THE CARLO ALBERTO NOTEBOOKS

Collegio Carlo Alberto

Modelling stochastic mortality for dependent lives

Elisa Luciano
Jaap Spreeuw
Elena Vigna

Working Paper No. 43
April 2007
www.carloalberto.org

Modelling stochastic mortality for dependent lives[1]

Elisa Luciano[2], Jaap Spreeuw[3]and Elena Vigna[4]

February 2007[5]

[1]The authors 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. They acknowledge support from the European Science Foundation (ESF) through the 'Advanced Mathematical Methods for Finance' project. Comments from participants to the 2006 IME conference are gratefully acknowledged. All remaining errors are ours.

[2]Collegio Carlo Alberto, University of Turin, ICER and FERC.

[3]Cass Business School, London.

[4]University of Turin.

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 represents a first attempt to model the mortality risk of couples of individuals, according to the stochastic intensity approach. We extend to couples the Cox processes set up, namely the idea that mortality is driven by a jump process whose intensity is itself a stochastic process, proper of a particular generation within each gender. 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) copula and the (analytical) margins. First, we calibrate and select the best fit copula according to the methodology of Wang and Wells (2000b) for censored data. Then, we provide a sample-based calibration for the intensity, using a time-homogeneous, non mean-reverting, affine process: this gives the marginal survival functions. By coupling the best fit copula with the calibrated margins we obtain a joint survival function which incorporates the stochastic nature of mortality improvements. Several measures of time dependent association can be computed out of it.

We apply the methodology to a well known insurance dataset, using a sample generation. The best fit copula turns out to be a Nelsen one, which implies not only positive dependency, but dependency increasing with age.

JEL Classification: G22

Keywords: stochastic mortality, bivariate mortality, copula functions, longevity risk.

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 as a doubly stochastic or Cox process, i.e. in interpreting death arrival as the first jump time of a Poisson-like process, whose intensity, contrary to the one of the standard Poisson, is a stochastic process. A priori then 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 behavior of the common risk factor, the intensity. The term "common" extends here to a whole generation within a gender.

The stochastic mortality approach has been proposed by Milevsky and Promislow (2001) and developed by Dahl (2004), Cairns et al. (2005), Biffis (2005), Schrager (2006), Luciano and Vigna (2005). The probabilistic setting however can be traced back to Brémaud (1981), and has been quite extensively applied in the financial literature on default arrival (see for instance the seminal works of Artzner and Delbaen (1992), Duffie and Singleton (1999) and Lando (1998)). Provided that univariate affine processes are used for the intensity, the approach leads to analytical representations of survival probabilities.

Up to now, no attempt has been made to model the survivorship of couples of individuals stochastically, in the sense just specified. This paper attempts to fill up this gap, making use of the copula approach. We model and calibrate the marginal survival functions and the copula separately. In doing that, we do not impose a specific copula; at the opposite, we select a best fit one in a group of Archimedean ones. Having selected and calibrated it, by coupling it with sample-calibrated margins, we get a fully analytical survival function. Since in the end we work with analytical marginal survival functions as well as analytic copulas, the joint survival function can be extended to durations longer than the observation period and measures of age-depedent association can be discussed.

We apply our modelling and calibration procedure to a huge sample of joint survival data, belonging to a Canadian insurer, which has been used in order to discuss (non stochastic) joint mortality in Frees et al. (1996), Carriere (2000), Shemyakin and Youn (2001) and Youn and Shemyakin (1999, 2001).

The outline of the paper is as follows: in Section 2 we recall the copula approach to joint survivorship and justify the copula class we are going to adopt, the Archimedean one. In Section 3 we describe a copula calibration and selection methodology, consistent with the copula class suggested above, and originally proposed by Wang and Wells (2000b). Wang and Wells' methodology, which in turn extends the approach of Genest and Rivest (1993) to the case with censoring, has the advantage of allowing not only the calibration of the parameters for each Archimedean copula, but also of suggesting which is the best fit Archimedean copula in the calibrated group.

In Section 4 we review the stochastic mortality approach at the univariate level, and the particular marginal model we are going to adopt. We explain both the model and its calibration issues with uncensored and censored data.

From Section 5 onwards we apply the theoretical framework and the calibration method to the data sample: we present the data set, find the empirical margins with the Kaplan-Meier methodology, apply the Wang and Wells' copula calibration and selection procedure, and compare its results with the ones of the omnibus or pseudo maximum likelihood procedure. We then derive the marginal survival functions, adapting the procedure in Luciano and Vigna (2005). In Section 6 the specific best fit copula obtained, together with the analytical margins, enables us to present an estimate of the joint survival function and to discuss the corresponding measures of time-dependent association, following the results in Spreeuw (2006). Section 7 concludes.

2 Modelling bivariate survival functions with copulas

Suppose that the heads (x) and (y), belonging respectively to the gender m (males) and f (females), have remaining lifetimes T_x^m and T_y^f, respectively, both with continuous distributions. We denote the marginal survival functions by S_x^m and S_y^f, respectively, so that, for all $t \geq 0$, $S_x^m(t) = \Pr[T_x^m > t]$ and $S_y^f(t) = \Pr\left[T_y^f > t\right]$. By Sklar's theorem, there exists a copula, denoted by C, such that for all $(s,t) \in \mathbb{R}^{2+}$ the joint survival function of x and y, denoted by S, can be represented in terms of the marginal ones:

$$S(s,t) = C(S_x^m(s), S_y^f(t)).$$

This representation is unique over the range of the margins.

The copula approach has become a popular method of modelling the (non stochastic) bivariate survival function of the two lives of one couple. Working on the same data set that we will use, 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's copula and couple the two lives from the time of birth. Carriere (2000) on the other hand, discusses several copulas with more than one parameter (Frank, Clayton, Normal, Linear Mixing, Correlated Frailty), and couples the lives at the start of the observation period. 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 member 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. This process cannot be other than independent of the specification of the margins: Genest and Rivest (1993) have shown that this is feasible for Archimedean copulas, as long as data are complete, i.e. uncensored. Denuit et al. (2001) managed to get hold of complete data by visiting cemeteries. Applying the method developed by Genest and Rivest (1993), they established a weak correlation of lifetimes between males and females, and identified several plausible candidates for the copula.

Genest and Rivest's method cannot be used if data are censored. This is the case for the data set from the large Canadian insurer which we are going to use. The period of observation is slightly longer than five years, and most lives were still alive at the end of the period of observation. Wang and Wells (2000b) have extended Genest and Rivest's method to bivariate right-censored data. The procedure requires a nonparametric estimator of the joint bivariate survival function. A popular candidate of such an estimator is Dabrowska (1988), which needs estimates of the margins in accordance with Kaplan-Meier.

We are going to apply the Wang and Wells' method for the data set at hand, since their methodology allows

- the calibration of the copula parameters - for any given copula family in the Archimedean class – and
- the choice of the best fit copula among the calibrated ones.

This paper then differs from the aforementioned papers on bivariate survival models (Frees et al., 1996, Carriere, 2000, Shemyakin and Youn, 2001, Youn and Shemyakin, 1999, 2001, Denuit et al., 2001) not only because we include stochastic mortality improvements at the marginal level, but also because, instead of assuming a specific copula, we select a best fitting one by following the Wang and Wells procedure for censored data. Using Wang and Wells means that we maintain the Archimedean assumption for the copula.

Archimedean copulas may be constructed using a function $\phi : I \rightarrow \Re^{*+}$, 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:

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

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 $\phi(t)$	$C(u,v)$	**Kendall's τ**
1	Clayton	$t^{-\theta}-1$	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1}{\theta}}$	$\frac{\theta}{\theta+2}$
2	Gumbel-Hougaard	$(-\ln t)^{\theta}$	$\exp\left[-\left((-\ln u)^{\theta}+(-\ln v)^{\theta}\right)^{\frac{1}{\theta}}\right]$	$1-\frac{1}{\theta}$
3	Frank	$-\ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$-\frac{1}{\theta}\ln\left[1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{e^{-\theta}-1}\right]$	$1-\frac{4}{\theta}\left(\int_{t=0}^{\theta}\frac{t}{\theta(e^{t}-1)}dt-1\right)$
4	Nelsen	$\exp\left[t^{-\theta}\right]-e$	$\left[\ln\left(\exp\left(u^{-\theta}\right)+\exp\left(v^{-\theta}\right)-e\right)\right]^{-\frac{1}{\theta}}$	$1-\frac{4}{\theta}\left(\frac{1}{\theta+2}\right.$ $\left.-\int_{t=0}^{1}t^{\theta+1}\exp\left[1-t^{-\theta}\right]\right)$
5	Special	$\frac{1}{t^{\theta}}-t^{\theta}$	$2^{-\frac{1}{\theta}}\left(-W+\sqrt{4+W^{2}}\right)$; $W=\phi(u)+\phi(u)$	Complicated form

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. We will deal with all of them in Section 6.

First of all, Anderson et al. (1992) introduced the rescaled conditional probability, denoted by ${}_1(s,t)$:

$$
{}_1(s,t)=\frac{S(s,t)}{S_x^m(s)\,S_y^f(t)}, \tag{2}
$$

for fixed t and s. If T_x^m and T_y^f are independent, then ${}_1(s,t)=1$ for all $s\geq 0$ and $t\geq 0$. If T_x^m and T_y^f are positively associated, then ${}_1(s,t)>1$ for all $s>0$ and $t>0$, with ${}_1$ monotone nondecreasing in each argument. This measure has also an interpretation in terms of conditional probabilities, since

$$
{}_1(s,t)=\frac{\Pr\left[T_x^m>s\left|T_y^f>t\right.\right]}{\Pr\left[T_x^m>s\right]}=\frac{\Pr\left[T_y^f>t\left|T_x^m>s\right.\right]}{\Pr\left[T_y^f>t\right]}
$$

Secondly Anderson et al. (1992) discuss the conditional expected residual lifetimes of (x) and (y), which we will specify as ${}_{2x}(s,t)$ and ${}_{2y}(s,t)$, respectively

$$
\begin{aligned}
{}_{2x}(s,t) &= \frac{E\left[T_x^m-s\left|T_x^m>s,T_y^f>t\right.\right]}{E\left[T_x^m-s\left|T_x^m>s\right.\right]}\\
{}_{2y}(s,t) &= \frac{E\left[T_y^f-t\left|T_x^m>s,T_y^f>t\right.\right]}{E\left[T_y^f-t\left|T_y^f>t\right.\right]}.
\end{aligned} \tag{3}
$$

The measure $\;_{2x}(s,t)$ ($\;_{2y}(s,t)$) describes how the knowledge that $T_y^f > t$ ($T_x^m > s$) affects the expected lifetime of T_x^m (T_y^f). Independence of T_x^m and T_y^f implies $\;_{2x}(s,t) = \;_{2y}(s,t) = 1$, while if T_x^m and T_y^f are positively associated, then $\;_{2x}(s,t) > 1$ and $\;_{2y}(s,t) > 1$ for all $s > 0$ and $t > 0$, with $\;_{2x}(s,t)$ ($\;_{2y}(s,t)$) monotone nondecreasing in t (s). In this paper we will concentrate on the behaviour of the functions $\;_{2x}(0,t)$ and $\;_{2y}(s,0)$.

The third measure is the cross-ratio function $CR(S(t_t,t_2))$, defined in Clayton (1978) and studied by Oakes (1989):

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

Spreeuw (2006) has shown that for Archimedean copulas and $u = s = t$, the cross ratio definition reduces to an expression in terms of the inverse of the generator:

$$CR(S(u,u)) = \left(\frac{\phi^{-1}(v)\left(\phi^{-1}\right)''(v)}{\left(\left(\phi^{-1}\right)'(v)\right)^2} \right)_{v=\phi(S(u,u))}. \tag{4}$$

Oakes (1994) derived a similar expression for frailty models (which are a subclass of Archimedean copula models).

The cross-ratio function specifies the relative increase of the force of mortality of the survivor, immediately upon death of the partner. If $CR(S(u,u))$ increases (decreases) as a function of u, this means that members of a couple become more (less) dependent on each other as they age. Manatunga and Oakes (1996) have demonstrated that increasing dependence with age entails an increasing plot of $CR(v)$ versus $1 - v$, for $v \in [0,1]$ (Note that $S(0,0) = 1$ and $\lim_{u\to\infty} S(u,u) = 0$.)

The first copula in Table 1, Clayton, will be studied because it is well known and bears the special property of the association remaining constant over time. Copulas 2 (Gumbel-Hougaard) and 3 (Frank) share the characteristics of being well known as well. Moreover, unlike Clayton, the association is 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 “Nelsen copula”. Copula 5, which is also due to Chapter 4 of Nelsen (2006), will be labelled as the “Special copula”. It was studied in Spreeuw (2006). Copulas 4 and 5, unlike the first three copulas, have association increasing over time.

3 Copula estimate and best fit choice

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

3.1 The distribution function of the Archimedean copula

Let $Z = S\left(T_x^m, T_y^f\right)$. Define K as the distribution function of Z. Note that we have that $Z = C\left(U, V\right)$ where $\left(U, V\right)$ is a random couple with unit uniform margins, and C the copula.

Genest and Rivest (1993) have shown that, for Archimedean copulas, with generator ϕ, this distribution function K is given by

$$K\left(z\right) = z - \lambda\left(z\right) \tag{5}$$

where

$$\lambda\left(z\right) = z - \frac{\phi\left(z\right)}{\phi'\left(z\right)}, \quad 0 < z \leq 1. \tag{6}$$

and ϕ' is the generator derivative. The function K 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 application of the current paper.

3.1.1 General principle without censoring

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

$$\widehat{K}_n\left(z\right) = \frac{1}{n}\#\left\{i\left|z_i \leq z\right.\right\} \text{ where } z_i = \frac{1}{n-1}\#\left\{\left(x_{(j)}, y_{(j)}\right)\left|x_{(j)} < x_{(i)}, y_{(j)} < y_{(i)}\right.\right\},$$

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

3.1.2 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. Since K can be written as

$$K\left(v\right) = \Pr\left[S\left(T_x^m, T_y^f\right) \leq v\right] = \mathbb{E}\left[\mathbb{I}_{\left\{S\left(T_x^m, T_y^f\right) \leq v\right\}}\right],$$

the estimator is given by

$$\widehat{K}_n\left(v\right) = \int_0^\infty \int_0^\infty \mathbb{I}_{\left\{\widehat{S}(s,t) \leq v\right\}} d\widehat{S}\left(s, t\right), \tag{7}$$

where $\widehat{S}$ stands for a nonparametric estimator of the joint survival function, taking censoring into account. For $\widehat{S}$ we will use the estimator introduced in Dabrowska (1988).

3.1.3 Dabrowska's estimator

Denote by $\widehat{S}^m$ and $\widehat{S}^f$ the Kaplan-Meier estimates of the univariate survival functions of T_x^m and T_y^f, and, for $i \in \{1, .., n\}$, let δ_{1i} and δ_{2i} be the indicators of the event that observations $x_{(i)}$ and $y_{(i)}$, respectively, will be uncensored. Furthermore, define

$$\begin{aligned}
\widehat{H}(s,t) &= \frac{1}{n}\#\left\{i \left| x_{(i)} > s, y_{(i)} > t\right.\right\};\\
\widehat{K}_1(s,t) &= \frac{1}{n}\#\left\{i \left| x_{(i)} > s, y_{(i)} > t, \delta_{1i} = 1, \delta_{2i} = 1\right.\right\};\\
\widehat{K}_2(s,t) &= \frac{1}{n}\#\left\{i \left| x_{(i)} > s, y_{(i)} > t, \delta_{1i} = 1\right.\right\};\\
\widehat{K}_3(s,t) &= \frac{1}{n}\#\left\{i \left| x_{(i)} > s, y_{(i)} > t, \delta_{2i} = 1\right.\right\},
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\Lambda}_{11}(s,t) &= \int_{u=0}^{s}\int_{v=0}^{t} \widehat{K}_1(du,dv) \Big/ \widehat{H}\left(u^-, v^-\right);\\
\widehat{\Lambda}_{10}(s,t) &= -\int_{u=0}^{s} \widehat{K}_2(du,t) \Big/ \widehat{H}\left(u^-, t\right);\\
\widehat{\Lambda}_{01}(s,t) &= -\int_{v=0}^{t} \widehat{K}_3(s,dv) \Big/ \widehat{H}\left(s, v^-\right).
\end{aligned}$$

Dabrowska's estimator is:

$$\widehat{S}(s,t) = \widehat{S}^m(s)\,\widehat{S}^f(t) \prod_{\substack{0<u\leq s\\ 0<v\leq t}} \left(1 - L(\triangle u, \triangle v)\right), \tag{8}$$

with

$$L(\triangle u, \triangle v) = \frac{\widehat{\Lambda}_{10}(\triangle u, v^-)\,\widehat{\Lambda}_{01}(u^-, \triangle v) - \widehat{\Lambda}_{11}(\triangle u, \triangle v)}{\left(1 - \widehat{\Lambda}_{10}(\triangle u, v^-)\right)\left(1 - \widehat{\Lambda}_{01}(u^-, \triangle v)\right)}, \tag{9}$$

with $\triangle u = u - u^-$, and $\triangle v = v - v^-$. Then $\widehat{\Lambda}_{11}(\triangle u, \triangle v)$ 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). The principle of equation (9) 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

$$F(s,t) = \prod_{\substack{0<u\leq s\\ 0<v\leq t}} \left(1 - L(\triangle u, \triangle v)\right), \tag{10}$$

as the multiplier by which the joint survival function differs from the one under independence (see equation (8)). It follows that positive association is implied if $F(s,t) \geq 1$.

3.2 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. This version can be compared with the empirical one for each copula, under censored data, and provides a corresponding best fit selection criterium among different copulas. As is known, the original procedure in Genest and Rivest (1993) for Archimidean copula selection consists in

1) determining - for each candidate copula - the parameter value $\hat{\theta}$ which corresponds to a (common) estimate $\hat{\tau}$ of the Kendall's tau coefficient, by working the parameter out of the relationship

$$\hat{\tau} = 4 \int_{\xi}^{1} \lambda(v) dv + 1 \tag{11}$$

where $\lambda(v)$ is given by (6);

2) building - again for each copula - a theoretical K, $K_{\phi_{\widehat{\theta}}}$, by substituting in (5), for a given generator, the estimate $\hat{\theta}$;

3) selecting as best fit copula the one whose theoretical K is the least distant - according to the L^2 or other norms - from the empirical one, $\widehat{K}_n$.

This procedure is appropriate for complete data, but 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). It is, however, not applicable when, as in our case, both observations can be censored. This is due to the fact that a consistent estimator for Kendall's tau does not exist in the latter case. Therefore, we adopt the modified Wang and Well's 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 and the empirical K, namely $K_{\phi_{\widehat{\theta}}}$ and $\widehat{K}_n$,

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, inverting the relationships used sub 1) above.

In symbols, at stage 1') we define $K_{\phi_{\widehat{\theta}}}(v) = v \square \lambda_{\phi_{\widehat{\theta}}}(v)$, and 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, $\widehat{K}_n$. 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:

$$error\,(\phi_\theta) = \int_{\xi}^{1} \left(K_{\phi_\theta}(v) \square \widehat{K}_n(v) \right)^2 dv. \tag{12}$$

Therefore

$$\widehat{\theta} = \arg\min_{\theta} \int_{\xi}^{1} \left(K_{\phi_\theta}(v) \square \widehat{K}_n(v) \right)^2 dv \tag{13}$$

In turn, the lower bound for the computation of the error, ξ, will 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 is positive:

$$\xi = \min\left\{\nu : K(\nu) > 0\right\} \tag{14}$$

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

At stage 2'), we select the copula which minimizes the (minimum) error:

$$error\left(\phi_{\widehat{\theta}}\right) = \int_{\xi}^{1} \left(K_{\phi_{\widehat{\theta}}}(v) - \widehat{K}_n(v)\right)^2 dv. \tag{15}$$

As a robustness check[1], we suggest double checking the result with another distance definition. A natural candidate is the distance of the sup norm, namely:

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

At stage 3'), we get the corresponding dependence measure by using, in correspondence to the best fit copula, the general relationship (11), which, for the estimated values, becomes

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

3.3 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, Luciano, Vecchiato (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:

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

where $(u_i, v_i) = \left(\widehat{S}^m(x_i), \widehat{S}^f(y_i)\right)$, $C_\theta(u_i, v_i)$ is the copula under consideration, $c_\theta(u_i, v_i)$ its density (i.e. the derivative with respect to both arguments) and $\delta_{1i}\ \delta_{2i}$ are as defined in section (8). Note that this procedure could also be applied to non Archimedean copulas; it leads to

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

[1] We do not provide a formal test of the hypothesis that the resulting copula is the population one, since the bootstrap methodology would be based on a variance estimate, the Wang and Wells' one, which has been proved by Genest, Quessy and Rémillard (2006) not to be valid. We thank B. Rémillard for having signalled to us this limit of the formal test.

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

Similarly to the Wang and Well's method, also the omnibus relies on empirical margins. Both therefore guarantee independency of the copula selection from the margin representation. We now turn to the margin selection procedure.

4 Marginal stochastic mortality

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 this phenomenon is provided by Cairns et al. (2005), 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. In order to define it appropriately, in what follows we briefly describe the doubly stochastic approach to mortality modelling. Then we summarize some previous findings, which justify the modelling choice for the intensity made in this paper.

4.1 Theoretical framework

4.1.1 Cox processes

Following Lando (1998, 2004), let us assume a complete probability space $(\quad , \mathcal{F}, \mathbb{P})$, a process X_t of $\mathbb{R}^d$ -valued state variables $(t \leq T)$ and the filtration $\{\mathcal{G}_t : t \geq 0\}$ of sub-σ-algebras of $\mathcal{F}$ generated by X, i.e. $\mathcal{G}_t = \sigma\{X_s; 0 \leq s \leq t\}$, satisfying the usual conditions.

Let Λ be a nonnegative measurable function s.t. $\int_0^t \Lambda(X_s)ds < \infty$ almost surely and define the first jump time of a nonexplosive adapted counting process N_t as follows:

$$\tau = \inf \left\{ t : \int_0^t \Lambda(X_s)ds \geq E_1 \right\} \tag{16}$$

where E_1 is an exponential random variable with unit parameter. In addition, let us consider the enlarged filtration $\mathcal{F}_t$, generated by both the state variable and the jump processes:

$$\begin{aligned} \mathcal{F}_t &= \mathcal{G}_t \vee \mathcal{H}_t, \\ \mathcal{H}_t &= \sigma\{N_s; 0 \leq s \leq t\} \end{aligned}$$

and assume that the $\mathcal{H}_0$ filtration is trivial, in that no jump occurs at time 0. Under this construction, the process N_t is said to admit the intensity $\Lambda(X_s)$, if the compensator of N_t admits the representation $\int_0^t \Lambda(X_s)ds$, i.e. if

$$M_t = N_t \square \int_0^t \Lambda(X_s)ds$$

is a local martingale. If the stronger condition $\mathbb{E}\left(\int_0^t \Lambda(X_s)ds\right) < \infty$ is satisfied, $M_t = N_t \square \int_0^t \Lambda(X_s)ds$ is a martingale.

Intuitively, this implies that, given the history of the state variables up to time t, the counting process is "locally" an inhomogeneous Poisson process, which jumps according to the intensity $\Lambda(X_t)$:

$$\mathbb{E}(N_{t+\Delta t} \square N_t|\mathcal{G}_t) = \Lambda(X_t)\Delta t + o(\Delta t).$$

Formally, the construction (16) implies that the survival function of the first jump time τ, evaluated at time 0, and conditional on knowledge of the state process up to time t, is

$$\Pr(\tau > t|\mathcal{G}_t) = \exp\left(\square \int_0^t \Lambda(X_s)ds\right)$$

where $\Pr(.)$ is the probability associated to the measure $\mathbb{P}$. It can also be shown, by simple conditioning, that the time 0 unconditional survival probability, which we will denote as $S(t)$, is

$$S(t) = \Pr(\tau > t) = \mathbb{E}\left[\exp\left(\square \int_0^t \Lambda(X_s)ds\right)\right]. \tag{17}$$

The unconditional probability at any date t' greater than 0 can be shown to be

$$\Pr(\tau > t \mid \mathcal{F}_{t'}) = \mathbb{I}_{\{\tau > t'\}}\mathbb{E}\left[\exp\left(\square \int_{t'}^t \Lambda(X_s)ds\right) \mid \mathcal{G}_{t'}\right]$$

where $\mathbb{I}_{\{\tau > t'\}}$ is the indicator function of the event $\tau > t'$.

A nonexplosive counting process N_t constructed as above is said to be a *Cox* or *doubly stochastic process driven by* $\{\mathcal{G}_t : t \geq 0\}$. The corresponding first jump time is *doubly stochastic* with intensity $\Lambda(X_s)$.

As a particular case, any Poisson process is a doubly stochastic process driven by the filtration $\mathcal{G}_t = (\emptyset, \ ) = \mathcal{G}_0$ for any $t \geq 0$, in that the intensity is deterministic.

These results can be naturally applied in the actuarial domain: if τ is the future lifetime of a head aged x, T_x, his/her survival function, $S_x(t)$, is

$$S_x(t) = \Pr(T_x > t) = \mathbb{E}\left[\exp\left(\square \int_0^t \Lambda(X_s)ds\right)\right]. \tag{18}$$

4.1.2 Affine processes

In general, the expectations (17) and (18) are not known in closed form: however, a remarkable exception is the case in which the dynamics of X is given by the SDE:

$$dX(t) = f(X(t))dt + g(X(t))d\tilde{W}(t) + dJ(t),$$

where $\tilde{W}$ is an n-dimensional Brownian motion, J is a pure jump process, and, above all, the drift $f(X(t))$, the covariance matrix $g(X(t))g(X(t))'$ and the jump measure associated with J have affine dependence on $X(t)$. Such a process is named an affine process, and a thorough treatment of these processes is in Duffie et al. (2003).

The convenience of adopting affine processes in modelling the intensity lies in the fact that, under technical conditions, it yields:

$$S_x(t) = \mathbb{E}\left[e^{\int_0^t - \Lambda(X(s))ds}\right] = e^{\alpha(t)+\beta(t)\Lambda(X(0))}, \tag{19}$$

where the coefficients $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy generalized Riccati ODEs (see for instance Duffie et al., 2000). The latter can be solved at least numerically and in some cases analytically. Therefore, the problem of finding the survival function becomes tractable, whenever affine processes for X are employed.

4.2 Selection of the intensity

In the existing actuarial literature, the Λ function has been chosen to be the identity, that is the mortality intensity is the direct driving force of the double counting process, and different classes of affine processes have been chosen for it. For example, Milevsky and Promislow (2001) investigate a so-called mean reverting Brownian Gompertz specification: the intensity h_t is given by

$$h_t = h_0 e_t^{gt+\sigma \int_0^t e^{-b(t-u)dW_u^h}},$$

with g, σ, b constant and the Brownian motion W uni-dimensional.

Dahl (2004) selects an extended Cox-Ingersoll-Ross (CIR) process, i.e. a time-inhomogeneous process μ, reverting to a deterministic function of time

$$d\mu_{x+t} = (\beta^{\mu}(t,x) - \;^{\mu}(t,x)\mu_{x+t})dt + \rho^{\mu}(t,x)\sqrt{\mu_{x+t}}dW_t,$$

where x is the initial age.

Biffis (2005) chooses two different specifications for the intensity process. In the first one, the intensity μ_t is given by a deterministic function of time, $m(t)$, plus a mean reverting jump diffusion process Y_t, with dynamics given by the SDE

$$dY_t = \;(\overline{y}(t) - Y_t)dt + \sigma dW_t - dJ_t.$$

In the second one, which is a two factor model, the intensity μ_t is a CIR-like process, mean reverting to another process $\overline{\mu}_t$. The dynamics of the two processes are given by

$$\begin{aligned} d\mu_t &= \;_1(\overline{\mu}_t - \mu_t)dt + \sigma_1\sqrt{\mu_t}dW_t^1 \\ d\overline{\mu}_t &= \;_2(m(t) - \overline{\mu}_t)dt + \sigma_2\sqrt{\overline{\mu}_t - m^*(t)}dW_t^2. \end{aligned}$$

Schrager (2006) proposes an M-factor affine mortality model, whose general form is given by

$$\mu_x(t) = g_0(x) + \sum_{i=1}^{M} Y_i(t)g_i(x),$$

where the factors Y_i are mean reverting.

Luciano and Vigna (2005) explore the following models: an Ornstein Uhlenbeck, a mean reverting with jumps and a CIR process as concerns the mean-reverting group, a Gaussian and a non Gaussian Feller type process without mean reversion, but with and without jumps, as concerns the non-mean reverting set.

Among the one-factor models, Biffis (2005) fits his mean reverting time inhomogeneous intensity to some Italian mortality tables, while Luciano and Vigna (2005) calibrate their time-homogeneous, simpler processes to the Human Mortality database for the UK population. In doing the calibration, they assume negative jumps, so as to incorporate sudden improvements in non-diversifiable mortality. As a whole, they show that, among time-homogeneous diffusion and jump diffusion processes, the ones with constant drift "beat" the ones with mean reversion, as descriptors of population mortality. Both the fit and the predictive power of the non mean reverting processes - when they are used for mortality forecasting within a given cohort - are very satisfactory, in spite of the analytical simplicity and limitations of the theoretical models. Among them, no one seems to outperform the others. Moreover, for different generations, different estimates of parameters are obtained: this confirms that generation effects cannot be ignored.

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

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

where $a_x > 0$ and $\sigma_x \geq 0$. The corresponding survival probability[2] is given by (19), with $\Lambda(X) = \mu_x$, i.e.

$$S_x(t) = \mathbb{E}\left[e^{\int_0^t -\mu_x(s)ds}\right] = e^{\alpha_x(t)+\beta_x(t)\mu_x(0)}, \tag{20}$$

where, omitting the dependence on the cohort or generation x for simplicity

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

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

The parameters a and σ can be obtained either from mortality tables, or, as we will do below, on sample, censored data. In both cases they can be calibrated by minimizing the mean squared error between the theoretical and actual probabilities: 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.

[2] 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 $\sigma^2 - 2dc < 0$.

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). The original data set concerns 14,947 contracts in force with a large Canadian insurer. The period of observation runs from December 29, 1988, until December 31, 1993. Like the aforementioned papers, we have eliminated same-sex contracts (58 in total). Besides, like Youn and Shemyakin (1999, 2001), for couples with more than one policy, we have eliminated all but one contracts (3,435 contracts). This has left us with a set of 11,454 married couples and contracts.

Since, as explained above, the methodology for the marginal survival functions applies to single generations, 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 women in the sample, obtained after eliminating same sex and double contracts, is 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 for the estimate of the marginal survival functions. Then, in order to estimate joint survival probabilities, we have further concentrated on the couples whose members belong to the generation 07-20 for males and 10-23 for females. This subset includes a total of 3,931 couples. Both individuals and couples are observable for nineteen years, because they were born during a fourteen year period and the observation period is five years. In focusing on a generation and allowing for the three-year age difference, we have considered only one illustrative example; however, the procedure can evidently be repeated for any other couple of generations.

On the chosen generation, we adopt the general procedure sketched in Section 4 for the margins and the one in Section 3 for the joint survival function.

We first obtain the empirical margins, using the Kaplan-Meier methodology. These margins feed the Dabrowska estimate for the empirical joint survival function. Starting from it, the best fit analytical copula is estimated using the Wang and Wells (2000b) method. Like Denuit et al. (2004), we perform a check of the parameters and of the best fit choice using the omnibus procedure.

The marginal Kaplan-Meier data are used also as inputs for the calibration of the analytical marginal survival functions, according to the methodology in Luciano and Vigna (2005).

The final step of the calibration procedure involves obtaining the joint analytical survival function from the best fit copula and the calibrated margins.

5.2 Kaplan-Meier estimates of marginal survival functions

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

	MALES	FEMALES
t	${}_tp_{68}$	${}_tp_{65}$
1	0.972253	0.9877123
2	0.96103	0.9818795
3	0.938278	0.977377
4	0.913871	0.970495
5	0.89417	0.9646967
6	0.869726	0.9572001
7	0.845971	0.947749
8	0.815979	0.9322838
9	0.783494	0.9199416
10	0.758918	0.9073177
11	0.730908	0.8941103
12	0.696391	0.8814861
13	0.657758	0.8654661
14	0.603822	0.8494678
15	0.557302	0.829017
16	0.518074	0.7921956
17	0.483845	0.7559616
18	0.401803	0.7205523
19	0.331582	0.6826285

Table 2

We notice that, differently from both Carriere (2000) and Frees et al. (1996), we can calculate the empirical survival probabilities ${}_tp_x$ only until $t = 19$. This is due to the limited range of birth dates of our generations, coupled with the five year length of observation. Based on the explanation above, we take the initial age x to be 68 for males, 65 for females.

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.

In Table 3 we present the multipliers $F(s,t)$, as defined in equation (10). As usual with censoring, due to the time frame of observation of five years, we cannot explicitly compute the multipliers for durations greater than five: 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.

We notice that all the multipliers are greater than one. This indicates positive association and confirms our intuition about the dependency of the lifetimes of couples. Later on, we will provide an appropriate measure (Kendall tau) of the amount of association.

F-function	0	1	2	3	4	>=5
0	1	1	1	1	1	1
1	1	1.000637	1.000892	1.001329	1.001972	1.002155
2	1	1.001055	1.004109	1.005851	1.006285	1.007077
3	1	1.001509	1.004665	1.00909	1.009978	1.010515
4	1	1.001524	1.004547	1.008826	1.011508	1.012414
>=5	1	1.001966	1.00483	1.009402	1.012536	1.017135

Table 3

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 in turn is used as an input for $\hat{K}$, the empirical version of the K function, according to a discretized version of formula (7). In order to obtain the latter we divide the unit interval into a thousand subintervals[3]. Figure 1 presents the empirical estimate for K, $\hat{K}$.

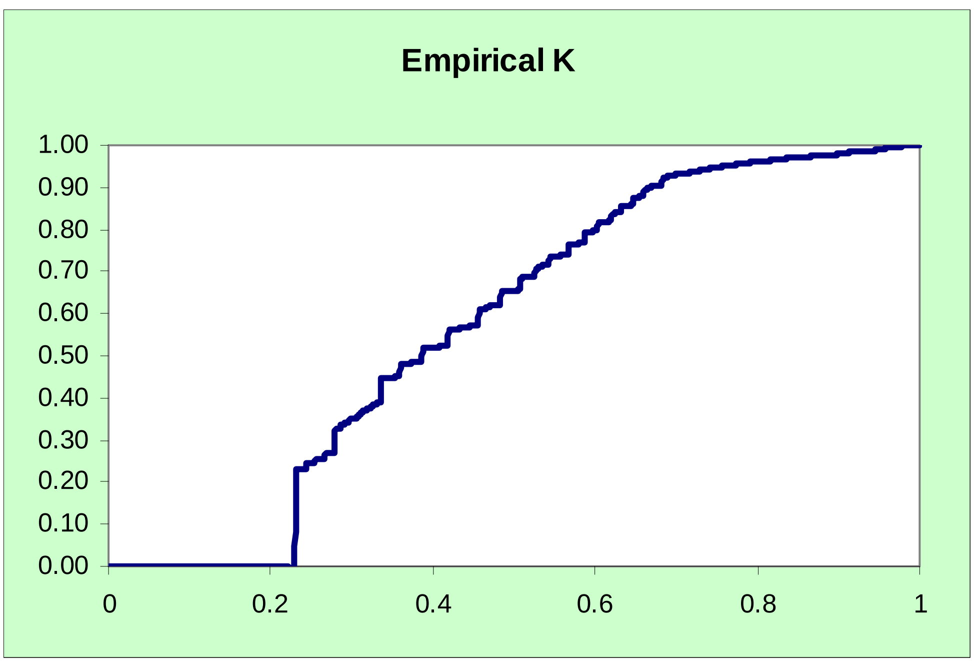

Figure 1

[3]We checked the robustness of the procedure by changing the discretization step.

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 (15). 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 3.2.

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 2, where we present the theoretical K's and the empirical one.

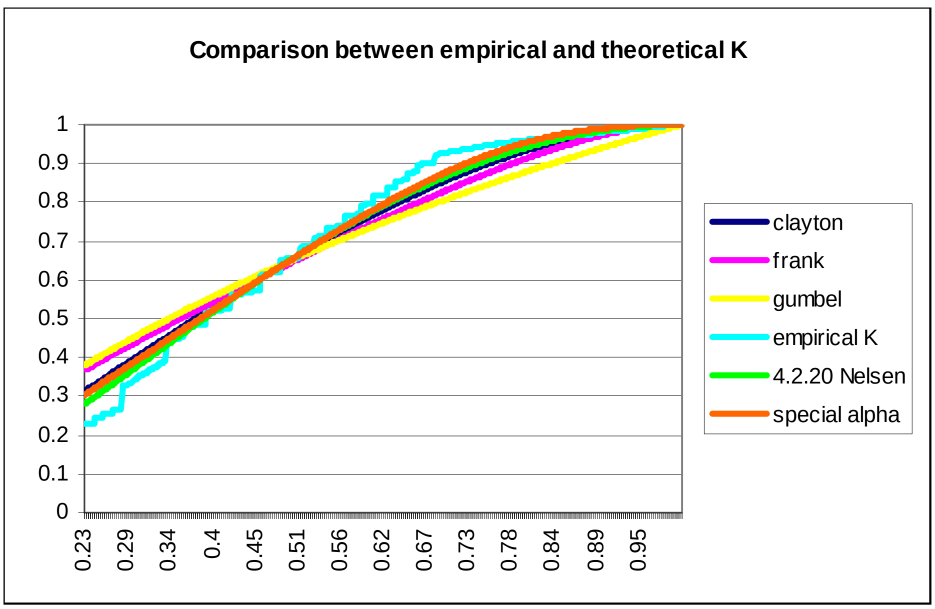

Figure 2

We also compute the distance of each theoretical function from the empirical one, i.e. the minimized distance in (15). This gives the errors in Table 4.

Error (L^2 -norm distance)				
Clayton	Frank	Gumbel	4.2.20 in Nelsen	Special alfa
1.336382	3.095018	4.777058	0.720027337	0.8110124

Table 4

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 Nelsen copula we also get an estimate of Kendall's tau, as explained under 3') of section 3.2: 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, Quessy, Rémillard (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 as best fit copula the Nelsen 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 use again 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.

Copula	Theta via omn. proc.	Standard error	Theta via W&W proc.	Max-likelihood
Clayton	2.2325	0.3290	2.731165	-734.698
Frank	3.4892	0.4154	6.313338	-735.268
Gumbel	1.1292	0.0217	2.2612029	-750.297
4.2.20 Nelsen	1.0402	0.1427	1.004763	-734.573
Special	4.3734	0.42495	3.0966724	-740.396

Table 5

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

Also, the omnibus approach 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 Nelsen or Clayton copula. However, if one repeats the test using the estimated parameters of the sup norm distance, he finds that the Nelsen and Special estimates from the Wang and Well's methodology fall within the maximum likelihood significance bounds: therefore, the Nelsen is the only one which passes the test for both norms.

5.6 The analytical marginal survival functions

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. On the contrary, in this paper we distinguish different generation survival probabilities and intensity processes. We take as a generation not a single age of birth, but thirteen consecutive of them: this assumption is based on the one side on the possibilities of reliable calibration (number of data) offered by the present data set; on the other side, 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. It is evident however that the specific choice adopted here is purely illustrative.

We have chosen the generation 1907-20 for males, initial age 68, and 1910-23 for females, initial age 65. We therefore present only two survival functions, which will be denoted as $S^{m}_{68}(t), S^{f}_{65}(t)$ respectively. Their analytical expression is given by (20). 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, respectively for males and females

$$a_{68} = 0.0810021, \sigma_{68} = 0.00005, a_{65} = 0.124979, \sigma_{65} = 0.00005$$

while the initial intensity values are[4]

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

The two survival functions are presented in Figure 3.

[4] The values of $\mu_{68}(0)$ and $\mu_{65}(0)$, according to Luciano and Vigna (2005), should be □ $\ln(p_{68})$ 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. However, these data are not available. Therefore, using the data set we have estimated with the Kaplan Meier method p_{68} males and p_{65} females, without restrictions on the generation. This has been done in order to have an estimate of those survival probabilities as accurate as possible (also considering the fact that the observation period is only five years, and therefore the individuals entering the calculation of the survival probabilities were born in a six years interval).

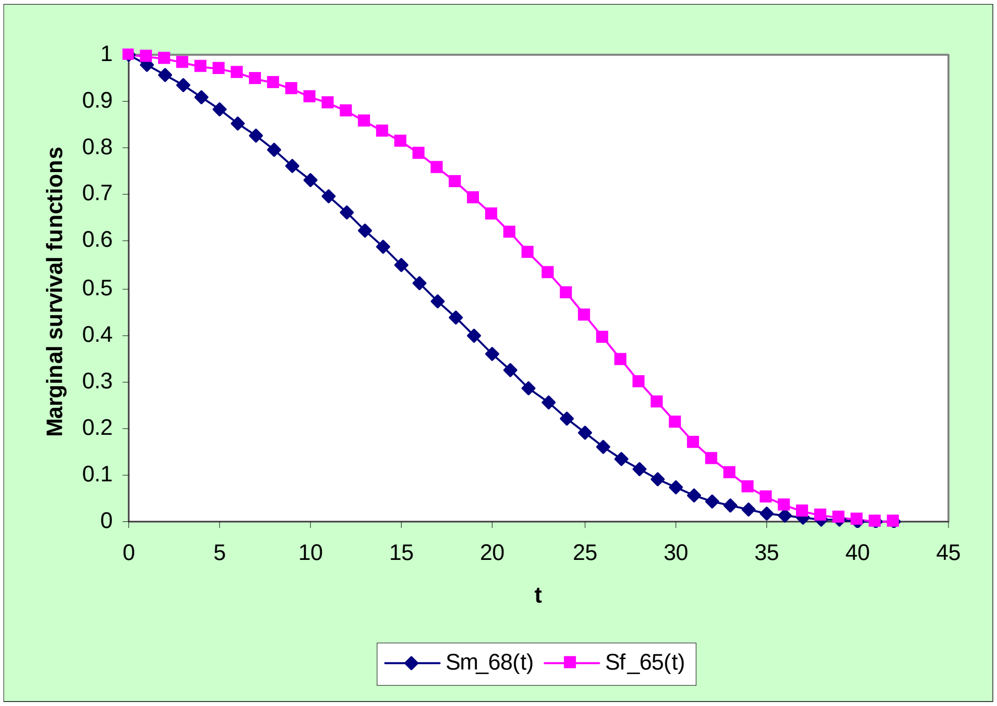

Figure 3

6 The analytical joint survival function and its time-dependent association

We couple the fitted marginal survival functions of Section 5.6 with the best fit copula choice of Section 5.4, according to the formula

$$S(x,y) = C(S_{68}^{m}(x), S_{65}^{f}(y))$$

and using the Nelsen's copula:

$$C_\theta(u,v) = \left[\ln\left(\exp(u^{-\theta}\right) + \exp(v^{-\theta}) - e)\right]^{-\frac{1}{\theta}}$$

By doing so, we obtain the joint survival function $S(x,y)$ of Figure 4, some of whose sections are presented in Figures 5 and 6 respectively

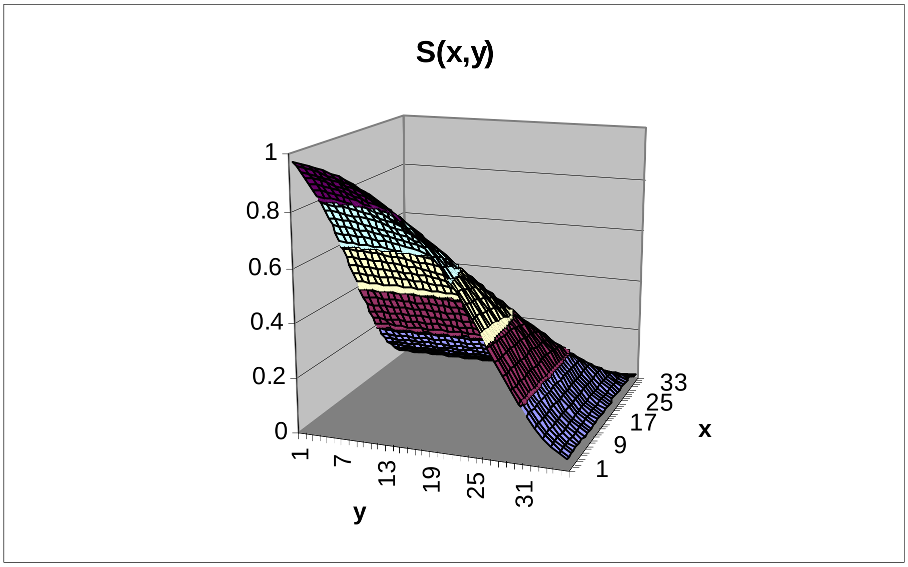

Figure 4

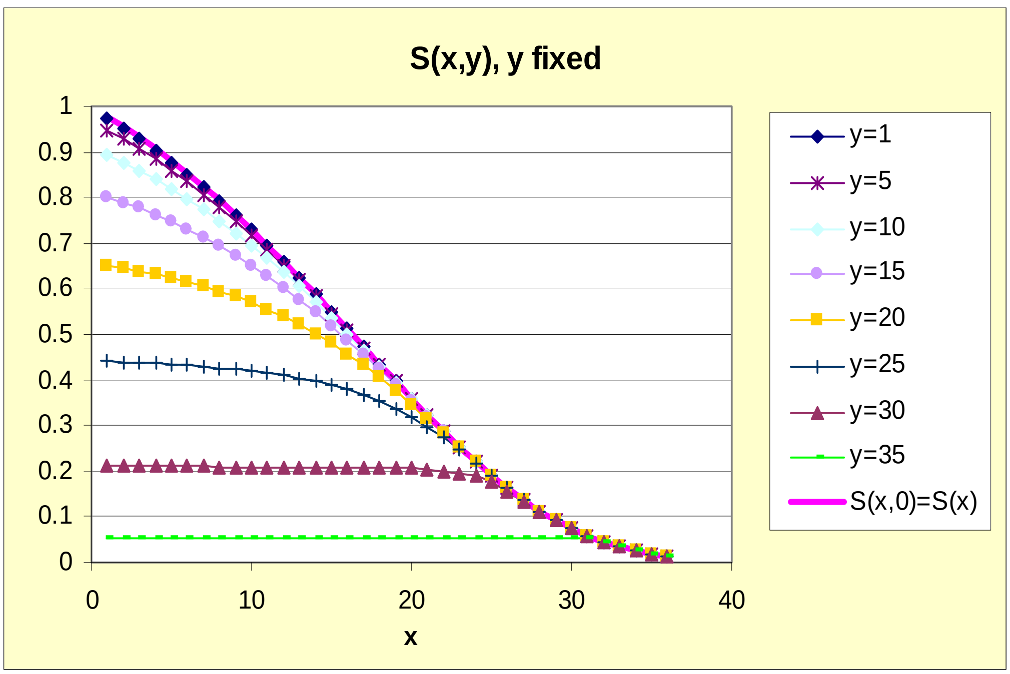

Figure 5

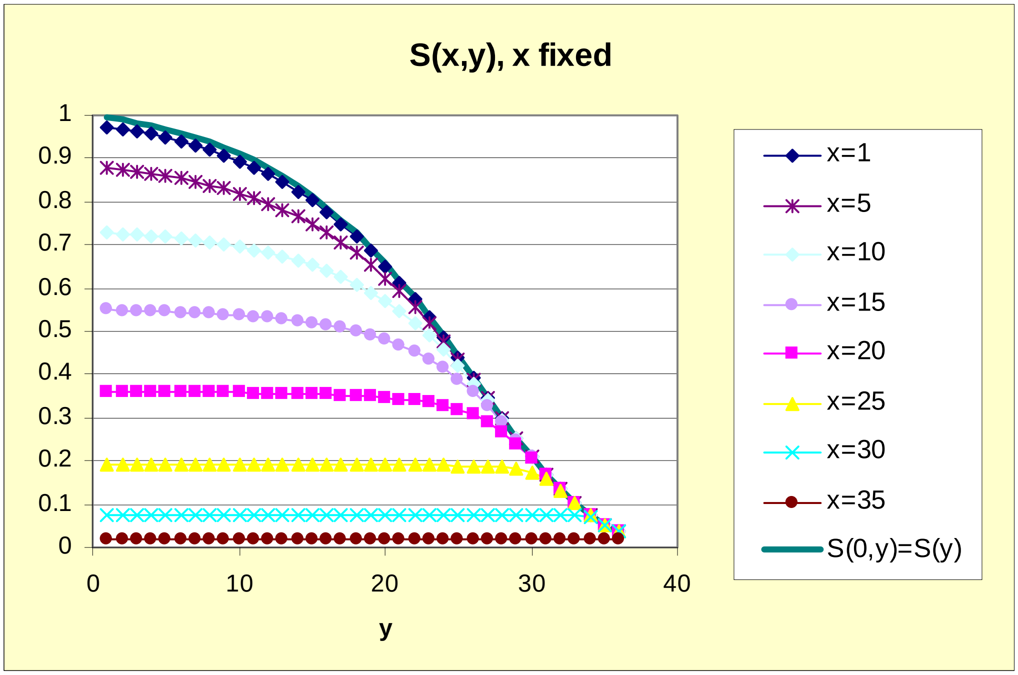

Figure 6

Looking at Figure 5, we notice that , if y is high, $S(x, y)$ is almost flat until a certain age $\widehat{x}$ after which it decreases. This is due to the fact that the probability for the female of surviving y years, with high y, is very low: this affects to a great extent the joint probability of surviving x years for the male and y years for the female (even when the probability $S(x, 0)$ is very high because x is small). After age $\widehat{x}$ the joint probability starts to decrease because of the joint effect of low probability of surviving y years for the female and x years for the male.

For Figure 6 the same comments made for Figure 5 apply. Notice that, while the age $\widehat{x}$ after which $S(x, y)$, y fixed, starts to decrease is always smaller than the fixed value of y, here the age $\widehat{y}$ after which $S(x, y)$, x fixed, starts to decrease is always higher than the fixed value of x. This is probably due to the difference in death rates for a male and a female with the same age. Evidence of this can be also found in the different level of the sections when we change sex: for instance, $S(x, 35)$ lies at a higher level than $S(35, y)$, $S(x, 30)$ lies at a higher level than $S(30, y)$, etc.

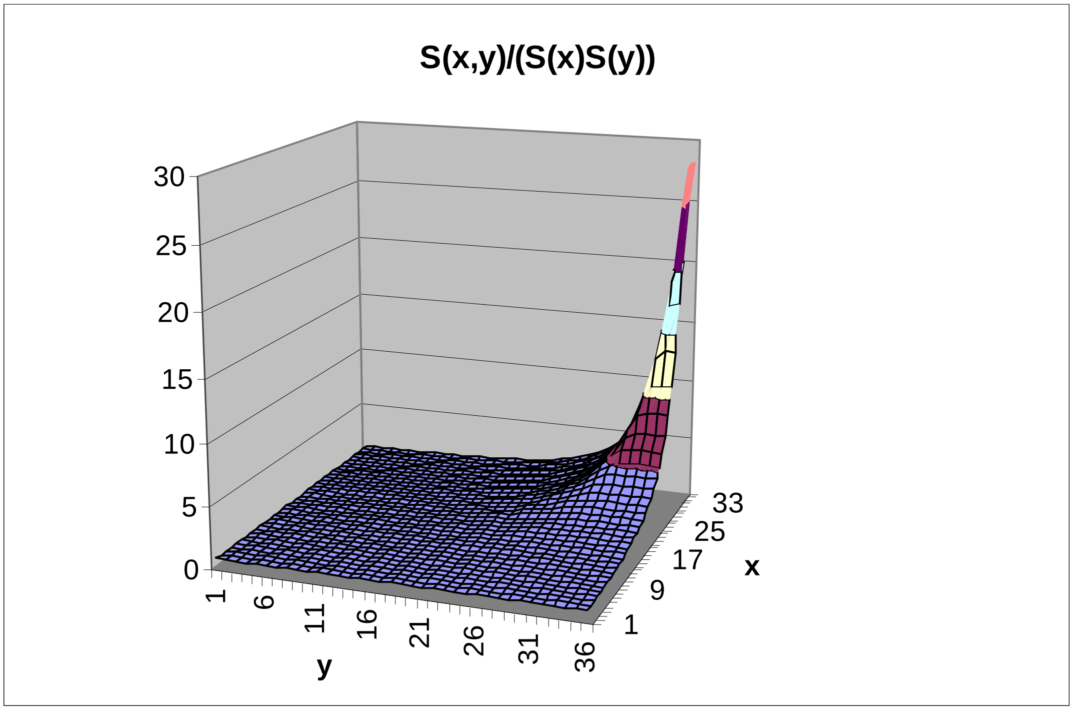

Figure 7

In Figure 7, we report the ratio between the joint survival function $S(x, y)$ and the probability which we would obtain under the assumption of independence, namely the product copula one, $S(x)S(y)$. In doing this, please notice that we use the short notation $S_{68}^{m}(x) = S(x), S_{65}^{f}(y) = S(y)$. Figure 7 therefore reports the time dependent measure of association $\ _1(x, y)$ as defined in (2). The ratio takes values greater than one, because of positive dependence, is monotone in each argument, as expected from the copula selected, and reaches very large values for large x and y.

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

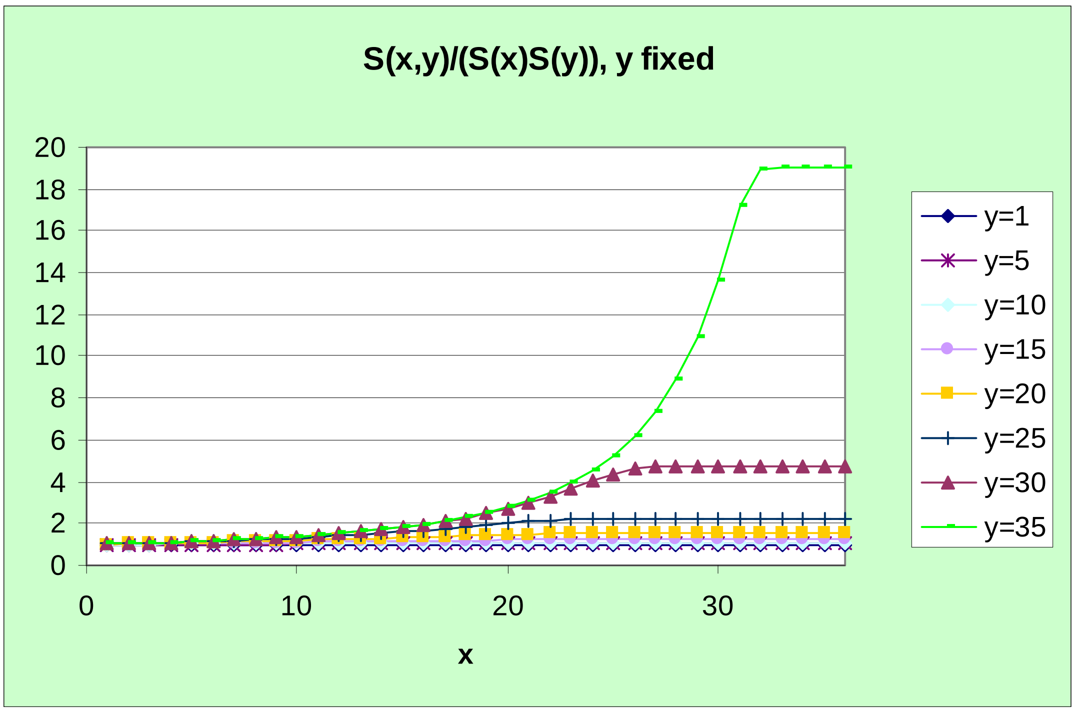

Figure 8

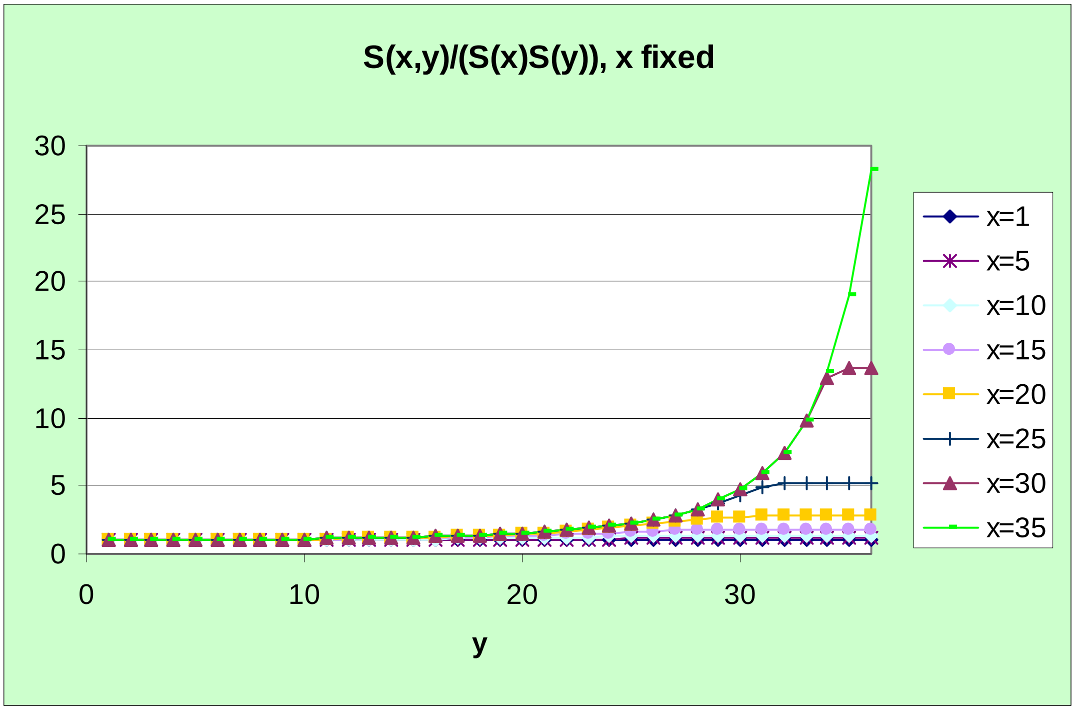

Figure 9

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

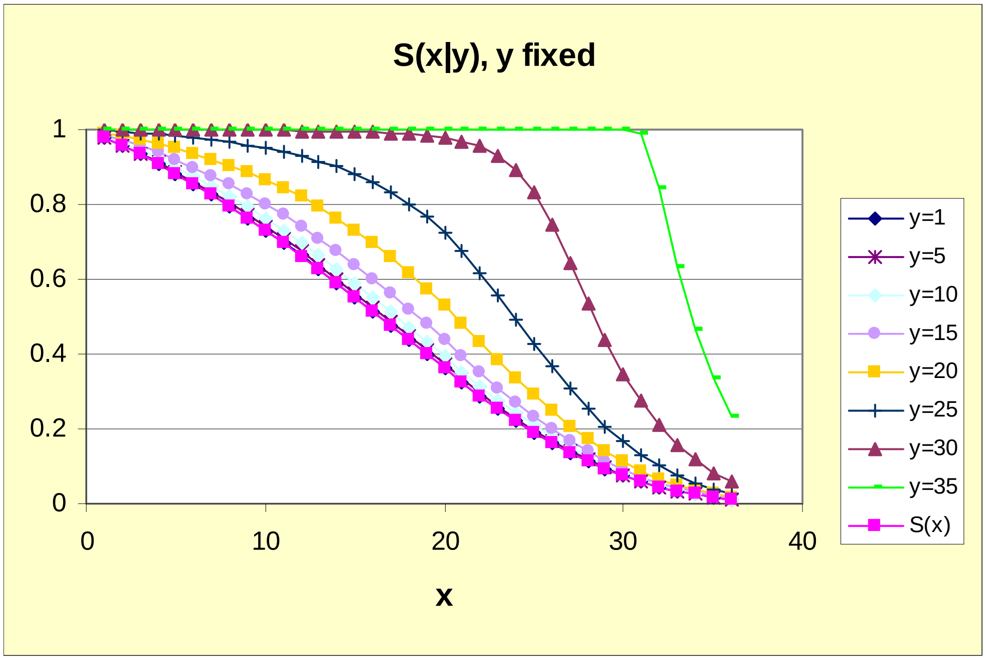

Figure 10

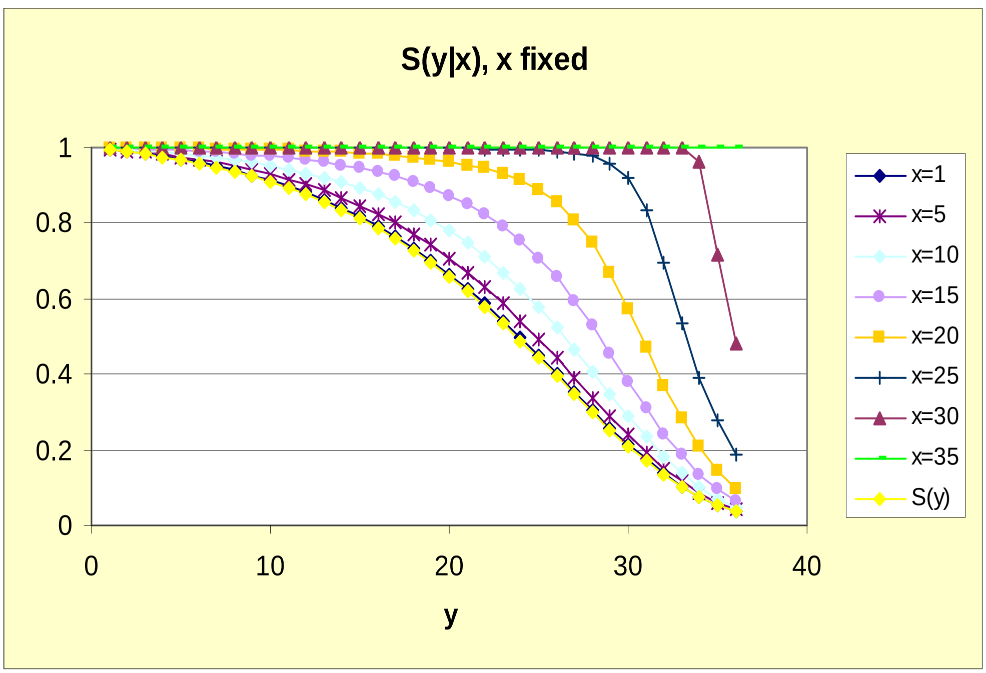

Figure 11

As for the second measure of time-dependent association in Section 2, table 6 illustrates the measures $\ _{2x}(0,y)$ and $\ _{2y}(x,0)$ as defined in equation (3). The unconditional life expectancy $E\left[T_x^m\right]$ and $E\left[T_y^f\right]$ are respectively equal to 16.51 and 21.92. Column 2 displays the relative increase of the conditional expected remaining lifetime of (x), given that (y) survives to y, with respect to $E\left[T_x^m\right]$: as explained in Section 2, in correspondence to our copula, it increases as a function of y. Similarly, column 4 shows the relative increase of the conditional expected remaining lifetime of (y), given that (x) survives to x, with respect to $E\left[T_y^f\right]$: it is increasing as a function of x, as expected. We observe that, for $x = y$, ${}_{2x}(0, y) < {}_{2y}(x, 0)$ for small values of x or y, but the inequality sign is reversed for large values of this argument. 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.

y	E(T_x|T_y>y)/E(T_x)	x	E(T_y|T_x>x)/E(T_y)
1	1.002	1	1.006
5	1.015	5	1.028
10	1.044	10	1.056
15	1.097	15	1.089
20	1.199	20	1.130
25	1.381	25	1.175
30	1.632	30	1.219

Table 6

As for the third measure of time-dependent association in Section 2, the cross-ratio function for the Nelsen copula, as a function of $S(u, u)$, is

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

As the previous measures, and as shown in Spreeuw (2006), it is increasing as a function of age (u): differently from the other measures however it does not depend on the margins. Its measures the relative increase in the survivor force of mortality. Figure 12 gives a plot of $CR(v)$ versus $1 \square v$: notice that $CR(1) = 3.00953$, $CR(v)$ is increasing in $1 \square v$ (as expected from the previous reasoning on u) and takes very large values for v close to 0.

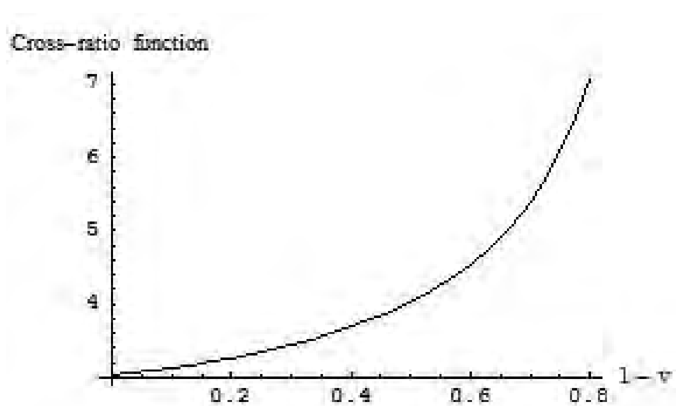

Figure 12

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

7 Conclusions

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

On the theoretical side, we extend the Cox processes setup to couples, where Cox processes are based on the idea that mortality is driven by a jump process whose intensity is itself a stochastic process, proper of a particular generation within a gender. The dependency 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.

On the empirical side, we fit the joint survival function by calibrating separately the (analytical) margins and both calibrating and selecting the best fit (analytical) copula. The calibration of the margins, due to the fact that the individual intensity of mortality in stochastic intensity models is generation dependent, must be performed on a given generation: as an example, we choose two generations which are in their retirement age during the observation period.

First, we parametrize and select the best fit copula in a group of Archimedean ones, according to the methodology of Wang and Wells (2000b) for censored data. We obtain as best fit copula the so-called Nelsen one and we confirm its appropriateness with the pseudo maximum likelihood or omnibus procedure. The best copula is far from representing independence: this confirms both intuition and the results of all the existing studies on the same data set. In addition, since the best fit copula is the Nelsen one, dependency is increasing with age.

Then, we provide a calibration of the marginal survival functions of males and females. We select time-homogeneous, non mean-reverting, affine processes for the intensity and give the corresponding survival functions in analytical form. Differently from Luciano and Vigna (2005), we base the calibration on sample insurance data and not on mortality tables.

Coupling the best fit calibrated copula with the calibrated margins we obtain a joint survival function which is fully analytical and therefore can be extended, for the chosen generation, to durations longer than the observation period. This permits to compute time dependent association measures.

The main contribution of the paper is in the selection of a joint survival function which incorporates stochastic future mortality for both individuals in a couple, and which is analytically tractable. The approach seems to be manageable and flexible, and lends itself to extensive applications for pricing and reserving purposes. These are in the agenda for future research.

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

[5] Cairns, A. J. G., Blake, D., and Dowd, K. (2005). 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*, 17-31.

[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. *The 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., 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.

[12] 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.

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

[14] Duffie, D., Pan, J. and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions, *Econometrica*, 68, 1343–1376.

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

[16] 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.

[17] 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.

[18] 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.

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

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

[21] Lando, D. (2004). *Credit Risk Modeling*, Princeton: Princeton Univ. Press.

[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, *Submitted.*

[23] Manatunga, A.K. and Oakes, D. (1996). A measure of association for bivariate frailty distributions. *Journal of Multivariate Analysis* 56, 60-74.

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

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

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

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

[28] 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.

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

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

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

[32] 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.

[33] 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.

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