Modelling dependent yearly claim totals including zero-claims in private health insurance

Vinzenz Erhardt, Claudia Czado

Technische Universität München, Zentrum Mathematik,
Boltzmannstr. 3, 85747 Garching, Germany
erhardt@ma.tum.de, cczado@ma.tum.de

Abstract

In insurance applications yearly claim totals of different coverage fields are often dependent. In many cases there are numerous claim totals which are zero. A marginal claim distribution will have an additional point mass at zero, hence this probability function will not be continuous at zero and the cumulative distribution functions will not be uniform. Therefore using a copula approach to model dependency is not straightforward. We will illustrate how to express the joint probability function by copulas with discrete and continuous margins. A pair copula construction will be used for the fit of the continuous copula allowing to choose appropriate copulas for each pair of margins.

Keywords: dependence modeling, health insurance, pair-copula-constructions, zero-claim.

1 Introduction

Dependencies in insurance data may occur in many fields. Claim frequencies and sizes are likely to be dependent. A copula approach to this issue applied to car insurance data has been developed by Kastenmeier (2008). Specifically in the field of health insurance, dependencies between inpatient and outpatient treatments are considered by Frees et al. (2007). Pitt et al. (2006) discuss multi-dimensional measures of health care utilization. They model the dependency between six measures of medical care demand, which are categorized numbers of visits to physicians. Zimmer & Trivedi (2006) use a copula for three simultaneously determined outcomes, i.e. the health insurance status for married couples and their individual health care demand. Dependencies between the number of visits of insured and uninsured persons per year have been considered by Deb et al. (2006). Spatial clustering is investigated by Brezger & Lang (2006) where treatment costs are assumed to be influenced by time, age, sex and spatial effects. A longitudinal model for normalized patient days per year in Wisconsin nursing homes from 1995 through 2001 using copulas was developed by Sun et al. (2008).

The aim of this paper is to develop a collective model of yearly claim totals capable of reflecting the dependency between different coverage fields. This will allow to appropriately predict which yearly total amount an insurer needs to reserve in order to cover expenditures for an insured person depending on age, sex and other attributes. This is crucial for the pricing of premiums and for risk management. Neglecting dependency between dependent fields may result for example in a misspecification of significant policy characteristics or in false reserving calculus, since diversification effects are neglected. Applications for such a model abound: in operational risk, losses of different dependent types occur very seldom, hence many loss totals are zero. Whenever policies cover different risks claim totals may be zero for some risk and positive for other risks: a car insurance contract may cover vehicle damages of different subclasses or third-party liabilities.

Claim frequency and claim size models are standard tools in non-life actuarial science. For claim frequencies often many zeros are observed caused for example by deductibles. Specifically in private health insurance there is an additional incentive for excess zeros: the policyholder receives a premium refund by the end of the year when not claiming a single reimbursement. Based on claim frequency and claim size models one can construct models for the yearly total claim which will be a continuous random variable only given that at least one claim occured. The interest, however, often lies in modelling claims in general. If one allows for claim frequencies of zero, the yearly total claim distribution will have an additional point mass at zero. Using a copula approach to model dependency of the yearly total claims thus requires the use of discrete as well as continuous copula properties. We will develop such an approach and estimate parameters using maximum likelihood.

In this paper, we also utilize pair-copula constructions (PCC's) of general multivariate distributions. We model multivariate data using a cascade of pair-copulas, acting on two variables at a time. Pair-copula decompositions build on the work on vines of Joe (1996), Bedford & Cooke (2001a), Bedford & Cooke (2001b) and Bedford & Cooke (2002). For high-dimensional distributions there are many possible pair-copula decompositions for the same multivariate distribution. Bedford & Cooke (2001b) introduced a graphical model called regular vine to help organize them. They also identified two important subclasses of regular vines, which they called C- and D-vines. Pair-copula decomposed models also represent a very flexible class of higher-dimensional copulas. While Kurowicka & Cooke (2006) considered nonstandard estimation methods, Aas et al. (2009) used maximum likelihood for statistical inference and explored the flexibility to model financial time series. There are several advantages of using PCC's: a T-dimensional multivariate density of continuous margins will be expressed as a product of marginal densities and bivariate copulas with individual parameters each. Therefore, in high dimensions T the numerical evaluation of the joint density is very tractable. Each pair of margins can be modelled separately, i.e. the copula class and hence tail dependence properties can be chosen individually. Also, since Archimedean copulas (see e.g. Nelsen (2006, Chapter 4)) are capable only of modelling exchangeable correlation structures, PCC's provide a possibility for generalizing the correlation structure. Finally, model selection in the sense of eliminating weakly correlated copula densities from the joint density can be facilitated.

The paper is innovative with regard to the following aspects: first of all, we present a novel opportunity for modeling the joint density of total claims including zero claims based on copulas for binary and continuous margins. We illustrate how PCC's can be utilized under marginals. Finally we present a novel approach to choose the copula when the margins are discrete. Our model will allow to model the dependency of large claim portfolios in the presence of zero observations.

This paper is organized as follows: in Section 2 we will give a short review of the concept of copulas and illustrate how multivariate distributions can be constructed using pair-copula constructions. In Section 3 an appropriate model for dependent yearly total claims including the zero will be developed: while Subsection 3.1 deals with the aggregation to yearly totals, Subsection 3.2 addresses the problem of specifying a copula based model dependent for claim totals and zero-claim events. An application to health insurance including a detailed illustration of how to deal with the copula choice problem will be given in Section 4. We conclude with a summary and discussion.

2 Copulas and multivariate distributions

A J-dimensional copula C_J is a multivariate cdf $C_J : [0,1]^J \to [0,1]$ whose univariate margins are uniform on $[0,1]$, i.e. $C_J(1,\dots,1,u_j,1,\dots,1) = u_j \quad \forall j \in \{1,\dots,J\}$. For J continuous random variables (rv) $\boldsymbol{X} := (X_1,\dots,X_J)'$ with marginal distributions $F_1,\dots,$ F_J and densities $f_1,\dots,f_J$, all transformed rv's $U_j := F_j(X_j)$, $j = 1,\dots,J$ are uniform on $[0,1]$, hence while F_j reflects the marginal distribution of X_j, C_J reflects the dependency. Sklar (1959) shows that

$$F_{\boldsymbol{X}}(x_1,\dots,x_J) = C_J(F_1(x_1),\dots,F_J(x_J)|\boldsymbol{\zeta}), \tag{1}$$

where $\boldsymbol{\zeta}$ are the corresponding copula parameters. If a multivariate cdf of $\boldsymbol{X}$ exists, there also exists a copula C_J which separates the dependency structure from the marginal distributions. If the margins are continuous, C_J is unique. Vice versa, according to (1) we can construct a multivariate cdf from J marginal distributions using a J-dimensional copula C_J. For a more detailed introduction to copulas, see for instance Joe (1997) or Nelsen (2006). Definitions of some elliptical and Archimedean copulas together with their bivariate densities can be found in Appendix A.

While in this paper we use J dimensional copulas to model dependent discrete margins, a pair-copula construction (PCC) of the joint density will be utilized to describe the dependence of continuous margins. A PCC consists of a cascade of pair-copulas, acting on two variables at a time. In high dimensions there are many different PCC's possible. Bedford & Cooke (2001b) and Bedford & Cooke (2002) show that they can represent such a PCC in a sequence of nested trees with undirected edges, which they call regular vine. One distinguishes between the classes of C and D vines where in the trivariate case these classes coincide. In the following we will illustrate the construction of a C-vine: a multivariate density can be expressed as a product of conditional densities, i.e.

$$f(x_1,...,x_J) = f(x_J|x_1,\cdots,x_{J-1})f(x_1,\cdots,x_{J-1}) = \prod_{j=2}^{J} f(x_j|x_1,\cdots,x_{J-1}) \cdot f(x_1). \tag{2}$$

Here $F(\cdot|\cdot)$ and $f(\cdot|\cdot)$ denote conditional cdf's and densities, respectively. Using Sklar's theorem applied to conditional bivariate densities we can express $f(x_J|x_1,\cdots,x_{J-1})$ as

$$f(x_J|x_1,\cdots,x_{J-1}) \quad = \quad \frac{f(x_{J-1},x_J|x_1,\cdots,x_{J-2})}{f(x_{J-1}|x_1,\cdots,x_{J-2})}$$

$$= c_{J-1,J|1,\cdots,J-2} \cdot f(x_J|x_1,\cdots,x_{J-2}). \tag{3}$$

Here we use for arbitrary distinct indices $i, j, i_1, \cdots, i_k$ with $i < j$ and $i_1 < \cdots < i_k$ the following abbreviation for a bivariate conditional copula density evaluated at conditional cdf's:

$$c_{i,j|i_1,\cdots,i_k} := c_{i,j|i_1,\cdots,i_k}(F(x_i|x_{i_1},\cdots,x_{i_k}), F(x_j|x_{i_1},\cdots,x_{i_k})).$$

Joe (1996) showed that for a d-dimensional vector $\boldsymbol{\nu}$ and a reduced vector $\boldsymbol{\nu}_{-j}$ equal to $\boldsymbol{\nu}$ but without component j the conditional cdf can be obtained recursively by

$$F(x|\boldsymbol{\nu}) = \frac{\partial C\left(F(x|\boldsymbol{\nu}_{-j}), F(\nu_j|\boldsymbol{\nu}_{-j})\right)}{\partial F(x|\boldsymbol{\nu}_{-j})}. \tag{4}$$

A detailed proof of this can be found for example in Czado et al. (2009). For the special case where $\boldsymbol{\nu} = \{\nu\}$ it follows that $F(x|\nu) = \frac{\partial C(F(x),F(\nu))}{\partial F(\nu)}$. For the uniform margins $U := F(x)$ and $V := F(v)$ we define a function

$$h(u|v) := \frac{\partial\, C(u,v)}{\partial v}. \tag{5}$$

This h function has been derived explicitly for many copulas by Aas et al. (2009). A summary of the ones used in this paper is given in Table 12 in Appendix A. By recursive use of (3) one can express the product of conditional densities (2) by

$$\begin{aligned} f(x_1,...,x_J) &= f(x_1) \cdot \prod_{j=2}^{J} \prod_{k=1}^{j-1} c_{j-k,j|1,\cdots,j-k-1} \cdot f(x_j) \\ &= \prod_{r=1}^{J} f(x_r) \cdot \prod_{j=2}^{J} \prod_{k=1}^{j-1} c_{j-k,j|1,\cdots,j-k-1}. \end{aligned} \tag{6}$$

For $k = j-1$ the conditioning set in $c_{j-k,j|1,\cdots,j-k-1}$ is empty, i.e. we set $c_{1,j|10} := c_{1,j}$. In the trivariate there are only three theoretical decompositions of $f(x_1, x_2, x_3)$ (ignoring the possibility of choosing different bivariate copula classes), whereas in higher dimensions, there are many more. On possible decomposition is obtained by using $f(x_1, x_2, x_3) = f_1(x_1)f_{2|1}(x_2|x_1)f_{3|12}(x_3|x_1, x_2)$. Then the PCC is given by

$$\begin{aligned} f(x_1,x_2,x_3) &= c_{12}(F_1(x_1), F_2(x_2))\; c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)) \\ &\quad \cdot c_{13}(F_1(x_1), F_3(x_3)) \prod_{j=1}^{3} f_j(x_j). \end{aligned} \tag{7}$$

The joint density of pairs of margins corresponding to the PCC in (7) can be written as

$$\begin{aligned} f(x_1,x_2) &= \int f_1(x_1)f_{2|1}(x_2|x_1)f_{3|12}(x_3|x_1,x_2)dx_3 = f_1(x_1)f_{2|1}(x_2|x_1) \\ &= c_{12}(F_1(x_1), F_2(x_2)) \prod_{j=1}^{2} f_j(x_j), \end{aligned}$$

and similarly $f(x_1, x_3) = c_{13}(F_1(x_1), F_3(x_3)) \cdot f_1(x_1)f_3(x_3)$. The final margin requires integration, i.e.

$$\begin{aligned} f(x_2, x_3) &= \int f_1(x_1) f_{2|1}(x_2|x_1) f_{3|12}(x_3|x_1, x_2) dx_1 \\ &= \int c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23|1}(h(F_2(x_2)|F_1(x_1)), h(F_3(x_3)|F_1(x_1))) \\ &\quad \cdot c_{13}(F_1(x_1), F_3(x_3)) \prod_{j=1}^{3} f_j(x_j) dx_1 \\ &= \int_0^1 c_{12}(u_1, u_2) \cdot c_{23|1}(h(u_2|u_1), h(u_3|u_1)) \cdot c_{13}(u_1, u_3) \\ &\quad \cdot \prod_{j=2}^{3} f_j(F_j^{-1}(u_j)) du_1, \end{aligned} \tag{8}$$

where we substitute $u_j = F_j(x_j)$ and transform $dx_1 = \frac{1}{f_1(x_j)} du_1$. For most copula choices, the integral in (8) can only be calculated numerically.

3 A model for dependent yearly claim totals

In this section we aim to develop a joint model for yearly dependent total claims including zero claims. One possible approach to this has been developed by Frees & Valdez (2008) and has been applied to car accident claims where payments may occur in three different correlated random classes. They point out that this is a nonstandard problem since all three claim types are rarely observed simultaneously. In their approach, the combination of claims and zero claims is modelled by a multinomial logit model. We will model yearly total claims for a certain claim type and utilize a copula to obtain a joint model. In general, we assume that we have J dependent yearly claims available. For a population of insured individuals this may be J different treatment fields. A zero claim may arise for different reasons: first of all, a healthy patient simply had no need to see a physician. Secondly, the invoice may be below a deductible. Thirdly, the insured person will get a premium refund when not recovering a single bill throughout the year and opts for this when expecting the refund to be higher than the invoice. In health insurance we consider the dependence betweeen ambulant, inpatient and dental treatments. Zero claim events will certainly be dependent due to the health status of an insured person. The deductible will not have an impact on the dependency of the zero claim events since they apply separately for the three fields. The premium refund, however, will only be paid if no reimbursement is claimed in either of these fields. Therefore, it will also influence the dependence.

3.1 Aggregation of claim frequencies and sizes to yearly totals

We will express the yearly total claim T_j in field j as

$$T_j := W_j \cdot 0 + (1 - W_j) N_j \bar{S}_j = (1 - W_j) T_j^+ \geq 0.$$

Here W_j is a binary indicator for the zero claim event, i.e. $W_j = 1$ if the claim is zero and $W_j = 0$ else. Also, $\boldsymbol{W} := (W_1, \ldots, W_T)'$ and $\boldsymbol{W}_{-j} := (W_1, \ldots, W_{j-1}, W_{j+1}, \ldots, W_T)'$. Further, N_j is the positive number of claims and $\bar{S}_j$ the average claim size (also strictly positive). In general we will observe only $\bar{S}_j|\{W_j = 0\}$ but we assume that $\bar{S}_j|\{W_j = 1\}$ exists. Let the total claim $T_j^+ := N_j\bar{S}_j > 0$ if $W_j = 0$ and unknown but positive if $W_j = 1$. A more general case is given when the single claims S_{kj}, $k = 1, \ldots, N_j$, which contribute to the yearly total, are known and are i.i.d., i.e. if $T_j^+ := \sum_{k=1}^{N_j} S_{kj}$. Then the distribution of T_j^+ will be obtained by convolution and can be approximated for example using the methods summarized in the R package 'actuar' (see Dutang et al. (2008)). This paper, however, will focus on average claims but apart from the expression for total claims all approaches carried out in this paper apply in a similar way.
We denote probability mass functions (pmf) by p and their cumulative distribution functions (cdf) by P. Probability density functions (pdf) and cdf of continuous random variables are denoted by f and F, respectively. The following distributions for the zero claim event, claim frequencies and sizes are assumed:

$$\begin{aligned} W_j &\sim \mathit{binary}(p_{W_j}(1)), \text{ with pmf } p_{W_j}, \text{ cdf } P_{W_j}, \\ N_j|\{W_j = 0, \boldsymbol{W}_{-j}\} &\sim P_{N_j|\{W_j=0,\boldsymbol{W}_{-j}\}}, \text{ positive no. of claims, pmf } p_{N_j|\{W_j=0,\boldsymbol{W}_{-j}\}}, \\ \bar{S}_j|\{N_j, W_j = 0, \boldsymbol{W}_{-j}\} &\sim F_{\bar{S}_j|\{N_j,W_j=0,\boldsymbol{W}_{-j}\}}, \text{ average claim size, pdf } f_{\bar{S}_j|\{N_j,W_j=0,\boldsymbol{W}_{-j}\}}. \end{aligned}$$

For data following these distributions regression models may be fitted. Then the realizations $\boldsymbol{w}_{-j}$ of $\boldsymbol{W}_{-j}$ are used as regressors in the latter two models and additionally the realizations n_j of N_j in the last one. It follows that $T_j^+ := N_j\bar{S}_j$ is conditionally independent of $\boldsymbol{W}_{-j}$. Note that one only fits these two regression models to data with $W_j = 0$. The distribution of N_j may be modelled using a zero-truncated count distribution (see for example Zuur et al. (2009, Chapter 11)) which can be constructed based on any count distribution. For example, let N_c follow some count distribution (Poisson, Negative-Binomial etc.) with pmf p_c, then $N \sim P_N$ with pmf

$$p_N(n) := \frac{p_c(n)}{1 - p_c(0)}, \quad n = 1, 2, \ldots$$

will be the zero-truncated representative of this count distribution.

Example 3.1. *For the Negative Binomial (NB) distribution with mean parameter $\mu > 0$, shape parameter $r > 0$ and variance $\mu\left(1 + \frac{\mu}{r}\right)$ a zero-truncated Negative Binomial (ZTNB) distribution has the pmf*

$$p_N(n) \ := \ \frac{1}{1 - \left(\frac{r}{\mu+r}\right)^r} \cdot \frac{\Gamma(n+r)}{\Gamma(r)\Gamma(n!)} \cdot \left(\frac{r}{\mu+r}\right)^r \cdot \left(\frac{\mu}{\mu+r}\right)^n, \quad n = 1, 2, \ldots.$$

We can now model and quantify the dependency of the vectors $\boldsymbol{W}$ and $\boldsymbol{T}$. On the one hand, the number of zero claims $\boldsymbol{W}$ reflects the impact of the health insurancer's incentives for not having a single claim throughout the year, which the insurer wants to know about in order to arrange its deductibles and premium refund policy. On the other hand, the dependence of $\boldsymbol{T}$ can be used for premium and risk capital calculation.

Let F_{T_j} the cdf of T_j, then a derivative at $t = 0$ does not exist and therefore only the derivative conditional on $\{W_j = 0\}$ may be called a density function. Similar to Heller et al. (2007) we simply refer to f_{T_j} as the probability function (pf) of T_j. Jørgensen & de Souza (1994) and Smyth & Jørgensen (2002) consider models for continuous claim sizes including zero claims. These are based on the class of Tweedie distributions (Tweedie (1984)), which are members of the exponential family. In particular they use a compound Poisson-gamma distribution which is contained in the class of the Tweedie distributions. Belasco & Ghosh (2008) develop a model based on the Tobit model (Tobin (1958)) in which a zero outcome arises from left-censoring. The marginal distributions of T_j^+ given N_j, $W_j = 0$ and $\boldsymbol{W}_{-j}$ will be denoted by

$$T_j^+|\{N_j, W_j = 0, \boldsymbol{W}_{-j}\} \sim F_{T_j^+|\{N_j,W_j=0,\boldsymbol{W}_{-j}\}}, \text{ with pdf } f_{T_j^+|\{N_j,W_j=0,\boldsymbol{W}_{-j}\}}.$$

For the moment, we drop the index j for field j and also the dependency on $\boldsymbol{W}_{-j}$.

Lemma 3.2. *For average claims $\bar{S}$, cdf and pdf of $T^+|\{W = 0\}$ are given by*

$$\begin{aligned} F_{T^+|\{W=0\}}(t^+) &= \sum_{k=1}^{\infty} F_{\bar{S}|\{N,W=0\}}\left(\frac{t^+}{k}|\{N = k\}\right) p_{N|\{W=0\}}(k) \\ f_{T^+|\{W=0\}}(t^+) &= \sum_{k=1}^{\infty} f_{\bar{S}|\{N,W=0\}}\left(\frac{t^+}{k}|\{N = k\}\right) p_{N|\{W=0\}}(k). \end{aligned}$$

Proof. See Appendix B. □

Lemma 3.3. *The cdf of the yearly total claim T at t is*

$$F_T(t) = p_W(0) + \mathbb{1}_{\{t>0\}}(1 - p_W(0))F_{T^+|\{W=0\}}(t). \tag{9}$$

Proof. See Appendix B. □

3.2 A joint distribution of yearly total claims based on copulas

In this section we develop a joint distribution of $\boldsymbol{T} := (T_1, \ldots, T_J)'$. Utilizing copulas in order to model dependency between $\boldsymbol{T} := (T_1, \ldots, T_J)'$ is nonstandard since according to Lemma 3.3, $U_j := F_{T_j}(T_j)$ will have a point mass at 0 and hence U_j will not be uniform on $[0, 1]$. Nevertheless, we will develop a joint distribution of T_j, $j = 1, \ldots, J$, based on two copula constructions, one with discrete margins $\boldsymbol{W} := (W_1, \ldots, W_J)'$ and one with continuous margins $\boldsymbol{T}^+ := (T_1^+, \ldots, T_J^+)'$. We allow $\boldsymbol{W}$ and $\boldsymbol{T}^+$ to be dependent random vectors and use the conditional independence between $\boldsymbol{W}$ and $\boldsymbol{T}|\boldsymbol{W}$, i.e. we use

$$P(T_j \le t_j, W_j = w_j, \forall j) = P((1 - W_j)T_j^+ \le t_j|\{W_j = w_j\}, \forall j) \cdot P(W_j = w_j, \forall j).$$

Here $p_{\boldsymbol{W}} := P(W_j = w_j, \forall j = 1, \ldots, J)$ can be obtained by constructing $P_{\boldsymbol{W}}$ by a J dimensional copula and using the formula of Song (2007)[p. 128] to obtain the joint pmf. For $\boldsymbol{T}^+$ a joint pdf $f_{\boldsymbol{T}^+|\boldsymbol{W}}$ and hence a joint cdf $F_{\boldsymbol{T}^+|\boldsymbol{W}}$ may be constructed using a PCC. We stress that the PCC is utilized for $\boldsymbol{T}^+$, which is unobserved for some observations of t_j^+ but nevertheless we use the conditional independence of $\boldsymbol{W}$ and $\boldsymbol{T}|\boldsymbol{W}$. The joint distribution of yearly total claims $\boldsymbol{T}$ and zero claim events $\boldsymbol{W}$ will be given in Proposition 3.4.

Proposition 3.4. *Let* $T_j = (1 - W_j) \cdot T_j^+$ *and* $J^0(\boldsymbol{w}) := \{j \in \{1, \ldots, J\} : W_j = 0\} = \{j_{1(\boldsymbol{w})}, \ldots, j_{n(\boldsymbol{w})}\}$ *with* $n(\boldsymbol{w})$ *the cardinality of* $J^0(\boldsymbol{w})$. *Then the joint probability function of* $\boldsymbol{T}$ *and* $\boldsymbol{W}$ *is given by*

$$f_{\boldsymbol{T},\boldsymbol{W}}(\boldsymbol{t}, \boldsymbol{w}) = p_{\boldsymbol{W}}(\boldsymbol{w}) f_{T^+_{1(\boldsymbol{w})},\ldots,T^+_{n(\boldsymbol{w})}|\boldsymbol{W}}(t^+_{1(\boldsymbol{w})}, \ldots, t^+_{n(\boldsymbol{w})}|\boldsymbol{w})$$

where $f_{T^+_{1(\boldsymbol{w})},\ldots,T^+_{n(\boldsymbol{w})}|\boldsymbol{W}}(t^+_{1(\boldsymbol{w})}, \ldots, t^+_{n(\boldsymbol{w})}|\boldsymbol{w})$ *is the joint pdf of* $\boldsymbol{T}^+$ *where all margins with* $W_j = 1$, $j = 1, \ldots, J$ *are integrated out.*

Proof. We consider

$$\begin{aligned} P(T_j \leq t_j, j = 1, \ldots, J, \boldsymbol{W} = \boldsymbol{w}) &= P((1 - W_j)T_j^+ \leq t_j, \forall j|\boldsymbol{W} = \boldsymbol{w}) \cdot p_{\boldsymbol{W}}(\boldsymbol{w}) \\ &= P(T^+_{j_k} \leq t_{j_k}, j_k \in J^0(\boldsymbol{w})|\boldsymbol{W} = \boldsymbol{w}) \cdot p_{\boldsymbol{W}}(\boldsymbol{w}). \end{aligned}$$

The joint probability function is obtained by deriving for t_{j_k}, $j_k \in J^0(\boldsymbol{w})$, i.e.
$f_{T^+_{1(\boldsymbol{w})},\ldots,T^+_{n(\boldsymbol{w})}|\boldsymbol{W}}(t^+_{1(\boldsymbol{w})}, \ldots, t^+_{n(\boldsymbol{w})}|\boldsymbol{w}) = \frac{\partial}{\partial t_{1(\boldsymbol{w})}} \cdots \frac{\partial}{\partial t_{n(\boldsymbol{w})}} P(T^+_{j_k} \leq t_{j_k}, j_k \in J^0(\boldsymbol{w})|\boldsymbol{W} = \boldsymbol{w})$. □

So whenever an observation $T_j^+|\{W_j = 1\}$ is unknown, the margin in the corresponding PCC is integrated out. Hence the distribution of the vector $\boldsymbol{T}|\boldsymbol{W}$ is defined for strictly positive numbers.

Example 3.5. *For* $J = 3$,

$$\begin{aligned} f_{\boldsymbol{T},\boldsymbol{W}}(t_1, t_2, t_3, w_1, w_2, w_3) &= p_{\boldsymbol{W}}(w_1, w_2, w_3) \cdot \Big[\mathbb{1}_{\{\boldsymbol{w}=(1,1,1)'\}} + \mathbb{1}_{\{\boldsymbol{w}=(1,1,0)'\}} f_{T_3^+}(t_3) \\ &\quad + \mathbb{1}_{\{\boldsymbol{w}=(1,0,1)'\}} f_{T_2^+}(t_2) + \mathbb{1}_{\{\boldsymbol{w}=(0,1,1)'\}} f_{T_1^+}(t_1) \\ &\quad + \mathbb{1}_{\{\boldsymbol{w}=(1,0,0)'\}} f_{T_2^+,T_3^+}(t_2, t_3) + \mathbb{1}_{\{\boldsymbol{w}=(0,1,0)'\}} f_{T_1^+,T_3^+}(t_1, t_3) \\ &\quad + \mathbb{1}_{\{\boldsymbol{w}=(0,0,1)'\}} f_{T_1^+,T_2^+}(t_1, t_2) + \mathbb{1}_{\{\boldsymbol{w}=(0,0,0)'\}} f_{\boldsymbol{T}^+}(t_1, t_2, t_3)\Big]. \end{aligned}$$

We define $P_{\boldsymbol{W}}(w_1, \ldots, w_J|\boldsymbol{\zeta}^W) := C_J(P_{W_1}(w_1), \ldots, P_{W_J}(w_J)|\boldsymbol{\zeta}^W)$ by a copula cdf C_J in dimension J with copula parameters $\boldsymbol{\zeta}^W$. For binary margins,

$$\begin{aligned} p_{\boldsymbol{W}}(w_1, \ldots, w_J|\boldsymbol{\zeta}^W) &= \sum_{j_1=0}^{w_1} \cdots \sum_{j_J=0}^{w_J} (-1)^{\sum_{k=1}^J (j_k + w_k)} \\ &\quad \cdot C_J(P_{W_1}(j_1), \ldots, P_{W_J}(j_J)|\boldsymbol{\zeta}^W). \end{aligned} \tag{10}$$

Proof. According to Song (2007)[p. 128]

$$p_{\boldsymbol{W}}(w_1, \ldots, w_J|\boldsymbol{\zeta}^W) = \sum_{k_1=0}^{1} \cdots \sum_{k_J=0}^{1} (-1)^{k_1+\ldots+k_J} C_J\big(u_{1k_1}(w_1), \ldots, u_{Jk_J}(w_J)|\boldsymbol{\zeta}^W\big),$$

where $u_{t0}(w_t) := P_{W_t}(w_t)$ and $u_{t1}(w_t) := P_{W_t}(w_t - 1)$. Now $u_{t0}(1) = P_{W_t}(1) = 1$, $u_{t1}(1) = P_{W_t}(0)$ and $u_{t0}(0) = P_{W_t}(0)$. Since $u_{t1}(0) = P_{W_t}(-1) = 0$ and $C_J(\ldots, 0, \ldots|\boldsymbol{\zeta}^W) = 0$ we only need to consider $k_t \leq w_t$. By transforming $j_t := w_t - k_t \geq 0$ we obtain the required result. □

Note that $P_{W_j}(1) = 1$, $j = 1, \ldots, J$. In this case and given we use an elliptical or Archimedean copula, the copula in (10) at such a marginal probability will be of the same class only with decreased dimension. On the other hand, for the continuous random vector $\boldsymbol{T}^+$ we define $F_{\boldsymbol{T}^+}(t_1^+, \ldots, t_J^+|\boldsymbol{\zeta}^{T^+})$ by a PCC introduced in (7).
We return to the trivariate case ($J = 3$). The bivariate marginal distributions are defined according to (8), where $f_{T_2^+,T_3^+}(t_{i2}^+, t_{i3}^+)$ requires numerical integration. Let $\boldsymbol{\zeta}^W$ the parameters of the copula of $\boldsymbol{W}$ and $\boldsymbol{\zeta}^{T^+} := (\boldsymbol{\zeta}_{12}^{T^+}, \boldsymbol{\zeta}_{13}^{T^+}, \boldsymbol{\zeta}_{23|1}^{T^+})'$ the parameters of the PCC of $\boldsymbol{T}^+$. Since the expression depending on $\boldsymbol{\zeta}^W$ is independent of the expression depending on $\boldsymbol{\zeta}^{T^+}$, i.e. for $\boldsymbol{\zeta} := (\boldsymbol{\zeta}^W, \boldsymbol{\zeta}^{T^+})'$, the log-likelihood is $l(\boldsymbol{\zeta}) = l(\boldsymbol{\zeta}^W) + l(\boldsymbol{\zeta}^{T^+})$, which in a maximum likelihood context can be fitted separately over those two parameter sets. For observations $i = 1, \ldots, I$,

$$
\begin{aligned}
l(\boldsymbol{\zeta}^W) &= \sum_{i=1}^{I} \log(p_{\boldsymbol{W}}(w_{i1}, w_{i2}, w_{i3})) \\
&= \sum_{i=1}^{I} \log\left(\sum_{j_1=0}^{w_{i1}} \sum_{j_2=0}^{w_{i2}} \sum_{j_3=0}^{w_{i3}} (-1)^{\sum_{k=1}^{3}(j_k + w_{ik})} C_J(P_{W_1}(j_1), P_{W_2}(j_2), P_{W_3}(j_3)|\boldsymbol{\zeta}^W) \right), \\
l(\boldsymbol{\zeta}^{T^+}) &= \sum_{i=1}^{I} \Big[\mathbb{1}_{\{w_{i1}=1, w_{i2}=0, w_{3}=0\}} \cdot \log(f_{T_2^+,T_3^+}(t_{i2}^+, t_{i3}^+|\boldsymbol{\zeta}^{T^+})) \\
&\quad + \mathbb{1}_{\{w_{i1}=0, w_{i2}=1, w_{i3}=0\}} \cdot \log(c_{13}(F_{T_1^+}(t_{i1}^+), F_{T_3^+}(t_{i3}^+)|\boldsymbol{\zeta}_{13}^{T^+})) \\
&\quad + \mathbb{1}_{\{w_{i1}=0, w_{i2}=0, w_{i3}=1\}} \cdot \log(c_{12}(F_{T_1^+}(t_{i1}^+), F_{T_2^+}(t_{i2}^+)|\boldsymbol{\zeta}_{12}^{T^+})) \\
&\quad + \mathbb{1}_{\{w_{i1}=0, w_{i2}=0, w_{i3}=0\}} \cdot \Big[\log(c_{12}(F_{T_1^+}(t_{i1}^+), F_{T_2^+}(t_{i2}^+)|\boldsymbol{\zeta}_{12}^{T^+})) \\
&\quad + \log(c_{23|1}(h(F_{T_2^+}(t_{i2}^+)|F_{T_1^+}(t_{i1}^+), \boldsymbol{\zeta}_{12}^{T^+}), h(F_{T_3^+}(t_{i3}^+)|F_{T_1^+}(t_{i1}^+), \boldsymbol{\zeta}_{13}^{T^+})|\boldsymbol{\zeta}_{23|1}^{T^+})) \\
&\quad + \log(c_{13}(F_{T_1^+}(t_{i1}^+), F_{T_3^+}(t_{i3}^+)|\boldsymbol{\zeta}_{13}^{T^+})) \Big] \Big] + \text{ const. independent of } \boldsymbol{\zeta}^{T^+}.
\end{aligned}
$$

4 Application to health insurance data

We will consider data from a German private health insurer. Each record represents one out of 37 819 insured persons. Claim frequencies will be the number of benefits received by an insured person, where a benefit may be any treatment or prescription balanced to a patient, i.e. a patient usually gets charged for several benefits during one visit. Claim sizes will be the average invoice, i.e. the yearly total amount divided by the number of benefits. Responses as well as explanatory variables have been observed in the ambulant (i.e. outpatient), inpatient and dental field over three years from 2005 to 2007. We will abbreviate the treatment fields by 'A' for ambulant, 'I' for inpatient and 'D' for dental or indices $j = 1, \ldots, 3$, respectively. Around 76% of the insured persons are male, which is typical for the policy line considered. All policyholders in the population are covered in all three fields. The private German health care system allows for deductibles, which - depending on policy type and treatment - may be a specific amount for a certain benefit or a percentage of the amount invoiced. Policyholders not handing in a single bill for a whole

Variable	Description
Responses	
W_{ijt}	Zero claim event (1 if zero claim) by patient i in treatment field j and year t.
$N_{ijt} \| \{W_{ijt} = 0\}$	Total positive number of benefits received by patient i in treatment field j and year t.
T_{ijt}	Total invoice for patient i in treatment field j and year t (including deductibles).
$\bar{S}_{ijt} \| \{W_{ijt} = 0\}$	T_{ijt}/N_{ijt}, average invoice of patient i in treatment field j and year t.
Covariates	
	CATEGORICAL
SEX_i	Dummy for gender of patient i.
	DISCRETE
AGE_{it}	Age of patient i at December, 31 in year t.
	CONTINUOUS
DED_{ijt}	Total of all deductibles of $\bar{S}_{ijt}$ of patient i in treatment field j and year t.
$\overline{DED}_{ijt}$	Average deductible of patient i in treatment field j and year t
	SPATIAL
ZIP_i	ZIP code of the home address of patient i as of Dec. 31, 2007.
$D(i)$	Dummy for home district of patient i as of Dec. 31, 2007. There are 439 German districts. Individuals are spread over all districts.
	CONTINUOUS WITH SPATIAL INFORMATION
$PHYS.INH_{D(i)}$	(number of physicians in district $D(i)$ listed in the yellow pages as of April 15, 2008 divided by the number of inhabitants in district $D(i)$ in 2007) $\cdot 100$.
$URBAN_{D(i)}$	Number of inhabitants per square kilometer in district $D(i)$ in 2007
$BP_{D(i)}$	Average buying power in Euro in 2007 in district $D(i)$ on a scale of nine scoring levels. Buying power has been determined as the average net income per district + public transfer payments.

Table 1: Description of variables considered for claim frequencies and claim size models for the health insurance data

year in any of the three fields will get a premium refund. Therefore, we might not see the actual treatment numbers and amounts invoiced in the data. A variable description including responses and explanatory variables will be given in Table 1. The data has been supplemented by data from different sources:

- a mapping of ZIP codes to 439 districts not including corporate ZIP codes (*http://www.manfrin-it.com/postleitzahlen/plz.html*), completed by single queries for missing ZIP codes from *http://w3logistics.com/infopool/plz/index.php*,
- number of physicians per ZIP code listed in the yellow pages from 8233 automated web requests searching for 'Arzt' (physician) to *http://web2.cylex.de*,

- number of inhabitants and area in square kilometers of each of the 439 German districts according to the GfK GeoMarketing GmbH (*http://www.gfk-geomarketing.de/marktdaten/samples.php*),
- data transcribed from a map displaying the buying power per district by the GfK GeoMarketing GmbH *http://www.gfk-geomarketing.de/presse/bdm/html/01_2007.html*, reference *Grafik: GfK GeoMarketing.*

Fitting marginal distributions first and fitting the copula parameters for fixed margins afterwards is known as inference functions for margins or the IFM method (see for example Joe (1997, Section 10.1)). In the following subsections we will briefly summarize the regression models chosen for W_{jt}, N_{jt} and $\bar{S}_{jt}$, $j, t = 1, 2, 3$.

4.1 Marginal zero claim event models

Consider a logistic regression model for W_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, J$, i.e.

$$W_{ij} \sim binary\left(\frac{\exp(\boldsymbol{x}_{ij}^{Wt}\boldsymbol{\beta}_j)}{1+\exp(\boldsymbol{x}_{ij}^{Wt}\boldsymbol{\beta}_j)}\right).$$

We choose variables by backward selection based on the Wald test with a 5% significance level. The model equations of the reduced designs are given in Table 2. An exemplary summary of the regression model for W_{A7} is given in Table 3.

Model equations
$W_{A5} \sim 1 + \mathbb{1}_{DED_{A5} \leq 100} + \mathbb{1}_{AGE_5 \leq 32} \cdot AGE_5 + \mathbb{1}_{AGE_5 > 32} \cdot AGE_5 + SEX + BP$
$W_{A6} \sim 1 + \mathbb{1}_{DED_{A6} \leq 100} + \mathbb{1}_{AGE_6 \leq 32} \cdot AGE_6 + \mathbb{1}_{AGE_6 > 32} \cdot AGE_6 + SEX + BP$
$W_{A7} \sim 1 + \mathbb{1}_{DED_{A7} \leq 100} + \mathbb{1}_{AGE_7 \leq 32} \cdot AGE_7 + \mathbb{1}_{AGE_7 > 32} \cdot AGE_7 + SEX + BP$
$W_{I5} \sim 1 + \mathbb{1}_{DED_{I5} = 0} + \mathbb{1}_{AGE_5 \leq 32} \cdot AGE_5 + \mathbb{1}_{AGE_5 > 32} \cdot AGE_5 + SEX$
$W_{I6} \sim 1 + \mathbb{1}_{DED_{I6} = 0} + \mathbb{1}_{AGE_6 \leq 32} \cdot AGE_6 + \mathbb{1}_{AGE_6 > 32} \cdot AGE_6 + SEX$
$W_{I7} \sim 1 + \mathbb{1}_{DED_{I7} = 0} + \mathbb{1}_{AGE_7 \leq 32} \cdot AGE_7 + \mathbb{1}_{AGE_7 > 32} \cdot AGE_7 + SEX + PHYS.INH$
$W_{D5} \sim 1 + \mathbb{1}_{DED_{D5} = 0} + \mathbb{1}_{AGE_5 > 32} \cdot AGE_5 + SEX + PHSY.INH + BP$
$W_{D6} \sim 1 + \mathbb{1}_{DED_{D6} = 0} + \mathbb{1}_{AGE_6 \leq 32} \cdot AGE_6 + \mathbb{1}_{AGE_6 > 32} \cdot AGE_6 + SEX + BP$
$W_{D7} \sim 1 + \mathbb{1}_{DED_{D7} = 0} + \mathbb{1}_{AGE_7 \leq 32} \cdot AGE_7 + \mathbb{1}_{AGE_7 > 32} \cdot AGE_7 + SEX + PHYS.INH + BP$

Table 2: Reduced model equations for each of the nine logistic regression models for W_{jt}, $j = 1, 2, 3 = A, I, D$; $t = 5, 6, 7$ after applying sequential backward selection based on the Wald test

4.2 Marginal claim frequency models

Let N_{ij}, $i \in \mathcal{I}_j := \{i = 1, \ldots, n, W_{ij} = 0\}$, $j = 1, \ldots, J$ follow the zero-truncated negative binomial distribution (ZTNB) defined in Example 3.1. Further let $\boldsymbol{w}_{i(-j)} :=$

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-9.4914	0.9951	-9.54	$< 2 \cdot 10^{-16}$
$\mathbb{1}_{DED_{A7} \leq 100}$	10.9182	0.9945	10.98	$< 2 \cdot 10^{-16}$
$\mathbb{1}_{AGE_7 \leq 32} \cdot AGE_7$	0.3480	0.0284	12.26	$< 2 \cdot 10^{-16}$
$\mathbb{1}_{AGE_7 > 32} \cdot AGE_7$	-0.6850	0.0354	-19.35	$< 2 \cdot 10^{-16}$
SEX	0.1965	0.0425	4.63	$3.7 \cdot 10^{-6}$
BP	-0.0795	0.0169	-4.71	$2.5 \cdot 10^{-6}$

Table 3: Model summary for the reduced logistic regression model of W_{A7}.

$(w_{i1}, \ldots, w_{i(j-1)}, w_{i(j+1)}, \ldots, w_{iJ})'$. Then a ZTNB regression model (see e.g. Cruyff & van der Heijden (2008)) is given by

$$\begin{aligned} N_{ij} | \{\boldsymbol{x}_{ij}^N, \boldsymbol{w}_{i(-j)}\} &\sim ZTNB\left(\mu_{ij}(\boldsymbol{x}_{ij}^N, \boldsymbol{w}_{i(-j)}), r_j\right), \\ \mu_{ij}(\boldsymbol{x}_{ij}^N, \boldsymbol{w}_{i(-j)}) &= \exp(\boldsymbol{x}_{ij}^{Nt} \boldsymbol{\gamma}_j^1 + \boldsymbol{w}_{i(-j)} \boldsymbol{\gamma}_j^2). \end{aligned}$$

We utilize the Wald test for backward selection. Thereby we use the observed Fisher

Model equations
$N_{A5} \sim 1 + DED_{A5} + poly(AGE_5)[,1] + poly(AGE_5)[,2] + SEX + DED_{A5} : poly(AGE_5)[,1] + DED_{A5} : poly(AGE_5)[,2] + SEX : poly(AGE_5)[,1] + SEX : poly(AGE_5)[,2] + W_{I5} + W_{D5}$ $N_{A6} \sim 1 + DED_{A6} + poly(AGE_6)[,1] + poly(AGE_6)[,2] + SEX + URBAN + DED_{A6} : poly(AGE_6)[,1] + DED_{A6} : poly(AGE_6)[,2] + SEX : poly(AGE_6)[,1] + SEX : poly(AGE_6)[,2] + W_{I6} + W_{D6}$ $N_{A7} \sim 1 + DED_{A7} + poly(AGE_7)[,1] + poly(AGE_7)[,2] + SEX + URBAN + BP + DED_{A7} : poly(AGE_7)[,1] + DED_{A7} : poly(AGE_7)[,2] + SEX : poly(AGE_7)[,1] + SEX : poly(AGE_7)[,2] + URBAN : BP + W_{I7} + W_{D7}$
$N_{I5} \sim DED_{I5} + AGE_5 + SEX + URBAN + BP + SEX : URBAN + W_{A5} + W_{D5}$ $N_{I6} \sim 1 + DED_{I6} + AGE_6 + BP + W_{A6}$ $N_{I7} \sim DED_{I7} + AGE_7 + BP + DED_{I7} : AGE_7 + W_{A7}$
$N_{D5} \sim 1 + \log(DED_{D5}) + poly(AGE_5)[,1] + poly(AGE_5)[,2] + SEX + URBAN + \log(DED_{D5}) : poly(AGE_5)[,1] + \log(DED_{D5}) : poly(AGE_5)[,2] + SEX : poly(AGE_5)[,2] + poly(AGE_5)[,1] : URBAN + W_{A5} + W_{I5}$ $N_{D6} \sim 1 + \log(DED_{D6}) + poly(AGE_6)[,1] + poly(AGE_6)[,2] + SEX + \log(DED_{D6}) : poly(AGE_6)[,1] + \log(DED_{D6}) : poly(AGE_6)[,2] + SEX : poly(AGE_6)[,1] + SEX : poly(AGE_6)[,2] + W_{A6} + W_{I6}$ $N_{D7} \sim 1 + \log(DED_{D7}) + poly(AGE_7)[,1] + poly(AGE_7)[,2] + SEX + URBAN + BP + log(DED_{D7}) : poly(AGE_7)[,1] + \log(DED_{D7}) : poly(AGE_7)[,2] + SEX : poly(AGE_7)[,1] + SEX : poly(AGE_7)[,2] + poly(AGE_7)[,1] : URBAN + W_{A7}$

Table 4: Reduced model equations for each of the nine ZTNB claim frequency models after applying sequential backward selection based on the Wald test

information based on the numerical Hessian matrix obtained by the R routine *optim*. The reduced model equations are given in Table 4. For N_{A7}, a summary of the reduced design is given in Table 5.

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept	2.216	0.015	151.123	$< 2 \cdot 10^{-16}$
DED_{A7}	0.469	0.006	72.379	$< 2 \cdot 10^{-16}$
$poly(AGE_7)[,1]$	52.296	1.705	30.672	$< 2 \cdot 10^{-16}$
$poly(AGE_7)[,2]$	9.455	1.908	4.955	$7.2 \cdot 10^{-7}$
SEX	−0.286	0.012	−23.794	$< 2 \cdot 10^{-16}$
$URBAN$	0.007	0.005	1.307	0.191
BP	0.003	0.005	0.641	0.521
$DED_{A7} : poly(AGE_7)[,1]$	−8.003	1.148	−6.972	$3.1 \cdot 10^{-12}$
$DED_{A7} : poly(AGE_7)[,2]$	−6.585	1.202	−5.477	$4.3 \cdot 10^{-8}$
$SEX : poly(AGE_7)[,1]$	−14.489	1.887	−7.680	$1.6 \cdot 10^{-14}$
$SEX : poly(AGE_7)[,2]$	21.686	1.992	10.884	$< 2 \cdot 10^{-16}$
$URBAN : BP$	−0.009	0.004	−2.359	0.018
W_{I7}	0.618	0.014	45.614	$< 2 \cdot 10^{-16}$
W_{D7}	0.237	0.011	21.081	$< 2 \cdot 10^{-16}$

Table 5: Model summary for the reduced ZTNB regression model of the claim frequencies for treatment field ambulant in 2007 with dispersion parameter θ is estimated to be 2.15.

4.3 Marginal claim size models

As marginal models for the claim sizes $\bar{S}_{ij}$, $i \in \mathcal{I}_j := \{i = 1, \ldots, n, W_{ij} = 0\}$, $j = 1, \ldots, J$ we aim to use weighted log normal models given by

$$\bar{S}_{ij} | \{\boldsymbol{x}_{ij}^{\bar{S}}, \boldsymbol{w}_{i(-j)}, n_j\} \quad \sim \quad lognormal(\boldsymbol{x}_{ij}^{\bar{S}t} \boldsymbol{\alpha}_j^1 + \boldsymbol{w}_{i(-j)} \boldsymbol{\alpha}_j^2 + n_j \alpha_j^3, \sigma_j, \text{ weights } \omega_{ij}^{\bar{S}}). \quad (11)$$

Since we model average claims rather than actual claim sizes we observe high heteroscedasticity in $\bar{S}_{ij}$ which will depend on the number of claims per year for each observation. As for the logarithmic transformation of the responses in the linear model the exact theoretical influence of N on the heteroscedasticity cannot be determined. We perform a three step approach based on ordinary least square (OLS) regression and weighted least square (WLS) regression (DeMaris (2004, p.201)) with unknown weights. First we fit a log normal OLS regression model based on $\{\boldsymbol{x}_{ij}^{\bar{S}}, \boldsymbol{w}_{i(-j)}, n_j\}$. Now we want to allow for heteroscedasticy using a WLS approach. In order to determine weights $\omega_{ij}^{\bar{S}}$, we regress the OLS squared residuals (as responses) on the model's predictors in another lognormal OLS regression model and use fitted values from this run as variance estimates (see DeMaris (2004, p.205)). In a third step, we replace the OLS model from the first step by the weighted regression (11). Variable selection is carried out using backward selection based on the Wald test. For every design which we consider new weights are determined, i.e. we update the estimated coefficients in order to predict variances. The choice of regressors for determining the weights, however, is not changed throughout the backward selection procedure. The model equations of the reduced fitted models are given in Table 6. A model summary for $\bar{S}_{A7}$ is given in Table 7.

Model equations
$\bar{S}_{A5} \sim 1 + \overline{DED}_{A5} + poly(AGE_5)[,1] + poly(AGE_5)[,2] + SEX + URBAN + BP + \overline{DED}_{A5} : SEX + URBAN : BP + W_{I5} + W_{D5} + N_{A5}$
$\bar{S}_{A6} \sim 1 + \overline{DED}_{A6} + poly(AGE_6)[,1] + poly(AGE_6)[,2] + SEX + PHYS.INH + URBAN + BP + URBAN : BP + W_{I6} + W_{D6} + N_{A6}$
$\bar{S}_{A7} \sim 1 + \overline{DED}_{A7} + poly(AGE_7)[,1] + poly(AGE_7)[,2] + SEX + PHYS.INH + URBAN + BP + \overline{DED}_{A7} : SEX + PHYS.INH : URBAN + W_{I7} + W_{D7} + N_{A7}$
$\bar{S}_{I5} \sim 1 + poly(\overline{DED}_{I5})[,1] + poly(\overline{DED}_{I5})[,2] + poly(AGE_5)[,1] + poly(AGE_5)[,2] + BP + W_{A5}$
$\bar{S}_{I6} \sim 1 + poly(\overline{DED}_{I6})[,1] + poly(\overline{DED}_{I6})[,2] + poly(AGE_6)[,1] + SEX + BP + SEX : BP + W_{A6} + N_{I6}$
$\bar{S}_{I7} \sim 1 + poly(\overline{DED}_{I7})[,1] + poly(\overline{DED}_{I7})[,2] + poly(AGE_7)[,1] + SEX + PHYS.INH + URBAN + BP + SEX : BP$
$\bar{S}_{D5} \sim 1 + \log(\overline{DED}_{D5}) + poly(AGE_5)[,1] + poly(AGE_5)[,2] + SEX + PHYS.INH + URBAN + BP + W_{A5} + N_{D5}$
$\bar{S}_{D6} \sim 1 + \log(\overline{DED}_{D6}) + poly(AGE_6)[,1] + poly(AGE_6)[,2] + BP + W_{A6} + N_{D6}$
$\bar{S}_{D7} \sim 1 + \log(\overline{DED}_{D7}) + poly(AGE_7)[,1] + poly(AGE_7)[,2] + PHYS.INH + URBAN + BP + URBAN : BP + W_{A7} + W_{I7} + N_{D7}$

Table 6: Reduced model equations for each of the nine average claim size models after applying sequential backward selection based on the Wald test

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept	3.4401	0.0093	369.23	$< 2 \cdot 10^{-16}$
$\overline{DED}_{A7}$	0.1445	0.0057	25.51	$< 2 \cdot 10^{-16}$
$poly(AGE_7)[,1]$	19.5573	0.6597	29.65	$< 2 \cdot 10^{-16}$
$poly(AGE_7)[,2]$	-6.4546	0.6748	-9.56	$< 2 \cdot 10^{-16}$
SEX	0.0263	0.0084	3.13	0.0018
$PHYS.INH$	-0.0057	0.0042	-1.36	0.1732
$URBAN$	0.0315	0.0038	8.26	$< 2 \cdot 10^{-16}$
BP	0.0291	0.0037	7.92	$< 2 \cdot 10^{-16}$
$\overline{DED}_{A7} : SEX$	-0.0140	0.0067	-2.09	0.0369
$PHYS.INH : URBAN$	-0.0057	0.0025	-2.34	0.0195
W_{I7}	0.0952	0.0101	9.40	$< 2 \cdot 10^{-16}$
W_{D7}	-0.0166	0.0069	-2.39	0.0167
N_{A7}	0.0066	0.0003	22.36	$< 2 \cdot 10^{-16}$

Table 7: Model summary for the reduced ZTNB regression model of the claim frequencies for treatment field ambulant in 2007, θ estimated to be 2.41.

4.4 Results of fitting copulas to the binary and continuous margins

We model the dependency between the three treatment fields ambulant, inpatient and dental. The years 2005 to 2007 will be investigated separately.

4.4.1 Binary margins

The distribution of eight combinations of zero-claims over the three fields in 2005 to 2007 is listed in Table 8. More than 40% of the insured persons in every year did not claim any reimbursement whatsoever. Recall that $\{W_j = 0\}$ refers to *not* having a zero claim. The copula arguments for (10) will be determined using predicted cdf $\hat{P}(W_{ij} \le 0|\boldsymbol{x}_{ij}^W) =$

A	I	D	2005	2006	2007
1	1	1	44.12%	41.27%	40.22%
1	1	0	2.49%	2.54%	2.55%
1	0	1	0.46%	0.39%	0.46%
0	1	1	13.73%	14.60%	13.73%
1	0	0	0.05%	0.04%	0.04%
0	1	0	30.20%	31.84%	33.40%
0	0	1	3.22%	3.35%	3.28%
0	0	0	5.74%	5.97%	6.32%

Table 8: Distribution of outcomes of $\boldsymbol{W}$ in the data for 2005 - 2007

$\frac{1}{1+\exp(\boldsymbol{x}_{ij}^{W t}\hat{\boldsymbol{\beta}}_j)}$ and $\hat{P}(W_{ij} \le 1|\boldsymbol{x}_{ij}^W) = 1$. In Table 9 the fitted copula parameters for the independence copula as well as the trivariate Gaussian, Student t, Clayton and Gumbel copulas are given. Note that we are not using a PCC for modelling the dependency between the binary margins. This would imply multiple integration of the PCC with different upper boundaries in order to obtain joint cdfs of these margins, which would then be used in (10) for calculating the joint pmf. In order to compare those fits we utilize a test proposed by Vuong (1989) and the distribution-free test (Clarke (2007)) for nonnested model comparison. Vuong defines the statistics

$$m_i := \log\left(\frac{p_{\boldsymbol{W}}^1(w_{i1}, \dots, w_{iJ}|\hat{\boldsymbol{\zeta}}^1)}{p_{\boldsymbol{W}}^2(w_{i1}, \dots, w_{iJ}|\hat{\boldsymbol{\zeta}}^2)}\right), \quad i = 1, \dots, n, \tag{12}$$

where $p_{\boldsymbol{W}}^1(\cdot)$ and $p_{\boldsymbol{W}}^2(\cdot)$ are the pmf of two different (copula) models for $\boldsymbol{W}$ and $\hat{\boldsymbol{\zeta}}^1$ and $\hat{\boldsymbol{\zeta}}^2$ the copula parameter estimates, respectively. For details we refer to Vuong (1989). Let $\mathbf{m} = (m_1, \dots, m_n)^t$ and $E(\mathbf{m}) := \boldsymbol{\mu}_0^m = (\mu_1^m, \dots, \mu_n^m)$. Model 1 (2) is closer to the true model if $\boldsymbol{\mu}_0^m > 0$ (< 0) and falls together with the true model if $\boldsymbol{\mu}_0^m > 0$. Hence, we can test the null hypothesis $H_0 : \boldsymbol{\mu}_0^m = \mathbf{0}$ against $H_1 : \boldsymbol{\mu}_0^m \neq \mathbf{0}$. Using the central limit theorem Vuong (1989) shows that under H_0

$$\nu := \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n m_i\right]}{\sqrt{\frac{1}{n}\sum_{i=1}^n (m_i - \bar{m})^2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad \text{as } n \to \infty$$

where $\bar{m} := \frac{1}{n}\sum_{i=1}^n m_i$. Then an asymptotic α-level test rejects H_0 if and only if $|\nu| \ge z_{1-\frac{\alpha}{2}}$, where $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution. The test chooses model 1 over 2, if $\nu \ge z_{1-\frac{\alpha}{2}}$. Similarly, model 2 is chosen if $\nu \le -z_{1-\frac{\alpha}{2}}$. No model is preferred for $-z_{1-\frac{\alpha}{2}} < \nu < z_{1-\frac{\alpha}{2}}$. Clarke (2007) proposes a distribution-free test based on a modified paired sign test to the differences in the individual log-likelihoods. Clarke considers as test statistic the number of positive differences, i.e. $B := \sum_{i=1}^{n} \mathbb{1}_{\{0,+\infty\}}(m_i)$. Let M_i a rv with value m_i, then the null hypothesis of the distribution-free test is

$$H_0^{DF} : P\left[M_i > 0\right] = 0.5 \quad \forall i = 1, \ldots, n.$$

Hence under the null hypothesis B is Binomial distributed with parameters n and probability 0.5. For the test problem H_0^{DF} versus $H_{1+}^{DF} : P\left[M_i > 0\right] > 0.5,\ i = 1, \ldots, n$, the corresponding α - level upper tail test rejects H_0^{DF} versus H_{1+}^{DF} if and only if $B \geq c_{\alpha+}$, where $c_{\alpha+}$ is the smallest integer such that $\sum_{c=c_{\alpha+}}^{n} \binom{n}{c} 0.5^n \leq \alpha$. If the upper tail test rejects H_0^{DF} we decide that model 1 is preferred over model 2. For the alternative $H_{1-}^{DF} : P\left[M_i > 0\right] < 0.5,\ i = 1, \ldots, n$, the α - level lower tail test rejects H_0^{DF} versus H_{1-}^{DF} if and only if $B \leq c_{\alpha-}$, where $c_{\alpha-}$ is the largest integer such that $\sum_{c=0}^{c_{\alpha-}} \binom{n}{c} 0.5^n \leq \alpha$. If H_0^{DF} versus H_{1-}^{DF} is rejected, then model 2 is preferred over model 1. If H_0^{DF} cannot be rejected, no model is preferred. The test decisions applied to our data are given in Table 10. Note

		Year	MLE
Gaussian	$(\hat{\tau}_{AI}^{W}, \hat{\tau}_{AD}^{W},$	2005	$(0.373, 0.886, 0.420)'$
	$\hat{\tau}_{ID}^{W})'$	2006	$(\mathbf{0.319}, \mathbf{0.816}, \mathbf{0.384})'$
		2007	$(\mathbf{0.382}, \mathbf{0.886}, \mathbf{0.410})'$
Student t	$(\hat{\psi}_{AI}^{W}, \hat{\psi}_{AD}^{W},$	2005	$(\mathbf{0.329}, \mathbf{0.908}, \mathbf{0.366}, \mathbf{19.86})'$
	$\hat{\psi}_{ID}^{W}, \hat{\nu}^{W})'$	2006	$(0.408, 0.759, 0.382, 18.73)'$
		2007	$(0.405, 0.771, 0.387, 18.84)'$
Clayton	$\hat{\theta}^{W}$	2005	0.642
		2006	0.623
		2007	0.640
Gumbel	$\hat{\lambda}^{W}$	2005	1.917
		2006	1.837
		2007	1.838

Table 9: Fitted copula parameters for different trivariate copula families with binary margins in 2005 - 2007. The preferred models according to Vuong and Clarke tests (see Table 10) are highlighted in boldtype.

that we also apply the Schwarz correction described in these papers for the number of parameters. In each cell the decisions toward model (I) labelled row-wise or (II) labelled column-wise are given. The decision of the Vuong test together with its p value is given in the first row of each cell. The decision of the Clarke test with the p value in brackets are given in the second row. We see that the independence copula is not preferred over any other copula for both the Vuong and the Clarke test in any year. Also the Clayton and Gumbel are not preferred over the Gaussian and Student t copula fit. Between these two classes the Student t copula is preferred according to the Clarke test in 2005, whereas the Vuong test decision is less significant. For 2006 and 2007 the Clarke test chooses the Gaussian model with very low p-value. For all three years we see in Table 9 strong correlation between the binary margins. It is driven not only by the health status of the insured person but also by the incentive the insurer sets: if no bill is refunded in any of the three fields throughout the year, the policyholder will receive a premium refund. The more policyholders can "optimize" their medical treatment patterns, the higher the correlation between these fields will be. This explains the high correlation between ambulant and dental treatments. Since the policyholders' influence on whether or not they have to go to a hospital (inpatient treatments) will be very low, the correlations between the ambulant/ dental field and the inpatient field are relatively low.

(I) \ (II)	2005			
	Gaussian	**Student t**	Clayton	Gumbel
Indep.	V: (II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$
	C: (II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$
Gaussian		(II) 0.0004	(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
		(II) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
Student t			(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
			(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
Clayton				(II) $< 2 \cdot 10^{-16}$
				(II) $< 2 \cdot 10^{-16}$
(I) \ (II)	2006			
	Gaussian	Student t	Clayton	Gumbel
Indep.	V: (II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$
	C: (II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$
Gaussian		(II) 0.2063	(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
		(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
Student t			(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
			(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
Clayton				(II) $< 2 \cdot 10^{-16}$
				(II) $< 2 \cdot 10^{-16}$
(I) \ (II)	2007			
	Gaussian	Student t	Clayton	Gumbel
Indep.	V: (II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$
	C: (II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$	(II) $< 2 \cdot 10^{-16}$
Gaussian		(II) 0.0001	(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
		(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
Student t			(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
			(I) $< 2 \cdot 10^{-16}$	(I) $< 2 \cdot 10^{-16}$
Clayton				(I) $< 2 \cdot 10^{-16}$
				(I) $< 2 \cdot 10^{-16}$

Table 10: Preferred model according to the tests proposed by Vuong ("V", first row of each cell) and Clarke ("C", second row) followed by p-values for different copula choices modeling the dependence structure of the binary margins $\boldsymbol{W}$. The preferred models are highlighted in boldtype.

4.4.2 Continuous margins

The arguments of the PCC model given in (7) will be estimated using Lemma 3.2, i.e. for $i \in \mathcal{I}_j$

$$\hat{F}_{T_{ij}^+}(t_{ij}^+|\boldsymbol{x}_{ij}^N, \boldsymbol{x}_{ij}^{\bar{S}}, \boldsymbol{w}_{i(-j)}, n_{ij}) := \sum_{k=1}^{\infty} \hat{F}_{\bar{S}} \left(\frac{t_{ij}^+}{k} | \boldsymbol{x}_{ij}^N, \boldsymbol{x}_{ij}^{\bar{S}}, \boldsymbol{w}_{i(-j)}, n_{ij} \right) \hat{p}_N(k). \qquad (13)$$

Two additional choices have to be made in order to fully specify the PCC. First one needs to determine which pairs of margins will be modelled by the unconditional copulas c_{12} and c_{13} and which by $c_{23|1}$, i.e. the problem of choosing a good permutation of the margins. Further one needs to pick appropriate copula families to describe the dependency structure between pairs of margins. The first problem may be addressed for example by performing a simple a priori fit. Thereby we fit three arbitrary but identical bivariate Gaussian copulas on the subset of the data, where all observations with at least one zero claim in either of the two margins have been taken out. The two pairs of margins with the strongest fitted correlation parameter will be modelled by c_{12} and c_{13}. For the data at hand, there is no permutation necessary for any of the three years, i.e. we choose treatment fields A, I and D to be the margins $1, 2$ and 3, respectively, hence $c_{12} = c_{AI}$, $c_{13} = c_{AD}$ and $c_{23|1} = c_{ID|A}$. The second problem may be addressed by looking at scatterplots for the same reduced data subsets. Since it is hard to detect typical copula structures from scatterplots based on marginally transformed uniform margins $u_{ij} := \hat{F}_{T_{ij}^+}(t_{ij}^+|\boldsymbol{x}_{ij}^N, \boldsymbol{x}_{ij}^{\bar{S}}, \boldsymbol{w}_{i(-j)}, n_{ij})$, $j = 1, 2, 3$, $i \in \mathcal{I}_j$, we consider scatterplots of $z_{ij} := \Phi^{-1}(u_{ij}))$, where $\Phi^{-1}(\cdot)$ is the quantile of the standard normal distribution. We will compare these plots to contour plots of the corresponding theoretical copulas with standard normal margins at the maximum likelihood estimate of the empirical data. In Figure 1 scatterplots of Z_{j_1} and Z_{j_2} are plotted for the pairs of margins AI and AD in 2005 to 2007. Additionally kernel density estimates are added to these scatterplots. The theoretical contour plots for an appropriate copula choice are plotted to the right of each scatterplot. The copula parameters are the MLE obtained from the corresponding data conditional on $\{W_{j_1} = 0\}$ and $\{W_{j_2} = 0\}$. Based on these copula choices the conditional arguments of $c_{23|1}$ can be calculated. For example, for 2005 we have to determine $u_{iI|A5} := h^C(u_{iI5}|u_{iA5}, \hat{\theta} = 0.33)$ and $u_{iD|A5} := h^{Ga}(u_{iD5}|u_{iA5}, \hat{\rho} = 0.10)$, $i \in \mathcal{I}_A \cap \mathcal{I}_I \cap \mathcal{I}_D$, where h^C and h^{Ga} are the h functions w.r.t. to the Clayton and the Gaussian copula, respectively (see Appendix A), and 0.33 and 0.10 are the MLE of these copulas determined in the previous step. We will plot $z_{iI|A5} := \Phi^{-1}(u_{iI|A5})$ and $z_{iD|A5} := \Phi^{-1}(u_{iD|A5})$ in Figure 2 and proceed similarly as before in order to choose appropriate copulas. The maximum likelihood estimates when jointly estimating the copula parameters for the PCC's are given in Table 11. Since in 2006 the parameter of $c_{23|1}$ of the Gumbel copula is close to 1 and the parameter of the Gaussian copula for $c_{23|1}$ in 2007 is close to 0, we replace these copulas by the independence copulas. The optimal model choices are typed bold. For these copulas there is a one-to-one relationship to Kendall's τ, i.e. we can determine theoretical Kendall's τ corresponding to the ML copula parameters and compare them to the empirical Kendall's τ. For the Gaussian and the Student t copulas we transform $\tau := 2/\pi \cdot \sin^{-1}(\rho)$, for the Clayton we need to calculate $\tau := \theta/(2 + \theta)$ and for the Gumbel we have $\tau := 1 - 1/\lambda$ (see for instance Frees & Valdez

Copulas	Year	$\hat{\zeta}^{T+}_{AI}$	$\hat{\zeta}^{T+}_{AD}$	$\hat{\zeta}^{T+}_{ID\|A}$
C / C / Ga	2005	**0.333**	**0.096**	**−0.041**
C / C / Gu	2006	0.346	0.176	1.010
C / C / Ind		**0.345**	**0.176**	
t / C / Ga	2007	0.272, *df* = 20.9	0.144	−0.005
t / C / Ind		**0.272, df = 20.9**	**0.144**	
corresponding Kendall's τ				
theor.	2005	0.143	0.061	−0.026
empir.		*0.171*	*0.058*	*−0.027*
theor.	2006	0.147	0.081	0.010
theor.		0.147	0.081	
empir.		*0.178*	*0.082*	*0.027*
theor.	2007	0.175	0.067	−0.003
theor.		0.175	0.067	
empir.		*0.173*	*0.066*	*−0.013*

Table 11: Maximum likelihood estimates of the copula parameters for the Gaussian (Ga), Student t (t), Clayton (C) and Gumbel copula (Gu). Corresponding theoretical Kendall's τ and empirical Kendall's τ of copula data. Updated fit using the independence copula for $ID|A$ in 2006 and 2007.

(1998, Appendix B)). The empirical Kendall's τ is based on the uniformely transformed margins. The results concerning Kendall's τ are given in the lower panel of Table 11: the theoretical and empirical Kendall's τ are quite close which confirms the results of our fitting approach. There is a positive correlation between ambulant and inpatient as well as for ambulant and dental treatments for all three years, which is driven by the health status of the insured person. Given ambulant treatments, the correlation between inpatient and dental treatments is close to zero and is set to zero for 2006 and 2007.

4.5 Model interpretation

For the year 2007 we aim to investigate the influence of AGE on the predicted probability of a refund $\hat{P}_W(1,1,1|\boldsymbol{x}_j^W)$. Thereby we fix all other covariates, i.e. we fix the applied deductible DED_{A7} at its median value 34.85, whereas DED_{I7} and DED_{D7} will be fixed at 0. The buying power will be fixed at its mode 19499.40 and the urbanity at its median 396.35. The number of physicians per inhabitants we set to its mode 0.223. Modes are estimated using kernel density estimates of histograms of the covariates. For men and women, the influence of AGE on $\hat{P}_W(1,1,1|\boldsymbol{x}_j^W)$ both under the joint model and assuming independence are graphed in the left panel of Figure 3.

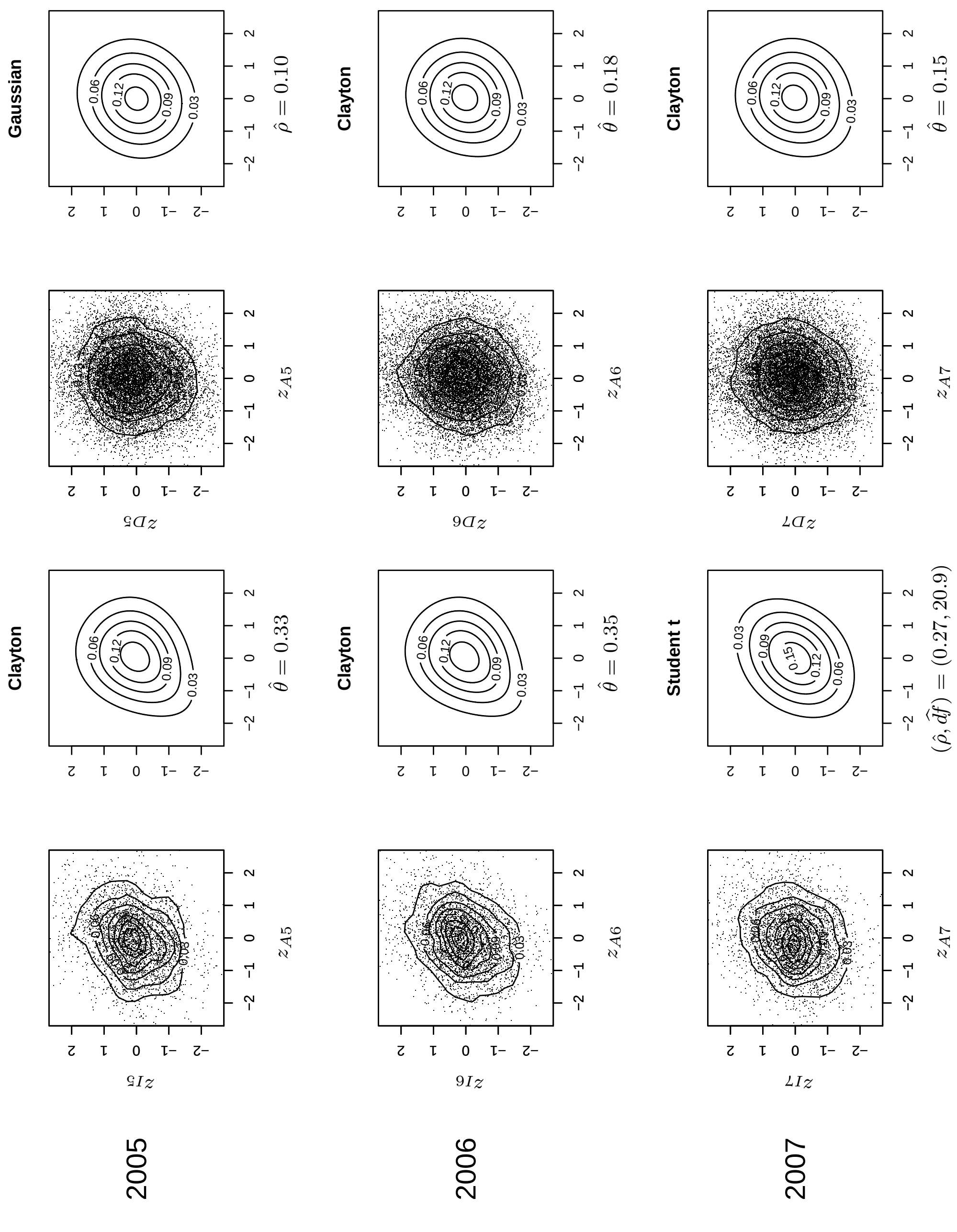


Figure 1: Scatterplots of pairs of $z_{ij} := \Phi^{-1}\left(\hat{F}_{T_{ij}^{+}}(t_{ij}^{+}|\boldsymbol{x}_{ij}^{N}, \boldsymbol{x}_{ij}^{\bar{S}}, \boldsymbol{w}_{i(-j)}, n_{ij})\right)$, $j = 1, 2, 3$ with contour plots of bivariate kernel density estimates for ambulant / inpatient margins (first column) and for ambulant / dental margins (third column). In column two (four) we show theoretical contour plots based on a chosen pair copula family for ambulant / hospital (ambulant / dental) margins.

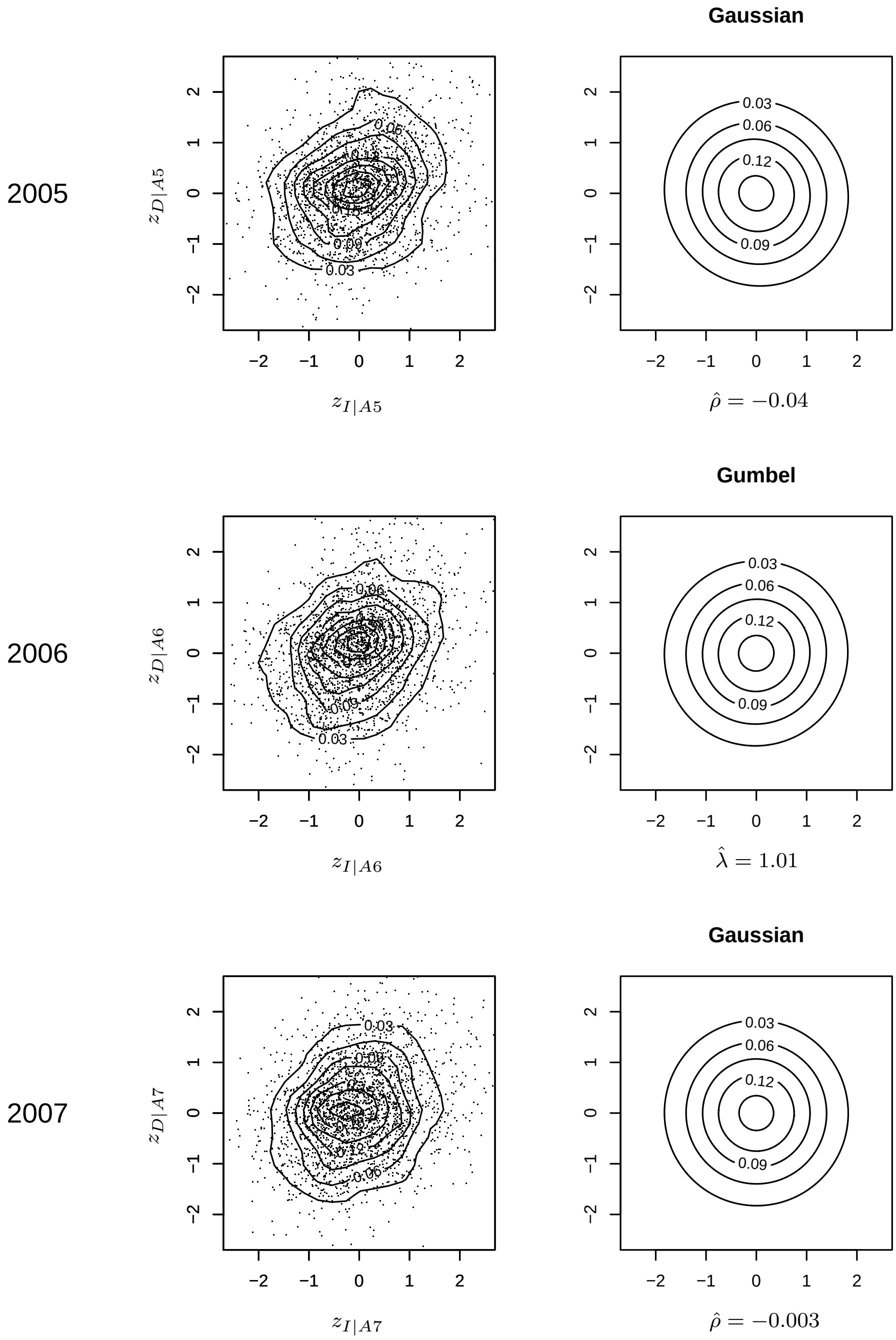
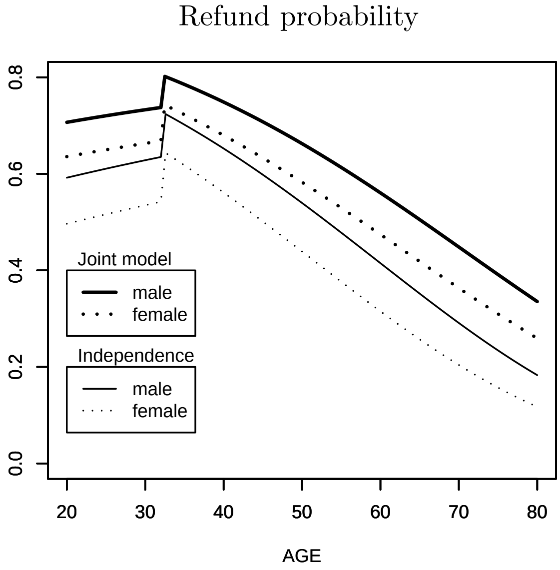


Figure 2: Scatterplots of conditional pairs of $z_{ij_1|j_2} := \Phi^{-1}\left(h(u_{ij_1}|u_{ij_2})\right)$, $j_k = 1, 2, 3$ with contour plots of bivariate kernel density estimates for inpatient / dental margins given the ambulant margin (first column). In column two we show theoretical contour plots based on a chosen pair copula family for each year.

Male insured persons have a higher refund probability in general. Since AGE was taken into our models as a piecewise linear function there is a jump at 32. Whereas earlier than 32 the refund probability slightly increases, it rapidly falls when the person gets



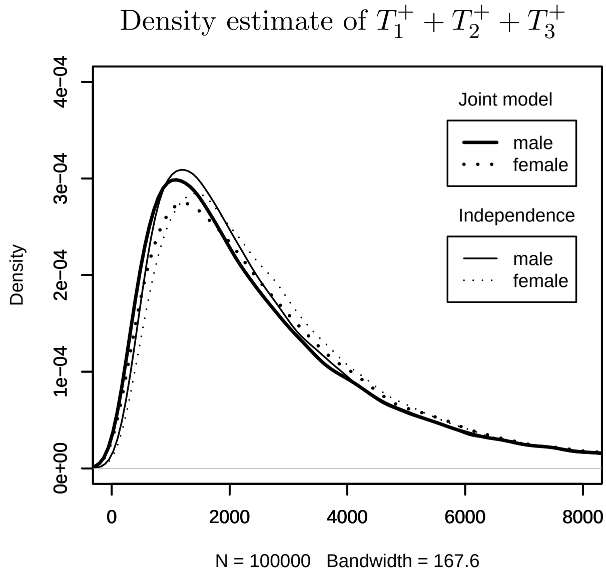


Figure 3: Influence of AGE on the refund probability when assuming independence and using the joint fitted probability $p_{\boldsymbol{W}}(\boldsymbol{w})$ while fixing other covariates; density estimates of sums of claims.

older, hence it becomes increasingly difficult to get the premium refund. In a second step we are interested in estimating the density of $T_1^+ + T_2^+ + T_3^+$, therefore we additionally fix AGE at its mode of 40.79. Further we assume we have a rather sick person and set $\boldsymbol{W} := (0,0,0)'$, i.e. we assume a claim occured in each treatment field. The arguments of the PCC will be predictive cdfs of T_j^+, $j = 1,2,3$ determined according to (13). We approximate quantile functions for T_j^+ using the R function "approxfun" in package *stats*. Then we proceed by sampling $(t_{r1}^+, t_{r2}^+, t_{r3}^+)'$, $r = 1, \ldots, 100\ 000$ from $\boldsymbol{T}$. Sampling from a C-vine is straightforward, we refer to Aas et al. (2009) for details. Finally we compute $t_r^+ := t_{r1}^+ + t_{r2}^+ + t_{r3}^+$ and plot its density estimate using the *stats* function "density". On the right panel of Figure 3 we see that the highest predicted density of $T_1^+ + T_2^+ + T_3^+$ when using the joint model for males lies around 1600 Euro (1750 Euro for females). Under the independence assumption the peaks of the estimated densities are even higher, therefore the joint model also reflects diversification effects.

Summary and Discussion

For the first time, a multivariate analysis of claims including zero claims based on PCC's was carried out. We have fitted separately a joint distribution for total claims given zero claim events and for the zero claims. The total claims given zero claim events can be expressed as a PCC under margins. Whatever combination of zero claims occurs one gains knowledge in terms of a likelihood contribution either on the correlation of the total positive claims or on the correlation of the zero claims. Even if the percentage of positive claims in one or more margins is very low, our approach yet allows to fit these data. In higher dimensional problems, however, the computational effort of numerically integrating margins out, increases. Other approaches for approximating high-dimensional integrals may be more efficient for the problem at hand and may decrease the computational time. Such approaches might also allow to efficiently approximate the joint cdf of the PCC and hence to model the dependency of the binary margins also based on PCC's. The choice of the bivariate copula families of such a PCC with binary margins is still an open question. In an application to health insurance we saw that the zero claim events between ambulant and dental treatments show a large positive correlation. There is a positive correlation also for the positive claims fitted by the pair-copula construction. The correlation is driven by the health status of the insured person. Given ambulant treatments, the correlation between inpatient and dental treatments is very low and needs not be fitted by a copula for 2006 and 2007, i.e. we may assume independence between the conditional margins.

Acknowledgement:

C. Czado is supported by DFG (German Science Foundation) grant CZ 86/1-3. V. Erhardt is supported by a stipend.

References

K. Aas, et al. (2009). 'Pair-copula constructions of multiple dependence'. *Insurance: Math. and Econom.* **44**(2):182–198.

T. Bedford & R. M. Cooke (2001a). *Monte Carlo simulation of vine dependent random variables for applications in uncertainty analysis.* 2001 Proceed. of ESREL2001, Turin, Italy.

T. Bedford & R. M. Cooke (2001b). 'Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines'. *Ann. Math. Artif. Intell.* **32**(1-4):245–268.

T. Bedford & R. M. Cooke (2002). 'Vines -a new graphical model for dependent random variables'. *Ann. Statist* **30**:1031–1068.

E. J. Belasco & S. K. Ghosh (2008). 'Modeling Censored Data Using Mixture Regression Models with an Application to Cattle Production Yields'. 2008 Annual Meeting, Orlando, Florida 6341, Agricultural and Applied Economics Association.

A. Brezger & S. Lang (2006). 'Generalized structured additive regression based on Bayesian P-splines'. *Comput. Stat. Data Anal.* **50**(4):967–991.

K. A. Clarke (2007). 'A Simple Distribution-Free Test for Nonnested Model Selection'. *Polit. Anal.* **15**(3):347–363.

M. J. L. F. Cruyff & P. G. M. van der Heijden (2008). 'Point and Interval Estimation of the Population Size Using a Zero-Truncated Negative Binomial Regression Model'. *Biom. J.* **50**(6):1035–1050.

C. Czado, et al. (2009). 'Pair-copula constructions for modeling exchange rate dependence'. *Submitted for publication* Preprint available at *http://www-m4.ma.tum.de/Papers/index.html*.

P. Deb, et al. (2006). 'Private Insurance, Selection, and Health Care Use: A Bayesian Analysis of a Roy-Type Model'. *J. Bus. & Econ. Stat.* **24**(4):403–415.

A. DeMaris (2004). *Regression with social data: modeling continuous and limited response variables.* Hoboken, N.J. : John Wiley & Sons, Inc.

C. Dutang, et al. (2008). 'actuar: An R Package for Actuarial Science'. *Journal of Statistical Software* Preprint available from: http://www.cran.r-project.org/web/packages/actuar/.

E. Frees & E. A. Valdez (1998). 'Understanding Relationships Using Copulas'. *N. Amer. Actuarial J.* **2**:1–25.

E. W. Frees, et al. (2007). 'Predicting the Frequency and Amount of Health Care Expenditures' *http://research3.bus.wisc.edu/file.php/129/Papers/AggLossExpenditures24Aug2007.pdf*.

E. W. Frees & E. A. Valdez (2008). 'Hierarchical Insurance Claims Modeling'. *J. Amer. Statistical Assoc.* **103**(484):1457–1469.

G. Z. Heller, et al. (2007). 'Mean and dispersion modelling for policy claims costs'. *Scand. Actuarial J.* **2007**(4):281–292.

H. Joe (1996). *Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters.* In L. Rüschendorf and B. Schweizer and M. D. Taylor (Ed.), Distributions with Fixed Marginals and Related Topics.

H. Joe (1997). *Multivariate models and dependence concepts.* Monographs on Statistics and Applied Probability. 73. London: Chapman and Hall. xviii, 399 p. .

B. Jørgensen & M. C. P. de Souza (1994). 'Fitting Tweedie's compound Poisson model to insurance claims data'. *Scand. Actuarial J.* pp. 69–93.

R. Kastenmeier (2008). 'Joint Regression Analysis of Insurance Claims and Claim Sizes'. Master's thesis, Technische Universität München (*www-m4.ma.tum.de/Diplarb/*).

D. Kurowicka & R. Cooke (2006). *Uncertainty analysis with high dimensional dependence modelling.* Wiley series in probability and statistics, Chichester, England: Wiley.

R. B. Nelsen (2006). *An introduction to copulas. 2nd ed.* Springer Series in Statistics. New York, NY: Springer. xiii, 269 p.

M. Pitt, et al. (2006). 'Efficient Bayesian inference for Gaussian copula regression models'. *Biometrika* **93**(3):537–554.

A. Sklar (1959). 'Fonctions de répartition à n dimensions et leurs marges'. *Publications de l'Institut de Statistique de L'Université de Paris* **8**:229–231.

G. K. Smyth & B. Jørgensen (2002). 'Fitting tweedie's compound Poisson model to insurance claims data: dispersion modelling'. *ASTIN Bull.* **32**(1):143–157.

P. X.-K. Song (2007). *Correlated Data Analysis: Modeling, Analytics and Applications*, vol. 1. Springer-Verlag, New York.

J. Sun, et al. (2008). 'Heavy-tailed longitudinal data modeling using copulas'. *Insurance: Math. and Econom.* **42**(2):817–830.

J. Tobin (1958). 'Estimation of Relationships for Limited Dependent Variables'. *Econometrica* **26**(1):24–36.

M. C. K. Tweedie (1984). 'An index which distinguishes between some important exponential families'. In J. K. Ghosh & J. Roy (eds.), *Statistics: Applications and New Directions*. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Calcutta: Indian Statistical Institute.

G. G. Venter (2001). 'Tails of copulas'. In *Proceed. ASTIN Washington, USA*, pp. 68–113.

Q. H. Vuong (1989). 'Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses'. *Econometrica* **57**(2):307–333.

D. M. Zimmer & P. K. Trivedi (2006). 'Using Trivariate Copulas to Model Sample Selection and Treatment Effects: Application to Family Health Care Demand'. *J. Bus. Econ. Statist.* **24**:63–76.

A. F. Zuur, et al. (2009). *Mixed effects models and extensions in ecology with R (in: Stat. Biol. Health)*. Springer New York.

Appendix

A Definition of selected copulas

Definition A.1 (Gaussian copula)**.** *The J-dimensional Gaussian copula with association matrix $\mathbf{\Sigma} = (\tau_{ij})_{i,j=1,\ldots,J}$ is given by*

$$C_J^G(u_1, \ldots, u_J|\mathbf{\Sigma}) := \Phi_J\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_J)|\mathbf{\Sigma}\right), \tag{14}$$

where $\Phi_J(\cdot|\mathbf{\Sigma})$ is the cdf of the J-dimensional normal distribution with mean $\boldsymbol{\mu} = \mathbf{0}_J$ and covariance $\mathbf{\Sigma}$, $\phi_J(\cdot|\mathbf{\Sigma})$ its density and $\Phi^{-1}(\cdot)$ is the quantile of the standard normal distribution.

In the special case of $J = 2$ we use notation $C_{12}^G(u_1, u_2|\tau_{12}) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)|\tau_{12})$ instead of (14).

Definition A.2 (Student t copula). *The J-dimensional t copula with parameters ν and $\mathbf{\Psi} = (\psi_{ij})_{i,j=1,\ldots,J}$ is given by*

$$\begin{aligned} C_J^t(u_1,\ldots,u_J|\nu,\mathbf{\Psi}) &:= F_J(t_\nu^{-1}(u_1),\ldots t_\nu^{-1}(u_J)|\nu,\mathbf{\Psi}) \\ &= \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_J)} \frac{\Gamma(\frac{\nu+J}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^J|\mathbf{\Psi}|}} \left(1+\frac{\boldsymbol{x}'\mathbf{\Psi}^{-1}\boldsymbol{x}}{\nu}\right)^{-\frac{\nu+J}{2}} d\boldsymbol{x}, \end{aligned} \tag{15}$$

where $F_J(\cdot|\nu,\mathbf{\Psi})$ is the joint cdf of a t distributed random vector with mean $\mathbf{0}$, covariance $\mathbf{\Psi}$ and ν degrees of freedom, $f_J(\cdot|\nu,\mathbf{\Psi})$ its density, and t_ν^{-1} denotes the quantile function of a standard univariate t_ν distribution.

For $J = 2$ we write $C_{12}^t(u_1,u_2|\nu,\psi_{12})$ instead of (15).

Definition A.3 (Archimedean copula). *Archimedean copulas are defined as*

$$C_J(u_1,\ldots,u_J|\theta) = \varphi^{-1}\big(\sum_{j=1}^{J}\varphi(u_j)\big), \tag{16}$$

where function φ is called generator. Further $\varphi : [0,1] \to [0,\infty)$ is a continuous, strictly monotonic decreasing convex function with $\varphi(1) = 0$.

We consider in particula the Clayton and the Gumbel copula. The generator for the J-dimensional Clayton copula with parameter $\theta > 0$ is $\varphi^C(u) := \frac{1}{\theta}(u^{-\theta}-1)$. For the J-dimensional Gumbel copula with parameter $\lambda \geq 1$ is $\varphi^{Gu}(u) := (-\log(u))^\lambda$. The bivariate copula densities (for the Clayton and Gumbel see Venter (2001)) together with h functions defined in 4 (see Aas et al. (2009)) are given in Table 12.

	Bivariate copula density	$h(u_1\|u_2)$
Gaussian	$\phi_2\left(\Phi^{-1}(u_1),\Phi^{-1}(u_2)\|\tau_{12}\right)\cdot\prod_{j=1}^{2}\frac{1}{\phi(\Phi^{-1}(u_j))}$	$\Phi\left(\frac{\Phi^{-1}(u_1)-\tau_{12}\Phi^{-1}(u_2)}{\sqrt{1-\tau_{12}}^2}\right)$
Student t	$f_2(t_\nu^{-1}(u_1),t_\nu^{-1}(u_2)\|\nu,\psi_{12})\cdot\prod_{j=1}^{2}\frac{1}{f_\nu(t_\nu^{-1}(u_j))}$	$t_{\nu+1}\left(\frac{t_\nu^{-1}(u_1)-\psi_{12}t_\nu^{-1}(u_2)}{\sqrt{\frac{\nu+\left(t_\nu^{-1}(u_2)\right)^2(1-(\psi_{12})^2)}{\nu+1}}}\right)$
Clayton	$(1+\theta)(u_1u_2)^{-1-\theta}\left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1/\theta-2}$	$u_2^{-\theta-1}\left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1-1/\theta}$
Gumbel	$C_{12}(u_1,u_2)\,(u_1u_2)^{-1}\,((-\log u_1)^\lambda + (-\log u_2)^\lambda)^{-2+2/\lambda}(\log u_1\log u_2)^{\lambda-1}\left[1+(\lambda-1)((-\log u_1)^\lambda+(-\log u_2)^\lambda)^{-1/\lambda}\right]$, where $C_{12}(u_1,u_2) = \exp\left(-\left[(-\log u_1)^\lambda + (-\log u_2)^\lambda\right]^{1/\lambda}\right)$	$C_{12}(u_1,u_2)\frac{1}{u_2}(-\log u_2)^{\lambda-1}\left[(-\log u_1)^\lambda+(-\log u_2)^\lambda\right]^{1/\lambda-1}$

Table 12: Bivariate copula densities and h functions for selected copulas

B Proofs of Lemmas and Propositions

Proof. (Lemma 3.2)

$$\begin{aligned}
F_{T^+|\{W=0\}}(t^+) &= P(N\bar{S} \leq t^+|\{W=0\}) \\
&= \sum_{k=1}^{\infty} P(N\bar{S} \leq t^+|\{N=k, W=0\})P(N=k|\{W=0\}) \\
&= \sum_{k=1}^{\infty} P(\bar{S} \leq \frac{t^+}{k}|\{N=k, W=0\})P(N=k|\{W=0\}) \\
&= \sum_{k=1}^{\infty} F_{\bar{S}|\{N,W=0\}}\left(\frac{t^+}{k}|\{N=k\}\right) p_{N|\{W=0\}}(k).
\end{aligned}$$

□

Proof. (Lemma 3.3) For $t \geq 0$

$$\begin{aligned}
F_T(t) &= P(T \leq t) \\
&= P(T \leq t|\{W=1\}) \cdot P(W=1) + P(T \leq t|\{W=0\}) \cdot P(W=0) \\
&= P((1-W)T^+ \leq t|\{W=1\}) \cdot P(W=1) \\
&\quad + P(W=0) \cdot P((1-W)T^+ \leq t|\{W=0\}) \\
&= P(0 \leq t|\{W=1\}) \cdot P(W=1) + P(W=0) \cdot P(T^+ \leq t|\{W=0\}) \\
&= P(W=1) + \mathbb{1}_{\{t>0\}} P(W=0) \cdot F_{T^+|\{W=0\}}(t) \\
&= p_W(0) + \mathbb{1}_{\{t>0\}}(1-p_W(0)) \cdot F_{T^+|\{W=0\}}(t).
\end{aligned}$$

□