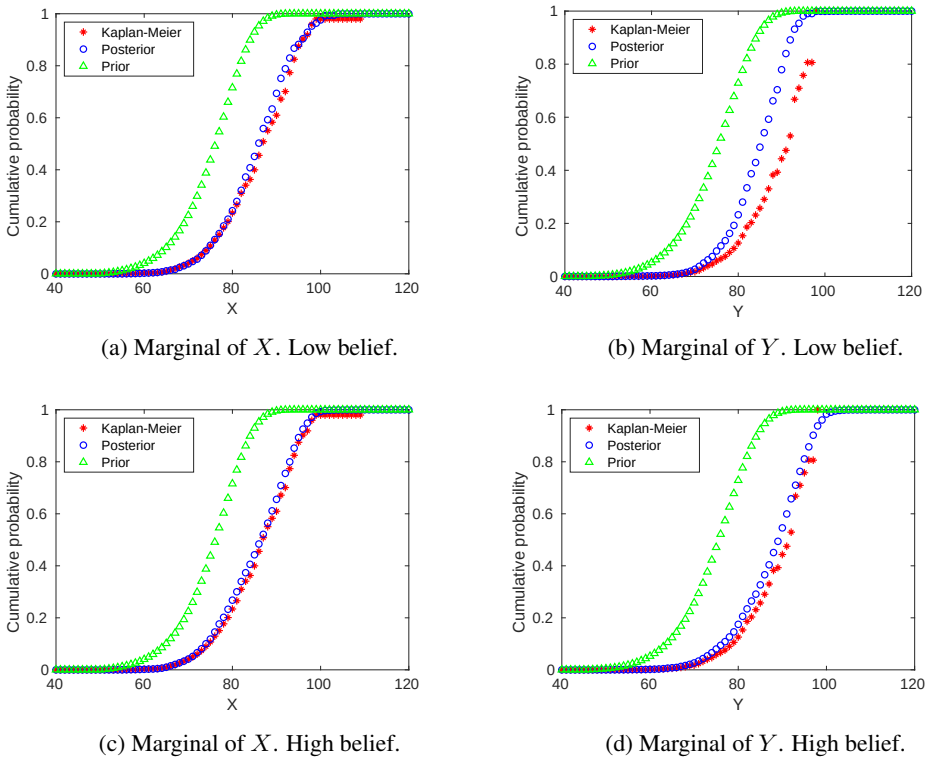
(d) Marginal of Y. High belief.

Figure 5: Comparison of the marginals of X and Y using a Poisson prior. In green we show the prior distribution, the Kaplan-Meier estimators in red and the B-RUP solution in blue. The upper row shows the results for the low belief scenario and the lower row the same for the high belief scenario. All distributions are conditioned on the minimum value of the uncensored observations. For this example those were $X_{min} = 51$ and $Y_{min} = 46$.

Table 5 presents the means and the variances of X and Y, and their correlation, for the raw uncensored data, the KM estimator (notice that in this case correlation is not obtained because independence is assumed), and the different B-RUPs. For completeness, we also provide all quantities when taking all observations as uncensored ("Whole"), as we did in the analytical example.

First notice that using only uncensored observations leads to an underestimation of the average lifetimes and to a likely overestimation of dependence. But this is no surprise, as it is known that ignoring censoring is highly misleading in survival analysis [Klein and Moeschberger, 2003]. Moreover, observe how, for low strengths of belief, the difference in the B-RUP results are mainly due to the choice of the initial sample for A, which in our case is generated from its prior distribution. Consistently with what we have already said, the uniform scenario with a high strength of belief generates the most radical results, including a correlation close to the one (0.8) we can estimate from the uncensored observations. All the other B-RUPs tend to suggest a less extreme correlation, in the vicinity of 0.5, in line with Frees et al. [1996] and Luciano et al. [2008]. In any case, a correlation of 0.5 should definitely not be ignored by assuming independence, as per the KM approach.

In Figures 8a and 8b we show the joint distributions obtained with the B-RUP model for the low and high belief scenarios, respectively, as well as the prior distribution for comparison purposes. Notice how, although their marginals are very similar–see Figure 5–the shape of the contours is considerably different around the areas of lower probability. Conversely the joints are more compatible around the inner contours, for the bulk of the distribution, sign that the strength of belief is not large enough to affect the areas where data are concentrated. In particular, for high strengths of



(a) Marginal of X. Low belief.

(b) Marginal of Y. Low belief.

(c) Marginal of X. High belief.

(d) Marginal of Y. High belief.

Figure 6: Comparison of the marginals of X and Y using a Gompertz prior. In green we show the prior distribution, the Kaplan-Meier estimators in red and the B-RUP solution in blue. The upper row shows the results for the low belief scenario and the lower row the same for the high belief scenario. All distributions are conditioned on the minimum value of the uncensored observations. For this example, $X_{min} = 51$ and $Y_{min} = 46$.

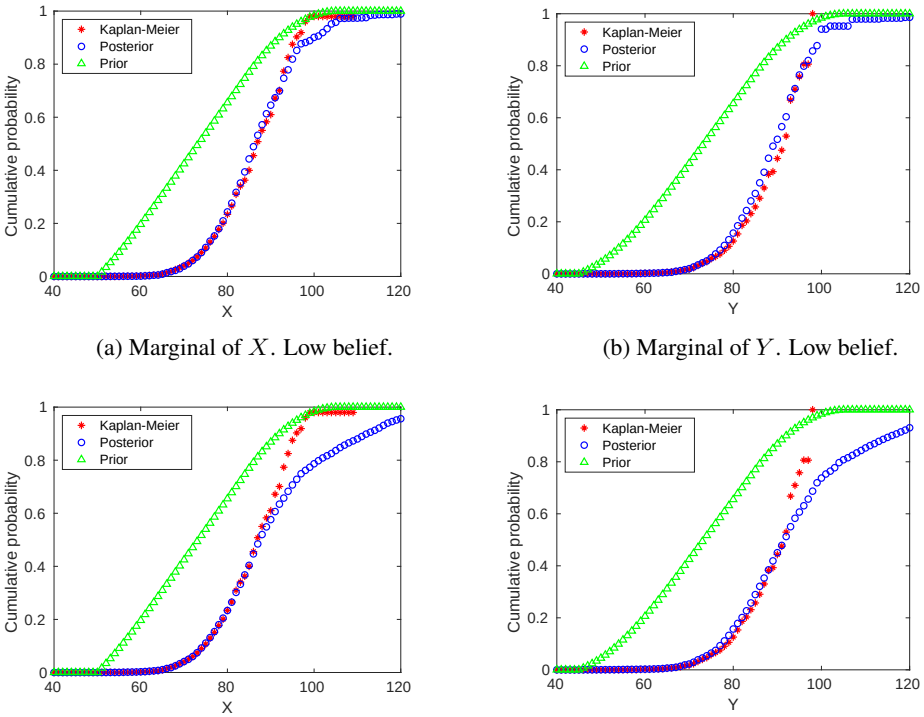
	Unc.	Whole	KM	Poi_l	Poi_h	Gomp_l	Gomp_h	Uni_l	Uni_h
Mean(X)	74.514	71.710	86.769	85.183	85.801	85.907	85.865	87.457	90.819
Mean(Y)	74.011	68.871	89.625	85.509	87.108	84.950	88.208	89.696	94.295
Var(X)	52.049	42.821	75.553	57.556	74.305	65.519	67.886	106.766	208.802
Var(Y)	61.707	54.118	56.316	37.034	54.947	45.047	64.267	91.114	217.951
Corr(X,Y)	0.820	0.776	–	0.477	0.421	0.475	0.414	0.611	0.810

Table 5: Comparison of the means, variances and correlation of X and Y using: the uncensored data, the whole dataset (wrongly assuming no censoring), the KM estimator and the B-RUP estimator for both low and high strength of belief, respectively.

belief the contours are somewhat smoother, a phenomenon which we already observed in the analytical example. This makes the distribution of Figure 8b easier to interpret[13].

A way to improve our previous results, obtaining a more tractable joint distribution, while maintaining the desired strength of belief, is to use a kernel bivariate density estimation. For example, we can employ a bivariate normal kernel, with bandwidth parameters consistent with Sylverman's rule of thumb [Sylverman, 1986]. The obtained distributions are presented in Figures 8c and 8d, for low and high strengths of belief, respectively. Note that the contours of the high belief kernel estimate are considerably more symmetric than those of the low belief kernel estimate, in accordance with the results of Figures 8a and 8b.

[13]Nevertheless, notice that this smoothing induced by the strength of belief has a secondary role, while the main purpose of a reliable prior is to embed behaviours that are not captured by the data. Moreover, whereas further increasing the strength of belief would surely result in smoother distributions, we could end up giving a larger weight to the prior distribution than originally intended.



(a) Marginal of X. Low belief.

(b) Marginal of Y. Low belief.

(c) Marginal of X. High belief.

(d) Marginal of Y. High belief.

Figure 7: Comparison of the marginals of X and Y using an Uniform prior. In green we show the prior distribution, the Kaplan-Meier estimators in red and the B-RUP solution in blue. The upper row shows the results for the low belief scenario and the lower row the same for the high belief scenario. All distributions are conditioned on the minimum value of the uncensored observations. Here $X_{min} = 51$ and $Y_{min} = 46$.

Notice that, by increasing the strength of belief, the areas with fewer observations are dominated by the prior, and this causes a decline in the correlation, which goes from 0.477 to 0.421, consistently with the findings in Table 5.

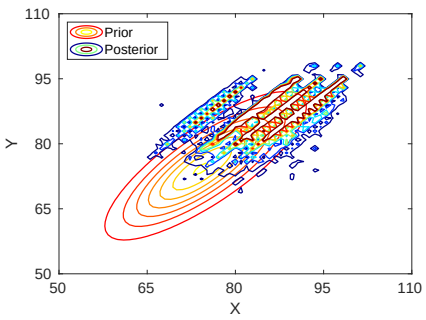
As observed in Carriere [2000] for this very same dataset–and more recently in Sanders and Melenberd [2016] for a dataset concerning the whole Dutch population–the positive dependence between the lifetimes of married couples increases with the ages of its members. In other words, people in their forties are expected to present a smaller positive dependence than a couple in their eighties. To check for this phenomenon, as in Sanders and Melenberd [2016] we compute the relative difference between the joint survival function and the product of the marginal survival distributions. The results are in Figures 9a and 9b for the low and high belief scenarios, respectively, and they are clearly in line with an increase of dependence with age. Moreover, from the same figures we can conclude that the relative difference is bigger under the low belief scenario, which would explain the slight increase in the overall correlation (Table 5).

Now that we have estimated the joint probability distribution, we can calculate the annuity ratio. We will still assume a fixed interest rate of 0.05 and show the annuity ratio for several initial ages of the annuitants. Next, when we want to compute the annuity ratio as a function of the interest rate, we will hypothesize that annuitants enter the contract at the same age to simplify the exposition, so that we can easily plot the results as a function of the interest rate only.

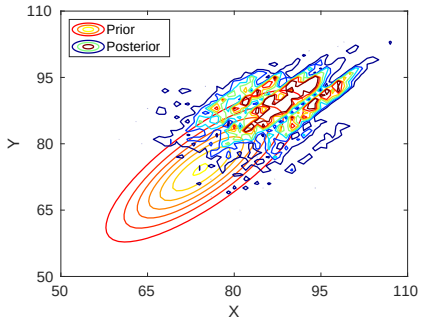
Figure 10 shows the results for the low and high strengths of belief, and the fixed interest rate. In line with expectations, even for the annuity ratio, the high belief scenario gives again a smoother surface.

Notice that, in both cases, the annuity ratio increases with the age difference between the annuitants. When one of the annuitants dies at a very old age, the probability for the surviving spouse to also die at an old age is larger under positive correlation than under independence. This effect seems to be bigger when the female is the surviving annuitant, probably because of the larger life expectancy of females with respect to males [Carriere, 2000].

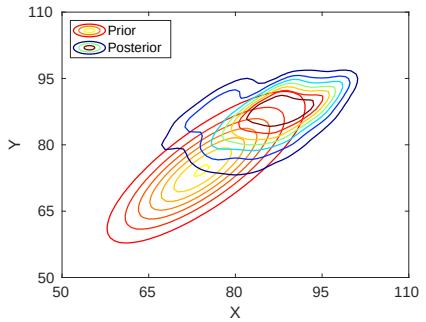
When both spouses are around the same age, the annuity ratio takes values below one, specially for very high entry ages. If one annuitant dies at a young age, the probability of the remaining annuitant to also die in the near future is higher when positive dependence is assumed, but in the long term the surviving annuitant will count as being independent from



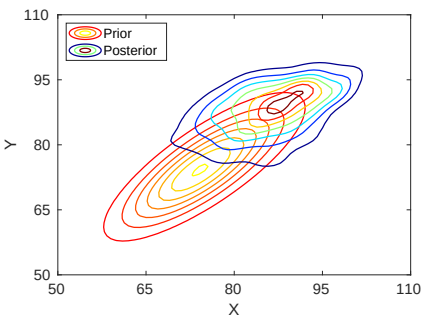
(a) Low strength of belief.



(b) High strength of belief.

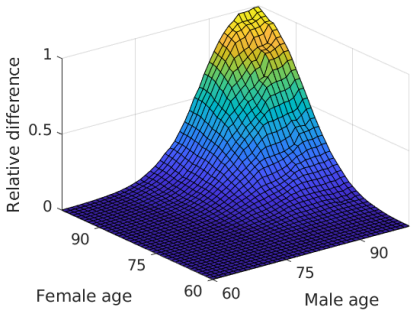


(c) Low strength of belief. Kernel estimate

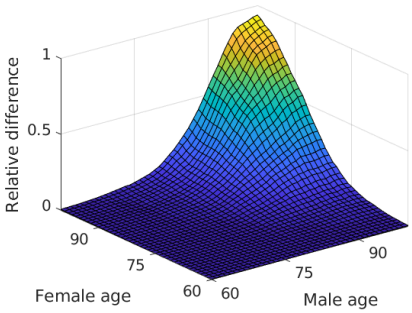


(d) High strength of belief. Kernel estimate.

Figure 8: Contour plot showing the joint distribution obtained via the B-RUP model for the low (a) and high (b) belief scenarios and their respective densities via bivariate kernel estimation ((c) and (d)) using a Poisson prior. All distributions are conditioned on the minimum value of the uncensored observations, i.e. $X_{min} = 51$ and $Y_{min} = 46$.



(a) Low strength of belief.

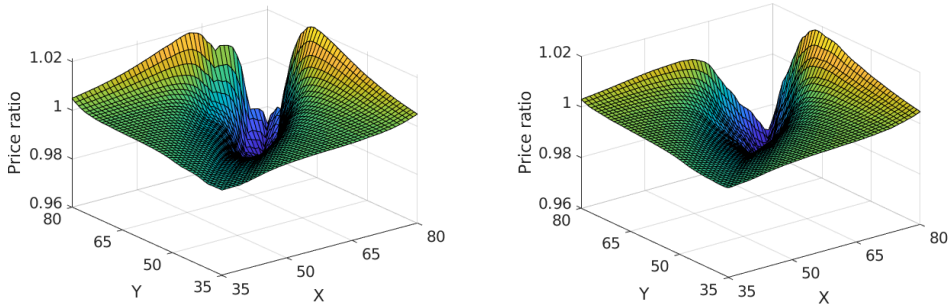


(b) High strength of belief.

Figure 9: Surface plot showing the relative difference between the joint survival distribution and the product of the marginal survival distributions (independence) for the low (a) and high (b) belief scenarios.

the deceased spouse, and the annuity ratio is closer to 1. Conversely, for high entry ages, the impact of the deceased spouse upon the surviving annuitant is more relevant in the near future than in the long term—according to the meaning of "high" entry ages, there is no long term survival to be expected in the first place—and the annuity ratio keeps on decreasing below the unit. All these findings are in line with the results obtained in Frees et al. [1996] for the same dataset, when using copula models with Gompertz and Weibull marginals.

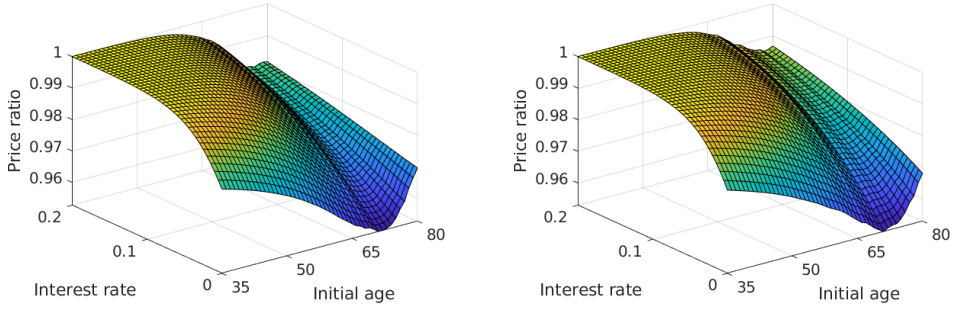
Also notice the similarity of these surfaces with those of Figure 3: qualitatively the behaviour is mostly the same. This is because in both cases we have positive dependence, and thus all the reasoning developed in this section also applies for the analytical example. Moreover, since the correlation levels are very similar (around 0.5), we obtain annuity ratios of comparable orders.



(a) Low strength of belief. (b) High strength of belief.

Figure 10: Ratio between the annuity values of the joint and independent survival models using the B-RUP estimators. The left plot (a) shows the dependency with the initial age of the annuitants an interest rate of 0.05. In the right plot we assume that the annuitants have the same age and plot the price ratio for several values of the interest rate

The results of the annuity ratio as a function of the interest rate can be found in Figure 11. Please notice how the overall shape is in accordance with what we already observed in Figure 4. For small interest rates and entry ages, the annuity ratio is below one, meaning that the annuity price with independent mortality is overestimated. As we increase the entry age, the annuity ratio decreases up to a minimum (0.9527 for and interest rate equal to 0 and an entry age of 71 years), from where it starts to increase again, also reaching values above the unit.



(a) Low strength of belief. (b) High strength of belief.

Figure 11: Ratio between the annuity values of the joint and independent survival models using the B-RUP with different strengths of belief.

$X \backslash Y$	40	50	60	70	80
40	0.993	0.996	0.999	1.001	1.007
50	0.996	0.989	0.992	1.000	1.012
60	0.999	0.993	0.979	0.988	1.019
70	1.001	1.001	0.990	0.966	1.009
80	1.007	1.013	1.021	1.015	0.977

Table 6: Annuity ratio obtained with the B-RUP model with low strength of belief parameters for several initial ages of the annuitants with an interest rate of 0.05. Assuming independence can definitely be a bad choice in pricing.

We would like to stress that, while annuity ratios of 1.02 or 0.98 may seem close enough to 1, and this could lead to the conclusion that it is ok to assume independence, the nominal value for an annuity contract is usually considerably large, and thus the final difference in the prices may not be negligible in monetary terms. In Table 6 we show some values of the annuity ratio for particular entry ages.

All the results above are in line with what Frees et al. [1996] and Luciano et al. [2008] found on the same data, using different approaches like copulas. While working nicely on data, copulas do not have the same flexibility of the B-RUP. In particular when a priori knowledge can be used to improve fitting, or to deal with specific characteristics of the data—like a severe lack of observations in the tails—the B-RUP clearly has an edge over the other commonly used approaches. And this without taking into consideration the model risk involved in every parametric choice behind the use of copulas [Hull, 2015, Mikosch, 2006], and parametric methodologies in general. This is why we believe that the nonparametric approach the B-RUP proposes, which combines a priori knowledge (when meaningful) and the ability of learning from the data, represents a viable and powerful alternative to the existing approaches to the modeling of coupled lifetimes.

6 Conclusions

In this paper we have proposed the bivariate reinforced urn process (B-RUP) as a way of modeling dependent mortality, to price joint and survivor annuities.

The main advantages of the B-RUP lie in its intuitive construction, in the possibility of using experts' judgements—also combining them with empirical evidence—in the ability of the model to learn and improve its performances over time, like in many machine learning approaches, but without "black boxes" [Knight, 2017], and in the successful treatment of right-censored observations, very common in annuity modeling.

In the absence of a credible a priori, which would give a strong competitive edge to its use, the B-RUP is nevertheless able to replicate the performances of other commonly used approaches (e.g. copulas as in Frees et al. [1996] or Luciano et al. [2008]), thus showing an interesting flexibility. Differently from other models, however, the B-RUP can be used on a continuous basis, as it automatically updates its parameters whenever new data become available, without the necessity of re-estimating the model entirely. This can be extremely useful in an online learning environment. Finally, being nonparametric, the B-RUP is less subject to model risk than copulas or similar approaches [Hull, 2015], especially if a clever use of priors can also account for data problems and historical bias.

In terms of performances, the model provides results in a very reasonable time, spanning from a few minutes to a couple of hours, depending on the size of the dataset and the amount of predictions required. In particular, for the Canadian data set considered and 10^4 iterations of the Gibbs sampler, the algorithm took approximately 1 minute of computing time in a C++ environment and an Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz processor.

Using artificial data anda well-known Canadian dataset of annuities [Frees et al., 1996], we have discussed the performances of the bivariate reinforced urn process, which appear definitely satisfactory. In analysing those performances, we have also discussed how a sufficient number of data can correct a wrong a priori, but also how strong priors may try to correct for historical bias and deal with tail events.

In the future, it would be extremely useful to go deeper into the study of credible and reliable a priori beliefs for annuity modeling, by gathering opinions and recommendations from experts. From a computational point of view, conversely, it could be meaningful to find ways of introducing parallelization in the simulations of the B-RUP. While not immediately relevant to the application described in this paper, such a possibility could dramatically expand the applicability and the performances of reinforced urn models in insurance and finance [Amerio et al., 2004, Cheng and Cirillo, 2018, 2019, Peluso et al., 2015].

Acknowledgements

This research has been financed by the European Union, under the H2020-EU.1.3.1. MSCA-ITN-2018 scheme, thanks to Grant 813261.

References

Emanuele Amerio, Pietro Muliere, and Piercesare Secchi. Reinforced Urn Processes for Modeling Credit Default Distributions. *International Journal of Theoretical and Applied Finance*, 07(04):407–423, 2004. URL `https://doi.org/10.1142/S0219024904002505`.

Davis Blackwell and James B. MacQueen. Ferguson distributions via pólya urn schemes. *The Annals of Statistics*, 1(2): 353–355, 1973.

J. R. Brown and J. M. Poterba. Joint life annuities and annuity demand by married couples. *Journal of Risk and Insurance*, 74(4):527–553, 2000.

Paolo Bulla. *Application of Reinforced Urn Processes to Survival Analysis*. PhD thesis, Bocconi University, 2005.

Paolo Bulla, Pietro Muliere, and Steven Walker. Bayesian Nonparametric Estimation of a Bivariate Survival Function. *Statistica Sinica*, 17(2):427–444, 2007.

Jacques F. Carriere. Bivariate survival models for coupled lives. *Scandinavian Actuarial Journal*, 1:17–32, 2000.

Dan Cheng and Pasquale Cirillo. A Reinforced Urn Process Modeling of Recovery Rates and Recovery Times. *Journal of Banking & Finance*, 96:1–17, 2018. ISSN 0378-4266. URL `http://www.sciencedirect.com/science/article/pii/S037842661830178X`.

Dan Cheng and Pasquale Cirillo. A Urn-Based Nonparametric Modeling of the Dependence Between PD and LGD with an Application to Mortgages. 2019.

Pasquale Cirillo, Jürg Hüsler, and Pietro Muliere. A Nonparametric Urn-based Approach to Interacting Failing Systems with an Application to Credit Risk Modeling. *International Journal of Theoretical and Applied Finance*, 41:1–18, 2010.

Pasquale Cirillo, Jürg Hüsler, and Pietro Muliere. Alarm Systems and Catastrophes from a Diverse Point of View. *Methodology and Computing in Applied Probability*, 15(4):821–839, 2013.

D. R. Cox and D. Oakes. *Analysis of survival data*. Chapman & Hall, 1984.

Dorota M. Dabrowska. Kaplan-meier estimate on the plane. *The Annals of Statistics*, 16(4):1475–1489, 1988.

B. de Finetti. *Theory of Probability: A critical introductory treatment*. Wiley Series in Probability and Statistics. Wiley, 2017. ISBN 9781119286370.

James Derbyshire. The Siren Call of Probability: Dangers Associated with Using Probability for Consideration of the Future. *Futures*, 88:43–54, 2017. URL `http://www.sciencedirect.com/science/article/pii/S0016328716303056`.

Kjell Doksum. Tailfree and neutral random probabilities and their posterior distributions. *The Annals of Probability*, 2 (2):183–201, 1974.

R. C. Elandt-Johnson and N. L. Johnson. *Survival models and data analysis*. John Wiley, 1980.

Thomas S. Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

Sandra Fortini and Sonia Petrone. Hierarchical Reinforced Urn Processes. *Statistics & Probability Letters*, 82 (8):1521–1529, 2012. ISSN 0167-7152. URL `http://www.sciencedirect.com/science/article/pii/S0167715212001551`.

Edward W. Frees, Jacques Carriere, and Emiliano Valdez. Annuity valuation with dependent mortality. *The Journal of Risk and Insurance*, 63(2):229–261, 1996.

Phillip Good. *Permutation, Parametric and Bootstrap Tests of Hypotheses*. Springer New York, 1994.

Svetlana Gribkova and Olivier Lopez. Non-parametric copula estimation under bivariate censoring. *Scandinavian Journal of Statistics*, 42(4):925–946, 2015.

John C. Hull. *Risk Management and Financial Institutions*. Wiley, New York, 4th edition, 2015.

C. Jagger and C. J. Sutton. Death after Marital Bereavement—Is the Risk Increased? *Statistics in Medicine*, 10(3): 395–404, 1991.

E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

John P. Klein and Melvin L. Moeschberger. *Survival Analysis Techniques for Censored and Truncated Data*. Second edition, 2003.

Will Knight. The Dark Secret at the Heart of AI. *Technology Review*, 120:54–63, 2017.

D. Y. Lin. A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, 80:573–581, 1993.

Olivier Lopez. A generalization of kaplan-meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications to model-checking for survival copula models. *Insurance: Mathematics and Economics*, 51(3):505–516, 2012.

Elisa Luciano, Jaap Spreeuw, and Elena Vigna. Modelling stochastic mortality for dependent lives. *Insurance: Mathematics and Economics*, 43:234–244, 2008.

Hosam Mahmoud. *Polya urn models*. Chapman & Hall/CRC, Boca Raton, 2008.

Thomas Mikosch. Copulas: Tales and facts. *Extremes*, (9):3–20, 2006.

O. Mitchell, J. M. Poterba, M.J. Warshawsky, and J.R. Brown. New evidence on the money's worth of individual annuities. *The American Economic Review*, 89(5):1299–1318, 1999.

Pietro Muliere, Piercesare Secchi, and Stephen G Walker. Urn Schemes and Reinforced Random Walks. *Stochastic Processes and their Applications*, 88(1):59–78, 2000. ISSN 03044149.

Pietro Muliere, Piercesare Secchi, and Stephen G Walker. Reinforced Random Processes in Continuous Time. *Stochastic Processes and their Applications*, 104(1):117–130, 2003.

Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge MA, 2012. ISBN 0262018020, 9780262018029.

Stefano Peluso, Antonietta Mira, and Pietro Muliere. Reinforced Urn Processes for Credit Risk Models. *Journal of Econometrics*, 184(1):1–12, 2015.

R. C. Pruitt. On negative mass assigned by the bivariate kaplan-meier estimator. *Annals of Statistics*, 19:443–453, 1991.

Lisanne Sanders and Bertrand Melenberd. Estimating the joint survival probabilities of married individuals. *Insurance: Mathematics and Economics*, 67:88–106, 2016.

George Lennox Sharman Shackle. *Uncertainty in Economics and Other Reflections*. Cambridge University Press, Cambridge, 1955.

Pao-Sheng Shen and Ya-Fang Yan. Nonparametric estimation of the bivariate survival function with left-truncated and right-censored data. *Journal of Statistical Planning and Inference*, 138(12):4041–4054, 2008.

Eytan Sheshinski. *The Economic Theory of Annuities*. Number 8536 in Economics Books. Princeton University Press, 2007.

B. W. Sylverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, 1986.

Nassim Nicholas Taleb. *The Black Swan: The Impact of the Highly Improbable*. Random House, 2007.

Stephen Walker and Pietro Muliere. Beta-Stacy Processes and a Generalization of the Pólya-Urn Scheme. *The Annals of Statistics*, 25(4):1762–1780, 1997.

W. Wang and M. T. Wells. Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika*, 84:863–880, 1997.

H.E. Winklevoss. *Pension Mathematics with Numerical Illustrations*. Pension Research Council, Wharton School, University of Pennsylvania, 1977.

H. Youn and A. Shemaykin. Statistical aspects of joint life insurance pricing. *1999 Proceedings of the Business and Statistics Section of the American Statistical Association*, pages 34–38, 1999.

H. Youn and A. Shemaykin. Pricing practices for joint last survivor insurance. *Actuarial Research Clearing House, 2001.1*, 2001.

Heekyung Youn, Arkady Shemyakin, and Edwin Herman. A Re-Examination of the Joint Mortality Functions . *North American Actuarial Journal*, 6(1):166–170, 2002.