DOCUMENT DE TREBALL

XREAP2017-07

MULTIVARIATE COUNT DATA GENERALIZED LINEAR MODELS: THREE APPROACHES BASED ON THE SARMANOV DISTRIBUTION

Catalina Bolancé (RISKCENTER, XREAP)
Raluca Vernic

Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution

Catalina Bolancé* and Raluca Vernic**
*Department of Econometrics, Riskcenter-IREA
University of Barcelona
**Faculty of Mathematics and Informatics
Ovidius University of Constanta

November 10, 2017

Abstract

Starting from the question: "What is the accident risk of an insured?", this paper considers a multivariate approach by taking into account three types of accident risks and the possible dependence between them. Driven by a real data set, we propose three trivariate Sarmanov distributions with generalized linear models (GLMs) for marginals and incorporate various individual characteristics of the policyholders by means of explanatory variables. Since the data set was collected over a longer time period (10 years), we also added each individual's exposure to risk. To estimate the parameters of the three Sarmanov distributions, we analyze a pseudo-maximum-likelihood method. Finally, the three models are compared numerically with the simpler trivariate Negative Binomial GLM.

Keywords: multivariate counting distribution, Sarmanov distribution, Negative Binomial distribution, Generalized Linear Model, ML estimation algorithm

1 Introduction

Quantifying the risk of an accident is essential for pricing in insurance markets. Insurers tended to focus on the risk associated with the policy contracted during a certain period, typically one year in non-life lines. Indeed, various studies adopt this approach when dealing with pricing in auto insurance lines (see, for example, Abdallah et al., 2016; Boucher and Inoussa, 2014; Bolancé et al., 2008; Boucher et al., 2007; Bolancé et al., 2003; Pinquet et al., 2001). However, few papers to date have attempted to analyse the policyholder's accident risk from a multivariate perspective. Two examples of the use of bivariate count data models to tackle pricing in the auto insurance line are provided by Abdallah et al. (2016) and Bermudez and Karlis (2011), while Shi and Valdez (2014) use copula-based models to a trivariate analysis in this same line.

Therefore, we seek to address the following question: What is the client's accident risk when he has more than one type of coverage from his insurance company? This question could be answered by using univariate generalized linear models (GLMs), i.e., we could estimate one model for each coverage assuming the independent behavior of this policyholder in relation to each coverage. Alternatively, we can use a multivariate GLM that allows us to obtain a joint distribution associated with each individual, which also takes into account the fact that the risks of accident covered by the insurance company are dependent.

In this paper, we analyze different multivariate models for claims frequencies with GLMs for marginals, and we propose three multivariate GLMs based on the Sarmanov distribution, on the grounds that they are better alternatives to the multivariate Negative Binomial (NB) model. The multivariate models proposed allow us to fit the multivariate accident rate of the policyholders that have contracted different risk coverages associated with different policies in distinct non-life lines. Our aim is to capture the relationship between the behavior of a policyholder as regards the different risk coverages he has contracted. We show that the approaches based on the Sarmanov distribution allow us to model the dependence under different assumptions, i.e. we can directly assume that the dependence exists between the number of claims in each coverage and, alternatively, we can assume that the dependence exists between the random effects associated with unobserved factors.

Taking the global perspective of the client of an insurance company, some analyses have attempted to consider the information contained in the different policies of the same policyholder (representing different insurance coverages) with regards to the policyholder's profitability and loyalty. For example, Guelman and Guillén (2014) analyzed the price elasticity associated with an insurance contract maximizing the overall profitability of the policyholder (see also, Guelman et al., 2014). Other analyses have focused on the policyholder lapse as they study the relationship that exists between the cancellations of different policies by the same policyholder (see, e.g., Guillén et al., 2012; Brockett et al., 2008).

To have sufficient multivariate information about the policyholders' behavior, we need to observe this behaviour over a long period, i.e., more than one year. Typically, insurance companies use annual information for pricing in non-life insurance lines. However, insurance premiums are subject to different adjustments, some of which may be related to the risk quantification and others to marketing strategies or customer selection. Thus, insurers need to analyze portfolio information over several periods and so, here we analyze 10 years of claims information on auto and home lines.

Univariate mixed Poisson GLMs have been widely used in non-life insurance pricing (see Frees, 2009, Chapter 12, for a review). In this paper, we study three trivariate GLMs based on the trivariate Sarmanov distribution, and we use them to model trivariate count data corresponding to frequency of claims in non-life insurance. The first model consists of the discrete trivariate Sarmanov with NB GLM marginals. The second model is similar to that proposed by Abdallah et al. (2016) for the bivariate case, although it incorporates certain modification. Thus, we mix the trivariate model of independent Poisson distributions with a trivariate Sarmanov distribution with Gamma distributed marginals. In the third model, we mix a discrete trivariate Sarmanov distribution with Poisson marginals with three independent Gamma distributions. The main difference between these models lies in their respective dependence structures. In all three we assume that each policyholder has a given exposure to risk which can differ for the distinct insurance lines. Moreover, the expected number of claims associated with the analyzed risks depends on a set of explanatory variables. In our case, these explanatory variables are related to the customer's characteristics and are the same for each counting variable, but they can change in function of the analyzed risk.

The maximum likelihood (ML) estimation of all the parameters of a model based on the trivariate Sarmanov distribution is far from straightforward. It requires adding different restrictions to the parameters and an optimal solution is not readily found. Alternatively, we analyze a pseudo-maximum-likelihood estimation method based on a conditional likelihood that allows us to estimate all three trivariate models obtained from the Sarmanov distribution.

We also compare the three Sarmanov's distributions with the well known alternative multivariate Poisson GLM mixed with Gamma that is, with the trivariate NB GLM. The numerical study is conducted on a set of trivariate claims data from auto and home insurance lines, collected over a period of 10 years from a portfolio belonging to an international insurance company operating in the Spanish market. In both lines we select claims at fault linked to civil liability coverage. Moreover, in the case of the auto insurance line we specifically differentiated two types of claims: only property damage and bodily injury. This distinction has been used previously in other studies focused on the severity of auto insurance claims (see Bahraoui et al., 2015; Bolancé et al., 2014; Bahraoui et al., 2014; Bolancé et al., 2008).

The rest of this paper is structured as follows: In Section 2, we review some univariate and trivariate mixed Poisson GLMs and introduce the main notation. In Section 3, we present the three mixed models which result in three trivariate Sarmanov with NB GLMs as marginal distributions. We analyse some properties and propose an algorithm for estimating the models based on the specificity of the Sarmanov distribution. In Section 4, we describe the data and discuss the results of the numerical application. Finally, we draw some conclusions in Section 5.

2 Mixed Poisson distributions

A mixed Poisson distribution is a generalization of the Poisson distribution that can overcome the restriction that the mean is equal to the variance, a restriction that is inappropriate for most counting random variables. A key property of this distribution is that it can be easily expressed as a GLM.

A well-known example of the mixed Poisson distribution is the NB distribution, which mixes the Poisson and Gamma distributions.

2.1 Univariate case

Let N be the random variable (r.v.) total number of a certain type of claims of one insured for a given period. We assume that $N \sim Poisson(\theta)$, where θ is the realization of a positive and continuous r.v. Θ having a probability density function (p.d.f.) h; hence, N follows a mixed Poisson distribution with a probability function (p.f.) given by:

$$\Pr(N = n) = \int_0^\infty e^{-\theta} \frac{\theta^n}{n!} h(\theta)\, d\theta. \tag{1}$$

We recall that the expected value, variance and Laplace transform of this distribution are, respectively:

$$\mathbb{E}N = \mathbb{E}\Theta, VarN = \mathbb{E}\Theta + Var\Theta, \mathscr{L}_N(t) = \mathbb{E}\left(e^{-tN}\right) = \mathscr{L}_\Theta\left(1 - e^{-t}\right), \tag{2}$$

where $\mathscr{L}_\Theta$ denotes the Laplace transform of Θ.

In the following case, for sake of consistency with the GLMs, we shall consider the parameterization of the mixed Poisson distribution such that $\mathbb{E}N = \mu\theta$. Moreover, since in our numerical example we have three different types of claims, we shall index the r.v. N with the index j denoting the claims type wherever necessary.

2.1.1 Negative Binomial case

This distribution can be obtained by mixing Poisson and Gamma distributions. Hence, for consistency with the NB GLM, we assume that $N \sim Poisson(\mu\theta)$, where $\mu > 0$ is a fixed parameter and θ the realization of a Gamma distributed r.v. with mean 1 and variance $1/\alpha$, i.e., $\Theta \sim Gamma(\alpha, \alpha), \alpha > 0$. We easily obtain that:

$$\begin{aligned} \Pr(N = n) &= \int_0^\infty e^{-\mu\theta} \frac{(\mu\theta)^n}{n!} h(\theta)\, d\theta \\ &= \frac{\Gamma(\alpha + n)}{n!\Gamma(\alpha)} \left(\frac{\alpha}{\alpha + \mu}\right)^\alpha \left(\frac{\mu}{\alpha + \mu}\right)^n, n \in \mathbb{N}, \end{aligned} \tag{3}$$

hence $N \sim NB(\alpha, \tau)$, where $\tau = \frac{\alpha}{\alpha+\mu}$. In this case, from the properties of the NB distribution we have that:

$$\mathbb{E}N = \mu,\ VarN = \mu + \frac{\mu^2}{\alpha},$$
$$\mathscr{L}_N(t) = \mathscr{L}_\Theta\left(\mu\left(1-e^{-t}\right)\right) = \left(\frac{\alpha}{\alpha+\mu\left(1-e^{-t}\right)}\right)^{\alpha}, t > \ln\frac{\alpha}{\alpha+\mu}.$$

2.1.2 Adding exposure and explanatory variables. GLMs

Recall that in the numerical example we have different types of claim; hence, we let N_j denote the r.v. total number of claims of type j, $j = 1,\dots,m$, where m is the number of different claim types. At this point, we also introduce subscript i related to individual ($i = 1,\dots,I$). We know that during the period analyzed, the policyholders could have contracted more than one policy in the same line and, furthermore, that the duration of one contract could be shorter than that of the period analyzed. This means that the policyholder's exposure to risk may differ. Let E_{ij} be the exposure of individual i in the contracted coverage j. We define $E_{ij} = 1$ if the policyholder has contracted exactly one policy during the entire period under analysis; otherwise, we obtain $E_{ij} > 1$ if the policyholder has contracted more than one policy and the total duration is longer than that of the period analyzed and, alternatively, we obtain $E_{ij} < 1$ if the total duration is shorter than that of the period analyzed.

Additionally, we shall now consider the more general situation when the total number of a certain type of claim, N_{ij}, depends on certain individual characteristics of the policyholder i, i.e., we include explanatory variables (covariates) in GLM form. There are three components to GLM:

1. A stochastic component, which states that the observed r.v.s N_{ij} are independent and distributed in the exponential family.

2. A systematic component, according to which a set of covariates $X_{i0},\dots,X_{ip}$, where $X_{i0} = 1,\ \forall i$ is a constant term, produces a linear predictor with parameters $\beta_{0j},\dots,\beta_{pj}$ for each observation, i.e., $\eta_{ij} = \sum_{k=0}^{p} X_{ik}\beta_{kj}$.

3. A link function g relating the expected value of the stochastic component to the systematic component by $\eta_{ij} = g\left(\mu_{ij}\right)$, where $\mu_{ij} = \mathbb{E}(N_{ij})$.

For counting variables, a Poisson GLM is the first choice, in which case the canonical link function is the logarithmic function, i.e., $\eta_{ij} = \ln\left(\mu_{ij}\right) \Leftrightarrow \mu_{ij} = \exp\left(\eta_{ij}\right)$. However, in practice, the Poisson GLM does not usually provides a good fit because of the overdispersion that occurs when the response variance is greater than the mean. Alternatively, NB GLMs have been developed using the same link function (see McCullagh and Nelder, 1989).

Negative Binomial GLM with exposure. Such a model can be considered to arise wen we mix the Poisson distribution with a $Gamma(\alpha, \alpha)$ distribution as in formula (3). We denote by β_0 the intercept coefficient and, in view of the numerical study, we shall also introduce the exposure (note that the exposure frequently appears in GLMs as weights).

Let $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{ip})'$ be a column vector with the values of the explanatory variables of individual i and $\beta_j = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{pj})'$ the parameters vector associated with the coverage j. We assume the logarithmic link function $\mathbf{X}_i'\beta_j = \ln(\mu_{ij})$ or, inversely, $\mu_{ij} = \exp(\mathbf{X}_i'\beta_j)$; moreover, by including exposure E_{ij}, the individual expected value becomes:

$$\mathbb{E}(N_{ij}) = E_{ij}\mu_{ij} = E_{ij}\exp(\mathbf{X}_i'\beta_j).$$

Therefore, based on formula (3), we obtain for coverage j (hence, with α_j denoting the Gamma parameter) of the ith insured:

$$\begin{aligned} \Pr\left(N_{ij} = n\right) &= \frac{\Gamma\left(\alpha_j + n\right)}{n!\Gamma\left(\alpha_j\right)} \left(\frac{\alpha_j}{\alpha_j + E_{ij}\mu_{ij}}\right)^{\alpha_j} \left(\frac{E_{ij}\mu_{ij}}{\alpha_j + E_{ij}\mu_{ij}}\right)^{n} \qquad (4) \\ &= \frac{\Gamma\left(\alpha_j + n\right)}{n!\Gamma\left(\alpha_j\right)} \frac{\alpha_j^{\alpha_j}\exp\left\{n\left(\ln\left(E_{ij}\right) + \mathbf{X}_i'\beta_j\right)\right\}}{\left(\alpha_j + \exp\left\{\ln\left(E_{ij}\right) + \mathbf{X}_i'\beta_j\right\}\right)^{\alpha_j + n}}. \end{aligned}$$

In this case, the likelihood function is:

$$L\left(\alpha_j, \beta_j\right) = \prod_{i=1}^{I} \Pr\left(N_{ij} = n_{ij}\right) = \prod_{i=1}^{I} \frac{\Gamma\left(\alpha_j + n_{ij}\right)}{n_{ij}!\Gamma\left(\alpha_j\right)} \frac{\alpha_j^{\alpha_j}\exp\left\{n_{ij}\left(\ln\left(E_{ij}\right) + \mathbf{X}_i'\beta_j\right)\right\}}{\left(\alpha_j + \exp\left\{\ln\left(E_{ij}\right) + \mathbf{X}_i'\beta_j\right\}\right)^{\alpha_j + n_{ij}}},$$

where n_{ij} is the number of observed claims of policyholder i related to coverage j.

Also, the Laplace transform of N_{ij} becomes:

$$\begin{aligned} \mathscr{L}_{N_{ij}}(t) &= \left(\frac{\alpha_j}{\alpha_j + E_{ij}\mu_{ij}\left(1 - e^{-t}\right)}\right)^{\alpha_j} \\ &= \left(\frac{\alpha_j}{\alpha_j + \left(1 - e^{-t}\right)\exp\left\{\ln\left(E_{ij}\right) + \mathbf{X}_i'\beta_j\right\}}\right)^{\alpha_j}. \end{aligned}$$

2.2 Multivariate case

To obtain a multivariate mixed Poisson distribution, we let $N_j \sim Poisson\left(\mu_j\theta\right)$ with $\mu_j > 0$ fixed parameters, $j = 1, \ldots, m$, and consider θ to be the realization of some positive r.v. Θ with pdf h. We also assume that, conditionally on $\Theta = \theta$, the r.v.s N_j are independent. In the case of the numerical study, in what follows we shall only consider the NB case.

Multivariate Negative Binomial case. Under the assumptions outlined above, let

$\Theta \sim Gamma(\alpha,\alpha), \alpha > 0$. In this case, the joint probabilities of $\mathbf{N} = (N_1,...,N_m)$ are:

$$\begin{aligned} \Pr(\mathbf{N}=\mathbf{n}) &= \int_0^\infty \Pr(\mathbf{N}=\mathbf{n}\,|\Theta=\theta)\, h(\theta)\, d\theta \\ &= \frac{\alpha^\alpha}{\Gamma(\alpha)} \left(\prod_{j=1}^m \frac{\mu_j^{n_j}}{n_j!}\right) \int_0^\infty \theta^{\sum_{j=1}^m n_j+\alpha-1} e^{-\theta\left(\sum_{j=1}^m \mu_j+\alpha\right)} d\theta \\ &= \frac{\Gamma\left(\alpha+\sum_{j=1}^m n_j\right)}{\Gamma(\alpha)\prod_{j=1}^m n_j!} \left(\frac{\alpha}{\alpha+\sum_{k=1}^m \mu_k}\right)^\alpha \prod_{j=1}^m \left(\frac{\mu_j}{\alpha+\sum_{k=1}^m \mu_k}\right)^{n_j}, \mathbf{n} \in \mathbb{N}^m, \end{aligned}$$

which is the p.f. of the multivariate NB distribution defined as (see, e.g., Johnson et al., 1997):

$$NB_m\left(\alpha; \frac{\alpha}{\alpha+\sum_{j=1}^m \mu_j}, \left(\frac{\mu_j}{\alpha+\sum_{j=1}^m \mu_j}\right)_{j=1,\ldots,m}\right).$$

For our numerical application, we shall need the trivariate NB distribution ($m = 3$), in which we also include the exposure of each individual; i.e., introducing the subscript i related to the individual, we have that for $\mathbf{N}_i = (N_{i1}, N_{i2}, N_{i3}), i = 1,...,I$,

$$\mathbf{N}_i \sim NB_3\left(\alpha; \frac{\alpha}{\alpha+\sum_{k=1}^3 (E_{ik}\mu_{ik})}, \left(\frac{(E_{ij}\mu_{ij})}{\alpha+\sum_{k=1}^3 (E_{ik}\mu_{ik})}\right)_{j=1,2,3}\right). \quad (5)$$

The correlation coefficient between two marginals for individual i is:

$$corr\left(N_{ij}, N_{ik}\right) = \sqrt{\frac{E_{ij}\mu_{ij}E_{ik}\mu_{ik}}{\left(E_{ij}\mu_{ij}+\alpha\right)\left(E_{ik}\mu_{ik}+\alpha\right)}}, 1 \le j < k \le 3. \quad (6)$$

Let $(n_{i1}, n_{i2}, n_{i3})_{i=1}^I$ be a trivariate data sample with the corresponding exposures $(E_{i1}, E_{i2}, E_{i3})_{i=1}^I$ and we denote $\mu_i = (\mu_{i1}, \mu_{i2}, \mu_{i3}), i = 1,...,I$. Then the likelihood function with exposure is

$$L(\alpha, \mu_1, \ldots, \mu_n) = \prod_{i=1}^I \frac{\Gamma\left(\alpha+\sum_{j=1}^3 n_{ij}\right)}{\Gamma(\alpha)\prod_{j=1}^3 n_{ij}!} \left(\frac{\alpha}{\alpha+\sum_{k=1}^3 \mu_{ik}E_{ik}}\right)^\alpha \prod_{j=1}^3 \left(\frac{\mu_{ij}E_{ij}}{\alpha+\sum_{k=1}^3 \mu_{ik}E_{ik}}\right)^{n_{ij}}.$$

To estimate the parameters, we shall use the EM method proposed in Ghitany et al. (2012), which facilitates the ML estimation for multivariate mixed Poisson GLMs.

The classical multivariate NB model in (5) assumes that the dependence is based on a common random factor Θ, an assumption that implies a lack of flexibility in the dependence structure. Shi and Valdez (2014) considered some alternative multivariate models based on the NB and copulae that allow for a generalization of the dependence structure. Alternatively, we shall study a different method for generalizing the dependence structure by using the Sarmanov distribution.

3 Models based on the trivariate Sarmanov distribution

3.1 Trivariate Sarmanov distribution

This distribution can be defined in the discrete, as well as in the continuous case. In general, it is known (see Kotz et al., 2000) that a trivariate random vector $\mathbf{Y}$ follows a continuous trivariate Sarmanov distribution if its joint p.d.f. is given for $\mathbf{y} \in \mathbb{R}^3$ by:

$$h_{Sarm}(\mathbf{y}) = \prod_{j=1}^{3} h_j(y_j) \times \left(1 + \sum_{1 \leq j < k \leq 3} \omega_{jk} \phi_j(y_j) \phi_k(y_k) + \omega_{123} \phi_1(y_1) \phi_2(y_2) \phi_3(y_3)\right), \quad (7)$$

where $(h_j)_{j=1}^3$ are the marginal pdfs, $(\phi_j)_{j=1}^3$ are bounded non-constant kernel functions and $\omega_{jk},\ \omega_{123}$ are real numbers such that:

$$\begin{cases} \int_{\mathbb{R}} \phi_j(y) h_j(y) = 0, \text{ for } j = 1,2,3 \\ 1 + \sum_{1 \leq i < j \leq 3} \omega_{jk} \phi_j(y_j) \phi_k(y_k) + \omega_{123} \phi_1(y_1) \phi_2(y_2) \phi_3(y_3) \geq 0, \ \forall \mathbf{y} \in \mathbb{R}^3 \end{cases}. \quad (8)$$

The correlation coefficient between two marginal variables is related to the parameters ω_{jk} and the kernel functions ϕ_j by:

$$corr(Y_j, Y_k) = \omega_{jk} \frac{\mathbb{E}[Y_j \phi_j(Y_j)]\, \mathbb{E}[Y_k \phi_k(Y_k)]}{\sqrt{Var(Y_j)\, Var(Y_k)}}. \quad (9)$$

Proposition 1 *Concerning the parameter ω_{123}, it holds that:*

$$\omega_{123} = \frac{\mathbb{E}\left[\prod_{j=1}^{3} (Y_j - \mathbb{E}Y_j)\right]}{\prod_{j=1}^{3} \mathbb{E}[Y_j \phi_j(Y_j)]}. \quad (10)$$

Proof 1 *Let $j_3 = 6 - j_1 - j_2$, then we can write:*

$$\begin{aligned} \mathbb{E}\left[\prod_{j=1}^{3} (Y_j - \mathbb{E}Y_j)\right] &= \prod_{j=1}^{3} \int_{\mathbb{R}} (y_j - \mathbb{E}Y_j) h_j(y_j) dy_j \\ &+ \sum_{1 \leq j_1 < j_2 \leq 3} \omega_{j_1 j_2} \prod_{k=1}^{2} \int_{\mathbb{R}} \phi_{j_k}(y_{j_k})(y_{j_k} - \mathbb{E}Y_{j_k}) h_{j_k}(y_{j_k}) dy_{j_k} \int_{\mathbb{R}} (y_{j_3} - \mathbb{E}Y_{j_3}) h_{j_3}(y_{j_3}) dy_{j_3} \\ &+ \omega_{123} \prod_{j=1}^{3} \int_{\mathbb{R}} \phi_j(y_j)(y_j - \mathbb{E}Y_j) h_j(y_j) dy_j \\ &= \omega_{123} \prod_{j=1}^{3} \mathbb{E}[Y_j \phi_j(Y_j)], \end{aligned}$$

which easily yields the result. □

As for the form of the kernel functions ϕ_j, several alternatives are used in the literature: for example, the Farlie-Gumbel-Morgenstern (FGM) distribution obtained for $\phi_j = 1 - 2F_j$, where F_j is the cumulative distribution function of the marginal Y_j. Unfortunately, this FGM is restricted by the fact that the correlation coefficient of any two marginals cannot exceed $1/3$ in absolute value; hence, we do not consider it any further here. Another form of the kernel function is $\phi_j(y) = y - \mathbb{E}Y_j$, but this case usually involves truncated marginals to satisfy the conditions (8), which complicates computations. Therefore, here, we shall consider a third choice, which we call the exponential kernel, i.e., $\phi_j(y) = e^{-y} - \mathscr{L}_{Y_j}(1)$.

Since we shall work solely with nonnegative values, this last function $\phi_j(y) = e^{-y} - \mathscr{L}_{Y_j}(1)$ will be bounded. Note that it is also decreasing; hence, we denote:

$$\begin{aligned} m_j &= \inf_{y \geq 0} \phi_j(y) = \phi_j(\infty) = -\mathscr{L}_{Y_j}(1), \\ M_j &= \sup_{y \geq 0} \phi_j(y) = \phi_j(0) = 1 - \mathscr{L}_{Y_j}(1), j = 1, 2, 3. \end{aligned}$$

Then the conditions (8) yield the restrictions:

$$1 + \omega_{jk}\varepsilon_j\varepsilon_k \geq 0, 1 \leq j < k \leq 3, \quad (11)$$

$$1 + \sum_{1 \leq j < k \leq 3} \omega_{jk}\varepsilon_j\varepsilon_k + \omega_{123}\varepsilon_1\varepsilon_2\varepsilon_3 \geq 0, \quad (12)$$

where $\varepsilon_j \in \{m_j, M_j\}$, $j = 1, 2, 3$. From these conditions we can deduce bounds for the parameters ω_{jk} and ω_{123}.

In the discrete case, the joint probabilities of the trivariate Sarmanov distribution are given for $\mathbf{n} \in \mathbb{N}^3$ by

$$\begin{aligned} \Pr_{Sarm}(\mathbf{N} = \mathbf{n}) &= \prod_{j=1}^{3} \Pr(N_j = n_j) \\ &\times \left(1 + \sum_{1 \leq j < k \leq 3} \omega_{jk}\phi_j(n_j)\phi_k(n_k) + \omega_{123}\phi_1(n_1)\phi_2(n_2)\phi_3(n_3)\right). \quad (13) \end{aligned}$$

In this paper, we compare three models (see below) based on the trivariate Sarmanov distribution. All three models have the same marginals, but different kernel functions, and hence a different dependence structure.

3.2 Model I

For each individual i, we shall consider that $\mathbf{N}_i$ follows the discrete trivariate Sarmanov distribution expressed in (13) with type (4) NB GLM distributed marginals and kernel functions of exponential type

$$\phi_{ij}(n_j) = e^{-n_j} - \mathscr{L}_{N_{ij}}(1), \ j = 1, 2, 3.$$

In the case of this expression, note that when we use GLM marginals in (13), the kernel functions depend on individual i and, therefore, from conditions (11) and (12), we have to calculate different bounds of ω_{jk} and ω_{123} for each i (the ω's do not depend on individual i, but their limits do). In practice, we shall need to select the narrowest bounds.

3.3 Model II

We assume that $\mathbf{N}_i$ follows a trivariate Poisson distribution with independent marginals, which is mixed with a trivariate Sarmanov distribution with Gamma marginals, i.e. we assume that the dependence is given by the unobserved factor $\Theta_j,\ j=1,2,3$. Our model is a version of that proposed by Abdallah et al. (2016) for the bivariate case with a different parametrization. More specifically, we use $Gamma(\alpha_j,\alpha_j)$ marginals for the Sarmanov distribution, and we extend the model to the trivariate case. Since we also need to introduce the exposure, the p.f. of the mixed distribution is obtained by solving the following triple integral:

$$\Pr(\mathbf{N}_i=\mathbf{n})=\int_0^\infty\int_0^\infty\int_0^\infty\left(\prod_{j=1}^{3}e^{-E_{ij}\mu_{ij}\theta_j}\frac{(E_{ij}\mu_{ij}\theta_j)^{n_j}}{n_j!}\right)h_{Sarm}(\theta_1,\theta_2,\theta_3)\,d\theta_1 d\theta_2 d\theta_3,$$

where h_{Sarm} is given in (7) with h_j the pdf of the mixing marginal r.v. $\Theta_j\sim Gamma\left(\alpha_j,\alpha_j\right),\ j=1,2,3$, and the kernel functions $\phi_j\left(\theta_j\right)=e^{-\theta_j}-\mathscr{L}_{\Theta_j}(1)$. Note that in this model, the original trivariate Poisson distribution corresponds to the independence case. Then the resulting p.f. $\Pr(\mathbf{N}_i=\mathbf{n})$ is also of the Sarmanov type, but with more complex kernel functions, as can be seen from the following proposition.

Proposition 2 *Under the above assumptions, it holds that the mixed distribution of* $\mathbf{N}_i$ *has the p.f.:*

$$\begin{aligned}\Pr(\mathbf{N}_i=\mathbf{n}) &= \prod_{j=1}^{3}\Pr(N_{ij}=n_j)\\ &\left[1+\sum_{1\le j_1<j_2\le 3}\omega_{j_1 j_2}\prod_{k=1}^{2}\left(\left(\frac{\alpha_{j_k}+E_{ij_k}\mu_{ij_k}}{\alpha_{j_k}+E_{ij_k}\mu_{ij_k}+1}\right)^{\alpha_{j_k}+n_{j_k}}-\left(\frac{\alpha_{j_k}}{\alpha_{j_k}+1}\right)^{\alpha_{j_k}}\right)\right.\\ &\left.+\omega_{123}\prod_{j=1}^{3}\left(\left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j+n_j}-\left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}\right)\right],\end{aligned}$$

where the marginals $N_{ij}\sim NB\left(\alpha_j,\tau_{ij}\right)$ *with* $\tau_{ij}=\frac{\alpha_j}{\alpha_j+E_{ij}\mu_{ij}},\ j=1,2,3.$

Proof 2 *We have:*

$$\Pr(\mathbf{N}_i=\mathbf{n})=\int_0^\infty\int_0^\infty\int_0^\infty\left(\prod_{j=1}^{3}e^{-E_{ij}\mu_{ij}\theta_j}\frac{\left(E_{ij}\mu_{ij}\theta_j\right)^{n_j}}{n_j!}h_j\left(\theta_j\right)\right) \tag{14}$$
$$\times\left(1+\sum_{1\le j_1<j_2\le 3}\omega_{j_1 j_2}\prod_{k=1}^{2}\left(e^{-\theta_{j_k}}-\mathscr{L}_{\Theta_{j_k}}(1)\right)+\omega_{123}\prod_{j=1}^{3}\left(e^{-\theta_j}-\mathscr{L}_{\Theta_j}(1)\right)\right).$$

Recalling that $\mathscr{L}_{\Theta_j}(1) = \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}$ *is the Laplace transform of the Gamma* (α_j, α_j) *distribution, then we first evaluate:*

$$\begin{aligned}
& \int_0^\infty e^{-E_{ij}\mu_{ij}\theta_j} \frac{(E_{ij}\mu_{ij}\theta_j)^{n_j}}{n_j!} h_j(\theta_j) \left(e^{-\theta_j} - \mathscr{L}_{\Theta_j}(1)\right) d\theta_j \\
= & \int_0^\infty e^{-E_{ij}\mu_{ij}\theta_j} \frac{(E_{ij}\mu_{ij}\theta_j)^{n_j}}{n_j!} \frac{\alpha_j^{\alpha_j}}{\Gamma(\alpha_j)} \theta_j^{\alpha_j-1} e^{-\alpha_j\theta_j} \left[e^{-\theta_j} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}\right] d\theta_j \\
= & \frac{\alpha_j^{\alpha_j}(E_{ij}\mu_{ij})^{n_j}}{\Gamma(\alpha_j) n_j!} \left[\int_0^\infty \theta_j^{\alpha_j+n_j-1} e^{-\left(E_{ij}\mu_{ij}+\alpha_j+1\right)\theta_j} d\theta_j - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j} \int_0^\infty \theta_j^{\alpha_j+n_j-1} e^{-\left(E_{ij}\mu_{ij}+\alpha_j\right)\theta_j} d\theta_j\right] \\
= & \frac{\Gamma(\alpha_j+n_j)}{\Gamma(\alpha_j) n_j!} \left(\frac{\alpha_j}{\alpha_j+E_{ij}\mu_{ij}}\right)^{\alpha_j} \left(\frac{E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}}\right)^{n_j} \left[\left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j+n_j} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}\right] \\
= & \frac{\Gamma(\alpha_j+n_j)}{\Gamma(\alpha_j) n_j!} \tau_{ij}^{\alpha_j} (1-\tau_{ij})^{n_j} \left[\left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j+n_j} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}\right],
\end{aligned}$$

where $\tau_{ij} = \frac{\alpha_j}{\alpha_j+E_{ij}\mu_{ij}}$. *Inserting this formula into (14) we obtain the stated form of* $\Pr(\mathbf{N}_i = \mathbf{n})$, *which is also of the Sarmanov type with NB* (α_j, τ_{ij}) *marginals and kernel functions* $\phi_{ij}(n_j) = \left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j+n_j} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}$ *(it can be easily shown that* $\mathbb{E}\left[\phi_{ij}(N_{ij})\right] = 0, \forall i, j$*).*

In the following result we present restrictions for the parameters ω_{jk} and ω_{123}.

Proposition 3 *Under the assumptions of Model II, the following conditions must be fulfilled for all* $i = 1, \ldots, I$:

$$\max_{1\le j<k\le 3} \left\{\frac{-1}{M_{ij}M_{ik}}, \frac{-1}{m_j m_k}\right\} \le \omega_{jk} \le \min_{1\le j<k\le 3} \left\{\frac{-1}{M_{ij}m_k}, \frac{-1}{m_j M_{ik}}\right\},$$

$$\max_{\substack{1\le j<k\le 3 \\ h=6-j-k}} \left\{\frac{-1}{\prod_{l=1}^{3} M_{il}} - \sum_{\substack{1\le l_1<l_2\le 3 \\ l_3=6-l_1-l_2}} \frac{\omega_{l_1 l_2}}{M_{il_3}}, \frac{-1}{m_j m_k M_{ih}} - \frac{\omega_{jk}}{M_{ih}} - \frac{\omega_{jh}}{m_k} - \frac{\omega_{kh}}{m_j}\right\} \le \omega_{123},$$

$$\omega_{123} \le \min_{\substack{1\le j<k\le 3 \\ h=6-j-k}} \left\{\frac{-1}{\prod_{l=1}^{3} m_l} - \sum_{\substack{1\le l_1<l_2\le 3 \\ l_3=6-l_1-l_2}} \frac{\omega_{l_1 l_2}}{m_{l_3}}, \frac{-1}{M_{ij}M_{ik}m_h} - \frac{\omega_{jk}}{m_h} - \frac{\omega_{jh}}{M_{ik}} - \frac{\omega_{kh}}{M_{ij}}\right\},$$

where $m_j = -\left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}$, $M_{ij} = \left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}$, $j = 1, 2, 3, i = 1, \ldots, I$, *and, by convention,* $\omega_{jk} = \omega_{kj}$.

Proof 3 *The kernel function* $\phi_{ij}(n) = \left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j+n} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j}, n \in \mathbb{N}$, *is bounded and decreasing in n, hence its maximum is* $M_{ij} = \phi_{ij}(0) = \left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}+1}\right)^{\alpha_j} - \left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j} >$

0 *and its infimum is* $m_j = \phi_{ij}(\infty) = -\left(\frac{\alpha_j}{\alpha_j+1}\right)^{\alpha_j} < 0, \forall i = 1,\dots,I$. *Consequently, this type of kernel (8) also yields the restrictions (11)-(12), and from (11), we easily obtain the stated bounds of* ω_{jk}. *Regarding* ω_{123}, *(12) is equivalent to* $\omega_{123}\varepsilon_1\varepsilon_2\varepsilon_3 \geq -1 - \sum_{1\leq j<k\leq 3}\omega_{jk}\varepsilon_j\varepsilon_k$, *and by replacing each* ε_j *with the current* m_j *and* M_{ij}, *we obtain the result.*

We note that, since the maximum $M_{ij} = \phi_{ij}(0)$ depends on the individual expected value, the intervals for the parameters ω_{jk} and ω_{123} can differ for each i; hence, in practice, we need to select the narrowest interval.

3.4 Model III

For this model, we let $\mathbf{N}_i$ follow a mixed discrete trivariate Sarmanov distribution with independent Gamma mixing distributions, i.e.,

$$\Pr(\mathbf{N}_i = \mathbf{n}) = \int_0^\infty\int_0^\infty\int_0^\infty \Pr_{Sarm}(\mathbf{N}_i = \mathbf{n})\left(\prod_{j=1}^{3} h_j(\theta_j)\right) d\theta_1 d\theta_2 d\theta_3, \tag{15}$$

where $\Pr_{Sarm}(\mathbf{N}_i = \mathbf{n})$ is the discrete trivariate Sarmanov distribution (13) with Poisson marginals given by

$$\Pr_{Sarm}(\mathbf{N}_i = \mathbf{n}) = \left(\prod_{j=1}^{3} e^{-E_{ij}\mu_{ij}\theta_j}\frac{(E_{ij}\mu_{ij}\theta_j)^{n_j}}{n_j!}\right)\left(1 + \sum_{1\leq j_1<j_2\leq 3}\omega_{j_1j_2}\phi_{ij_1}(n_{j_1})\phi_{ij_2}(n_{j_2}) + \omega_{123}\prod_{j=1}^{3}\phi_{ij}(n_j)\right)$$

and with kernel function $\phi_{ij}(n) = e^{-n} - \mathscr{L}_{Po\left(E_{ij}\mu_{ij}\theta_j\right)}(1)$. The mixing distributions h_j are *Gamma* (α_j, α_j).

As can be seen from the following proposition, the resulting distribution is also of the Sarmanov type with the same marginals as in Model II, but with different kernel functions.

Proposition 4 *Under the above assumptions, the p.f. of the mixed distribution of* $\mathbf{N}_i$ *is:*

$$\begin{aligned}\Pr(\mathbf{N}_i = \mathbf{n}) \quad = \quad & \prod_{i=1}^{3}\Pr(N_{ij} = n_j)\left[1 + \sum_{1\leq j_1<j_2\leq 3}\omega_{j_1j_2}\prod_{k=1}^{2}\left(e^{-n_{j_k}} - \left(\frac{\alpha_{j_k} + E_{ij_k}\mu_{ij_k}}{\alpha_{j_k} + E_{ij_k}\mu_{ij_k}(2-e^{-1})}\right)^{\alpha_{j_k}+n_{j_k}}\right)\right.\\ & \left. + \omega_{123}\prod_{j=1}^{3}\left(e^{-n_j} - \left(\frac{\alpha_j + E_{ij}\mu_{ij}}{\alpha_j + E_{ij}\mu_{ij}(2-e^{-1})}\right)^{\alpha_j+n_j}\right)\right],\end{aligned}$$

where, as before, the marginals $N_{ij} \sim NB(\alpha_j, \tau_{ij})$ *with* $\tau_{ij} = \frac{\alpha_j}{\alpha_j + E_{ij}\mu_{ij}}$, $j = 1,2,3$.

Proof 4 *Since* $\phi_{ij}(n) = e^{-n} - e^{-E_{ij}\mu_{ij}\theta_j\left(1-e^{-1}\right)}$, *we obtain*

$$\begin{aligned}
& \int_0^\infty e^{-E_{ij}\mu_{ij}\theta_j} \frac{\left(E_{ij}\mu_{ij}\theta_j\right)^{n_j}}{n_j!} h_j(\theta_j)\,\phi_{ij}(n_j)\,d\theta_j \\
= & \int_0^\infty e^{-E_{ij}\mu_{ij}\theta_j} \frac{\left(E_{ij}\mu_{ij}\theta_j\right)^{n_j}}{n_j!} \frac{\alpha_j^{\alpha_j}}{\Gamma(\alpha_j)} \theta_j^{\alpha_j-1} e^{-\alpha_j\theta_j} \left[e^{-n_j} - e^{-E_{ij}\mu_{ij}\theta_j\left(1-e^{-1}\right)}\right] d\theta_j \\
= & \frac{\alpha_j^{\alpha_j}\left(E_{ij}\mu_{ij}\right)^{n_j}}{\Gamma(\alpha_j)\,n_j!} \int_0^\infty \left[\theta_j^{\alpha_j+n_j-1} e^{-\left(E_{ij}\mu_{ij}+\alpha_j\right)\theta_j} e^{-n_j} - \theta_j^{\alpha_j+n_j-1} e^{-\left(E_{ij}\mu_{ij}+\alpha_j+E_{ij}\mu_{ij}\left(1-e^{-1}\right)\right)\theta_j}\right] d\theta_j \\
= & \frac{\Gamma(\alpha_j+n_j)}{\Gamma(\alpha_j)\,n_j!} \left[e^{-n_j} \left(\frac{\alpha_j}{\alpha_j+E_{ij}\mu_{ij}}\right)^{\alpha_j} \left(\frac{E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}}\right)^{n_j} - \frac{\alpha_j^{\alpha_j}\left(E_{ij}\mu_{ij}\right)^{n_j}}{\left(E_{ij}\mu_{ij}+\alpha_j+E_{ij}\mu_{ij}\left(1-e^{-1}\right)\right)^{\alpha_j+n_j}}\right] \\
= & \frac{\Gamma(\alpha_j+n_j)}{\Gamma(\alpha_j)\,n_j!} \tau_{ij}^{\alpha_j}\left(1-\tau_{ij}\right)^{n_j} \left[e^{-n_j} - \left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}\left(2-e^{-1}\right)}\right)^{\alpha_j+n_j}\right].
\end{aligned}$$

Inserting this into (15), with some straightforward calculations, we obtain the stated formula.

The restrictions on the parameters ω_{jk} and ω_{123} are similar to these given in Proposition 3 with the maximums:

$$M_{ij} = \phi_{ij}(0) = 1 - \left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}\left(2-e^{-1}\right)}\right)^{\alpha_j}.$$

However, in this case, the minimums

$$m_{ij} = \min_{n_j\in\mathbb{N}} \left\{ e^{-n_j} - \left(\frac{\alpha_j+E_{ij}\mu_{ij}}{\alpha_j+E_{ij}\mu_{ij}\left(2-e^{-1}\right)}\right)^{\alpha_j+n_j} \right\}$$

also depend on individual i; moreover, they are obtained for some value in $\mathbb{N}$, and not by letting $n_j \to \infty$ as before.

3.5 Estimation procedure for Models I, II and III

Given the restricted shape of the parameters space, it is not easy to estimate all the parameters of Model I together. The same holds for the parameters of Model II and Model III. However, the specific shape of the Sarmanov distribution, which clearly splits into two parts -the marginal distributions and the dependence structure-, allows for the following approach:

- First, we estimate the parameters β_j, $j = 1,2,3$, associated with the expected values, i.e., with $\mu_{ij} = \exp(\mathbf{X}_i'\beta_j)$; the resulting estimations are denoted by $\hat{\beta} = \left(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3\right)$.
- Second, the parameters that define the variance and covariance matrix, i.e. α_1, α_2, α_3, ω_{12}, ω_{13}, ω_{13} and ω_{123} can also be estimated.

The parameters $\beta_j,\ j=1,2,3$ are estimated from the marginal distributions, i.e., we obtain $\hat{\beta}_j,\ j=1,2,3$ by maximizing the likelihood of the Poisson or NB GLM for each univariate marginal. If the NB model is the true one, both estimations are unbiased. From the ML estimations of the univariate NB GLM we also obtain the initial estimated values $\hat{\alpha}_1^0$, $\hat{\alpha}_2^0$, $\hat{\alpha}_3^0$, which are the starting values of the following iterative algorithm.

To estimate the dependence parameters, we define the following two conditional likelihoods: $L\left(\hat{\omega}|\hat{\alpha},\hat{\beta}\right)$ and $L\left(\hat{\alpha}|\hat{\omega},\hat{\beta}\right)$, where $\hat{\alpha}=(\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3)$ and $\hat{\omega}=(\hat{\omega}_{12},\hat{\omega}_{13},\hat{\omega}_{23},\hat{\omega}_{123})$ are two vectors with estimated parameters that allows us to obtain an estimation for the variance and covariance matrix.

The starting values for the dependence parameters, $\hat{\omega}^0=\left(\hat{\omega}_{12}^0,\hat{\omega}_{13}^0,\hat{\omega}_{23}^0,\hat{\omega}_{123}^0\right)$, are obtained by maximizing the first conditional likelihood given $\hat{\beta}$ and $\hat{\alpha}^0=\left(\hat{\alpha}_1^0,\hat{\alpha}_2^0,\hat{\alpha}_3^0\right)$. To this end, we need to define the parameters space (i.e., the current restrictions on the ωs). To do this, we determine the signs of the parameters in $\hat{\omega}^0$ and their intervals (note that this procedure is general for any iteration l in the following procedure). To find the signs, we use sample estimators based on (9) and (10). Taking these signs into account, we define variation intervals using Proposition 3. Starting with $l=0$, the rest of the procedure is divided in two steps:

- **Step 1** (iteration l) Within the parameter space obtained based on the estimated signs and intervals (cf. to the procedure described above), find $\hat{\omega}^l$ by maximizing the conditional likelihood $L\left(\hat{\omega}^l|\hat{\alpha}^l,\hat{\beta}\right)$.

- **Step 2** Obtain $\hat{\alpha}^{l+1}$ by maximizing the conditional likelihood $L\left(\hat{\alpha}^{l+1}|\hat{\omega}^l,\hat{\beta}\right)$.

If in Step 2 we have that $L\left(\hat{\alpha}^{l+1}|\hat{\omega}^l,\hat{\beta}\right)\leq L\left(\hat{\omega}^l|\hat{\alpha}^l,\hat{\beta}\right)$, we stop and consider the solution from the last iteration, Step 1; otherwise, we return to Step 1 for the next iteration.

Once we have estimated the parameters, we can calculate their standard errors using the approximate Hessian by Richardson's method implemented in R Software.

By way of alternative, in our algorithm, we also used an approach based on the EM algorithm proposed in Ghitany et al. (2012) to estimate new values of the β_js, $j=1,2,3$ in Step 2 at each iteration. However, the results are practically the same and computational time did not inprove.

4 Numerical Study

We fitted the models presented above to a data set from a multinational insurance company. The data come from the Spanish insurance market and consist of a random sample of $162,019$ policyholders who had had one or more auto and home policies during the decade 2006-2015. We used three dependent variables: the number of claims in auto insurance at fault involving only property damage (PD); the number of claims in auto insurance at fault with bodily injury (BI); and, the number of claims in home (H) insurance at fault. In Table 1 we show the claims frequency for each type of risk. For BI the maximum number of claims reported by a policyholder was six. For PD and H this maximum value reached 40 and 23, respectively.

In Table 2 we show four different statistics used to measure the dependence between the number of claims for each risk type. We also show the $p-$value associated with the significance test of each statistic. The statistics used were the following: chi$-$square (left upper triangle), to test for independence between discrete variables; the Pearson coefficient (left lower triangle), to test linear dependence; and, the Kendall and Spearman coefficients (right upper and lower triangles, respectively) to test non-linear dependence. To calculate the chi$-$square test statistic, we considered values of the number of claims from 0 to 6 or more and computed the $p-$values using the Monte Carlo method (see Hope, 1968).

From Table 2, note that all the statistics indicate that the different types of accident rate are dependent, with the exception of the Chi-square statistic for BI and H.

Table 1: Claims frequency.

Number of claims	0	1	2	3	4	5	≥ 6
Auto Property Damage	137437	15650	5247	1946	847	371	521
Auto Bodily Injury	156928	4586	440	50	10	2	3
Home	138694	17206	4125	1268	435	169	122

Table 2: Dependence analysis ($p-$value).

	Chi$-$Squared Statistics ($p-$value)				Kendall (p-value)		
	PD	BI	H		PD	BI	H
PD		42462 (0.0005)	64.731 (0.013)	PD		0.395 (0.000)	0.006 (0.010)
BI	0.444 (0.000)		25.261 (0.441)	BI	0.403 (0.000)		0.006 (0.009)
H	0.006 (0.003)	0.004 (0.049)		H	0.006 (0.010)	0.006 (0.009)	
	Pearson (p-value)				Spearman (p-value)		

The explanatory variables (covariates) used are listed in Table 3 with their respective means and variances. The values of these variables correspond to the latest available information for each policyholder. Although the models allow different covariates associated with each dependent variable to be used, here we opted for the same covariates for all three dependent variables, choosing them in relation to the policyholders' characteristics. Among the explanatory variables, we included "Gender" (note that while in the Spanish insurance market this variable cannot be included to calculate the insurance premium, it should be considered in the risk analysis). We also delimited three zones as areas of residence. The first zone consists of the big cities, which in Spain correspond to Barcelona and Madrid; the second corresponds to the north, given its specific weather; while the third corresponds to the rest of the country, and is defined as the reference zone.

Finally, we also included the age of the policyholder and the fact that the policyholder has contracted policies in other lines (e.g., accident insurance, life insurance, pension plans, etc.).

Table 3: Explanatory variables in the models (the values correspond to the latest available information for each policyholder).

Variable	**Description**	Mean	Variance
X_1	Gender of the policyholder: =1 if woman, =0 if man	0.237	0.181
X_2	Area of residence: =1 if big city, =0 if other	0.195	0.157
X_3	Area of residence: =1 if north, =0 if other	0.291	0.206
X_4	Age of policyholder	53.270	172.087
X_5	Client has other polices in the same company: =1 if yes, =0 if no	0.220	0.433

Table 4 presents the results for the estimated parameters of the trivariate NB GLM according to model (5), which includes the dependence between the numbers of claims in different types of insurance coverages. In Table 4 we have also included the estimated parameters obtained when fitting three independent univariate NB GLM distributions (i.e., the model with independent marginals). Then, in Table 5, we show the results for the estimated parameters of Models I, II and III, i.e., the three models based on the trivariate Sarmanov distribution. The Akaike information criterion (AIC) indicates that all the trivariate Sarmanov models improve the trivariate NB GLM model. Moreover, note that both Models II and III considerably improve Model I and, although the difference is small, Model II fits better than Model III.

Table 4: Estimation results of the trivariate Negative Binomial GLM assuming dependence (left) and independence (right).

	Dependent Marginal Distributions						Independent Marginal Distributions					
	Estimated Parameters			Standard Errors			Estimated Parameters			Standard Errors		
	PD	BI	H	PD	BI	H	PD	BI	H	PD	BI	H
Intercept	-1.2980	-3.0075	-1.7524	0.0320(***)	0.0711(***)	0.0343(***)	-1.3111	-3.0214	-1.7960	0.0334(***)	0.0714(***)	0.0354(***)
X_1	-0.2760	-0.1493	-0.113	0.0697(***)	0.1606	0.0668(**)	-0.2747	-0.1459	-0.0758	0.0723(***)	0.1602	0.0689
X_2	-0.0057	0.1031	-0.0301	0.0171	0.0368(***)	0.0174	-0.0040	0.0986	-0.0275	0.0183	0.0378(***)	0.0183(*)
X_3	-0.0211	-0.0797	-0.1138	0.0144(*)	0.0325(***)	0.0154(*)	-0.0183	-0.0888	-0.1056	0.0153	0.0332(***)	0.0161(***)
X_4	-0.0091	-0.0138	0.0017	0.0006(***)	0.0013(***)	0.0006(***)	-0.0093	-0.0130	0.0016	0.0006(***)	0.0013(***)	0.0006(***)
X_5	0.0291	-0.0917	0.0145	0.0077(***)	0.0183(***)	0.0087(**)	0.0243	-0.0687	0.0058	0.008(***)	0.0188(***)	0.0092
$X_1 \times X_4$	0.0011	-0.0014	0.0024	0.0013	0.0032	0.0012(**)	0.0010	-0.0015	0.0019	0.0014	0.003143	0.0013(*)
	$\alpha = 0.6110$									$\alpha_1 = 0.5060$	$\alpha_2 = 0.5066$	$\alpha_3 = 0.46754$
	AIC: 377654.3						AIC: 380184.5					
(***) significant at 1%, (**) significant at 5% and (*) significant at 10%.												

Table 5: Estimation results of the three models based on the Sarmanov distribution with NB GLM for marginals.

	Estimated Parameters			Standard Errors								
				Model I			Model II			Model III		
	$\hat{\beta}_1$ (PD)	$\hat{\beta}_2$ (BI)	$\hat{\beta}_3$ (H)	PD	BI	H	PD	BI	H	PD	BI	H
Intercept	-1.3111	-3.0214	-1.7960	0.0336 (***)	0.0716 (***)	0.0356 (***)	0.0360 (***)	0.0818 (***)	0.0358 (***)	0.0325 (***)	0.0698 (***)	0.0355 (***)
X_1	-0.2747	-0.1459	-0.0758	0.0723 (***)	0.1606	0.0693	0.0757 (***)	0.1782 (**)	0.0696	0.0704 (***)	0.1575	0.0691
X_2	-0.0040	0.0986	-0.0275	0.0180	0.0373 (***)	0.0183 (*)	0.0184	0.0428	0.0184 (*)	0.0174	0.0363 (***)	0.0182 (***)
X_3	-0.0183	-0.0888	-0.1056	0.0151	0.0328 (***)	0.0162 (***)	0.0150	0.0375 (***)	0.0162 (***)	0.0146	0.0319 (***)	0.0161(***)
X_4	-0.0093	-0.0130	0.0016	0.0006 (***)	0.0013 (***)	0.0006 (***)	0.0007 (***)	0.0015 (***)	0.0006 (***)	0.0006 (***)	0.0013 (***)	0.0006 (***)
X_5	0.0243	-0.0687	0.0058	0.0082 (***)	0.0183 (***)	0.0092	0.0088 (***)	0.0216 (***)	0.0093	0.0078 (***)	0.0178 (***)	0.0092
$X_1 \times X_4$	0.0010	-0.0015	0.0019	0.0014	0.0031	0.0013 (*)	0.0015	0.0035	0.0013 (*)	0.0014	0.0031	0.0013 (*)
				α_1=0.5122	α_2=0.4761	α_3=0.4770	α_1=0.3995	α_2=0.1185	α_3=0.4525	α_1=0.6277	α_2=0.9066	α_3=0.4780
				ω_{12}=1.9497, ω_{13}=0.4121, ω_{23}=0.9330 ω_{123}=0.6489			ω_{12}=4.9153, ω_{13}=1.4624, ω_{23}=0.0000 ω_{123}=-17.0000			ω_{12}=6.8222, ω_{13}=0.8181, ω_{23}=1.2438 ω_{123}=0.1863		
				AIC: 375230.0			AIC: 371987.4			AIC: 372415.6		
(***) significant at 1%, (**) significant at 5% and (*) significant at 10%.												

Note that, for the three Sarmanov models shown in Table 5, the values of the estimated parameters in the vectors $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ associated with the covariates are the same, since they are obtained by ML estimation of univariate marginal NB GLMs. Moreover, these estimated parameters are very similar to those in the trivariate NB GLM presented in Table 4. The differences between the models are given by the values of parameters associated with the variance and covariance matrix of $\mathbf{N}_i = (N_{i1}, N_{i2}, N_{i3})$, which changes affecting the standard errors of the parameters in $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$. The trivariate NB GLMs with dependent marginals is the model for which most of these standard errors have the lowest values; consequently, this estimated model tends not to reject the individual significance of the estimated parameters associated with the explanatory variables.

A further difference between the four multivariate models estimated involves the dependence assumed between the coverages analyzed. Focusing on the Sarmanov models, we note that the values of the dependence parameters ω_{jk} differ considerably between models, and that this happens because each model is associated with a different dependency. More precisely, Model I assumes dependence between NB variables, Model II between Gamma variables and Model III between Poisson variables. To compare the dependence structures of the models, we calculated the correlation coefficient of each individual according to formulas (6) (for the trivariate NB model) and (9) (for the Sarmanov models) and then we calculated the mean of these individual correlations (see Table 6). It can be seen that the correlations estimated for Model II are the ones most similar to those observed in the data (see Table 2).

Table 6: Correlations deduced from the four trivariate models estimated.

		Model I	
	PD	BI	H
PD		0.4562115	0.02477425
BI	0.1102308		0.02502437
H	0.2284269	0.09967611	
	Trivariate Negative Binomial		
		Model II	
	PD	BI	H
PD		0.5682647	0.02380325
BI	0.827436		0.01684441
H	0.02585861	0.02475868	
	Model III		

Even though Model II provides the best fit, all three models based on the Sarmanov distribution yield similar results with respect to the significance of the parameters $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$. These results indicate that the effect of the covariates depends on the type of coverage. For example, the effect of gender is negative and significant only in the case of property damage, i.e., women make less claims of this type. Living in big cities positively affects the number of bodily injury claims and negatively affects the number of home claims; however, living in the north of the country negatively affects both types of claims. The increasing age negatively affects the number of claims in the auto line and positively affects thats of claims in the home line. However, the parameter associated with the interaction of age and gender is significant at 10% only for the home line -the fact that this parameter is positive indicates that the effect of age is greater in the case of the women. Finally, the fact of having contracted more products with the same company only affects the auto line, where this effect is positive for property damage and negative for bodily injury.

5 Conclusions

In this paper, we have been able to identify the factors that affect each risk type by taking into account that the risks under analysis are dependent. To do this, we introduced three trivariate models with the same NB GLM marginals, but different dependence structures based on the Sarmanov distribution. Thus, where the first model (Model I) is simply a trivariate Sarmanov distribution with NB GLM marginals, the other two models where obtained by mixing three independent Poisson distributions with a Sarmanov distribution with Gamma distributed marginals (Model II) and a Sarmanov with Poisson marginals with independent Gamma distributions (Model III). These models were considered in connection with the number of claims made in three types of risks, two associated with the auto line (property damage and bodily insurance) and one associated with the home line.

Using a real data set from the Spanish insurance market, we also compared our proposed models with the trivariate NB GLM model and concluded that the two mixing models based on the Sarmanov distribution (Models II and III) improve the fit.

Moreover, we have proposed an algorithm for estimating the parameters in the Sarmanov based models. The expected number of claims estimated by each of the four models is practically the same. The main differences between the models are given by the values of the parameters associated with the dependence between the claims frequencies analyzed. These differences affect the risk quantification that depends on the correlation between the risk factors, and also the inference of the parameters associated with the covariates.

In conclusion, the mixing models based on the multivariate Sarmanov distribution add flexibility to the associated matrix of variances and covariances between dependent variables, resulting in a significant improvement in the fit compared to that obtained by simpler models including the multivariate NB GLM model and the multivariate discrete Sarmanov distribution with NB GLM marginals (Model I).

Acknowledgements

Catalina Bolancé acknowledges the Spanish Ministry of Education and the ERDF for grant ECO2016-76203-C2-2.

References

Abdallah, A., Boucher, J., Cossette, H., 2016. Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. Insurance: Mathematics and Economics 68, 120–133.

Bahraoui, Z., Bolancé, C., Pelican, E., Vernic, R., 2015. On the bivariate distribution and copula. an application on insurance data using truncated marginal distributions. Statistics and Operations Research Transactions, SORT 39, 209–230.

Bahraoui, Z., Bolancé, C., Pérez-Marín, A., 2014. Testing extreme value copulas to estimate the quantile. Statistics and Operations Research Transactions, SORT 38, 89–102.

Bermudez, L., Karlis, D., 2011. Bayesian multivariate poisson models for insurance ratemaking. Insurance: Mathematics and Economics 48, 226–236.

Bolancé, C., Bahraoui, Z., Artís, M., 2014. Quantifying the risk using copulae with nonparametric marginal. Insurance: Mathematics and Economics 58, 46–56.

Bolancé, C., Guillén, M., Pinquet, J., 2003. Time-varying credibility for frequency risk models. Insurance: Mathematics and Economics 33, 273–282.

Bolancé, C., Guillén, M., Pinquet, J., 2008. On the link between credibility and frequency premium. Insurance: Mathematics and Economics 43, 209–213.

Bolancé, C., Guillén, M., Pelican, E., Vernic, R., 2008. Skewed bivariate models and nonparametric estimation for cte risk measure. Insurance: Mathematics and Economics 43, 386–393.

Boucher, J.P., Denuit, M., Guillen, M., 2007. Risk classification for claim counts: A comparative analysis of various zero-inflated mixed Poisson and hurdle models. North American Actuarial Journal 11, 110–131.

Boucher, J.P., Inoussa, R., 2014. A posteriori ratemaking with panel data. ASTIN Bulletin 44, 587–612.

Brockett, P.L., Golden, L., Guillén, M., Nielsen, J., Parner, J., Pérez-Marín, A., 2008. Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection? Journal of Risk and Insurance 75, 713–737.

Frees, E., 2009. Regression Modelling with Actuarial and Financial Applications. Cambridge University Press.

Ghitany, M., Karlis, D., D.K., A.M., Al-Awadhi, F., 2012. An EM algorithm for multivariate mixed Poisson regression models and its application. Applied Mathematical Sciences 6, 6843–6856.

Guelman, L., Guillén, M., 2014. A causal inference approach to measure price elasticity in automobile insurance. Expert Systems with Applications 41, 387–396.

Guelman, L., Guillén, M., Pérez-Marín, A., 2014. A survey of personalized treatment models for pricing strategies in insurance. Insurance: Mathematics and Economics 58, 68–76.

Guillén, M., Nielsen, J., Scheike, T., Pérez-Marín, A., 2012. Time-varying effects in the analysis of customer loyalty: A case study in insurance. Expert Systems with Applications 39, 3551–3558.

Hope, A.C.A., 1968. A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society. Series B 30, 582–598.

Johnson, N., Kotz, S., Balakrishnan, N., 1997. Discrete multivariate distributions. Wiley.

Kotz, S., Balakrishnan, N., Johnson, N., 2000. Continuous Multivariate Distributions. Vol.1: Models and Applications. Wiley.

McCullagh, P., Nelder, J.A., 1989. Generalized linear models. Vol.37. CRC Press.

Pinquet, J., Guillén, M., Bolancé, C., 2001. Long-range contagion in automobile insurance data: estimation and implications for experience rating. ASTIN Bulletin 31, 337–348.

Shi, P., Valdez, E., 2014. Multivariate negative binomial models for insurance claim counts. Insurance: Mathematics and Economics 55, 18–29.

2006

CREAP2006-01
Matas, A. (GEAP)**; Raymond, J.Ll.** (GEAP)
"Economic development and changes in car ownership patterns"
(Juny 2006)

CREAP2006-02
Trillas, F. (IEB)**; Montolio, D.** (IEB)**; Duch, N.** (IEB)
"Productive efficiency and regulatory reform: The case of Vehicle Inspection Services"
(Setembre 2006)

CREAP2006-03
Bel, G. (PPRE-IREA)**; Fageda, X.** (PPRE-IREA)
"Factors explaining local privatization: A meta-regression analysis"
(Octubre 2006)

CREAP2006-04
Fernàndez-Villadangos, L. (PPRE-IREA)
"Are two-part tariffs efficient when consumers plan ahead?: An empirical study"
(Octubre 2006)

CREAP2006-05
Artís, M. (AQR-IREA)**; Ramos, R.** (AQR-IREA)**; Suriñach, J.** (AQR-IREA)
"Job losses, outsourcing and relocation: Empirical evidence using microdata"
(Octubre 2006)

CREAP2006-06
Alcañiz, M. (RISC-IREA)**; Costa, A.; Guillén, M.** (RISC-IREA)**; Luna, C.; Rovira, C.**
"Calculation of the variance in surveys of the economic climate"
(Novembre 2006)

CREAP2006-07
Albalate, D. (PPRE-IREA)
"Lowering blood alcohol content levels to save lives: The European Experience"
(Desembre 2006)

CREAP2006-08
Garrido, A. (IEB)**; Arqué, P.** (IEB)
"The choice of banking firm: Are the interest rate a significant criteria?"
(Desembre 2006)

CREAP2006-09
Segarra, A. (GRIT)**; Teruel-Carrizosa, M.** (GRIT)
"Productivity growth and competition in spanish manufacturing firms:
What has happened in recent years?"
(Desembre 2006)

CREAP2006-10
Andonova, V.; Díaz-Serrano, Luis. (CREB)
"Political institutions and the development of telecommunications"
(Desembre 2006)

CREAP2006-11
Raymond, J.L.(GEAP)**; Roig, J.L..** (GEAP)
"Capital humano: un análisis comparativo Catalunya-España"
(Desembre 2006)

CREAP2006-12
Rodríguez, M.(CREB)**; Stoyanova, A.** (CREB)
"Changes in the demand for private medical insurance following a shift in tax incentives"
(Desembre 2006)

CREAP2006-13
Royuela, V. (AQR-IREA)**; Lambiri, D.**; **Biagi, B.**
"Economía urbana y calidad de vida. Una revisión del estado del conocimiento en España"
(Desembre 2006)

CREAP2006-14

Camarero, M.; Carrion-i-Silvestre, J.LL. (AQR-IREA)**.**;**Tamarit**, **C.**
"New evidence of the real interest rate parity for OECD countries using panel unit root tests with breaks"
(Desembre 2006)

CREAP2006-15
Karanassou, M.; Sala, H. (GEAP)**.**;**Snower** , **D. J.**
"The macroeconomics of the labor market: Three fundamental views"
(Desembre 2006)

2007

XREAP2007-01
Castany, L (AQR-IREA)**; López-Bazo, E.** (AQR-IREA)**.**;**Moreno** , **R.** (AQR-IREA)
"Decomposing differences in total factor productivity across firm size"
(Març 2007)

XREAP2007-02
Raymond, J. Ll. (GEAP)**; Roig, J. Ll.** (GEAP)
"Una propuesta de evaluación de las externalidades de capital humano en la empresa"
(Abril 2007)

XREAP2007-03
Durán, J. M. (IEB)**; Esteller, A.** (IEB)
"An empirical analysis of wealth taxation: Equity vs. Tax compliance"
(Juny 2007)

XREAP2007-04
Matas, A. (GEAP)**; Raymond, J.Ll.** (GEAP)
"Cross-section data, disequilibrium situations and estimated coefficients: evidence from car ownership demand"
(Juny 2007)

XREAP2007-05
Jofre-Montseny, J. (IEB)**; Solé-Ollé, A.** (IEB)
"Tax differentials and agglomeration economies in intraregional firm location"
(Juny 2007)

XREAP2007-06
Álvarez-Albelo, C. (CREB)**; Hernández-Martín, R.**
"Explaining high economic growth in small tourism countries with a dynamic general equilibrium model"
(Juliol 2007)

XREAP2007-07
Duch, N. (IEB)**; Montolio, D.** (IEB); **Mediavilla, M.**
"Evaluating the impact of public subsidies on a firm's performance: a quasi-experimental approach"
(Juliol 2007)

XREAP2007-08
Segarra-Blasco, A. (GRIT)
"Innovation sources and productivity: a quantile regression analysis"
(Octubre 2007)

XREAP2007-09
Albalate, D. (PPRE-IREA)
"Shifting death to their Alternatives: The case of Toll Motorways"
(Octubre 2007)

XREAP2007-10
Segarra-Blasco, A. (GRIT); **Garcia-Quevedo, J.** (IEB); **Teruel-Carrizosa, M.** (GRIT)
"Barriers to innovation and public policy in catalonia"
(Novembre 2007)

XREAP2007-11
Bel, G. (PPRE-IREA); **Foote, J.**
"Comparison of recent toll road concession transactions in the United States and France"
(Novembre 2007)

XREAP2007-12
Segarra-Blasco, A. (GRIT);
"Innovation, R&D spillovers and productivity: the role of knowledge-intensive services"
(Novembre 2007)

XREAP2007-13
Bermúdez Morata, Ll. (RFA-IREA); **Guillén Estany, M.** (RFA-IREA), **Solé Auró, A**. (RFA-IREA)
"Impacto de la inmigración sobre la esperanza de vida en salud y en discapacidad de la población española"
(Novembre 2007)

XREAP2007-14
Calaeys, P. (AQR-IREA); **Ramos, R.** (AQR-IREA), **Suriñach, J**. (AQR-IREA)
"Fiscal sustainability across government tiers"
(Desembre 2007)

XREAP2007-15
Sánchez Hugalbe, A. (IEB)
"Influencia de la inmigración en la elección escolar"
(Desembre 2007)

2008

XREAP2008-01
Durán Weitkamp, C. (GRIT)**; Martín Bofarull, M.** (GRIT) ; **Pablo Martí, F.**
"Economic effects of road accessibility in the Pyrenees: User perspective"
(Gener 2008)

XREAP2008-02
Díaz-Serrano, L.; Stoyanova, A. P. (CREB)
"The Causal Relationship between Individual's Choice Behavior and Self-Reported Satisfaction: the Case of Residential Mobility in the EU"
(Març 2008)

XREAP2008-03
Matas, A. (GEAP)**; Raymond, J. L.** (GEAP)**; Roig, J. L.** (GEAP)
"Car ownership and access to jobs in Spain"
(Abril 2008)

XREAP2008-04
Bel, G. (PPRE-IREA) **; Fageda, X.** (PPRE-IREA)
"Privatization and competition in the delivery of local services: An empirical examination of the dual market hypothesis"
(Abril 2008)

XREAP2008-05
Matas, A. (GEAP); **Raymond, J. L.** (GEAP); **Roig, J. L.** (GEAP)
"**Job accessibility and employment probability**"
(Maig 2008)

XREAP2008-06
Basher, S. A.; Carrión, J. Ll. (AQR-IREA)
Deconstructing Shocks and Persistence in OECD Real Exchange Rates
(Juny 2008)

XREAP2008-07
Sanromá, E. (IEB)**; Ramos, R.** (AQR-IREA); Simón, H.
Portabilidad del capital humano y asimilación de los inmigrantes. Evidencia para España
(Juliol 2008)

XREAP2008-08
Basher, S. A.; Carrión, J. Ll. (AQR-IREA)
Price level convergence, purchasing power parity and multiple structural breaks: An application to US cities
(Juliol 2008)

XREAP2008-09
Bermúdez, Ll. (RFA-IREA)
A priori ratemaking using bivariate poisson regression models
(Juliol 2008)

XREAP2008-10
Solé-Ollé, A. (IEB), **Hortas Rico, M.** (IEB)
Does urban sprawl increase the costs of providing local public services? Evidence from Spanish municipalities
(Novembre 2008)

XREAP2008-11
Teruel-Carrizosa, M. (GRIT), **Segarra-Blasco, A.** (GRIT)
Immigration and Firm Growth: Evidence from Spanish cities
(Novembre 2008)

XREAP2008-12
Duch-Brown, N. (IEB), **García-Quevedo, J.** (IEB), **Montolio, D.** (IEB)
Assessing the assignation of public subsidies: Do the experts choose the most efficient R&D projects?
(Novembre 2008)

XREAP2008-13
Bilotkach, V., **Fageda, X.** (PPRE-IREA), **Flores-Fillol, R.**
Scheduled service versus personal transportation: the role of distance
(Desembre 2008)

XREAP2008-14
Albalate, D. (PPRE-IREA), **Gel, G.** (PPRE-IREA)
Tourism and urban transport: Holding demand pressure under supply constraints
(Desembre 2008)

2009

XREAP2009-01
Calonge, S. (CREB)**; Tejada, O.**
"A theoretical and practical study on linear reforms of dual taxes"
(Febrer 2009)

XREAP2009-02
Albalate, D. (PPRE-IREA)**; Fernández-Villadangos, L.** (PPRE-IREA)
"Exploring Determinants of Urban Motorcycle Accident Severity: The Case of Barcelona"
(Març 2009)

XREAP2009-03
Borrell, J. R. (PPRE-IREA)**; Fernández-Villadangos, L.** (PPRE-IREA)
"Assessing excess profits from different entry regulations"
(Abril 2009)

XREAP2009-04
Sanromá, E. (IEB)**; Ramos, R.** (AQR-IREA), **Simon, H.**
"Los salarios de los inmigrantes en el mercado de trabajo español. ¿Importa el origen del capital humano?"
(Abril 2009)

XREAP2009-05
Jiménez, J. L.; Perdiguero, J. (PPRE-IREA)
"(No)competition in the Spanish retailing gasoline market: a variance filter approach"
(Maig 2009)

XREAP2009-06
Álvarez-Albelo,C. D. (CREB)**, Manresa, A.** (CREB)**, Pigem-Vigo, M.** (CREB)
"International trade as the sole engine of growth for an economy"
(Juny 2009)

XREAP2009-07
Callejón, M. (PPRE-IREA)**, Ortún V, M.**
"The Black Box of Business Dynamics"
(Setembre 2009)

XREAP2009-08
Lucena, A. (CREB)
"The antecedents and innovation consequences of organizational search: empirical evidence for Spain"
(Octubre 2009)

XREAP2009-09
Domènech Campmajó, L. (PPRE-IREA)
"Competition between TV Platforms"
(Octubre 2009)

XREAP2009-10
Solé-Auró, A. (RFA-IREA),**Guillén, M.** (RFA-IREA)**, Crimmins, E. M.**
"Health care utilization among immigrants and native-born populations in 11 European countries. Results from the Survey of Health, Ageing and Retirement in Europe"
(Octubre 2009)

XREAP2009-11
Segarra, A. (GRIT)**, Teruel, M.** (GRIT)
"Small firms, growth and financial constraints"
(Octubre 2009)

XREAP2009-12
Matas, A. (GEAP)**, Raymond, J.Ll.** (GEAP), **Ruiz, A.** (GEAP)
"Traffic forecasts under uncertainty and capacity constraints"
(Novembre 2009)

XREAP2009-13
Sole-Ollé, A. (IEB)
"Inter-regional redistribution through infrastructure investment: tactical or programmatic?"
(Novembre 2009)

XREAP2009-14
Del Barrio-Castro, T., **García-Quevedo, J.** (IEB)
"The determinants of university patenting: Do incentives matter?"
(Novembre 2009)

XREAP2009-15
Ramos, R. (AQR-IREA), **Suriñach, J.** (AQR-IREA)**, Artís, M.** (AQR-IREA)
"Human capital spillovers, productivity and regional convergence in Spain"
(Novembre 2009)

XREAP2009-16
Álvarez-Albelo, C. D. (CREB), **Hernández-Martín, R.**
"The commons and anti-commons problems in the tourism economy"
(Desembre 2009)

2010

XREAP2010-01
García-López, M. A. (GEAP)
"The Accessibility City. When Transport Infrastructure Matters in Urban Spatial Structure"
(Febrer 2010)

XREAP2010-02
García-Quevedo, J. (IEB), **Mas-Verdú, F.** (IEB), **Polo-Otero, J.** (IEB)
"Which firms want PhDs? The effect of the university-industry relationship on the PhD labour market"
(Març 2010)

XREAP2010-03
Pitt, D., **Guillén, M.** (RFA-IREA)
"An introduction to parametric and non-parametric models for bivariate positive insurance claim severity distributions"
(Març 2010)

XREAP2010-04
Bermúdez, Ll. (RFA-IREA), **Karlis, D.**
"Modelling dependence in a ratemaking procedure with multivariate Poisson regression models"
(Abril 2010)

XREAP2010-05
Di Paolo, A. (IEB)
"Parental education and family characteristics: educational opportunities across cohorts in Italy and Spain"
(Maig 2010)

XREAP2010-06
Simón, H. (IEB), **Ramos, R.** (AQR-IREA), **Sanromá, E.** (IEB)

"Movilidad ocupacional de los inmigrantes en una economía de bajas cualificaciones. El caso de España"
(Juny 2010)

XREAP2010-07
Di Paolo, A. (GEAP & IEB), **Raymond, J. Ll.** (GEAP & IEB)
"Language knowledge and earnings in Catalonia"
(Juliol 2010)

XREAP2010-08
Bolancé, C. (RFA-IREA), **Alemany, R.** (RFA-IREA), **Guillén, M.** (RFA-IREA)
"Prediction of the economic cost of individual long-term care in the Spanish population"
(Setembre 2010)

XREAP2010-09
Di Paolo, A. (GEAP & IEB)
"Knowledge of catalan, public/private sector choice and earnings: Evidence from a double sample selection model"
(Setembre 2010)

XREAP2010-10
Coad, A., **Segarra, A.** (GRIT), **Teruel, M.** (GRIT)
"Like milk or wine: Does firm performance improve with age?"
(Setembre 2010)

XREAP2010-11
Di Paolo, A. (GEAP & IEB), **Raymond, J. Ll.** (GEAP & IEB), **Calero, J.** (IEB)
"Exploring educational mobility in Europe"
(Octubre 2010)

XREAP2010-12
Borrell, A. (GiM-IREA), **Fernández-Villadangos, L.** (GiM-IREA)
"Clustering or scattering: the underlying reason for regulating distance among retail outlets"
(Desembre 2010)

XREAP2010-13
Di Paolo, A. (GEAP & IEB)
"School composition effects in Spain"
(Desembre 2010)

XREAP2010-14
Fageda, X. (GiM-IREA), **Flores-Fillol, R.**
"Technology, Business Models and Network Structure in the Airline Industry"
(Desembre 2010)

XREAP2010-15
Albalate, D. (GiM-IREA), **Bel, G.** (GiM-IREA), **Fageda, X.** (GiM-IREA)
"Is it Redistribution or Centralization? On the Determinants of Government Investment in Infrastructure"
(Desembre 2010)

XREAP2010-16
Oppedisano, V., **Turati, G.**
"What are the causes of educational inequalities and of their evolution over time in Europe? Evidence from PISA"
(Desembre 2010)

XREAP2010-17
Canova, L., **Vaglio, A.**
"Why do educated mothers matter? A model of parental help"
(Desembre 2010)

2011

XREAP2011-01
Fageda, X. (GiM-IREA), **Perdiguero, J.** (GiM-IREA)
"An empirical analysis of a merger between a network and low-cost airlines"
(Maig 2011)

XREAP2011-02
Moreno-Torres, I. (ACCO, CRES & GiM-IREA)
"What if there was a stronger pharmaceutical price competition in Spain? When regulation has a similar effect to collusion"
(Maig 2011)

XREAP2011-03
Miguélez, E. (AQR-IREA); **Gómez-Miguélez, I.**
"Singling out individual inventors from patent data"
(Maig 2011)

XREAP2011-04
Moreno-Torres, I. (ACCO, CRES & GiM-IREA)
"Generic drugs in Spain: price competition vs. moral hazard"
(Maig 2011)

XREAP2011-05
Nieto, S. (AQR-IREA)**, Ramos, R.** (AQR-IREA)
"¿Afecta la sobreeducación de los padres al rendimiento académico de sus hijos?"
(Maig 2011)

XREAP2011-06
Pitt, D., **Guillén, M.** (RFA-IREA), **Bolancé, C.** (RFA-IREA)
"Estimation of Parametric and Nonparametric Models for Univariate Claim Severity Distributions - an approach using R"
(Juny 2011)

XREAP2011-07
Guillén, M. (RFA-IREA), **Comas-Herrera, A.**
"How much risk is mitigated by LTC Insurance? A case study of the public system in Spain"
(Juny 2011)

XREAP2011-08
Ayuso, M. (RFA-IREA), **Guillén, M.** (RFA-IREA), **Bolancé, C.** (RFA-IREA)
"Loss risk through fraud in car insurance"
(Juny 2011)

XREAP2011-09
Duch-Brown, N. (IEB), **García-Quevedo, J.** (IEB), **Montolio, D.** (IEB)
"The link between public support and private R&D effort: What is the optimal subsidy?"
(Juny 2011)

XREAP2011-10
Bermúdez, Ll. (RFA-IREA), **Karlis, D.**
"Mixture of bivariate Poisson regression models with an application to insurance"
(Juliol 2011)

XREAP2011-11
Varela-Irimia, X-L. (GRIT)
"Age effects, unobserved characteristics and hedonic price indexes: The Spanish car market in the 1990s"
(Agost 2011)

XREAP2011-12
Bermúdez, Ll. (RFA-IREA), **Ferri, A.** (RFA-IREA), **Guillén, M.** (RFA-IREA)
"A correlation sensitivity analysis of non-life underwriting risk in solvency capital requirement estimation"
(Setembre 2011)

XREAP2011-13
Guillén, M. (RFA-IREA), **Pérez-Marín, A.** (RFA-IREA), **Alcañiz, M.** (RFA-IREA)
"A logistic regression approach to estimating customer profit loss due to lapses in insurance"
(Octubre 2011)

XREAP2011-14
Jiménez, J. L., **Perdiguero, J.** (GiM-IREA), **García, C.**
"Evaluation of subsidies programs to sell green cars: Impact on prices, quantities and efficiency"
(Octubre 2011)

XREAP2011-15
Arespa, M. (CREB)
"A New Open Economy Macroeconomic Model with Endogenous Portfolio Diversification and Firms Entry"
(Octubre 2011)

XREAP2011-16
Matas, A. (GEAP), **Raymond, J. L.** (GEAP), **Roig, J.L.** (GEAP)
"The impact of agglomeration effects and accessibility on wages"
(Novembre 2011)

XREAP2011-17
Segarra, A. (GRIT)
"R&D cooperation between Spanish firms and scientific partners: what is the role of tertiary education?"
(Novembre 2011)

XREAP2011-18
García-Pérez, J. I.; **Hidalgo-Hidalgo, M.**; **Robles-Zurita, J. A.**
"Does grade retention affect achievement? Some evidence from PISA"
(Novembre 2011)

XREAP2011-19
Arespa, M. (CREB)
"Macroeconomics of extensive margins: a simple model"
(Novembre 2011)

XREAP2011-20
García-Quevedo, J. (IEB), **Pellegrino, G.** (IEB), **Vivarelli, M.**
"The determinants of YICs' R&D activity"
(Desembre 2011)

XREAP2011-21
González-Val, R. (IEB), **Olmo, J.**
"Growth in a Cross-Section of Cities: Location, Increasing Returns or Random Growth?"
(Desembre 2011)

XREAP2011-22
Gombau, V. (GRIT), **Segarra, A.** (GRIT)
"The Innovation and Imitation Dichotomy in Spanish firms: do absorptive capacity and the technological frontier matter?"
(Desembre 2011)

2012

XREAP2012-01
Borrell, J. R. (GiM-IREA), **Jiménez, J. L.**, **García, C.**
"Evaluating Antitrust Leniency Programs"
(Gener 2012)

XREAP2012-02
Ferri, A. (RFA-IREA), **Guillén, M.** (RFA-IREA), **Bermúdez, Ll.** (RFA-IREA)
"Solvency capital estimation and risk measures"
(Gener 2012)

XREAP2012-03
Ferri, A. (RFA-IREA), **Bermúdez, Ll.** (RFA-IREA), **Guillén, M.** (RFA-IREA)
"How to use the standard model with own data"
(Febrer 2012)

XREAP2012-04
Perdiguero, J. (GiM-IREA), **Borrell, J.R.** (GiM-IREA)
"Driving competition in local gasoline markets"
(Març 2012)

XREAP2012-05
D'Amico, G., **Guillen, M.** (RFA-IREA), Manca, R.
"Discrete time Non-homogeneous Semi-Markov Processes applied to Models for Disability Insurance"
(Març 2012)

XREAP2012-06
Bové-Sans, M. A. (GRIT), Laguado-Ramírez, R.
"Quantitative analysis of image factors in a cultural heritage tourist destination"
(Abril 2012)

XREAP2012-07
Tello, C. (AQR-IREA), **Ramos, R.** (AQR-IREA), **Artís, M.** (AQR-IREA)
"Changes in wage structure in Mexico going beyond the mean: An analysis of differences in distribution, 1987-2008"
(Maig 2012)

XREAP2012-08
Jofre-Monseny, J. (IEB), **Marín-López, R.** (IEB), **Viladecans-Marsal, E.** (IEB)
"What underlies localization and urbanization economies? Evidence from the location of new firms"
(Maig 2012)

XREAP2012-09
Muñiz, I. (GEAP), **Calatayud, D.**, **Dobaño, R.**
"Los límites de la compacidad urbana como instrumento a favor de la sostenibilidad. La hipótesis de la compensación en Barcelona medida a través de la huella ecológica de la movilidad y la vivienda"
(Maig 2012)

XREAP2012-10
Arqué-Castells, P. (GEAP), **Mohnen, P.**
"Sunk costs, extensive R&D subsidies and permanent inducement effects"
(Maig 2012)

XREAP2012-11
Boj, E. (CREB), **Delicado, P.**, **Fortiana, J**., **Esteve, A.**, **Caballé, A.**
"Local Distance-Based Generalized Linear Models using the dbstats package for R"
(Maig 2012)

XREAP2012-12
Royuela, V. (AQR-IREA)
"What about people in European Regional Science?"
(Maig 2012)

XREAP2012-13
Osorio A. M. (RFA-IREA)**, Bolancé, C.** (RFA-IREA), **Madise, N.**
"Intermediary and structural determinants of early childhood health in Colombia: exploring the role of communities"
(Juny 2012)

XREAP2012-14
Miguelez. E. (AQR-IREA)**, Moreno, R.** (AQR-IREA)
"Do labour mobility and networks foster geographical knowledge diffusion? The case of European regions"
(Juliol 2012)

XREAP2012-15
Teixidó-Figueras, J. (GRIT)**, Duró, J. A.** (GRIT)
"Ecological Footprint Inequality: A methodological review and some results"
(Setembre 2012)

XREAP2012-16
Varela-Irimia, X-L. (GRIT)
"Profitability, uncertainty and multi-product firm product proliferation: The Spanish car industry"
(Setembre 2012)

XREAP2012-17
Duró, J. A. (GRIT), **Teixidó-Figueras, J.** (GRIT)
"Ecological Footprint Inequality across countries: the role of environment intensity, income and interaction effects"
(Octubre 2012)

XREAP2012-18
Manresa, A. (CREB), **Sancho, F.**
"Leontief versus Ghosh: two faces of the same coin"
(Octubre 2012)

XREAP2012-19
Alemany, R. (RFA-IREA), **Bolancé, C.** (RFA-IREA), **Guillén, M.** (RFA-IREA)
"Nonparametric estimation of Value-at-Risk"
(Octubre 2012)

XREAP2012-20
Herrera-Idárraga, P. (AQR-IREA), **López-Bazo, E.** (AQR-IREA), **Motellón, E.** (AQR-IREA)
"Informality and overeducation in the labor market of a developing country"
(Novembre 2012)

XREAP2012-21
Di Paolo, A. (AQR-IREA)
"(Endogenous) occupational choices and job satisfaction among recent PhD recipients: evidence from Catalonia"
(Desembre 2012)

2013

XREAP2013-01
Segarra, A. (GRIT), **García-Quevedo, J.** (IEB), **Teruel, M.** (GRIT)
"Financial constraints and the failure of innovation projects"
(Març 2013)

XREAP2013-02
Osorio, A. M. (RFA-IREA), **Bolancé, C.** (RFA-IREA), Madise, N., Rathmann, K.
"Social Determinants of Child Health in Colombia: Can Community Education Moderate the Effect of Family Characteristics?"
(Març 2013)

XREAP2013-03
Teixidó-Figueras, J. (GRIT), **Duró, J. A.** (GRIT)
"The building blocks of international ecological footprint inequality: a regression-based decomposition"
(Abril 2013)

XREAP2013-04
Salcedo-Sanz, S., **Carro-Calvo, L.**, **Claramunt, M.** (CREB), **Castañer, A.** (CREB), **Marmol, M.** (CREB)
"An Analysis of Black-box Optimization Problems in Reinsurance: Evolutionary-based Approaches"
(Maig 2013)

XREAP2013-05
Alcañiz, M. (RFA), **Guillén, M.** (RFA), **Sánchez-Moscona, D.** (RFA), **Santolino, M.** (RFA), **Llatje, O., Ramon, Ll.**
"Prevalence of alcohol-impaired drivers based on random breath tests in a roadside survey"
(Juliol 2013)

XREAP2013-06
Matas, A. (GEAP & IEB), **Raymond, J. Ll.** (GEAP & IEB), **Roig, J. L.** (GEAP)
"How market access shapes human capital investment in a peripheral country"
(Octubre 2013)

XREAP2013-07
Di Paolo, A. (AQR-IREA), **Tansel, A.**
"Returns to Foreign Language Skills in a Developing Country: The Case of Turkey"
(Novembre 2013)

XREAP2013-08
Fernández Gual, V. (GRIT), **Segarra, A.** (GRIT)
"The Impact of Cooperation on R&D, Innovation andProductivity: an Analysis of Spanish Manufacturing and Services Firms"
(Novembre 2013)

XREAP2013-09
Bahraoui, Z. (RFA); **Bolancé, C.** (RFA); **Pérez-Marín. A. M.** (RFA)
"Testing extreme value copulas to estimate the quantile"
(Novembre 2013)

2014

XREAP2014-01
Solé-Auró, A. (RFA), **Alcañiz, M.** (RFA)
"Are we living longer but less healthy? Trends in mortality and morbidity in Catalonia (Spain), 1994-2011"
(Gener 2014)

XREAP2014-02

Teixidó-Figueres, J. (GRIT), **Duro, J. A.** (GRIT)
"Spatial Polarization of the Ecological Footprint distribution"
(Febrer 2014)

XREAP2014-03
Cristobal-Cebolla, A.; Gil Lafuente, A. M. (RFA), **Merigó Lindhal, J. M.** (RFA)
"La importancia del control de los costes de la no-calidad en la empresa"
(Febrer 2014)

XREAP2014-04
Castañer, A. (CREB); **Claramunt, M.M.** (CREB)
"Optimal stop-loss reinsurance: a dependence analysis"
(Abril 2014)

XREAP2014-05
Di Paolo, A. (AQR-IREA); **Matas, A.** (GEAP); **Raymond, J. Ll.** (GEAP)
"Job accessibility, employment and job-education mismatch in the metropolitan area of Barcelona"
(Maig 2014)

XREAP2014-06
Di Paolo, A. (AQR-IREA); **Mañé, F.**
"Are we wasting our talent? Overqualification and overskilling among PhD graduates"
(Juny 2014)

XREAP2014-07
Segarra, A. (GRIT); **Teruel, M.** (GRIT); **Bové, M. A.** (GRIT)
"A territorial approach to R&D subsidies: Empirical evidence for Catalonian firms"
(Setembre 2014)

XREAP2014-08
Ramos, R. (AQR-IREA); **Sanromá, E.** (IEB); **Simón, H.**
"Public-private sector wage differentials by type of contract: evidence from Spain"
(Octubre 2014)

XREAP2014-09
Bel, G. (GiM-IREA); **Bolancé, C.** (Riskcenter-IREA); **Guillén, M.** (Riskcenter-IREA)**; Rosell, J.** (GiM-IREA)
"The environmental effects of changing speed limits: a quantile regression approach"
(Desembre 2014)

2015

XREAP2015-01
Bolance, C. (Riskcenter-IREA); **Bahraoui, Z.** (Riskcenter-IREA), **Alemany, R.** (Risckcenter-IREA)
"Estimating extreme value cumulative distribution functions using bias-corrected kernel approaches"
(Gener 2015)

XREAP2015-02
Ramos, R. (AQR-IREA); **Sanromá, E.** (IEB), **Simón, H.**
"An analysis of wage differentials between full- and part-time workers in Spain"
(Agost 2015)

XREAP2015-03
Cappellari, L.; **Di Paolo, A.** (AQR-IREA)
"Bilingual Schooling and Earnings: Evidence from a Language-in-Education Reform"
(Setembre 2015)

XREAP2015-04
Álvarez-Albelo, C. D., Manresa, A. (CREB)**, Pigem-Vigo, M.** (CREB)
"Growing through trade: The role of foreign growth and domestic tariffs"
(Novembre 2015)

XREAP2015-05
Caminal, R., Di Paolo, A. (AQR-IREA)
Your language or mine?
(Novembre 2015)

XREAP2015-06
Choi, H. (AQR-IREA)**, Choi, A.** (IEB)
When one door closes: the impact of the hagwon curfew on the consumption of private tutoring in the Republic of Korea

(Novembre 2015)

2016

XREAP2016-01
Castañer, A. (CREB, XREAP); **Claramunt, M M.** (CREB, XREAP), **Tadeo, A.**, **Varea, J.** (CREB, XREAP)
Modelización de la dependencia del número de siniestros. Aplicación a Solvencia II
(Setembre 2016)

XREAP2016-02
García-Quevedo, J. (IEB, XREAP); **Segarra-Blasco, A.** (GRIT, XREAP), **Teruel, M.** (GRIT, XREAP)
Financial constraints and the failure of innovation projects
(Setembre 2016)

XREAP2016-03
Jové-Llopis, E. (GRIT, XREAP); **Segarra-Blasco, A.** (GRIT, XREAP)
What is the role of innovation strategies? Evidence from Spanish firms
(Setembre 2016)

XREAP2016-04
Albalate, D. (GiM-IREA, XREAP); **Rosell, J.** (GiM-IREA, XREAP)
Persistent and transient efficiency on the stochastic production and cost frontiers – an application to the motorway sector
(Octubre 2016)

XREAP2016-05
Jofre-Monseny, J. (IEB, XREAP), **Silva, J. I.**, **Vázquez-Grenno, J.** (IEB, XREAP)
Local labor market effects of public employment
(Novembre 2016)

XREAP2016-06
Garcia-López, M. A. (IEB, XREAP), **Hemet, C.**, **Viladecans-Marsal, E.** (IEB, XREAP)
Next train to the polycentric city: The effect of railroads on subcenter formation
(Novembre 2016)

XREAP2016-07
Vayá, E. (AQR-IREA, XREAP)**, García, J. R.** (AQR-IREA, XREAP)**, Murillo, J.** (AQR-IREA, XREAP)**, Romaní, J.** (AQR-IREA, XREAP)**, Suriñach, J.** (AQR-IREA, XREAP)**,**
Economic impact of cruise activity: the port of Barcelona
(Desembre 2016)

XREAP2016-08
Ayuso, M. (Riskcenter, XREAP)**, Guillen, M.** (Riskcenter, XREAP)**, Nielsen, J. P.**
Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data
(Desembre 2016)

XREAP2016-09
Ruíz, A. (GEAP, XREAP), **Matas, A.** (GEAP, XREAP), **Raymond, J. Ll.**
How do road infrastructure investments affect the regional economy? Evidence from Spain
(Desembre 2016)

2017

XREAP2017-01
Bernardo, V. (GiM-IREA, XREAP); **Fageda, X.** (GiM-IREA, XREAP)
Globalization, long-haul flights and inter-city connections
(Octubre 2017)

XREAP2017-02
Di Paolo, A. (AQR-IREA, XREAP); **Tansel, A.**
Analyzing Wage Differentials by Fields of Study: Evidence from Turkey
(Octubre 2017)

XREAP2017-03
Melguizo, C. (AQR-IREA, XREAP); **Royuela, V.** (AQR-IREA, XREAP)
What drives migration moves across urban areas in Spain? Evidence from the great recession
(Octubre 2017)

XREAP2017-04
Boonen, T.J., Guillén, M. (RISKCENTER, XREAP); **Santolino, M.** (RISKCENTER, XREAP)
Forecasting compositional risk allocations
(Octubre 2017)

XREAP2017-05
Curto-Grau, M. (IEB, XREAP)**, Solé-Ollé, A.** (IEB, XREAP)**, Sorribas-Navarro, P.** (IEB, XREAP)
Does electoral competition curb party favoritism?
(Novembre 2017)

XREAP2017-06
Esteller, A. (IEB, XREAP)**, Piolatto, A.** (IEB, XREAP)**, Rablen, M. D.**
Taxing high-income earners: tax avoidance and mobility
(Novembre 2017)

XREAP2017-07
Bolancé, C. (RISKCENTER, XREAP), **Vernic, R**
Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution
(Novembre 2017)

XARXA
DE REFERÈNCIA
EN ECONOMIA APLICADA

xarxa.xreap@gmail.com