Modelling stochastic mortality for dependent lives

Elisa Luciano[^], Jaap Spreeuw[+] and Elena Vigna[**]
[^] University of Turin, Icer and Collegio Carlo Alberto, Turin
+ Cass Business School, London ** University of Turin

March 26, 2008

Abstract

Stochastic mortality, i.e. modelling death arrival via a jump process with stochastic intensity, is gaining increasing reputation as a way to represent mortality risk. This paper is a first attempt to model the mortality risk of couples of individuals, according to the stochastic intensity approach. Dependence between the survival times of the members of a couple is captured by an Archimedean copula.
We also provide a methodology for fitting the joint survival function by working separately on the (analytical) marginals and on the (analytical) copula. First, we provide a sample-based calibration for the intensity, using a time-homogeneous, non mean-reverting, affine process: this gives the marginal survival functions. Then we calibrate and select the best fit copula according to the Wang and Wells (2000b) methodology for censored data. By coupling the calibrated marginals with the best fit copula, we obtain a joint survival function which incorporates the stochastic nature of mortality improvements.

We apply the methodology to a well known insurance data set, using a sample generation. The best fit copula turns out to be one listed in Nelsen (2006), which implies not only positive dependence, but dependence increasing with age.

Acknowledgement. We wish to acknowledge the Society of Actuaries, through the courtesy of Edward (Jed) Frees and Emiliano Valdez, for providing the data used in this paper. We acknowledge support from the European Science Foundation (ESF) through the 'Advanced Mathematical Methods for Finance' project. Comments from the participants to the 2006 IME conference, 2007 AFIR conference, 2008 CREST-AXA workshop as well as from an anonymous referee are gratefully acknowledged. All remaining errors are ours.

1 Introduction

Longevity risk, that is the tendency of individuals to live longer and longer, has been increasingly attracting the attention of the actuarial literature. More generally, mortality risk, that is the occurrence of unexpected changes in survivorship, is a well accepted phenomenon.

One way to incorporate improvements in survivorship, especially at old ages, is to introduce the so called stochastic mortality. This boils down to describing death arrival using a doubly stochastic or Cox process, i.e. in interpreting death arrival as the first jump time of a Poisson-like process, the intensity of which, contrary to that of the standard Poisson, is a stochastic process. Then, a priori two sources of uncertainty impinge on each individual: a common one, represented by the intensity, and an idiosyncratic one, represented by the actual jump time, for a given intensity. Mortality risk is captured by the behaviour of the common risk factor, the intensity. The term "common" extends here to a whole generation within a gender. The approach leads to generation-dependent survival functions and therefore allows for survivorship improvements. It has been proposed by Milevsky and Promislow (2001) and developed by Dahl (2004), Biffis (2005), Denuit and Devolder (2005), Luciano and Vigna (2005), Cairns et al. (2006) and Schrager (2006).

Up to now, no attempt has been made to model the survivorship of couples of individuals stochastically, in the sense above specified. Bivariate mortality has been described in the nonstochastic case only. The copula approach has become the most popular methodology in this context. Working on a data set of mortality of couples from a large Canadian insurer, both Frees et al. (1996) and Carriere (2000) present fully parametric models using maximum likelihood, where the marginal distribution functions (Frees et al.) or survival functions (Carriere) are assumed to be of Gompertz type. Frees et al. (1996) use the Frank copula. Carriere (2000), on the other hand, discusses several copulas with more than one parameter (Frank, Clayton, Normal, Linear Mixing, Correlated Frailty). Using the same data set in an attempt to address the issue of different types of dependence, Youn and Shemyakin (1999, 2001) refine Frees et al.'s method by classifying individuals according to the age difference between the female and the male members of each couple. Shemyakin and Youn (2001) adopt a Bayesian methodology as an alternative. All three papers use the Gumbel-Hougaard copula.

With the exception of Carriere (2000), the existing literature based on the same sample does not perform a best fit copula choice. However, since different copulas entail different characteristics regarding the type of dependence and aging properties, as shown in Spreeuw (2006), the choice of an appropriate copula is essential. Ideally, one should use the best copula among all possible ones. Practically, the process of choosing a copula must be restricted to a finite number of them.

For complete data, Genest and Rivest (1993) have developed a procedure where the process of choosing a copula in the Archimedean class is independent

from the specification of the margins. Denuit et al. (2001) have managed to get hold of complete survival data of coupled lives by visiting cemeteries. Applying the method developed by Genest and Rivest (1993), they have established a weak correlation of lifetimes between males and females, and identified several plausible candidates for the copula.

Genest and Rivest method cannot be used if data are right-censored. This is the case for the data set used in Frees et al. (1996), Carriere (2000), Shemyakin and Youn (2001) and Youn and Shemyakin (1999, 2001). Wang and Wells (2000b) have extended Genest and Rivest method to bivariate right-censored data. This technique has not been applied to coupled lives yet.

In this paper, we couple the stochastic mortality univariate approach with the best fit copula technique of Wang and Wells in order to introduce stochastic mortality on two dependent lives. To our knowledge, this is the first attempt to model the mortality of couples in a stochastic way. As a consequence, the main contributions in this paper are twofold:

- We incorporate the random evolution of mortality intensity and generation effects in the mortality of couples.
- We perform the best fit selection so as to infer from the data how association evolves when the couple ages. We do this by choosing candidate copulas which incorporate different behaviours in respect to aging.

We apply our modelling and calibration procedure to a generation-based sample of joint survival data, belonging to the Canadian data set used by Frees et al. (1996), Carriere (2000), Shemyakin and Youn (2001) and Youn and Shemyakin (1999, 2001). We select and calibrate the copula function and check its robustness through pseudo-maximum likelihood.

The outline of the paper is as follows: in Section 2 we review the stochastic mortality approach at the univariate level, and the particular marginal model that we have selected. In Section 3 we justify the copula class that we have adopted - the Archimedean one - and list the different copulas which serve as candidates for our choice. In Section 4 we describe the copula calibration and selection methodology, as proposed by Wang and Wells (2000b). In Section 5 we apply the theoretical framework and the calibration method to the data sample. In Section 6 the specific best fit copula obtained, together with the analytical margins, enables us to present the analytical joint survival function and to discuss the corresponding measures of time-dependent association. Section 7 concludes.

2 Marginal stochastic mortality

2.1 Theoretical framework

It has been widely accepted that mortality has improved over time, and different generations have different mortality patterns: according to the standard

terminology, we will call this phenomenon mortality risk. Evidence of mortality risk is provided by Cairns et al. (2006), who present also a very detailed discussion of the different existing approaches for modelling it. Essentially, most of these approaches rely on a continuous-time stochastic process for the instantaneous mortality intensity, which can be interpreted as a stochastic force of mortality. These works exploit – on one hand – the similarity between time to default and remaining lifetime and – on the other – the similarity between short-term interest rate and force of mortality. Milevsky and Promislow (2001) have used a stochastic force of mortality, the expectation of which at any future date – under an appropriate choice of parameters – has a Gompertz specification. Dahl (2004), Biffis (2005), Denuit and Devolder (2005), Luciano and Vigna (2005) and Schrager (2006) have applied the mathematical tools of Cox processes used in the credit risk literature when modelling the time to default (see, for instance, Artzner and Delbaen (1992), Lando (1998) and Duffie and Singleton (1999)). Since the approach is in reduced form, one is allowed to parallel the financial approach even though the factors underlying the death of an individual – essentially biological – and the factors driving the default of a firm – essentially economical – are different. In this setting, the remaining lifetime of an individual, τ, is a doubly stochastic stopping time with intensity μ, which is itself a stochastic process. Intuitively, this means that conditionally on μ, the jump process driving mortality is Poisson. For a thorough treatment of the mathematical framework, we refer the reader to Brémaud (1981). It is possible to show that if $\tau \geq t$ is a stopping time doubly stochastic with intensity μ, then:

$$P(\tau > s|\mathcal{G}_t) = E\left[e^{-\int_t^s \mu(u)du}|\mathcal{G}_t\right] \tag{1}$$

where $\mathcal{G}_t$ describes the information at time t. The parallel with the interest-rate literature consists also in the similarity between the survival probability given by (1) and the price at current time t of a zero-coupon bond with maturity at $s > t$ when the short-term interest rate is given by the process μ. All the Mathematical Finance literature about interest rate models can be retrieved in this setting.

It is convenient to specify the stochastic intensity μ as an affine process in $\mathbb{R}$, the dynamics of which are given by the SDE:

$$d\mu(t) = f(\mu(t))dt + g(\mu(t))d\tilde{W}(t) + dJ(t) \tag{2}$$

where $\tilde{W}$ is an n-dimensional Brownian motion, J is a pure jump process and where the drift $f(\mu(t))$, the covariance matrix $g(\mu(t))g(\mu(t))'$ and the jump measure associated with J have affine dependence on $\mu(t)$. Interested readers can find a thorough treatment of affine processes in Duffie et al. (2003). The convenience of adopting an affine process in modelling the intensity lies in the fact that, under technical conditions (see Duffie and Singleton (2003)), it yields:

$$E\left[e^{\int_t^T -\mu(u)du}|\mathcal{G}_t\right] = e^{\alpha(T-t)+\beta(T-t)\mu(t)}, \tag{3}$$

where the coefficients $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy generalized Riccati ODEs. The latter can be solved at least numerically and in some cases analytically. Therefore,

the problem of calculating the expectation (1), which in general is very hard if not impossible to solve, becomes easier to solve whenever affine processes for $\mu(t)$ are employed.

2.2 Selection of the intensity and calibration

In the existing actuarial literature, different classes of affine processes have been explored for the mortality intensity μ. For example, Dahl (2004) selects an extended Cox-Ingersoll-Ross (CIR) process, i.e. a time-inhomogeneous process reverting to a deterministic function of time. Biffis (2005) chooses two different specifications for the intensity process. In the first one, the intensity is given by a deterministic function of time, plus a mean reverting jump diffusion process. In the second one, which is a two factor model, the intensity is a CIR-like process, mean-reverting to another process. Schrager (2006) proposes an M-factor affine stochastic intensity, which includes the sum of mean-reverting factors. Denuit and Devolder (2005) introduce various stochastic models for the force of mortality, which are continuous-time versions of the Lee Carter model and include a mean reversion effect to a limit table. Luciano and Vigna (2005) explore time-homogeneous, non mean-reverting affine processes the deterministic part of which increases exponentially. They consider a few natural extensions of the Gompertz model for the force of mortality, which differ from each other only in their stochastic part. In particular, they propose three models. The first one is an Ornstein-Uhlenbeck Gaussian process, the second one is an Ornstein-Uhlenbeck process with jumps, and the third one is a Feller type model. For each model, they assume different intensity processes for different generations. This allows them to model survival improvements over generations.

Among the one-factor models, Biffis (2005) fits his mean-reverting time-inhomogeneous intensity to Italian mortality tables, while Luciano and Vigna (2005) calibrate their time-homogeneous simpler processes to the Human Mortality database for the UK population. They show that time-homogeneous non mean-reverting affine processes describe appropriately the stochastic force of mortality. Indeed, both the fit and the predictive power of these processes are satisfactory, in spite of their analytical simplicity. A relevant result is that the addition of a stochastic part to the basic Gompertz model is needed. Indeed, they show – by means of simulations – that the volatility of the intensity process is directly transferred to the variability of the simulated number of deaths in a population. Moreover, for different generations, different estimates of parameters, and therefore different processes, are obtained. These two results confirm that both the diffusive part and the generation effects should not be ignored.

The results obtained in Luciano and Vigna (2005) justify the choice made in this paper of an affine, time-homogeneous intensity process, without mean reversion. In particular, we use a non Gaussian Feller model, since in this case the intensity can never become negative. The Feller intensity, for the cohort or generation born x years ago, follows the equation

$$d\mu_x(s) = a_x\mu_x(s)ds + \sigma_x\sqrt{\mu_x(s)}dW_x(s),$$

where $a_x > 0$ and $\sigma_x \geq 0$. The corresponding survival probability[1] is given by (1), i.e.

$$S_x(t) = \mathbb{E}\left[e^{\int_0^t -\mu_x(s)ds}\right] = e^{\alpha_x(t)+\beta_x(t)\mu_x(0)}, \tag{4}$$

where, omitting the dependence on x for simplicity

$$\begin{cases} \alpha(t) = 0 \\ \beta(t) = \frac{1-e^{bt}}{c+de^{bt}} \end{cases},$$

with

$$\begin{cases} b = -\sqrt{a^2+2\sigma^2} \\ c = \frac{b+a}{2} \\ d = \frac{b-a}{2} \end{cases}.$$

The parameters a and σ can be obtained either from mortality tables or, as we do below, from sample censored data. In both cases they can be calibrated by minimizing the mean square error between the theoretical and actual probabilities[2]: in the mortality table case the actual probabilities are the table ones, while in the sample case they are the empirical ones, as obtained, for instance, by the classical Kaplan-Meier procedure for censored data.

3 Modelling bivariate survival functions with copulas

3.1 The copula representation

Suppose that the lives (x) and (y), belonging respectively to the gender m (males) and f (females), with ages x and y at the start of the observation, have remaining lifetimes T_x^m and T_y^f, both with continuous distributions. We denote the marginal survival functions by S_x^m and S_y^f, so that, for all $s, t \geq 0$, $S_x^m(s) = \Pr[T_x^m > s]$ and $S_y^f(t) = \Pr\left[T_y^f > t\right]$. By Sklar's theorem, there exists a copula, denoted by C_{xy}, such that for all $(s,t) \in \mathbb{R}^+ \times \mathbb{R}^+$ the joint survival function of x and y, denoted by S_{xy}, can be represented in terms of the marginal ones:

$$S_{xy}(s,t) = C_{xy}(S_x^m(s), S_y^f(t)). \tag{5}$$

This representation is unique over the range of the margins. Note that the suffix xy for both the copula function and the joint survival function reflects the dependence of both these functions on the generation and the starting ages

[1] These probabilities are decreasing in age t if and only if

$$e^{bt}(\sigma^2 + 2d^2) > \sigma^2 - 2dc$$

A sufficient condition for this is that $\sigma^2 - 2dc < 0$.

[2] In Luciano and Vigna (2005) the authors perform the calibration both with the maximum likelihood (ML) and mean square error (MSE) procedure. Since they find equivalence between the two methodologies over a number of different generations, here we have chosen the MSE approach.

x and y. Indeed, a fundamental issue for copula models applying to coupled lives concerns the time of coupling. Unlike twins, members of a couple do not have common genes and usually do not grow up in the same house. It does not seem too unrealistic to assume that the remaining lifetimes of coupled lives are independent until they meet each other for the first time. This is one of the reasons why in our setup the copula will not take effect at birth, as in Frees et al. (1996), but only later in life. Ideally we would couple lives at the time two people start living together permanently. Unfortunately, couples in our data set only start getting observed after this event. This is the left-truncation issue which we have had to cope with. Carriere (2000) couples lives as soon as they start getting observed. Carriere argues that in this way, unlike Frees et al. (1996), the marginal survival probabilities of either life at the start of the observation remain preserved, in the sense that they do not depend on the survival status of the other life. For e.g. males, this would mean

$$S_x(s) = \frac{S_0(x+s)}{S_0(x)}.$$

Carriere also states that the issue that the remaining lifetime of the living person depends on that of the yet unborn partner is avoided. We share Carriere's view in this respect.

3.2 Candidate copulas

Using the Wang and Wells procedure implies that all copulas which we are going to consider belong to the Archimedean class. Archimedean copulas may be constructed using a function $\phi : I \to \mathbb{R}^+$, continuous, decreasing, convex and such that $\phi(1) = 0$. Such a function ϕ is called a generator. It is called a strict generator whenever $\phi(0) = +\infty$. Having defined the pseudo-inverse of ϕ, ϕ^{-1}, in such a way that, by composition with the generator, it gives the identity:

$$\phi^{-1}\left(\phi\left(v\right)\right) = v.$$

An Archimedean copula C^A is generated as follows:

$$C^A(v,z) = \phi^{-1}\left(\phi(v) + \phi(z)\right). \tag{6}$$

Archimedean copulas have been widely used, due to their mathematical tractability. The Archimedean class is rich, so allowing for Archimedean copulas does not seem to be very restrictive. We refer the reader to the book by Nelsen (2006) for a review of Archimedean copulas definition and properties, and to Cherubini et al. (2004) for their applications.

In the Archimedean class in particular we will take into consideration the copulas in Table 1.

We have selected these families following the results in Spreeuw (2006), who studied the type of time-dependent association between lives implied by many Archimedean copulas.

No.	**Name**	**Generator** $\phi(t)$	$C(u,v)$
1	Clayton	$t^{-\theta}-1$	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1}{\theta}}$
2	Gumbel-Hougaard	$(-\ln t)^{\theta}$	$\exp\left[-\left((-\ln u)^{\theta}+(-\ln v)^{\theta}\right)^{\frac{1}{\theta}}\right]$
3	Frank	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$-\frac{1}{\theta}\ln\left[1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right]$
4	4.2.20 Nelsen	$\exp\left[t^{-\theta}\right]-e$	$\left[\ln\left(\exp\left(u^{-\theta}\right)+\exp\left(v^{-\theta}\right)-e\right)\right]^{-\frac{1}{\theta}}$
5	Special	$\frac{1}{t^{\theta}}-t^{\theta}$	$2^{-\frac{1}{\theta}}\left(-W+\sqrt{4+W^2}\right)$; $W=\phi(u)+\phi(u)$

Table 1: Archimedean copula families

Three measures of time-dependent association between T_x^m and T_y^f have been introduced in the literature. They all stem from Anderson et al. (1992). We will apply them in Section 6.

The first measure is the rescaled conditional probability, denoted by $\psi_1(s,t)$:

$$\psi_1(s,t)=\frac{S_{xy}(s,t)}{S_x^m(s)\,S_y^f(t)}=\frac{\Pr\left[T_x^m>s\left|T_y^f>t\right.\right]}{\Pr\left[T_x^m>s\right]}=\frac{\Pr\left[T_y^f>t\left|T_x^m>s\right.\right]}{\Pr\left[T_y^f>t\right]}. \quad (7)$$

The second measure is the conditional expected residual lifetime of (x) and (y), which we will specify as $\psi_{2x}(s,t)$ and $\psi_{2y}(s,t)$, respectively

$$\begin{aligned}\psi_{2x}(s,t) &= \frac{E\left[T_x^m-s\left|T_x^m>s,T_y^f>t\right.\right]}{E\left[T_x^m-s\left|T_x^m>s\right.\right]}\\ \psi_{2y}(s,t) &= \frac{E\left[T_y^f-t\left|T_x^m>s,T_y^f>t\right.\right]}{E\left[T_y^f-t\left|T_y^f>t\right.\right]}.\end{aligned} \quad (8)$$

In this paper we will concentrate on the behaviour of the functions $\psi_{2x}(0,t)$ and $\psi_{2y}(s,0)$.

The third measure is the cross-ratio function $CR(S_{xy}(s,t))$, defined in Clayton (1978) and studied by Oakes (1989).

$$CR(S_{xy}(s,t))=\frac{S_{xy}(s,t)\,\frac{d}{ds}\frac{d}{dt}S_{xy}(s,t)}{\frac{d}{ds}S_{xy}(s,t)\,\frac{d}{dt}S_{xy}(s,t)}.$$

It specifies the relative increase of the force of mortality of the survivor (immediately) upon death of the partner. If $CR(S_{xy}(u,u))$ increases (decreases) as a

function of u, this means that members of a couple become more (less) dependent on each other as they age. Spreeuw (2006) has shown that for Archimedean copulas and $u = s = t$, the cross ratio definition comes down to an expression in terms of the inverse of the generator:

$$CR\left(S_{xy}\left(u,u\right)\right) = \left(\frac{\phi^{-1}\left(v\right)\left(\phi^{-1}\right)''\left(v\right)}{\left(\left(\phi^{-1}\right)'\left(v\right)\right)^{2}} \right)_{v=\phi(S_{xy}(u,u))} . \tag{9}$$

The candidate copulas have been selected in regard to association, as measured by the cross-ratio function. The first copula in Table 1, Clayton, bears the special property of the association remaining constant over time. Copulas 2 (Gumbel-Hougaard) and 3 (Frank) both present association decreasing over time. Copula families 4 and 5 are due to Nelsen (2006). Family 4 can be identified as "Family 4.2.20" in Chapter 4 of Nelsen (2006) and will henceforth be referred to as the "4.2.20 Nelsen copula". Copula 5, which is also due to Chapter 4 of Nelsen (2006), will be labelled as the "Special copula". It is the interior power family $\phi(t^{\alpha})$ associated with $\phi(t) = \frac{1}{t} - t$, which is one of the generators listed in Nelsen (2006). It was studied in Spreeuw (2006). Copulas 4 and 5, unlike the first three copulas, have association increasing over time, which is what one would expect for coupled lives.

4 Copula estimate and best fit choice

In this section we describe the procedure we will adopt in order to select and calibrate an Archimedean copula under double right censoring.

4.1 The distribution function of the Archimedean copula

For a couple with lives aged x and y at the start of observation, let $Z_{xy} = S_{xy}\left(T_x^m, T_y^f\right)$. Define K_{xy} as the distribution function of Z_{xy}. Note that $Z_{xy} = C_{xy}\left(U,V\right)$ where (U,V) is a random couple with unit uniform margins.

Genest and Rivest (1993) have shown that, for Archimedean copulas with generator ϕ, the distribution function K_{xy} is given by

$$K_{xy}\left(z\right) = z - \lambda_{xy}\left(z\right) \tag{10}$$

where

$$\lambda_{xy}\left(z\right) = \frac{\phi_{xy}\left(z\right)}{\phi_{xy}^{'}\left(z\right)}, \quad 0 < z \leq 1. \tag{11}$$

and $\phi_{xy}^{'}$ is the generator derivative. The function K_{xy} is to be estimated from the data. We will make a distinction between complete data, such as in Denuit et al. (2001), and censored data, such as in the current paper.

4.1.1 General principle without censoring

Genest and Rivest (1993) have shown that, for complete data of size n, K_{xy} can be estimated using its empirical counterpart, $\widehat{K}_{n(xy)}$, defined as

$$\widehat{K}_{n(xy)}(z) = \frac{1}{n}\#\{i \,|z_i \leq z\} \text{ where } z_i = \frac{1}{n-1}\#\{(s_j,t_j)\,|s_j < s_i, t_j < t_i\},$$

where the symbol # indicates the cardinality of a set and $\{(s_i,t_i), i = 1,...,n\}$ are the observed data.

4.1.2 Wang-Wells empirical version of the generator in the presence of censored data

Wang and Wells (2000b) have proposed a modified estimator of K for censored data. K can be written as

$$K_{xy}(v) = 1 - \Pr\left[S_{xy}\left(T_x^m, T_y^f\right) > v\right] = 1 - \mathbb{E}\left[\mathbb{I}_{\left\{S_{xy}\left(T_x^m,T_y^f\right)>v\right\}}\right],$$

Assume we have n observations $\left\{\left(\widetilde{s}_i,\widetilde{t}_i\right), i = 1,..,n\right\}$ Let $\widetilde{s}_1 \leq ... \leq \widetilde{s}_n$ and $\widetilde{t}_1 \leq ... \leq \widetilde{t}_n$ be their ordered versions. Then the estimator is given by

$$\begin{aligned}\widehat{K}_{n(xy)}(v) &= 1 - \int_0^\infty \int_0^\infty \mathbb{I}_{\left\{\widehat{S}_{xy}(s,t)>v\right\}} d\widehat{S}_{xy}(ds,dt) \\ &= 1 - \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{I}_{\left\{\widehat{S}_{xy}\left(\widetilde{s}_i,\widetilde{t}_j\right)>v\right\}} \widehat{S}_{xy}\left(\triangle\widetilde{s}_i, \triangle\widetilde{t}_j\right),\end{aligned} \tag{12}$$

where $\widehat{S}_{xy}$ stands for a nonparametric estimator of the joint survival function, taking censoring into account, and $\widehat{S}_{xy}\left(\triangle\widetilde{s}_i, \triangle\widetilde{t}_j\right) = \widehat{S}_{xy}\left(\widetilde{s}_i,\widetilde{t}_j\right) - \widehat{S}_{xy}\left(\widetilde{s}_{i-1},\widetilde{t}_j\right) - \widehat{S}_{xy}\left(\widetilde{s}_i,\widetilde{t}_{j-1}\right) + \widehat{S}_{xy}\left(\widetilde{s}_{i-1},\widetilde{t}_{j-1}\right)$ is the estimated mass at $\left(\widetilde{s}_i,\widetilde{t}_j\right)$. We therefore need a nonparametric estimate of the bivariate survival function, denoted by $\widehat{S}_{xy}$, which depends on the margins. For $\widehat{S}_{xy}$ we will use the estimator introduced in Dabrowska (1988).

4.1.3 Dabrowska's estimator

Let us denote by $\widehat{S}_x^m$ and $\widehat{S}_y^f$ the Kaplan-Meier estimates of the univariate survival functions of T_x^m and T_y^f. Dabrowska's estimator is:

$$\widehat{S}_{xy}(s,t) = \widehat{S}_x^m(s)\,\widehat{S}_y^f(t) \prod_{\substack{0<u\leq s\\ 0<v\leq t}} \left(1 - L_{xy}(\triangle u, \triangle v)\right), \tag{13}$$

with

$$L_{xy}(\triangle u, \triangle v) = \frac{\widehat{\Lambda}_{10}(\triangle u, v^-)\,\widehat{\Lambda}_{01}(u^-, \triangle v) - \widehat{\Lambda}_{11}(\triangle u, \triangle v)}{\left(1 - \widehat{\Lambda}_{10}(\triangle u, v^-)\right)\left(1 - \widehat{\Lambda}_{01}(u^-, \triangle v)\right)}, \tag{14}$$

with $\triangle u = u - u^-$, and $\triangle v = v - v^-$. $\widehat{\Lambda}_{11}(\triangle u, \triangle v)$ is defined as the estimated hazard function of double failures (i.e. deaths) at point (u,v), while

$\widehat{\Lambda}_{10}(\triangle u, v^-)$ and $\widehat{\Lambda}_{01}(u^-, \triangle v)$ are the estimated hazard functions of failures of (x) at u and (y) at v respectively, given the exposed to risk defined at (u, v) [3]. The principle of equation (14) can be derived from the numerator. We match the expected number of joint failures in case of independence with the actual number of joint failures. A negative difference implies positive association. We define

$$H_{xy}(s,t) = \prod_{\substack{0<u\leq s \\ 0<v\leq t}} (1 - L_{xy}(\triangle u, \triangle v)), \tag{15}$$

as the multiplier by which the joint survival function differs from the one under independence (see equation (13)). It follows that positive association is implied if $H_{xy}(s,t) \geq 1$.

4.2 Wang-Wells theoretical version of the generator in the presence of censored data

Wang and Wells also suggested a procedure for obtaining the theoretical version of K_{xy}. For each copula the theoretical version can be compared with the empirical one under censored data, and the comparison provides a best fit selection criterium among different copulas. As is known, the original procedure in Genest and Rivest (1993) for Archimedean copula selection consists in

1) obtaining an estimated Kendall's $\hat{\tau}$ coefficient;

2) determining – for each candidate copula – the parameter value $\hat{\theta}$ which corresponds to the $\hat{\tau}$ by working the parameter out of the relationship

$$\hat{\tau} = 4\int_0^1 \lambda_{xy}(v)dv + 1 \tag{16}$$

where $\lambda_{xy}(v)$ is given by (11);

3) building – again for each copula – a theoretical K_{xy}, $K_{\phi_{\hat{\theta}}(xy)}$, by substituting, for a given generator, the estimate $\hat{\theta}$ in (10);

4) selecting as best fit copula the one where the theoretical K_{xy} is the least distant – according to the L^2 or other norms – from the empirical one, $\widehat{K}_{n(xy)}$.

This procedure is appropriate for complete data, but it is not applicable without provisos in the bivariate censored case. It is still applicable when the greatest observations are not censored, as shown by Wang and Wells (2000a) and done by Denuit et al. (2004). However, it is not applicable when, as in our case, the greatest observations are not uncensored. This is due to the fact that a consistent estimator for Kendall's tau does not exist in the latter case.

[3] For $i \in \{1, .., n\}$, let δ_{1i} and δ_{2i} be the indicators of the event that observations s_i and t_i respectively, will be uncensored. We define $\widehat{\Lambda}_{11}(s,t) = \int_{u=0}^{s}\int_{v=0}^{t} \widehat{K}_1(du, dv) \Big/ \widehat{M}\ u^-, v^-$; $\widehat{\Lambda}_{10}(s,t) = -\int_{u=0}^{s} \widehat{K}_2(du, t) \Big/ \widehat{M}\ u^-, t$; $\widehat{\Lambda}_{01}(s,t) = -\int_{v=0}^{t} \widehat{K}_3(s, dv)\, \widehat{M}\ s, v^-$, where $\widehat{M}(s,t) = \frac{1}{n}\#\{i \,|s_i > s, t_i > t\}$; $\widehat{K}_1(s,t) = \frac{1}{n}\#\{i \,|s_i > s, t_i > t, \delta_{1i} = 1, \delta_{2i} = 1\}$; $\widehat{K}_2(s,t) = \frac{1}{n}\#\{i \,|s_i > s, t_i > t, \delta_{1i} = 1\}$; $\widehat{K}_3(s,t) = \frac{1}{n}\#\{i \,|s_i > s, t_i > t, \delta_{2i} = 1\}$

Therefore, we adopt the modified Wang and Wells procedure, which comprises the following steps:

1') choosing as parameter value $\hat{\theta}$ for each copula the one which minimizes the distance between the corresponding theoretical $K_{\phi_{\hat{\theta}}(xy)}$ and the empirical $\widehat{K}_{n(xy)}$,

2') selecting as best fit copula the one which minimizes such a distance,

3') getting an estimate of Kendall's tau from the parameter value of the best fit copula.

In symbols, at stage 1') we define $K_{\phi_{\hat{\theta}}(xy)} = v - \lambda_{\phi_{\hat{\theta}}(xy)}(v)$, and choose as parameter estimate $\widehat{\theta}$ the one which makes the corresponding theoretical K_{xy}, $K_{\phi_{\hat{\theta}}(xy)}$, the least distant from the empirical K_{xy}, $\widehat{K}_{n(xy)}$. In the present paper, as in Wang and Wells, the distance or error is defined in the usual quadratic sense, i.e. it is taken under the L^2 norm:

$$error\left(\phi_{\theta(xy)}\right) = \int_{\xi}^{1} \left(K_{\phi_{\theta}(xy)} - \widehat{K}_{n(xy)}(v)\right)^2 dv. \tag{17}$$

Therefore

$$\widehat{\theta} = \arg\min_{\theta} \left[error\left(\phi_{\theta(xy)}\right)\right]. \tag{18}$$

In turn, the lower bound for the computation of the error, ξ, is taken to be the minimum value admissible according to Wang and Wells in the presence of censoring, that is the smallest value for which the empirical K_{xy} is positive:

$$\xi = \min\{\nu : K_{xy}(\nu) > 0\} \tag{19}$$

In this way we use all the available information, given double censoring.

At stage 2'), we select the copula which provides the minimum (minimized) error, $error\left(\phi_{\widehat{\theta}(xy)}\right)$. As a robustness check, we suggest double checking the result with another distance definition. A natural candidate is the distance of the sup norm, namely:

$$error'\left(\phi_{\widehat{\theta}(xy)}\right) = \sup_{\xi<\nu<1} \left|K_{\phi_{\hat{\theta}}(xy)}(v) - \widehat{K}_{n(xy)}(v)\right|.$$

At stage 3'), we get the corresponding dependence measure by using, in correspondence to the best fit copula, the general relationship (16), which, for the estimated values, becomes

$$\hat{\tau} = 4\int_{\xi}^{1} \left[v - K_{\phi_{\hat{\theta}}(xy)}(v)\right] dv + 1.$$

Wang and Wells (2000b) have derived the asymptotic properties of $\widehat{\theta}$, under the condition of $K_{\phi_{\hat{\theta}}(xy)}(v)$ being twice differentiable with respect to $\widehat{\theta}$, with bounded derivatives and weak convergence of Kendall's process

$$\sqrt{n}\left(\widehat{K}_{n(xy)}(\cdot) - K_{\phi_{\theta}(xy)}(\cdot)\right)$$

to a pointwise limit identified in Barbe et al. (1996).

4.3 Omnibus procedure

In order to confirm the results of the procedure described above, we estimate the dependence parameter and compare the copula fit through the pseudo-maximum likelihood or omnibus procedure. This method has been described in broad terms by Oakes (1994). Its statistical properties are analyzed in Genest et al. (1995). It is discussed in Cherubini et al. (2004).

The procedure treats marginal distributions as nuisance parameters of infinite dimension. The margins are estimated nonparametrically by rescaled versions of the Kaplan-Meier estimators, with the rescaling factor (multiplier) equal to $n/(n+1)$. The loglikelihood function to be maximized, denoted by $L(\theta)$, has the following shape:

$$L(\theta) = \sum_{i=1}^{n} \left[\begin{array}{c} \delta_{1i}\ \delta_{2i}\ \ln\left[c_\theta\left(u_i, v_i\right)\right] + \left(1-\delta_{1i}\right)\delta_{2i} \ln\left[\frac{\partial C_\theta(u_i,v_i)}{\partial v_i}\right] \\ +\delta_{1i}\left(1-\delta_{2i}\right)\ln\left[\frac{\partial C_\theta(u_i,v_i)}{\partial u_i}\right] + \left(1-\delta_{1i}\right)\left(1-\delta_{2i}\right)\ln\left[C_\theta\left(u_i, v_i\right)\right] \end{array} \right],$$

where $(u_i, v_i) = \left(\widehat{S}^m_{x_i}(s_i), \widehat{S}^f_{y_i}(t_i)\right)$, $C_\theta(u_i, v_i)$ is the copula being considered, $c_\theta(u_i, v_i)$ its density (i.e. the derivative with respect to both arguments) and $\delta_{1i}\ \delta_{2i}$ are as defined in section 4.1. Note that this procedure could also be applied to non Archimedean copulas; it leads to

$$\hat{\theta} = \arg\max_\theta L(\theta)$$

and to selecting the copula family whose optimal loglikelihood, $L(\hat{\theta})$, is maximal.

Similarly to the Wang and Wells method, the omnibus procedure relies on empirical margins. Both procedures therefore guarantee independency of the copula selection from the margin representation.

5 Application to the Canadian data set

5.1 Description of the data set

We use the same data set as Frees et al. (1996), Carriere (2000) and Youn and Shemyakin (1999, 2001). We have eliminated same-sex contracts and for couples with more than one policy we have eliminated all but one contract. This has left us with a set of 11,454 contracts on couples which were observed for slightly more than five years (the observation period runs from December 29, 1988, until December 31, 1993). The couples of the original Canadian data set have dates of birth between 1884 and 1993: in the papers which have dealt with it, the same law of mortality is assumed to apply for all the individuals of the same gender. Generation effects are therefore neglected. In this paper, on the other hand, we distinguish different generation survival probabilities and intensity processes. Since, as explained above, the methodology for the marginal survival functions applies to single generations x and y, we focus on a limited range of birth dates, both for males and females. In doing this, we have

also taken into consideration the fact that the average age difference between married man and woman in our sample is of three years. We have selected the generation of males born between January 1st, 1907 and December 31, 1920 and those of females born between January 1st, 1910 and December 31, 1923. These two subsets, which amount to 5,025 and 5,312 individuals respectively, have been used to estimate the marginal survival functions [4]. Then, in order to estimate joint survival probabilities, we have further concentrated on the couples whose members belong to the generation 1907-1920 for males and 1910-1923 for females. For the sake of simplicity, instead of labeling the generations with the whole range of initial ages, we denote the male generation with $x = 68$ and the female one with $y = 65$. This subset includes a total of 3,931 couples. Since both individuals and couples were born during a fourteen year period, and the observation period is five years, we have data available for nineteen years of age. So, entry ages vary between 68 and 82 for males and between 65 and 79 for females. In focusing on a generation and allowing for the three-year age difference, we have considered only one illustrative example. Obviously, the procedure explained above can be repeated for any other couple of generations.

For the chosen generation, we adopt the general procedure sketched in Section 2 for the margins and the one in Section 4 for the joint survival function.

We first obtain the empirical margins, using the Kaplan-Meier methodology. The marginal Kaplan-Meier data are used as inputs for the calibration of the analytical marginal survival functions, according to the methodology in Luciano and Vigna (2005). The KM margins also feed the Dabrowska estimate for the empirical joint survival function.

5.2 Marginal survival functions

The Kaplan-Meier estimates of the marginal survival probabilities are collected in Table 2.

[4] As a generation, we take a series of fourteen successive years of birth, rather than a single year. This assumption is based on one hand on the possibilities of reliable calibration (number of data) offered by the present data set; on the other, by the fact that there is not a unique definition of generation and, generally speaking, persons with ages of birth close to each other can be considered to belong to the same generation.

t	MALES $_tp_{68}$	FEMALES $_tp_{65}$
1	0.97225	0.98771
2	0.96103	0.98188
3	0.93828	0.97738
4	0.91387	0.9705
5	0.89417	0.9647
6	0.86973	0.9572
7	0.84597	0.94775
8	0.81598	0.93228
9	0.78349	0.91994
10	0.75892	0.90732
11	0.73091	0.89411
12	0.69639	0.88149
13	0.65776	0.86547
14	0.60382	0.84947
15	0.5573	0.82902
16	0.51807	0.7922
17	0.48384	0.75596
18	0.4018	0.72055
19	0.33158	0.68263

Table 2. Kaplan-Meier marginal survival probabilities.

We notice that, differently from both Carriere (2000) and Frees et al. (1996), we can calculate the empirical survival probabilities $_tp_x$ only up to $t = 19$. This, as already mentioned, is due to the limited range of birth dates of our generations, coupled with the five year length of observation. The analytical survival functions will be denoted as $S^m_{68}(t), S^f_{65}(t)$ respectively, with expression given by (4). The corresponding parameters are estimated by minimizing the mean square error between the Kaplan Meier and the analytical survival functions, similarly to Luciano and Vigna (2005). The estimated parameters are[5]

$$a_{68} = 0.0810051, \sigma_{68} = 0.00024, a_{65} = 0.124923, \sigma_{65} = 0.0034,$$

while the initial intensity values are[6]

$$\mu_{68}(0) = 0.0204276, \mu_{65}(0) = 0.0046943.$$

[5]Note that the values for the male, a_{68} and σ_{68}, satisfy the sufficient condition reported in Footnote 1, while the values for the female, a_{65} and σ_{65}, satisfy the necessary and sufficient condition of Footnote 1 for all relevant values of t.

[6]The values of $\mu_{68}(0)$ and $\mu_{65}(0)$, according to Luciano and Vigna (2005), are equal to $-\ln(p_{68})$ and $-\ln(p_{65})$ respectively, with p_{68} being the survival probability of a Canadian insured male born in 1920 and aged 68 and with p_{65} being the survival probability of a Canadian insured female born in 1923 and aged 65. Using the data set we have estimated p_{68} males and p_{65} females with the Kaplan-Meier method, without restrictions on the generation. This has been done in order to have an estimate of those survival probabilities as accurate as possible.

The two survival functions are presented in Figure 1.

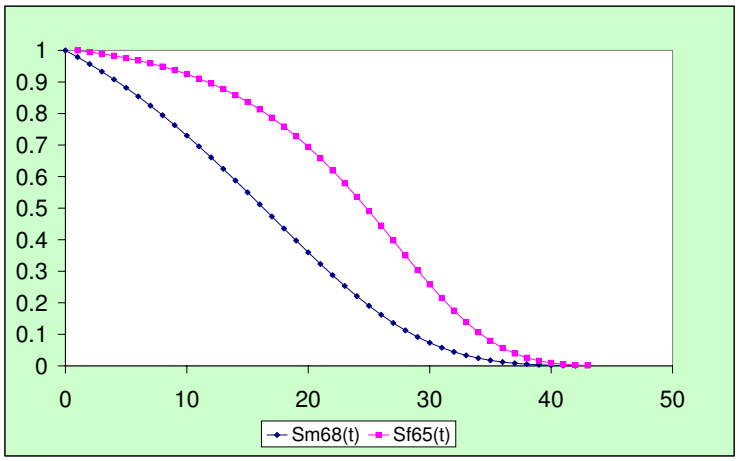


Figure 1. Analytical marginal survival functions $S_{68}^{m}(t)$ and $S_{65}^{f}(t)$.

5.3 The bivariate survival function (Dabrowska)

Given the empirical margins in Table 2, provided by the Kaplan-Meier method, we reconstruct the joint empirical survival function using the Dabrowska estimator. We have simplified the estimator by truncating to integer durations. This means that e.g. a duration of k (integer) corresponds to death between k and $k+1$. As data of death between durations 5 and 6 were incomplete (due to the maximal period of observation of 5.0075 years), we have not considered any deaths more than five years after the start of the observation. For durations greater than the observation period, we take the multiplier computed for the maximal duration. Because of this, our estimate of the joint survival function will be conservative.

In Table 3 we present the multipliers $H_{xy}(s,t)$, as defined in equation (15). The estimate $H_{xy}(s,t)$ is based on the generation data set as specified. We notice that all the multipliers are greater than one. This indicates positive association and confirms our intuition about the dependence of the lifetimes of couples. Later on, we will provide an appropriate measure (Kendall's tau) of the amount of association.

$s\backslash t$	0	1	2	3	4	≥ 5
0	1	1	1	1	1	1
1	1	1.00064	1.00089	1.00133	1.00197	1.00216
2	1	1.00106	1.00411	1.00585	1.00629	1.00708
3	1	1.00151	1.00466	1.00909	1.00998	1.01051
4	1	1.00152	1.00455	1.00883	1.01151	1.01241
≥ 5	1	1.00197	1.00483	1.0094	1.01254	1.01713

Table 3. Function $H_{68,65}(s,t)$.

Another relevant feature of the data which can be captured from the table is the fact that the multipliers are generally increasing per row and per column: this means that the amount of association is increasing. Namely, it means that, for given survival time of one individual in the couple, the conditional survival probability of the other member is more and more different from the unconditional one as time goes by.

5.4 The copula choice (Wang & Wells)

The Dabrowska empirical estimate of the joint survival function is in turn used as an input for $\hat{K}$, the empirical version of the K function, according to a discretized version of formula (12). In order to obtain the latter we divide the unit interval into a thousand subintervals. Figure 2 presents the empirical estimate for K, $\hat{K}$.

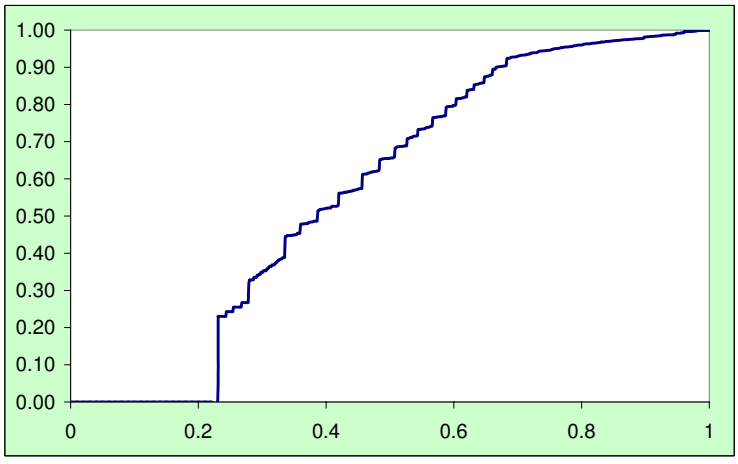


Figure 2. Empirical K, $\hat{K}$.

We observe that $\hat{K}(v)$ is zero for $v < 0.23$, because the smallest value of $S(s,t)$ is $S(19,19) = 0.23$ (Let us recall that this minimum is due to censoring and to the restriction to one generation, which reduces the observation window to 19 years).

As stated above, the empirical K is used, together with the theoretical ones, in order to

a) select the θ parameter value for each copula and

b) select the best fit copula.

At both stages we use the L^2 norm, and then we check the result using the sup norm.

For each copula, we choose as parameter estimate $\widehat{\theta}$ the one which makes the corresponding theoretical K, $K_{\phi_{\widehat{\theta}}}$, the least distant from the empirical K, $\widehat{K}_n$. The distance is first appreciated graphically, then computed by discretizing the integral appearing in $error\left(\phi_{\widehat{\theta}(xy)}\right)$. The discretization has step 1/1000, the one of the empirical K. The lower bound for the computation of the error is taken to be $\xi = 0.231$, according to the criterion in section 4.2.

We therefore obtain a different theoretical K function for each copula, and we are ready to compare them in order to assess their goodness-of-fit and to select the best copula. The graphical comparison can be done using Figure 3, where we present the functions $\lambda(z) = z - K(z)$ using the theoretical K functions and the empirical one (the graphical comparison results easier to perform with the λ functions than with the K ones).

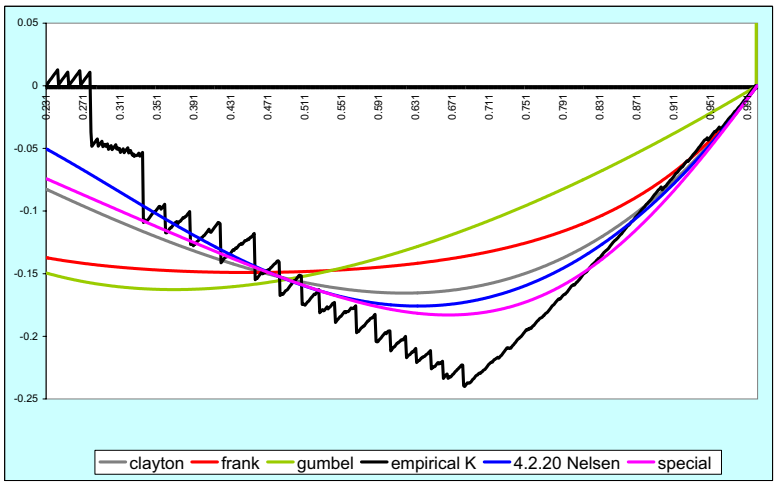


Figure 3. Graphical comparison among theoretical and empirical $\lambda(z) = z - K(z)$ functions.

Table 4 reports for each copula the distance between theoretical K function and empirical one, i.e. the minimized distance $error\left(\phi_{\widehat{\theta}(xy)}\right)$.

Clayton	Frank	Gumbel	4.2.20 Nelsen	Special
1.33638	3.09502	4.77706	0.72003	0.81101

Table 4. Error $\phi_{\widehat{\theta}(xy)}$, according to L^2- norm distance.

Both from the graph and the errors, we conclude that the best fit copula is the 4.2.20 Nelsen one.

By inverting the parameter value of the 4.2.20 Nelsen copula we also get an estimate of Kendall's tau, as explained under 3') of section 4.2: this results in $\hat{\tau} = 0.6039$, roughly in line with the values obtained for the same Canadian set, but without focusing on a generation, by other authors (Frees et al., 1996, Carriere, 2000, Youn and Shemyakin, 1999, 2001, Shemyakin and Youn, 2001).

In the absence of a formal test for censored data (see Genest et al., 2006), we also check the correctness of the copula choice by repeating the procedure - namely, points 1') and 2') above - with the sup norm: we again obtain the 4.2.20 Nelsen copula as the best fit one.

5.5 Omnibus procedure

As a further check of our selection, we implement the omnibus or pseudo-maximum likelihood procedure. As inputs for it, we again use the rescaled Kaplan-Meier marginal probabilities in Table 2. Table 5 presents the estimated parameters $\hat{\theta}$ for each copula, their standard errors and the maximized likelihood function.

	θ via omnibus procedure	standard error	θ via Wang and Wells procedure	Maximum-likelihood
Clayton	2.23253	0.32899	2.73117	-734.698
Frank	3.48923	0.4154	6.31334	-735.268
Gumbel	1.12921	0.02172	2.2612	-750.297
4.2.20 Nelsen	1.04017	0.14272	1.00476	-734.573
Special	4.37336	0.42495	3.09667	-740.396

Table 5. Estimated θ via omnibus procedure and Wang and Wells procedure, together with their standard errors and maximized likelihood.

The likelihood is maximized in correspondence to the 4.2.20 Nelsen copula: this procedure then confirms the results of the Wang and Wells one.

The omnibus approach also confirms the validity of the Kendall's tau estimates obtained with the Wang and Wells approach: using the above standard errors, for each copula parameter - and consequently for the Kendall's tau - we computed a 95% confidence interval around the maximum likelihood one. Both the copula parameter and the Kendall's tau of the Wang and Wells method fall

in the 95% confidence interval of the omnibus procedure estimate, if one considers the 4.2.20 Nelsen or Clayton copula. However, if the test is repeated using the estimated parameters of the sup norm distance, it results that the 4.2.20 Nelsen and Special estimates from the Wang and Wells methodology fall within the maximum likelihood significance bounds. Therefore, the 4.2.20 Nelsen is the only one which passes the test for both norms.

6 Analysis of joint survivorship

6.1 The analytical joint survival function

We couple the fitted marginal survival functions of Section 5.2 with the best fit copula choice of Section 5.4, according to the formula

$$S_{68,65}(s,t) = C_{68,65}(S^m_{68}(s), S^f_{65}(t))$$

and using the 4.2.20 Nelsen copula, as specified in Table 1. In order to simplify the notation, from now onwards we will drop the indices, so that

$$S_{68,65}(s,t) = S(s,t);\ S^m_{68}(s) = S(s);\ S^f_{65}(t) = S(t)$$

By coupling marginals and copula, we obtain the joint survival function $S(s,t)$ of Figure 4. Some of its sections are presented in Figures 5 and 6 respectively.

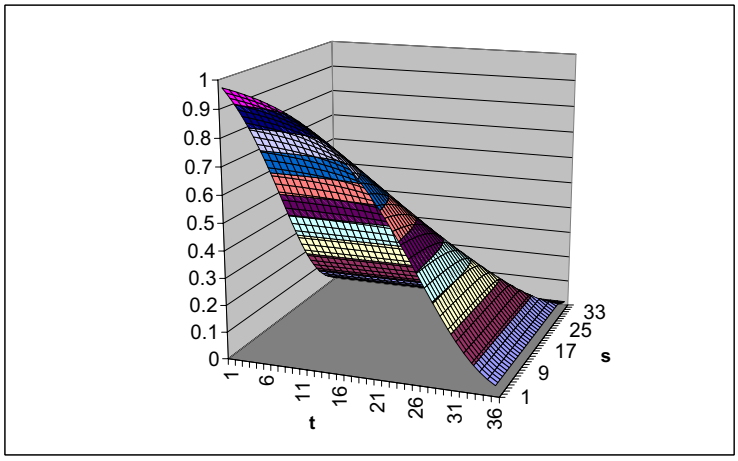


Figure 4. Joint analytical survival function, $S_{68,65}(s,t)$.

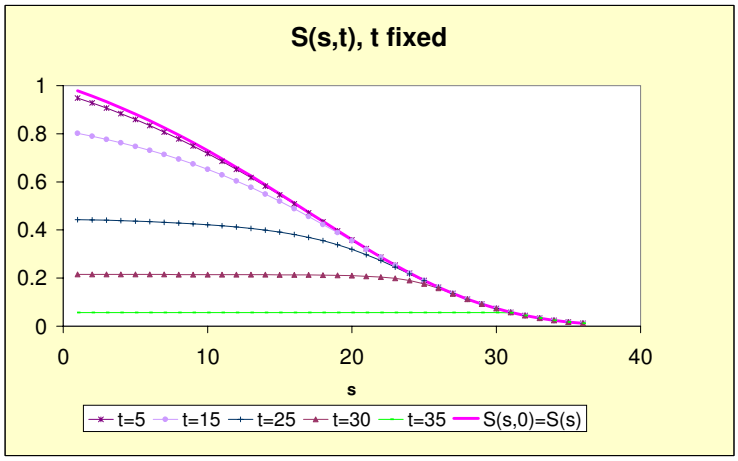


Figure 5. Sections of the joint analytical survival function, for female fixed duration.

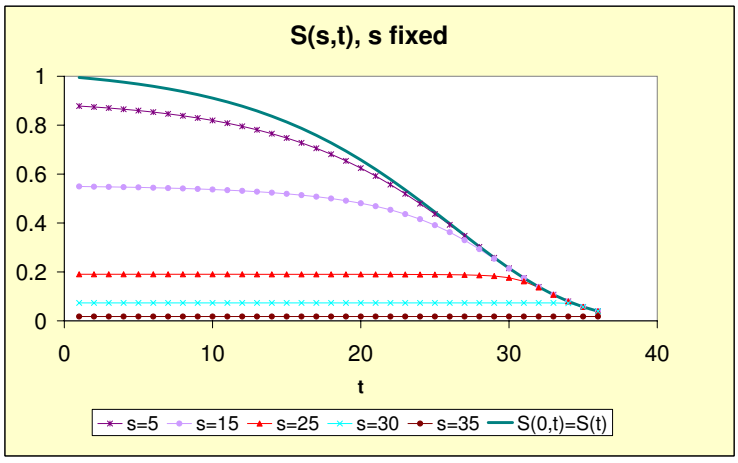


Figure 6. Sections of the joint analytical survival function, for male fixed duration.

Looking at Figure 5, we notice that, if t is high, $S(s,t)$ is almost flat until a certain duration s^* after which it decreases. This is due to the fact that the probability for the female of surviving t years, with high t, is very low: this

affects to a great extent the joint probability of surviving s years for the male and t years for the female (even when the probability $S(s,0)$ is very high because s is small). After duration s^*, the joint probability starts to decrease because of the joint effect of low probability of surviving t years for the female and s years for the male.

The same comments made for Figure 5 apply to Figure 6. Note that, while the duration s^* after which $S(s,t)$ t fixed starts to decrease is always lower than the fixed value of t, here the duration t^* after which $S(s,t)$ s fixed starts to decrease is always higher than the fixed value of s. This is probably due to the difference in death rates for a male and a female with the same age. Evidence of this can also be found in the different level of the sections when we change sex: for instance, $S(s,35)$ lies at a higher level than $S(35,t)$, $S(s,30)$ lies at a higher level than $S(30,t)$, etc.

6.2 Time-dependent association

In this section we report and discuss the measures of time-dependent association.

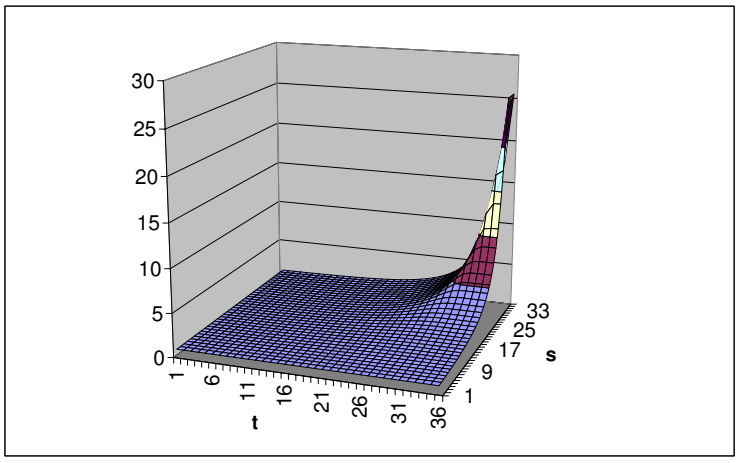


Figure 7. Association measure $\psi_1(s,t) = \frac{S(s,t)}{S(s)S(t)}$.

In Figure 7, we report the ratio between the joint survival function $S(s,t)$ and the probability which we would obtain under the assumption of independence, namely the product copula one, $S(s)S(t)$. Note that we use the short notation $S^m_{68}(s) = S(s), S^f_{65}(t) = S(t)$. Figure 7 therefore reports the time dependent measure of association $\psi_1(s,t)$ as defined in (7). The ratio has values greater than one because of positive dependence, is monotone in each argument, as

expected from the copula selected, and reaches very high values for large s and t.

The sections of the dependence measure in Figure 7 are in Figures 8 and 9. All the curves start at 1 for $s = 0$ or $t = 0$ and increase monotonically until a certain value, defined as s^* in Figure 8 and t^* in Figure 9, from which they remain constant. The ratio of the conditional to unconditional survival probability for men, given a female age, is then stable above s^*, while the corresponding ratio for women, given a male age, is stable over t^*. Comparing the sections of Figure 8 with those of Figure 9 for the same fixed value, we observe that $s^* < t^*$. This is a distinctive feature of the mortality experienced by males compared to females, which the specific joint survival function permits to highlight.

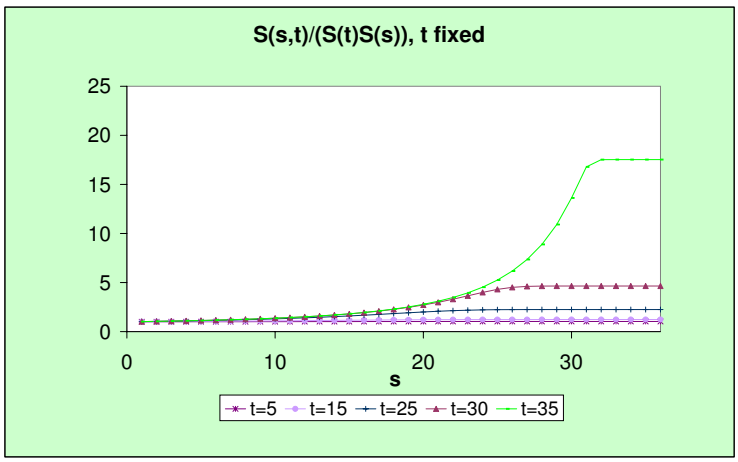


Figure 8. Sections of the association measure $\psi_1(s,t) = \frac{S(s,t)}{S(s)S(t)}$, for fixed female duration t.

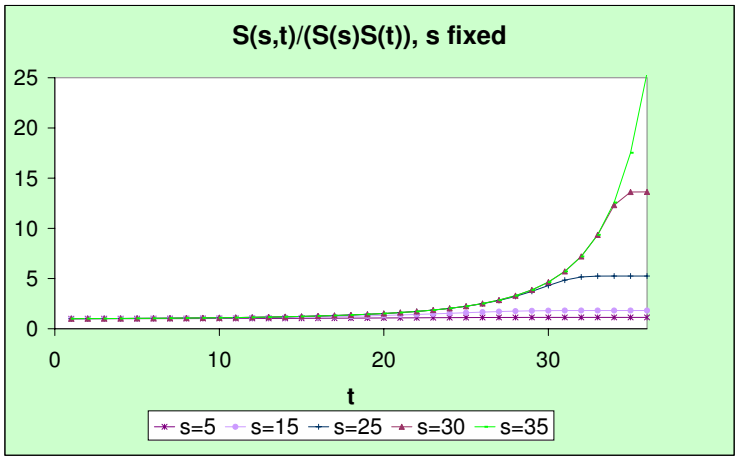


Figure 9. Sections of the association measure $\psi_1(s,t) = \frac{S(s,t)}{S(s)S(t)}$, for fixed male duration s.

Starting from the previous age dependent association measure, we compute the conditional survival probabilities resulting from our estimates, $S^m_{68}(s \mid t)$ and $S^f_{65}(t \mid s)$ respectively. For the sake of brevity, we denote them as $S^m_{68}(s \mid t) = S(s \mid t), S^f_{65}(t \mid s) = S(t \mid s)$ and present them in Figures 10 and 11 respectively. In Figure 10, for small values of t, $S(s|t)$ approaches the marginal distribution $S(s)$ as expected . For high values of t the level of $S(s|t)$ increases, and is even equal to 1 for a considerable period of time, if $t = 30, 35$. This means that the probability of surviving for a long time for the male is actually one, given that the female survives even longer. Similar comments apply to Figure 11. Here we notice that with high values of s, $S(t|s)$ is 1 for durations longer than s. Generally speaking, the fact that the male survives s years seems to guarantee that the female survives at least s years.

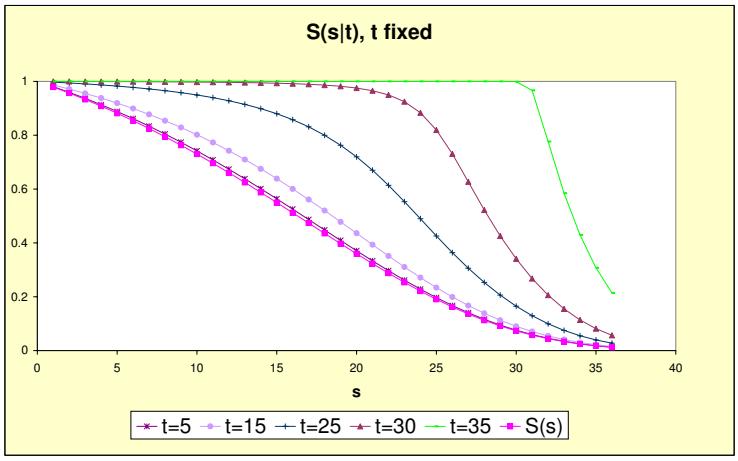


Figure 10. Male conditional survival probability for fixed female duration t.

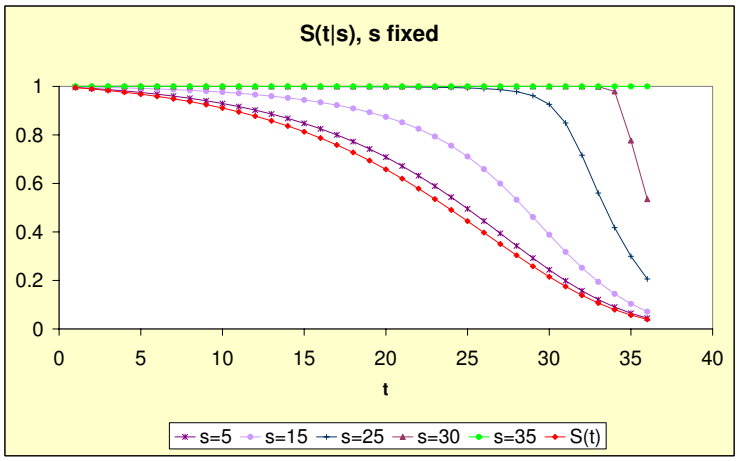


Figure 11. Female conditional survival probability for fixed male duration s.

Regarding the second measure of time-dependent association in Section 3, Table 6 illustrates the measures $\psi_{2x}(0,t)$ and $\psi_{2y}(s,0)$. The unconditional life expectancies $E[T_x^m]$ and $E\left[T_y^f\right]$ are respectively equal to 16.51 and 21.84.

Column 2 displays the relative increase of the conditional expected remaining lifetime of (x), given that (y) survives to age $y+t$, with respect to $E[T_x^m]$: as explained in Section 3, in correspondence to our copula, it increases as a function of t. Similarly, column 4 shows the relative increase of the conditional expected remaining lifetime of (y), given that (x) survives to age $x+s$, with respect to $E\left[T_y^f\right]$: it is increasing as a function of s, as expected. We observe that, for $s=t$, $\psi_{2x}(0,t) < \psi_{2y}(s,0)$ for small values of s or t, but the inequality sign is reversed for large values of s or t. Knowledge of the fact that the female survives a given number of years affects the remaining survivorship of the male less than the opposite knowledge, for short maturities (1, 5, 10 respectively). The opposite applies to long maturities (more than 10 years).

Even this second measure then gives us a very specific information on the sample survivorship.

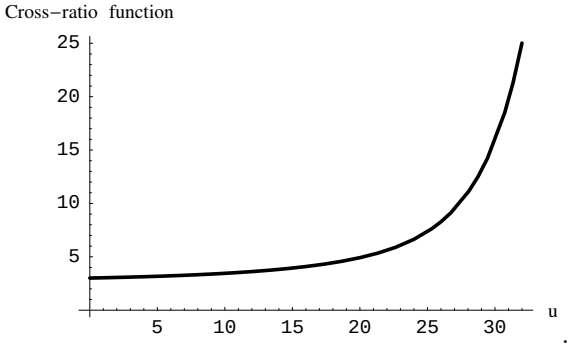
t	$\psi_2(0,t)=\frac{E(T_x\|T_y>t)}{E(T_x)}$	s	$\psi_2(s,0)=\frac{E(T_y\|T_x>s)}{E(T_y)}$
1	1.002	1	1.006
5	1.015	5	1.028
10	1.044	10	1.055
15	1.097	15	1.088
20	1.198	20	1.127
25	1.379	25	1.171
30	1.627	30	1.213

Table 6. Association measures $\psi_2(0,t)$ (column 2) and $\psi_2(s,0)$ (column 4).

The third measure of time-dependent association in Section 3, the cross-ratio function for the 4.2.20 Nelsen copula, as a function of $S(u,u)$, is

$$CR(S(u,u)) = 1+\theta\left(1+[S(u,u)]^{-\theta}\right).$$

Like the previous measures, and as shown in Spreeuw (2006), it is increasing as a function of duration u. It measures the relative increase in the survivor force of mortality. Figure 12 gives a plot of $CR(S(u,u))$ versus u: note that, as expected, $CR(S(u,u))$ is increasing in u and takes very high values for u greater than 30.



.

Figure 12. Cross ratio function or association measure $CR(S(u,u))$.

To sum up, since the 4.2.20 Nelsen copula is the best fit one for the sample at hand, members of a couple become more dependent on each other as they age. The measures illustrated above give different perspectives on this age dependence, based respectively on conditional survival probabilities, expected lifetimes and their conditional version, as well as relative increase of the survivor's force of mortality.

7 Summary and conclusions

This paper represents a first attempt to model the mortality risk of couples of individuals according to the stochastic intensity approach.

From a theoretical point of view, we extend the Cox processes setup to couples. Cox processes are based on the idea that mortality is driven by a jump process the intensity of which is itself a stochastic process, pertaining to a gender's particular generation. The dependence between the survival times of members of a couple is captured by a copula, which we assume to be of the Archimedean class, as in the previous literature on bivariate mortality.

From an empirical point of view, we present an application on a data set of couples from a large Canadian insurer, well known in the literature on bivariate survival functions. First, we choose two generations of males and females in their retirement age and provide a calibration of their marginal survival functions. We select time-homogeneous non mean-reverting affine processes for the intensity and give the corresponding survival functions in analytical form. Then, we parametrize and select the best fit copula, according to the methodology of Wang and Wells (2000b) for censored data. We obtain the 4.2.20 Nelsen copula as the best fit one and we confirm its appropriateness with the pseudo-maximum likelihood procedure. Since the 4.2.20 Nelsen copula is the best fit one, dependence increases with age.

Coupling the best fit copula with the calibrated margins, we obtain a joint survival function which is fully analytical and incorporates the stochastic terms of the mortality intensities of both members.

We believe that the main contributions in this paper are twofold. Firstly, generation effects are taken into account in modelling mortality of coupled lives, since the methodology of stochastic intensities is applied to the univariate margins. Secondly, we perform the best fit selection on candidate copulas which incorporate different time-dependent associations and, therefore, we infer from the data how association evolves when the couple ages. Although in this paper only one generation is analyzed, the procedure can evidently be repeated for several generations. In further research, it would be interesting to investigate the impact of generation on bivariate mortality, by comparing the results for different generations.

References

[1] Anderson, J.E., Louis, T.A., Holm, N.V. and Harvald, B. (1992). Time-dependent association measures for bivariate survival distributions, *Journal of the American Statistical Association*, 87 (419), 641-650.

[2] Artzner, P. and Delbaen, F. (1992). Credit risk and prepayment option, *ASTIN Bulletin*, 22, 81-96.

[3] Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations, *Insurance: Mathematics and Economics*, 37, 443–468.

[4] Brémaud, P. (1981). *Point processes and queues - martingale dynamics*, New York: Springer Verlag.

[5] Cairns, A. J. G., Blake, D. and Dowd, K. (2006). Pricing death: Framework for the valuation and securitization of mortality risk, *ASTIN Bulletin*, 36, 79-120.

[6] Carriere, J.F. (2000). Bivariate survival models for coupled lives, *Scandinavian Actuarial Journal*, Issue 1, 17-32.

[7] Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula methods in Finance*, Chichester: John Wiley.

[8] Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, 65, 141-151.

[9] Dabrowska, D.M. (1988). Kaplan-Meier estimate on the plane, *Annals of Statistics*, 16, 1475-1489.

[10] Dahl, M. (2004) Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts, *Insurance: Mathematics and Economics*, 35, 113–136.

[11] Denuit, M. and Devolder, P. (2005). Continuous time stochastic mortality and securitization of longevity risk, *mimeo*.

[12] Denuit, M., Dhaene, J., Le Bailly de Tilleghem, C. and Teghem, S. (2001). Measuring the impact of dependence among insured lifelengths, *Belgian Actuarial Bulletin*, 1 (1), 18-39.

[13] Denuit, M., Purcaru, O. and Van Keilegom, I. (2004). Bivariate Archimedean copula modelling for loss-ALAE data in non-life insurance, *Discussion Paper* 0423, Institut de Statistique, Université Catholique de Louvain, Louvain-La-Neuve, Belgium.

[14] Duffie, D. Filipovič, D. and Schachermayer, W. (2003). Affine processes and applications in finance, *Annals of Applied Probability*, 13, 984–1053.

[15] Duffie, D. and Singleton, K. (1999). Modelling term structures of defaultable bonds, *Review of Financial Studies, 12,* 687-720.

[16] Duffie, D. and Singleton, K. (2003). *Credit risk*, Princeton University Press.

[17] Frees, E.W., Carriere, J.F. and Valdez, E.A. (1996). Annuity valuation with dependent mortality, *Journal of Risk and Insurance*, 63 (2), 229-261.

[18] Genest, C, Ghoudi, K. and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika,* 82, 543-552.

[19] Genest, P., Quessy, J.F. and Rémillard, B. (2006). Goodness of fit procedures for copula models based on the probability integral transformation, *Scandinavian Journal of Statistics,*1-30.

[20] Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas, *Journal of the American Statistical Association*, 88, 1034-1043.

[21] Lando, D. (1998). On Cox processes and credit risky securities, *Review of Derivatives Research*, 2, 99–120.

[22] Luciano, E. and Vigna, E. (2005). Non mean reverting affine processes for stochastic mortality, ICER working paper and Proceedings of the XVth International AFIR Colloquium, Zurich, *Under review.*

[23] Milevsky, M.A. and Promislow, S. D. (2001). Mortality derivatives and the option to annuitise, *Insurance: Mathematics and Economics*, 29, 299–318.

[24] Nelsen, R.B. (2006). *An Introduction to Copulas*, Springer Series, Second edition.

[25] Oakes, D. (1989). Bivariate survival models induced by frailties, *Journal of the American Statistical Association*, 84 (406), 487-493.

[26] Oakes, D. (1994). Multivariate survival distributions, *Journal of Nonparametric Statistics* 3, 343-354.

[27] Shemyakin, A. and Youn, H. (2001). Bayesian estimation of joint survival functions in life insurance, In: *Monographs of Official Statistics. Bayesian Methods with applications to science, policy and official statistics*, European Communities, 489-496.

[28] Schrager, D. F. (2006). Affine stochastic mortality, *Insurance: Mathematics and Economics*, 40, 81-97.

[29] Spreeuw, J. (2006). Types of dependence and time-dependent association between two lifetimes in single parameter copula models, *Scandinavian Actuarial Journal*, Issue 5, 286-309.

[30] Wang, W. and Wells, M.T. (2000a). Estimation of Kendall's tau under censoring, *Statistica Sinica*, 10, 1199-1215.

[31] Wang, W. and Wells, M.T. (2000b). Model selection and semiparametric inference for bivariate failure-time data, *Journal of the American Statistical Association*, 95, 62-72.

[32] Youn, H. and Shemyakin, A. (1999). Statistical aspects of joint life insurance pricing, *1999 Proceedings of the Business and Statistics Section of the American Statistical Association*, 34-38.

[33] Youn, H. and Shemyakin, A. (2001). Pricing practices for joint last survivor insurance, *Actuarial Research Clearing House*, 2001.1.