Multivariate claim count regression model with varying dispersion and dependence parameters

Citation for published version:
Jeong, H, Tzougas, G & Fung, TC 2023, 'Multivariate claim count regression model with varying dispersion and dependence parameters', *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 186, no. 1, pp. 61-83. https://doi.org/10.1093/jrsssa/qnac010

Digital Object Identifier (DOI):
10.1093/jrsssa/qnac010

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Journal of the Royal Statistical Society Series A: Statistics in Society

Multivariate claim count regression model with varying dispersion and dependence parameters

Himchan Jeong*, George Tzougas†, Tsz Chai Fung‡

December 2022

Abstract

The aim of this paper is to present a regression model for multivariate claim frequency data with dependence structures across the claim count responses, which may be of different sign and range, and overdispersion from the unobserved heterogeneity due to systematic effects in the data. For illustrative purposes, we consider the bivariate Poisson-lognormal regression model with varying dispersion. Maximum likelihood estimation of the model parameters is achieved through a novel Monte Carlo Expectation-Maximization algorithm, which is shown to have a satisfactory performance when we exemplify our approach to Local Government Property Insurance Fund data from the state of Wisconsin.

Keywords: Multivariate claim frequency modelling; Overdispersion; Correlations of different signs and magnitude; Copulas; Regression models for the mean, dispersion and dependence parameters; Bivariate Poisson-lognormal regression model with varying dispersion; Monte Carlo Expectation-Maximization algorithm

Running head: MULTI-COUNT MODEL WITH VARYING DEPENDENCE

1 Introduction

Over the last few decades, there has been a vast increase in the literature concerning multivariate count regression models that can permit inferences about the dependence structure between multiple response variables based on covariate information. In particular, the number of possible ways in which univariate count regression models have been generalized to their multivariate counterparts is not exhaustive, as is evident from the plethora of their applications in several distinct fields, such as, for instance, sociology, biology, biometrics, genetics, medicine, marketing, epidemiology, marketing, criminology, engineering and insurance. An interested reader is referred to the recent editions of the book by Winkelmann [2008] and the latest edition of the book of Cameron and Trivedi [2013] for thorough reviews of regression models for multivariate count data. Broadly speaking, among the multivariate count regression models, those that have been growing in importance in recent years in both theoretical and applied work can be categorized into three main classes:

(i) Multivariate Poisson models.

(ii) Multivariate mixed Poisson models.

(iii) Copula-based models.

The literature on multivariate Poisson (MVP) distributions, which belong to class (i), started growing nine decades ago, see M'Kendrick [1925], and has now reached solid and extensive foundations. Many references for MVP distributions, including historical remarks, can be found in Johnson et al. [1997] and Krummenauer [1998]. In the context of regression analysis, different versions of the MVP regression model

*Address: Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada. Email: himchan_jeong@sfu.ca

†Address: Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, EH14 4AS, UK. George.Tzougas@hw.ac.uk

‡Address: Department of Risk Management and Insurance, 35 Broad Street NW, Georgia State University, Atlanta, GA 30303, USA. Email: tfung@gsu.edu.

have also been studied by many authors. The most basic bivariate Poisson regression model, see, for example, Jung and Winkelmann [1993], Ho and Singer [2001] and Kocherlakota and Kocherlakota [2001], can be derived based on the trivariate reduction method, see Kocherlakota [1992]. Furthermore, the MVP regression model can be obtained by using a general multivariate reduction scheme, see, for example, Winkelmann [2008]. An extension of the MVP regression model, which permits a larger covariance structure between the variables, can be found in Karlis and Meligkotsidou [2005]. The aforementioned MVP models allow for positive dependencies between variables and have the advantage that they have closed-form densities. Finally, a bivariate Poisson model that is not based on trivariate reduction and can handle negative, zero or positive correlations was introduced by Lakshminarayana et al. [1999], who defined the distribution from the product of Poisson marginals with a multiplicative factor parameter, see also Famoye [2010] who considered the same model for the case with covariates.

Multivariate mixed Poisson (MVMP) models, which belong to class (ii), represent an important class of models that can allow for flexible correlation structures, as well as overdispersion by including unobserved heterogeneity variables in the independent Poisson marginals following some mixing distributions, see, for example, Cameron and Trivedi [2013]. Within the MVMP models, i.e., class (ii) above, we distinguish two different types which have been systematically studied in the statistical literature:

(a) When a shared random effect is distributed according to a univariate continuous mixing distribution.

(b) When multiple random effects are distributed according to a multivariate continuous mixing distribution.

The models of type (a) allow for positive dependencies between variables. The literature along this line includes, for example, the works of Stein and Juritz [1987], Stein et al. [1987] and Kocherlakota [1988] for the case without explanatory variables. Also, Munkin and Trivedi [1999], Gurmu and Elder [2000] and Ghitany et al. [2012] considered regression specifications for the marginal means. Specifically, Stein and Juritz [1987] developed the bivariate negative binomial distribution using a gamma distribution and four bivariate versions of the Sichel (Poisson-generalized inverse Gaussian) distribution, while the multivariate Sichel distribution was studied in Stein et al. [1987]. Kocherlakota [1988] described the bivariate versions of the negative binomial, Neyman type A, Hermite and Poisson-inverse Gaussian distributions respectively. Munkin and Trivedi [1999] estimated a bivariate negative binomial regression model based on the simulated likelihood (SML) method. Gurmu and Elder [2000] presented an alternative estimation approach for the bivariate negative binomial regression model based on a series expansion of the random effects distribution, where the leading term in the expansion was a gamma density. Ghitany et al. [2012] developed an Expectation-Maximization (EM) algorithm for maximum likelihood (ML) estimation of MVMP regression models of type (a). They demonstrated the feasibility of their approach with a multivariate negative binomial (MVNB), a multivariate Poisson-inverse Gaussian (MVPIG), and a multivariate Poisson-lognormal (MVPLN) regression model respectively.

Regarding models of type (b), the MVPLN model, which can be derived using the multivariate lognormal mixing distribution, see Aitchison and Ho [1989], instead of its univariate version in type (a) approach, has been widely studied in the literature as it can cope with both positive and negative dependencies. On the other hand, similarly to the MVPLN model of type (a), the estimation of the MVPLN model of type (b) involves numerical integration, because both models do not have their densities in closed from. However, the numerical integration that is required to evaluate the MVPLN likelihood increases the complexity of the estimation procedure under type (b) approach. It is because when the number of dimensions increases, so does the order of numerical integration involved, which in practice can be more time consuming than type (a) approach, especially for very large high-dimensional datasets when covariates are available. Nevertheless, regardless of the high computational intensity as mentioned in Cameron and Trivedi [2013], high-dimensional integrals can be efficiently evaluated using either deterministic or Monte Carlo integration, with the latter method preferred as the number of dimensions increases. These issues were addressed also within the context of regression techniques, for example, Munkin and Trivedi [1999] computed the ML estimates of the parameters of the bivariate Poisson-lognormal (BVPLN) regression model based on the SML procedure. Also, Chib and Winkelmann [2001] extended the previous model by resorting to the Bayesian Markov chain Monte Carlo (MCMC) ML estimation approach, which can deal with a high-dimensional outcome vector, to fit the MVPLN regression model. Additionally, many authors estimated the parameters of several extensions of the MVPLN regression model within the Bayesian paradigm, see, for instance, Park and Lord [2007], Ma et al. [2008], El-Basyouny and Sayed [2009], Aguero-Valverde and Jovanis [2009] and Zhan et al. [2015]. Additionaly, Silva et al. [2019] estimated the parameters of a finite MVPLN mixture regression model via a Markov chain Monte Carlo Expectation-Maximization (MCMC-EM) algorithm. Recently, Chiquet et al. [2020] proposed a variational EM algorithm for ML estimation of the MPLN random effects model.

However, despite its popularity, the caveat with the MPLN model in type (b) approach is that it lacks generality in correlation structure in the the sense that the covariance matrix of the Gaussian distributed latent vector describes the underlying structure of dependence between the multiple responses.

Finally, another more recently developed approach is to model dependence between multiple responses through the use of copula-based models, which belong to class (iii). A multivariate count distribution can be viewed as a continuous copula distribution paired with discrete marginals. The copula functions can fully specify the dependence structure separately from the univariate marginals see, for example, Joe [1997]. The literature on copula-based regression models includes, among others, Lee [1999], Cameron et al. [2004], Nikoloulopoulos and Karlis [2010], Nikoloulopoulos [2013], and Nikoloulopoulos [2016]. Lee [1999] and Cameron et al. [2004] considered bivariate copula-based regression models. Nikoloulopoulos and Karlis [2010] presented a series of different multivariate copulas of varying dependence structure for modelling vectors of count response variables whose marginal distributions depend on covariates through negative binomial regressions. Nikoloulopoulos [2013] explored the asymptotic and small-sample efficiency of the SML method for the multivariate normal (MVN) copula with univariate binary Poisson and negative binomial regressions. Nikoloulopoulos [2016] studied the estimation of the MVN copula model with discrete responses via SML method which was proposed in the above study. The main advantage of copula-based models is that they can cope with both positive and negative dependence structures. A review of employing copulas to specify correlation structures can be found in Chen and Hanson [2017]. Nevertheless, the computational burden for copula models paired with discrete marginal distributions becomes cumbersome if the dimension of the model increases. This is discussed, for instance, in Zimmer and Trivedi [2006] and Genest and Nešlehová [2007]. However, simplifications and computationally efficient estimation methods, can be used to overcome these issues, see Rüschendorf [2013], Nikoloulopoulos [2013] and Nikoloulopoulos [2016]. Finally, Marra and Wyszynski [2016] proposed semi-parametric copula sample selection models for count responses, which consist of two regressions, a binary selection equation, and an outcome equation, to model discrete response. Their approach allows for the use of several dependence structures, potentially any discrete outcome margin and for flexible covariate effects. Furthermore, parameter estimation was based on the penalized likelihood estimation framework.

As far as modelling multivariate claim count data in non-life insurance, which is the main research focus of this study, is concerned, it is quite common to observe the existence of dependence structures between different types of claim counts either from the same type of coverage, such as, for example, bodily injury and property damage claims in motor third party liability (MTPL) insurance, or claim counts from multiple types of coverage, such as, for instance, motor and home insurance bundled into one single policy. For more details, the interested reader can refer, for instance, to the articles by Bermúdez and Karlis [2011], Bermúdez and Karlis [2012], Shi and Valdez [2014], Abdallah et al. [2016], Bermúdez and Karlis [2017], Bermúdez et al. [2018], Pechon et al. [2018], Pechon et al. [2019], Bolancé and Vernic [2019], Denuit et al. [2019], Fung et al. [2019] and Bolancé et al. [2020] who introduced different multivariate count regression, based on a single or multiple random effects, and copula based models. Recently, Jeong and Dey [Forthcoming] considered a multivariate quasi-Poisson regression model with shared gamma random effect to capture positive but heterogeneous associations among the claims from different lines of business while Pechon et al. [2021] derived a multivariate credibility premium for home and motor insurance joined at a household level using the MVPLN model.

At this point, it should be noted that modelling the joint dynamics of different claim types and their associated counts is far from straightforward since there are systematic effects in the data due to unobserved heterogeneity, which gives rise to overdispersion, and because the correlations between claims from different types of perils may be of different signs and magnitude. Furthermore, the explanatory power of predictors, which are considered in a regression analysis of multivariate claim count data, should always be investigated as it is very important to understand, if in addition to the mean components, which are traditionally modelled using covariate information, the dispersion and dependence components are influenced by different individual and coverage-type risk characteristics.

Therefore, even if the literature on models of types (i), (ii) and (iii) is abundant, as it can be clearly understood, there is no guarantee that these modelling approaches will always adequately describe diverse situations in alternative high-dimensional claim count datasets. In particular, the three main classes of multivariate claim count regression models have the following limitations, which should be taken into consideration when analyzing the relationship between explanatory variables and a set of claim count variables:

- The model specification of type (i) is too restrictive for multivariate claim count data since it does not take into account the overdispersion phenomenon, which is a problem that was inherited from the univariate case.

- The model specification of type (ii), category (a) allows only positive correlation between multiple types of claims but in some cases, negative correlations may be of interest as well.
- The model specification of type (ii), category (b) implies that one cannot use coverage specific covariate information to model the mean components and dependence among claim types from multiple types of insurance coverage. Furthermore, this model specification does not allow for varying dispersions or dependence upon individual characteristics and the characteristics of the coverage types.
- The model specification of type (iii) may not fully specify the dependence structure, since as opposed to the case with continuous marginals, identifiability issues can arise when a continuous copula distribution is paired with discrete marginals, see Genest and Nešlehová [2007]. Further, the density function of a copula with discrete marginals usually involves both summation and integration so that it suffers from computational burden due to multivariate numerical integration [Oh et al., 2021].

At the same time, the prediction accuracy of any model is of vital importance, as this will determine how useful the model is for a wide range of actuarial applications, such as risk management and pricing of (re-)insurance contracts. Thus, a critical task of actuaries is to construct more representative models that can capture all the important aspects of multivariate claim count data since the accuracy of their predictive performance depends on the reliability of the statistical method which was used to develop them.

In this study, we introduce a multivariate claim count regression model with varying dispersion and dependence parameters for modelling claim counts jointly from multiple types of claims based on the use of available covariate information. The proposed modelling framework is suitable for addressing all the challenges which may often be encountered in high-dimensional non-life insurance datasets. In particular, we make the following contributions:

- The model is constructed based on a mixing between multiple marginal Poisson distributions and unit mean continuous prior, or mixing distributions which are combined with a copula distribution to form a joint multivariate continuous mixing density.
- The unit mean requirement for the continuous mixing densities ensures that the marginal mixed Poisson models are identifiable see, for example, Karlis [2001], Ghitany et al. [2012], Barreto-Souza and Simas [2016] and Tzougas [2020], among many others. The margins are continuous and hence the corresponding copula is unique [Sklar, 1973] so that one can effectively optimize the joint likelihood by avoiding having finite differences in the likelihood.
- In order to illustrate the versatility of our approach, we consider the bivariate Poisson-lognormal (BPLN) regression model with varying dispersion and dependence, extending the setup of Ghitany et al. [2012] and Chiquet et al. [2020] who, as was previously mentioned, considered versions of the model which belong to class (ii), in categories (a) and (b) respectively.
- The proposed modelling framework has sufficient flexibility for accommodating multivariate overdispersion and accounting for wide range of dependence, allowing both positive and negative correlation between the different types of claims counts.
- Regressors are allowed on the mean, dispersion and dependence parameters of the model in order to fully exploit the influence of important individual and coverage type covariate information on the mean and dispersion components and the dependence structure among multiple claim count responses. Thus, this approach enhances the predictive performance of the model.
- Finally, we develop a novel Monte Carlo Expectation-Maximization (MCEM) algorithm for estimating the parameters of the BPLN regression model with varying dispersion and dependence parameters, which is easily implementable and is demonstrated to perform well when the model is fitted to the Local Government Property Insurance Fund (LGPIF) data from the state of Wisconsin.

The remainder of this paper is organized as follows. In Section 2, we present the derivation of the proposed multivariate claim count regression model with varying dispersion and dependence parameters. Section 3 describes the ML estimation procedure for our proposed model via the MCEM algorithm. Section 4 contains an application to the LGPIF dataset that we use for our empirical analysis and we fit the BPLN claim count regression model with varying dispersion and dependence parameters, followed by estimation results and model comparison for the BPLN regression models of Ghitany et al. [2012] and Chiquet et al. [2020] that we use as benchmarks for comparison. Finally, concluding remarks are given in Section 5.

2 Multivariate mixed Poisson regression model with varying dispersion and dependence

2.1 Modelling framework

Suppose that there are n policyholder contracts, each of which involves in L types of claims (or perils). Denote $\boldsymbol{N}_i = (N_i^{(1)}, \ldots, N_i^{(L)})$ and $\boldsymbol{n}_i = (n_i^{(1)}, \ldots, n_i^{(L)})$ respectively as the number of claims vector (for each of the L claim types) and its corresponding realizations. Corresponding to each contract, several explanatory variables $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iP})$ are available for us to analyze the observed heterogeneities of policyholder's risk profiles. The policyholder contracts are assumed to be independent of each other.

Suppose that given a continuous random variable $Z_i^{(l)} > 0$, $N_i^{(l)}|Z_i^{(l)}$ follows a Poisson distribution with probability mass function (pmf) given by

$$p\left(n_i^{(l)}|z_i^{(l)}\right) = \frac{(\mu_i^{(l)} z_i^{(l)})^{n_i^{(l)}} e^{-\mu_i^{(l)} z_i^{(l)}}}{n_i^{(l)}!}, \tag{1}$$

where $\mu_i^{(l)} > 0$ is the mean parameter and $z_i^{(l)} > 0$ is the realization of latent variable $Z_i^{(l)}$ with $\mathbb{E}[N_i^{(l)}|Z_i^{(l)}] = Var(N_i^{(l)}|Z_i^{(l)}) = \mu_i^{(l)} Z_i^{(l)}$ for $l = 1, \ldots, L$. Here, L is total number of alternative types of claims. Denote $F_l(z_i^{(l)})$ as the marginal distribution of $Z_i^{(l)}$ which is parameterized by the dispersion parameter $\sigma_i^{(l)}$. We also assume that $\mathbb{E}[Z_i^{(l)}] = 1$ for the sake of model identifiability.

The dependence of $\boldsymbol{N}_i$ among the claim types is modelled through the dependence of the latent variables $\boldsymbol{Z}_i := (Z_i^{(1)}, \ldots, Z_i^{(L)})$ using a copula, with the joint distribution of $\boldsymbol{Z}_i$ given by

$$\pi(\boldsymbol{z}_i) = \prod_{l=1}^{L} f_l(z_i^{(l)}) \times c_{\boldsymbol{\Phi}_i}\left(F_1(z_i^{(1)}),\, \ldots,\, F_L(z_i^{(L)})\right), \tag{2}$$

where $\boldsymbol{z}_i = (z_i^{(1)}, \ldots, z_i^{(L)})$, f_l and F_l are marginal density and distribution functions of $z^{(l)}$, respectively. $c_{\boldsymbol{\Phi}_i}$ is a copula density function that models dependence among the latent variables. While it is possible to use an arbitrary family of copulas to model the joint distribution, we use elliptical copulas as candidates of $c_{\boldsymbol{\Phi}_i}$ in order to consider flexible magnitudes and signs of associations among the latent variables as in Frees and Wang [2006]. In this case, $\boldsymbol{\Phi}_i := \{\phi_i^{(l,l')}\}_{l,l'=1,\ldots,L}$ is a symmetric parameter matrix influencing the correlations among $\boldsymbol{Z}_i$.

Now, the joint pmf of $(N_i^{(1)}, \ldots, N_i^{(L)})$ is given by the following:

$$p\left(n_i^{(1)}, \ldots, n_i^{(L)}\right) = \int \prod_{l=1}^{L} p\left(n_i^{(l)}|z_i^{(l)}\right) \pi(\boldsymbol{z}_i) d\boldsymbol{z}_i. \tag{3}$$

Note that if C_{ϕ_i} is an independent copula, then $p\left(n_i^{(1)}, \ldots, n_i^{(L)}\right) = \prod_{l=1}^{L} p\left(n_i^{(l)}\right)$ so that the number of claims from different types of perils are assumed to be independent. While one can directly model dependence among the count variables, here we do not use a copula model with discrete marginals because it has been known that optimization of marginal likelihood for the copula with discrete marginals involves summation with sign, which usually ends up with substantial computation time and numerical instability in optimization. Further, unidentifiability issue often arises in copula models directly applied to discrete distributions (Genest and Nešlehová [2007]), hindering model interpretability and statistical tractability (e.g. quantification of parameter uncertainties).

To allow for the mean, dispersion and dependence parameters to be modelled as functions of explanatory variables with parametric linear functional forms, we assume that

$$\mu_i^{(l)} = \exp\left(\boldsymbol{x}_{1,i}^{(l)T} \boldsymbol{\beta}_1^{(l)}\right), \tag{4}$$

$$\sigma_i^{(l)} = \exp\left(\boldsymbol{x}_{2,i}^{(l)T} \boldsymbol{\beta}_2^{(l)}\right) \text{ and} \tag{5}$$

$$\phi_i^{(l,l')} = g(\boldsymbol{x}_{3,i}^{(l,l')T} \boldsymbol{\beta}_3^{(l,l')}), \tag{6}$$

where $\boldsymbol{x}_{1,i}^{(l)}$, $\boldsymbol{x}_{2,i}^{(l)}$ and $\boldsymbol{x}_{3,i}^{(l,l')}$ are covariate vectors being (potentially different) subsets of $\boldsymbol{x}_i$ with dimensions $P_1^{(l)} \times 1$, $P_2^{(l)} \times 1$ and $P_3^{(l,l')} \times 1$ respectively for $l, l' = 1, \ldots, L$, and $\boldsymbol{\beta}_1^{(l)} = \left(\beta_{1,1}^{(l)}, ..., \beta_{1,P_1^{(l)}}^{(l)}\right)^T$, $\boldsymbol{\beta}_2^{(l)} =$

$\left(\beta_{2,1}^{(l)},...,\beta_{2,P_2^{(l)}}^{(l)}\right)^T$ and $\boldsymbol{\beta}_3^{(l,l')} = \left(\beta_{3,1}^{(l,l')},...,\beta_{3,P_3^{(l,l')}}^{(l,l')}\right)^T$ are the corresponding parameter vectors. We also denote the $\boldsymbol{X}_1^{(l)}$, $\boldsymbol{X}_2^{(l)}$ and $\boldsymbol{X}_3^{(l,l')}$ as the designed matrices with rows given by $\boldsymbol{x}_{1,i}^{(l)}$, $\boldsymbol{x}_{2,i}^{(l)}$ and $\boldsymbol{x}_{3,i}^{(l,l')}$ respectively, and assume that they are of full rank, for $i = 1,...,n$ and $l, l' = 1, \ldots, L$. Further, $g(\cdot)$ is a monotone link function which depends on the family of copula used by the modeler.

Our proposed multivariate mixed Poisson regression model has a natural model interpretation in insurance perspective. The latent variables $\boldsymbol{Z}_i$ represent the unobserved policyholder information during insurance underwriting process, and capture the unobserved heterogeneities among different policyholders. We assume that the unobserved heterogeneities fully drive the over-dispersion and dependence of claim counts across different perils.

2.2 Model properties

In this section we discuss some desirable properties which ensure the flexibility and mathematical tractability of the proposed model.

Property 1 *(Moment property) The expectation, variance, covariance and correlations of the number of claims are given by*

$$\mathbb{E}[N_i^{(l)}] = \mu_i^{(l)}, \quad Var(N_i^{(l)}) = \mu_i^{(l)} + {\mu_i^{(l)}}^2 Var(Z_i^{(l)}), \quad Cov(N_i^{(l)}, N_i^{(l')}) = \mu_i^{(l)}\mu_i^{(l')}Cov(Z_i^{(l)}, Z_i^{(l')}), \quad (7)$$

$$\begin{aligned} Corr(N_i^{(l)}, N_i^{(l')}) &= \frac{Cov(Z_i^{(l)}, Z_i^{(l')})}{\sqrt{\left(1/\mu_i^{(l)} + Var(Z_i^{(l)})\right)\left(1/\mu_i^{(l')} + Var(Z_i^{(l')})\right)}} \\ &= \frac{Corr(Z_i^{(l)}, Z_i^{(l')})}{\sqrt{\left(1/\left[\mu_i^{(l)}Var(Z_i^{(l)})\right] + 1\right)\left(1/\left[\mu_i^{(l')}Var(Z_i^{(l')})\right] + 1\right)}}, \end{aligned} \quad (8)$$

and hence we have: (i) $sign(Corr(N_i^{(l)}, N_i^{(l')})) = sign(Corr(Z_i^{(l)}, Z_i^{(l')}))$; and (ii) $Corr(N_i^{(l)}, N_i^{(l')}) = \pm 1$ iff $Corr(Z_i^{(l)}, Z_i^{(l')}) = \pm 1$, $\mu_i^{(l)}Var(Z_i^{(l)}) \to \infty$ and $\mu_i^{(l')}Var(Z_i^{(l')}) \to \infty$.

The above property shows ability of the proposed model to capture both positive and negative dependence of claim counts across claim types. Also, it gives the necessary conditions such that the model will be flexible enough to capture a full range of dependence. This provides a guidance on specifying the distributions classes of copula model $C_{\boldsymbol{\Phi}_i}$ and latent variables $\boldsymbol{Z}_i$.

Property 2 *(Marginalization property) The proposed joint pmf is closed to marginalization, i.e. the marginal distribution of $N_i^{(l)}$ is given by*

$$p(n_i^{(l)}) = \int p(n_i^{(l)}|z_i^{(l)})dF_l(z_i^{(l)}), \quad (9)$$

which is a univariate mixed Poisson regression model with varying dispersion. In general, any L'-variate response marginal (with $L' < L$) is still an L'-variate mixed Poisson regression model with varying dispersion and dependence.

Proof. For conciseness we only derive the results for univariate case.

$$\begin{aligned} p(n_i^{(l)}) &= \sum_{n_i^{(1)}=0}^{\infty} \cdots \sum_{n_i^{(l-1)}=0}^{\infty} \sum_{n_i^{(l+1)}=0}^{\infty} \cdots \sum_{n_i^{(L)}=0}^{\infty} p\left(n_i^{(1)}, \ldots n_i^{(L)}\right) \\ &= \int \left[\prod_{j\neq l}^{L} \sum_{n_i^{(j)}=0}^{\infty} p\left(n_i^{(j)}|z_i^{(j)}\right)\right] p(n_i^{(l)}|z_i^{(l)})\pi(\boldsymbol{z}_i)d\boldsymbol{z}_i \\ &= \int p(n_i^{(l)}|z_i^{(l)})\pi(\boldsymbol{z}_i)d\boldsymbol{z}_i = \int p(n_i^{(l)}|z_i^{(l)})dF_l(z_i^{(l)}). \end{aligned} \quad (10)$$

■

The above marginalization property enables insurers to easily price the insurance contracts for which the policyholder only participates in some but not all lines of businesses (claim types). We finally present the identifiability of the proposed model, ensuring uniqueness of the regression model and avoiding the issue of multiple interpretations on the same model.

Property 3 *(Identifiability of the joint distribution) Suppose the following conditions hold for $l, l' = 1, \ldots, L$:*

(i) $Var(Z_i^{(l)})$ is a monotonic strictly increasing or decreasing function of the dispersion parameter $\sigma_i^{(l)}$ only; and

(ii) $Corr(Z_i^{(l)}, Z_i^{(l')})$ is a monotonic strictly increasing or decreasing function of the dependence parameter $\phi_i^{(l,l')}$ given that other parameters $\mu_i^{(l)}$, $\mu_i^{(l')}$, $\sigma_i^{(l)}$ and $\sigma_i^{(l')}$ are fixed.

Then, the proposed model is identifiable, i.e. the two joint distributions (from the proposed model class with different parameterizations) match with

$$p\left(n_i^{(1)}, \ldots, n_i^{(L)}\right) = \tilde{p}\left(n_i^{(1)}, \ldots, n_i^{(L)}\right)$$

where p and $\tilde{p}$ are parameterized by $\theta = \{\boldsymbol{\beta}_1^{(l)}, \boldsymbol{\beta}_2^{(l)}, \boldsymbol{\beta}_3^{(l,l')} \,|\, l, l' = 1, \ldots, L\}$ and $\tilde{\theta} = \{\tilde{\boldsymbol{\beta}}_1^{(l)}, \tilde{\boldsymbol{\beta}}_2^{(l)}, \tilde{\boldsymbol{\beta}}_3^{(l,l')} \,|\, l, l' = 1, \ldots, L\}$, if and only if $\theta = \tilde{\theta}$.

Proof. Since all covariates spaces are of full rank and the link functions in Equations (4) to (6) are monotone, it is apparent that the identifiability problem for regression is reduced to the identifiability problem for multivariate mixed Poisson distribution (without regression) only. In other words, p and $\tilde{p}$ are re-parameterized by $\theta^* = (\mu_i^{(l)}, \sigma_i^{(l)}, \phi_i^{(l,l')} \,|\, l, l' = 1, \ldots, L)$ and $\tilde{\theta}^* = (\tilde{\mu}_i^{(l)}, \tilde{\sigma}_i^{(l)}, \tilde{\phi}_i^{(l,l')} \,|\, l, l' = 1, \ldots, L)$, and we want to show that $p(\cdot) = \tilde{p}(\cdot)$ if and only if $\theta^* = \tilde{\theta}^*$. The "if" statement is trivial. For the "only if" statement, we note that in order for the two joint distributions match, all moments (including expectation, variance and correlations) must reconcile. From the moment property above, matching the mean $\mathbb{E}[N_i^{(l)}]$ requires that $\mu_i^{(l)} = \tilde{\mu}_i^{(l)}$. Similarly, matching the variance $Var(N_i^{(l)})$ requires that $\sigma_i^{(l)} = \tilde{\sigma}_i^{(l)}$ given that condition (i) holds and $\mu_i^{(l)}$ is fixed. Finally, matching the correlation $\text{Corr}(N_i^{(l)}, N_i^{(l')})$ requires that $\phi_i^{(l,l')} = \tilde{\phi}_i^{(l,l')}$ given that condition (ii) holds and $\mu_i^{(l)}$ and $\sigma_i^{(l)}$ are fixed. ■

2.3 Model specifications

For the copula function $C_{\boldsymbol{\Phi}_i}(\cdot)$, we choose a Gaussian copula. In this case, the copula density is given as the following closed form:

$$c_{\boldsymbol{\Phi}_i}(u_{i1}, \ldots, u_{iL}) = |\boldsymbol{\Phi}_i|^{-1/2} \exp(-\Phi^{-1}(\boldsymbol{u}_i)^\top (\boldsymbol{\Phi}_i^{-1} - I_L)\Phi^{-1}(\boldsymbol{u}_i))/2),$$

where $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iL})$, I_L is the $L \times L$ identity matrix, and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. For expository purposes, from now on we specialize with the bivariate case $L = 2$.

When $L = 2$, we only have one dependence parameter $\phi_i^{(1,2)}$ in $\boldsymbol{\Phi}_i$. Therefore, from now on, we write $\phi_i := \phi_i^{(1,2)}$ and $C_{\phi_i}(\cdot) := C_{\boldsymbol{\Phi}_i}(\cdot)$ for simplicity. Further, we simplify the notations of Equation (6) as $\phi_i = g(\boldsymbol{x}_{3,i}^T \boldsymbol{\beta}_3)$. As ϕ_i is a correlation parameter with $\phi_i \in (-1, 1)$, we use $g(x) = \frac{2}{\pi} \arctan x$ as a natural link function. Since $g: (-\infty, \infty) \Rightarrow (-1, 1)$, there is no restriction in the range of $\boldsymbol{x}_{3,i}^T \boldsymbol{\beta}_3$, which is beneficial in terms of optimization.

We assume lognormal distribution[1] $\mathcal{LN}(-\frac{\sigma_i^{(l)2}}{2}, \sigma_i^{(l)2})$ is selected as $F_l(\cdot)$, the marginal distribution of the latent random variable $Z_i^{(l)}$, so that we have $\mathbb{E}[Z_i^{(l)}] = 1$. The variance of $Z_i^{(l)}$ is given by $Var(Z_i^{(l)}) = \exp\{\sigma_i^{(l)2}\} - 1$, which is a monotonic strictly increasing function of $\sigma_i^{(l)}$, and hence condition (i) of the marginalization property is satisfied. In our proposed general framework on multivariate claim frequencies modelling, it is possible to use alternative marginal distributions of $Z_i^{(l)}$ as long as $\mathbb{E}[Z_i^{(l)}] = 1$ and $Var(Z_i^{(l)})$ is a monotone function of $\sigma_i^{(l)}$. For example, another model specification and corresponding estimation scheme where $Z_i^{(l)}$ marginally follows gamma distribution, a classical choice of random effects for frequency, are provided in Appendix A. However, as shown in Equation (17), the joint density function of latent variables $\pi(\boldsymbol{z}_i)$ under gamma marginals cannot be expressed as a closed-form formula, due to the lack of an analytical expression for $\Phi^{-1}(F_l(z_i^{(l)}))$ with gamma cdf F_l. This will significantly increase the computational burden for parameter estimations. In contrast, the choice of log-normal marginals results to analytically tractable $\pi(\boldsymbol{z}_i)$ (See Equation (11) below), making the proposed model more computationally desirable.

[1]Note that the proposed modelling framework has sufficient flexibility to be used for any continuous and at least twice differentiable mixing distribution, including but not limited to the exponential dispersion family distributions which may be more suitable for alternative datasets.

Once the copula and marginal distributions are specified as above, $\boldsymbol{Z}_i$ has the following joint density function, which turns out to be the density of a bivariate lognormal distribution:

$$
\begin{aligned}
\pi(\boldsymbol{z}_i) &= f_1(z_i^{(1)})f_2(z_i^{(2)}) \times c_{\boldsymbol{\Phi}_i}\left(F_1(z_i^{(1)}),\, F_2(z_i^{(2)})\right) \\
&= \frac{1}{2z_i^{(1)}z_i^{(2)}\pi\sigma_i^{(1)}\sigma_i^{(2)}} \exp\left[-\frac{1}{2}\left(\frac{\log z_i^{(1)}+0.5\sigma_i^{(1)2}}{\sigma_i^{(1)}}\right)^2 - \frac{1}{2}\left(\frac{\log z_i^{(2)}+0.5\sigma_i^{(2)2}}{\sigma_i^{(2)}}\right)^2\right] \times \\
&\frac{1}{\sqrt{1-\phi_i^2}} \exp\left(-\frac{\phi_i}{2(1-\phi_i^2)}\begin{pmatrix}\frac{\log z_i^{(1)}+0.5\sigma_i^{(1)2}}{\sigma_i^{(1)}}\\ \frac{\log z_i^{(2)}+0.5\sigma_i^{(2)2}}{\sigma_i^{(2)}}\end{pmatrix}^{\top}\begin{pmatrix}\phi_i & -1\\ -1 & \phi_i\end{pmatrix}\begin{pmatrix}\frac{\log z_i^{(1)}+0.5\sigma_i^{(1)2}}{\sigma_i^{(1)}}\\ \frac{\log z_i^{(2)}+0.5\sigma_i^{(2)2}}{\sigma_i^{(2)}}\end{pmatrix}\right) \qquad (11)\\
&= \frac{1}{2z_i^{(1)}z_i^{(2)}\pi\sigma_i^{(1)}\sigma_i^{(2)}\sqrt{1-\phi_i^2}} \times \\
&e^{-\frac{1}{2(1-\phi_i^2)}\left[\left(\frac{\log z_i^{(1)}+0.5\sigma_i^{(1)2}}{\sigma_i^{(1)}}\right)^2 - 2\phi_i\left(\frac{\log z_i^{(1)}+0.5\sigma_i^{(1)2}}{\sigma_i^{(1)}}\right)\left(\frac{\log z_i^{(2)}+0.5\sigma_i^{(2)2}}{\sigma_i^{(2)}}\right) + \left(\frac{\log z_i^{(2)}+0.5\sigma_i^{(2)2}}{\sigma_i^{(2)}}\right)^2\right]}.
\end{aligned}
$$

The correlation coefficient between $Z_i^{(1)}$ and $Z_i^{(2)}$ is given by

$$
\mathrm{Corr}(Z_i^{(1)}, Z_i^{(2)}) = \frac{\left(\exp(\phi_i\sigma_i^{(1)}\sigma_i^{(2)}) - 1\right)}{\sqrt{\left[\exp(\sigma_i^{(1)2}) - 1\right]\left[\exp(\sigma_i^{(2)2}) - 1\right]}}, \qquad (12)
$$

which is a strictly increasing function of ϕ_i given fixed dispersion parameters. Therefore, condition (ii) of identifiability property also satisfies and hence the proposed specifications lead to an identifiable model. Further, note from the above equation that the sign of $\mathrm{Corr}(Z_i^{(1)}, Z_i^{(2)})$ is fully determined by the sign of ϕ_i, and hence from the aforementioned moment property, the sign of ϕ_i also fully determines the sign of $\mathrm{Corr}\left(N_i^{(1)}, N_i^{(2)}\right)$.

We also examine the correlation coefficient between $N_i^{(1)}$ and $N_i^{(2)}$ as follows:

$$
\mathrm{Corr}\left(N_i^{(1)}, N_i^{(2)}\right) = \frac{\left(\exp(\phi_i\sigma_i^{(1)}\sigma_i^{(2)}) - 1\right)}{\sqrt{\left[1/\mu_i^{(1)} + \exp(\sigma_i^{(1)2}) - 1\right]\left[1/\mu_i^{(2)} + \exp(\sigma_i^{(2)2}) - 1\right]}}. \qquad (13)
$$

Hence, the proposed model can capture a full range of dependence, given from the moment property that $\mu_i^{(1)}\exp\{\sigma_i^{(1)2}\} \to \infty$, $\mu_i^{(2)}\exp\{\sigma_i^{(2)2}\} \to \infty$, $\sigma_i^{(1)} \to \infty$ and $\sigma_i^{(2)} \to \infty$.

3 Statistical inference: The MCEM algorithm

We introduce a sophisticated way to optimize the joint likelihood that consists of discrete marginals and a continuous mixing copula via an MCEM algorithm. In particular, the MCEM algorithm works around the difficulty of finding the ML estimates of the model which has a joint pmf in Equation (3) that cannot be written in closed form by applying Monte Carlo techniques at the E-Step to obtain an approximate expected log-likelihood function and subsequently by solving the corresponding optimization problem at the M-Step. For more details regarding the MCEM algorithm see Wei and Tanner [1990], Booth et al. [2001]. Also, Chan and Ledolter [1995] and Fort and Moulines [2003] studied its convergence properties.

By augmentation of unobserved multivariate random effects $Z_i^{(l)}$ for $i = 1, \ldots, n$ and $l = 1, \ldots, L$, one can write the complete log-likelihood as follows:

$$
\ell_c(\theta) = \sum_{i=1}^{n}\left[\left(\sum_{l=1}^{L} n_i^{(l)}\log(\mu_i^{(l)}z_i^{(l)}) - \mu_i^{(l)}z_i^{(l)} - \log n_i^{(l)}!\right) + \log\pi(\boldsymbol{z}_i)\right], \qquad (14)
$$

where $\theta = (\boldsymbol{\beta}_1^{(1)}, \ldots, \boldsymbol{\beta}_1^{(L)}, \boldsymbol{\beta}_2^{(1)}, \ldots, \boldsymbol{\beta}_2^{(L)}, \boldsymbol{\beta}_3)$ includes all parameters to be estimated.

- **E-step:** Evaluate the following Q-function given $\theta^{(r)}$, estimated value of θ at the r^{th} iteration.

$$
\begin{aligned}
Q(\theta;\theta^{(r)}) &= \mathbb{E}_{z_i}[\ell_c(\theta)|\mathcal{D},\theta^{(r)}] \\
&\propto \sum_{i=1}^{n}\sum_{l=1}^{L} n_i^{(l)} \log \mu_i^{(l)} - \mu_i^{(l)} \mathbb{E}_{z_i}[z_i^{(l)}|\mathcal{D},\theta^{(r)}] + \sum_{i=1}^{n} \mathbb{E}_{z_i}[\log \pi(\boldsymbol{z}_i)|\mathcal{D},\theta^{(r)}].
\end{aligned} \tag{15}
$$

Here we denote $w_i^{(l;r)} = \mathbb{E}_{z_i}[z_i^{(l)}|\mathcal{D},\theta^{(r)}]$, which needs to be numerically evaluated since

$$
\pi(\boldsymbol{z}_i|\mathcal{D}) = \frac{\pi(\boldsymbol{z}_i)\prod_{l=1}^{L} p\left(n_i^{(l)}|z_i^{(l)}\right)}{\int \pi(\boldsymbol{z}_i)\prod_{l=1}^{L} p\left(n_i^{(l)}|z_i^{(l)}\right) d\boldsymbol{z}_i}
$$

might not have a closed form. However, one can obtain posterior expectation of a function of $\boldsymbol{z}_i$ using a Monte Carlo sum as follows:

$$
\begin{aligned}
\mathbb{E}_{z_i}[f(\boldsymbol{z}_i)|\mathcal{D},\theta^{(r)}] &= \int f(\boldsymbol{z}_i)\pi(\boldsymbol{z}_i|\mathcal{D})d\boldsymbol{z}_i \\
&= \frac{\int f(\boldsymbol{z}_i)\pi(\boldsymbol{z}_i)\prod_{l=1}^{L} p\left(n_i^{(l)}|z_i^{(l)}\right) d\boldsymbol{z}_i}{\int \pi(\boldsymbol{z}_i)\prod_{l=1}^{L} p\left(n_i^{(l)}|z_i^{(l)}\right) d\boldsymbol{z}_i} \\
&\simeq \frac{\sum_{s=1}^{S} f(\boldsymbol{z}_{i[s]})\prod_{l=1}^{L} p\left(n_i^{(l)}|z_{i[s]}^{(l)}\right)}{\sum_{s=1}^{S}\prod_{l=1}^{L} p\left(n_i^{(l)}|z_{i[s]}^{(l)}\right)},
\end{aligned} \tag{16}
$$

where $\boldsymbol{z}_{i[s]} \sim \pi(\boldsymbol{z}_i)$ for $s = 1, \ldots, S$. If $\boldsymbol{z}_i$ follows the bivariate lognormal distribution as mentioned above, then

$$
\begin{aligned}
\mathbb{E}_{z_i}[\log \pi(\boldsymbol{z}_i)|\mathcal{D},\theta^{(r)}] = &-\log 2\pi - \log \sigma_i^{(1)} - \log \sigma_i^{(2)} - 0.5\log(1-{\phi_i}^2) \\
&- \mathbb{E}_{z_i}\left[\log z_i^{(1)} + \frac{(\log z_i^{(1)})^2}{2(\sigma_i^{(1)})^2(1-\phi_i^2)}\middle|\mathcal{D},\theta^{(r)}\right] \\
&- \mathbb{E}_{z_i}\left[\log z_i^{(2)} + \frac{(\log z_i^{(2)})^2}{2(\sigma_i^{(2)})^2(1-\phi_i^2)}\middle|\mathcal{D},\theta^{(r)}\right] \\
&+ \mathbb{E}_{z_i}\left[\frac{\phi_i \log z_i^{(1)} \log z_i^{(2)}}{\sigma_i^{(1)}\sigma_i^{(2)}(1-\phi_i^2)}\middle|\mathcal{D},\theta^{(r)}\right] \\
&- \mathbb{E}_{z_i}\left[\frac{\log z_i^{(1)}}{2(1-\phi_i^2)} + \frac{\log z_i^{(2)}}{2(1-\phi_i^2)}\middle|\mathcal{D},\theta^{(r)}\right] \\
&- \mathbb{E}_{z_i}\left[\frac{(\sigma_i^{(1)})^2}{8(1-\phi_i^2)} + \frac{(\sigma_i^{(2)})^2}{8(1-\phi_i^2)}\middle|\mathcal{D},\theta^{(r)}\right] \\
&+ \mathbb{E}_{z_i}\left[\frac{\phi_i \log z_i^{(1)}\sigma_i^{(2)}}{2\sigma_i^{(1)}(1-\phi_i^2)} + \frac{\phi_i \log z_i^{(2)}\sigma_i^{(1)}}{2\sigma_i^{(2)}(1-\phi_i^2)}\middle|\mathcal{D},\theta^{(r)}\right] \\
&+ \mathbb{E}_{z_i}\left[\frac{\phi_i\sigma_i^{(1)}\sigma_i^{(2)}}{4(1-\phi_i^2)}\middle|\mathcal{D},\theta^{(r)}\right].
\end{aligned}
$$

- **M-step:** In this step, we want to find the updated parameters $\theta^{(r+1)}$ such that the Q-function is increased with respect to θ, in other words, $Q(\theta^{(r+1)};\theta^{(r)}) \geq Q(\theta^{(r)};\theta^{(r)})$. To do so, we update the parameters $\boldsymbol{\beta}_1^{(1)}, \boldsymbol{\beta}_1^{(2)}, \boldsymbol{\beta}_2^{(1)}, \boldsymbol{\beta}_2^{(2)}$ and $\boldsymbol{\beta}_3$ sequentially using Newton-Raphson algorithm as follows:
  1. Set $\theta \leftarrow \theta^{(r)}$. Recall that $\theta = (\boldsymbol{\beta}_1^{(1)}, \boldsymbol{\beta}_1^{(2)}, \boldsymbol{\beta}_2^{(1)}, \boldsymbol{\beta}_2^{(2)}, \boldsymbol{\beta}_3)$.
  2. For $l = 1, 2$ and $j = 1, 2$, update the parameters sequentially as $\boldsymbol{\beta}_j^{(l)} \leftarrow \boldsymbol{\beta}_j^{(l)} - [H_j^{(l;r)}(\theta)]^{-1} h_j^{(l;r)}(\theta)$, where $h_j^{(l;r)}(\theta)$ is a $p_j^{(l)}$-column vector and $H_j^{(l;r)}(\theta)$ is a $p_j^{(l)} \times p_j^{(l)}$ matrix which will be defined below.

3. Update the regression parameters for dependence $\boldsymbol{\beta}_3 \leftarrow \boldsymbol{\beta}_3 - [H_3^{(r)}(\theta)]^{-1} h_j^{(r)}(\theta)$, where $h_j(\theta)$ is a p_3-column vector and $H_j(\theta)$ is a $p_3 \times p_3$ matrix defined below.
4. Retrieve the updated parameters $\theta^{(r+1)} \leftarrow \theta$.

From the above, the vectors $h_j^{(l;r)}(\theta)$ $(j = 1, 2;\ l = 1, 2)$, $h_3^{(r)}(\theta)$ and the matrices $H_j^{(l;r)}(\theta)$ $(j = 1, 2;\ l = 1, 2)$, $H_3^{(r)}(\theta)$ are expressed as follows:

$$h_j^{(l;r)}(\theta) = \frac{\partial Q(\theta;\theta^{(r)})}{\partial \boldsymbol{\beta}_j^{(l)}} := \boldsymbol{W}_{j1}^{(l;r)} \boldsymbol{X}_j^{(l)}; \qquad H_j^{(l;r)}(\theta) = \frac{\partial^2 Q(\theta;\theta^{(r)})}{\partial \boldsymbol{\beta}_j^{(l)} \partial {\boldsymbol{\beta}^{(l)}}'_j} = \boldsymbol{X}_j^{(l)T} \boldsymbol{W}_{j2}^{(l;r)} \boldsymbol{X}_j^{(l)}$$

for $j = 1, 2$ and $l = 1, 2$, and

$$h_3^{(r)}(\theta) = \frac{\partial Q(\theta;\theta^{(r)})}{\partial \boldsymbol{\beta}_3} := \boldsymbol{W}_{31}^{(r)} \boldsymbol{X}_3; \qquad H_3^{(r)}(\theta) = \frac{\partial^2 Q(\theta;\theta^{(r)})}{\partial \boldsymbol{\beta}_3 \partial \boldsymbol{\beta}_3'} = \boldsymbol{X}_3^T \boldsymbol{W}_{32}^{(r)} \boldsymbol{X}_3,$$

where $\boldsymbol{W}_{j1}^{(l;r)}$, $\boldsymbol{W}_{j2}^{(l;r)}$ $(j = 1, 2; l = 1, 2)$, $\boldsymbol{W}_{31}^{(r)}$ and $\boldsymbol{W}_{32}^{(r)}$ are the matrices given by

$$\boldsymbol{W}_{11}^{(l;r)} = J_n^T \operatorname{diag}\left(\{n_i^{(l)} - \mu_i^{(l)} w_i^{(l;r)}\}_{i=1,\dots,n}\right),$$

$$\boldsymbol{W}_{12}^{(l;r)} = \operatorname{diag}\left(\{-\mu_i^{(l)} w_i^{(l;r)}\}_{i=1,\dots,n}\right),$$

$$\begin{aligned}\boldsymbol{W}_{21}^{(l;r)} = J_n^T \operatorname{diag}\Bigg(\Bigg\{ & -1 + \frac{\eta_i^{(ll;r)}}{(\sigma_i^{(l)})^2(1-\phi_i^2)} - \frac{\phi_i \eta_i^{(12;r)}}{\sigma_i^{(1)}\sigma_i^{(2)}(1-\phi_i^2)} - \frac{(\sigma_i^{(l)})^2}{4(1-\phi_i^2)} + \frac{\phi_i \sigma_i^{(1)}\sigma_i^{(2)}}{4(1-\phi_i^2)} \\ & - \frac{\phi_i \eta_i^{(l;r)} \sigma_i^{(1)}\sigma_i^{(2)}}{2(\sigma_i^{(l)})^2(1-\phi_i^2)} + \frac{\phi_i \eta_i^{(l';r)} (\sigma_i^{(l)})^2}{2(\sigma_i^{(1)}\sigma_i^{(2)})(1-\phi_i^2)} \Bigg\}_{i=1,\dots,n}\Bigg), \qquad l' \neq l,\end{aligned}$$

$$\begin{aligned}\boldsymbol{W}_{22}^{(l;r)} = \operatorname{diag}\Bigg(\Bigg\{ & \frac{-2\eta_i^{(ll;r)}}{(\sigma_i^{(l)})^2(1-\phi_i^2)} + \frac{\phi_i \eta_i^{(12;r)}}{\sigma_i^{(1)}\sigma_i^{(2)}(1-\phi_i^2)} - \frac{(\sigma_i^{(l)})^2}{2(1-\phi_i^2)} + \frac{\phi_i \sigma_i^{(1)}\sigma_i^{(2)}}{4(1-\phi_i^2)} \\ & + \frac{\phi_i \eta_i^{(l;r)} \sigma_i^{(1)}\sigma_i^{(2)}}{2(\sigma_i^{(l)})^2(1-\phi_i^2)} + \frac{\phi_i \eta_i^{(l';r)} (\sigma_i^{(l)})^2}{2(\sigma_i^{(1)}\sigma_i^{(2)})(1-\phi_i^2)} \Bigg\}_{i=1,\dots,n}\Bigg),\end{aligned}$$

$$\boldsymbol{W}_{31}^{(r)} = J_n^T \operatorname{diag}\left(\left\{\frac{2/\pi}{1+(\boldsymbol{x}_{3,i}^T\boldsymbol{\beta}_3)^2} \times h_{3,i}^{(r)}\right\}_{i=1,\dots,n}\right)$$

with

$$\begin{aligned} h_{3,i}^{(r)} = & \frac{\phi_i}{1-{\phi_i}^2} - \frac{\phi_i \eta_i^{(11;r)}}{(\sigma_i^{(1)})^2(1-{\phi_i}^2)^2} - \frac{\phi_i \eta_i^{(22;r)}}{(\sigma_i^{(2)})^2(1-{\phi_i}^2)^2} + \frac{(1+\phi_i^2)\eta_i^{(12;r)}}{\sigma_i^{(1)}\sigma_i^{(2)}(1-{\phi_i}^2)^2} \\ & - \frac{((\sigma_i^{(1)})^2 + (\sigma_i^{(2)})^2)\phi_i}{4(1-\phi_i^2)^2} - \frac{\phi_i(\eta_i^{(1;r)} + \eta_i^{(2;r)})}{(1-\phi_i^2)^2} + \frac{(1+\phi_i^2)\sigma_i^{(1)}\sigma_i^{(2)}}{4(1-\phi_i^2)^2} \\ & + \frac{(1+\phi_i^2)(\eta_i^{(1;r)}\sigma_i^{(2)}/\sigma_i^{(1)} + \eta_i^{(2;r)}\sigma_i^{(1)}/\sigma_i^{(2)})}{2(1-{\phi_i}^2)^2},\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{W}_{32}^{(r)} = \operatorname{diag}\Bigg(\Bigg\{ & \frac{4/\pi^2}{(1+(\boldsymbol{x}_{3,i}^T\boldsymbol{\beta}_3)^2)^2} \times \\ & \Bigg(\frac{1+\phi_i^2}{(1-{\phi_i}^2)^2} - \frac{(1+3\phi_i^2)\eta_i^{(11;r)}}{(\sigma_i^{(1)})^2(1-\phi_i^2)^3} - \frac{(1+3\phi_i^2)\eta_i^{(22;r)}}{(\sigma_i^{(2)})^2(1-\phi_i^2)^3} - \frac{2\phi_i(\phi_i^2+3)\eta_i^{(12;r)}}{\sigma_i^{(1)}\sigma_i^{(2)}(1-\phi_i^2)^3} \\ & + \frac{((\sigma_i^{(1)})^3 + (\sigma_i^{(2)})^3)(1+3\phi_i^2)}{4(1-\phi_i^2)^3} - \frac{(1+3\phi_i^2)(\eta_i^{(1;r)}\sigma_i^{(1)} + \eta_i^{(2;r)}\sigma_i^{(2)})}{(1-\phi_i^2)^3} - \pi \boldsymbol{x}_{3,i}^T\boldsymbol{\beta}_3 h_{3,i}^{(r)} \\ & + \frac{\phi_i(3+\phi_i^2)\sigma_i^{(1)}\sigma_i^{(2)}}{2(1-\phi_i^2)^3} + \frac{(1+\phi_i^2)(\eta_i^{(1;r)}\sigma_i^{(2)}/\sigma_i^{(1)} + \eta_i^{(2;r)}\sigma_i^{(1)}/\sigma_i^{(2)})}{(1-{\phi_i}^2)^2}\Bigg)\Bigg\}_{i=1,\dots,n}\Bigg).\end{aligned}$$

Note that J_n is a $n \times 1$ matrix where all components are 1. From the above expressions, the notations $\eta_i^{(1;r)}, \eta_i^{(2;r)}, \eta_i^{(11;r)}, \eta_i^{(22;r)}, \eta_i^{(12;r)}$ are the expectations given as follows:

$$\eta_i^{(1;r)} = \mathbb{E}_{z_i}\left[\log z_i^{(1)} \middle| \mathcal{D}, \theta^{(r)}\right], \qquad \eta_i^{(2;r)} = \mathbb{E}_{z_i}\left[\log z_i^{(2)} \middle| \mathcal{D}, \theta^{(r)}\right],$$
$$\eta_i^{(11;r)} = \mathbb{E}_{z_i}\left[(\log z_i^{(1)})^2 \middle| \mathcal{D}, \theta^{(r)}\right], \quad \eta_i^{(22;r)} = \mathbb{E}_{z_i}\left[(\log z_i^{(2)})^2 \middle| \mathcal{D}, \theta^{(r)}\right],$$
$$\eta_i^{(12;r)} = \mathbb{E}_{z_i}\left[\log z_i^{(1)} \log z_i^{(2)} \middle| \mathcal{D}, \theta^{(r)}\right],$$

which can be approximated by a Monte Carlo sum described in Equation (16).

Finally, standard errors of regression coefficients can be estimated by computing Hessian matrix numerically as shown in Le Kang et al. [2013].

4 Empirical analysis

4.1 Data description

In this section, we examine the Local Government Property Insurance Fund (LGPIF) data from the state of Wisconsin. The LGPIF provides property insurance to various local government units, including cities, counties, towns and schools. Local government entities pay premiums of about a total of \$25 million each year in exchange of receiving a total coverage of about \$75 billion. One apparent difference between LGPIF and private property insurance companies is that LGPIF is not allowed to deny coverage of local government entity. The LGPIF contains three major groups of property insurance coverage, namely building and contents (BC), inland marine (IM) covering contractor's equipment, and motor vehicles. In this application, we specially focus on modelling jointly the claim frequencies of inland marine (IM, denoted as $N_i^{(1)}$) and new vehicle collisions (CN, denoted as $N_i^{(2)}$). There are a total of $n = 5,240$ entity-years (from now on we call it "policies") over a period of 5 years from 2006-2010 for model training purpose. The remaining $n^o = 1,025$ policies of year 2011 are treated as a test set for model validation purpose. The dataset is also adopted by e.g. Ahn et al. [2021] and is publicly available in `https://sites.google.com/a/wisc.edu/jed-frees/home`.

Table 1 presents the summary statistics for the claim frequencies from each of the two coverages. We observe that new vehicle collisions are more prevalent than inland marine claims. While over-dispersion exists for both claim types, the dispersion ratio of vehicle collision claims (3.854) is much larger than that of inland marine claims (1.746). These illustrate the heterogeneities across the two claim types. All three dependence measures (correlation, Kendall's tau and Spearman's rho) suggest highly significant dependence between the two claim types, highlightling the importance of appropriately modelling the claim frequencies jointly instead of separately assuming independence.

Table 2 presents the summary statistics of the explanatory variables. Variables 1–5 are the categorical variables to classify policies into one of the six entity types. Variables 6 and 9 are included as the coverage amounts for each of the two claim types. Since the coverage amounts for the original dataset are very heavy-tailed in nature where the outliers will drastically distort the model fitting, we transform these variables using a rank-based standardization procedure, so the effect of outliers is mitigated. Deductible amount is included as variable 7 while it is only relevant for IM claims. Variables 8 and 10 indicate whether or not each type of claims occurred in the previous year. Since variables 6–8 are the information specifically for IM claims only, we include only variables 1–8 in explaining the mean and dispersion of IM claims (i.e. $\boldsymbol{x}_{1,i}^{(1)}$ and $\boldsymbol{x}_{2,i}^{(1)}$ in Equations (4) and (5) contain only variables 1–8). Similarly, $\boldsymbol{x}_{1,i}^{(2)}$ and $\boldsymbol{x}_{2,i}^{(2)}$ in Equations (4) and (5) contain only variables 1–5 and 9–10 explaining the CN claims. On the other hand, we select only variables 1–5 as $\boldsymbol{x}_{3,i}$ in Equation (6) to explain the dependence heterogeneity between two coverages.

4.2 In-sample estimation

To assess the novelty of the proposed model, we compare the fitting results of the proposed model and the benchmark models both in-sample estimation and out-of-sample validation. For detailed description of the benchmark models, see Appendix B.

Throughout this section, our proposed model with varying dispersion and dependence, the BPLN regression model with the shared random effect, and the BPLN regression model with common covariates

Table 1: Summary statistics for the claim frequencies

	Prop. of 0s	Mean	SD	Max	Correlation	Kendall's tau	Spearman's rho
IM	0.9574	0.0571	0.3156	6			
CN	0.9353	0.1273	0.7004	15	0.2443	0.1981	0.2007

Table 2: Summary statistics for the explanatory variables

Variable index	Variable name	Type	Description	Proportion/ Mean
1	TypeCity	Categorical	Indicator for city entity.	0.1450
2	TypeCounty	Categorical	Indicator for county entity.	0.0592
3	TypeMisc	Categorical	Indicator for miscellaneous entity.	0.1078
4	TypeSchool	Categorical	Indicator for school entity.	0.2910
5	TypeTown	Categorical	Indicator for town entity.	0.1660
–	TypeVillage	Categorical	Indicator for village entity (reference category).	0.2309
6	CoverageIM	Continuous	Coverage amount of IM (transformed).	0.0000
7	lnDeductIM	Continuous	Log deductible amount for inland marine.	5.3440
8	NoClaimCreditIM	Binary	Indicator for no IM claims in prior year.	0.4399
9	CoverageCN	Continuous	Coverage amount of CN (transformed).	0.0000
10	NoClaimCreditCN	Binary	Indicator for no CN claims in prior year.	0.0945

are referred as **Proposed**, **Shared**, and **Common** models, respectively. Table 3 summarizes the results of estimation for the models. Note that $\hat{\sigma} = 0.5448$ for the shared model and $\hat{\sigma}^{(1)} = 1.2186$, $\hat{\sigma}^{(2)} = 1.3250$, and $\hat{\phi} = 0.0087$ for the common model.

Here are some findings from Table 3:

- Regression coefficients for mean components are similar for the proposed and shared model, while the the coefficients from the common model deviates from such trend.
- The common model suffers from lack of fit mainly due to the omission of coverage specific covariates.
- As compared to the shared model, which implicitly assumes the perfect positive correlation between the latent variables $Z^{(1)}$ and $Z^{(2)}$, the proposed model apparently has a better goodness-of-fit in terms of the log-likelihood. After taking account of model complexities, our proposed model still shows slight improvement in AIC yet produces inferior BIC. This reflects that the fitting performance can be significantly enhanced by adopting less than perfectly correlated latent variables and incorporating covariates influences to multiple model components (dispersion of marginals and dependence structure), but such improvements are not large enough to obtain a better BIC. We briefly explain this issue through two different aspects: BIC's nature and model characteristics.

  By its definition, BIC penalizes the model complexity more heavily than AIC, i.e., the penalty term is $\log(n)$ (for BIC) versus two (for AIC) multiplied by the number of parameters. An excess penalty disfavors our proposed model, which is more complex than the shared random effect model. The proposed model may require a larger sample size to justify its outperformance in terms of BIC. Furthermore, note that the original purpose of BIC is to identify the true model (Kuha [2004]), i.e., BIC is often more accurate in choosing the correct model complexity. Yet, this is not the aim in insurance practice since the true model is never known for any real datasets. Kuha [2004] also argues that AIC is a more appropriate criterion if our goal is to predict future data, even if AIC itself may not be consistent. This is in line with the concept of the so-called "*algorithmic modelling culture*" according to Breiman [2001] that has to do with the ability of a given model or algorithm to generalize well to external data, which in this case are unseen insurance policies. Therefore, one should not simply conclude that a model generally does not perform as expected solely based on an inferior BIC.

  From the proposed model, $\hat{\beta}_3$ is significantly positive for all types of locations so that the estimated correlation, $\hat{\phi}_i$, is positive for all $i = 1, \ldots, n$. With positive dependence captured for both proposed and shared models, these two models are close in capturing the dependence structures across perils.

However, one cannot preclude the possibility of having negatively correlated claim frequencies that dampens the applicability of the shared random effect model especially when more than two types of coverage are jointly modeled. In this case the advantage of using our proposed model will be even more apparent.

- Observing the estimated regression coefficients ($\hat{\beta}_2^{(1)}$, $\hat{\beta}_2^{(2)}$ and $\hat{\beta}_3$) for the dispersion parameters ($\hat{\sigma}_i^{(1)}$, $\hat{\sigma}_i^{(2)}$) and correlation ($\hat{\phi}_i$) parameter with the corresponding standard errors, we find that $\hat{\sigma}_i^{(1)}$, $\hat{\sigma}_i^{(2)}$ and $\hat{\phi}_i$ are all significantly influenced by many explanatory variables. This result has an important implication in insurance pricing perspective, as insurance premiums are often determined not only by the expected claims, but also by their uncertainties. As insurance companies are typically risk adverse, higher dispersion or higher correlation on claim frequencies usually result to higher premium due to increased risk or reduced diversification. For example, it may be reasonable to charge an increased premium on policyholders with school entity type even if it does not have significant impacts to the average claim frequencies of both perils ($\hat{\beta}_1^{(1)}$ and $\hat{\beta}_1^{(2)}$ do not significantly deviate from zero), due to its positive effects to the dispersion parameters (significantly positive $\hat{\beta}_2^{(1)}$ and $\hat{\beta}_2^{(2)}$ resulting to higher uncertainties on the claim counts) and correlation parameter (significantly positive $\hat{\beta}_3$ resulting to reduced diversification).

Table 3: Estimation results

	Proposed					Shared		Common	
	$\beta_1^{(1)}$	$\beta_1^{(2)}$	$\beta_2^{(1)}$	$\beta_2^{(2)}$	β_3	$\beta_1^{(1)}$	$\beta_1^{(2)}$	$\beta_1^{(1)}$	$\beta_1^{(2)}$
(Intercept)	-4.0163	-5.4606	-1.3651	-0.1861	0.2011	-4.1238	-5.3784	-4.1178	-3.9479
	(0.4249)	(0.8398)	(0.0366)	(0.0153)	(0.0119)	(0.4224)	(0.2923)	(0.1546)	(0.1301)
TypeCity	-0.2121	0.372	0.6322	-0.0931		-0.1461	0.3671	1.047	0.9493
	(0.1891)	(0.156)	(0.0354)	(0.0272)		(0.2103)	(0.1536)	(0.1896)	(0.1626)
TypeCounty	0.7295	0.975	-0.133	0.1343	0.2034	0.5447	0.9306	2.615	3.4519
	(0.1909)	(0.1283)	(0.05)	(0.0477)	(0.0839)	(0.2237)	(0.138)	(0.177)	(0.1389)
TypeMisc	-2.1581	-0.8179	0.4907	0.0801		-2.1404	-0.8255	-2.9629	-2.1741
	(1.0123)	(0.6054)	(0.0379)	(0.0294)		(1.0098)	(0.5825)	(1.0118)	(0.5918)
TypeSchool	-0.0174	-0.2059	0.9499	0.0815	0.0796	0.1547	-0.2219	-1.0404	-0.0088
	(0.1815)	(0.1739)	(0.0306)	(0.0221)	(0.0348)	(0.2968)	(0.1693)	(0.2765)	(0.1755)
TypeTown	-0.3916	-1.413	-0.0408	0.0787		-0.3823	-1.3951	-0.4621	-1.6276
	(0.2764)	(0.4745)	(0.0315)	(0.0254)		(0.274)	(0.3713)	(0.2767)	(0.3768)
Coverage	1.4547	2.439	1.1862	-0.0638		1.5764	2.4129		
	(0.1153)	(0.4562)	(0.0187)	(0.0077)		(0.1758)	(0.1489)		
lnDeduct	0.0296	-0.5795	-0.1439	0.1692		0.0381	-0.6244		
	(0.0627)	(0.0452)	(0.006)	(0.0259)		(0.0616)	(0.1314)		
NoClaimCredit	-0.3689		-0.8581			-0.4757			
	(0.1088)		(0.0197)			(0.1231)			
Loglikelihood		-1840.45				-1861.96		-2399.70	
AIC		3754.91				3759.92		4829.40	
BIC		3997.78				3878.07		4927.86	

To assess the distributional goodness-of-fit of the proposed fitted model, we also depict two Q-Q plots (separately for each marginal) constructed as follows: Consider the normalized randomized quantile residuals for the marginal distributions resulting from the proposed models are defined as $\hat{r}_i^l = \phi^{-1}(u_i^l)$, where ϕ^{-1} is the inverse cumulative distribution function of a standard normal distribution, u_i^l is a uniformly distributed random value on $\left[P_i(n_i^l - 1|\boldsymbol{\theta}^{(r+1)}), P_i(n_i^l|\boldsymbol{\theta}^{(r+1)})\right]$, P_i is the cumulative distribution function estimated for the i^{th} policyholder, $\boldsymbol{\theta}^{(r+1)}$ is the vector of the estimated model parameters and n_i^l is the corresponding observation for $i = 1, \ldots, n$ per claim type $l = 1, 2$. Figure 1 depicts the normalized (random) quantiles for the proposed model per claim type $l = 1, 2$. From Figure 1, we see that the residuals of the proposed model are close to the diagonal which indicate a good fit to the distribution of the claim frequencies in the body and tail areas. Thus, the choice of the mixing density is adequate for this data, since the tails of mixed Poisson distributions are equivalent to the tails of their mixing distributions, see, for example, Perline [1998].

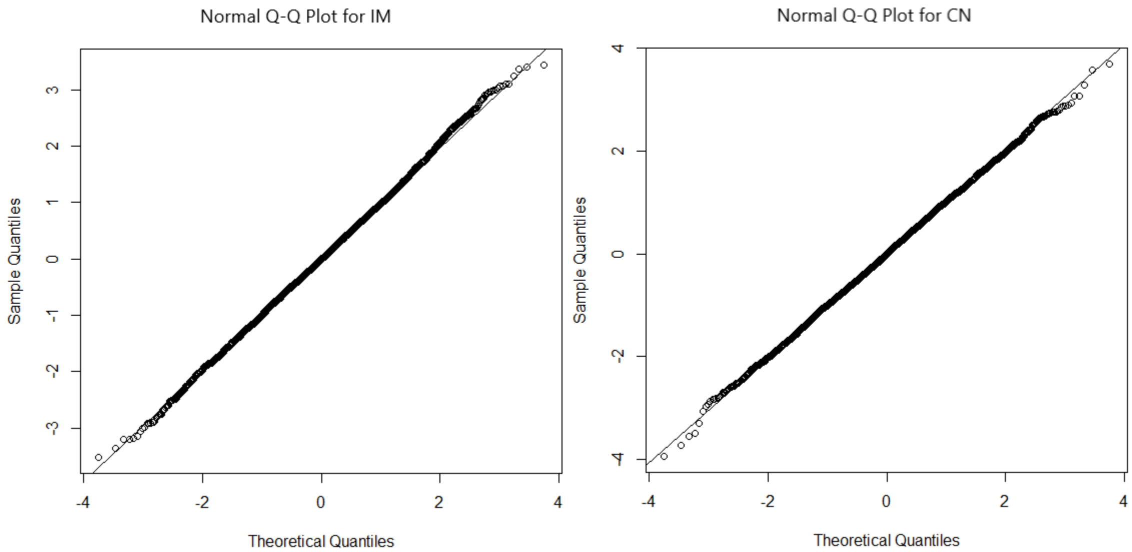

Figure 1: Normalized quantiles of the proposed model per claim type

To further assess the adequateness of the proposed fitted model to capture the regression structures, we perform a normalized residuals analysis, which plots $\hat{r}_i^l$ against each variable. In Figure 2, we demonstrate two violin plots (one for each claim type) of normalized residuals for each of the 6 entity types. As the violins do not differ significantly across levels, we conclude that our proposed model is sufficient in capturing the influence of covariates. Similar plots are performed against other variables and they all yield very similar results. For conciseness purpose, we refrain from presenting them one-by-one in this paper.

Figure 2: Residual violin plots against entity type per claim type

4.3 Analysis of dependence and out-of-sample validation

As the responses are discrete, it is more appropriate to analyze rank based dependence measures such as Kendall's tau (instead of linear correlation) to examine the goodness-of-fit of the fitted model in terms of dependence modelling. Table 4 compares the Kendall's taus between empirical data and fitted model under two approaches: "with covariates influence" and "without covariates influence". "With covariates influence" simply computes the empirical Kendall's tau (unconditioned on covariates) between $N_i^{(1)}$ and $N_i^{(2)}$ from the dataset, compared with that obtained by simulating $N_i^{(1)}$ and $N_i^{(2)}$ from the fitted model. However, such an approach also includes the dependence caused by the influence of the covariates simultaneously to the marginal distributions of both claim types. To evaluate the "real" intrinsic dependence between the two claim types, we present the Kendall's tau "without covariates influence" by applying a probability transformation technique. For details, readers are referred to Badescu et al. [2015] and Fung et al. [2019]. From both approaches, we can see that the fitted model matches decently to the empirical data in terms of Kendall's taus, suggesting the capability of the proposed multivariate count model to adequately capture the dependence structure of the dataset.

Table 4: Kendall's tau for the fitted model versus the empirical dataset

Kendall's tau	Empirical dataset	Proposed model
With covariates influence	0.198	0.182
Without covariates influence	0.321	0.322

Once the models are fitted with the training set, prediction performances of the models are assessed via out-of-sample validation. To measure the prediction performances, we used root-mean squared error (RMSE) and deviance statistic. While the difference of RMSEs between the proposed and shared models are negligible, one can see that the proposed model significantly outperforms the other models in terms of deviance, as shown in Table 5.

Table 5: Out-of-sample validation results

	Proposed	Shared	Common
RMSE	0.4672	0.4664	0.5276
Deviance	444.0522	471.4090	633.0516

5 Concluding Remarks

In this article, we considered a multivariate claim count regression model with varying dispersion and dependence parameters for addressing several challenges when modelling multivariate claim count data, such as systematic effects in the data, due to unobserved heterogeneity which leads overdispersion, and correlations of different signs and magnitude among the multiple claim count responses. Unlike many existing copula based methods for discrete marginals, we accommodate a continuous mixing density to capture the dependence that allows us to avoid finite differences in the likelihood, which trigger exponentially increasing computation times and numerical instability. Furthermore, our approach takes into account the impact of individual and coverage type covariates on the mean, dispersion and dependence components increasing the model prediction accuracy while maintaining its tractability. Therefore, the setup we proposed is fully flexible and can be efficiently employed for modelling diverse high-dimensional claim count data and hence it can be applied in various non-life insurance contexts. The bivariate Poisson-lognormal regression model with varying dispersion and dependence parameters was considered for expository purposes. For model fitting, we developed an efficient Monte Carlo Expectation-Maximization algorithm which reduced the computational burden for maximum likelihood estimation of the parameters of the model. The implementation of the algorithm was illustrated by fitting the model to Local Government Property Insurance Fund data from the state of Wisconsin.

Finally, it is worth noting that an interesting line for further research would be to incorporate time series components into the model equations to take into account both cross-sectional and temporal dependence between multiple claim types proceeding along similar lines as in Bermúdez et al. [2018].

Data Availability

The data underlying this article are available in its online supplementary material. The datasets were derived from the Wisconsin Property Fund data, which are also publicly available at `https://sites.google.com/a/wisc.edu/jed-frees/home`.

Acknowledgements and Funding

The authors warmly thank two anonymous referees and the editors for their numerous constructive comments that greatly helped to improve the paper compared to its initial version. The authors also declare no funding support for this work.

Conflict of Interest

The authors declare that there is no conflict of interest.

List of Figure Legends

References

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

J. Aguero-Valverde and P. P. Jovanis. Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. *Transportation Research Record*, 2136(1):82–91, 2009.

J. Y. Ahn, H. Jeong, and Y. Lu. On the ordering of credibility factors. *Insurance: Mathematics and Economics*, 101:626–638, 2021.

J. Aitchison and C. Ho. The multivariate Poisson-lognormal distribution. *Biometrika*, 76(4):643–653, 1989.

A. L. Badescu, X. S. Lin, D. Tang, and E. A. Valdez. Multivariate Pascal mixture regression models for correlated claim frequencies. *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2618265*, 2015.

W. Barreto-Souza and A. B. Simas. General mixed Poisson regression models with varying dispersion. *Statistics and Computing*, 26(6):1263–1280, 2016.

L. Bermúdez and D. Karlis. Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics*, 48(2):226–236, 2011.

L. Bermúdez and D. Karlis. A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Computational Statistics & Data Analysis*, 56(12):3988–3999, 2012.

L. Bermúdez and D. Karlis. A posteriori ratemaking using bivariate Poisson models. *Scandinavian Actuarial Journal*, 2017(2):148–158, 2017.

L. Bermúdez, M. Guillén, and D. Karlis. Allowing for time and cross dependence assumptions between claim counts in ratemaking models. *Insurance: Mathematics and Economics*, 83:161–169, 2018.

C. Bolancé and R. Vernic. Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution. *Insurance: Mathematics and Economics*, 85:89–103, 2019.

C. Bolancé, M. Guillen, and A. Pitarque. A Sarmanov distribution with beta marginals: An application to motor insurance pricing. *Mathematics*, 8(11), 2020.

J. G. Booth, J. P. Hobert, and W. Jank. A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. *Statistical Modelling*, 1(4):333–349, 2001.

L. Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231, 2001.

A. C. Cameron and P. K. Trivedi. *Regression analysis of count data*, volume 53. Cambridge University Press, 2013.

A. C. Cameron, T. Li, P. K. Trivedi, and D. M. Zimmer. Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *The Econometrics Journal*, 7 (2):566–584, 2004.

K. Chan and J. Ledolter. Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, 90(429):242–252, 1995.

Y. Chen and T. Hanson. Copula regression models for discrete and mixed bivariate responses. *Journal of Statistical Theory and Practice*, 11(4):515–530, 2017.

S. Chib and R. Winkelmann. Markov Chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435, 2001.

J. Chiquet, M. Mariadassou, and S. Robin. The Poisson-lognormal model as a versatile framework for the joint analysis of species abundances. *bioRxiv*, 2020.

M. Denuit, M. Guillen, and J. Trufin. Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. *Annals of Actuarial Science*, 13(2):378–399, 2019.

K. El-Basyouny and T. Sayed. Accident prediction models with random corridor parameters. *Accident Analysis & Prevention*, 41(5):1118–1123, 2009.

F. Famoye. On the bivariate negative binomial regression model. *Journal of Applied Statistics*, 37(6): 969–981, 2010.

G. Fort and E. Moulines. Convergence of the Monte Carlo expectation maximization for curved exponential families. *Annals of Statistics*, 31(4):1220–1259, 2003.

E. W. Frees and P. Wang. Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics*, 38(2):360–373, 2006.

T. C. Fung, A. L. Badescu, and X. S. Lin. A class of mixture of experts models for general insurance: Application to correlated claim frequencies. *ASTIN Bulletin: The Journal of the IAA*, 49(3):647–688, 2019.

C. Genest and J. Nešlehová. A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA*, 37(2):475–515, 2007.

M. Ghitany, D. Karlis, D. Al-Mutairi, and F. Al-Awadhi. An EM algorithm for multivariate mixed Poisson regression models and its application. *Applied Mathematical Sciences*, 6(137):6843–6856, 2012.

S. Gurmu and J. Elder. Generalized bivariate count data regression models. *Economics Letters*, 68(1): 31–36, 2000.

L. L. Ho and J. d. M. Singer. Generalized least squares methods for bivariate Poisson regression. *Communications in Statistics - Theory and Methods*, 30(2):263–277, 2001.

H. Jeong and D. K. Dey. Multi-peril frequency credibility premium via shared random effects. *Variance*, Forthcoming. Available at SSRN 3825435.

H. Joe. *Multivariate models and multivariate dependence concepts*. CRC Press, 1997.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Discrete multivariate distributions*, volume 165. Wiley New York, 1997.

R. C. Jung and R. Winkelmann. Two aspects of labor mobility: a bivariate Poisson regression approach. *Empirical economics*, 18(3):543–556, 1993.

D. Karlis. A general EM approach for maximum likelihood estimation in mixed Poisson regression models. *Statistical Modelling*, 1(4):305–318, 2001.

D. Karlis and L. Meligkotsidou. Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15(4):255–265, 2005.

N. R. Kocherlakota. Bubbles and constraints on debt accumulation. *Journal of Economic theory*, 57(1): 245–256, 1992.

S. Kocherlakota. On the compounded bivariate Poisson distribution: A unified treatment. *Annals of the institute of Statistical Mathematics*, 40(1):61–76, 1988.

S. Kocherlakota and K. Kocherlakota. Regression in the bivariate Poisson distribution. *Communications in Statistics - Theory and Methods*, 30(1):815–825, 2001.

F. Krummenauer. Representation of multivariate discrete distributions by probability generating functions. *Statistics & Probability Letters*, 39(4):327–331, 1998.

J. Kuha. AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research*, 33(2):188–229, 2004.

J. Lakshminarayana, S. N. Pandit, and K. Srinivasa Rao. On a bivariate Poisson distribution. *Communications in Statistics - Theory and Methods*, 28(2):267–276, 1999.

R. C. Le Kang, K. Darcy, J. Kauderer, and S.-Y. Liao. A fast Monte Carlo EM algorithm for estimation in latent class model analysis with an application to assess diagnostic accuracy for cervical neoplasia in women with AGC. *Journal of Applied Statistics*, 40(12):2699, 2013.

A. Lee. Applications: Modelling rugby league data via bivariate negative binomial regression. *Australian & New Zealand Journal of Statistics*, 41(2):141–152, 1999.

J. Ma, K. M. Kockelman, and P. Damien. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention*, 40(3): 964–975, 2008.

G. Marra and K. Wyszynski. Semi-parametric copula sample selection models for count responses. *Computational Statistics & Data Analysis*, 104:110–129, 2016.

A. M'Kendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1925.

M. K. Munkin and P. K. Trivedi. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. *The Econometrics Journal*, 2(1):29–48, 1999.

A. K. Nikoloulopoulos. Copula-based models for multivariate discrete response data. In *Copulae in Mathematical and Quantitative Finance*, pages 231–249. Springer, 2013.

A. K. Nikoloulopoulos. Efficient estimation of high-dimensional multivariate normal copula models with discrete spatial responses. *Stochastic Environmental Research and Risk Assessment*, 30(2):493–505, 2016.

A. K. Nikoloulopoulos and D. Karlis. Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, 37(9):1555–1568, 2010.

R. Oh, H. Jeong, J. Y. Ahn, and E. A. Valdez. A multi-year microlevel collective risk model. *Insurance: Mathematics and Economics*, 100:309–328, 2021.

E. S. Park and D. Lord. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record*, 2019(1):1–6, 2007.

F. Pechon, J. Trufin, and M. Denuit. Multivariate modelling of household claim frequencies in motor third-party liability insurance. *ASTIN Bulletin: The Journal of the IAA*, 48(3):969–993, 2018.

F. Pechon, M. Denuit, and J. Trufin. Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal*, 9(2):575–602, 2019.

F. Pechon, M. Denuit, and J. Trufin. Home and motor insurance joined at a household level using multivariate credibility. *Annals of Actuarial Science*, 15(1):82–114, 2021.

R. Perline. Mixed Poisson distributions tail equivalent to their mixing distributions. *Statistics & Probability Letters*, 38(3):229–233, 1998.

L. Rüschendorf. Copulas, Sklar's theorem, and distributional transform. In *Mathematical Risk Analysis*, pages 3–34. Springer, 2013.

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

A. Silva, S. J. Rothstein, P. D. McNicholas, and S. Subedi. A multivariate Poisson-lognormal mixture model for clustering transcriptome sequencing data. *BMC Bioinformatics*, 20(1):1–11, 2019.

A. Sklar. Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6):449–460, 1973.

G. Z. Stein and J. M. Juritz. Bivariate compound Poisson distributions. *Communications in Statistics - Theory and Methods*, 16(12):3591–3607, 1987.

G. Z. Stein, W. Zucchini, and J. M. Juritz. Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82(399):938–944, 1987.

G. Tzougas. EM estimation for the Poisson-inverse gamma regression model with varying dispersion: an application to insurance ratemaking. *Risks*, 8(3):97, 2020.

G. C. Wei and M. A. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704, 1990.

R. Winkelmann. *Econometric analysis of count data*. Springer Science & Business Media, 2008.

X. Zhan, H. A. Aziz, and S. V. Ukkusuri. An efficient parallel sampling technique for multivariate Poisson-lognormal model: Analysis with two crash count datasets. *Analytic Methods in Accident Research*, 8:45–60, 2015.

D. M. Zimmer and P. K. Trivedi. Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business & Economic Statistics*, 24(1): 63–76, 2006.

Appendices

Appendix A. Model specification and estimation scheme for random effects that follow Gaussian copula paired with gamma marginals

We use gamma distribution $\mathcal{G}(\sigma_i^{(l)}, 1/\sigma_i^{(l)})$ as $F_l(\cdot)$, the marginal distribution of the latent random variable $Z_i^{(l)}$ so that we have $\mathbb{E}[Z_i^{(l)}] = 1$. The variance of $Z_i^{(l)}$ is given by $Var(Z_i^{(l)}) = 1/\sigma_i^{(l)}$, which is a monotonic strictly decreasing function of $\sigma_i^{(l)}$, and hence condition (i) of the marginalization property is satisfied.

We use the same copula function $C_{\phi_i}(\cdot)$ as in (11) where $\phi_i = \frac{2}{\pi}\arctan(\boldsymbol{x}_{3,i}^T\boldsymbol{\beta}_3)$. Define uniform random vector $\boldsymbol{u}_i = (u_i^{(1)}, u_i^{(2)})$ where $u_i^{(l)} = F_l(z_i^{(l)})$ for $l = 1, 2$. Then we have the following joint density function of $\boldsymbol{Z}_i$:

$$\begin{aligned}\pi(\boldsymbol{z}_i) &= \frac{f_{\phi_i}\left(\Phi^{-1}(u_i^{(1)}), \Phi^{-1}(u_i^{(2)})\right)}{f_0\left(\Phi^{-1}(u_i^{(1)}), \Phi^{-1}(u_i^{(2)})\right)} f_1(z_i^{(1)}) f_2(z_i^{(2)}), \\ &:= c_{\phi_i}(\mathbf{u}_i) f_1(z_i^{(1)}) f_2(z_i^{(2)}),\end{aligned} \tag{17}$$

where $\Phi(\cdot)$ is the cdf of standard normal random variable with $\Phi^{-1}(\cdot)$ as its quantile function, $f_\rho(\cdot,\cdot)$ is the density function of bivariate normal random variable with the following expression:

$$f_\rho(v_1, v_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{v_1^2 - 2\rho v_1 v_2 + v_2^2}{2(1-\rho^2)}\right\}.$$

Finally, the joint pmf of $(N_i^{(1)}, N_i^{(2)})$ is given by the following double integral:

$$\begin{aligned} p\left(n_i^{(1)}, n_i^{(2)}\right) &= \int p\left(n_i^{(1)} | z_i^{(1)}\right) p\left(n_i^{(2)} | z_i^{(2)}\right) \pi(\boldsymbol{z}_i) d\boldsymbol{z}_i \\ &= \frac{{\mu_i^{(1)}}^{n_i^{(1)}}}{n_i^{(1)}!} \frac{{\mu_i^{(2)}}^{n_i^{(2)}}}{n_i^{(2)}!} \int_0^\infty \int_0^\infty \exp\left(-\mu_i^{(1)} z_i^{(1)} - \mu_i^{(2)} z_i^{(2)}\right) {z_i^{(1)}}^{n_i^{(1)}} {z_i^{(2)}}^{n_i^{(2)}} \pi(\boldsymbol{z}_i) dz_i^{(1)} dz_i^{(2)}, \end{aligned}$$

which can be evaluated by Gauss-Legendre quadrature. The extension of the Gauss-Legendre quadrature rule into high dimensional situation is much more straightforward. The following m-dimensional integral can be approximated as

$$\begin{aligned} &\int_{a_1}^{b_1} \cdots \int_{a_m}^{b_m} h(y_1, \ldots, y_n) dy_1 \ldots dy_n \\ \approx &\left(\prod_{i=1}^m \frac{b_i - a_i}{2}\right) \sum_{i_1=1}^n \cdots \sum_{i_m=1}^n h\left(\frac{b_1 - a_1}{2}\xi_{i_1} + \frac{b_1 + a_1}{2}, \ldots, \frac{b_n - a_n}{2}\xi_{i_n} + \frac{b_n + a_n}{2}\right) w_{i_1} \ldots w_{i_k}, \end{aligned}$$

where $\xi_\cdot$ are roots of Legendre polynomials of degree n and $w_\cdot$ are the corresponding weights. These can also be found easily in R by the 'gauss.quad' function from the package 'statmod'. Similarly, the h function can be well-approximated by a polynomial of degree $2n-1$ or less to ensure accuracy of the approximation.

For statistical inference of the above model, we can apply an MCEM algorithm to estimate the parameters $\theta = (\boldsymbol{\beta}_1^{(1)}, \boldsymbol{\beta}_1^{(2)}, \boldsymbol{\beta}_2^{(1)}, \boldsymbol{\beta}_2^{(2)}, \boldsymbol{\beta}_3)$ using the complete log-likelihood as in (14).

- **E-step:** Evaluate the Q-function in (15) given $\theta^{(r)}$, estimated value of θ at the r^{th} iteration.
- **M-step:** As in Section 3, we update the parameters $\boldsymbol{\beta}_1^{(1)}, \boldsymbol{\beta}_1^{(2)}, \boldsymbol{\beta}_2^{(1)}, \boldsymbol{\beta}_2^{(2)}$ and $\boldsymbol{\beta}_3$ sequentially using Newton-Raphson algorithm so that it suffices to find the form of $\boldsymbol{W}_{j1}^{(l;r)}$, $\boldsymbol{W}_{j2}^{(l;r)}$ $(j = 1, 2; l = 1, 2)$, $\boldsymbol{W}_{31}^{(r)}$ and $\boldsymbol{W}_{32}^{(r)}$ for iterative update. One can easily check that the diagonal matrices to update $\boldsymbol{\beta}_1^{(l)}$ are given by

$$\begin{aligned} \boldsymbol{W}_{11}^{(l;r)} &= \text{diag}\left(\{n_i^{(l)} - \mu_i^{(l)} w_i^{(l;r)}\}_{i=1,\ldots,n}\right), \\ \boldsymbol{W}_{12}^{(l;r)} &= \text{diag}\left(\{-\mu_i^{(l)} w_i^{(l;r)}\}_{i=1,\ldots,n}\right), \end{aligned}$$

by the same underlying mean structure of Poisson distributions.

By letting $\phi(\cdot)$ and $\psi(\cdot)$ be the standard normal density and digamma functions, respectively, one can show that

$$\begin{aligned} \frac{\partial Q(\theta; \theta^{(r)})}{\partial \sigma_i^{(l)}} &= \left(-\frac{\phi_i^2}{1-\phi_i^2} \frac{\Phi^{-1}(u_i^{(l)})}{\phi(\Phi^{-1}(u_i^{(l)}))} + \frac{\phi_i}{1-\phi_i^2} \frac{\Phi^{-1}(u_i^{(l')})}{\phi(\Phi^{-1}(u_i^{(l)}))}\right) \frac{\partial u_i^{(l)}}{\partial \sigma_i^{(l)}} \\ &\quad + 1 + \log(\sigma_i^{(l)}) - \psi\left(\sigma_i^{(l)}\right) + \eta_i^{(l;r)} - w_i^{(l;r)}, \qquad l' \neq l, \\ \frac{\partial Q(\theta; \theta^{(r)})}{\partial \phi_i} &= \frac{\phi_i}{1-\phi_i^2} - \frac{\phi_i}{(1-\phi_i^2)^2}\left(\Phi^{-1}(u_i^{(1)}) + \Phi^{-1}(u_i^{(2)})\right) + \frac{1+\phi_i^2}{(1-\phi_i^2)^2}\Phi^{-1}(u_i^{(1)})\Phi^{-1}(u_i^{(2)}), \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 Q(\theta; \theta^{(r)})}{\partial \sigma_i^{(l)} \partial \sigma_i^{(l)}} &= \left(-\frac{\phi_i^2}{1-\phi_i^2} \frac{1 + \Phi^{-1}(u_i^{(l)})^2}{\phi(\Phi^{-1}(u_i^{(l)}))^2} + \frac{\phi_i}{1-\phi_i^2} \frac{\Phi^{-1}(u_i^{(1)})\Phi^{-1}(u_i^{(2)})}{\phi(\Phi^{-1}(u_i^{(l)}))}\right) \left(\frac{\partial u_i^{(l)}}{\partial \sigma_i^{(l)}}\right)^2 \\ &\quad + \frac{\partial Q(\theta; \theta^{(r)})}{\partial \sigma_i^{(l)}} \frac{\partial^2 u_i^{(l)}}{\partial \sigma_i^{(l)} \partial \sigma_i^{(l)}} + \frac{1}{\sigma_i^{(l)}} - \psi'(\sigma_i^{(l)}), \\ \frac{\partial^2 Q(\theta; \theta^{(r)})}{\partial \phi_i^2} &= \frac{1+\phi_i^2}{(1-\phi_i^2)^2} - \frac{1+3\phi_i^2}{(1-\phi_i^2)^3}\left(\Phi^{-1}(u_i^{(1)}) + \Phi^{-1}(u_i^{(2)})\right) + \frac{2\phi_i^3 + 6\phi_i}{(1-\phi_i^2)^3}\Phi^{-1}(u_i^{(1)})\Phi^{-1}(u_i^{(2)}), \\ \frac{\partial^2 Q(\theta; \theta^{(r)})}{\partial \sigma_i^{(1)} \partial \sigma_i^{(2)}} &= \frac{\phi_i}{1-\phi_i^2} \frac{\frac{\partial u_i^{(1)}}{\partial \sigma_i^{(1)}} \frac{\partial u_i^{(2)}}{\partial \sigma_i^{(2)}}}{\phi(\Phi^{-1}(u_i^{(1)}))\phi(\Phi^{-1}(u_i^{(2)}))}. \end{aligned}$$

Based on the above computation, the diagonal matrices to update $\boldsymbol{\beta}_2^{(l)}$ and $\boldsymbol{\beta}_3$ are given as follows:

$$\boldsymbol{W}_{21}^{(l;r)} = \text{diag}\left(\left\{\sigma_i^{(l)} \times \frac{\partial Q(\theta;\theta^{(r)})}{\partial \sigma_i^{(l)}}\right\}_{i=1,\ldots,n}\right),$$

$$\boldsymbol{W}_{22}^{(l;r)} = \text{diag}\left(\left\{\sigma_i^{(l)} \times \frac{\partial Q(\theta;\theta^{(r)})}{\partial \sigma_i^{(l)}} + \sigma_i^{(l)2} \times \frac{\partial^2 Q(\theta;\theta^{(r)})}{\partial \sigma_i^{(l)} \partial \sigma_i^{(l)}} + \sigma_i^{(1)}\sigma_i^{(2)} \times \frac{\partial^2 Q(\theta;\theta^{(r)})}{\partial \sigma_i^{(1)} \partial \sigma_i^{(2)}}\right\}_{i=1,\ldots,n}\right),$$

$$\boldsymbol{W}_{31}^{(r)} = \text{diag}\left(\left\{\frac{2/\pi}{1+(\boldsymbol{x}_{3,i}^T\boldsymbol{\beta}_3)^2} \times \frac{\partial Q(\theta;\theta^{(r)})}{\partial \phi_i}\right\}_{i=1,\ldots,n}\right),$$

$$\boldsymbol{W}_{32}^{(r)} = \text{diag}\left(\left\{\frac{4/\pi^2}{(1+(\boldsymbol{x}_{3,i}^T\boldsymbol{\beta}_3)^2)^2} \times \frac{\partial^2 Q(\theta;\theta^{(r)})}{\partial \phi_i^2} + \frac{4\boldsymbol{x}_{3,i}^T\boldsymbol{\beta}_3/\pi}{(1+(\boldsymbol{x}_{3,i}^T\boldsymbol{\beta}_3)^2)^2} \times \frac{\partial Q(\theta;\theta^{(r)})}{\partial \phi_i}\right\}_{i=1,\ldots,n}\right).$$

Appendix B. Benchmarks for comparison

The BPLN regression model with the shared random effect

As a special case of the proposed model, consider a multivariate Poisson-lognormal mixture model with the shared random effect where $Z_i^{(l)} = Z_i$ and $\sigma_i = \sigma$ for $l = 1,\ldots,L$ and $i = 1,\ldots,n$ in (3) so that

$$p\left(n_i^{(l)}|z_i\right) = \frac{\exp[-(\mu_i^{(l)}z_i)](\mu_i^{(l)}z_i)^{n_i^{(l)}}}{n_i^{(l)}!}, \tag{18}$$

and

$$\pi(z_i;\sigma) = \frac{\exp\left[-\frac{(\log(z_i)+\sigma^2/2)^2}{2\sigma^2}\right]}{\sqrt{2\pi}\sigma z_i}, \tag{19}$$

where

$$\mu_i^{(l)} = \exp\left(\boldsymbol{x}_{1,i}^{(l)T}\boldsymbol{\beta}_1^{(l)}\right), \quad \mathbb{E}(z_i) = 1 \quad \text{and} \quad \text{Var}(z_i) = \exp(\sigma^2) - 1, \tag{20}$$

for $l = 1,\ldots,L$ and $i = 1,...,n$.

Thus, based on Equations (18 and 19), it is easy to see that the resulting distribution is the BNB distribution with the following joint pmf:

$$p(n_i^{(1)}, n_i^{(2)}) = \int_0^\infty \prod_{l=1}^2 \frac{\exp[-\mu_i^{(l)}z_i](\mu_i^{(l)}z_i)^{n_i^{(l)}}}{n_i^{(l)}!} \frac{\exp\left[-\frac{(\log(z_i)+\sigma^2/2)^2}{2\sigma^2}\right]}{\sqrt{2\pi}\sigma z_i} dz_i, \tag{21}$$

which could not be written in closed form and hence numerical integration is required.

Note that the correlation between $n_i^{(1)}$ and $n_i^{(2)}$ is given as

$$\text{Corr}\left(n_i^{(1)}, n_i^{(2)}\right) = \frac{\left(\exp(\sigma^2) - 1\right)}{\sqrt{\left[1/\mu_i^{(1)} + \exp(\sigma^2) - 1\right]\left[1/\mu_i^{(2)} + \exp(\sigma^2) - 1\right]}}. \tag{22}$$

As it can be seen from Equation (22), the bivariate Poisson regression model with shared random effects allows only positive correlation between the two types of claims.

The BPLN regression model with common covariates

As another special case of the proposed model, consider the multivariate Poisson-lognormal random effects model by Chiquet et al. [2020], which is equivalent to the model where $\boldsymbol{x}_{1,i}^{(l)T} = \boldsymbol{x}_{1,i}^T$, $\sigma_i^{(l)} = \sigma^{(l)}$, and $\phi_i = \phi$ for $l = 1,\ldots,L$ and $i = 1,\ldots,n$ in (3) so that the joint pmf is given as

$$p\left(n_i^{(1)}, n_i^{(2)}\right) = \int \prod_{l=1}^2 p\left(n_i^{(l)}|z_i^{(l)}\right)\pi(z_i)d\boldsymbol{z}_i, \tag{23}$$

where $\pi(\boldsymbol{z}_i)$ is defined by (11).

In this case, the correlation coefficient is given as

$$\mathrm{Corr}\left(n_i^{(1)}, n_i^{(2)}\right) = \frac{\left(\exp(\phi\sigma^{(1)}\sigma^{(2)}) - 1\right)}{\sqrt{\left[1/\mu_i^{(1)} + \exp(\sigma^{(1)2}) - 1\right]\left[1/\mu_i^{(2)} + \exp(\sigma^{(2)2}) - 1\right]}}. \tag{24}$$

As we observe Equation (24), this model specification does not allow the dispersions or dependence to vary through covariate information regarding different coverage types and the policyholders.