Multivariate claim count regression model with varying dispersion and dependence parameters

Himchan Jeong
(joint work with George Tzougas and Tsz Chai Fung)

Simon Fraser University

22 February, 2022

Structure of multivariate claim count data

- 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 $\boldsymbol{n}_i = (n_i^{(1)}, \ldots, n_i^{(L)})$ for $i = 1, \ldots, n$ 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.

Models for multivariate claim count data

(i) Multivariate Poisson models

Models for multivariate claim count data

(i) Multivariate Poisson models
(ii) Multivariate mixed Poisson models

Models for multivariate claim count data

(i) Multivariate Poisson models

(ii) Multivariate mixed Poisson models

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

Models for multivariate claim count data

(i) Multivariate Poisson models

(ii) Multivariate mixed Poisson models

   (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

Models for multivariate claim count data

(i) Multivariate Poisson models
(ii) Multivariate mixed Poisson models
    (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
(iii) Copula-based models

Challenges in the pre-existing models

- 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.

Challenges in the pre-existing models

- 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.

Challenges in pre-existing models

- 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.

Challenges in pre-existing models

- 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.
- 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.

The proposed model

- Mainly, we consider a multivariate mixed Poisson model with multiple random effects.

The proposed model

- Mainly, we consider a multivariate mixed Poisson model with multiple random effects.
- We also use coverage specific covariate information to model not only the mean component, but also varying dispersions or dependence upon individual characteristics and the characteristics of the coverage types.

Multivariate mixed Poisson regression model

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

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

Multivariate mixed Poisson regression model

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

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

- $Z_i^{(l)}$ is the random effect for l^{th} claim of poliyholder i that accounts for the associated unobserved heterogeneity.
- $\mathbb{E}[N_i^{(l)}|Z_i^{(l)}] = Var(N_i^{(l)}|Z_i^{(l)}) = \mu_i^{(l)} Z_i^{(l)}$
- Note that we assume that $\mathbb{E}[Z_i^{(l)}] = 1$ for the sake of model identifiability.

Multivariate mixed Poisson regression model

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

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

where

Multivariate mixed Poisson regression model

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

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

where

- $\boldsymbol{z}_i = (z_i^{(1)}, \ldots, z_i^{(L)})$,
- c_{Φ_i} is an elliptical copula density function that models dependence among the latent variables,
- f_l and F_l are marginal density and distribution functions of $z^{(l)}$, respectively, which are parameterized by the dispersion parameter $\sigma_i^{(l)}$.

Multivariate mixed Poisson regression model

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

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

Multivariate mixed Poisson regression model

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

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

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.

Multivariate mixed Poisson regression model

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

Multivariate mixed Poisson regression model

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

$$\mu_i^{(l)} \quad = \quad \exp\left(\boldsymbol{x}_{1,i}^{(l)\,T}\beta_1^{(l)}\right), \tag{4}$$

Multivariate mixed Poisson regression model

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

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

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

Multivariate mixed Poisson regression model

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

$$\mu_i^{(l)} = \exp\left(\boldsymbol{x}_{1,i}^{(l)T}\beta_1^{(l)}\right), \quad (4)$$

$$\sigma_i^{(l)} = \exp\left(\boldsymbol{x}_{2,i}^{(l)T}\beta_2^{(l)}\right) \text{ and} \quad (5)$$

$$\phi_i^{(l,l')} = g(\boldsymbol{x}_{3,i}^{(l,l')T}\beta_3^{(l,l')}), \quad (6)$$

Multivariate mixed Poisson regression model

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.

Model properties: Moments

The expectation, variance, covariance and correlations of the number of claims are given by

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

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

Model properties: Moments

It implies

Model properties: Moments

It implies

(i) $\text{sign}(\text{Corr}(N_i^{(l)}, N_i^{(l')})) = \text{sign}(\text{Corr}(Z_i^{(l)}, Z_i^{(l')}))$,

Model properties: Moments

It implies

(i) $\text{sign}(\text{Corr}(N_i^{(l)}, N_i^{(l')})) = \text{sign}(\text{Corr}(Z_i^{(l)}, Z_i^{(l')}))$,

(ii) $\text{Corr}(N_i^{(l)}, N_i^{(l')}) = \pm 1$ iff $\text{Corr}(Z_i^{(l)}, Z_i^{(l')}) = \pm 1$, $\mu_i^{(l)} Var(Z_i^{(l)}) \to \infty$ and $\mu_i^{(l')} Var(Z_i^{(l')}) \to \infty$.

Model properties: Marginalization

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

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

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.

Model properties: 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)

Model properties: 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) $\text{Corr}(Z_i^{(l)}, Z_i^{(l')})$ 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.

Model properties: 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) $\text{Corr}(Z_i^{(l)}, Z_i^{(l')})$ 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 $\theta = \{\boldsymbol{\beta}_1^{(l)}, \boldsymbol{\beta}_2^{(l)}, \boldsymbol{\beta}_3^{(l,l')} \mid l, l' = 1, \ldots, L\}$ and $\tilde{\theta} = \{\tilde{\boldsymbol{\beta}}_1^{(l)}, \tilde{\boldsymbol{\beta}}_2^{(l)}, \tilde{\boldsymbol{\beta}}_3^{(l,l')} \mid l, l' = 1, \ldots, L\}$, if and only if $\theta = \tilde{\theta}$.

Model specifications

Choice of copula

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

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

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.

Model specifications

Choice of copula

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

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

  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 $L = 2$.

Model specifications

Link function for dependence parameter

- When $L = 2$, we only have one dependence parameter $\phi_i^{(1,2)}$ in $\mathbf{\Phi}_i$. Therefore, from now on, we write $\phi_i := \phi_i^{(1,2)}$ for simplicity.

Model specifications

Link function for dependence parameter

- When $L = 2$, we only have one dependence parameter $\phi_i^{(1,2)}$ in $\mathbf{\Phi}_i$. Therefore, from now on, we write $\phi_i := \phi_i^{(1,2)}$ for simplicity.
- 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.

Model specifications

Marginal distribution of Z

- We assume lognormal distribution $\mathcal{LN}(-\frac{\sigma_i^{(l)2}}{2}, \sigma_i^{(l)2})$ is selected as $F_l(\cdot)$, the marginal distribution of the latent random variable $Z_i^{(l)}$, so that we have $\mathbb{E}[Z_i^{(l)}] = 1$ and $Var(Z_i^{(l)}) = \exp\{\sigma_i^{(l)2}\} - 1$.

Model specifications

Marginal distribution of Z

- We assume lognormal distribution $\mathcal{LN}(-\frac{\sigma_i^{(l)2}}{2}, \sigma_i^{(l)2})$ is selected as $F_l(\cdot)$, the marginal distribution of the latent random variable $Z_i^{(l)}$, so that we have $\mathbb{E}[Z_i^{(l)}] = 1$ and $Var(Z_i^{(l)}) = \exp\{\sigma_i^{(l)2}\} - 1$.
- Note that 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 $\sigma_i^{(l)}$. (e.g., gamma)

Model specifications

With the model specifications, we have the correlation coefficient between $N_i^{(1)}$ and $N_i^{(2)}$ as follows:

$$\text{Corr}\left(N_i^{(1)}, N_i^{(2)}\right) = \frac{\left(\exp(\phi_i \sigma_i^{(1)} \sigma_i^{(2)}) - 1\right)}{\sqrt{\left[1/\mu_i^{(1)} + \exp(\sigma_i^{(1)2}) - 1\right]\left[1/\mu_i^{(2)} + \exp(\sigma_i^{(2)2}) - 1\right]}}. \tag{10}$$

Statistical inference: The MCEM algorithm

By augmentation of the 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:

$$\ell_c(\theta) = \sum_{i=1}^{n} \left[\left(\sum_{l=1}^{L} n_i^{(l)} \log(\mu_i^{(l)} z_i^{(l)}) - \mu_i^{(l)} z_i^{(l)} - \log n_i^{(l)}! \right) + \log \pi(\boldsymbol{z}_i) \right], \tag{11}$$

where $\theta = (\beta_1^{(1)}, \ldots, \beta_1^{(L)}, \beta_2^{(1)}, \ldots, \beta_2^{(L)}, \beta_3)$ includes all parameters to be estimated.

Statistical inference: The MCEM algorithm

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

$$
\begin{aligned}
Q(\theta;\theta^{(r)}) &= \mathbb{E}_{z_i}[\ell_c(\theta)|\mathcal{D},\theta^{(r)}] \\
&\propto \sum_{i=1}^{n}\sum_{l=1}^{L} n_i^{(l)} \log \mu_i^{(l)} - \mu_i^{(l)}\mathbb{E}_{z_i}[z_i^{(l)}|\mathcal{D},\theta^{(r)}] \\
&+ \sum_{i=1}^{n}\mathbb{E}_{z_i}[\log \pi(\boldsymbol{z}_i)|\mathcal{D},\theta^{(r)}].
\end{aligned}
\tag{12}
$$

Statistical inference: The MCEM algorithm

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

$$\begin{aligned} Q(\theta;\theta^{(r)}) &= \mathbb{E}_{z_i}[\ell_c(\theta)|\mathcal{D},\theta^{(r)}] \\ &\propto \sum_{i=1}^{n}\sum_{l=1}^{L} n_i^{(l)}\log\mu_i^{(l)} - \mu_i^{(l)}\mathbb{E}_{z_i}[z_i^{(l)}|\mathcal{D},\theta^{(r)}] \\ &+ \sum_{i=1}^{n}\mathbb{E}_{z_i}[\log\pi(\boldsymbol{z}_i)|\mathcal{D},\theta^{(r)}]. \end{aligned} \tag{12}$$

where

$$\begin{aligned} \mathbb{E}_{z_i}[f(\boldsymbol{z}_i)|\mathcal{D},\theta^{(r)}] &= \int f(\boldsymbol{z}_i)\pi(\boldsymbol{z}_i|\mathcal{D})d\boldsymbol{z}_i \\ &\simeq \frac{\sum_{s=1}^{S} f(\boldsymbol{z}_{i[s]})\prod_{l=1}^{L} p\left(n_i^{(l)}|z_{i[s]}^{(l)}\right)}{\sum_{s=1}^{S}\prod_{l=1}^{L} p\left(n_i^{(l)}|z_{i[s]}^{(l)}\right)}. \end{aligned} \tag{13}$$

Statistical inference: The MCEM algorithm

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, $Q(\theta^{(r+1)}; \theta^{(r)}) \geq Q(\theta^{(r)}; \theta^{(r)})$.

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:

Statistical inference: The MCEM algorithm

1. Set $\theta \leftarrow \theta^{(r)}$. Recall that $\theta = (\beta_1^{(1)}, \beta_1^{(2)}, \beta_2^{(1)}, \beta_2^{(2)}, \beta_3)$.

1. Set $\theta \leftarrow \theta^{(r)}$. Recall that $\theta = (\beta_1^{(1)}, \beta_1^{(2)}, \beta_2^{(1)}, \beta_2^{(2)}, \beta_3)$.
2. For $l = 1, 2$ and $j = 1, 2$, update the parameters sequentially as $\beta_j^{(l)} \leftarrow \beta_j^{(l)} - [H_j^{(l;r)}(\theta)]^{-1} h_j^{(l;r)}(\theta)$, 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.

Statistical inference: The MCEM algorithm

1. Set $\theta \leftarrow \theta^{(r)}$. Recall that $\theta = (\beta_1^{(1)}, \beta_1^{(2)}, \beta_2^{(1)}, \beta_2^{(2)}, \beta_3)$.
2. For $l = 1, 2$ and $j = 1, 2$, update the parameters sequentially as $\beta_j^{(l)} \leftarrow \beta_j^{(l)} - [H_j^{(l;r)}(\theta)]^{-1} h_j^{(l;r)}(\theta)$, 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 $\beta_3 \leftarrow \beta_3 - [H_3^{(r)}(\theta)]^{-1} h_j^{(r)}(\theta)$, where $h_j(\theta)$ is a p_3-column vector and $H_j(\theta)$ is a $p_3 \times p_3$ matrix defined below.

Statistical inference: The MCEM algorithm

1. Set $\theta \leftarrow \theta^{(r)}$. Recall that $\theta = (\beta_1^{(1)}, \beta_1^{(2)}, \beta_2^{(1)}, \beta_2^{(2)}, \beta_3)$.
2. For $l = 1, 2$ and $j = 1, 2$, update the parameters sequentially as $\beta_j^{(l)} \leftarrow \beta_j^{(l)} - [H_j^{(l;r)}(\theta)]^{-1} h_j^{(l;r)}(\theta)$, 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 $\beta_3 \leftarrow \beta_3 - [H_3^{(r)}(\theta)]^{-1} h_j^{(r)}(\theta)$, 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 $\theta^{(r+1)} \leftarrow \theta$.

Data description

- We examine the Local Government Property Insurance Fund (LGPIF) data from the state of Wisconsin.

Data description

- We examine the Local Government Property Insurance Fund (LGPIF) data from the state of Wisconsin.
- While the LGPIF contains multiple groups of property insurance coverage, 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 $N_i^{(2)}$).

Data description

- We examine the Local Government Property Insurance Fund (LGPIF) data from the state of Wisconsin.
- While the LGPIF contains multiple groups of property insurance coverage, 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.

Data description

- We examine the Local Government Property Insurance Fund (LGPIF) data from the state of Wisconsin.
- While the LGPIF contains multiple groups of property insurance coverage, 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.

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

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 $i = 1, \ldots, n$.

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 $i = 1, \ldots, n$.

- The BPLN regression model with **common** covariates

  As another special case of the proposed model, consider the multivariate Poisson-lognormal random effects 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-sample 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	

Findings from the estimation results

- 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.

Findings from the estimation results

- 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.

Findings from the estimation results

- After taking account of model complexities, our proposed model still shows slight improvement in AIC yet produces inferior BIC compared to the shared model.

Findings from the estimation results

- After taking account of model complexities, our proposed model still shows slight improvement in AIC yet produces inferior BIC compared to the shared model.
- We briefly explain this issue as follows: $\hat{\beta}_3$ is significantly positive for all types of location so that $\hat{\phi}_i > 0$ 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.

Findings from the estimation results

- After taking account of model complexities, our proposed model still shows slight improvement in AIC yet produces inferior BIC compared to the shared model.
- We briefly explain this issue as follows: $\hat{\beta}_3$ is significantly positive for all types of location so that $\hat{\phi}_i > 0$ 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.

Findings from the estimation results

- 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.

Findings from the estimation results

- 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.

Findings from the estimation results

- 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),

Findings from the estimation results

- 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),
- because of 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).

Analysis of dependence

- 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.

Analysis of dependence

- 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.
- 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.

Analysis of dependence

- 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.

Analysis of dependence

- 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.

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

Out-of-sample validation

- 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.

Out-of-sample validation

- 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 following table.

	Proposed	Shared	Common
RMSE	0.4672	0.4664	0.5276
Deviance	444.0522	471.4090	633.0516

Concluding Remarks

- In this article, we considered a multivariate claim count regression model with varying dispersion and dependence parameters.

Concluding Remarks

- In this article, we considered a multivariate claim count regression model with varying dispersion and dependence parameters.
- 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.

Concluding Remarks

- 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.

Concluding Remarks

- 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.