Concomitant Information in a Bivariate Model of Claim Frequencies and Severities

Gabriel Escarela[1] Jacques F. Carrière[2]

August 8, 2005

Suggested Running Head: **A Bivariate Model of Claim Frequencies and Severities**

Address for correspondence:
Dr. Gabriel Escarela
Departamento de Matematicas
Universidad Autonoma Metropolitana
Unidad Iztapalapa, AT-223
Av. San Rafael Atlixco No. 186
Col. Vicentina
C.P. 09340, Mexico D.F.
Mexico

Email: ge@xanum.uam.mx

[1]Departamento de Matemáticas, Universidad Autónoma Metropolitana, Unidad Iztapalapa, AT-321, Av. San Rafael Atlixco No. 186, Col. Vicentina, C.P. 09340, México D.F., Mexico.
[2]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

Concomitant Information in a Bivariate Model of Claim Frequencies and Severities

Abstract

Bivariate claim data come from a population that consists of insureds who may claim either one, both or none of the two types of benefits covered by a policy. In the present paper, we develop a statistical procedure to fit bivariate distributions of claims in presence of covariates. This allows for a more accurate study of insureds' choice and size in the frequency and severity of the two types of claims. A generalised logistic model is employed to examine the frequency probabilities, whilst the three parameter Burr distribution is suggested to model the underlying severity distributions. The bivariate copula model is exploited in such a way that it allows us to adjust for a range of frequency dependence structures; a method for assessing the adequacy of the fitted severity model is outlined. A health claims dataset illustrates the methods; we describe the use of orthogonal polynomials for characterising the relationship between age and the frequency and severity models.

Key words: Bivariate loss distribution; Frank's copula; Survival copula; Burr regression; Diagnostics.

Concomitant Information in a Bivariate Model of Claim Frequencies and Severities

1 Introduction

In this article we deal with a mixture model of claim frequencies and severity distributions for the analysis of bivariate loss data. Many studies in the statistical literature have focused on the fit of both univariate and bivariate loss distributions (e.g. Hogg & Klugman, 1984; Panjer & Willmot, 1992; Klugman *et al.*, 1998; Klugman & Parsa, 1999; Watkins, 1999). More sophisticated analyses have also used regression models for portfolio segmentation purposes (e.g. Beirlant *et al.*, 1998). Little attention, however, has been given to the case when we can attribute variability in cause-specific loss outcomes. This is of particular interest when analysing risk characteristics on specific types of claims.

Consider the analysis of an insurance portfolio where the interest centres on the estimation of the joint distribution of losses corresponding to two types of claims. In many practical situations, this loss data comes with associated or *concomitant* information on which the losses are thought to depend. In health insurance data, for example, we may measure the size of drug claims and of "other" claims paid by the insurance company to N insureds during a determined amount of time, and the corresponding joint distribution of these losses may depend on age, gender and other *auxiliary variables*.

Let X and Y denote the random variables corresponding to the amounts of two types of claims. The joint cumulative distribution function (CDF) is then

defined as

$$H(x,y) = \Pr\{X \leq x, Y \leq y\} \qquad x,y \geq 0. \tag{1}$$

Our goal is to model, fit and calibrate the CDF in Eq. (1) in presence of covariate information.

Because the insureds may not claim both types of benefits, we define the *frequency probabilities* as follows:

$$\begin{aligned} p_{00} &= \Pr\{X=0, Y=0\}, & p_{01} &= \Pr\{X=0, Y>0\}, \\ p_{10} &= \Pr\{X>0, Y=0\}, & p_{11} &= \Pr\{X>0, Y>0\}, \\ p_{0\bullet} &= \Pr\{X=0\}, \quad \text{and} & p_{\bullet 0} &= \Pr\{Y=0\}. \end{aligned}$$

Notice that $p_{0\bullet} = p_{00} + p_{01}$, $p_{\bullet 0} = p_{00} + p_{10}$ and $p_{00} + p_{01} + p_{10} + p_{11} = 1$.

We then define the corresponding *severity distributions* as follows:

$$H_{01}(y) = \Pr\{Y \leq y \mid X=0, Y>0\}, \tag{2}$$

$$H_{10}(x) = \Pr\{X \leq x \mid X>0, Y=0\}, \tag{3}$$

$$H_{11}(x,y) = \Pr\{X \leq x, Y \leq y \mid X>0, Y>0\}. \tag{4}$$

It follows that, if $x \geq 0$ and $y \geq 0$, the joint PDF in Eq. (1) can be expressed in terms of the frequency probabilities and the severity distributions as follows

$$H(x,y) = p_{00} + p_{01}H_{01}(y) + p_{10}H_{10}(x) + p_{11}H_{11}(x,y). \tag{5}$$

Thus, the estimation of $H(x,y)$ is reduced to the joint estimation of the frequency and severity models.

In this study, we assume that the distributions H_{01}, H_{10}, H_{11} are absolutely continuous with probability density functions (PDFs) h_{01}, h_{10} and h_{11} respectively. Thus, $H_{10}(x) = \int_0^x h_{10}(u)\,\mathrm{d}u$, $H_{01}(y) = \int_0^y h_{01}(v)\,\mathrm{d}v$ and $H_{11}(x,y) = \int_0^x \int_0^y h_{11}(u,v)\,\mathrm{d}u\,\mathrm{d}v$.

2 The mixture model

Let x_k and y_k be the amounts claimed by the k-th insured, $k = 1, ..., N$, and let $\mathbf{z}_k = (1, z_{1k}, z_{2k}, ..., z_{pk})$ be the corresponding vector of covariates. Define *status indicator vectors* as follows:

$$c_{00,k} = I(x_k = 0, y_k = 0) \qquad c_{01,k} = I(x_k = 0, y_k > 0),$$
$$c_{10,k} = I(x_k > 0, y_k = 0), \qquad c_{11,k} = I(x_k > 0, y_k > 0).$$

Here, $I(A)$ denotes the indicator function of the event A. Notice that $c_{00,k} + c_{01,k} + c_{10,k} + c_{11,k} = 1$.

The likelihood function corresponding to the N individuals can then be written as follows:

$$\begin{aligned} L &= \prod_{k=1}^{N} [p_{00,k}]^{c_{00,k}} [p_{01,k}\, h_{01}(y_k)]^{c_{01,k}} [p_{10,k}\, h_{10}(x_k)]^{c_{10,k}} [p_{11,k}\, h_{11}(x_k, y_k)]^{c_{11,k}} \\ &= \prod_{k=1}^{N} [p_{00,k}]^{c_{00,k}} [p_{01,k}]^{c_{01,k}} [p_{10,k}]^{c_{10,k}} [p_{11,k}]^{c_{11,k}} \times \qquad (6) \\ &\qquad \times \prod_{k=1}^{N} [h_{01}(y_k)]^{c_{01,k}} [h_{10}(x_k)]^{c_{10,k}} [h_{11}(x_k, y_k)]^{c_{11,k}} \qquad (7) \\ &= L_f \times L_s\,. \end{aligned}$$

Thus, the likelihood function can be split into the likelihood corresponding to the frequency probabilities, denoted here as L_f, and the likelihood corresponding to the severity distributions, denoted as L_s.

The presence of effects of covariates on the frequency probabilities can be modelled using a *generalised logistic model* (Cox & Snell, 1989, pp. 155–157), so that the *frequency model* has the following form:

$$p_{ij,k} = \frac{\exp(\boldsymbol{\beta}_{ij}^T \mathbf{z}_k)}{\exp(\boldsymbol{\beta}_{00}^T \mathbf{z}_k) + \exp(\boldsymbol{\beta}_{01}^T \mathbf{z}_k) + \exp(\boldsymbol{\beta}_{10}^T \mathbf{z}_k) + \exp(\boldsymbol{\beta}_{11}^T \mathbf{z}_k)} \qquad (8)$$

where $ij \in \{00, 01, 10, 11\}$ and $\boldsymbol{\beta}_{00}$, $\boldsymbol{\beta}_{01}$, $\boldsymbol{\beta}_{10}$, $\boldsymbol{\beta}_{11}$ are the corresponding vectors of parameters. To avoid redundancy, we set $\boldsymbol{\beta}_{00} = \mathbf{0}$.

Using the probability model defined by Eq. (8), the log of the likelihood in Eq. (6) is

$$\begin{aligned} l_f(\boldsymbol{\beta}) = \log L_f(\boldsymbol{\beta}) &= \sum_{k=1}^{N} c_{00,k} \log p_{00,k} + c_{01,k} \log p_{01,k} + c_{10,k} \log p_{10,k} + c_{11,k} \log p_{11,k} \\ &= \sum_{k=1}^{N} c_{01,k} \left(\boldsymbol{\beta}_{01}^T \mathbf{z}_k\right) + c_{10,k} \left(\boldsymbol{\beta}_{10}^T \mathbf{z}_k\right) + c_{11,k} \left(\boldsymbol{\beta}_{11}^T \mathbf{z}_k\right) - \log \psi_k(\boldsymbol{\beta}), \end{aligned}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_{01}^T, \boldsymbol{\beta}_{10}^T, \boldsymbol{\beta}_{11}^T)^T$, and $\psi_k(\boldsymbol{\beta}) = 1 + \exp(\boldsymbol{\beta}_{01}^T \mathbf{z}_k) + \exp(\boldsymbol{\beta}_{10}^T \mathbf{z}_k) + \exp(\boldsymbol{\beta}_{11}^T \mathbf{z}_k)$.

In most situations, however, there will be relatively very few observations of insureds who do not claim any of the benefits covered by the policy. Although the maximum likelihood estimates (MLE's) of the frequency model can be found by direct maximisation, the model as it stands will tend to be poorly conditioned whenever $(1/N) \sum_{k=0}^{N} c_{00,k} \approx 0$. The result is that the maximum likelihood estimates of the components of $\boldsymbol{\beta}$ tend to be highly correlated and have large standard errors. A more practical problem is that the probability of zero-claim is set to zero, i.e. $p_{00} = 0$. The problem is solved by removing the event 00 in the frequency model, namely

$$p_{ij,k} = \frac{\exp(\boldsymbol{\beta}_{ij}^T \mathbf{z}_k)}{\exp(\boldsymbol{\beta}_{01}^T \mathbf{z}_k) + \exp(\boldsymbol{\beta}_{10}^T \mathbf{z}_k) + \exp(\boldsymbol{\beta}_{11}^T \mathbf{z}_k)} \tag{9}$$

for $ij \in \{01, 10, 11\}$, where $\boldsymbol{\beta}_{11} = \mathbf{0}$. The log-likelihood for this constrained model is:

$$l_f(\boldsymbol{\beta}) = \sum_{k=1}^{N'} c_{01,k} \left(\boldsymbol{\beta}_{01}^T \mathbf{z}_k\right) + c_{10,k} \left(\boldsymbol{\beta}_{10}^T \mathbf{z}_k\right) - \log \psi_k(\boldsymbol{\beta}), \tag{10}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_{01}^T, \boldsymbol{\beta}_{10}^T)^T$, $N' = N - \sum_{k=0}^{N} c_{00,k}$ and $\psi_k(\boldsymbol{\beta}) = 1 + \exp(\boldsymbol{\beta}_{01}^T \mathbf{z}_k) + \exp(\boldsymbol{\beta}_{10}^T \mathbf{z}_k)$.

When it comes to modelling the severity distributions, it is important to consider that the paired positive observations may possess certain degree of correlation; therefore, the form of the bivariate distribution $H_{11}(x, y)$ must take into account that each marginal loss might be related to the other. An approach to create an "appropriate" joint distribution that incorporates dependence among the variables is via a bivariate copula.

The use of the copula, C_θ, is appealing since the elimination of the marginals through the copula helps to model and understand dependence structures effectively, as the dependence has no relationship with the marginal behaviour of individual characteristics. The copula is able to capture a range of global association between X and Y, $X, Y > 0$, and does so through a dependence parameter θ. Because of its flexixibility to model bivaraite loss data, copulas have received considerable attention in recent years. Here we can refer to Frees & Valdez (1998) and Klugman & Parsa (1999) and the references therein. A good introduction to the theory of copulas can be found in Nelsen (1999) and in Benes & Stephan (1997).

The definition of joint distribution function $H_{11}(x, y)$ with marginals $F_{21}(x) = H_{11}(x, \infty)$ and $F_{12}(y) = H_{11}(\infty, y)$ is implicitly defined through the identity $H_{11}(x, y) = C_\theta [F_{21}(x), F_{12}(y)]$. There are situations, however, when it is easier to find analytical expressions for the corresponding *bivariate survival function*, which can be written as follows

$$\begin{aligned} S_{11}(x, y) &= \Pr\{X > x, Y > y \mid X > 0, Y > 0\} \\ &= C_\theta [S_{21}(x), S_{12}(y)] . \end{aligned} \tag{11}$$

Here $S_{21}(x) = 1 - F_{21}(x)$ and $S_{12}(y) = 1 - F_{12}(y)$ are the marginal survival functions.

In this study, we adopt Frank's (1979) family of two-dimensional archimedean

copulas which is defined as

$$C_\theta(u,v) = -\frac{1}{\theta}\log\left(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}\right), \qquad \theta \in \mathbf{R}-\{0\}, \tag{12}$$

where $\mathbf{R}$ denotes the ordinary real line. This copula was introduced by Frank (1979) in a context of probabilistic metric spaces and has been studied by Nelsen (1986, 1999) and Genest (1987). The use of Frank's copula is appealing since it is able to capture the full range of dependence; it includes the Fréchet upper and lower bound copulas as well as the product copula, i.e. the independent structure.

To assess the degree of dependence between the marginals in the copula model, we use Kendall's τ measure of association which is defined as the difference between the probabilities of concordance and discordance for two independent and identically distributed pairs of random variables. Genest (1987) and Nelsen (1986) showed that the Kendall's τ corresponding to a member of the Frank family of Archimedean copulas C_θ can be expressed as follows:

$$\tau_\theta = 1-\frac{4}{\theta}\left(1-\frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}\,\mathrm{d}t\right).$$

The joint survival function of the paired positive random variables given by Frank's copula in Eq. (12) can then be written as

$$S_{11}(x,y) = \begin{cases} -\dfrac{1}{\theta}\log\left(1+\dfrac{(e^{-\theta S_{21}(x)}-1)(e^{-\theta S_{12}(y)}-1)}{\left(e^{-\theta}-1\right)}\right) & : \theta \neq 0, \\ S_{21}(x)\,S_{12}(y) & : \theta = 0. \end{cases} \tag{13}$$

Differentiating the bivariate survival function in Eq. (11), when $\theta \neq 0$, with respect to x and y we find that the corresponding pdf is of the form

$$h_{11}(x,y) = h_{21}(x)h_{12}(y)\,\frac{\theta\exp\left\{-\theta\left[S_{21}(x)+S_{12}(y)-2S_{11}(x,y)\right]\right\}}{(1-e^{-\theta})}; \tag{14}$$

when $\theta = 0$, we obtain $h_{11}(x,y) = h_{21}(x)h_{12}(y)$, the independent model. Thus, the pdf in Eq. (14) is non-negative for all $x, y > 0$. Therefore, the bivariate survival

function defined in Eq. (13) is a valid model with arbitrary marginal survival functions S_{21} and S_{12}.

In order to model the underlying distributions H_{01}, H_{10}, H_{21} and H_{12} in the severity model, we consider the three-parameter Burr family of distributions whose survival distribution function is specified by $S(t) = [1+\gamma(t\lambda)^{\alpha}]^{-1/\gamma}$, for $t > 0$ and $\lambda, \gamma, \alpha > 0$, and by the following pdf:

$$h(t) = \frac{(t\lambda)^{\alpha}\alpha}{t[1+\gamma(t\lambda)^{\alpha}]^{1+1/\gamma}}. \tag{15}$$

Burr's loss model is a flexible and tractable family of distributions. It has the ability to accommodate many shapes of distributions. It contains the Pareto distribution when $\alpha = 1$ and the log-logistic distribution when $\gamma = 1$, while as $1/\gamma$ approaches infinity the Weibull distribution is obtained.

To allow for the effects of covariates, and to ensure that the parameters λ, γ and α remain positive, a log link is applied as follows:

$$\lambda = \exp\left\{\mathbf{a}^T\,\mathbf{z}\right\}, \quad \gamma = \exp\left\{\mathbf{b}^T\,\mathbf{z}\right\}, \quad \alpha = \exp\left\{\mathbf{c}^T\,\mathbf{z}\right\},$$

where $\mathbf{z}$ is the vector of covariates and $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are the corresponding vectors of parameters.

A similar model for the Burr regression of univariate loss data was proposed by Beirlant *et al.* (1998); unlike their approach, which takes into account covariate information in one of the parameters only, our set up has sufficient flexibility to vary both the tail and the mode behaviours in presence of risk factors.

The log of the likelihood corresponding to the severity model in Eq. (7) can be written as follows

$$l_s(\boldsymbol{\nu}) = l_{01}^{(s)}(\boldsymbol{\nu}_{01}) + l_{10}^{(s)}(\boldsymbol{\nu}_{10}) + l_{11}^{(s)}(\boldsymbol{\nu}_{11}), \tag{16}$$

where

$$l_{01}^{(s)}(\boldsymbol{\nu}_{01}) = \sum_{k=1}^{N} c_{01,k} \log h_{01}(y_k; \boldsymbol{\nu}_{01}), \tag{17}$$

$$l_{10}^{(s)}(\boldsymbol{\nu}_{10}) = \sum_{k=1}^{N} c_{10,k} \log h_{10}(x_k; \boldsymbol{\nu}_{10}), \tag{18}$$

$$l_{11}^{(s)}(\boldsymbol{\nu}_{11}) = \sum_{k=1}^{N} c_{11,k} \log h_{11}(x_k, y_k; \boldsymbol{\nu}_{11}), \tag{19}$$

and

$$\begin{aligned} \boldsymbol{\nu} &= (\boldsymbol{\nu}_{01}, \boldsymbol{\nu}_{10}, \boldsymbol{\nu}_{11}), \\ \boldsymbol{\nu}_{01} &= (\mathbf{a}_{01}, \mathbf{b}_{01}, \mathbf{c}_{01}), \\ \boldsymbol{\nu}_{10} &= (\mathbf{a}_{10}, \mathbf{b}_{10}, \mathbf{c}_{10}), \\ \boldsymbol{\nu}_{11} &= (\mathbf{a}_{21}, \mathbf{b}_{21}, \mathbf{c}_{21}, \mathbf{a}_{12}, \mathbf{b}_{12}, \mathbf{c}_{12}, \theta). \end{aligned} \tag{20}$$

It follows that the MLE's of $\boldsymbol{\nu}$ can be obtained by optimising the log-likelihood functions in Eqs. (17), (18) and (19) separately.

Both the frequency and severity models are not linear, so numerical techniques need to be used in order to find the MLE's of the corresponding parameters. In this study, we used the `Surv` function in the statistical package S-PLUS to create design matrices. We also made use of the function `nlminb` to minimise minus two times the log-likelihood functions in Eqs. (10), (17), (18) and (19). In this study, the covariance matrix will be calculated from the inverse of the observed information matrices of the MLE's $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\nu}}$, denoted here as $\mathbf{I}_{\hat{\beta}}$ and $\mathbf{I}_{\hat{\nu}}$ respectively. For this, we used the mathematical package MAPLE (Char *et al.*, 1991) to obtain the Hessian matrix and translated the code to S-PLUS. Thus, inferences will be based on the approximations $\boldsymbol{\beta} \sim \text{MVN}(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\hat{\beta}})$ and $\boldsymbol{\nu} \sim \text{MVN}(\hat{\boldsymbol{\nu}}, \mathbf{V}_{\hat{\nu}})$, where $\mathbf{V}_{\hat{\beta}} = \mathbf{I}_{\hat{\beta}}^{-1}$, $\mathbf{V}_{\hat{\nu}} = \mathbf{I}_{\hat{\nu}}^{-1}$ and MVN denotes the multivariate normal distribution.

Taking the theory of generalised linear models as a start point, it is desirable

to determine if all the parameters in the chosen model are needed. The required hypothesis tests can be combinations taken from the following positions:

- $H_0 : \beta_j = 0$, i.e. no effects of the j-th covariate on the frequency probabilities.
- $H_0 : a_j = b_j = c_j = 0$, i.e. no effects of the j-th covariate on loss rate.

Let $M1$ and $M2$ be two nested models. Model $M2$ has q_2 parameters and model $M1$ has q_1 parameters, where $q_2 > q_1$. Model $M1$ is nested in model $M2$ if the parameters that correspond to specific covariates considered in $M1$ are contained in $M2$. Let $\hat{\boldsymbol{\omega}}_{M1}$ and $\hat{\boldsymbol{\omega}}_{M2}$ be the maximum likelihood estimates of the parameters in models $M1$ and $M2$, respectively. The corresponding *deviance statistic* is defined as

$$d = -2\left(\log L_{M1}(\hat{\boldsymbol{\omega}}_{M1}) - \log L_{M2}(\hat{\boldsymbol{\omega}}_{M2})\right),$$

which has an asymptotic distribution $d \sim \chi^2_{df=q_2-q_1}$. This statistic is then used to test the hypothesis $H_0 : \boldsymbol{\omega} = \mathbf{0}$ such that $\boldsymbol{\omega} \in M2 - M1$, i.e. the parameters that are contained in $M2$ but not in $M1$ have no effects. The hypothesis is rejected at a significance level ρ if

$$d > \chi^2_{\rho,q_2-q_1}.$$

We turn now to model diagnostics. If T is Burr distributed as in Eq. (15), by assumption $-\log S(T) = (1/\gamma)\log[1+\gamma(T\lambda)^\alpha]$ is unit exponential; therefore, we could rank the n observed values of $u = (1/\gamma)\log[1+\gamma(t\lambda)^\alpha]$ into the order statistics $u_{(j)}$ and plot

$$\frac{n+1-j}{n+1} \qquad \text{against} \qquad \exp(-u_{(j)}),$$

for $j = 1, ..., n$, where n is the number of observations in each of the severity models. This is basically a quantile-quantile plot.

The joint severities of the paired positive claims can be assessed by considering

$$\begin{aligned} G(v,w) &= \Pr\{X > v \mid Y < w\} \\ &= \frac{S_{21}(v) - S_{11}(v,w)}{1 - S_{12}(w)}. \end{aligned}$$

We then order the values of v into order statistics $v_{(j)}$ and plot

$$\frac{n_{11}+1-j}{n_{11}+1} \qquad \text{against} \qquad \frac{1}{n_{11}} \sum_{i=1}^{n_{11}} G(v_{(j)}, w_i; \lambda_{21_i}, \gamma_{21_i}, \alpha_{21_i}, \lambda_{12_i}, \gamma_{12_i}, \alpha_{12_i}, \theta)$$

for $j = 1, ..., n_{11}$, where n_{11} is the number of paired positive observations. If the marginal distributional assumptions about the marginals in H_{11} seem reasonable in the light of the data, then assessing their joint distribution is essentially assessing the dependence structure.

3 Example

We analyse a series of $N = 19827$ health policies in a insurance portfolio. We are interested in modelling the drug claims paid in one year for the k-th policy, denoted by x_k, and the total amount on all other claims, denoted by y_k. We consider the continuous covariate `age` and the factor `GENDER` (male=0; female=1).

To visualise and assess the effects of age in both the frequency and severity models, we use an orthogonal polynomial model. There are considerable advantages to be gained using orthogonal polynomials. It brings to account one independent variate at a time, so that the successive terms of the final regression equation are orthogonal to one another. The likelihood ratio method described above still makes sense; the significance of a model of certain degree can be determined by comparing the likelihood with and without the variables associated

with the polynomial. Furthermore, using likelihood techniques, significance of non-linearity of the relationship can be assessed by comparing the log-likelihood for a model of higher degree polynomial to the log-likelihood for a model with the "classic" linear effect (see e.g. Seber, 1977). In S-PLUS, the function `poly` generates the basis matrix for representing such a family of orthogonal polynomials; the primary use of this function is in a model formula, allowing the user to directly specify a polynomial term in a regression model.

3.1 Frequency Model

We consider first the *null model* of the frequency probabilities, i.e. the model in which only the intercept terms appear in Eq. (8). The MLEs and standard errors are

$$\hat{\beta}_{01} = 3.94223\ (0.18434), \quad \hat{\beta}_{10} = 6.23657\ (0.18275), \quad \hat{\beta}_{11} = 4.57746\ (0.18351).$$

Estimates of each of the probabilities can be obtained by substituting previous estimates in Eq. (8). Estimating SE's for these probabilities is a complex problem since Eq. (8) involves several parameters and their correlation structure must be taken into account to do so. One simple way to solve this problem is by generating a sample of multivariate normal random vectors with mean equal to $\hat{\boldsymbol{\beta}}$ and covariance equal to the corresponding observed matrix; each vector is evaluated in Eq. (8) and the Monte Carlo method is applied. We used the S-PLUS function `rmvnorm` and evaluated the standard deviation of the simulated probabilities; in a similar way, 95% confidence intervals (CI's) can be obtained by using the 2.5% and the 97.5% percentiles of the simulated probabilities, or simply by evaluating the Monte Carlo SE's in the standard normal CI's.

Thus, we obtain the following estimators of the frequency probabilities, together with the corresponding 95% confidence intervals:

$$\hat{p}_{00} = 0.00151\ (0.00105, 0.00216), \qquad \hat{p}_{01} = 0.07797\ (0.07425, 0.08173),$$
$$\hat{p}_{10} = 0.77334\ (0.76738, 0.77918), \qquad \hat{p}_{11} = 0.14717\ (0.14264, 0.15233).$$

These point estimators are very similar to the frequency probabilities Carrière (1997) reported in the study of a similar dataset.

In his article, Carrière also carried out tests about the null hypothesis H_0 : $p_{00} = p_{0\bullet}\, p_{\bullet 0}$, the *independence of claim frequency*; with the implementation of a bootstrap method, he found strong evidence to reject such hypothesis. Here, under the frequency model represented in Eq. (8), it is possible to carry out frequency independence tests in presence of auxiliary information by using some well-known asymptotic results. It can be shown that such hypothesis tests can be stated as

$$H_0 :\ \mathbf{A}\boldsymbol{\beta} = \mathbf{0} \qquad \text{vs.} \qquad H_1 :\ \mathbf{A}\boldsymbol{\beta} \neq \mathbf{0}, \tag{21}$$

where $\mathbf{A} = (\mathbf{I}_{1+p}, \mathbf{I}_{1+p}, -\mathbf{I}_{1+p})$, and $\mathbf{I}_{1+p}$ represents the $(1+p) \times (1+p)$ identity matrix.

To test H_0 in Eq. (21), we use a Wald test assuming the usual regularity conditions. It follows that the test statistic

$$Z^2 = (\mathbf{A}\hat{\boldsymbol{\beta}})^T (\mathbf{A}\mathbf{V}_{\hat{\beta}}\mathbf{A}^T)^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}})$$

has asymptotically a χ^2 distribution with $p+1$ degrees of freedom under H_0. We therefore reject H_0 at a significance level ρ if $Z^2 > \chi^2_{\rho, df=p+1}$. Under the null frequency model, i.e. when $p = 0$, we found that $Z^2 = 912.38$, which corresponds to a p-value of 0.00000. Thus, there is strong evidence to reject the hypothesis of frequency independence.

Because there are very few zero paired observations, we decided to fit covariate information to the frequency model in Eq. (9) – we thus fit and test difference in deviance of nested models against a critical value of $\chi^2_{0.95,df=2} = 5.99$. Table 1 shows the results of a number of main effects models. It is possible to observe in models 1 to 4 that both variables are marginally and additively significant; it is also interesting to notice that the inclusion of the polynomial terms improve the frequency model significantly when compared with the traditional linear model 4. The inclusion of the 7th and 8th degrees in Models 10 and 11 respectively, do not show any improvement to the fit though. We therefore decided to take model 9 as the "best" regression equation since it gives the most significant reduction of deviance.

It is clear that age and gender are important predictors of frequency since they present very significant effects. The plots of the corresponding fit of the polynomial regression of age in the best frequency model are shown in Fig. 1; here, the dashed lines are 95% confidence bands. The fits show that the effect of age is nonlinear and that the plots widen at the left and right extremes where less data is available. The coefficients of the remaining parameters in our best frequency model are displayed in Table 2. We can notice that these coefficients are individually significant.

In order to assess the differences between groups of insureds, it is informative to look at the fitted frequency probabilities which can be evaluated through Eq. (9). The probabilities are displayed in Fig. 2 which allow us to compare the three events. The panels show the fitted probabilities in function of age and classified by gender. We can observe that, although `GENDER` was statistically important, the shapes do not seem to differ too much between gender groups.

Table 1: Values of deviance for some nested frequency probability models fitted to the health policies data.

	Model	$-2 \times \hat{l}_f$	Np
1	*null*	26893.83	2
2	`age`	26824.97	4
3	`GENDER`	26830.88	4
4	`age + GENDER`	26743.34	6
5	`poly(age,2) + GENDER`	26694.50	8
6	`poly(age,3) + GENDER`	26659.21	10
7	`poly(age,4) + GENDER`	26655.79	12
8	`poly(age,5) + GENDER`	26642.99	14
9	`poly(age,6) + GENDER`	26637.06	16
10	`poly(age,7) + GENDER`	26635.77	18
11	`poly(age,8) + GENDER`	26633.90	20

Np=Number of parameters

Table 2: Estimates of the coefficients and standard errors of the covariates for the frequency model.

	$\boldsymbol{\beta}_{01}$		$\boldsymbol{\beta}_{10}$	
Parameter	Est.	S.E.	Est.	S.E.
`Intercept`	-0.985	(0.053)	1.537	(0.031)
`GENDER`	0.573	(0.068)	0.245	(0.042)

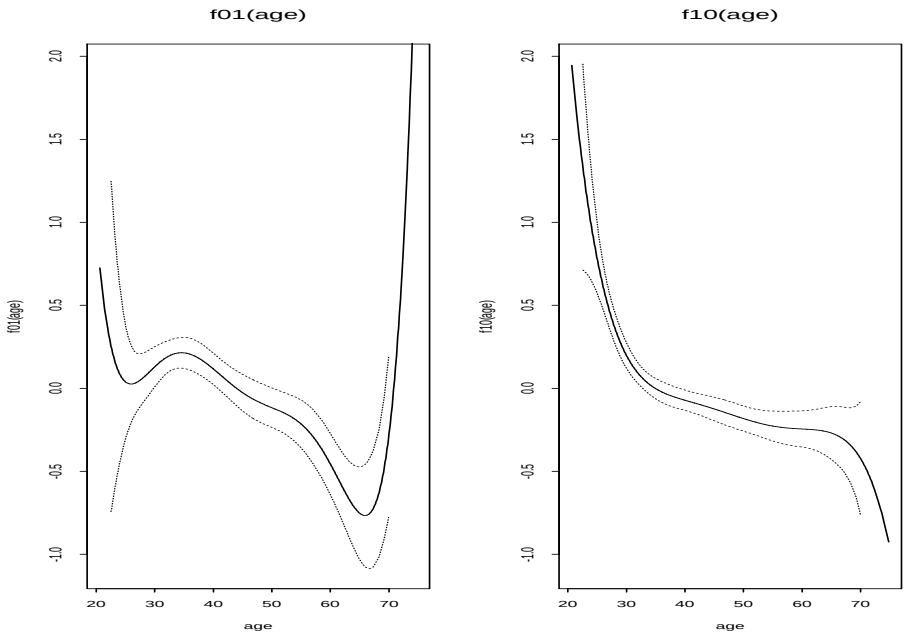

Figure 1: Polynomial fit of age for the model `poly(age,6) + GENDER`.

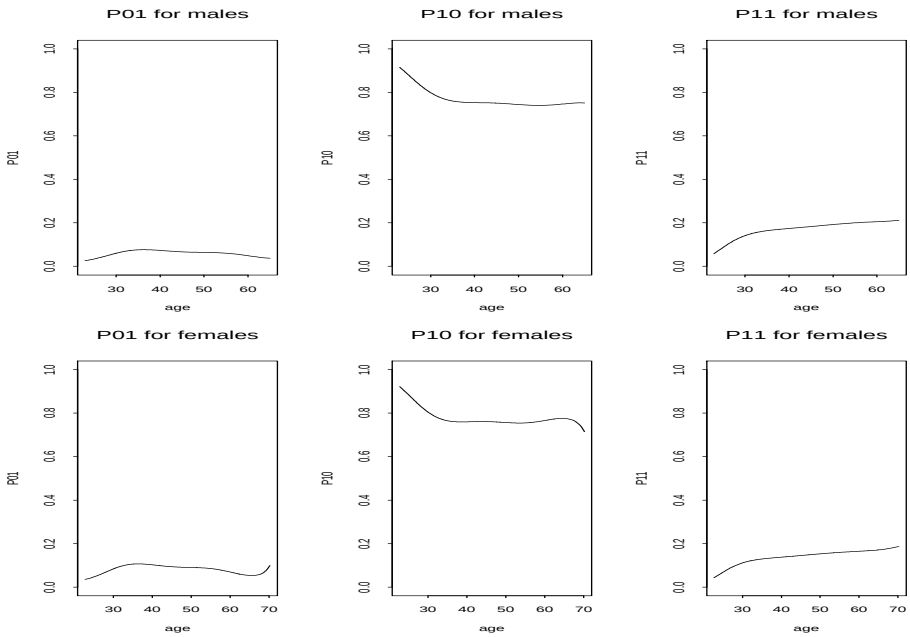

Figure 2: Estimated probabilities of frequencies related to age and gender.

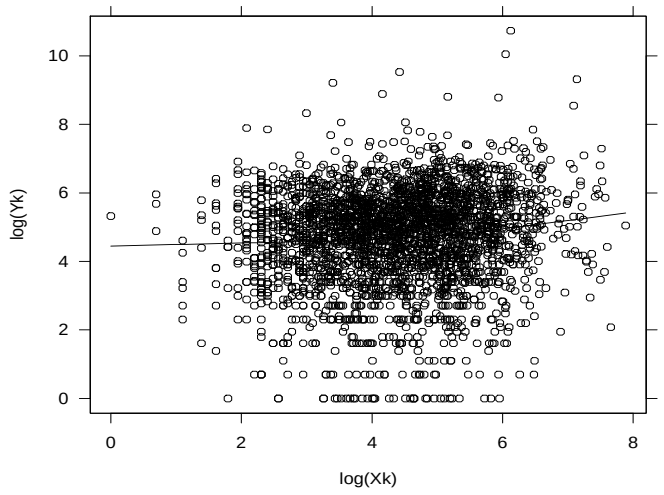

Figure 3: Scatter plot of the natural logarithm of $x_k = \text{drugs}_k$ and $y_k = \text{other}_k$ where $x_k, y_k > 0$.

3.2 Severity Model

Before fitting and carrying out tests about the severity distributions, it is useful to explore the nature of the correlation structure of the paired positive observations. This initial analysis can provide a feel for the choice of the form in which the marginals of H_{11} associate. For this, we plot the pair $(\log x_k, \log y_k)$, for $x_k, y_k > 0$, which is displayed in Fig. 3. This scattered plot shows a locally weighted regression smooth obtained using the function `loess.smooth` in the package S-PLUS. We can observe that the observations show a very modest positive association between the two types of claims.

At this point, there is weak evidence to assume that there is certain association

Table 3: Estimates of the coefficients and standard errors of the covariates for the frequency model.

Parameter	01	10	21	12
$\log \lambda$	$-4.296\,(0.016)$	$-5.103\,(0.037)$	$-4.511\,(0.038)$	$-5.105\,(0.031)$
$\log \gamma$	$-0.334\,(0.043)$	$-0.861\,(0.132)$	$-0.402\,(0.098)$	$-1.075\,(0.101)$
$\log \alpha$	$0.321\,(0.014)$	$0.266\,(0.033)$	$0.244\,(0.030)$	$0.030\,(0.022)$
θ			0.586	(0.109)

between the two types of claims for the non-zero paired observations. This might suggest that the choice of the *severity independent model*, namely $H_{11}(x, y) = F_{21}(x)F_{12}(y)$, may be adequate. However, it is important to consider that the inclusion of auxiliary variables in the marginals, as specified by our severity model, may change the degree of association between both marginals. Such degree can be assessed through the dependence parameter θ, as indicated above; furthermore, it is possible to carry out *severity independence* tests in the presence of covariates in the form of $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$. This will indicate whether the complication of the copula model is necessary.

The parameter estimates for the null model of the frequency distributions are displayed in Table 3. We observe that, although the point estimator of the dependence parameter represents a very modest association, $\tau_{\hat{\theta}} = 0.065$, the Wald test $|\hat{\theta}/\mathrm{SE}_{\hat{\theta}}| = 5.38$ suggests the rejection of the severity independence hypothesis. The dependence between the marginals of H_{11} is significant and, therefore, the parameter θ must be accounted for.

We can also observe in Table 3 that in general all remaining parameter estimates are individually significant. The only parameter that does not seem to be

significant is $\log \alpha_{12}$; therefore, if we test $H_0 : \log \alpha_{12} = 0$, i.e. the underlying distribution F_{12} is Pareto, we do not reject such hypothesis. The hypotheses that state that H_{01}, H_{10} and F_{21} can be parameterised as either Pareto or log-logistic can be rejected at a level of significance 0.05.

It is interesting to notice that the parameter estimates corresponding to H_{01} and F_{21} are alike; likewise the shapes of H_{10} and F_{12} show certain similarities. This puts forward the hypothesis that *the cause-specific severities are the same given frequency*, which can be stated as:

$$H_0 : \{h_{10}(x) = f_{21}(x)\} \wedge \{h_{01}(y) = f_{12}(y)\}$$

$$H_1 : \{h_{10}(x) \neq f_{21}(x)\} \vee \{h_{01}(y) \neq f_{12}(y)\};$$

under the Burr model described in the previous section, it is equivalent to

$$H_0 : \mathbf{B}\boldsymbol{\nu} = \mathbf{0} \qquad \text{vs.} \qquad H_1 : \mathbf{B}\boldsymbol{\nu} \neq \mathbf{0},$$

where $\mathbf{B} = (\mathbf{I}_{6(p+1)}, -\mathbf{I}_{6(p+1)}, \mathbf{0}_{6(p+1)\times 1})$. Here, $\mathbf{I}_{6(p+1)}$ represents the $6(p+1)\times 6(p+1)$ identity matrix, $\mathbf{0}_{6(p+1)\times 1}$ is the $6(p+1)$ zero column and $\boldsymbol{\nu}$ is the vector of parameters in the severity model as specified in Eq. (20). Calculating the statistic:

$$Z^2 = (\mathbf{B}\hat{\boldsymbol{\nu}})^T (\mathbf{B}\mathbf{V}_{\hat{\nu}}\mathbf{B}^T)^{-1} (\mathbf{B}\hat{\boldsymbol{\nu}}),$$

yields $Z^2 = 144.01$. Using the asymptotic χ^2 distribution with $6(p+1)$ degrees of freedom, we found that the p-value was very close to zero. Therefore, there is a preponderance of evidence to reject the equal cause-specific severities hypothesis.

The deviances of the maximised log-likelihoods in Eqs. (17), (18) and (19) for several regression models are presented in Table 4. We first observe that both age and gender are important for H_{01} and that model 9 shows the most significant deviance reduction for this distribution and, thus, take this regression model as

Table 4: Values of deviance for some nested models in the severity model.

		$-2\times\hat{l}_{01}^{(s)}$	Np	$-2\times\hat{l}_{10}^{(s)}$	Np	$-2\times\hat{l}_{11}^{(s)}$	Np	$\hat{\theta}$	95% CI for θ
1	*null*	175878.63	3	19611.28	3	72345.44	7	0.585	(0.371, 0.799)
2	`age`	175651.19	6	19609.62	6	72212.11	13	0.562	(0.348, 0.776)
3	`GENDER`	174830.79	6	19599.47	6	72224.83	13	0.595	(0.381, 0.808)
4	`age+GENDER`	174417.09	9	19597.97	9	72069.23	19	0.575	(0.361, 0.789)
5	`poly(age,2)+GENDER`	174167.16	12	19589.15	12	72026.88	25	0.584	(0.370, 0.798)
6	`poly(age,3)+GENDER`	174159.85	15	19587.13	15	72010.32	31	0.595	(0.381, 0.808)
7	`poly(age,4)+GENDER`	174067.81	18	19583.81	18	71991.35	37	0.587	(0.373, 0.801)
8	`poly(age,5)+GENDER`	174064.65	21	19582.94	21	71987.95	43	0.586	(0.371, 0.800)
9	`poly(age,6)+GENDER`	174052.71	24	19579.14	24	71980.06	49	0.590	(0.374, 0.806)
10	`poly(age,7)+GENDER`	174050.37	27	19577.21	27	71974.98	55	0.591	(0.375, 0.807)

the best. We also observe that, while Gender's contribution is significant for H_{10}, age seems unimportant, even after adding higher degree polynomials. We then take the regression model 3 for H_{10}.

Age and gender are significant in H_{11}; here, the regression model 8 shows the most significant deviance reduction. Notice that we include the same covariates in F_{21} and F_{12}; in this case, we observed that the polynomial of age turned out to be significant in both marginal distributions (p-value < 0.0001). We also observed that, while gender is significant for H_{01}, H_{10} and F_{21}, its contribution is unimportant for F_{12} (p-value=0.65). This suggests that there is an important discrepancy between the fits of H_{10} and F_{12}.

Table 4 also shows estimated dependence parameters and 95% confidence intervals, based on the approximation $(\hat{\theta}-\theta)/\widehat{\text{se}}_{\hat{\theta}} \sim \text{N}(0,1)$, for Frank's cop-

Table 5: Estimates of the coefficients and standard errors of the covariates for the best frequency regression models.

Parameter	01	10	21	12
$a_{\texttt{Intercept}}$	$-4.694\,(0.020)$	$-5.027\,(0.060)$	$-4.783\,(0.043)$	$-5.144\,(0.043)$
$a_{\texttt{GENDER}}$	$0.827\,(0.030)$	$-0.110\,(0.075)$	$0.594\,(0.072)$	$0.006\,(0.061)$
$b_{\texttt{Intercept}}$	$-0.840\,(0.076)$	$-0.573\,(0.203)$	$-0.764\,(0.150)$	$-1.564\,(0.216)$
$b_{\texttt{GENDER}}$	$1.046\,(0.092)$	$-0.417\,(0.266)$	$0.499\,(0.184)$	$0.237\,(0.237)$
$c_{\texttt{Intercept}}$	$0.335\,(0.018)$	$0.419\,(0.061)$	$0.315\,(0.038)$	$0.001\,(0.031)$
$c_{\texttt{GENDER}}$	$0.190\,(0.029)$	$-0.217\,(0.072)$	$0.005\,(0.060)$	$0.017\,(0.043)$
θ			0.586	(0.109)

ula model in each of the regression models of H_{11}. We can observe that, although the estimated dependences represent modest measures of severity association ($0.062 < \tau_{\hat{\theta}} < 0.066$), none of the confidence bands displayed include zero; therefore, there is evidence to reject the hypothesis of frequency independence.

Figure 4 displays the polynomials of age fitted in the severity model. The remaining coefficients of the best regressions obtained above are given in Table 5. We can observe that the coefficients corresponding to gender in the F_{12} are not statistically signifficant.

The diagnostics plots for both the marginal and dependence fits of the severity model are shown in Figs. 5 and 6. These graphs suggests that the severity model proposed here gives a good fit to the data.

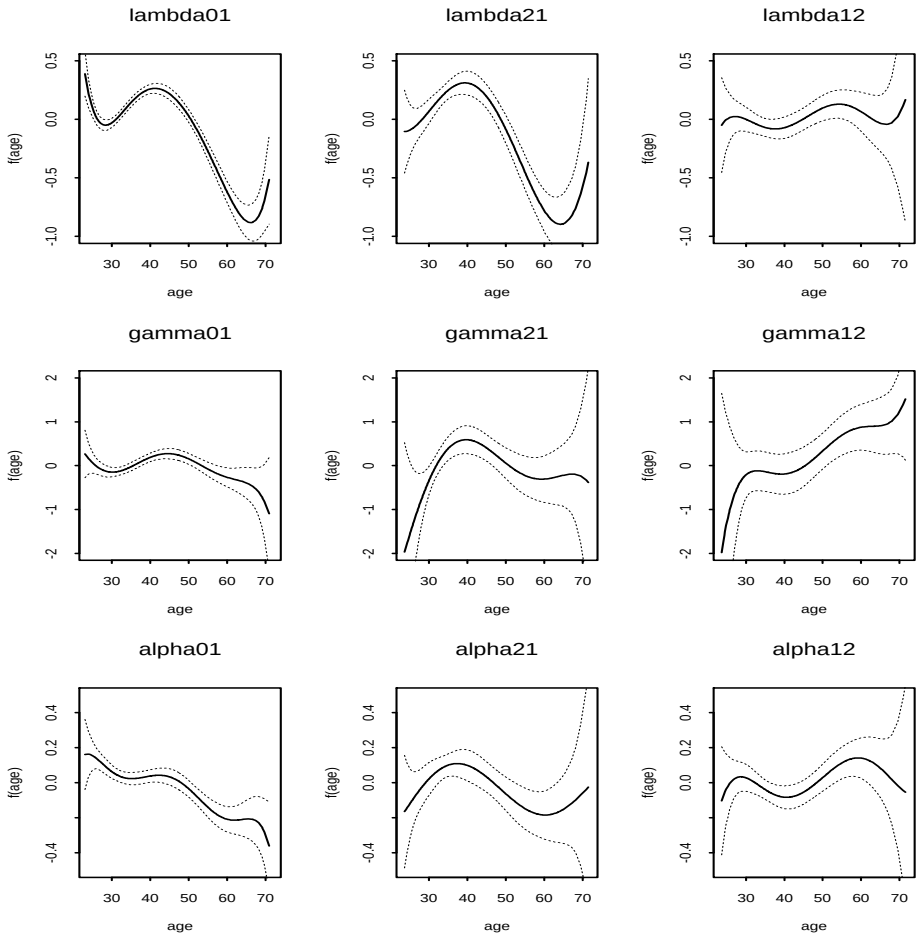

Figure 4: Polynomials of age for the best regression models in the severity distributions.

4 Conclusions

We have proposed a regression methodology for bivariate claims using a mixture model of frequency probabilities and Burr distributed severities. We have paid attention to both the fitting procedure and the model evaluation. Assuming maximum likelihood theory and some asymptotic results, we have carried out hypothesis tests concerning the choice of covariates and the parametric form of the

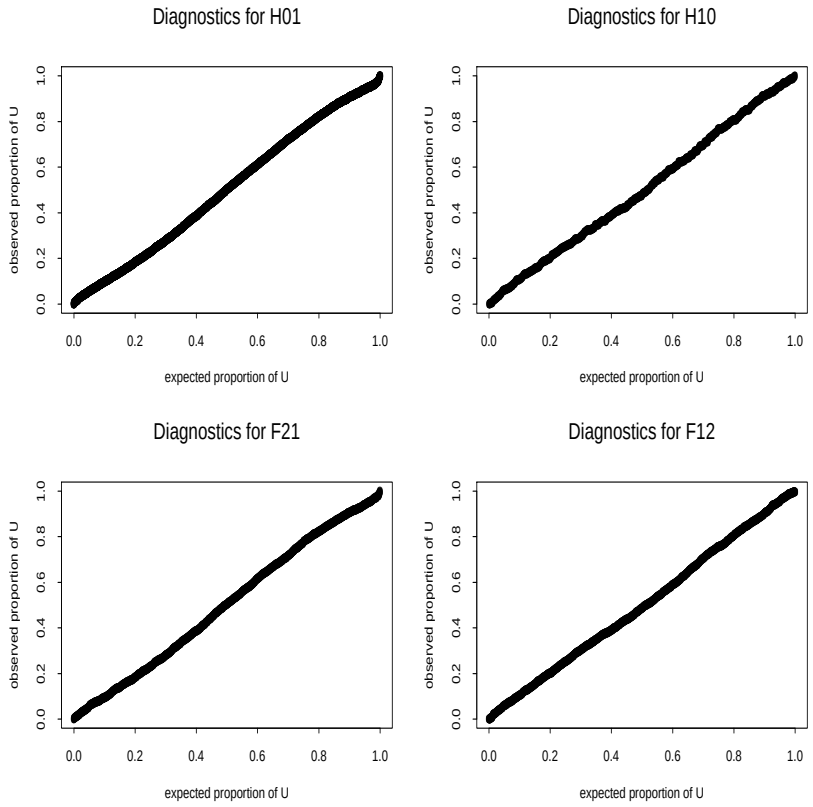

Figure 5: Diagnostic plots for the fitted marginal severity distributions.

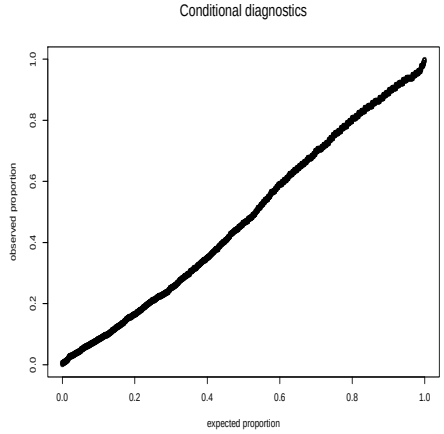

Figure 6: Diagnostic plot for the conditional distribution of the paired positive claims.

underlying severities.

A general feeling in the insurance business is that individual characteristics affect frequency distributions much more than severity distributions, and that severity is mostly independent of the individuals' type (Doherty, 1995). In the present study, we have observed that this conception may not be adequate. Our case study, which considers health claims, which are classified as drugs and "other", has proved that auxiliary variables such as age and gender are important predictors for both the frequencies and severities of the claims. Therefore, it is important to take into account concomitant information not only in the frequencies but also in the severities of the losses.

The strengths of the dataset we analysed are the comparatively large number of policies involved and the follow-up period. These strengths enabled us to investigate the data in more detail and with rather more confidence than studies with a more modest number would allow. We have shown that the use of polynomial regression to model age avoids the problems that result from inappropriate

linearity assumptions.

With regard to the severity model, other copula and distribution models could have been used. Thus, an interesting question would be to see if similar inferences can be obtained using different specifications. That question can be answered only by conducting similar research with this and other datasets and in different bivariate loss specifications to chart the boundaries of the results we have identified.

REFERENCES

BEIRLANT, J., GOEGEBEUR, Y., VERLAAK, R., & VYNCKIER, P. (1998) Burr regression and portfolio segmentation. **Insurance: Mathematics and Economics**, 23, pp. 231-250.

BENES, V. & STEPHAN, J. (1997) **Distributions with Given Marginals and Moment Problems** (Dordrecht, Kluwer).

CARRIÈRE J. (1997) Testing independence in bivariate distributions of claim frequencies and severities. **Insurance: Mathematics and Economics**, 21, pp. 81-89.

CHAR, B., GEDDES, K., GONNET, B., LEONG, B., MONAGAN, M., & WATT, S. (1991) **Maple V Library Reference Manual** (New York, Springer-Verlag).

COX, D. R. & SNELL, E. J. (1989) **Analysis of Binary Data**, 2nd ed. (London, Chapman & Hall).

DOHERTY, N. A. (1995) Severity risk and the adverse selection of frequency risk. **The Journal of Risk and Insurance**, 62, pp. 649-665.

FRANK, M. (1979) On the simultaneous associativity of $f(x, y)$ and $x+y-f(x, y)$. **Aequationes Mathematicae**, 19, pp. 194-226.

FREES, E. & VALDEZ, E. (1998) Understanding relationships using copulas. **North American Actuarial Journal**, 2, pp. 1-25.

GENEST, C. (1987) Frank's family of bivariate distributions. **Biometrika**, 74, pp. 549-555.

HOGG, R. & KLUGMAN, S. (1984) **Loss Distributions** (New York, Wiley).

KLUGMAN, S., PANJER, H., & WILLMONT, G. (1998) **Loss Models: From Data to Decisions** (New York, Wiley).

KLUGMAN, S. & PARSA, R. (1999) Fitting bivariate loss distributions with copulas. **Insurance: Mathematics and Economics**, 24, pp. 139-148.

NELSEN, R. (1986) Properties of a one-parameter family of bivariate distributions with specified marginals. **Communications in Statistics - Theory and Methods**, 15, pp. 3277-3285.

NELSEN, R. (1999) **An Introduction to Copulas** (New York, Springer).

PANJER, H. & WILLMOT, G. (1992) **Insurance Risk Models** (Schaumburg, Society of Actuaries).

SEBER, G. (1977). **Linear Regression Analysis**. (New York, Wiley).

WATKINS, A. (1999) An algorithm for maximum likelihood estimation in the three parameter Burr XII distribution. **Computational Statistics and Data Analysis**, 32, pp. 19-27.