A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking

Lluís Bermúdez[a,*], Dimitris Karlis[b]

[a] *Risc en Finances i Assegurances-IREA, Universitat de Barcelona, Spain*
[b] *Athens University of Economics and Business, Greece*

Abstract

Bivariate Poisson regression models for ratemaking in car insurance has been previously used. They included zero-inflated models to account for the excess of zeros and the overdispersion in the data set. These models are now revisited in order to consider alternatives. A 2-finite mixture of bivariate Poisson regression models is used to demonstrate that the overdispersion in the data requires more structure if it is to be taken into account, and that a simple zero-inflated bivariate Poisson model does not suffice. At the same time, it is shown that a finite mixture of bivariate Poisson regression models embraces zero-inflated bivariate Poisson regression models as a special case. Finally, an EM algorithm is provided in order to ensure the models' ease-of-fit. These models are applied to an automobile insurance claims data set and it is shown that the modelling of the data set can be improved considerably.

Keywords: Zero-inflation, Overdispersion, EM algorithm, Automobile insurance, *A priori* ratemaking

1. Introduction

In a recent paper Bermúdez (2009) describes bivariate Poisson (BP) regression models for ratemaking in car insurance. The central idea is that the dependence between two different types of claim must be taken into account to achieve better ratemaking. BP regression models are presented, therefore, as an instrument that can account for the underlying correlation between

*Corresponding Author. Departament de Matemàtica Econòmica, Financera i Actuarial, Universitat de Barcelona, Diagonal 690, 08034-Barcelona, Spain. Tel.:+34-93-4034853; fax: +34-93-4034892; e-mail: `lbermudez@ub.edu`

two types of claim arising from the same policy (i.e. third-party liability claims and all other automobile insurance claims). The paper concludes that even when there are small correlations between the claims, major differences in ratemaking can nevertheless appear. Thus, using a BP model results in ratemaking that has larger variances and, hence, larger loadings in premiums than those obtained under the independence assumption.

The paper also includes zero-inflated bivariate Poisson (ZIBP) models so as to inflate the (0,0) cell and to account for the excess of zeros and overdispersion typically observed in this type of data set. This produces the best goodness of fit among the bivariate Poisson models considered. In conclusion, the independence assumption should be rejected when using either BP or ZIBP regression models, but one question still remains unresolved: do ZIBP models constitute the best option for dealing with the unobserved heterogeneity usually observed in such a data? The aim of the present paper is to examine this question further by considering alternative bivariate models that might account for both these features of the data, i.e. the excess of zeros and overdispersion.

In the context of automobile insurance, the problem of unobserved heterogeneity is caused by the differences in driving behavior among policyholders that cannot be observed or measured by the actuary, such as a driver's reflexes, his or her aggressiveness, or knowledge of the Highway Code, among others. The main consequence of unobserved heterogeneity is overdispersion, i.e. data exhibiting variance larger than mean, which it invalidates the use of a simple Poisson distribution. The presence of excess of zeros in most insurance data sets can be also seen as a consequence of unobserved heterogeneity.

In the univariate case, Lambert (1992) introduced the zero-inflated Poisson regression model. Since then, there has been a considerable increase in the number of applications of zero-inflated regression models based on several different distributions. A comprehensive discussion of these applications can be found in Winkelmann (2008) and a specific application to insurance ratemaking is addressed in Boucher et al. (2007). Zero inflated negative binomial regression models have been also described as for example in Wang (2003) and Garay et al. (2011). See again Winkelmann (2008) for a description of a variety of such models and Denuit et al. (2007) for an exhaustive review of the models used in ratemaking systems for automobile insurance.

In the bivariate (or multivariate) case, the literature analysing the excess of zeros and overdispersion is less developed. For example, zero-inflation in the bivariate case is examined in Gurmu and Elder (2008), Karlis and

Ntzoufras (2003) and the references therein, while in the multivariate case it is analysed in Li et al. (1999). Recently, in the actuarial literature and for ratemaking purposes, Bermúdez (2009) and Bermúdez and Karlis (2011) deal with the bivariate and multivariate versions of the zero-inflated Poisson regression models, respectively. They tackle overdispersion via the excess of zeros, i.e. zero-inflated models.

A natural approach for accounting for overdispersion is to consider models with some overdispersed marginal distribution, as opposed to bivariate Poisson models. In this paper we consider an m-finite mixture of bivariate Poisson regressions (m-FMBP) extending the no-covariate cases presented in Karlis and Meligkotsidou (2007). This model has a number of interesting features: first, the zero-inflated model represents a special case; second, it allows for overdispersion; and, third, it allows for an elegant interpretation based on the typical clustering application of finite mixture models. To the best of our knowledge, this model is new to the literature, so in what follows we seek to explain its properties as well as to discuss appropriate estimation approaches.

The rest of of the paper proceeds as follows. The new model is described in the next section, followed by the development of an EM algorithm for parameter estimation. The model is then applied to the same data set as in Bermúdez (2009). Finally, we conclude with some remarks.

2. The proposed model

2.1. A bivariate Poisson distribution

Consider random variables X_k, $k = 1, 2, 3$ which follow independent Poisson distributions with parameters $\lambda_k \geq 0$, respectively. Then the random variables $Y_1 = X_1 + X_3$ and $Y_2 = X_2 + X_3$ jointly follow a bivariate Poisson distribution, denoted as $BP(\lambda_1, \lambda_2, \lambda_3)$, with joint probability function given by

$$\begin{aligned} P_{Y_1,Y_2}(y_1, y_2) &= P(Y_1 = y_1, Y_2 = y_2) = \\ &= BP(y_1, y_2; \lambda_1, \lambda_2, \lambda_3) \\ &= e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^{y_1}}{y_1!} \frac{\lambda_2^{y_2}}{y_2!} \sum_{s=0}^{min(y_1,y_2)} \binom{y_1}{s} \binom{y_2}{s} s! \left(\frac{\lambda_3}{\lambda_1\lambda_2}\right)^s . \end{aligned}$$

The above bivariate distribution allows for dependence between the two random variables. It is also related to a common shock model. Marginally each

random variable follows a Poisson distribution with $E(Y_1) = \lambda_1 + \lambda_3$ and $E(Y_2) = \lambda_2 + \lambda_3$. Moreover, $Cov(Y_1, Y_2) = \lambda_3$, and hence λ_3, is a measure of dependence between the two random variables. If $\lambda_3 = 0$ then the two variables are independent and the bivariate Poisson distribution reduces to the product of two independent Poisson distributions (also known as a double Poisson distribution). For a comprehensive treatment of the bivariate Poisson distribution and its multivariate extensions the reader is referred to Kocherlakota and Kocherlakota (1992) and Johnson et al. (1997).

For greater flexibility, we can assume a bivariate Poisson regression model where each of the parameters of the BP is related to some covariates through a log link function, i.e. by assuming

$$\log \lambda_{ki} = \boldsymbol{\beta}_k^T \boldsymbol{x}_{ki}, \quad k = 1, 2, 3, \quad i = 1, \ldots, n,$$

where $\boldsymbol{x}_{ki}$ is a vector of covariates for the i-th observation related to the k-th parameter and $\boldsymbol{\beta}_k$ is the associated vector of regression coefficients. Note that $\boldsymbol{x}$ does not need to be the same for all the parameters. Likewise note that according to Karlis and Ntzoufras (2003), it is perhaps a good idea not to use the same covariates in all the parameters since this may lead to problems in their interpretation. For example, since the marginal mean for Y_1 is $\lambda_1 + \lambda_3$ using the same covariates in both may create problems of interpretation especially if the signs of the regression coefficients differ. `R` package `bivpois` can be used to fit this model based on an EM algorithm.

In this model, and since the marginal distributions are Poisson, the marginal means and variances are equal. Moreover, that the correlation is positive. Therefore, there we need to consider extensions to allow for overdispersion (variance greater than the mean) and a possible negative correlation.

2.2. Mixed bivariate Poisson models

A natural way to allow for overdispersion is to consider mixtures of a simpler model. This is best achieved in the univariate setting by moving from the simple Poisson model to the negative binomial model. Such an approach, while applicable in the bivariate setting, is not an easy task especially because there is not a single correct way to do it and, in consequence, questions of ease and interpretation come into play.

Mixtures of BP distribution can be considered in at least two different ways. In the first we start with a $BP(a\lambda_1, a\lambda_2, a\lambda_3)$ distribution where a follows some distribution. We can assume $\lambda_3 = 0$ which makes the calculation

much easier and assumes that all the correlation comes from the common a. If $\lambda_3 > 0$ then the correlation is twofold, due to λ_3 (known as an intrinsic correlation) and due to the common a. This complicates the interpretation of the parameters. A natural assumption in this case is that $E(a) = 1$ so a does not inflate the means. This is a very typical extension of a simple mixed Poisson regression models. One drawback, however, is that the model only allows a positive correlation. The literature on this approach includes the works of Stein et al. (1987), **?**, Stein and Yuritz (1987) and Kocherlakota (1988) for the case without covariates. Munkin and Trivedi (1999) described multivariate mixed Poisson regression models based on this type of mixing and a gamma mixing distribution. Gurmu and Elder (2000) used an extended gamma density as a mixing distribution. This approach also has a random-effect representation if covariates are used. This assumes that

$$\begin{aligned} Y_{1i}, Y_{2i} &\sim BP(\lambda_{1i}, \lambda_{2i}, \lambda_{3i}), \\ \log \lambda_{ki} &= \boldsymbol{\beta}_k^T \boldsymbol{x}_{ki} + u_i, \quad k = 1, 2, 3, \quad i = 1, \ldots, n \qquad \text{and} \\ u_i &\sim G(u), \end{aligned}$$

where u_i is the random effect associated with the i-th observation, common to all the parameters. In fact this approach is equivalent to a frailty model.

In the second case, we start with a $BP(a_1\lambda_1, a_2\lambda_2, a_3\lambda_3)$ distribution, but now the a's are different. We need to assume that they jointly follow a trivariate (or bivariate if we assume that $\lambda_3 = 0$) distribution. Clearly such a construction is much more complicated and, in practice, not especially useful. The case when $\lambda_3 = 0$ has received attention primarily because it can induce negative correlation between counts. Steyn (1976) proposed the use of a bivariate normal distribution as the mixing distribution. Some years later, Aitchinson and Ho (1989) proposed using the bivariate lognormal distribution instead of the simple bivariate normal distribution. For a Bayesian application of this distribution see Chib and Winkelmann (2001).

To put it in a random effect format the above model is equivalent to assuming

$$\begin{aligned} Y_{1i}, Y_{2i} &\sim BP(\lambda_{1i}, \lambda_{2i}, \lambda_{3i}), \\ \log \lambda_{ki} &= \boldsymbol{\beta}_k^T \boldsymbol{x}_{ki} + u_{ki}, \quad k = 1, 2, 3, \quad i = 1, \ldots, n \qquad \text{and} \\ u_{1i}, u_{2i}, u_{3i} &\sim G(\cdot), \end{aligned}$$

where now $G(\cdot)$ is a trivariate distribution and, hence, the random effects are different, albeit related, for each parameter. Again for purposes of iden-

tifiability, it must be assumed that the expectation for each random effect is 1.

In both of the above models the specification of the random-effects distribution $G(\cdot)$ can be a continuous, a discrete or a finite distribution. Here we consider the latter case assuming that the joint distribution for the random effects is a finite distribution, i.e. the case in which only a finite number of points have positive probabilities. Such an assumption gives rise to finite mixture models, which are very popular in a range of disciplines. These models, i.e. finite mixtures of multivariate Poisson distributions, have been described in Karlis and Meligkotsidou (2007).

The novelty of our approach lies in the fact that we assume different regression lines for each component in the mixture, extending the finite mixture Poisson regression model of Wang et al. (1998) (see Grun and Leisch (2007) for the implementation of models of this type) in two dimensions. Thus, in the next section we introduce the finite mixture of bivariate Poisson regressions.

2.3. The finite mixture of bivariate Poisson regressions

Let the $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \lambda_3)$ denote the vector of parameters. We define as an m-finite mixture of bivariate Poisson distributions the distribution with joint probability function

$$P(y_1, y_2) = \sum_{j=1}^{m} p_j BP(y_1, y_2; \boldsymbol{\theta}_j),$$

where $p_j > 0, \;\; j = 1, \ldots, m$ are the mixing proportions with $\sum_{j=1}^{m} p_j = 1$ and $\boldsymbol{\theta}_j$ are the component-specific vectors of parameters, namely $\boldsymbol{\theta}_j = (\lambda_{1j}, \lambda_{2j}, \lambda_{3j})$. In the sequel the first subscript denotes the parameter and the second the component, while if we require a further subscript to indicate the observation we will use a third one.

In this mixture model, the marginal expectations are given by

$$E(Y_k) = \sum_{j=1}^{m} p_j(\lambda_{kj} + \lambda_{3j}), \quad k = 1, 2,$$

while its variance covariance matrix of $\boldsymbol{Y} = (Y_1, Y_2)^T$ is given by

$$Var(\boldsymbol{Y}) = \boldsymbol{A}\left[\sum_{j=1}^{m} p_j \boldsymbol{\Sigma}_j - \left(\sum_{j=1}^{m} p_j \boldsymbol{\theta}_j\right)\left(\sum_{j=1}^{m} p_j \boldsymbol{\theta}_j\right)^T\right]\boldsymbol{A}^T,$$

where

$$\boldsymbol{\Sigma}_j = \begin{bmatrix} \lambda_{1j}^2 + \lambda_{1j} & \lambda_{1j}\lambda_{2j} & \lambda_{1j}\lambda_{3j} \\ \lambda_{1j}\lambda_{2j} & \lambda_{2j}^2 + \lambda_{2j} & \lambda_{2j}\lambda_{3j} \\ \lambda_{1j}\lambda_{3j} & \lambda_{2j}\lambda_{3j} & \lambda_{3j}^2 + \lambda_{3j} \end{bmatrix}$$

and

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

This can be written in the following interesting form

$$Var(\boldsymbol{Y}) = \boldsymbol{A}\boldsymbol{D}(\boldsymbol{\theta})\boldsymbol{A}^T,$$

where

$$\boldsymbol{D}(\boldsymbol{\theta}) = \begin{bmatrix} Var(\lambda_1) + E(\lambda_1) & Cov(\lambda_1, \lambda_2) & Cov(\lambda_1, \lambda_3) \\ Cov(\lambda_1, \lambda_2) & Var(\lambda_2) + E(\lambda_2) & Cov(\lambda_2, \lambda_3) \\ Cov(\lambda_1, \lambda_3) & Cov(\lambda_2, \lambda_3) & Var(\lambda_3) + E(\lambda_3) \end{bmatrix},$$

which results in

$$Cov(Y_1, Y_2) = Cov(\lambda_1, \lambda_2) + Cov(\lambda_2, \lambda_3) + Cov(\lambda_1, \lambda_3) + Var(\lambda_3) + E(\lambda_3).$$

Thus if the λ's are negatively correlated we can end up with negative correlation.

The above model has some interesting properties. First, as shown in Karlis and Meligkotsidou (2007), even if $\lambda_3 = 0$, i.e. within each component the two variables are uncorrelated, the Y's are correlated due to the correlation induced by the finite distribution of the λ's. Such a model, with $\lambda_3 = 0$ for all the components, actually assumed independence within each component, but again overall we can have correlation. Second, the correlation between Y_1 and Y_2 can be negative, while Y_1 and Y_2 are overdispersed if $m > 1$. Note also that the marginal distributions are finite Poisson mixtures. Finally, as we prove in Appendix A, mixed bivariate Poisson distributions always give equal or greater probability to the (0,0) cell from the corresponding bivariate Poisson with the same marginal means. Furthermore, zero-inflated

bivariate Poisson models can be considered a special case of this model, when the first component has $\lambda_1 = \lambda_2 = \lambda_3 = 0$ and, hence, all the probability mass is given in the (0,0) cell. This also suggests why zero-inflated models are overdispersed and can induce different correlation structures.

In Appendix B, we summarize some of the moments of the finite mixture of bivariate Poisson distribution. These quantities can be used for actuarial purposes as in Bermúdez (2009) and Bermúdez and Karlis (2011). Basically, following the net premium principle, the total net premium is obtained as $E(Y_1+Y_2)$. However, following the variance premium principle, i.e. including a risk loading to the net premium that is proportional to the variance of the risk, $Var(Y_1 + Y_2)$ must be obtained.

In order to include covariates and thus allow for greater flexibility we assume that each parameter is associated to a vector of regressors. Namely our model takes the form

$$\boldsymbol{Y}_i = (Y_{1i}, Y_{2i}) \sim \sum_{j=1}^{m} p_j BP(y_1, y_2; \lambda_{1ji}, \lambda_{2ji}, \lambda_{3ji}), \;\; i = 1, \ldots, n, \;\; j = 1, \ldots, m,$$

where

$$\log(\lambda_{kji}) = \boldsymbol{\beta}_{kj}^T \boldsymbol{x}_{kji}, \quad k = 1, 2, 3, \quad j = 1, \ldots, m, \tag{1}$$

where $\boldsymbol{x}_{kji}$ is a vector of covariates for the i-th observation associated with the k-th parameter of the j-th component of the mixture and $\boldsymbol{\beta}_{kj}$ is the set of regression coefficients. It is clear that the covariates can differ for different parameters.

This model extends the finite mixture of Poisson regression model of Wang et al. (1998). The model assumes that for each variable we have m distinct Poisson regression models that relate the variable of interest with different covariates. Hence, we assume that the population has several distinct clusters presenting different behaviour. The added feature is that now we model two variables together and so we are able to take into account their relationships and correlation. Moreover, starting from a bivariate Poisson model, within each group we may assume a different correlation structure. A natural extension of the model is to use covariates also in the mixing proportions, i.e. the vector of probabilities $(p_1, \ldots, p_m)$. A typical choice is to assume a multinomial logistic model for the vector of mixing proportions (reducing to simple logistic regression if only two components are present).

In the next section, we provide an EM algorithm to allow for a relatively simple maximum likelihood (ML) estimation of the model. It is based on the

standard EM for finite mixtures but also takes into account the trivariate reduction derivation of the bivariate Poisson model.

3. ML estimation via an EM algorithm

In this section we develop an EM algorithm. The parameters to be estimated are the mixing proportions p_j, $j = 1, \ldots, m-1$, and the component-specific vector of regression coefficients $\boldsymbol{\beta}_{kj}$, $k = 1, 2, 3,\ j = 1, \ldots, m$.

Being a finite mixture, standard missing data representation is possible. Let $\boldsymbol{Z}_i = (Z_{1i}, \ldots, Z_{mi})$ be a vector with $Z_{ji} = 1$ if the i-th observation belongs to the j-th group and 0 elsewhere. We also introduce component-specific latent variables, i.e. for the j-th component we use the unobservable vectors $\boldsymbol{Y}_i^{j*} = (T_{1ji}, T_{2ji}, S_{ji})$ such as $Y_{1i} = T_{1ji} + S_{ji}$ and $Y_{2i} = T_{2ji} + S_{ji}$, as the trivariate reduction derivation implies. The algorithm is similar to that described in Brijs et al. (2004), but here we also have regressors. Clearly if $\boldsymbol{Z}_i$ and S_{ji} were observables then estimation would have been a simple task, since at the E-step we need to obtain the conditional expectations. The algorithm is now given by:

E-step: Given the values of the parameters after the rth iteration we obtain from (1), $\boldsymbol{\lambda}_{1j}^{(\tau)}$, $\boldsymbol{\lambda}_{2j}^{(\tau)}$ and $\boldsymbol{\lambda}_{3j}^{(\tau)}$ and then we calculate the expected values of the unobservables:

$$\begin{aligned} s_{ji} &= E(S_{ji} \mid Y_{1i}, Y_{2i}, \boldsymbol{\lambda}_{1j}^{(\tau)}, \boldsymbol{\lambda}_{2j}^{(\tau)}, \boldsymbol{\lambda}_{3j}^{(\tau)}) = \\ &= \begin{cases} \lambda_{3ji}^{(\tau)} \frac{BP(y_{1i}-1, y_{2i}-1; \lambda_{1ji}^{(\tau)}, \lambda_{2ji}^{(\tau)}, \lambda_{3ji}^{(\tau)})}{BP(y_{1i}, y_{2i}; \lambda_{1ji}^{(\tau)}, \lambda_{2ji}^{(\tau)}, \lambda_{3ji}^{(\tau)})}, & \text{if } y_{1i}y_{2i} > 0 \\ 0 & \text{if } y_{1i}y_{2i} = 0 \end{cases} \end{aligned}$$

and

$$w_{ji} = \frac{p_j^{(\tau)} BP(y_{1i}, y_{2i}; \lambda_{1ji}^{(\tau)}, \lambda_{2ji}^{(\tau)}, \lambda_{3ji}^{(\tau)})}{\sum_{j=1}^{m} p_j^{(\tau)} BP(y_{1i}, y_{2i}; \lambda_{1ji}^{(\tau)}, \lambda_{2ji}^{(\tau)}, \lambda_{3ji}^{(\tau)})}.$$

M-step: Update the estimates by

$$\begin{aligned} p_j^{(\tau+1)} &= \sum_{i=1}^{n} w_{ji}/n \\ \boldsymbol{\beta}_{1j}^{(\tau+1)} &= \hat{\boldsymbol{\beta}}(\boldsymbol{y}_1 - \boldsymbol{s}_j, \boldsymbol{x}_1, \boldsymbol{w}_j), \\ \boldsymbol{\beta}_{2j}^{(\tau+1)} &= \hat{\boldsymbol{\beta}}(\boldsymbol{y}_2 - \boldsymbol{s}_j, \boldsymbol{x}_2, \boldsymbol{w}_j) \qquad \text{and} \\ \boldsymbol{\beta}_{3j}^{(\tau+1)} &= \hat{\boldsymbol{\beta}}(\boldsymbol{s}_j, \boldsymbol{x}_3, \boldsymbol{w}_j), \end{aligned}$$

where $\boldsymbol{s}_j = [s_{j1}, \ldots, s_{jn}]^T$ is the $n \times 1$ vector, $\hat{\boldsymbol{\beta}}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{w})$ are the weighted maximum likelihood estimated parameters of a Poisson model with response to the vector $\boldsymbol{y}$, design or data matrix given by $\boldsymbol{x}$ and weights $\boldsymbol{w}$. Note that different covariates may be used for each λ, i.e. different design matrices.

The above procedure has all the advantages and drawbacks of the EM algorithm. For this reason, suitable terminating conditions should be considered carefully. In the case in which covariates are also used for the mixing proportions, then the M-step has to be replaced by one that fits a multinomial logistic (or a simple logistic if only two components are considered) regression using w_{ij} as response vector.

Finally, initial values can be obtained by fitting a simple univariate Poisson regression to each variable so as to obtain the fitted values. Then, by simply perturbing them (e.g. multiplying the lambda expressions by 0.8 and 1.2), we can obtain initial values for each component. Initial values for the mixing proportions are less important for initialization. Furthermore, as in other finite mixture settings, initial values can be obtained using a standard clustering algorithm. Note that obtaining initial estimates for the w_{ji} is sufficient to initialise the algorithm.

4. Application

4.1. The data

The original population comprised a ten-percent sample of the 1996 automobile portfolio (note, only automobiles categorized as being for private use were considered) of a major insurance company operating in Spain and contains information for 80,994 policyholders. The data have previously been also used in Bermúdez (2009) where bivariate Poisson models, including zero-inflated models, were fitted. The sample is not representative of the company's current portfolio, being drawn from a larger panel of policyholders that had been customers of the company for at least seven years; however, the sample should be helpful here for illustrative purposes. Twelve exogenous variables were considered plus the annual number of accidents recorded for both types of claim. For each policy, the information at the beginning of the period and the total number of claims from policyholders "at fault" were reported for each year. The exogenous variables, described in Table 1 outline the covariates, and this data set have previously been used in Pinquet et al. (2001), Brouhns et al. (2003), Bolancé et al. (2003), Bolancé et al. (2008), Boucher et al. (2007), Boucher and Denuit (2008) and in Boucher et al.

Variable	Definition
V1	equals 1 for women and 0 for men
V2	equals 1 when driving in urban area, 0 otherwise
V3	equals 1 when zone is medium risk (Madrid and Catalonia)
V4	equals 1 when zone is high risk (Northern Spain)
V5	equals 1 if the driving license is between 4 and 14 years old
V6	equals 1 if the driving license is 15 or more years old
V7	equals 1 if the client is in the company between 3 and 5 years
V8	equals 1 if the client is in the company for more than 5 years
V9	equals 1 of the insured is 30 years old or younger
V10	equals 1 if includes comprehensive coverage (except fire)
V11	equals 1 if includes comprehensive and collision coverage
V12	equals 1 if horsepower is greater than or equal to 5500cc

Table 1: Explanatory variables used in the models

(2009). In this study, all customers had held a policy with the company for at least three years. Thus, variable *V7* could be rejected and variable *V8* retained, the latter's baseline now being established as a customer who had been with the company for fewer than five years.

The meaning of the variables that refer to the policyholders' coverage should also be clarified. The classification adopted here responds to the most common types of automobile insurance policy available on the Spanish market. The simplest policy only includes just third-party liability (claimed and counted as Y_1) and a set of basic guarantees such as emergency road- side assistance, legal assistance or insurance covering medical costs but it does not include comprehensive coverage or collision coverage (claimed and counted as Y_2). This simplest type of policy makes up the baseline group, while variable *V10* denotes policies which, apart from the guarantees contained in the simplest policies, also include comprehensive coverage (except fire), and variable *V11* denotes policies which also include fire and collision coverage. Y_1 and Y_2 are the two response variables used in the paper.

4.2. Results

We fitted a 2-finite mixture of bivariate Poisson regressions to this data set. We have avoided running a model with three components as the interpretation of such a model would have been more difficult and because a 2-finite mixture allows sufficient interpretation of this particular data set.

Model	Log-Lik	Parameters	AIC
Double Poisson	-48,882.95	24	97,813.90
Bivariate Poisson (BP)	-48,135.98	25	96,321.96
BP (regressors on λ_3)	-47,873.37	26	95,798.74
Zero inflated BP (ZIBP)	-45,435.00	26	90,922.00
ZIBP (regressors on λ_3)	-45,414.80	27	90,883.60
2-finite mixture BP (2-FMBP1)	-44,927.01	51	89,956.02
2-FMBP2 (regressors on p)	-44,842.22	53	89,737.44
Bivariate Negative Binomial	-45,080.12	25	90,210,24
Bivariate Poisson Lognormal	-45,036.85	27	90,127.70

Table 2: Information criteria for selecting the best model for the data

However, models with more components can easily be fitted via the EM algorithm provided. Further, we estimated both λ_3 parameters as being equal to 0, implying that conditional on the component no correlation was present.

The first model fitted does not have covariates in the mixing proportion p while the second uses *V10* and *V11* as covariates in the mixing proportion. We used these covariates because when fitting the first model we noticed that there was a large difference in a posteriori probabilities when considering values 0 or 1 for *V10* and *V11*. We return to this issue later. Bermúdez (2009) also used $V10$ covariate to model λ_3 parameter. In the sequel, 2-FMBP1 is the name given to the first model without covariates on p and 2-FMBP2 is the name given to the second model with covariates on p. Models were fitted via the EM algorithm provided.

Table 2 presents the results from fitting various models to the data. We fitted models of increasing complexity, starting from a simple independent Poisson regression model. The first five models are the same as those fitted in Bermúdez (2009). It can be seen that the 2-finite mixture of bivariate Poisson regressions are by far the best models, especially the regression with covariates in the mixing proportion, which has the best AIC. We also include in the table some more models mentioned in Section 2.2 to show the gain offered by our model.

Table 3 shows the results from fitting the 2-finite mixture of bivariate Poisson model with the covariates in the mixing proportion. The p-value refers to the likelihood ratio test (LRT) statistic when the variable is included or excluded from the model. We prefer this approach as standard errors in fi-

	1st component ($j = 1$)				2nd component ($j = 2$)			
	Y_1		Y_2		Y_1		Y_2	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Intercept	0.071	< 0.001	-1.611	< 0.001	-3.118	< 0.001	-6.014	< 0.001
V1	-0.061	0.115	0.032	0.218	0.127	0.059	0.037	0.258
V2	-0.037	0.162	0.008	0.308	-0.076	0.123	0.179	0.001
V3	-0.090	0.027	0.106	0.003	0.197	0.006	0.242	< 0.001
V4	0.129	0.003	-0.043	0.166	0.284	< 0.001	-0.371	< 0.001
V5	-0.132	0.142	0.153	0.111	-0.346	0.016	0.452	0.003
V6	-0.216	0.052	0.027	0.313	-0.524	0.002	0.137	0.215
V8	0.101	0.022	0.135	0.002	0.190	0.013	0.326	< 0.001
V9	0.078	0.145	0.035	0.262	0.193	0.048	0.171	0.024
V10	-0.707	< 0.001	1.622	< 0.001	-2.676	< 0.001	2.953	< 0.001
V11	-0.361	< 0.001	1.069	< 0.001	-0.285	0.002	2.412	< 0.001
V12	0.036	0.224	0.102	0.028	0.079	0.145	0.397	< 0.001
λ_3	0.000	-			0.000	-		
Mixing Proportion (p)								
Intercept	-2.4595	< 0.001						
V10	1.447	< 0.001						
V11	0.680	< 0.001						

Table 3: Results from fitting the 2-FMBP2 model (with regressors on p)

nite mixtures are not easy to derive. In our case we would need to derive the Hessian of the log-likelihood function which is particularly time consuming and vulnerable to overflows as we have 53 parameters (12 regression coefficients for each variable for two of the components, plus three coefficients for the mixing proportion and two covariance parameters). Bootstrapping as an alternative can also be very slow. So, we removed each variable each time and calculated a LRT. The p-values reported correspond to this LRT.

Figures 1 and 2 help illustrate that the 2-finite mixture of bivariate Poisson regression (with covariates on p) is a good option, and better in all circumstances than a zero-inflated bivariate Poisson regression, for dealing with overdispersion and the excess of zeros present in the data set.

Figure 1 shows the components fitted. We plotted boxplots for the two components for the two variables under consideration. The boxplots represent the values of λ_{kji} for $k = 1, 2$, $j = 1, 2$, and $i = 1, \ldots, 80,994$. From this plot, it can be readily seen that the first component corresponds to policyholders with high rates of claims for both variables, Y_1 and Y_2, while the

second component corresponds to those with small claim rates. In fact, the second component has very small means for the underlying Poisson components, which implies a high probability of zeros. Thus, the second component introduces a large amount of zero inflation in our model.

Bermúdez (2009) fitted zero-inflated bivariate Poisson models to account for the excess of zeros found with respect to the simple bivariate Poisson model while at the same time, allowing for overdispersion. Here, we show that the problem is more than one of simple zero inflation. Thus, by assuming the existence of two types of policyholder described separately by each component in the mixture, we are able to improve considerably the modelling of the data set. Indeed zero-inflated models represent special instances of the finite mixture model presented here, which was considered, at least initially, to account for overdispersion.

In fact, in the univariate case, Lord et al. (2005) and Lord et al. (2007) criticize zero-inflated models when modelling the number of accidents owing to a dual-state process assumption. According to them, the claim is made that zero-inflated models assume two sources of zeros: "true" and "observed". The existence of "true" zeros may be too strong an assumption in some cases (see also Boucher and Santolino (2010)). However, as Park and Lord (2009) discuss in the univariate case, the two-component mixture model used here does not make this somewhat strict dual-state process assumption and allows mixing with respect to both zeros and positives. This interpretation is more flexible and it holds better in our case. The group separation is characterized by low mean with low variance (policyholders considered as a "good" drivers) and high mean with high variance (policyholders considered as a "bad" drivers).

From Figure 1, it is also interesting to note that third-party liability claims (Y_1) present greater separation between the two components than is shown by the rest of automobile claims (Y_2). For each observation, we also calculated the underlying variance and covariance. These are depicted in Figure 2. The horizontal line is the observed quantity and the boxplot refers to the values fitted for each individual based on the second model (the one with covariates on p). As for the covariance, we can see that the model captures this quite well. In the case of the variance, we can see that the model's prediction is somewhat smaller than that observed. This is perhaps an indication that some overdispersion remains uncaptured, either because a third component could be fitted or because we have overlooked some covariates.

Most of the parameters are significant. Note, however, that the sample

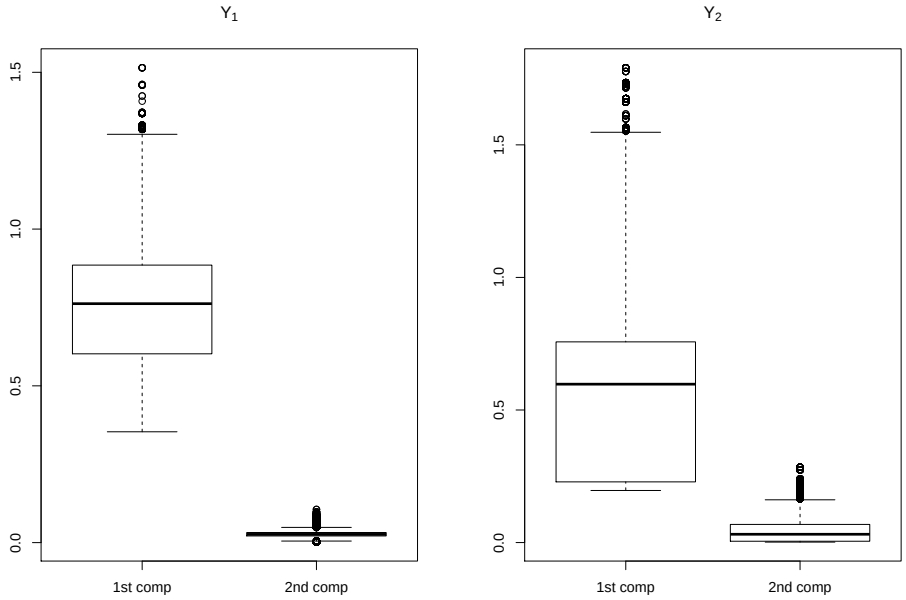

Figure 1: The fitted components for the two variables analyzed

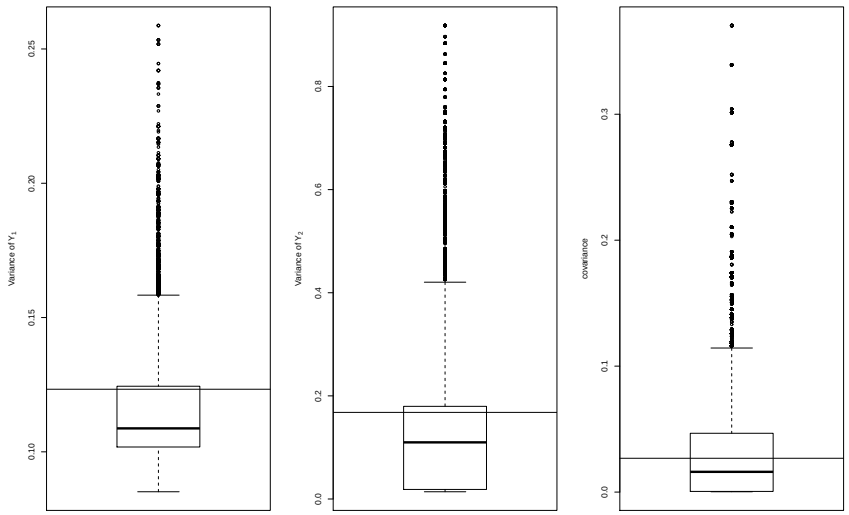

Figure 2: Variance and covariance for the fitted model

size was very large. Any variable selection technique could have been used to reduce the number of variables, however in this application we preferred to retain all the variables in order to see their effect. Recall that it is not necessary to use the same covariate vector for all the parameters. Only the parameter related to gender (*V1*) is not significant in all cases, i.e. for both components and both response variables. On the other hand, parameters related to the driving zone (*V3*), the number of years the customer has been with the company (*V8*), and the type of coverage (*V10* and *V11*) present significant coefficients for both components and both response variables. It is interesting to note that parameters *V10* and *V11* present coefficients of different signs for each response variable. For the Y_1 variable (third-party liability claims), the more policy guarantees the customers take out the fewer claims they report. The opposite is the case for the Y_2 variable (all other automobile insurance claims). Finally, parameters *V5* and *V9*, related to the policyholder's driving experience and age respectively, are significant only for the second component, while parameter *V12*, related to the car's horsepower, is only significant for the second response variable.

A more detailed explanation of the coefficients is of interest here to differ-

entiate between the two groups. Recall that the first component corresponds to the policyholders considered "good" drivers, characterized by a low mean with low variance, and the second component corresponds to the policyholders considered "bad" drivers, characterized by a high mean with high variance. Most of the parameters present the same behaviour for both "good" and "bad" drivers. This is the case of the parameters related to the driving zone (*V3* and *V4*), the type of coverage (*V10* and *V11*), and the car's horsepower (*V12*). Another example is the longer the customer has been with the company (*V8*) the more claims the policyholder reports, regardless of the group to which he or she belongs. By contrast, three parameters are only significant for the second component, and as such can be used to define a "bad" driver. These are basically the parameters related to a driver's age and driving experience. The fact of being thirty years old or younger (*V9*) results in the expected number of claims to increase for all types of claim. Driving experience (*V5*) reduces the expected number of third-party liability claims, but increases the expected number for all other automobile insurance claims. Moreover, "bad" drivers in urban areas (*V2*) only present a larger expected number of claims for Y_2 type claims. Finally, *V10* and *V11* are also highly significant for the mixing proportion, implying that the existence of *V10* and *V11* increases the probability of belonging to the first cluster. Hence, "good" drivers take out more guarantees in their policies than is the case with "bad" drivers.

Table 4 presents the observed and expected frequencies under the two 2-finite mixture of bivariate Poisson regressions. To obtain the expected frequencies, for each observations we calculated the probability table based on the estimated parameters and then we summed all these probability tables to obtain the one with the expected frequencies. The fit is quite good, while there are still a few cells that have large residuals. The results of the chi-square test show that only a few cells contribute to this goodness of fit, but owing to the very large sample size, rejecting the null hypothesis is somewhat artefactual. It is our belief that the fit is, in fact, very good given the size of the data set. Furthermore, note that a zero-inflated model would only correct with respect to the (0,0) cell and not to the entire probability table.

Finally, we present Figure 3 in an effort to see which variables characterize each cluster and which variables can be included as regressors in the mixing proportion p. Using the posterior probabilities, available on finishing the EM algorithm, we can classify each observation to a cluster, based (as usual) on the maximum posterior probability. Since all the variables were binary, for

Y_1		Y_2							
		0	1	2	3	4	5	6	7
0	Observed	71087	3722	807	219	51	14	4	0
	2-FMBP1	70992.70	3932.01	593.23	161.74	49.25	13.93	3.54	0.80
	2-FMBP2	71045.30	3806.84	644.29	191.87	54.33	13.26	2.81	0.53
1	Observed	3022	686	184	71	26	10	3	1
	2-FMBP1	3032.90	753.38	253.26	87.06	27.60	7.87	2.01	0.46
	2-FMBP2	3055.16	737.08	280.54	92.21	25.57	6.10	1.28	0.24
2	Observed	574	138	55	15	8	4	1	1
	2-FMBP1	580.86	200.93	71.19	24.73	7.87	2.25	0.58	0.13
	2-FMBP2	476.95	217.76	77.63	23.73	6.27	1.45	0.30	0.05
3	Observed	149	42	21	6	6	1	0	1
	2-FMBP1	107.79	38.11	13.61	4.75	1.52	0.44	0.11	0.03
	2-FMBP2	117.96	49.36	15.41	4.31	1.07	0.24	0.05	0.01
4	Observed	29	15	3	2	1	1	0	0
	2-FMBP1	15.41	5.50	1.98	0.69	0.22	0.06	0.02	0
	2-FMBP2	24.48	9.06	2.46	0.62	0.14	0.03	0.01	0
5	Observed	4	1	0	0	0	0	2	0
	2-FMBP1	1.79	0.65	0.23	0.08	0.03	0.01	0	0
	2-FMBP2	4.26	1.42	0.34	0.08	0.02	0	0	0
6	Observed	2	1	0	1	0	0	0	0
	2-FMBP1	0.18	0.06	0.02	0.01	0	0	0	0
	2-FMBP2	0.64	0.20	0.04	0.01	0	0	0	0
7	Observed	1	0	0	1	0	0	0	0
	2-FMBP1	0.02	0.01	0	0	0	0	0	0
	2-FMBP2	0.08	0.02	0	0	0	0	0	0
8	Observed	0	0	1	0	0	0	0	0
	2-FMBP1	0	0	0	0	0	0	0	0
	2-FMBP2	0.01	0	0	0	0	0	0	0

Table 4: Observed and expected frequencies

each cluster we considered the proportion of observations that belong to the first cluster for all the variables.

In Figure 3 the profiles of the two clusters are depicted for each model fitted, i.e. the mean for all the clients assigned to each cluster. The left-hand side plot corresponds to the 2-FMBP without covariates in the mixing proportion, while the right-hand side plot corresponds to the model with *V10* and *V11* as covariates in the mixing proportion. The dotted line represents the first cluster while the solid line represents the second. For the left-hand side plot, the main differences occur for variables *V10* and *V11* while some small difference is found for *V3*. In simple terms, these variables can be used to distinguish between the two clusters. Interestingly, these variables also have different signs in their regression coefficients for the two components (see Table 3). They are also the only variables that are statistically significant for both components and both response variables. For all the other variables, the profiles are the same which indicates their unsuitability for characterizing the clusters.

After including *V10* and *V11* as covariates in the mixing proportion, it can be seen that all the information regarding *V11* is now included within the mixing proportion parameter p. In other words, the p does not differ for the two components but it is significant when selecting the component. Thus, if we do not use *V11* in the p then the differences will be apparent in the means. By contrast, the inclusion of the comprehensive coverage (except fire) (*V10*) variable still characterizes the clusters. Differences exist for the other variables but are smaller.

Note also the differences in interpretation afforded by the two models. By using covariates in the mixing proportion, we model the effect of the covariate explicitly to the choice of component, while when using them only in the mean of the component we do so implicitly. It also helpful to consider how covariates directly affect the probability of each customer belonging to a group.

5. Concluding Remarks

We have proposed a new model of finite mixture of bivariate Poisson regressions. The idea is that the data consist of subpopulations of different regression structures. A potential use for such a model is for examining the clustering of observations, taking into consideration the effect of certain covariates while also taking into account the dependence between the response

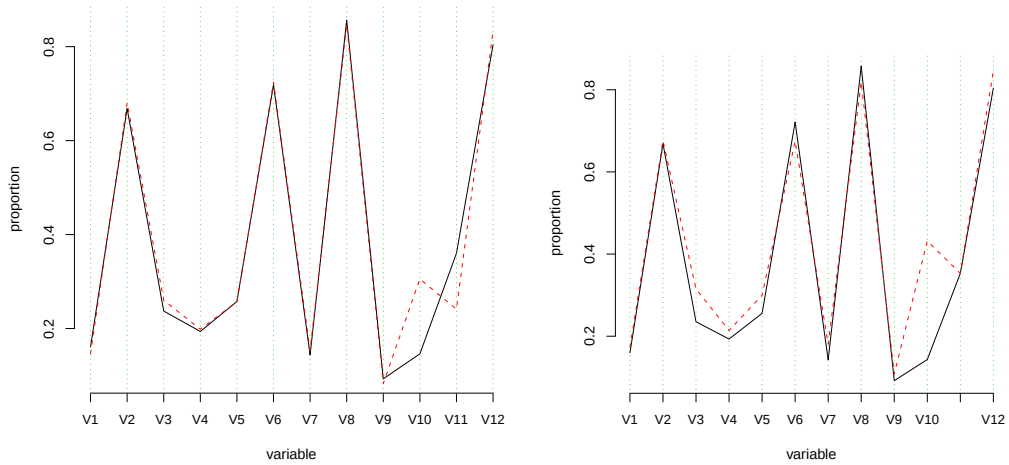

Figure 3: The profiles of the two clusters considered for 2-FMBP1 and 2-FMBP2 models

variables. The model corrects for the zero inflation and overdispersion present in the real automobile insurance data set used in the application. The model can also be used to model negative correlation.

The AIC reported here indicates that the 2-finite mixture of bivariate Poisson regression with covariates in the mixing proportion is the best model for describing the data set. This model has a number of interesting features:

Firstly being a finite mixture it has a nice cluster interpretation. Secondly it is semi-parametric, in the sense that we avoid a parametric assumption on the mixing distribution. Thirdly is flexible enough allowing for overdispersion, zero inflation but also for negative correlation. Finally, our EM algorithm makes its fit straightforward while for other models their fitting procedure is not easy for large samples and many covariates as in our case. Such an example is the bivariate Poisson lognormal model which needs the evaluation of a bivariate integral for each observation, and of course each iteration during maximization which leads to a lot of computational problems. Concluding the proposed model provides very interesting structure while being computationally feasible.

The problem of overdispersion arises because of the presence of unobserved heterogeneity in many real data sets. In insurance data sets, an in-

surance company cannot keep track of the many differences between policyholders. However, the model proposed in this paper accounts for unobserved heterogeneity by choosing a finite number of subpopulations. We assume the existence of two types of policyholder described separately according to each component in the mixture.

The phenomenon of excess of zeros may also be seen as a consequence of this unobserved heterogeneity. The model proposed here, as a finite mixture of bivariate Poisson regression model, embraces the zero-inflated bivariate Poisson regression model as a special case. The main difference with zero-inflated models is that the two-component mixture model reported here allows mixing with respect to both zeros and positives. This interpretation is more flexible and holds better in our application. The group separation is characterized by low mean (policyholders considered as a "good" drivers) and high mean (policyholders considered as a "bad" drivers).

Moreover, as it seems that the data set may have been generated from two distinct subpopulations, the model allows for a net interpretation of each cluster separately. Note that different regression coefficients can be used to account for the "observed" heterogeneity within each population.

Finally, we would like to mention various ways in which this paper might be extended. Although in the present paper we limit our analysis to the bivariate case, it could be extended to include larger dimensions. Following the general model presented by Karlis and Meligkotsidou (2007), covariates might be added and this finite mixture of multivariate Poisson regressions could be used to cluster high-dimensional data. A particularly interesting case occurs if we consider there to be no dependence within a cluster, whereby within-cluster independent Poisson regressions are considered.

To conclude this section, we should point out that the one of the limitations of the bivariate Poisson model is that it allows only for positive dependence within each component, owing to the properties of the bivariate (multivariate) Poisson distribution. To overcome this shortcoming, other bivariate models, such as the copula-based models defined in Nikoloulopoulos and Karlis (2010), might be considered as the component specific bivariate distributions.

Acknowledgements. Research for this paper was initiated while the second author was visiting the "Risk in Finance and Insurance" Research Group at the University of Barcelona. The authors would like to thank referees for

interesting comments that led to improvement from the previous version.

AppendixA. Zero inflation in mixed bivariate Poisson distributions

Lemma: Mixed bivariate Poisson distributions always give equal or greater probability to the (0,0) cell from the corresponding bivariate Poisson with the same marginal means.

Proof: It is straightforward to see that any mixed bivariate Poisson distribution has an excess of zeros compared to the bivariate Poisson distribution with the same marginal means. This result generalizes the known property in one dimension (Shaked's Two Crossings Theorem). To demonstrate this, consider for sake of simplicity the 2-finite bivariate Poisson mixture, with probability p and $(1-p)$ to the points $(\lambda_{11}, \lambda_{21}, \lambda_{31})$ and $(\lambda_{12}, \lambda_{22}, \lambda_{32})$. The marginal means are $p(\lambda_{11}+\lambda_{31})+(1-p)(\lambda_{12}+\lambda_{32})$ and $p(\lambda_{21}+\lambda_{31})+(1-p)(\lambda_{22}+\lambda_{32})$ respectively. Consider also the bivariate Poisson with the same marginal means. Under the 2-finite mixture case the $(0,0)$ probability is given by

$$P_2(0,0) = p\exp\left(-(\lambda_{11}+\lambda_{21}+\lambda_{31})\right) + (1-p)\exp\left(-(\lambda_{12}+\lambda_{22}+\lambda_{32})\right)$$

or

$$P_2(0,0) = p\exp(-\Lambda_1) + (1-p)\exp(-\Lambda_2),$$

while for the bivariate Poisson we have

$$P_{BP}(0,0) = \exp\left(-(p\Lambda_1) + (1-p)\Lambda_2\right).$$

By considering the random variable Q that takes value $-\Lambda_1$ and $-\Lambda_2$ with probabilities p and $1-p$ and considering the Jensen's inequality we have that

$$E(\exp(Q)) \geq \exp(E(Q))$$

and thus

$$P_2(0,0) \geq P_{BP}(0,0).$$

Thus, this mixing of this kind also results in zero inflation. The above result can be readily generalized to an infinite number of components as well as to more than two dimensions.

AppendixB. Some of the moments for the 2-finite mixture of bivariate Poisson distribution

It can readily be obtained that

$$E(Y_k) = \sum_{j=1}^{m} p_j(\lambda_{kj} + \lambda_{3j}),$$

$$E(Y_k^2) = \sum_{j=1}^{m} p_j \left[\lambda_{kj} + \lambda_{3j} + (\lambda_{kj} + \lambda_{3j})^2\right],$$

$$Var(Y_k) = E(Y_k^2) - \left[E(Y_k)\right]^2,$$

$$E(Y_1Y_2) = \sum_{j=1}^{m} p_j \left[\lambda_{3j} + (\lambda_{1j} + \lambda_{3j})(\lambda_{2j} + \lambda_{3j})\right],$$

and

$$Cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2).$$

For actuarial purposes one may well be interested in quantities such as $E(Y_1 + Y_2)$ and/or $Var(Y_1 + Y_2)$ (see, e.g. Bermúdez (2009)). These can be easily obtained from the above formulas.

Aitchinson, J., Ho, C., 1989. The multivariate Poisson-log normal distribution. Biometrika 75, 621–629.

Bermúdez, L., 2009. A priori ratemaking using bivariate Poisson regression models. Insurance: Mathematics and Economics 44, 135–141.

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

Bolancé, C., Guillén, M., Pinquet, J., 2003. Time-varying credibility for frequency risk models: Estimation and tests for autoregressive specifications on the random effects. 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.

Boucher, J.P., Denuit, M., 2008. Credibility premiums for the zero inflated Poisson model and new hunger for bonus interpretation. Insurance: Mathematics and Economics 42, 727–735.

Boucher, J.P., Denuit, M., Guillén, 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., Denuit, M., Guillén, M., 2009. Number of accidents or number of claims? an approach with zero-inflated Poisson models for panel data. Journal of Risk and Insurance 76, 821–846.

Boucher, J.P., Santolino, M., 2010. Discrete distributions when modeling the disability severity score of motor victims. Accident Analysis and Prevention 42, 2041–2049.

Brijs, T., Karlis, D., Swinnen, G., Vanhoof, K., Wets, G., Manchanda, P., 2004. A multivariate Poisson mixture model for marketing applications. Statistica Neerlandica 58, 322–348.

Brouhns, N., Guillén, M., Denuit, M., Pinquet, J., 2003. Bonus-malus scales in segmented tariffs with stochastic migration between segments. Journal of Risk and Insurance 70, 577–599.

Chib, S., Winkelmann, R., 2001. Markov Chain Monte Carlo Analysis of Correlated Count Data. Journal of Business and Economic Statistics 19, 428–435.

Denuit, M., Marechal, X., Pitrebois, S., Walhin, J.F., 2007. Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems. Wiley, New York.

Garay, A., Hashimoto, E., Ortega, E., Lachos, V., 2011. On estimation and influence diagnostics for zero-inflated negative binomial regression models. Computational Statistics and Data Analysis 55, 1304 – 1318.

Grun, B., Leisch, F., 2007. Fitting finite mixtures of generalized linear regressions in R. Computational Statistics and Data Analysis 51, 5247 – 5252.

Gurmu, S., Elder, J., 2000. Generalized bivariate count data regression models. Econometric Letters 68, 31–36.

Gurmu, S., Elder, J., 2008. A bivariate zero-inflated count data regression model with unrestricted correlation. Economics Letters 100, 245–248.

Johnson, N., Kotz, S., Balakrishnan, N., 1997. Multivariate Discrete Distributions. Wiley, New York.

Karlis, D., Meligkotsidou, L., 2007. Finite multivariate Poisson mixtures with applications. Journal of Statistical Planning and Inference 137, 1942–1960.

Karlis, D., Ntzoufras, L., 2003. Analysis of sports data by using bivariate Poisson models. Journal of the Royal Statistical Society Series D: The Statistician 52, 381–393.

Kocherlakota, S., 1988. On the compounded bivariate Poisson distribution: A unified treatment. Annals of the Institute of Statistical Mathematics 40, 61–76.

Kocherlakota, S., Kocherlakota, K., 1992. Bivariate Discrete Distributions, Statistics: textbooks and monographs. volume 132. Markel Dekker, New York.

Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14.

Li, C.S., Lu, J.C., Park, J., Kim, K., Brinkley, P., Peterson, J., 1999. Multivariate zero-inflated Poisson models and their applications. Technometrics 41, 29–38.

Lord, D., Washington, S., Ivan, J., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. Accident Analysis and Prevention 37, 35–46.

Lord, D., Washington, S., Ivan, J., 2007. Further notes on the application of zero-inflated models in highway safety. Accident Analysis and Prevention 39, 53–57.

Munkin, M., Trivedi, P., 1999. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. Econometrics Journal 2, 29–48.

Nikoloulopoulos, A.K., Karlis, D., 2010. Regression in a copula model for bivariate count data. Journal of Applied Statistics 37, 1555–1568.

Park, B.J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accident Analysis and Prevention 41, 683–691.

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.

Stein, G., Yuritz, J., 1987. Bivariate compound Poisson distributions. Communications in Statistics -Theory and Methods 16, 3591–3607.

Stein, G., Zucchini, W., Juritz, J., 1987. Parameter estimation for the Sichel distribution and its multivariate extension. Journal of the American Statistical Association 82, 938–944.

Steyn, H., 1976. On the multivariate Poisson normal distribution. Journal of the American Statistical Association 71, 233–236.

Wang, P., 2003. A bivariate zero-inflated negative binomial regression model for count data with excess zeros. Economics Letters 78, 373–378.

Wang, P., Cockburn, I., Puterman, M., 1998. Analysis of patent data: a mixed Poisson regression model approach. Journal of Business and Economic Statistics 16, 27–36.

Winkelmann, R., 2008. Econometric Analysis of Count Data, 4th edition. Springer, New York.