Article

Modeling the Conditional Dependence between Discrete and Continuous Random Variables with Applications in Insurance

Emilio Gómez-Déniz [1,*] and Enrique Calderín-Ojeda [2]

[1] Department of Quantitative Methods and TIDES Institute, Campus de Tafira s/n, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain

[2] Centre for Actuarial Studies, Department of Economics, The University of Melbourne, 3010 Victoria, Australia; enrique.calderin@unimelb.edu.au

* Correspondence: emilio.gomez-deniz@ulpgc.es

Abstract: We jointly model amount of expenditure for outpatient visits and number of outpatient visits by considering both dependence and simultaneity by proposing a bivariate structural model that describes both variables, specified in terms of their conditional distributions. For that reason, we assume that the conditional expectation of expenditure for outpatient visits with respect to the number of outpatient visits and also, the number of outpatient visits expectation with respect to the expenditure for outpatient visits is related by taking a linear relationship for these conditional expectations. Furthermore, one of the conditional distributions obtained in our study is used to derive Bayesian premiums which take into account both the number of claims and the size of the correspondent claims. Our proposal is illustrated with a numerical example based on data of health care use taken from Medical Expenditure Panel Survey (MEPS), conducted by the U.S. Agency of Health Research and Quality.

Keywords: bivariate distributions; conditional distributions; credibility; premium

1. Introduction

The use of bivariate distributions where the marginals are one continuous and the other one discrete are extremely rare in statistics and more specifically in applied statistics. Some examples in which these types of distributions have been previously considered are [1–3]. The latter work was applied in tourism settings and the first two papers were applied in actuarial context. Conditional specification structure has also been applied in insurance framework when both marginal distributions are continuous. See [4] for details.

The statistical literature contains numerous methods to obtain bivariate distributions, some of them within the copula framework. These are functions that enable us to separate the marginal distributions from the dependency structure of a given multivariate distribution (see [5,6], among others). Several flexible approaches based on different bivariate copula families such as [7–9] have been proposed in the literature, including the Plackett and Frank copulas. The methodology proposed in this work allows us to build a bivariate distribution based on a conditional specification technique (see for instance [10,11]). In general, although the models obtained through this methodology rely on formulations that incorporate many parameters, this requirement can be relaxed, as it will be shown below, to obtain simple and practical expressions. The major advantage of this methodology, as compared to copula approach, is that the latter one incorporates a parameter that controls the correlation which is difficult to estimate, since it is restricted to a range of allowed values.(In practice, this can often be solved by choosing a suitable parametrization (e.g., using the logit or arctan of the correlation as parameter)).

In this work, we jointly model amount of expenditure for outpatient visits and number of outpatient visits by taking into account both dependence and simultaneity. It is noted that both variables are important inputs in the structure of health budgets and in health

insurance. More specifically, on the one hand, we model dependence by proposing a bivariate structural model that describes both variables specified in terms of their conditional distributions [10,11]. In fact, the model proposed here is a special case of the one obtained in [3]. On the other hand, simultaneity is implemented into the bivariate distribution, by making two assumptions: (*i*) the conditional expectation expenditure for outpatient visits with respect to the number of outpatient visits is linear; and (*ii*) the conditional expectation of outpatient visits with respect to the expenditure for outpatient visits is also linear. Thus, we assume that more expenditure implies that the considered health system is capable of handling more visits (note that we are not assuring that more expenditure involves a greater number of visits, which is difficult to sustain in practice); on the other hand, more visits obviously imply more expense.

Furthermore, one of the conditional distributions obtained in our study is used to calculate Bayesian premiums which take into account both the number of claims and the size of the correspondent claim. A credibility expression for the Bayesian premium in which the credibility factor obeys the classical expression provided in [12] is also derived. This type of premiums can be useful in automobile insurance contract to take into consideration both the number of claims and the severity of claim.

The second major contribution of this work is based on the fact that the proposed model enables us to evaluate the effect of covariates on both expenditure and number of visits. Thus, we can distinguish between the effects of these explanatory variables on both responses and, therefore, we can determine which covariates simultaneously affect both expenditure and the number of visits.

The proposed model is empirically evaluated by using data of health care use taken from Medical Expenditure Panel Survey (MEPS), conducted by the U.S. Agency of Health Research and Quality. This set of data can be downloaded from the web page of Professor E. Frees. We jointly model the number of outpatient visits and the amount of expenditure of outpatient visits. Some factors affecting these two random variables will be considered.

The rest of this paper is organized as follows. Section 2 provides the general motivation of this study and the main results. The proposed model for number of visits and expenditure is introduced here. A bivariate model allowing for the implementation of fixed effects is provided in Section 3. Some methods of estimation are shown in Section 4. Next, an empirical analysis of these models is performed in Section 5 and finally Section 6 summarizes the main conclusions drawn from this work.

2. Motivation and Results

In this section, we propose a non-linear bivariate model of amount of expenditure for outpatient visits and number of outpatient visits. The model is a special case of the one provided in [3] which was built by using a characterization via conditional expectations. In this case it was constructed through a class of bivariate distributions such that the conditional distributions belong to a specified exponential family.

As is well known, multivariate distributions can be specified in terms of their conditional distributions rather than directly. In this case, the dependence structure can be modeled using bivariate copulas or other mathematical or statistical methods such as conditional distributions or mixing distributions. By assuming that the conditional distributions belong to certain parametric families of distributions, the joint distribution can be obtained as described in [10] (see also [11] for an introduction to this topic). For applied works in this setting, see [1,3,13], among others. To obtain the joint distribution, it is first necessary to determine the solution of certain functional equations, which facilitates the derivation of highly flexible multiparametric distributions. As [11] pointed out, when we wish to specify a bivariate distribution it is sometimes convenient to visualize conditional distributions rather than marginal or joint distributions. In this context, it is useful in statistical modeling to have tractable multivariate distributions with given marginals to quantify the dependence effect of the variables in the model.

Accordingly, let us assume that the amount of expenditure for outpatient visits is represented as a random variable X and that the number of outpatients visits to the hospital is also random, and denoted by N. We now wish to obtain the more general bivariate distribution (X, N) whose conditional distributions satisfy the following conditions based on expenditure, conditional on the number of visits and on the number of visits to the hospital conditional to expenditure.

On the one hand, the expenditure for outpatient visits depends on the number of outpatient visits, i.e., if $X|N = n$ is the expenditure conditional to $N = n$ visits is directly proportional to the number of outpatient visits. Hence,

$$E(X|N = n) = \alpha_1 + \alpha_2 n, \quad \alpha_1 > 0, \quad \alpha_2 > 0. \tag{1}$$

According to expression (1), assuming that n takes values within the set of integer numbers $\{0, 1, \ldots, n\}$ there is a minimum expense given by α_1 in the case of no outpatient visits. Then, it seems coherent to assume that the expense increases as the number of visits increases.

On the other hand, it also seems logical to believe that the number of outpatient visits will depend on the expenditure, i.e., the number of outpatient visits that the health care system can assist will depend on its capacity, which in turn is conditioned by the budget (in terms of expenditure) available. Hence,

$$E(N|X = x) = \alpha_3 x, \quad \alpha_3 > 0. \tag{2}$$

Although Equation (1) includes an intercept since some expenditure would be necessary even for zero visits, pointing out that the health system must be prepared for any future visit, this is not the case of Equation (2). This is mainly due to two reasons: (i) On the one hand, this modelization facilitates the construction of the mathematical model obtained from Equation (2); and (ii) it seems obvious that an initial expenditure of near-zero monetary units implies that the health system is unable to absorb patients and, therefore, the mean of visits, and therefore the conditional mean, will be close to zero.

In some circumstances it is certain that this linear dependence relationship is not easily to be assumed. For example, an increase in the number of patient visits to the hospital might be caused by the suffering of more serious diseases that, in turn, incur in more expensive treatments. In this case, it is likely that a non-linear or logarithmic polynomial relationship could be more realistic. Obviously, the modeling in this case would be different from the one considered in this work.

In this work, we propose the use of a bivariate distribution that satisfies the conditions (1) and (2). We will consider the case where one of the conditional random variable is Gamma and the other one Poisson. A discrete random variable is said to follow a Poisson distribution with parameter $\lambda > 0$ if its probability mass function (pmf) is given by

$$f_N(n) = \Pr(N = n) = \frac{\lambda^n \exp(-\lambda)}{n!}, \quad n = 0, 1, 2, \ldots \tag{3}$$

In the following, $N \sim \mathcal{P}o(\lambda)$ will denote a random variable following a Poisson distribution with pmf given by (3). Moreover, a continuous random variable follows a gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$ if its probability density function (pdf) is expressed as

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x > 0. \tag{4}$$

We will write in this case, $X \sim \mathcal{G}(\alpha, \beta)$. By searching for the bivariate conditional distribution, let us now assume that

$$
\begin{aligned}
X|N=n &\sim \mathcal{G}(\sigma(n), \eta(n)), & (5)\\
N|X=x &\sim \mathcal{P}o(\varphi(x)) & (6)
\end{aligned}
$$

for given functions $\varphi(x) : \mathbb{N} \longrightarrow \mathbb{R}^+$, $\sigma(n) : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ and $\eta(n) : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$. Then, the conditional distributions given in (5) and (6) have pdfs given by

$$
\begin{aligned}
f_{N|X}(N|X=x) &= \frac{1}{n!} \exp\{-n \log \varphi(x) - \varphi(x)\}, & (7)\\
f_{X|N}(X|N=n) &= \frac{\eta(n)^{\sigma(n)}}{\Gamma(\sigma(n))} \exp\{-\eta(n)x + \sigma(n) \log x\}, & (8)
\end{aligned}
$$

i.e., they are members of the exponential family of distributions. Ref. [3] (see also [10], Chapter 4, p. 98) studied the most general bivariate distribution with conditionals given by (7) and (8) and concluded that this has pdf given by,

$$
f_{X,N}(x,n) = \frac{\kappa(\beta - \gamma)^n x^{\alpha + \tau n - 1}}{n!} \exp\{-(\beta + \varepsilon n)x\}, \tag{9}
$$

with $x > 0$, $n = 0, 1, \ldots$, where $\alpha \geq 0$, $\beta > \gamma > 0$, $\varepsilon \geq 0$ and $\tau \geq 0$. The parameter κ is a normalizing constant, a function of the remaining parameters, satisfying

$$
\kappa^{-1} = \int_0^\infty \sum_{n=1}^\infty \frac{(\beta - \gamma)^n x^{\alpha + \tau n - 1}}{n!} \exp\{-(\beta + \varepsilon n)x\} \, dx,
$$

which can be obtained from one of the conditional distributions given in (5) and (6), where

$$
\begin{aligned}
\sigma(n) &= \alpha + \tau n, & (10)\\
\eta(n) &= \beta + \varepsilon n, & (11)\\
\varphi(x) &= (\beta - \gamma) \exp\{-\varepsilon x + \tau \log x\}. & (12)
\end{aligned}
$$

Certainly, the major limitation of this type of models, built by conditional specification, lies in the difficulty that both the marginal distributions and the joint distribution depend on a normalization constant that in most cases is very difficult or even impossible to calculate in closed form (see for example [14]). Obviously, the difficulty grows when the two variables on which the model depends have different support or it is even more complicated when one variable is discrete and the other one is continuous, as the case considered here. However, from a numerical point of view, this is not an issue of relative importance since there exist integration and summation techniques (more difficult in the latter case) that allow us to compute the constant term. The reader is referred to [15] for more details on how to proceed in these cases.

For practical purposes, an appropriate choice of the parameter could enable us to obtain a closed-form expression for the normalizing constant. For example, the case $\varepsilon = \tau = 0$ corresponds to that one where N and X are independent and the marginal distributions are Poisson and gamma as in (3) and (4), respectively. The cases $\beta = 1 + \gamma$, $\varepsilon = 0$, $\tau = 1$ and $\varepsilon = 0$, $\tau = 1$ provide a closed-form model with conditional means of the type of (1) and (2) and thus suitable for our purpose. This latter model that includes an additional parameter as compared to the previous one will be considered here. Then, the bivariate distribution given in (9) can be rewritten as

$$
f_{X,N}(x,n) = \frac{(\gamma x)^\alpha [x(\beta - \gamma)]^n \exp(-\beta x)}{x \Gamma(n+1)\Gamma(\alpha)}, \quad x > 0, \ n = 0, 1, 2, \ldots \tag{13}
$$

Please note that the bivariate distribution (13) has a marginal distribution with continuous support and the other one with discrete support. This type of bivariate distribution is uncommon in theoretical and applied statistical literature, allowing us to model phenomena that are not usual in practical situations but sometimes occur, as it is the case described in this paper. See for example [1,16,17].

2.1. Marginal Distributions

The marginal distributions of X and N are given by

$$
\begin{aligned}
X &\sim \mathcal{G}(\alpha, \gamma), \quad \alpha > 0, \ \gamma > 0, \\
N &\sim \mathcal{NB}\left(\alpha, p = \frac{\gamma}{\beta}\right), \quad \alpha > 0, \ \beta > \gamma, \ \gamma > 0,
\end{aligned}
$$

where $\mathcal{NB}$ refers to the negative binomial distribution, i.e.,

$$
\begin{aligned}
f_X(x) &= \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\gamma x), \quad x > 0, \\
f_N(n) &= \binom{\alpha + n - 1}{n} \left(\frac{\gamma}{\beta}\right)^\alpha \left(1 - \frac{\gamma}{\beta}\right)^n, \quad n = 0, 1, \dots
\end{aligned}
$$

The marginal distribution of N is obtained by integrating (13) with respect to x over $(0, +\infty)$ and the marginal distribution of X is calculated by summing (13) with respect to n in the support $\{0, 1, \dots\}$.

Furthermore, it is readily apparent that the marginal means and variances are given by

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\gamma}, \quad var(X) = \frac{\alpha}{\gamma^2}, \\
E(N) &= 1 + \frac{\alpha}{\gamma}(\beta - \gamma), \quad var(N) = \frac{\alpha\beta(\beta - \gamma)}{\gamma^2},
\end{aligned}
$$

while the cross moment of X and N is

$$
E(XN) = \frac{\alpha}{\gamma^2}[\beta + \alpha(\beta - \gamma)].
$$

Simple calculations provide the covariance and the coefficient of correlation. These are given by

$$
\begin{aligned}
cov(X, N) &= \frac{\alpha(\beta - \gamma)}{\gamma^2}, \\
\rho(X, N) &= \sqrt{1 - \frac{\gamma}{\beta}},
\end{aligned}
$$

respectively, which are always positive. The coefficient of correlation is bounded between 0 ($\gamma \approx \beta$) and 1 ($\gamma \approx 0$). Thus, parameter γ controls the dependence structure of the model.

If the researcher wanted to work with a model that allows for positive and negative correlations, it would be enough to choose in (9) a less restrictive model.

2.2. Conditional Distributions

The conditional distribution of X given N, shown in (5), is gamma with the parameters given in (10) and (11) and the conditional distribution of N given X, see (6), is Poisson with parameter given in (12). Thus, the conditional distribution of X given $N = n$ is $\mathcal{G}(\sigma(n), \eta(n))$, where

$$
\begin{aligned}
\sigma(n) &= \alpha + n, \\
\eta(n) &= \beta
\end{aligned}
$$

and the conditional distribution of N given $X = x$ is $\mathcal{P}o(\varphi(x))$, with $\varphi(x) = (\beta - \gamma)x$. Observe that under these conditional distributions, expressions (1) and (2) are guaranteed.

Both the moments and maximum likelihood methods are feasible ways of estimating the vector of parameters $\boldsymbol{\phi} = (\alpha, \beta, \gamma)$ of the distribution through sample observations.

2.3. Hypothesis Testing

By using bivariate data, hypothesis testing can be performed with the parameters α, β and γ. We may also be interested in determining, via the likelihood ratio test, when the model might depend only on two parameters, e.g., when $\beta = \gamma + 1$. In this case, the vector $\boldsymbol{\phi}$ can be estimated with the constraint that $\beta = \gamma + 1$. If we denote the new vector by $\boldsymbol{\phi}^*$, the critical region for the null hypothesis $H_0 : \beta = \gamma + 1$ is determined from the test statistic

$$2\left[\ell(\widehat{\boldsymbol{\phi}}; (\tilde{x}, \tilde{n})) - \ell(\widehat{\boldsymbol{\phi}}^*; (\tilde{x}, \tilde{n}))\right],$$

which asymptotically has a χ^2-squared distribution with one degree of freedom.

3. Allowing for Covariates

In this section, we introduce a more practical model where covariates may be easily implemented. The bivariate linear regression model, which makes no distributional assumptions, is likely to be unsatisfactory because certain combinations of parameters and regressors could violate the non-negative restriction of the mean of both variables and therefore prediction could give non-realistic values. To avoid this situation, we propose a parametric model based on the distributional assumptions presented in the previous section when the researcher wishes to examine the factors or covariates that can affect simultaneously both means.

For our regression analysis, we reparameterized (13) as a function of its marginal means $\mu_1 = E(X)$ and $\mu_2 = E(N)$, then α is replaced by $\gamma\mu_1$ and β is replaced by $\gamma + \mu_2/\mu_1$, where $\mu_1 > 0$ and $\mu_2 \geq 1$. Then, the pdf (13) can be rewritten as

$$f(x, n) = \frac{(x\gamma)^{\gamma\mu_1}}{x\Gamma(n+1)\Gamma(\gamma\mu_1)} \left(\frac{x\mu_2}{\mu_1}\right)^n \exp\left[-\frac{x}{\mu_1}(\mu_2 + \gamma\mu_1)\right], \tag{14}$$

for $x > 0$ and $n = 0, 1, \ldots$ After this reparameterization we also obtain the cross moment, the covariance and the correlation, which are given by

$$\begin{aligned} E(XN) &= \mu_1\mu_2 + \frac{\mu_2}{\gamma}, \\ cov(X, N) &= \frac{\mu_2}{\gamma}, \\ \rho(X, N) &= \sqrt{\frac{\mu_2}{\gamma\mu_1 + \mu_2}}. \end{aligned} \tag{15}$$

Thus, parameter γ controls the dependence or independence of the model. As $\gamma \to 0$, $\rho(X, N) \to 1$ while $\gamma \to \infty$, $\rho(X, N) \to 0$.

Therefore, the model provided in (14) is suitable for including covariates since the marginal means are given by μ_1 and μ_2. The expression $(X, N) \sim \mathcal{BGNB}(\gamma, \mu_1, \mu_2)$ is used to denote a bivariate random variable (X, N) that follows the distribution given in (14), with marginals gamma and negative binomial. Graphs of the density function for different parameter values and their corresponding contour plots are shown in Figure 1, revealing that the density seems to be unimodal for different values of its parameters.
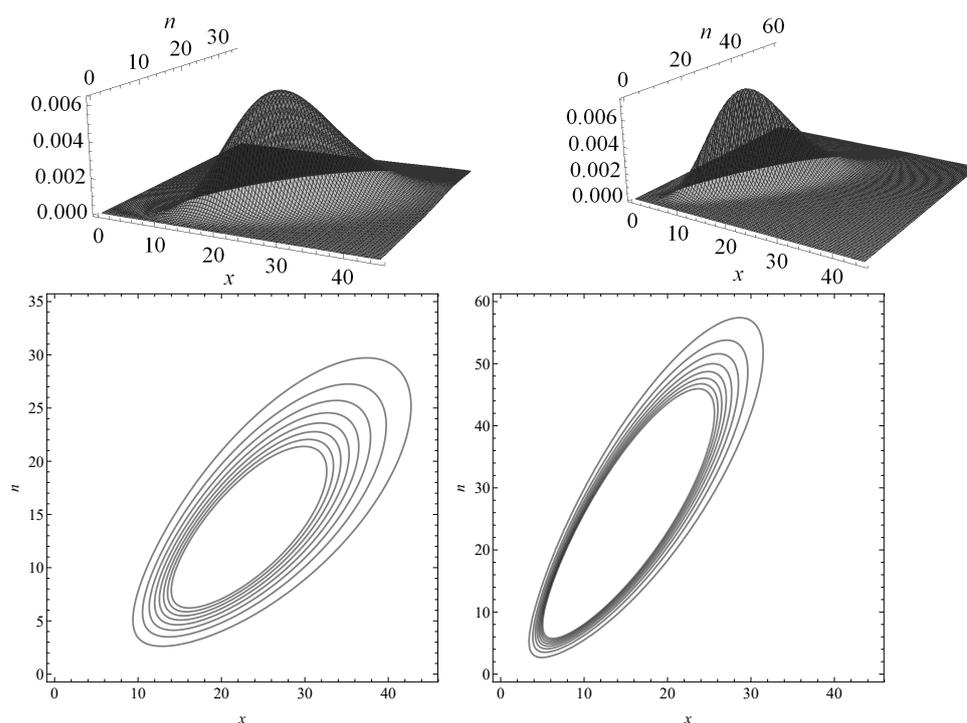
Figure 1. Graphs of the density for the bivariate distribution $f(x, n)$ in (14) and the corresponding contour plots for different values of the parameters. Left panel $(\gamma, \mu_1, \mu_2) = (0.5, 25, 15)$ and right panel $(\gamma, \mu_1, \mu_2) = (0.65, 15, 25)$.

Now, let $\boldsymbol{y}_i = (y_{1i}, \ldots, y_{ki})'$ and $\boldsymbol{w}_i = (w_{1i}, \ldots, w_{ki})'$ be two vectors of k covariates associated with the ith observation. They are two vectors of linearly independent regressors that are thought to determine (x, n). For the ith observation, the model takes the form

$$
\begin{aligned}
(X_i, N_i) &\sim \mathcal{BGNB}(\gamma, \mu_{1i}, \mu_{2i}), \\
\log(\mu_{1i}) &= \boldsymbol{y}_i'\boldsymbol{\delta}, \\
\log(\mu_{2i} - 1) &= \boldsymbol{w}_i'\boldsymbol{\eta},
\end{aligned}
$$

for $i = 1, \ldots, t$ and where t denotes the number of observations and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)'$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)'$ the corresponding vectors of regression coefficients.

Moreover, each of the variables, X and N, can be influenced by different factors, hence the explanatory variables that are taken to explain $\mu_{\kappa i}$, $\kappa = 1, 2$, could not be the same. Furthermore, observe that the log link assumed ensures that μ_{1i} falls within the interval $(0, \infty)$ and μ_{2i} within the interval $(1, \infty)$. Furthermore, in practice the model can handle factors that are only related to just one outcome and not to the other one.

In this study, all estimation computations were implemented by using the Wolfram Mathematica (v. 12.0, Wolfram Research Inc., Champaing, IL, USA) and RATS (v. 7.00 Estima, Evanston, IL, USA) software. In the latter case, the approximation given in (16) was used, obtaining the same results as those ones computed with Mathematica. For further information on these software packages, the reader is referred to [18,19].

4. Some Methods of Estimation

In the following, we derive estimators based on the moments and maximum likelihood methods for the model with and without covariates. We also provide closed-form expressions for the Fisher's information matrix.

4.1. Estimation of the Model without Covariates

The log-likelihood function and the normal equations to provide the corresponding estimates are almost given in closed form. Let us first consider the case of the model without covariates. If

$$(\tilde{x}, \tilde{n}) = \{(x_1, n_1), (x_2, n_2), \dots, (x_t, n_t)\}$$

is a sample obtained from the distribution with joint density (14) and $\bar{x} = (1/t) \sum_{i=1}^{t} x_i$, $\bar{n} = (1/t) \sum_{i=1}^{t} n_i$ and $\mu_{12} = (1/t) \sum_{i=1}^{t} x_i n_i$ are the corresponding sample moments, some computations provide the estimators based on these sample moments, which are given by

$$\widehat{\mu}_1 = \bar{x}, \quad \widehat{\mu}_2 = \bar{n}, \quad \widehat{\gamma} = \frac{\widehat{\mu}_2^2}{\widehat{\mu}_1(\mu_{12} - \widehat{\mu}_2(1 + \widehat{\mu}_2))}.$$

The Score Vector and Fisher Information Matrix

We now consider the method of maximum likelihood estimation. Let $\Theta = (\gamma, \mu_1, \mu_2)$ be the vector of parameters to be estimated. The log-likelihood function is proportional to

$$\begin{aligned} \ell((\tilde{x}, \tilde{n}); \Theta) \quad \propto \quad & t[\gamma \mu_1 (\bar{x}^* + \log \gamma) - \log \Gamma(\gamma \mu_1)] \\ & + t\bar{n}[\log \mu_2 - \log \mu_1] - \frac{t\bar{x}(\mu_2 + \gamma \mu_1)}{\mu_1}, \end{aligned}$$

where $\bar{x}^* = (1/t) \sum_{i=1}^{t} \log x_i$.

Thus, the normal equations which provide the estimators of the parameters are given by

$$\begin{aligned} \bar{x}^* + \log \gamma + 1 - \psi(\gamma \mu_1) - \frac{\bar{x}}{\mu_1} &= 0, \\ \mu_1^2 \gamma [\bar{x}^* + \log \gamma - \psi(\gamma \mu_1)] - \mu_1 \bar{n} + \bar{x} \mu_2 &= 0, \\ \mu_1 \bar{n} - \bar{x} \mu_2 &= 0, \end{aligned}$$

where $\psi(z)$ is the digamma function (the logarithmic derivative of the Euler gamma function). After some algebra the maximum likelihood estimators of μ_1 and μ_2 are obtained, which are given by $\widehat{\mu}_1 = \bar{x}$ and $\widehat{\mu}_2 = \bar{n}$. Finally, the estimator of the parameter γ is the solution of the equation

$$\log \gamma - \psi(\gamma \widehat{\mu}_1) + \bar{x}^* = 0,$$

which can be solved numerically.

The second partial derivatives are as follows:

$$\begin{aligned} \frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma^2} &= \frac{t\mu_1}{\gamma} - t\mu_1^2 \psi_1(\gamma \mu_1), \\ \frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma \partial \mu_1} &= t[\bar{1} + x^* + \log \gamma - \psi(\gamma \mu_1) - \gamma \mu_1 \psi_1(\gamma \mu_1)], \\ \frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma \partial \mu_2} &= 0, \\ \frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \mu_1^2} &= \frac{t\bar{n}}{\mu_1^2} - t\gamma^2 \psi_1(\gamma \mu_1) - \frac{2t\bar{x}}{\mu_1^3} \mu_2, \\ \frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \mu_1 \partial \mu_2} &= \frac{t\bar{x}}{\mu_1^2}, \\ \frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \mu_2^2} &= -\frac{t\bar{n}}{\mu_2^2}, \end{aligned}$$

where $\psi_1(\cdot)$ is the first derivative of the digamma function.

The entries of the Fisher's information matrix, $\mathcal{J}(\widehat{\Theta})$, are therefore

$$
\begin{aligned}
\mathcal{J}_{11}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial \gamma^2}\right)\bigg|_{\Theta=\widehat{\Theta}} = -\frac{t\widehat{\mu}_1}{\widehat{\gamma}} + t\widehat{\mu}_1^2 \psi_1(\widehat{\gamma}\widehat{\mu}_1),\\
\mathcal{J}_{12}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\gamma}\partial\widehat{\mu}_1}\right)\bigg|_{\Theta=\widehat{\Theta}} \approx -\sum_{i=1}^{t}\log x_i - t[(1+\log\widehat{\gamma}-\psi(\widehat{\gamma}\widehat{\mu}_1)\\
&\quad -\widehat{\gamma}\widehat{\mu}_1\psi_1(\widehat{\gamma}\widehat{\mu}_1)],\\
\mathcal{J}_{13}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\gamma}\widehat{\mu}_2}\right)\bigg|_{\Theta=\widehat{\Theta}} = 0,\\
\mathcal{J}_{22}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\mu}_1^2}\right)\bigg|_{\Theta=\widehat{\Theta}} = -\frac{t(\widehat{\mu}_2)}{\widehat{\mu}_1^2} + t\widehat{\gamma}^2\psi_1(\widehat{\gamma}\widehat{\mu}_1)\\
&\quad +\frac{2t\widehat{\mu}_2}{\widehat{\mu}_1^2},\\
\mathcal{J}_{23}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\mu}_1\widehat{\mu}_2}\right)\bigg|_{\Theta=\widehat{\Theta}} = -\frac{t}{\widehat{\mu}_1},\\
\mathcal{J}_{33}(\widehat{\Theta}) &= E\left(-\frac{\partial^2 \ell((\tilde{x},\tilde{n});\Theta)}{\partial\widehat{\mu}_2^2}\right)\bigg|_{\Theta=\widehat{\Theta}} = \frac{t}{\widehat{\mu}_2}.
\end{aligned}
$$

Here, $\widehat{\Theta}$ represents the maximum likelihood estimator of Θ. Observe that the analytic expression of $E(\sum_{i=1}^{t}\log x_i)$ is not feasible. For computational reasons, for large values of t, this is evaluated by ignoring the expectation operator and replacing it by $\sum_{i=1}^{t}\log x_i$. The asymptotic variance-covariance matrix of $\widehat{\Theta}$ is obtained by inverting the observed information matrix.

The above model has the advantage of its simplicity; however, the normal equations require the use of the digamma function, $\psi(z) = \frac{d}{dz}\log(\Gamma(z))$, $z > 0$. Although the digamma function is provided in most of the statistical software, in practical situations it is convenient to approximate the digamma function by the following expression,

$$
\log(\Gamma(z)) \approx \frac{1}{2}\log(2\pi) + \left(z-\frac{1}{2}\right)\log(z) - z + \frac{z}{2}\log\left(z\sinh\left(\frac{1}{z}\right)\right), \tag{16}
$$

which is well known in the statistical literature and it has been proven to perform well in practice.

4.2. Estimation of the Model with Covariates

For the sake of simplicity, we will assume (This assumption is only made to display the normal equations and the Fisher information matrix in a nicer format. In practice, as is well known, the maximum of the likelihood function can be obtained directly by maximizing the logarithm of the likelihood function, and therefore, the covariates used do not have to be the same for the two response variables.) $\boldsymbol{\eta} = \boldsymbol{\delta}$ and write $\mu_{1i} = \mu_{1i}(\boldsymbol{\delta})$ and $\mu_{2i} = \mu_{2i}(\boldsymbol{\delta})$. Let us also denote $\Theta = (\gamma, \boldsymbol{\delta})$, thus the log-likelihood function is proportional to

$$
\begin{aligned}
\ell((\tilde{x},\tilde{n});\Theta) &\propto \gamma\sum_{i=1}^{t}\mu_{1i}\log(\gamma x_i) - \sum_{i=1}^{t}\log\Gamma(\gamma\mu_{1i})\\
&\quad + \sum_{i=1}^{t}n_i[\log x_i + \log\mu_{2i} - \log\mu_{1i}]\\
&\quad - \sum_{i=1}^{t}\frac{x_i}{\mu_{1i}}(\mu_{2i}+\gamma\mu_{1i}).
\end{aligned}
$$

Thus, the normal equations, for $i = 1, \ldots, t$, are given by

$$\frac{\partial \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma} = \sum_{i=1}^{t} \mu_{1i}[1 + \log(\gamma x_i) - \psi(\gamma \mu_{1i})] - t\bar{x} = 0,$$

$$\frac{\partial \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \delta_j} = \gamma \sum_{i=1}^{t} \frac{\partial \mu_{1i}}{\partial \delta_j}[\log(\gamma x_i) - \psi(\gamma \mu_{1i})]$$
$$+ \sum_{i=1}^{t} n_i \left[\frac{1}{\mu_{2i}} \frac{\partial \mu_{2i}}{\partial \delta_j} - \frac{1}{\mu_{1i}} \frac{\partial \mu_{1i}}{\partial \delta_j} \right]$$
$$- \sum_{i=1}^{t} \frac{x_i}{\mu_{1i}^2} \left[\mu_{1i} \frac{\partial \mu_{2i}}{\partial \delta_j} - \mu_{2i} \frac{\partial \mu_{1i}}{\partial \delta_j} \right] = 0,$$

where $j = 1, \ldots, k$, $\frac{\partial \mu_{1i}}{\partial \delta_j} = y_{ij} \mu_{1i}$ and $\frac{\partial \mu_{2i}}{\partial \delta_j} = w_{ij} \mu_{2i}$. Again, the approximation given in (16) can be used instead of the digamma function. Finally, after computing the second partial derivatives we obtain the elements of the Fisher's information matrix, as follows:

$$\mathcal{J}_{11}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma^2} \right)\Bigg|_{\Theta = \widehat{\Theta}} = -\sum_{i=1}^{t} \widehat{\mu}_{1i} \left[\frac{1}{\widehat{\gamma}} - \widehat{\mu}_{1i} \psi_1(\widehat{\gamma} \widehat{\mu}_{1i}) \right],$$

$$\mathcal{J}_{12}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \gamma \partial \delta_j} \right)\Bigg|_{\Theta = \widehat{\Theta}} \approx -\sum_{i=1}^{t} y_{ij} \widehat{\mu}_{1i}[1 + \log(\widehat{\gamma} x_i)$$
$$-\psi(\widehat{\gamma} \widehat{\mu}_{1i}) - \widehat{\gamma} \widehat{\mu}_{1i} \psi_1(\widehat{\gamma} \widehat{\mu}_{1i})],$$

$$\mathcal{J}_{22}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \delta_j^2} \right)\Bigg|_{\Theta = \widehat{\Theta}} \approx -\sum_{i=1}^{t} \left\{ y_{ij}^2 \mu_{1i} \widehat{\gamma}[\log(\widehat{\gamma} x_i) - \psi(\widehat{\gamma} \widehat{\mu}_{1i})] \right.$$
$$\left. -y_{ij}^2 \widehat{\mu}_{1i}^2 \widehat{\gamma}^2 \psi_1(\widehat{\gamma} \widehat{\mu}_{1i}) + x_i \frac{\mu_{2i}}{\mu_{1i}} \left(w_{ji} - y_{ji} \right)^2 \right\},$$

$$\mathcal{J}_{23}(\widehat{\Theta}) = E\left(-\frac{\partial^2 \ell((\tilde{x}, \tilde{n}); \Theta)}{\partial \delta_j \partial \delta_l} \right)\Bigg|_{\Theta = \widehat{\Theta}} \approx -\sum_{i=1}^{t} \left\{ y_{ij} y_{il} \widehat{\gamma} \widehat{\mu}_{1i}[\log(\widehat{\gamma} x_i) - \psi(\widehat{\gamma} \widehat{\mu}_{1i})] \right.$$
$$\left. -y_{ij} y_{il} \widehat{\mu}_{1i}^2 \widehat{\gamma}^2 \psi_1(\widehat{\gamma} \widehat{\mu}_{1i}) + x_i \frac{\mu_{2i}}{\mu_{1i}} \left(w_{ij} w_{il} - w_{ij} y_{il} - y_{ij} w_{il} + y_{ij} y_{il} \right) \right\}, \quad j \neq l.$$

5. Empirical Analysis

To analyze the practical performance of the model proposed, two different examples are provided in this section.

5.1. Example 1

In this first example a set of data that can be downloaded from the web page of Professor E. Frees (https://instruction.bus.wisc.edu/jfrees/jfreesbooks/Regression%20 Modeling/BookWebDec2010/data.html) is considered. The dataset contains information about the Medical Expenditure Panel Survey (MEPS), conducted by the U.S. Agency of Health Research and Quality. This survey provides nationally representative estimates of health care use, expenditure, sources of payment, and insurance coverage for the U.S. civilian population. This survey collects detailed information on individuals (Please note that the unit of study is not the individual patient but a health system (public or private) which is the case considered here, although it could also be a hospital unit or similar) of each medical care episode by type of services including physician office visits, hospital emergency room visits, hospital outpatient visits, hospital inpatient stays, all other medical provider visits, and use of prescribed medicines. This detailed information allows the researcher to develop models of health care use to predict future expenditure (see http://www.meps.ahrq.gov/mepsweb/ for more details). We consider MEPS data consisting of 1352 individuals that have positive outpatient expenditure for individuals

between ages 18 and 65. In Table 1 the summary of descriptive statistics of these two random variables is displayed. Our dependent variables consist of the logarithm of the amounts of expenditure for outpatient $\log(EXPENDOP)$ visits (i.e., *Expenditure*) and number of *outpatient visits COUNTOP*, (i.e., Outpatient visits).

Table 1. Summary of descriptive statistics of the two response variables considered.

	Mean	Stand. Dev.	Minimum	Maximum
Expenditure	6.43	1.54	1.09	11.05
Outpatient visits	8.39	14.43	1	167

The degree of association between the two response variables *Expenditure* and *Outpatient visits* in the sample for the different levels is summarized in terms of some measures of correlation for bivariate data. In Table 2 Pearson's, Spearman's and Kendall's measures of correlation for these random variables are displayed. It is noticeable that there exists a moderate degree of correlation between these two variables. This is also confirmed by Figure 2 where the scatter plot of these two variables and the theoretical (smoothed) distribution are illustrated. In addition the fitted correlation obtained from (15) by using the maximum likelihood estimates of the parameters provided in Table 3 is 0.592.

Table 2. Pearson's, Spearman's and Kendall's measures of correlation for the two dependent variables considered.

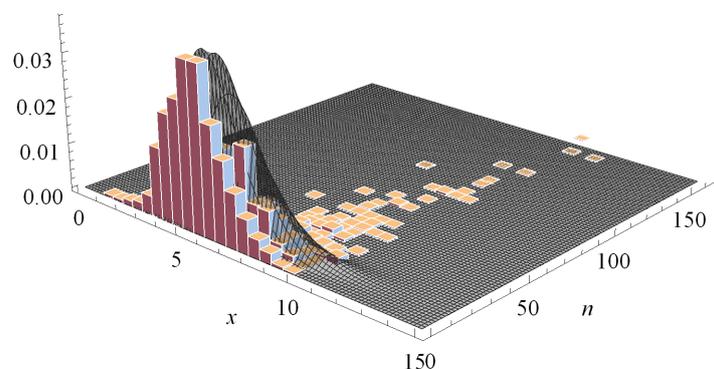| Measure of Correlation	Pearson's	Spearman's	Kendall's
	0.5232	0.7663	0.2856

Figure 2. Empirical (smoothed) distribution (histogram) and theoretical distribution model of the bivariate distribution considered.

For the bivariate model without covariates parameters estimates together with their corresponding standard errors are provided in Table 3 by using the parametrization of the bivariate model given in (14). In addition, the minimum value of the negative of the log-likelihood function (NLL) and the total sample size is displayed. Under this parametrization, $\widehat{\mu}_1$ and $\widehat{\mu}_2$ correspond to the estimates of the marginal means.

Parameter estimates and standard errors for the model with covariates are given below in Table 4 for the continuous random variable, *Expenditure*, and the discrete one, *Outpatient visits*. Explanation of the covariates considered can be found at the `Data Descriptions` file available on the same web page. As it can be seen many explanatory variables are statistically significant at the 5% level. Only the covariates *unemploy*, i.e., employment status of patients and *income*, i.e., income compared to poverty line do not have a significant impact simultaneously on the marginal means of the variables *Expenditure* and *Outpatient visits*. In addition, the dummy variable high school degree *highsch*, *managedcare*, *region* and marital status, *maristat* are not significant for the continuous response variable. On the other

hand, the explanatory variable *race*, i.e., race of the patient is not statistically significant at the same level for the *Outpatient visits* response variable.

Table 3. Parameter estimates, standard errors and negative of the log-likelihood (NLL) for the model without covariates.

Parameter	Estimate	Stand. Error
$\widehat{\gamma}$	2.412	0.094
$\widehat{\mu}_1$	6.428	0.044
$\widehat{\mu}_2$	8.394	0.098
NLL	11,014.532	
Observations	1352	

Table 4. Parameter estimates, *p*-values and the log-likelihood (NLL) for the model with covariates.

	Expenditure		*Outpatient Visits*	
Parameter	Estimate	*p*-Value	Estimate	*p*-Value
age	0.119	0.000	0.165	0.000
anylimit	0.075	0.000	0.721	0.000
college	0.014	0.092	0.218	0.000
highsch	−0.002	0.821	0.122	0.000
gender	0.041	0.000	0.253	0.000
mnhpoor	0.047	0.020	0.257	0.000
insure	0.099	0.000	0.105	0.002
usc	0.052	0.001	0.161	0.000
unemploy	−0.000	0.953	0.003	0.885
managedcare	0.008	0.541	−0.127	0.000
famsize	−0.010	0.006	−0.048	0.000
countop	0.070	0.000	0.199	0.000
race	0.017	0.003	0.015	0.148
region	−0.001	0.687	−0.054	0.000
education	0.009	0.033	0.094	0.000
maristat	−0.007	0.219	−0.058	0.000
income	−0.006	0.094	0.002	0.774
physical health	0.028	0.000	0.131	0.000
constant	1.172	0.000	0.594	0.000
γ	3.052 (0.000)			
NLL	9163.372			
Observations	1352			

5.2. Example 2

By using statistical tools from risk theory, the conditional distribution of N given $X = x$ provided in Section 2 can be used to compute bonus-malus premiums based on both number and size of the claims. The procedure of handling the variables number of claims and their respective claims size in the actuarial literature and, in particular, in automobile insurance, is usually based, given the difficulty involved in their study, on considering both variables separately (see [20,21]). The main advantage provided by the model proposed in this work is that it treats them together by taking into account their positive correlation. Please note that this model differs from the traditional collective risk model where the variable of interest is $S = \sum_{j=1}^{N} X_i$ and it is assumed that X_j are i.i.d. and independent of the variable N.

Let us now consider that the conditional distribution $N|X = x \sim \mathcal{P}(\mu_2 x / \mu_1)$ is defined in a way such that μ_2 follows a gamma prior distribution with shape parameter $a > 0$ and scale parameter $b > 0$. Then, it is straightforward that after observing the data

$\tilde{n} = (n_1, \ldots, n_t)$, the posterior distribution of μ_2 given the data $\tilde{n}$, is again gamma with updated parameters

$$a^* = a + \sum_{i=1}^{t} n_i, \tag{17}$$

$$b^* = b + \frac{tx}{\mu_1}. \tag{18}$$

Now, under squared-error loss function the net risk premium is just the mean of the random variable N, depending obviously on x, i.e.,

$$P(x, \mu_1, \mu_2) = E(N|X = x, \mu_1, \mu_2) = \frac{\mu_2 x}{\mu_1}$$

and the collective (prior) premium is the mean of the latter one and it results

$$P = E[E(N|X = x, \mu_1, \mu_2)] = \frac{ax}{b\mu_1}. \tag{19}$$

Since the posterior distribution is conjugate with respect to the likelihood, the Bayesian net premium is obtained directly from (19) by using (17) and (18) and results

$$P^* = \frac{(a + n^+)x}{\left(b + \frac{tx}{\mu_1}\right)\mu_1}, \tag{20}$$

where $n^+ = \sum_{i=1}^{t} n_i$.

In actuarial statistics, the Bayesian premium is a useful premium calculation tool, especially in the automobile insurance sector, since it combines sample information (of an insured) with a priori information (related to the group of insured). This Bayes premium can be easily calculated, i.e., the posterior mean of the parameter μ_2, when the prior distribution is conjugated with respect to the likelihood used since it is obtained directly from the collective premium by updating the parameters of this prior obtained from the data.

Observe that expression (20) increases with both the number of claims and the claims size, x. Additionally, this Bayesian premium can be rewritten as

$$P^* = Z_t \bar{n} + (1 - Z_t)P,$$

where $\bar{n} = n^+/t$ is the sample mean and

$$Z_t = \frac{tx}{b\mu_1 + tx}$$

is known in the actuarial literature as the credibility factor. It is straightforward to see that $Z_t \to 0$ as $t \to 0$, $Z_t \to 1$ as $t \to \infty$ and $Z_t = t/(t + K)$ where $K = var(E(N|X = x, \mu_1, \mu_2))/E(var(N|X = x, \mu_1, \mu_2))$. Thus, the credibility factor obeys the expression given in [12] (see also [22,23], among others).

A simple application of the methodology used in this paper is provided in the following. We suppose a motor vehicle insurance portfolio with mean value of 1 and 1/2 for the number of claims and the amount of claims, respectively.

Then, the expression (20) has been used to compute the Bayesian premiums of an automobile insurance portfolio for different values of the claim size x, by assuming that the mean of the number of claims is 1 (i.e., $a = b = 1$) and that the mean of claims size is $\mu_1 = 0.5$. Then, the Bayesian premiums obtained are shown in Table 5 for different values of the number of claims, $n^+ = \sum_{i=1}^{t} n_i$, observed over the time period (usually a year) t.

It is discernible that for all the values of x, i.e., number of claims, the premium increases with the value of the number of claims, n^+. In a similar way, the premium decreases with the period of time, t in which the policyholder does not declare a claim. Also, for a fixed value of k and t, the value of the premium increases with claim size x.

Table 5. Bayesian premiums for different values of claim size and number of claims.

		n^+					
$x = 0.50$	t	0	1	2	3	4	5
	0	1.00	2.00	3.00	4.00	5.00	6.00
	1	0.50	1.00	1.50	2.00	2.50	3.00
	2	0.33	0.66	1.00	1.33	1.66	2.00
	3	0.25	0.50	0.75	1.00	1.25	1.50
	4	0.20	0.40	0.60	0.80	1.00	1.20
	5	0.16	0.33	0.50	0.66	0.83	1.00
		n^+					
$x = 1.00$	t	0	1	2	3	4	5
	0	2.00	4.00	6.00	8.00	10.00	12.00
	1	0.66	1.33	2.00	2.66	3.33	4.00
	2	0.40	0.80	1.20	1.60	2.00	2.40
	3	0.28	0.57	0.85	1.14	1.42	1.71
	4	0.22	0.44	0.66	0.88	1.11	1.33
	5	0.18	0.36	0.54	0.72	0.90	1.09
		n^+					
$x = 1.50$	t	0	1	2	3	4	5
	0	3.00	6.00	9.00	12.00	15.00	18.00
	1	0.75	1.50	2.25	3.00	3.75	4.50
	2	0.42	0.85	1.28	1.71	2.14	2.57
	3	0.30	0.60	0.90	1.20	1.50	1.80
	4	0.23	0.46	0.69	0.92	1.15	1.38
	5	0.18	0.37	0.56	0.75	0.93	1.12

6. Final Comments

It is useful in applied statistics to obtain tractable bivariate distributions with given marginals that allow us to quantify the degree of dependence of the variables in the model. Although the most traditionally used methodology has been through copulas, in this article, we have examined this issue by using conditional specification structure. In general, the models derived via this approach, are based on expressions that might incorporate many parameters; however, this requirement may be relaxed to obtain simple and useful expressions. Furthermore, this methodology allows us to derive bivariate distributions with discrete and/or continuous marginals. In addition, one of the conditional distributions obtained in this study is used to calculate Bayesian premiums that take into account both the number of claims and the size of the correspondent claims. Also, a credibility expression for the Bayesian premium was derived. Finally, as a numerical application, an example that jointly explain the amount of expenditure for outpatient visits and number of outpatient visits by taking into account both dependence and simultaneity was considered. The model proposed in this paper enable us to evaluate and distinguish the effect of different explanatory variables on both number of visits and expenditure for outpatient visits.

References

1. Gómez-Déniz, E.; Calderín, E. Unconditional distributions obtained from conditional specifications models with applications in risk theory. *Scand. Actuar. J.* **2014**, *7*, 602–619. [CrossRef]
2. Gómez-Déniz, E.; Pérez-Rodríguez, J. Modelling dependence between daily tourist expenditure and length of stay. *Tour. Econ.* **2020**, in press.
3. Sarabia, J.; Gómez-Déniz, E.; Vázquez-Polo, F. On the use of conditional specification models in claim count distributions: An application to bonus-malus systems. *ASTIN Bull.* **2004**, *34*, 85–89. [CrossRef]
4. Sarabia, J.; Castillo, E.; Gómez-Déniz, E.; Vázquez-Polo, F. A class of conjugate priors for log–normal claims based on conditional specification. *J. Risk Insur.* **2005**, *72*, 479–495. [CrossRef]
5. Sklar, A. Fonctions de répartition á *n* dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris* **1959**, *8*, 229–231.
6. Sklar, A. Random variables, joint distributions, and copulas. *Kybernetika* **1973**, *9*, 449–460.
7. Farlie, D. The performance of some correlation coeficients for a general bivariate distribution. *Biometrika* **1960**, *47*, 307–323. [CrossRef]
8. Gumbel, E. Bivariate exponential distributions. *J. Am. Stat. Assoc.* **1960**, *55*, 698–707. [CrossRef]
9. Morgenstern, D. Einfache beispiele zweidimensionaler verteilungen. *Mitt. Math. Stat.* **1956**, *8*, 234–235.
10. Arnold, B.; Castillo, E.; Sarabia, J. *Conditional Specification of Statistical Models*; Springer: New York, NY, USA, 1999.
11. Arnold, B.; Strauss, D. Bivariate distributions with conditionals in prescribed exponential families. *J. R. Stat. Soc.* **1991**, *53*, 365–375. [CrossRef]
12. Bühlmann, H. Experience rating and credibility. *ASTIN Bull.* **1967**, *4*, 199–207. [CrossRef]
13. Sarabia, J.; Castillo, E. Bivariate distributions based on the generalized three-parameter Beta distribution. In *Advances on Distribution Theory, Order Statistics and Inference*; Balakrishnan, N., Castillo, E., Sarabia, J., Eds.; Birkhäuser: Boston, MA, USA, 2006; Chapter 2, pp. 85–110.
14. Moschopoulos, P.; Staniswalis, J.G. Estimation given conditionals from an exponential family. *Am. Stat.* **1994**, *48*, 271–275.
15. Davis, P.; Rabinowitz, P. *Methods of Numerical Integration*; Academic Press: San Diego, CA, USA, 1984.
16. Kotz, S.; Balakrishnan, N.; Johnson, N. Continuous Multivariate distributions. In *Models and Applications*; John Wiley and Sons: New York, NY, USA, 2000; Volume 1.
17. Spanos, A. *Probability Theory and Statistical Inference*; Cambridge University Press: Cambridge, UK, 1999.
18. Brooks, C. *RATS Handbook to Accompany Introductory Econometrics for Finance*; Cambridge University Press: Cambridge, UK, 2009.
19. Ruskeepaa, H. *Mathematica Navigator. Mathematics, Statistics, and Graphics*, 4th ed.; Academic Press: Cambridge, MA, USA, 2009.
20. Frangos, N.; Vrontos, S. Design of optimal Bonus-Malus systems with a frequency and a severity component on an individual basis in automobile insurance. *ASTIN Bull.* **2001**, *31*, 1–22. [CrossRef]
21. Mert, M.; Saykan, Y. On a bonus malus system where the claim frequency distribution is geometric and the claim severity distribution is Pareto. *Hacet. J. Math. Stat.* **2005**, *34*, 75–81.
22. Bühlmann, H.; Gisler, A. *A Course in Credibility Theory and its Applications*. Springer: Berlin/Heidelberg, Germany, 2005.
23. Gómez-Déniz, E. A generalization of the credibility theory obtained by using the weighted balanced loss function. *Insur. Math. Econ.* **2008**, *42*, 850–854. [CrossRef]