Pricing Lifelong Joint Annuity Insurances and Survival Annuity Insurances Using Copula Modeling of Bivariate Survival

Olof Werneman

Abstract

This thesis investigates the dependence in survival of married couples and its effect on the pricing of lifelong joint annuity insurances and survival annuity insurances. The dataset used contained 199127 observations of married couples where both husband and wife reached the age of 61 or more and the maximum difference in age between husband and wife was ten years. The Makeham distribution is proved to be a very good choice to model the marginal distributions. The hypothesis of independent survival for married couples is rejected at all reasonable levels of significance. The data show a weak positive dependence. The dependence structure is shown to best fit a Clayton copula model. For the lifelong joint annuity insurance the Clayton copula-model yields 1.6% lower discounted expected future payments than, when assuming independence. When pricing a survival annuity insurance it is important to distinguish between the case when the husband is insured and wife is co-insured and the case when the wife is insured and the husband is co-insured. If the ages of the husband and wife are approximately equal the Clayton copula-model yields 8% lower discounted expected future payments in the first case and 10% lower in the second case. In the first case, the difference expressed as a percentage, between the discounted expected future payments of the Clayton copula-model and the independent model is decreasing as the difference in age increases from -10 to 10. In the second case the difference expressed as a percentage is increasing.

Acknowledgements

I would like to thank Bengt von Bahr at The Premium Pension Authority for taking time to discuss problems with me, providing me with an office and financing the data from Statistics Sweden. I also would like to thank Prof. Boualem Djehiche at KTH for supervising my work.

Stockholm, January 2005

Olof Werneman

Contents

Chapter 1

Introduction

Lifelong joint annuity insurances pay monthly pension payments to the husband and wife as long as at least one of the persons is alive. Survival annuity insurances pay monthly pension payment from the time of death of the insured until the time of death of the co-insured. Traditionally, when pricing these insurances the survivals of the husband and wife are assumed to be independent. This thesis will investigate if there is any significant dependence structure in the survival and its effect on pricing of joint annuity insurances and survival annuity insurances. The dependence structure will be modeled using copulas. A copula is a function that links the marginal distributions to the joint distribution.

The reliability of the results in the analysis are effected by the limits provided by the data. The data concerns married couples in the Swedish population and not insured couples. The results of the study do however agree with other research made on insured couples, e.g. Carriere [1]. Moreover the maximum observed age of death is 94 years old. This should not have a crucial effect on the choice of copula since the dependence tends to decrease at high ages of death. The estimated parameters of the marginal distributions might however be effected causing too light tails.

Extensive research in the area of bivariate survival models include:
Carriere [1] studies the dependence between time of death of couples. Carriere suggests Gompertz marginal distributions and a linear mixture of Clayton and independent copula. Hougaard et al [8] analyzes the dependence in lifetimes and compare different distributions including Positive stable distribution and Gamma distribution. The dependence is, among others, measured by Kendall's tau and the data shows a weak positive dependence. Maguluri [11] analyzes the Clayton model for bivariate survival (the Clayton copula), by comparing the performance of different methods for estimating the copula parameter. Oakes [14] discusses a reparameterization of the Clayton copula, criticizes the by Clayton proposed likelihood method for estimating the copula parameter and suggest an alternative non-parametric estimator

based on Kendall's tau. Oakes [15] studies the model where two observed survival times depend via, a proportional-hazards model, on the same unobserved variable, called a frailty. The class of distributions arised in that way is treated. Shih and Louis [16] investigates two-stage parametric and two-stage semi-parametric estimation procedures for the copula parameter where censoring in either one or both of the components is allowed. The proposed methods are applied in an AIDS dataset. Zheng and Klein [17] assumes a known copula and introduces nonparametric estimators of the marginal survival functions. Further, bounds on the survival functions are provided when the copula only is known approximately.

Chapter 2

Dependence Structure

2.1 Copulas

Every joint distribution function for a random vector contains both description of the marginal behavior of the individual variables and a description of their dependence structure. In this section I will describe how copulas provide a way of isolating the description of the dependence structure. This is the mayor advantages of using a copula-approach when fitting data to a multivariate model; the marginal distributions can be fitted separately from the dependence structure. In this way copulas provide a way of constructing a wide variety of multivariate models. Copulas also help in the understanding of dependence at a deeper level.

Definition 2.1 *A two-dimensional copula is a distribution function on* $[0,1]^2$ *with standard uniform marginal distributions.*

The following three properties must hold:

1. $C(u_1, u_2)$ is increasing in each component u_i.
2. $C(1, u_i) = u_i$ for $i \in \{1, 2\}, u_i \in [0, 1]$.
3. For all $(a_1, a_2), (b_1, b_2) \in [0, 1]^2$with $a_i \leq b_i$ We have:

$$\sum_{i_1=1}^{2} \sum_{i_2=1}^{2} (-1)^{i_1+i_2} C(u_{1i_1}, u_{2i_2}) \geq 0,$$

   where $u_{j1} = a_j$ and $u_{j2} = b_j$ for $j \in \{1, 2\}$.

The first two properties are clear. The third property is required of any multivariate distribution function and ensures that if the random vector $(U_1, U_2)^T$ has distribution function C then $P(a_1 \leq U_1 \leq b_1, a_2 \leq U_2 \leq b_2)$ is non-negative. (Embrechts et al [4]).

2.1.1 Sklar's Theorem

Theorem 2.2 (Sklar's theorem) *Let F be a joint distribution function with margins F_1, F_2. Then there exists a copula C: $[0,1]^2 \to [0,1]$ such that for all x_1, x_2 in $\bar{\mathbf{R}} = [-\infty, \infty]$*

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \tag{2.1}$$

If the margins are continuous then C is unique; otherwise C is uniquely determined on $RanF_1 \times RanF_2$, where $RanF_i$ denotes the range of $F_i : RanF_i = F_i(\bar{\mathbf{R}})$. Conversely, if C is a copula and F_1, F_2 are distribution functions, then the function F defined in (2.1) is a joint distribution function with margins F_1, F_2.

(Nelsen [13]). Sklar's theorem says that the copula is a function that links the joint distribution function $F(x_1, x_2)$ to the marginal distributions $F_1(x_1), F_2(x_2)$. Another way of express this is through the following proposition

Proposition 2.3 *Let F be a joint distribution function with continuous margins F_1, F_2. Then the unique copula of F is given by*

$$C(u_1, u_2) = F(F_1^{\leftarrow}(u_1), F_2^{\leftarrow}(u_2))$$

where $F^{\leftarrow}$ denotes the generalized inverse.
(Embrechts et al [4]).

2.1.2 The Fréchet-Hoeffding Bounds

Consider the functions M, Π and W defined on $[0,1]^2$ as follows:

$$\begin{aligned} M(\mathbf{u}) &= min(u_1, u_2), \\ \Pi(\mathbf{u}) &= u_1 u_2, \\ W(\mathbf{u}) &= max(u_1 + u_2 - 1, 0). \end{aligned}$$

(Embrechts et al [5]).
The following theorem is called the Fréchet-Hoeffding bounds inequality.

Theorem 2.4 *If C is any copula, then for every $\boldsymbol{u}$ in $[0,1]^2$,*

$$W(\mathbf{u}) \le C(\mathbf{u}) \le M(\mathbf{u}).$$

Theorem 2.5 *Let $(X_1, X_2)^T$ be a vector of continuous random variables with copula C. Then X_1, X_2 are independent if and only if $C = \Pi$.*

Definition 2.6 *If $(X, Y)^T$ has the copula M then X and Y are said to be comonotonic; if it has the copula W they are said to be countermonotonic.*

2.1.3 Copula Densities

Copulas do not always have joint densities, the comonotonicity and countermonotonicity copulas are examples of copulas which are not absolutely continuous. However, most parametric copulas have densities given by

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}. \tag{2.2}$$

2.2 Dependence

When measuring dependence the most commonly used measure in the linear correlation. In this section I will explain some of the drawbacks of using the linear correlation and present the alternative dependence measure *rank correlation*.

2.2.1 Linear Correlation

Definition 2.7 *Let $(X, Y)^T$ be a vector of random variables with nonzero finite variances. The linear correlation coefficient for $(X, Y)^T$ is*

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

The linear correlation is a measure of the linear dependence talking values in [-1,1]. If X and Y are independent $\rho(X, Y) = 0$. An important property is that the converse is not true; $\rho(X, Y) = 0$ does not necessarily imply their independence. Consider for an example the case when $X \sim N(0, 1), Y = X^2$. The correlation between X and Y is zero although they are obviously not independent.

An important remark is that the linear correlation only is defined if the variances of X and Y are finite. The variance for heavy tail distributions may be infinite and the estimation of correlation may therefore be misleading.

The most important weakness of using the linear correlation is stated by the following theorem

Theorem 2.8 *The joint distribution of a random vector is not determined by the marginal distributions and the pairwise correlations.*

(Embrechts et al [4]). To understand this statement consider figure 2.1. The two bivariate random vectors have the same correlation and the same marginal distributions but two totally different dependence structures. This illustrates how using linear correlation might be misleading. Moreover the correlation between two random variables does not only depend on their copula but is also inextricably linked to the marginal distributions.

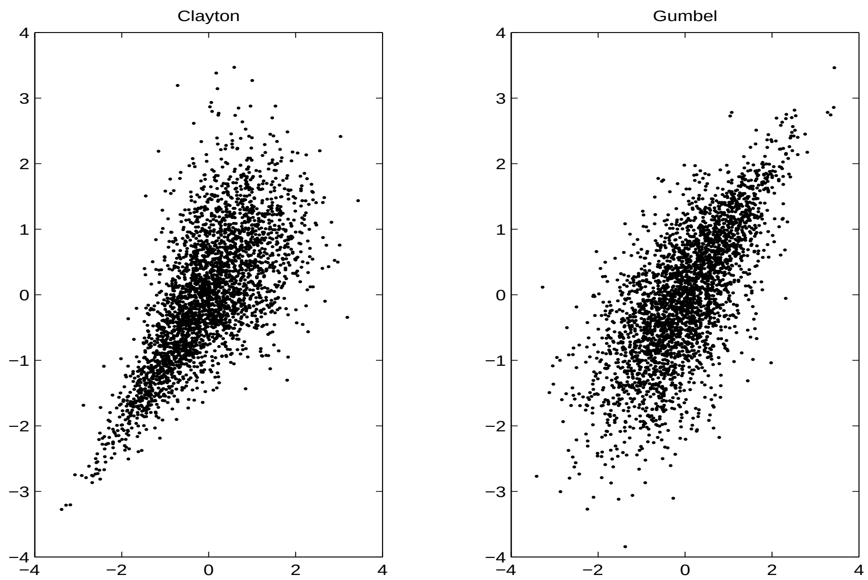

Figure 2.1: 3000 simulated points from a Clayton and a Gumbel copula with standard normal marginal distributions.

2.2.2 Rank Correlation

Rank correlation is a measure of *concordance* between two random variables. Consider two points (x, y) and $(\widetilde{x}, \widetilde{y})$. The two points are said to be *concordant* if $(x-\widetilde{x})(y-\widetilde{y}) > 0$ and are said to be *disconcordant* if $(x-\widetilde{x})(y-\widetilde{y}) < 0$.

Theorem 2.9 *Let $(X, Y)^T$ and $(\widetilde{X}, \widetilde{Y})^T$ be independent vectors of continuous random variables with joint distribution functions H and $\widetilde{H}$, respectively, with common margins F (of $X, \widetilde{X}$)) and G (of $Y, \widetilde{Y}$). Let C and $\widetilde{C}$ denote the copulas of $(X, Y)^T$ and $(\widetilde{X}, \widetilde{Y})^T$, respectively, so that $H(x, y) = C(F(x), G(y))$ and $\widetilde{H}(x, y) = \widetilde{C}(F(x), G(y))$. Let Q denote the difference between the probability of concordance and disconcordance of $(X, Y)^T$ and $(\widetilde{X}, \widetilde{Y})^T$, i.e. let*

$$Q = P\left((X - \widetilde{X})(Y - \widetilde{Y} > 0)\right) - P\left((X - \widetilde{X})(Y - \widetilde{Y} > 0)\right).$$

Then

$$Q = Q(C, \widetilde{C}) = 4 \int\int_{[0,1]^2} \widetilde{C}(u, v) dC(u, v) - 1,$$

(Nelsen [13]). This motivates *Kendall's rank correlation* (*Kendall's tau*), which is simply the probability of concordance minus the probability of disconcordance for these pairs.

Definition 2.10 *Kendall's tau for a random vector* $(X,Y)^T$ *is defined as*

$$\tau(X,Y) = P\left((X-\widetilde{X})(Y-\widetilde{Y}) > 0\right) - P\left((X-\widetilde{X})(Y-\widetilde{Y}) < 0\right)$$

where $(\widetilde{X},\widetilde{Y})^T$ *is an independent copy of* $(X,Y)^T$.

Theorem 2.11 *Let* $(X,Y)^T$ *be a vector of continuous random variables with copula C. Then Kendall's tau for* $(X,Y)^T$ *is given by*

$$\tau(X,Y) = Q(C,C) = 4\int\int_{[0,1]^2} C(u,v)dC(u,v) - 1.$$

(Embrechts et al [5]). Note that the integral above is the expected value of of the random variable *C(U,V)*, where $U,V \sim U(0,1)$, with joint distribution function C, i.e. $\tau(X,Y) = 4\mathbf{E}(C(U,V)) - 1$.

Theorem 2.12 *Let X and Y be continuous random variables with copula C, and let* τ *denote Kendall's tau. Then the following are true:*

1. $\tau(X,Y) = 1 \Longleftrightarrow C = M$
2. $\tau(X,Y) = -1 \Longleftrightarrow C = W$.

(Embrechts et al [5]). Kandall's tau is taking values in the interval $[-1,1]$. It gives the value zero for independent random variables, although a Kendall's tau of zero do not necessarily imply independence. But most importantly Kandall's tau does, unlike the linear correlation, only depend on the copula.

Let $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$ denote a random sample of n observations from a vector (X,Y) of continuous random variables. Then Kendall's tau for the sample is defined as

$$\hat{\tau} = \frac{c-d}{c+d} = (c-d)/\binom{n}{2} \tag{2.3}$$

where c is the number of concordant pairs and d is the number of disconcordant pairs. (Nelsen [13]).

2.3 Archimedean Copulas

Archimedean copulas are a class of copulas, which provide a great variety of dependence structures and have closed form expressions. Expressions of Archimedean copulas use the pseudo inverse.

Definition 2.13 *Let* φ *be a continuous, strictly decreasing function from* $[0,1]$ *to* $[0,\infty]$ *such that* $\varphi(1) = 0$. *The pseudo-inverse of* φ *is the function* $\varphi^{[-1]} : [0,\infty] \to [0,1]$ *given by*

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \le t \le \varphi(0) \\ 0, & \varphi(0) \le t \le \infty. \end{cases}$$

Note that $\varphi^{[-1]}$ is continuous and decreasing on $[0, \infty]$, and strictly decreasing on $[0, \varphi(0)]$. Furthermore, $\varphi^{[-1]}(\varphi(u)) = u$ on $[0, 1]$, and

$$\varphi(\varphi^{[-1]}(t)) = \begin{cases} t, & 0 \leq t \leq \varphi(0) \\ \varphi(0), & \varphi(0) \leq t \leq \infty. \end{cases}$$

Finally, if $\varphi(0) = \infty$, then $\varphi^{[-1]} = \varphi^{-1}$.
(Nelsen [13]).

Theorem 2.14 *Let φ be a continuous, strictly decreasing function from $[0, 1]$ to $[0, \infty]$ such that $\varphi(1) = 0$, and let $\varphi^{[-1]}$ be the pseudo-inverse of φ. Let C be the function from $[0, 1]^2$ to $[0, 1]$ given by*

$$C(u_1, u_2) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2)). \tag{2.4}$$

Then C is a copula if and only if φ is convex.

Copulas of the form (2.4) are called Archimedean copulas and the function φ is called the generator of the copula.

For Archimedean copulas Kendall's tau can be expressed as a one-dimensional integral of the generator

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt. \tag{2.5}$$

(Nelsen [13]).

2.3.1 Clayton Copula

The Clayton copula family has the generator $\varphi(t) = (t^{-\theta} - 1)/\theta$. By using (2.4) we get

$$C_\theta^{Cl}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}. \tag{2.6}$$

Using (2.5), Kendall's tau for a bivariate Clayton copula is

$$\tau = 1 + 4 \int_0^1 \frac{t^{\theta+1} - t}{\theta} dt = 1 + \frac{4}{\theta} \left(\frac{1}{\theta + 2} - \frac{1}{2} \right) = \frac{\theta}{\theta + 2}. \tag{2.7}$$

2.3.2 Gumbel Copula

The Gumbel copula family has the generator $\varphi(t) = (-\log t)^\theta$. By using (2.4) we get

$$C_\theta^{Gu}(u_1, u_2) = \exp\{-[(-\log u_1)^\theta + (-\log u_2)^\theta]^{1/\theta}\}. \tag{2.8}$$

Using (2.5), Kendall's tau for a bivariate Clayton copula is

$$\tau = 1 + 4 \int_0^1 \frac{t \log t}{\theta} dt = 1 + \frac{4}{\theta} \left(\left[\frac{t^2}{2} \log t \right]_0^1 - \int_0^1 \frac{t}{2} dt \right) = 1 - \frac{1}{\theta}. \tag{2.9}$$

2.3.3 Frank Copula

The Frank copula family has the generator $\varphi(t) = \log(\frac{e^{\theta t}-1}{e^{\theta}-1})$. By using (2.4) we get

$$C_\theta^{Fr}(u_1, u_2) = \frac{1}{\theta}\left(1 + \frac{\left(e^{\theta u_1} - 1\right)\left(e^{\theta u_2} - 1\right)}{(e^\theta - 1)}\right). \tag{2.10}$$

Using (2.5), it can be shown that Kendall's tau for a bivariate Frank copula is

$$\tau = 1 - \frac{4}{\theta}(1 - D_1(\theta)) \tag{2.11}$$

where $D_k(x)$ is the Debye function, given by

$$D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k}{e^t - 1}.$$

2.3.4 Linear Mixtures of Copulas

Let $C(u_1, u_2)$ and $\widetilde{C}(u_1, u_2)$ be two different copulas. Then the linear mixture C^m of the copulas is given by

$$C^m(u_1, u_2) = (1-p)C(u_1, u_2) + p\widetilde{C}(u_1, u_2), \qquad 0 \leq p \leq 1$$

is also a copula. (Nelsen [13]). In this thesis linear mixtures between independent and various copulas will be used.

Chapter 3

Fitting an Appropriate Model - Parameter Estimation

3.1 Fitting the Marginal Distributions

The parameters of the marginal distributions are estimated by the maximum likelihood method. Let $x_1, \ldots, x_n$ denote the data sample, and $\boldsymbol{\theta}$ denote the parameter vector of the distribution family. Then the MLE estimate of $\boldsymbol{\theta}$ is obtained by maximizing the log-likelihood function

$$\log L(\boldsymbol{\theta}; x_1, \ldots, x_n) = \sum_{i=1}^{n} \log f(\boldsymbol{\theta}; x_i) \tag{3.1}$$

where f is the marginal density function.

3.2 Fitting the Copula

3.2.1 The Standard Log-likelihood Method

Fitting a copula by the standard log-likelihood method is performed in two stages. In the first stage some appropriate parametric marginal distributions are selected and the parameters of the distributions are estimated. The marginal distributions are then used to construct a *pseudo-sample* of observations from the copula. In the second stage the copula parameters are estimated by maximum likelihood from the pseudo-sample. Let $\widehat{F}_1, \ldots, \widehat{F}_d$ denote the distribution functions of the selected margins. The pseudo-sample from the copula are the vectors $\widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n$ where

$$\widehat{\mathbf{U}}_i = (\widehat{U}_{1i}, \ldots, \widehat{U}_{di})^T = \left(\widehat{F}_1(X_{1i}), \ldots, \widehat{F}_d(X_{di})\right)^T \tag{3.2}$$

for $i = 1, \ldots, n$.
Let $C_{\boldsymbol{\theta}}$ denote a parametric copula where $\boldsymbol{\theta}$ is the vector of parameters to

be estimated. The MLE is obtained by maximizing

$$\log L(\boldsymbol{\theta}; \widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n) = \sum_{i=1}^{n} \log c_{\boldsymbol{\theta}}(\widehat{\mathbf{U}}_i) \tag{3.3}$$

with respect to $\boldsymbol{\theta}$, where $c_{\boldsymbol{\theta}}$ denotes the density as in (2.2), and $\widehat{\mathbf{U}}_i$ denotes the pseudo-observation from the copula as in (3.2).
(Embrects et al [4]). Obviously the statistical quality of the estimates of the copula parameters depend very much on the quality of the estimates of the marginal distributions used in the formation of the pseudo-sample from the copula. This is definitely a drawback and motivates the pseudo log-likelihood method.

3.2.2 The Pseudo Log-likelihood Method

Dias [3] describes the *pseudo log-likelihood method*, that avoids the crucial choice of parametric marginal distributions when estimating the pseudo-sample. The pseudo log-likelihood method has been proposed by several authors, among others: Clayton and Cuzick [2], Genest [6] and Oakes [14]. Suppose $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is a sample of n iid d-dimensional vectors, where $\mathbf{x_i} = (x_{1i}, \ldots x_{1i})^T$ for $i = 1, \ldots, n$. The difference from the standard log-likelihood method is that the distribution function of margin i is modeled by the rescaled empirical distribution function

$$F_{in}(x) = \frac{1}{n+1} \sum_{j=1}^{n} \mathbf{I}_{\{y \in \mathbf{R}: y \leq x\}}(x_{ij}), \tag{3.4}$$

assuming that the sample size will be large enough to enable statistically accurate non-parametric estimation. The n observed vectors are then transformed to pseudo-observations

$$\widetilde{\mathbf{U}}_i = (\widetilde{U}_{1i}, \ldots, \widetilde{U}_{di})^T = (F_{1n}(x_{1i}), \ldots, F_{di}(x_{di}))^T. \tag{3.5}$$

for $i = 1, \ldots, n$.
The pseudo log-likelihood estimates are obtained by maximizing

$$\log L(\boldsymbol{\theta}; \widetilde{\mathbf{U}}_1, \ldots, \widetilde{\mathbf{U}}_n) = \sum_{i=1}^{n} \log c_{\boldsymbol{\theta}}(\widetilde{\mathbf{U}}_i) \tag{3.6}$$

with respect to $\boldsymbol{\theta}$.

3.3 Goodness of Fit

This section explains how the marginal distributions and the copula are selected. To illustrate the fitting of the marginal distributions quantile-quantile-plots are performed. The best fitting model is based on two factors.

First the models are ranked by their Akaike information criterion. Secondly the hypothesis that one model is better than another is tested by a χ^2-test. The accuracy of the estimations of the parameters in the marginal distributions and the copula is determined by estimating the variances.

3.3.1 Quantile-Quantile-Plot

A quantile-quantile-plot (qq-plot) can be used to investigate if a sample $x_1, \ldots, x_n$ comes from a suggested reference distribution F. If we define the ordered sample as $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, then the qq-plot consists of the points

$$\left\{ \left(x_{(k)}, F^{\leftarrow} \left(\frac{k}{n+1} \right) \right) : k = 1, \ldots, n \right\}.$$

(Hult & Lindskog [10]). If the data has a similar distribution as the reference distribution the qq-plot is approximately linear. The plot will bend down if the reference distribution has too light tails and bend up if the reference distribution has too heavy tails.

3.3.2 Akaike Information Criterion

The log likelihood values $\log L$ shall not be used to compare the goodness-of-fit for different distributions since it produces a bias. Instead the Akaike information Criterion (AIC) may be used. AIC is given by

$$AIC = -2 \log L + 2q \tag{3.7}$$

where $\log L$ is the log-likelihood function and q is the number of parameters of the family of distribution fitted. The smaller the AIC-value is, the better the model fits the data.

3.3.3 Likelihood Ratio Test

Let $L_{D_1}(\widehat{\boldsymbol{\theta}}_{D_1})$ and $L_{D_2}(\widehat{\boldsymbol{\theta}}_{D_2})$ denote the likelihood functions of two distributions D_1 and D_2, evaluated at the maximum likelihood estimates $\widehat{\boldsymbol{\theta}}_{D_1}$ and $\widehat{\boldsymbol{\theta}}_{D_2}$. Further, let $Q = pD_1 + (1-p)D_2$ denote a mixture of the densities of distribution D_1 and distribution D_2 and let $\widehat{\boldsymbol{\theta}}_Q = (\widehat{\boldsymbol{\theta}}_{D_1}, \widehat{\boldsymbol{\theta}}_{D_2}, p)$ denote the parameters of Q. This mixture is called *the full model*. Next, let $L_Q(\widehat{\boldsymbol{\theta}}_Q)$ be the likelihood function for the full model evaluated in the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_Q$. Consider the null hypothesis H_0: distribution D_1, versus the alternate hypothesis H_a: distribution D_2. The test statistic is given by

$$T = -2 \log \left(\frac{L_{D_1}(\widehat{\boldsymbol{\theta}}_{D_1})}{L_{D_2}(\widehat{\boldsymbol{\theta}}_{D_2})} \right). \tag{3.8}$$

Suppose H_0 is rejected in favor for H_a whenever $T > \chi^2(r, \alpha)$, where the degrees of freedom r are the number of parameters in the full model less the number in the restricted model, and $\chi^2(r,\alpha)$ is the $[100(1-\alpha)]$:th percentile. Then

$$\Psi = -2\log\left(\frac{L_{D_1}(\widehat{\boldsymbol{\theta}}_{D_1})}{L_Q(\widehat{\boldsymbol{\theta}}_Q)}\right) \overset{a}{\sim} \chi^2(r).$$

That is, Ψ has an approximate chi-squared distribution, under the assumption that the data is actually distributed according to distribution D_1. Note that $T < \Psi$ and so

$$P\left(T > \chi^2(r,\alpha)\right) < P\left(\Psi > \chi^2(r,\alpha)\right) \approx \alpha,$$

assuming that distribution D_1 is the actual distribution.
(Carriere [1]). (3.8) may be expressed through the AIC-values

$$\begin{aligned} T &= -2\log\left(\frac{L_{D_1}(\widehat{\boldsymbol{\theta}}_{D_1})}{L_{D_2}(\widehat{\boldsymbol{\theta}}_{D_2})}\right) \\ &= -2\log L_{D_1}(\widehat{\boldsymbol{\theta}}_{D_1}) - (-2\log L_{D_2}(\widehat{\boldsymbol{\theta}}_{D_2})) \\ &= AIC_{D_1} - AIC_{D_2} - 2q_{D_1} + 2q_{D_2} \end{aligned} \tag{3.9}$$

where q_{D_1} and q_{D_2} are the number of parameters of distribution 1 and 2, respectively.

3.3.4 Accuracy of the Estimated Margin Parameters

To estimate the covariance matrix of the estimated parameters the so called *Fisher information matrix* is used. Let $\mathbf{X} = (X_1,\ldots,X_n)$ be a random sample, and let $f(\mathbf{X}|\boldsymbol{\theta})$ denote the probability density function for some model of the data, which has parameter vector $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_n)$. Then the Fisher information matrix $I_n(\boldsymbol{\theta})$ of the sample size n is given by the $k \times k$ symmetric matrix whose ij-th element is given by the first partial derivatives of the log-likelihood,

$$I_n(\boldsymbol{\theta})_{i,j} = Cov\left[\frac{\partial \log f(\mathbf{X}|\boldsymbol{\theta})}{\partial\theta_i}, \frac{\partial \log f(\mathbf{X}|\boldsymbol{\theta})}{\partial\theta_j}\right].$$

An alternative, but equivalent, definition for the Fisher information matrix is based on the expected values of the second partial derivatives, and is given by

$$I_n(\boldsymbol{\theta})_{i,j} = -\mathbf{E}\left[\frac{\partial^2 \log f(\mathbf{X}|\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right].$$

Strictly, this corresponds to the *expected* Fisher information. If no expectation is taken we obtain a data-dependent quantity called the *observed* Fisher information. It is worth noting that $I_n(\boldsymbol{\theta}) = nI_1(\boldsymbol{\theta})$, meaning that the expected Fisher information matrix for a sample of n independent observations is equivalent to n times the Fisher information matrix of a single observation. By the Fisher information matrix a lower bound for the variance of an arbitrary unbiased estimator may be determined.

Proposition 3.1 (The Cramer-Rao Inequality) *Let $T(\mathbf{X})$ be any statistic and let $\mu(\theta)$ be its expectation, such that $\mu(\theta) = \mathbf{E}[T(\mathbf{X})]$. Under some regularity conditions, it follows that for all θ,*

$$Var[T(\mathbf{X})] \geq \frac{\left(\frac{d\mu(\theta)}{d\theta}\right)^2}{I_n(\theta)}.$$

If $T(\mathbf{X})$ is an unbiased estimator for θ then the numerator becomes one and the lower bound is simply $1/I_n(\theta)$. For maximum likelihood estimator the following useful proposition can be used

Proposition 3.2 *The maximum likelihood estimator achieves the Cramer-Rao minimum variance asymptotically. That is, it follows under some regularity conditions that the sampling distribution of a maximum likelihood estimator $\boldsymbol{\theta}_{ML}$ is asymptotically unbiased and also asymptotically normal with its covariance matrix obtained from the inverse of the Fisher information matrix of size 1,*

$$\boldsymbol{\theta}_{ML} \to N\left(\boldsymbol{\theta}, \frac{I_1(\boldsymbol{\theta})^{-1}}{n}\right), \qquad n \to \infty$$

(Myung & Navarro [12]).

3.3.5 Accuracy of the Estimated Copula Parameters

In this section I will describe a procedure to estimate the covariance matrix of the estimated copula parameters when using the pseudo log-likelihood method.

Introduce the following notations

$$\begin{aligned}
l(\boldsymbol{\theta}; u_1, \ldots, u_d) &= \log c(\boldsymbol{\theta}; u_1, \ldots, u_d) \\
l_{\theta_i}(\boldsymbol{\theta}; u_1, \ldots, u_d) &= \frac{\partial}{\partial \theta_i} l(\boldsymbol{\theta}; u_1, \ldots, u_d) \quad i = 1, \ldots, q \\
l_{\theta_i\theta_j}(\boldsymbol{\theta}; u_1, \ldots, u_d) &= \frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\boldsymbol{\theta}; u_1, \ldots, u_d) \quad i, j = 1, \ldots, q \\
l_i(\boldsymbol{\theta}; u_1, \ldots, u_d) &= \frac{\partial}{\partial u_i} l(\boldsymbol{\theta}; u_1, \ldots, u_d) \quad i = 1, \ldots, d.
\end{aligned}$$

The functions $l_{\theta_i}(\boldsymbol{\theta}; u_1, \ldots, u_d)$ are the so called score functions.
For the case of a bivariate model where the copula has a scalar parameter, d=2 and q=1, we have the following proposition.

Proposition 3.3 *In the case θ is a scalar and d=2, the semi-parametric estimator $\widehat{\theta}$ obtained by maximizing (3.6) is consistent and $n^{1/2}(\widehat{\theta} - \theta)$ is asymptotically normal with variance $\nu^2 = \sigma^2/\beta^2$, where*

$$\beta = -\mathbf{E}\left(l_{\theta,\theta}(\theta; F_1(X_1), F_2(X_2))\right) = \mathbf{E}\left(l_\theta^2(\theta; F_1(X_1), F_2(X_2))\right)$$

and

$$\sigma^2 = var\left(l_\theta(\theta; F_{1n}(X_1), F_{2n}(X_2))\right) = W_1(X_1) + W_2(X_2)$$

for

$$W_i(X_i) = \int_{[0,1]^2} \mathbf{I}_{\{u\in[0,1]:F_i(X_i)\le u\}}(u_i) l_{\theta,i}(\theta; u_1, u_2) c(\theta; u_1, u_2) du_1 du_2 \quad (3.10)$$

with $i = 1,2$.

An alternative expression for (3.10) which will be useful for estimation purposes is given, for $i = 1, 2$, by

$$W_i(X_i) = -\int_{[0,1]^2} \mathbf{I}_{\{u\in[0,1]:F_i(X_i)\le u\}}(u_i) \\ l_\theta(\theta; u_1, u_2) l_i(\theta; u_1, u_2) c(\theta; u_1, u_2) du_1 du_2.$$

With proposition 3.3 we are equipped to estimate standard error for the estimate θ as long as we can estimate ν^2. β^2 can be viewed as the variance of the random variable

$$A(X_1, X_2) = l_\theta(\theta; F_1(X_1), F_2(X_2)),$$

given that the expected value of the score function $l_\theta(\theta; F_1(X_1), F_2(X_2))$ is zero, and σ^2 is the variance of

$$B(X_1, X_2) = A(X_1, X_2) + W_1(X_1) + W_2(X_2).$$

there exist estimators of those quantities, namely $\widehat{\sigma}^2$ and $\widehat{\beta}^2$, such that

$$\widehat{\nu}^2 = \frac{\widehat{\sigma}^2}{\widehat{\beta}^2} \quad (3.11)$$

is a consistent estimator of ν^2. The variables A and B are not observed but the pseudo-observations $\widehat{A}_i$ and $\widehat{B}_i$, $i = 1, 2, \dots, n$ can be computed. Estimates of σ^2 and β^2 are then obtained by the empirical sample variance of the pseudo-observations. Suppose that $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a sample of n bivariate observations. The pseudo-observations $\widehat{A}$ are given by

$$\widehat{A}_i(x_{1i}, x_{2i}) = l_\theta(\widehat{\theta}; F_{1n}(x_{1i}), F_{2n}(x_{2i})), \quad i = 1, 2, \dots, n.$$

For the $\widehat{B}$ pseudo-observations we need the rescaled empirical copula function of the sample,

$$C_n(u_1, u_2) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbf{I}_{\{(y_1,y_2)\in[0,1]^2: y_1\le u_1, y_2\le u_2\}}(u_{1i}, u_{2i}).$$

The pseudo-observations of $\widehat{B}$ are then obtained from

$$\widehat{B}_i(x_{1i}, x_{2i}) = \widehat{A}_i(x_{1i}, x_{2i}) + \widehat{W}_1(x_{1i}) + \widehat{W}_2(x_{2i}), \quad i = 1, 2, \dots, n,$$

with

$$\widehat{W}_j(x_{ji}) = -\int_{[0,1]^2} \mathbf{I}_{\{(y_1,y_2)\in[0,1]^2; F_{jn}(x_{ji})\leq y_j\}}(u_1,u_2)$$
$$l_\theta(\widehat{\theta};u_1,u_2) l_j(\widehat{\theta};u_1,u_2) dC_n(u_1,u_2)$$

for $j = 1,2$ and $i = 1,2,\ldots,n$. Now suppose we rearrange the sample $\{(x_{1i},x_{2i}) : i = 1,2,\ldots,n\}$ sorting the first components x_{1i} in increasing order. Denote the sample ordered in such a way by $\{(x_{1(i)},x_{2(i)}) : i = 1,2,\ldots,n\}$. As the pairs of observations must kept coupled, applying the marginal distributions, we obtain

$$\{(F_{1n}(x_{1(i)}),F_{2n}(x_{2(i)})) : i = 1,2,\ldots,n\} =$$
$$= \{(i/(n+1), S_i/(n+1)) : i = 1,2,\ldots,n\},$$

where S_i is the rank of x_{2i} within $\{x_{2i} : i = 1,2,\ldots,n\}$. So we can estimate the pseudo-observations $\widehat{A}$ as

$$\widehat{A}_i = l_\theta\left(\widehat{\theta}; \frac{i}{n+1}, \frac{S_i}{n+1}\right), \quad i = 1,2,\ldots,n \tag{3.12}$$

and $\widehat{B}$ by

$$\begin{aligned}\widehat{B}_i = \widehat{A}_i \quad &- \quad \frac{1}{n}\sum_{j=1}^{n} l_1\left(\widehat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right) l_\theta\left(\widehat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right) \\ &- \quad \frac{1}{n}\sum_{S_j\geq S_i} l_2\left(\widehat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right) l_\theta\left(\widehat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right)\end{aligned} \tag{3.13}$$

for $i = 1,2,\ldots,n$.

It is possible to generalize Proposition 3.3 yielding that the pseudo log-likelihood estimator for the vector parameter $\boldsymbol{\theta} \in \mathbf{R}^q$ is consistent and asymptotically normal. Moreover, the asymptotic covariance matrix of $n^{1/2}\boldsymbol{\theta}$ is

$$\mathbf{G}^{-1}\boldsymbol{\Sigma}\mathbf{G}^{-1} \tag{3.14}$$

where $\boldsymbol{G}$ is the information matrix associated with the copula and $\boldsymbol{\Sigma}$ is the covariance matrix of the q-dimensional random vector with kth component

$$\frac{\partial}{\partial\theta_k}\log c(F_1(X_1),F_2(X_2),\ldots,F_d(X_d);\boldsymbol{\theta}) + \sum_{i=1}^{d} W_{\theta_k i}(X_i), \tag{3.15}$$

with

$$W_{\theta_k i}(X_i) =$$
$$\int_{[0,1]^d} \mathbf{I}_{\{u\in[0,1]:F_i(X_i)\leq u\}}(u_i) l_{\theta_k i}(\boldsymbol{\theta},u_1,\ldots,u_d) dC(\boldsymbol{\theta};u_1,\ldots,u_d).$$

In order to clarify the generalization of the variance estimator, suppose now $\boldsymbol{\theta}$ has dimension q and $d = 2$. Given the properties for $q = 1$ we have as an estimator for the matrix $\boldsymbol{G}$ the following. The diagonal elements of $\boldsymbol{G}$, G_{kk} can be estimated as the variance of the pseudo-observations $\widehat{A}_{\theta_k i}$,

$$\widehat{A}_{\theta_k i} = l_{\theta_k}\left(\widehat{\theta}_k; \frac{i}{n+1}, \frac{S_i}{n+1}\right),$$

for $k = 1, 2, \ldots, q$ and $i = 1, 2, \ldots, n$ and assuming that we ordered the sample just like for (3.12). The off-diagonal elements of $\boldsymbol{G}$, G_{kj} for $k \neq j$ can be estimated by

$$\widehat{G}_{kj} = -\frac{1}{n}\sum_{i=1}^{n} l_{\theta_k\theta_j}\left(\widehat{\theta}_k, \widehat{\theta}_j; \frac{i}{n+1}, \frac{S_i}{n+1}\right).$$

The matrix $\boldsymbol{\Sigma}$ can be estimated as the sample covariance matrix of the q vectors of the pseudo-observations $\widehat{B}_{\theta_k i}$ for $k = 1, 2, \ldots, q$ and $i = 1, 2, \ldots, n$ namely

$$\begin{aligned}\widehat{B}_{\theta_k i} = \widehat{A}_{\theta_k i} \quad & - \quad \frac{1}{n}\sum_{j=1}^{n} l_1\left(\widehat{\theta}_k; \frac{j}{n+1}, \frac{S_j}{n+1}\right) l_{\theta_k}\left(\widehat{\theta}_k; \frac{j}{n+1}, \frac{S_j}{n+1}\right) \\ & - \quad \frac{1}{n}\sum_{S_j \geq S_i} l_2\left(\widehat{\theta}_k; \frac{j}{n+1}, \frac{S_j}{n+1}\right) l_{\theta_k}\left(\widehat{\theta}; \frac{j}{n+1}, \frac{S_j}{n+1}\right).\end{aligned}$$

(Dias [3]).

Chapter 4

The Insurance Models

4.1 Survival and Mortality

Traditionally, T_x denotes the additional time to death of a x-year old person.

Definition 4.1 *The survival function $\ell(x,t)$ of a x-year old person is given by*

$$\ell(x,t) = 1 - F_{T_x}(t) = 1 - P(T_x > t), \qquad t \geq 0.$$

Theorem 4.2 *The survival function of a x-year old person can be expressed by the survival function of a newborn $\ell(t)$*

$$\ell(x,t) = \frac{\ell(0, x+t)}{\ell(0,x)} = \frac{\ell(x+t)}{\ell(x)}, \qquad t \geq 0.$$

In this study we are only concerned with the survival of married persons that have reached the age of 61, i.e. $P(T_0 > 61) = 1$. The following notations will be used: T^m is length of life of 61-year old husband and T^f is the length of life of a 61-year old wife. T_x and T_y is the additional time to death of a x-year old husband and a y-year old wife respectively, where x and y are larger than 61.
Then

$$\begin{aligned} F_{T_x}(t) &= P(T_x \leq t) \\ &= 1 - P(T_x > t) \\ &= 1 - P(T^m > x + t | T^m > x) \\ &= 1 - \frac{P(T^m > x+t, T^m > x)}{P(T^m > x)} \\ &= 1 - \frac{P(T^m > x+t)}{P(T^m > x)} \\ &= \frac{F_{T^m}(x+t) - F_{T^m}(x)}{1 - F_{T^m}(x)}, \qquad t \geq 0. \end{aligned} \tag{4.1}$$

and

$$F_{T_y}(t) \quad = \quad \frac{F_{T^f}(y+t) - F_{T^f}(y)}{1 - F_{T^f}(y)}, \qquad t \geq 0. \tag{4.2}$$

Definition 4.3 *Let $F_T(x)$ denote the continuous distribution function for the time of death of a newborn person and let $f_T(x) = \frac{d}{dx}F_T(x)$ be the continuous density function. The mortality intensity $\mu(x)$ is the probability of death per time unit at the age of x, given by*

$$\mu(x) = \frac{f_T(x)}{1 - F_T(x)}.$$

Definition 4.4 *Makeham's formula of mortality intensity is defined as*

$$\mu(x) = a + be^{cx}.$$

Solving the differential equation of definition 4.3 for Makeham's formula gives

$$F_T(x) = 1 + k \exp\left\{ - \left(ax + \frac{b}{c}e^{cx} \right) \right\} \tag{4.3}$$

where k is a constant. The boundary condition $F_T(61) = 0$ gives

$$k = -\exp\left\{ 61a + \frac{b}{c}e^{61c} \right\}.$$

The probability of a person that have reached the age of 61 to die before the age of x is

$$F_T(x) = 1 - \exp\left\{ - \left(a(x-61) + \frac{b}{c}(e^{cx} - e^{61c}) \right) \right\} \tag{4.4}$$

and the density function is

$$f_T(x) = \frac{d}{dx}F_T(x) = (a + be^{cx}) \exp\left\{ - \left(a(x-61) + \frac{b}{c}(e^{cx} - e^{61c}) \right) \right\}. \tag{4.5}$$

Combining (4.1) and (4.2) with (4.3) gives

$$\begin{aligned}
& F_{T_x}(t) = \\
= \quad & \frac{F_{T^m}(x+t) - F_{T^m}(x)}{1 - F_{T^m}(x)} \\
= \quad & \frac{1 + k\exp\left\{ - \left(a(x+t) + \frac{b}{c}e^{c(x+t)} \right) \right\} - 1 - k\exp\left\{ - \left(ax + \frac{b}{c}e^{cx} \right) \right\}}{1 - 1 - k\exp\left\{ - \left(ax + \frac{b}{c}e^{cx} \right) \right\}} \\
= \quad & -\exp\left\{ - \left(a(x+t) + \frac{b}{c}e^{c(x+t)} - ax - \frac{b}{c}e^{cx} \right) \right\} + 1 \\
= \quad & 1 - \exp\left\{ \frac{b}{c}e^{cx}(1 - e^{ct}) - at \right\} \qquad (4.6)
\end{aligned}$$

and

$$F_{T_y}(t) = 1 - \exp\left\{\frac{b}{c}e^{cy}(1 - e^{ct}) - at\right\}. \tag{4.7}$$

4.2 The Expected Future Payments

4.2.1 Lifelong Joint Annuity Insurance

The probability of payment at t for a lifelong joint annuity insurance, as described in Section 1, is

$$\begin{aligned} P(\max[T_x, T_y] > t) &= 1 - P(\max[T_x, T_y] \le t) \\ &= 1 - P(T_x \le t, T_y \le t) \\ &= 1 - F_{T_x,T_y}(t, t). \end{aligned}$$

If K is the combined yearly pension payments to the husband and wife, the discounted expected future payments are

$$\begin{aligned} A &= K\int_0^\infty \left[1 - F_{T_x,T_y}(t,t)\right] e^{-\delta t} dt \\ &= K\int_0^\infty e^{-\delta t} dt - K\int_0^\infty F_{T_x,T_y}(t,t) e^{-\delta t} dt \\ &= \frac{K}{\delta} - K\int_0^\infty F_{T_x,T_y}(t,t) e^{-\delta t} dt. \end{aligned} \tag{4.8}$$

Using Sklar's Theorem, (4.6) and (4.7), (4.8) can be written as

$$A = \frac{K}{\delta} - K\int_0^\infty C(F_{T_x}(t), F_{T_y}(t)) e^{-\delta t} dt \tag{4.9}$$

where

$$\begin{aligned} F_{T_x}(t) = 1 - \exp\left\{\frac{b_1}{c_1}e^{c_1 x}(1 - e^{c_1 t}) - a_1 t\right\} \\ F_{T_y}(t) = 1 - \exp\left\{\frac{b_2}{c_2}e^{c_2 y}(1 - e^{c_2 t}) - a_2 t\right\} \end{aligned}$$

and C is the copula.

4.2.2 Survival Annuity Insurance

The probability of payment at t for a survival annuity insurance, as described in Section 1, is when the husband is insured and the wife is co-insured

$$\begin{aligned} P(T_x \le t, T_y > t) &= P(T_x \le t) - P(T_x \le t, T_y \le t) \\ &= F_{T_x}(t) - F_{T_x T_y}(t, t) \end{aligned}$$

and when the wife is insured and the husband is co-insured

$$P(T_x > t, T_y \leq t) = F_{T_y}(t) - F_{T_x T_y}(t,t) \tag{4.10}$$

Let K^m and K^f denote the yearly pension payments to the husband and wife, respectively. Then the discounted expected future payments, when the husband is insured and the wife is co-insured A^m and when the wife is insured and the husband is co-insured A^f, are given by

$$A^m = K^m \int_0^\infty \left[F_{T_x}(t) - F_{T_x,T_y}(t,t)\right] e^{-\delta t} dt \tag{4.11}$$

$$A^f = K^f \int_0^\infty \left[F_{T_y}(t) - F_{T_x,T_y}(t,t)\right] e^{-\delta t} dt. \tag{4.12}$$

Using Sklar's Theorem, (4.6) and (4.7), (4.11) and (4.12) can be written as

$$A^m = K^m \int_0^\infty \left[F_{T_x} - C(F_{T_x}(t), F_{T_y}(t))\right] e^{-\delta t} dt \tag{4.13}$$

$$A^f = K^f \int_0^\infty \left[F_{T_y} - C(F_{T_x}(t), F_{T_y}(t))\right] e^{-\delta t} dt \tag{4.14}$$

where

$$F_{T_x}(t) = 1 - \exp\left\{\frac{b_1}{c_1} e^{c_1 x}(1 - e^{c_1 t}) - a_1 t\right\}$$

$$F_{T_y}(t) = 1 - \exp\left\{\frac{b_2}{c_2} e^{c_2 y}(1 - e^{c_2 t}) - a_2 t\right\}$$

and C is the copula.

Chapter 5

Analysis

5.1 The Data

The data consists of 199127 pairs of Swedish citizens that fulfilled the following criteria

1. Were born during the period 1909-1942.
2. Got married.
3. Reached the age of 61 or more.
4. Had the maximum difference in age of 10 years to husband/wife.
5. Died during the period 1968-2003.

The ages are rounded off downwards i.e. a person that reached the age of 62 and 9 months is considered to have reached the age of 62. Couples where one or both of the persons has emigrated or in some other way disappeared from the national registration is leaved out from the data. Same-sex couples are not included in the data.

5.2 Modeling the Marginal Survival

Four different marginal distributions were tested: Makeham, Weibull, Gamma and Lognormal distribution. Quantile-quantile-plots against the different distributions were performed. Appendix A.1 shows the results. Based on the qq-plots the Makeham distribution seems to be the most satisfactory marginal distributions.

The distribution parameters were estimated by the log-likelihood method as described in Section 3.1. The covariance matrix of the parameters was estimated by the method described in Section 3.3.4. Table 5.1 shows the estimated distribution parameters and their variances and covariances.

Men		
Makeham	$\hat{a} = 0.0156$ $\hat{b} = 1.89 \cdot 10^{-6}$ $\hat{c} = 0.139$	$Var(\hat{a}) = 1.08 \cdot 10^{-7}$ $Var(\hat{b}) = 1.83 \cdot 10^{-14}$ $Var(\hat{c}) = 7.27 \cdot 10^{-7}$ $Cov(\hat{a}, \hat{b}) = -3.93 \cdot 10^{-11}$ $Cov(\hat{a}, \hat{c}) = 2.43 \cdot 10^{-7}$ $Cov(\hat{b}, \hat{c}) = -1.15 \cdot 10^{-10}$
Weibull	$\hat{a} = 1.07 \cdot 10^{-21}$ $\hat{b} = 11.0$	$Var(\hat{a}) = 1.05 \cdot 10^{-24}$ $Var(\hat{b}) = 2.41 \cdot 10^{-7}$ $Cov(\hat{a}, \hat{b}) = 2.54 \cdot 10^{-18}$
Gamma	$\hat{a} = 96.9$ $\hat{b} = 0.782$	$Var(\hat{a}) = 0.163$ $Var(\hat{b}) = 1.09 \cdot 10^{-5}$ $Cov(\hat{a}, \hat{b}) = -1.3 \cdot 10^{-3}$
Lognormal	$\hat{\mu} = 4.32$ $\widehat{\sigma}^2 = 0.102$	$Var(\hat{\mu}) = 5.44 \cdot 10^{-8}$ $Var(\widehat{\sigma}^2) = 4.62 \cdot 10^{-8}$ $Cov(\hat{\mu}, \widehat{\sigma}^2) = 9.77 \cdot 10^{-9}$
Women		
Makeham	$\hat{a} = 0.0138$ $\hat{b} = 3.76 \cdot 10^{-7}$ $\hat{c} = 0.158$	$Var(\hat{a}) = 5.51 \cdot 10^{-8}$ $Var(\hat{b}) = 6.21 \cdot 10^{-16}$ $Var(\hat{c}) = 6.26 \cdot 10^{-7}$ $Cov(\hat{a}, \hat{b}) = -4.92 \cdot 10^{-12}$ $Cov(\hat{a}, \hat{c}) = 1.53 \cdot 10^{-7}$ $Cov(\hat{b}, \hat{c}) = 6.26 \cdot 10^{-7}$
Weibull	$\hat{a} = 4.82 \cdot 10^{-23}$ $\hat{b} = 11.7$	$Var(\hat{a}) = 1.05 \cdot 10^{-24}$ $Var(\hat{b}) = 2.59 \cdot 10^{-7}$ $Cov(\hat{a}, \hat{b}) = 1.44 \cdot 10^{-18}$
Gamma	$\hat{a} = 100$ $\hat{b} = 0.768$	$Var(\hat{a}) = 0.169$ $Var(\hat{b}) = 1.03 \cdot 10^{-5}$ $Cov(\hat{a}, \hat{b}) = -1.3 \cdot 10^{-3}$
Lognormal	$\hat{\mu} = 4.34$ $\widehat{\sigma}^2 = 0.101$	$Var(\hat{\mu}) = 5.78 \cdot 10^{-8}$ $Var(\widehat{\sigma}^2) = 4.36 \cdot 10^{-8}$ $Cov(\hat{\mu}, \widehat{\sigma}^2) = 1.74 \cdot 10^{-8}$

Table 5.1: Estimated parameters of the marginal distributions and their variances and covariances.

Next, the AIC values were determined using (3.7). Table 5.2 shows the AIC-values.

	Men	Women
Makeham	1348698	1345575
Weibull	1382237	1370832
Gamma	1376508	1376133
Lognormal	1377743	1378591

Table 5.2: AIC-values for the marginal distributions.

The most satisfactory distribution based on the AIC-values is the Makeham distribution. The hypotheses that Weibull, Gamma or Lognormal distribution is significantly better than the Makeham distribution was tested using likelihood ratio tests as described in Section 3.3.3. Since the differences between the AIC-values of the Makeham distribution and the other distributions were very large ($> 10^4$) the hypotheses that Weibull, Gamma or Lognormal distribution is better than the Makeham distribution were rejected at all reasonable levels of significance, e.g. at the level of 99%; $\chi^2(4, 0.01) = 13.28$. Figure (5.1), (5.2), (5.3) and (5.4) shows plots of the empirical density function and the densities of the fitted marginal distributions. It is obvious that the Makeham distributions are the most satisfactory marginal distributions.

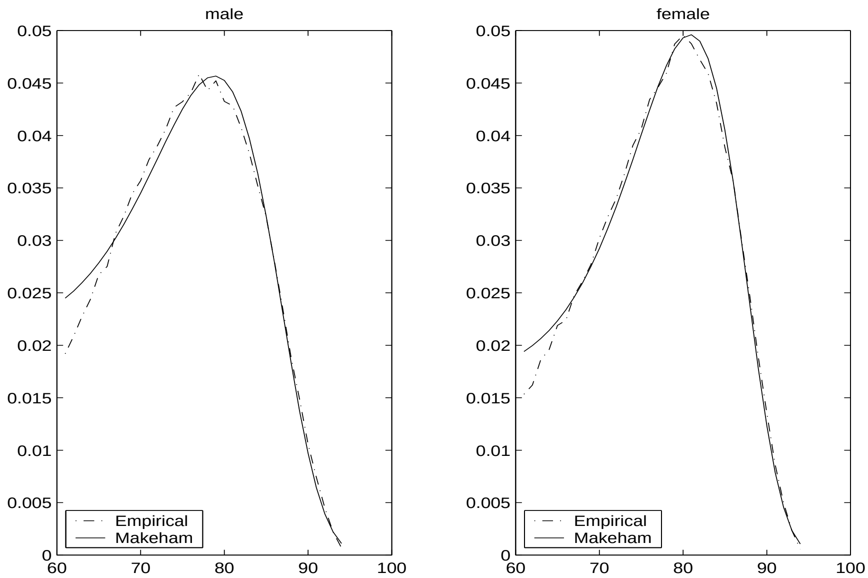

Figure 5.1: Empirical density functions and the density functions of the fitted Makeham distributions.

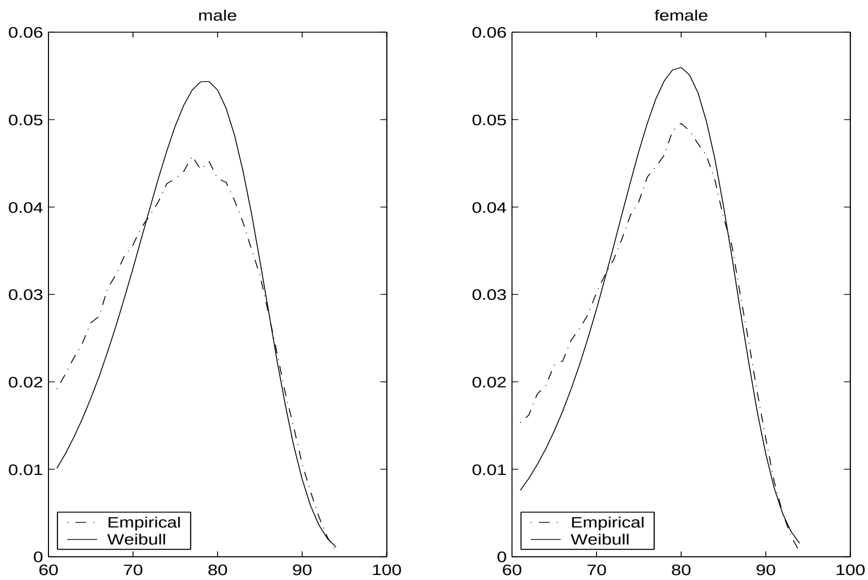

Figure 5.2: Empirical density functions and the density functions of the fitted Weibull distributions.

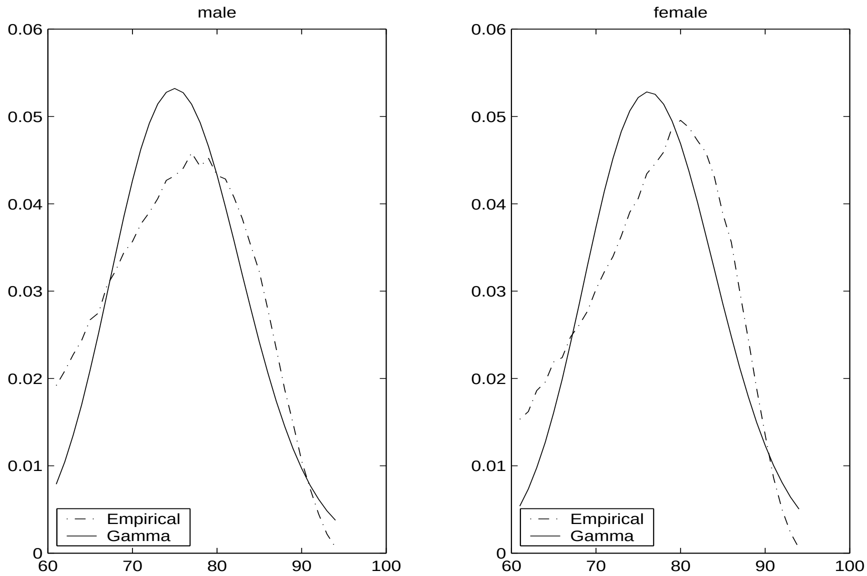

Figure 5.3: Empirical density functions and the density functions of the fitted Gamma distributions.

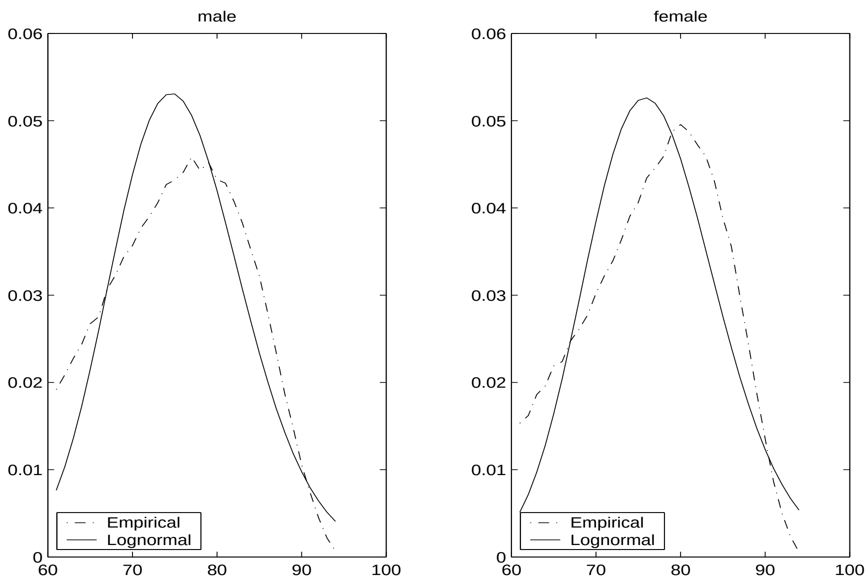

Figure 5.4: Empirical density functions and the density functions of the fitted Lognormal distributions.

5.3 Dependence Versus Difference in Age

The dataset was divided according to the difference in age between husband and wife (the difference in time between the dates of birth of the husband and wife). Let $z \in \{-10, -9, \ldots, 9, 10\}$ denote the difference in age, expressed in years (positive if husband older than wife and negative if wife older than husband).

The dependence in survival was measured by Kendall's tau rank correlation and linear correlation. Figure 5.5 shows Kendall's tau and the linear correlation plotted against z. The plot indicates a weak positive dependence that decreases as the difference in age is getting large.

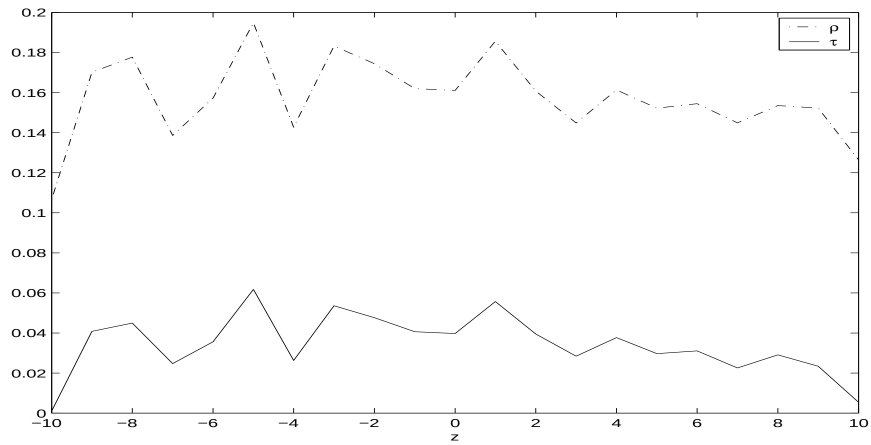

Figure 5.5: Linear correlation and Kendall's tau versus difference in age.

5.4 Modeling the Dependence Structure

To fit an appropriate copula the pseudo log-likelihood method, described in Section 3.2.2 was used. The pseudo-observations estimated by (3.5) forms the empirical copula. Figure 5.6 shows a two-dimensional histogram of the empirical copula of the total dataset.

Six different copulas were fitted: Frank, Gumbel, Clayton, Frank- Independent mixture, Gumbel-Independent mixture and Clayton-Independent mixture. The mixtures are linear, as described in Section 2.3.4. The variances and covariances of the copula parameters were estimated by the method described in Section 3.3.5. Based on the AIC-values Clayton is the most satisfactory fitting copula. The hypotheses that the other copulas fits better than Clayton were tested using likelihood ratio tests, as described in Section 3.3.3. Table 5.3 shows the estimated copula parameters, their variances and covariances, the AIC-values and the test statistics T. The hypotheses

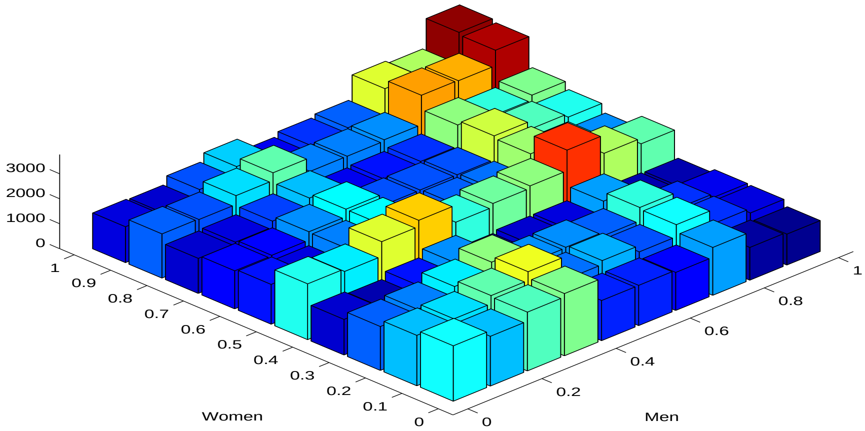

Figure 5.6: Histogram of the empirical copula.

were rejected at all reasonable levels of significance, e.g. at the level of 99% $\chi^2(2, 0.01) = 9.21$. Hence Clayton is proved to be the most satisfactory of the tested copulas.

Recall from section 3.3.5 that θ is asymptotically normal. A confidence interval for the Clayton copula parameter θ can be constructed by using the estimated variance.

$$I_{\hat{\theta}}^{99\%} = \hat{\theta} \pm \hat{\sigma}_{\hat{\theta}} \Phi^{-1}(0.995) = 0.1751 \pm 6.3 \cdot 10^{-3}$$

where Φ^{-1} denotes the inverse of standard normal distribution function. Note that zero does not fall within the interval. This is an important result since it rejects the hypothesis that the survival is independent.

To be able to generalize the conclusions, the total dataset was divided into sub datasets in two different ways. First, the total dataset was divided into 21 sub datasets, where each sub dataset contained notations of couples with the same difference in age z. The copula fitting procedure described above was repeated for all of the 21 datasets. Clayton was the best fitting copula, based on the AIC-values, for all of the datasets except for z=10 and z=-8, where Gumbel was the best one. However for z=10 and z=-8 the hypothesis that Clayton was a better fitting copula than Gumbel could not be rejected.

Secondly, the total dataset was divided into sub datasets containing couples where both persons have reached the age of $62, 63, \ldots$ or more. The copula fitting procedure was performed for each sub dataset. Clayton was the significantly best copula up to the minimum age of 70 and the copula parameter θ had a stable value of approximately 0.2 for the ages $62, 63, \ldots, 70$.

	F	FI	G	GI	C	CI
$\hat{\theta}$	0.881	2.57	1.08	1.22	0.175	0.605
$\hat{p}$		0.363		0.404		0.342
$\hat{\sigma}^2_{\hat{\theta}}$	$5.08 \cdot 10^{-6}$	$8.97 \cdot 10^{-4}$	$5.09 \cdot 10^{-6}$	$1.58 \cdot 10^{-5}$	$5.91 \cdot 10^{-6}$	$1.50 \cdot 10^{-3}$
$\hat{\sigma}^2_{\hat{p}}$		$1.59 \cdot 10^{-5}$		$3.73 \cdot 10^{-5}$		$3.93 \cdot 10^{-4}$
$\hat{\sigma}_{\hat{\theta}\hat{p}}$		$8.53 \cdot 10^{-5}$		$2.15 \cdot 10^{-5}$		$-7.21 \cdot 10^{-4}$
AIC	4275	4402	3653	3746	2872	3085
T	1403	1528	781.4	871.8		210.9

Table 5.3: Estimated parameters of the copulas: Frank, Frank-Independent mixture, Gumbel, Gumbel-Independent mixture, Clayton, and Clayton-Independent mixture, their variances and covariances, the AIC-values and the test statistics T. The parameters are estimated for the total dataset.

5.5 Pricing the Contracts

The aim of this section is to investigate how using the suggested Clayton copula-model, instead of assuming independence, effect the discounted expected future payments of the insurances. The discounted expected future payments of the insurances will further on be referred to as the prices. In the study below δ is set to 0.03 and K, K^m and K^f are set to one. The payments of the insurance, purchased at $t = 0$, are assumed to start immediately. The integrals in (5.1), (5.2), (5.3), (5.4), (5.5) and (5.6) have to be solved numerically since there exist no analytical solutions. This was performed by using the Matlab function *quad*, which uses adaptive Simpson quadrature where the absolute error tolerance was set to 10^{-11}.

5.5.1 Lifelong Joint Annuity Insurance

The pricing formula for the lifelong joint annuity insurance, when using the Clayton copula model, can be expressed by using (2.6) and (4.9)

$$A_C = \frac{K}{\delta} - K \int_0^{\infty} \left[F_{T_x}(t)^{-\theta} + F_{T_y}(t)^{-\theta} - 1 \right]^{-1/\theta} e^{-\delta t} dt. \tag{5.1}$$

The pricing formula for the independent model is

$$A_I = \frac{K}{\delta} - K \int_0^{\infty} F_{T_x}(t) F_{T_y}(t) e^{-\delta t} dt. \tag{5.2}$$

where F_{T_x} and F_{T_y} are the marginal distributions as defined in (4.9).

Let Δ^J be defined as

$$\Delta^J = \frac{A_I - A_C}{A_I} 100\%$$

i.e. Δ^J is the difference in price, expressed as a percentage, between the Clayton copula-model and the independent model.

Figure 5.7 shows the prices A_C and A_I plotted against the age of entrance of the couple. The Clayton copula parameter was set to $\theta = 0.2019$, i.e. the estimated parameter for $z = 0$. Figure 5.8 shows Δ^J plotted against the age of entrance of the couple. The ages of the coupled persons were assumed to be equal.

The price of the lifelong joint annuity insurance, when using the copula model instead of the independent model, is approximately 1.7% lower. This seemes like a good approximation for all z except $z =$ -10, where the survival can be assumed to be independent. This is motivated in Table 5.4, where Δ^J is listed for all z (estimated with the individual θ-values).

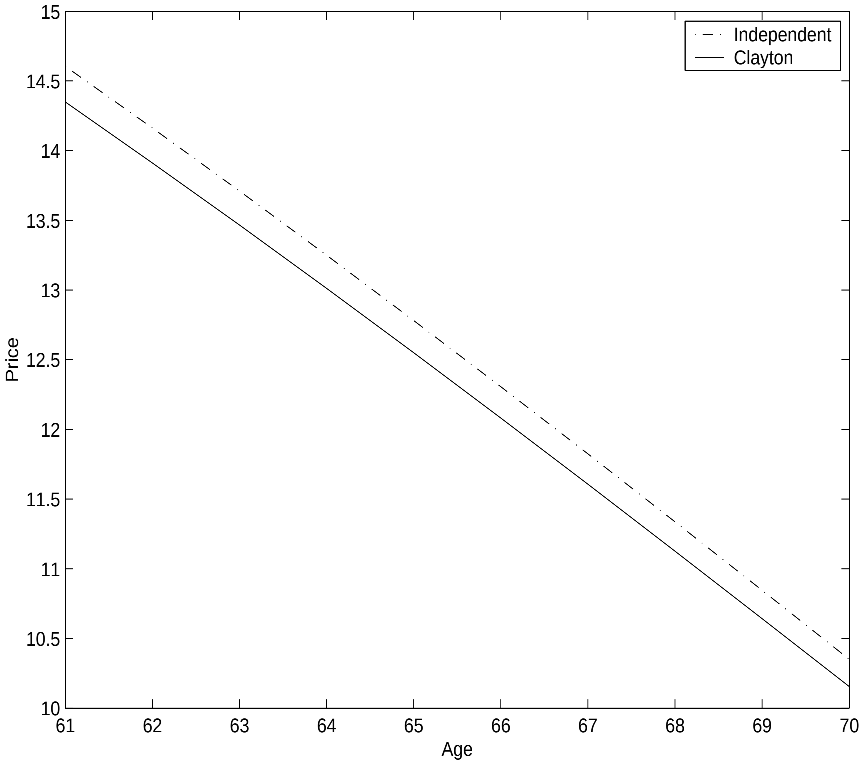

Figure 5.7: Price for the lifelong joint annuity insurance, when using the Clayton and the independent model, plotted against age of entrance. The ages of the coupled persons were assumed to be equal.

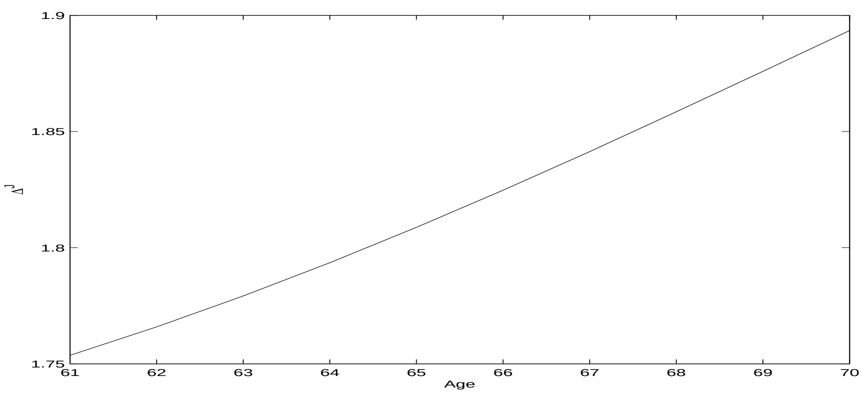

Figure 5.8: Δ^J plotted against age of entrance. The ages of the coupled persons were assumed to be equal.

5.5.2 Survival Annuity Insurance

The pricing formulas for the survival annuity insurances, when using the Clayton copula model, can be expressed by using (2.6), (4.13) and (4.14)

$$
\begin{aligned}
A_C^m &= K^m \int_0^\infty \left(F_{T_x}(t) - \left[F_{T_x}(t)^{-\theta} + F_{T_y}(t)^{-\theta} - 1 \right]^{-1/\theta} \right) e^{-\delta t} dt && (5.3)\\
A_C^f &= K^f \int_0^\infty \left(F_{T_y}(t) - \left[F_{T_x}(t)^{-\theta} + F_{T_y}(t)^{-\theta} - 1 \right]^{-1/\theta} \right) e^{-\delta t} dt && (5.4)
\end{aligned}
$$

and the pricing formulas, when using the independent model, are

$$
\begin{aligned}
A_I^m &= K^m \int_0^\infty \left(F_{T_x}(t) - F_{T_x}(t) F_{T_y}(t) \right) e^{-\delta t} dt && (5.5)\\
A_I^f &= K^f \int_0^\infty \left(F_{T_y}(t) - F_{T_x}(t) F_{T_y}(t) \right) e^{-\delta t} dt && (5.6)
\end{aligned}
$$

where F_{T_x} and F_{T_y} are the marginal distributions as defined in (4.9).

Let Δ^m and Δ^f be defined as

$$
\Delta^m = \frac{A_I^m - A_C^m}{A_I^m} 100\%, \qquad \Delta^f = \frac{A_I^f - A_C^f}{A_I^f} 100\%
$$

i.e. Δ^m and Δ^m are the differences in price, expressed as percentages, between the Clayton copula-models and the independent models.

Figure 5.9 shows A_C^m, A_I^m, A_C^f and A_I^f plotted against the age of entrance of the couple. Figure 5.10 shows Δ^m and Δ^f plotted against the age of entrance of the couple. The ages of the coupled persons are assumed to be equal. The prices of the survival annuity insurances, when using the copula models instead of the independent models, are approximately 7% lower when the husband is insured and 9% lower when the wife is insured (when $z = 0$).

The values of Δ^m and Δ^f for all z (estimated with the individual θ-values) are listed in Table 5.4. Notice that these values are estimated for the case when the youngest person is 61 years old.

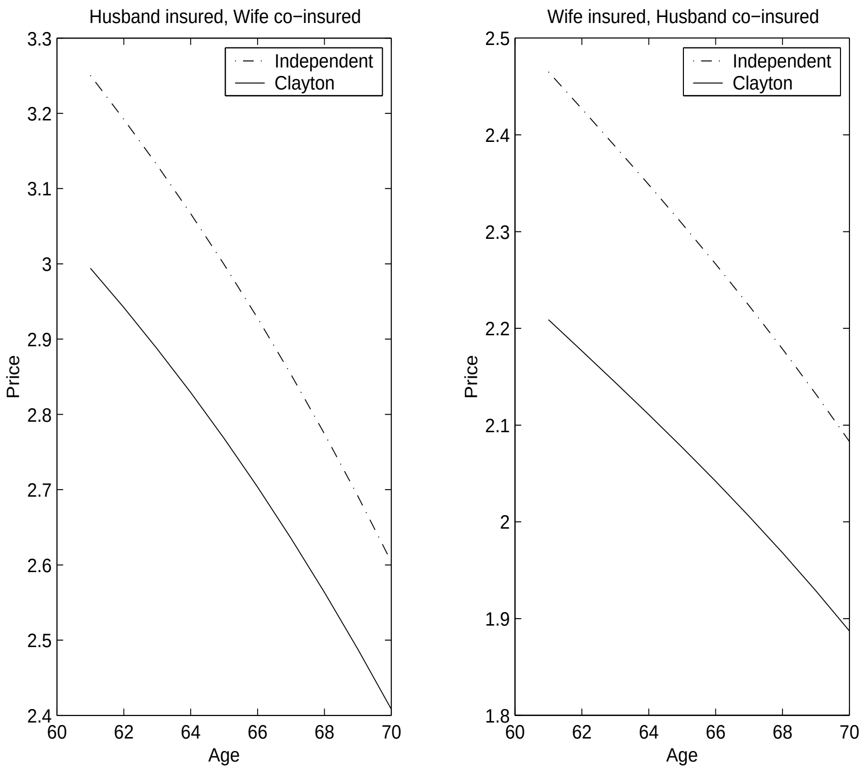

Figure 5.9: Price when using the Clayton and the independent model, when the husband is insured and the wife is co-insured (left) and when the wife is insured and the husband is co-insured (right), plotted against age of entrance. The ages of the coupled persons are assumed to be equal.

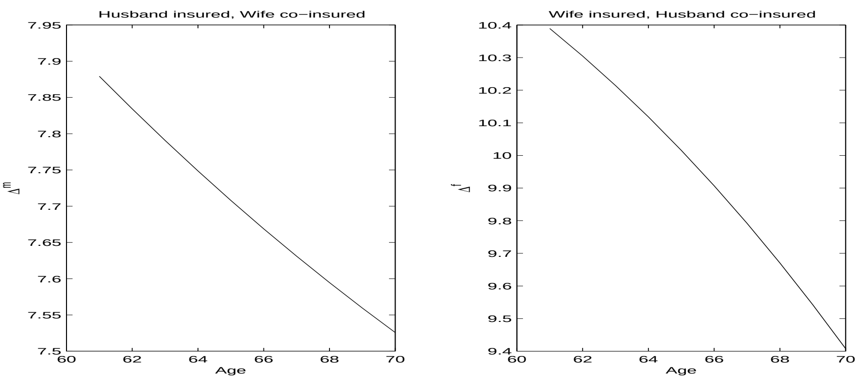

Figure 5.10: Δ plotted against age of entrance, when the husband is insured and the wife is co-insured (left) and when the wife is insured and the husband is co-insured (right). The ages of the persons are assumed to be equal.

z	n	θ	Δ^J	Δ^m	Δ^f
-10	361	0.0556	0.4431	4.4691	1.1678
-9	534	0.2281	1.7785	16.1641	5.0528
-8	866	0.2290	1.8471	15.2138	5.6703
-7	1252	0.1577	1.3409	10.0635	4.4566
-6	1965	0.1952	1.6734	11.5053	6.0321
-5	2917	0.2411	2.0663	13.0814	8.0885
-4	4706	0.1681	1.4898	8.7273	6.3381
-3	6977	0.2345	2.0425	11.1239	9.4470
-2	10387	0.2209	1.9326	9.8296	9.7160
-1	15155	0.1958	1.7188	8.1998	9.3863
0	21583	0.2019	1.7537	7.8792	10.3891
1	23519	0.2325	1.9877	8.3214	12.6050
2	23575	0.1869	1.6063	6.2636	10.9457
3	21304	0.1646	1.4046	5.1023	10.3270
4	18078	0.1975	1.6339	5.5321	13.0155
5	14429	0.1805	1.4663	4.6312	12.7092
6	11105	0.1807	1.4244	4.2016	13.4917
7	8110	0.1896	1.4374	3.9654	14.9436
8	5663	0.1923	1.3972	3.6113	16.0166
9	3988	0.1877	1.3028	3.1611	16.5446
10	2653	0.1679	1.1142	2.5433	15.7486
Tot	199127	0.1751	1.5342	6.8933	9.0891

Table 5.4: Properties of the sub datasets. n is the number of observations, θ is the estimated Clayton copula parameter and Δ^J, Δ^m and Δ^f are the differences in price, expressed as percentages, between the Clayton copula-model and the independent model. The Δ-values were estimated for a couple where the youngest person was 61 years old.

Chapter 6

Conclusion

Copulas have been applied on the pricing of lifelong joint annuity insurances and survival annuity insurances. The dataset used contained 199127 observations of married couples where both husband and wife reached the age of 61 or more and the maximum difference in age between husband and wife was ten years. The Makeham distribution is proved to be very good choices of marginal distributions and the hypotheses that the marginal survival of the husband and the wife better fit the Weibull, Gamma or Lognormal distributions than the Makeham distribution are rejected at all reasonable levels of significance.

The hypothesis of independent survival for married couples is also rejected at all reasonable levels of significance. The survival show a weak positive dependence, where Kendall's tau is approximately 0.05. The dependence is getting weaker as the difference in age between husband and wife is getting large, although it can be approximated to be equal for all of the differences in age, except when the female is ten years older than the male, where the survival is close to independent.

The dependence structure of the survival is shown to fit best a Clayton copula model since the hypotheses that the dependence structure fit the Gumbel, Gumbel-independent mixture, Frank, Frank-independent mixture or Clayton-independent mixture copulas better than the Clayton copula are rejected at all reasonable levels of significance. The dependence structures of the sub datasets, containing couples where both persons have reached the age of $62, 63, \ldots, 70$, are also shown to fit best the Clayton copula. The Clayton copula parameter θ can be approximated to 0.2 for all of the sub datasets.

For the lifelong joint annuity insurance the Clayton copula model yields a 1.7% lower discounted expected future payments than, when assuming independence.

When pricing a survival annuity insurance it is important to distinguish between the case when the husband is insured and wife is co-insured and the case when the wife is insured and the husband is co-insured. If the ages of the husband and wife are approximately equal the Clayton copula-model yields 8% lower discounted expected future payments in the first case and 10% lower in the second case. In the first case, the difference expressed as a percentage, between the discounted expected future payments of the Clayton copula-model and the independent model is decreasing as the difference in age increases from -10 to 10. In the second case the difference expressed as a percentage is increasing.

Suggested future research may be to apply the methods described in this thesis on a dataset of insured couples.

Appendix A

Appendix

A.1 Quantile-Quantile-Plots

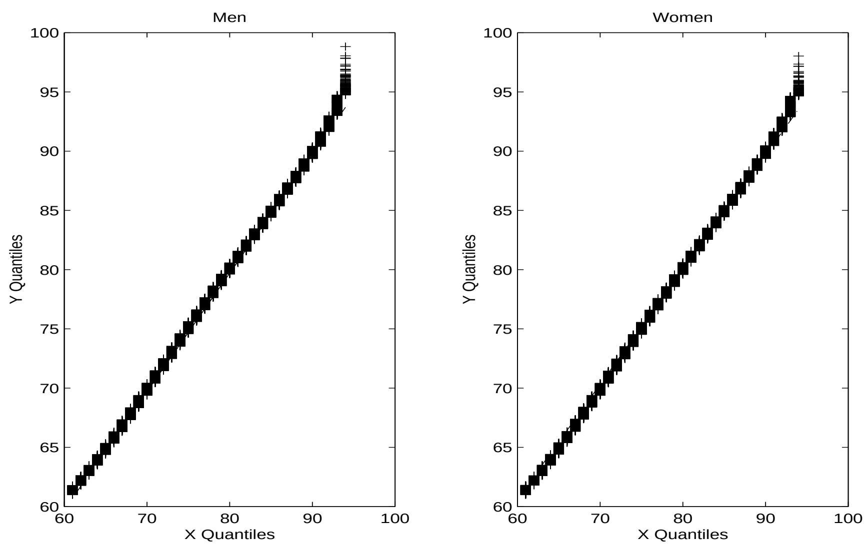

Figure A.1: QQ-plots against Makeham distribution.

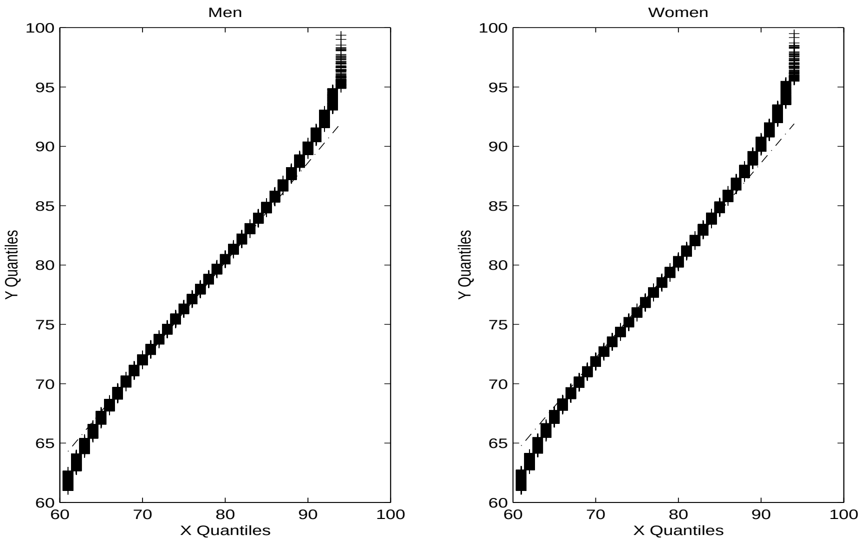

Figure A.2: QQ-plots against Weibull distribution.

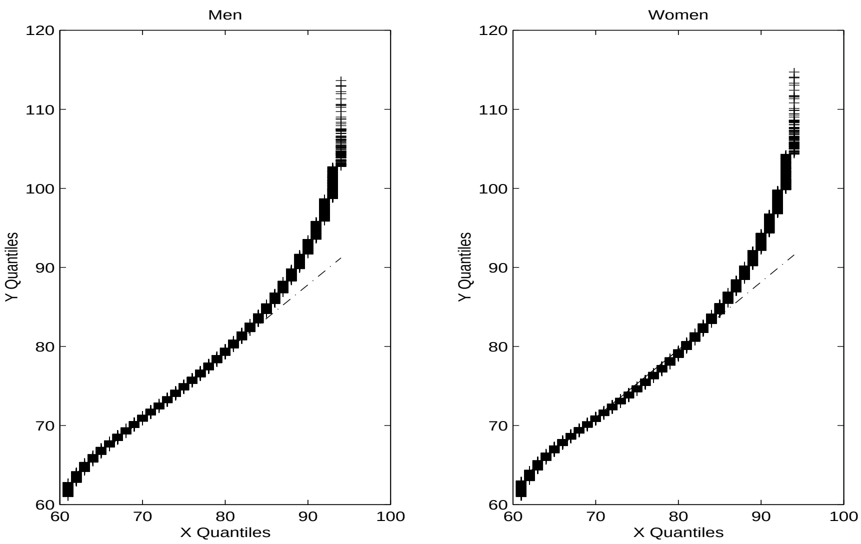

Figure A.3: QQ-plots against Gamma distribution.

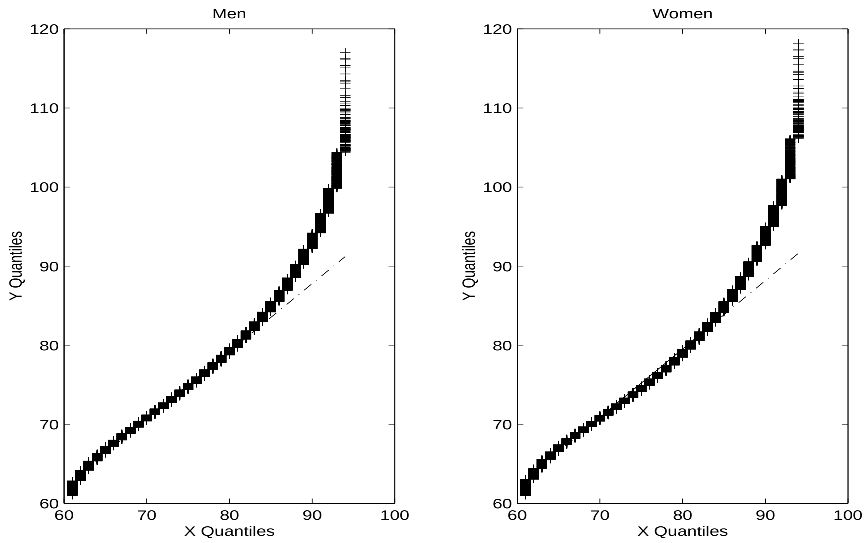

Figure A.4: QQ-plots against Lognormal distribution.

Bibliography

[1] Carriere, J.
Bivariate Survival Models for Coupled Lives, University of Alberta, (1998).

[2] Clayton, D., and Cuzick, J.
Multivariate generalizations of the proportional hazard model. J. Roy. Statist. Soc. Ser. A 148, 2 (1985), 82-117.

[3] Dias Da Costa, A.
Copula Inference for Financial and Insurance, Doctoral Thesis ETHZ (2004).

[4] Embrechts, P., Frey, R. and McNeil, A.
Quantitative Methods for Financial Risk Management, Extracts (2003).

[5] Embrechts, P., Lindskog, F. and McNeil, A.
Modeling Dependence with Copulas, ETHZ (2001).

[6] Genest, C.
Frank's family of bivariate distributions, Biometrica (1987) 549-555.

[7] Gut, A.
An Intermediate Course in Probability, Springer-Verlag (1995).

[8] Hougaard, P., Harvald, B., and Holm,. N.V.
Measuring the Similarities Between the Lifetimes of Adult Danish Twins Born Between 1881-1930, J. American Stat. Assoc., Vol 87, No. 417. March 1992, 17-24.

[9] Hougaard, P.
Modeling Multivariate Survival, Scand J Statist 14, 1987, 291-304.

[10] Hult,H. Lindskog, F.
Mathematical Methods in Risk Management, Lecture Notes, KTH (2004).

[11] Maguluri, G.
Semiparametric Estimation of Association in a Bivariate Survival Function, The Annals of Statistics, Vol 21, No 3, 1993, 1648-1662.

[12] Myung, J. I., Navarro, D. J.
Information Matrix, Encyclopedia of Behavioral Statistics, preprint, Wiley (2005).

[13] Nelsen, R.
An Introduction to Copulas, Springer-Verlag (1998).

[14] Oakes, D.
A model for association in bivariate survival data, J. Roy. Statist. Ser. B 44, 3 1982, 414-422.

[15] Oakes, D.
Bivariate Survival Models Induced by Frailties, J. American Stat. Assoc., Vol 84, No. 406. June 1989, 487-493.

[16] Shih, J. H., and Louis, T. A.
Inferences on the Association Parameter in Copula Models for Bivariate Survival Data, Biometrics 51, December 1995, 1384-1399.

[17] Zheng, M., and Klein,. J. P.
Estimates of Marginal Survival for Dependent Computing Risks Based on an Assumed Copula, Biometrica, 82, 1, 1995, 127-138.