SORT 44 (2) July-December 2020, 335-356 DOI: 10.2436/20.8080.02.105

Modelling multivariate, overdispersed count data with correlated and non-normal heterogeneity effects

Iraj Kazemi[1] and Fatemeh Hassanzadeh[2]

Abstract

Mixed Poisson models are most relevant to the analysis of longitudinal count data in various disciplines. A conventional specification of such models relies on the normality of unobserved heterogeneity effects. In practice, such an assumption may be invalid, and non-normal cases are appealing. In this paper, we propose a modelling strategy by allowing the vector of effects to follow the multivariate skew-normal distribution. It can produce dependence between the correlated longitudinal counts by imposing several structures of mixing priors. In a Bayesian setting, the estimation process proceeds by sampling variants from the posterior distributions. We highlight the usefulness of our approach by conducting a simulation study and analysing two real-life data sets taken from the German Socioeconomic Panel and the US Centers for Disease Control and Prevention. By a comparative study, we indicate that the new approach can produce more reliable results compared to traditional mixed models to fit correlated count data.

MSC: 60E05, 62J12, 62J99, 62H20.

Keywords: Bayesian computation, correlated random effects, hierarchical representation, longitudinal data, multivariate skew-normal distribution, over-dispersion.

1 Introduction

An important class of models for count data, in the presence of over-dispersion, is the mixed Poisson. The class includes several popular mixed-Poisson models in terms of choosing mixing priors for unobserved heterogeneity effects. The normal mixing prior was originally introduced by Bulmer (1974) and developed by many others, such as Guo and Trivedi (2002), Miller (2007), and Montesinos et al. (2017) among others. The mixing strategy generates a marginal distribution of longer-tailed than the routinely used Gamma prior, which creates the negative binomial (NB) model (Gonzales-Barron and Butler, 2011). It is also useful in analysing specific over-dispersed count response vari-

[1] Department of Statistics, University of Isfahan, Iran. Tel.: +98-31-37934596.
[2] Department of Statistics, University of Khansar, Iran. Tel.: +98-31-57995312.
Received: December 2019
Accepted: December 2020

ables (Izsák, 2008; Williams and Ebel, 2012). A familiar list of several mixed Poisson distributions is presented by Karlis and Xekalaki (2005), Nadarajah and Kotz (2006a), and Nadarajah and Kotz (2006b). Further models detailed in Kuba and Panholzer (2016) and Cameron and Trivedi (2013).

Count data analysis may involve dealing with both the occurrence of over-dispersion and the correlation between repeated outcomes. A comprehensive overview of the discrete correlated data analysis is provided by Molenberghs, Verbeke and Demetrio (2007) with a discussion on computational issues and the inclusion of many practical applications. In longitudinal studies, the presence of heterogeneity effects is an indication of correlated responses of each subject over time and possibly a sign of over-dispersion. In this scenario, a regular choice to explain variability is the Poisson-multivariate normal (PMN) model, wherein the distribution of effects is assumed to be multivariate normal (e.g., see Chib and Winkelmann, 2001; El-Basyouny and Sayed, 2009; Wu, Deng and Ramakrishnan, 2018). Then, the problem turns to solving an intractable marginal likelihood and requiring advanced computational techniques, such as the Markov chain Monte Carlo (MCMC) in the Bayesian framework.

The associated literature reveals that the multivariate normal is the most adopted mixing prior distribution to the heterogeneity effects. However, it is unlikely to lead always to the best-fitted model. It was our leading motivation to extend the PMN model by setting the multivariate skew-normal mixing prior (Azzalini, 1985; Sahu, Dey and Branco, 2003) for the conditional mean of the Poisson model. The proposed Poisson multivariate skew-normal (PMSN) regression model includes a vector of skewness parameters. Thus we can directly introduce it through an additional hierarchy level to the PMN model. Also, depending on the specific multivariate skew-normal mixing prior, we can define various types of the PMSN model. The proposed model includes Poisson and the PMN as its special cases. Also, the PMSN model reduces to the Poisson skew-normal (PSN) model when unobserved heterogeneity effects are assumed to be independent by introducing a skew-normal mixing prior distribution to the structure of the mixed Poisson model. Specifically, our findings show that the proposed model with various values of the skewness parameter has different performances. In particular, over-dispersion in counts increases as the value of the skewness parameter increases. Results reveal that the PSN over-dispersion is less (more) than the Poisson normal (PN) over-dispersion provided that the skewness parameter being negative (positive). It illustrates that the PSN regression model may be more flexible than the PN model if a count data set exhibits over-dispersion.

From a Bayesian perspective, the proposed models can appear hierarchically to ease the implementation of the Gibbs sampler technique. Also, we use a stochastic representation for the conditional mean of the Poisson regression. It simplifies Bayesian computations due to having the complete conditional posteriors, involved in the Gibbs sampler, in closed forms of known distributions. The Bayesian analysis of correlated count data by fitting the PMN model (e.g., Rizzato et al., 2016) is a specific case of our proposed model. The model fitting is performed by OpenBugs software version 3.2.3, which is an excellent platform for Bayesian inference using the Gibbs sampler algorithm (e.g., Lunn et al., 2009).

The article is organized as follows. In Section 2, we introduce the PSN model with independent heterogeneity effects for the analysis of count data. In Section 3, mixed-Poisson models with various multivariate skew-normal mixing priors are illustrated for longitudinal count data. In this section, we also emphasize the identification issue in mixed-Poisson models. In Section 4, we present Bayesian mixed models hierarchically to derive the complete conditional posteriors required to implement the Gibbs sampling approach. In Section 5, we conduct a simulation study to compare proposed models with some competing ones. In Section 6, we fit proposed models for the specific data sets taken from follow up studies on the national medical expenditure survey and the polio data. Section 7 gives some concluding remarks.

2 A new modelling methodology to the count data analysis

Assume that the count response Y_{it}, conditioned on the effect u_{it} for subject $i = 1, \cdots, n$ and at time $t = 1, \cdots, T$, follows a Poisson distribution with mean $\exp(\theta_{it})$, where $\theta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}$, $\mathbf{x}_{it}$ is a k-dimensional vector of covariates, and $\boldsymbol{\beta}$ is a k-dimensional vector of coefficients. Moreover, the effects u_{it}, defined on the whole real line, are assumed to follow a common probability distribution function (pdf) $G(u_{it}|\boldsymbol{\eta})$, where $\boldsymbol{\eta}$ is a vector of parameters that characterize $G(\cdot)$. The marginal density of Y_{it} is called a mixed Poisson density with the probability mass function (pmf) given by integrating out the effects u_{it}. The normal mixing prior for u_{it} leads to the well-known Poisson normal (PN) model. Here, we extend the methodology by letting the mixing prior be skew-normally distributed with the following specification.

Definition 1 *The random variable u_{it}, for subject $i = 1, \cdots, n$ and at time $t = 1, \cdots, T$, follows the skew-normal distribution, denoted by $u_{it} \overset{iid}{\sim} SN(\xi, \sigma^2, \delta)$, if the density function of u_{it} is given by*

$$g_{SN}\left(u_{it}|\xi, \sigma^2, \delta\right) = 2\varphi\left(u_{it}|\xi, \sigma^2 + \delta^2\right)\Phi\left(\frac{\delta\,(u_{it} - \xi)}{\sigma\sqrt{\sigma^2 + \delta^2}}\right), \tag{1}$$

with location parameter $\xi \in \mathbb{R}$, scale parameter $\sigma^2 \in \mathbb{R}^+$ and skewness parameter $\delta \in \mathbb{R}$, where $\varphi(\cdot)$ denotes the pdf of $N\left(\xi, \sigma^2 + \delta^2\right)$ and $\Phi(\cdot)$ denotes the cumulative density function (cdf) of the standard normal (Azzalini, 1985; Sahu et al., 2003).

Using usual statistical methods the following basic properties of density (1) hold.

Properties 1

i. For $\delta = 0$, the original normal mixing prior is retrieved; for $\delta > 0$, positively skewed and for $\delta < 0$, negatively skewed mixing priors are obtained. Figure 1 confirms these results.
ii. The hierarchical representation of u_{it} is shown to be $u_{it}|z_{it} \overset{ind}{\sim} N\left(\xi + \delta z_{it}, \sigma^2\right)$ with $Z_{it} \overset{iid}{\sim} HN(0,1)$, where HN denotes the half-normal distribution. This property helps us to generate a random variable that follows the skew-normal distribution and consequently to implement the McMC approach easily.
iii. The r-th moment of $w_{it} = \exp(u_{it})$, for any real r, is finite and equivalent to the moment generating function (MGF) of the skew-normal distribution. This is explicitly given by $m_r = E\left(w_{it}^r\right) = 2\Phi(\delta r)\, exp(r\xi + \frac{1}{2}r^2\left(\sigma^2 + \delta^2\right))$. In particular, the mean and variance of w are $\mu_w = m_1$ and $\sigma_w^2 = m_2 - m_1^2$, respectively.

Without loss of generality, we set $\xi = 0$ then in what follows we use notation $SN\left(\sigma^2, \delta\right)$ for simplicity. This defines the Poisson skew-normal (PSN) regression model as follows, where $'$ denotes vector transpose.

Definition 2 *Let for subject $i = 1, 2, \cdots, n$ and at time $t = 1, 2, \cdots, T$ the count variable $Y_{it}|u_{it} \overset{ind}{\sim} Pois(\exp(\theta_{it}))$, where $\theta_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$ and $u_{it} \overset{iid}{\sim} SN\left(\sigma^2, \delta\right)$. Then, the pmf of Y_{it} is of the form*

$$f_{PSN}\left(y_{it}|\boldsymbol{\beta}, \sigma^2, \delta\right) = \int_{-\infty}^{\infty} f_{Pois}\left(y_{it}|u_{it}, \boldsymbol{\beta}\right) g_{SN}\left(u_{it}|\sigma^2, \delta\right) du_{it}, \tag{2}$$

where $f_{Pois}\left(y_{it}|u_{it}, \boldsymbol{\beta}\right)$ is the conditional pmf of Poisson given u_{it}. We denote $Y_{it} \overset{ind}{\sim} PSN\left(\boldsymbol{\beta}, \sigma^2, \delta\right)$.

Clearly, the PN model is a special case of (2) when $\delta = 0$. Let $\mu_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta})$. By conducting algebraic operations, some properties of (2) are shown below in which any clear proof is omitted.

Properties 2

i. The mean and variance of Y_{it} are shown to be $E(Y_{it}) = \mu_{it}\mu_w$ and $\mathrm{var}(Y_{it}) = \mu_{it}\left(\mu_w + \mu_{it}\sigma_w^2\right)$ so that the heterogeneity factor is $\left(\mu_w + \mu_{it}\sigma_w^2\right)/\mu_w$.
ii. The PSN is unimodal.

   Proof. Since the skew-normal is unimodal thus the marginal mixed Poisson is also unimodal (Holgate, 1970).

iii. The PSN tends to $Pois(\mu_{it})$ as both σ^2 and δ tend to zero.

   Proof. The normal mixing prior is regained for u_{it} as $\delta \to 0$. Then, using the transformation $v_{it} = exp(u_{it}/\sigma)$, the pmf (2) can be written as $E_{v_{it}}\left\{f_{Pois}\left(y_{it}|\mu_{it}v_{it}^{\sigma}\right)\right\}$, where v_{it} is log-normally distributed with $f_{v_{it}}(v_{it}) = 2\varphi(log(v_{it}))$, $v_{it} \in R^+$. Taking limit as $\sigma^2 \to 0^+$ this expectation becomes $f_{Pois}(y_{it}|\mu_{it})$.

iv. For fixed μ_{it} and non-zero δ, let $\sigma^2 \to 0^+$. Then (2) tends to a mixed Poisson density with a truncated normal mixing density supported on the left-bounded interval $(0,\infty)$, for $\delta > 0$, and on the right bounded interval $(-\infty,0)$, for $\delta < 0$.

Proof. We first derive the limiting case of $g_{SN}\left(u_{it}|\sigma^2,\delta\right)$ as $\sigma^2 \to 0^+$. It is easy to show that the density tends to $g_1(u_{it}|\delta) = 2\varphi\left(u_{it}|0,\delta^2\right)$ for $\text{sign}(u_{it}\delta) = 1$, and to 0 otherwise, where $sign(\cdot)$ denotes the sign function. Thus, the random variable Y_{it} follows a mixed Poisson (2) with the mixing prior g_1.

v. For fixed μ_{it} and σ^2 the probability $Pr(Y_{it} = 0)$ is a decreasing function of δ.

Proof. By setting $y_{it} = 0$ in (2) and taking the transformation $Z_{it} = \log(w_{it})$, the first derivative of the probability of zero is given by

$$\frac{\partial}{\partial\delta} f_{PSN}\left(0|\boldsymbol{\beta},\sigma^2,\delta\right) = \frac{\sigma^2}{\sqrt{\sigma^2+\delta^2}} E_{Z_{it}}\left(Z_{it}e^{-\mu_{it}e^{Z_{it}}}\right),$$

where $Z_{it} \sim N\left(0,\sigma^2\right)$. The involved expectation is shown to be negative. Then, after some manipulation twice of this expectation turns into $E\left(|Z_{it}|e^{-\mu_{it}e^{|Z_{it}|}}\right) - E\left(|Z_{it}|e^{-\mu_{it}e^{-|Z_{it}|}}\right)$. This expression is negative since the first expectation is less than the second one. This property is also illustrated by Figure 1.

vi. The probability of Y_{it} being zero is greater than the corresponding probability for a Poisson distribution with the same mean $\mu_{it}\mu_w$.
Proof. We have $f_{PSN}\left(0|\boldsymbol{\beta},\sigma^2,\delta\right) = E_{w_{it}}\left(e^{-\mu_{it}w_{it}}\right)$ and by using the Jensen's inequality, this becomes greater than $e^{-\mu_{it}\mu_w} = f_{Pois}(0|\mu_{it}\mu_w)$.

Figure 1 indicates the pmf of PSN for $\mu_{it} \equiv \mu = 3$, $\sigma^2 = 1$ and $\delta = -2,-1,0,1,2$. It is seen that the PSN is skewed right. Also, the tail of the PSN distribution is longer than

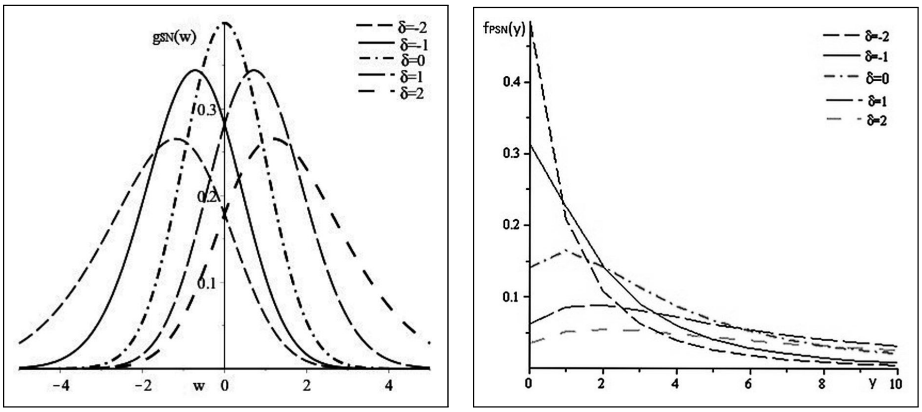

***Figure 1**: (a) Probability density functions of the skew-normal distribution (b) Probability mass functions of the PSN distribution.*

the tail of the PN distribution for positive values of δ, while is shorter for negative values of δ. Furthermore, the probability of zero counts increases as δ decreases.
The dispersion index, defined by the ratio of the variance to the mean, is given by

$$DI_{it}\left(\mu_{it},\sigma^2,\delta\right)=\frac{\operatorname{var}(Y_{it})}{E(Y_{it})}=1+\mu_{it}e^{\frac{1}{2}(\sigma^2+\delta^2)}\left\{\frac{\Phi(2\delta)\,e^{\left(\sigma^2+\delta^2\right)}-2\Phi^2(\delta)}{\Phi(\delta)}\right\}. \quad (3)$$

This indicates that $DI_{it}>1$, with strict inequality if the mixing distribution is non-degenerate, i.e., the mixing strategy can deal with additional variation present in count data. If $\delta=0$ then (3) reduces to the DI of the PN regression model, denoted by $DI_{it}\left(\mu_{it},\sigma^2,0\right)$. The difference between the dispersion index of two densities, $DDI_{it}\left(\mu_{it},\sigma^2,\delta\right)=DI_{it}\left(\mu_{it},\sigma^2,\delta\right)-DI_{it}\left(\mu_{it},\sigma^2,0\right)$, is shown in Figure 2. Negative and positive values of DDI_{it} show an advantage over the PN model. This indicates that the proposed model is more flexible than the PN model for dealing with over-dispersion in count data. Specifically, we set $\mu_{it}\equiv\mu=3$, $\sigma^2=1$ and $\delta\in(-1,1)$ that gives $DDI(\delta)\in(-2.446,47.769)$. Figure 2 illustrates that the PSN dispersion index is more than the PN dispersion index provided that $\delta>0$ while the difference $DDI(\delta)$ is negative for $\delta<0$. The differences increase as δ increases. We can also show by graphical techniques that if $\delta<0$ then the quantity $DDI(\delta)$ is positive over $\sigma^2\in\left(0,\sigma_0^2\right)$ for some small σ_0^2, whereas it is always negative over an interval $\sigma^2\in\left(\sigma_0^2,\infty\right)$. For any fixed δ the absolute value of $DDI(\delta)$ increases as σ^2 increases. Also, $DDI(\delta)<0$ for $\delta<0$, while $DDI(\delta)>0$ for $\delta>0$. These graphics are not shown here to save space.

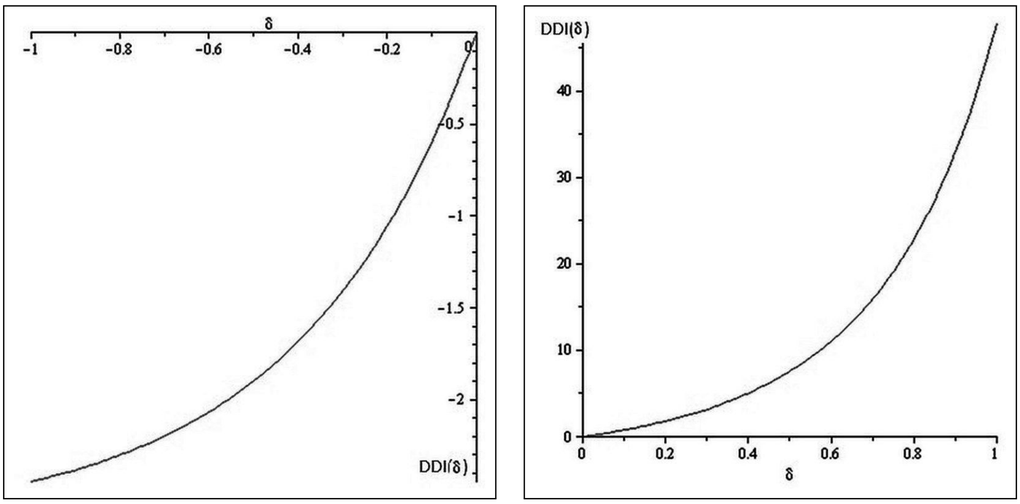

Figure 2: *Difference between the DIs for (a) negative and (b) positive values of δ.*

3 The proposed multivariate strategy for correlated count data

In longitudinal studies to count data, responses of each subject over time are usually correlated due to the existence of subject heterogeneity effects. Also, there may be evidence of over-dispersion in the structure of count responses. In many applications, there are situations where over-dispersion and the correlation between repeated outcomes can simultaneously occur. Here, we propose the multivariate skew-normal mixing prior distribution in the mean structure of mixed Poisson models to make a more adaptable analysis of correlated count responses. This strategy proceeds within the context of Bayesian hierarchical modelling together with several constructed specifications to fit related regression models. The specification of these proposed Poisson multivariate skew-normal (PMSN) models relies mostly on making different assumptions for the underlying multivariate skew-normal mixing priors.

3.1 The multivariate skew-normal mixing priors

Definition 3 *A T-dimensional random vector* $\mathbf{u}$ *follows the multivariate skew-normal distribution with location vector* $\boldsymbol{\xi} \in \mathbb{R}^T$*, positive-definite scale matrix* $\mathbf{V}$*, and skewness vector* $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_T)' \in \mathbb{R}^T$*, if its pdf is of the form*

$$f(\mathbf{u}_i|\boldsymbol{\xi}, \mathbf{V}, \boldsymbol{\delta}) = 2\varphi_T\left(\mathbf{u}_i|\boldsymbol{\xi}, V + \boldsymbol{\delta}\boldsymbol{\delta}'\right)\Phi\left(\frac{\boldsymbol{\delta}'\mathbf{V}^{-1}(\mathbf{u}_i - \boldsymbol{\xi})}{\sqrt{1+\boldsymbol{\delta}'\mathbf{V}^{-1}\boldsymbol{\delta}}}\right), \tag{4}$$

where $\varphi_T(.)$ *is the pdf of T-variate normal and* $\Phi(\cdot)$ *is the standard normal cdf. We denote* $\mathbf{u}_i \sim SN_T(\boldsymbol{\xi}, \mathbf{V}, \boldsymbol{\delta})$*.*

The density function (4) defines an attractive alternative to the multivariate skew-normal distribution introduced previously by Sahu et al. (2003) since instead of the evaluation of complex function $\Phi_T(\cdot)$, one needs only to compute one dimensional integral $\Phi(\cdot)$. The Poisson multivariate normal (PMN) model is a special case of (4) when $\boldsymbol{\delta} = \mathbf{0}$.

Properties 3 *The following properties hold for* $\mathbf{u}_i \sim SN_T(\boldsymbol{\xi}, \mathbf{V}, \boldsymbol{\delta})$*:*

i. The hierarchical representation is given by

$$\mathbf{u}_i|Z_i = z_i \overset{ind}{\sim} N_T(\boldsymbol{\xi} + \boldsymbol{\delta} z_i, \mathbf{V}) \text{ with } Z_i \overset{iid}{\sim} HN(0,1), \tag{5}$$

Thus, the mean vector and covariance matrix of $\mathbf{u}_i$ can be derived relatively easy. We obtain $E(\mathbf{u}_i) = \boldsymbol{\xi} + \boldsymbol{\delta}\sqrt{\frac{2}{\pi}}$, and $\text{var}(\mathbf{u}_i) = \mathbf{V} + \left(1 - \frac{2}{\pi}\right)\boldsymbol{\delta}\boldsymbol{\delta}'$.

ii. For any vector $\mathbf{r} = (r_1, \cdots, r_T)' \in \mathbb{R}^T$ the MGF is found to be

$$E\left(e^{\mathbf{r}'\mathbf{u}_i}\right) = 2\Phi\left(\mathbf{r}'\boldsymbol{\delta}\right)\exp\left\{\mathbf{r}'\boldsymbol{\xi} + \frac{1}{2}\mathbf{r}'\mathbf{V}\mathbf{r} + \left(\boldsymbol{\delta}'\mathbf{r}\right)^2\right\}. \tag{6}$$

Now, let $w_{it} = \exp(u_{it})$ be an element of the vector $w_i = (w_{i1}, \cdots, w_{iT})'$. Equation (6) is equivalent to $E\left(\prod_{t=1}^{T} w_{it}^{r_t}\right)$ which shows that all moments of w_{it}, including $E(w_i) = \boldsymbol{\mu}_w = \left(\mu_{w_{i1}}, \cdots, \mu_{w_{iT}}\right)'$ and $\text{var}(w_i) = D_w$, can be found easily. Specifically, putting the t-th element of $\mathbf{r}$ equal to one and zero otherwise, gives $\mu_{w_{it}}$, and when $r_t = r_s = 1$ and 0 otherwise, $E(w_{it}w_{is})$ is attained. In fact, we derive

$$\begin{aligned} \mu_{w_{it}} &= 2\Phi(\delta_t)\, e^{\frac{1}{2}\left(\delta_t^2 + \sigma_{tt}\right)}, \\ \sigma_{w_{ts}} &= 2e^{\frac{1}{2}\left(\delta_t^2 + \delta_s^2 + \sigma_{tt} + \sigma_{ss}\right)}\left\{e^{\delta_t\delta_s + \sigma_{ts}}\Phi(\delta_t + \delta_s) - 2\Phi(\delta_t)\Phi(\delta_s)\right\}, \end{aligned} \tag{7}$$

where the $\sigma_{w_{ts}}$ and σ_{ts} are, respectively, elements of $\mathbf{D}_w$ and $\mathbf{V}$.

iii. Let $c = \mathbf{a}'\mathbf{u}_i$ for any $\mathbf{a} \in \mathbb{R}^T$ then c follows the univariate skew-normal distribution, i.e. $c \sim SN\left(\mathbf{a}'\boldsymbol{\xi}, \mathbf{a}'\mathbf{V}\mathbf{a}, \mathbf{a}'\boldsymbol{\delta}\right)$.

Without loss of generality, in what follows we set $\boldsymbol{\xi} = \mathbf{0}$ and denote $\mathbf{u}_i \sim SN_T(\mathbf{V}, \boldsymbol{\delta})$ for simplicity. In model multivariate skew-normal specified by (4) no specific form of V and $\boldsymbol{\delta}$ is introduced in the data analysis process. It is mostly advisable in practice to explore possible causes of heterogeneity by allowing some specific forms for the u_{it}'s. Without having any knowledge on the source of heterogeneity, a priori justification is to allow u_{it}'s being into the one-way random effects framework. More specifically, let the $\mathbf{u}_i$ be of the familiar form $\mathbf{u}_i = \alpha_i\mathbf{1}_T + \boldsymbol{\varepsilon}_i$, where $\mathbf{1}_T$ denotes a unit vector of order T, the α_i represent the heterogeneity effects and the $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iT})'$ denote the residual terms that may reflect time-varying effects such as the effect of unobserved omitted covariates. In this setting, we specify the following types of the multivariate skew-normal distribution.

Remark 1 *For the above specified multivariate skew-normal model, let* $\alpha_i \overset{iid}{\sim} N\left(0, \sigma_\alpha^2\right)$ *and* $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} SN_T(\mathbf{V}_\varepsilon, \boldsymbol{\delta})$ *be all mutually independent. Then* $\mathbf{u}_i \overset{iid}{\sim} SN_T(\mathbf{D}, \boldsymbol{\delta})$ *where* $\mathbf{D} = \sigma_\alpha^2\mathbf{1}_T\mathbf{1}_T' + \mathbf{V}_\varepsilon$.

For the case with $\mathbf{V}_\varepsilon$ diagonal, we obtain

$$\text{corr}(w_{it}, w_{is}) = \frac{e^{\delta_t\delta_s + \sigma_\alpha^2}\Phi(\delta_t + \delta_s) - 2\Phi(\delta_t)\Phi(\delta_s)}{\sqrt{e^{\sigma_{tt} + \delta_t^2 + \sigma_\alpha^2}\Phi(2\delta_t) - 2\Phi^2(\delta_t)}\sqrt{e^{\sigma_{ss} + \delta_s^2 + \sigma_\alpha^2}\Phi(2\delta_s) - 2\Phi^2(\delta_s)}}, \tag{8}$$

for any $t \neq s$. Note that, the correlation coefficient (8) may take negative or positive values in the interval (-1,1).

Remark 2 *For the familiar form* $\mathbf{u}_i = \alpha_i \mathbf{1}_T + \boldsymbol{\varepsilon}_i$*, let* $\alpha_i \overset{iid}{\sim} SN\left(\sigma_\alpha^2, \delta\right)$ *and* $\boldsymbol{\varepsilon}_i \overset{iid}{\sim} N_T\left(\mathbf{0}, \mathbf{V}_\varepsilon\right)$ *be all mutually independent. Then* $\mathbf{u}_i \overset{iid}{\sim} SN_T\left(\mathbf{D}, \delta \mathbf{1}_T\right)$.

The correlation between w_{it} and w_{is} is a special case of (8) when δ_t and δ_s are replaced by constant δ for all t,s.

3.2 The Poisson multivariate skew-normal model

Let Y_{it} be the response variable and u_{it} be the corresponding heterogeneity effect of subject i at time period t for $i = 1,2,\cdots,n$ and $t = 1,2,\cdots,T$. The scheme of a PSN regression model in (2) allows for over-dispersion in the Poisson model but without taking into account the correlation among events. A common way to deal with this issue is to allow the vector $\mathbf{u}_i = (u_{i1},\cdots,u_{iT})'$ to follow a multivariate distribution with correlation amongst $u_{i1},\cdots,u_{iT}$, and consequently induce correlated $Y_{i1},\cdots,Y_{iT}$. A frequent assumption is multivariate normality of the $\mathbf{u}_i$. An alternative is to utilize a multivariate skew-normal distribution. Several versions of the multivariate skew-normal distribution, originally introduced by Azzalini and Dalla Valle (1996), have appeared in the literature. We present below a slight alteration of this distribution and provide its main properties that are related to our current work.

Definition 4 *Let the response vectors* $\mathbf{Y}_i = \{Y_{it}\}$ *of order T be independent for subjects* $i = 1,\cdots,n$ *and each* Y_{it} *conditioned on the effect* $\mathbf{u}_i$ *follows Poisson with the conditional mean exp*(θ_{it})*, where* $\theta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}$ *and* $\mathbf{u}_i = \{u_{it}\} \overset{iid}{\sim} SN_T\left(\mathbf{V}, \boldsymbol{\delta}\right)$*. Then the marginal pmf of* $\mathbf{Y}_i$ *is given by*

$$f_{PMSN}\left(\mathbf{y}_i|\boldsymbol{\beta},\mathbf{V},\boldsymbol{\delta}\right) = \int_{\mathbb{R}^T} \prod_{t=1}^{T} f_{Pois}\left(y_{it}|u_{it},\boldsymbol{\beta}\right) g_{MSN}\left(\mathbf{u}_i|\mathbf{V},\boldsymbol{\delta}\right) d\mathbf{u}_i, \tag{9}$$

where $g_{MSN}\left(\mathbf{u}_i|\mathbf{V},\boldsymbol{\delta}\right)$ *denotes the multivariate skew-normal density function for the i-th subject. We denote* (9) *as model PMSN*1.

The solution of (9) is not generally available in closed form. Thus, an MCMC scheme is implemented later to make statistical inferences. Furthermore, through standard calculation (see the supplementary Appendix A) we can straightforwardly show that

$$E\left(\mathbf{Y}_i\right) = \mathbf{M}_i \boldsymbol{\mu}_w, \text{ and } \operatorname{var}\left(\mathbf{Y}_i\right) = \mathbf{M}_i \mathbf{D}_w \mathbf{M}_i + \mathbf{M}_i \mathbf{M}_w, \tag{10}$$

where $\mathbf{M}_w$ and $\mathbf{M}_i$ are diagonal matrices with the elements $\mu_{w_{it}}$ and μ_{it} for $t = 1,2,\cdots,T$, respectively. The corresponding correlation coefficients between counts Y_{it} and Y_{is} are given by

$$\operatorname{corr}(Y_{it}, Y_{is}) = \operatorname{corr}(w_{it}, w_{is}) \sqrt{\frac{\mu_{it}}{\mu_{it} + \frac{\mu_{w_{it}}}{\sigma_{w_{tt}}}}} \sqrt{\frac{\mu_{is}}{\mu_{is} + \frac{\mu_{w_{is}}}{\sigma_{w_{ss}}}}}, \tag{11}$$

for all i, t, and s. Equation (11) shows that the two correlations $\operatorname{corr}(Y_{it}, Y_{is})$ and $\operatorname{corr}(w_{it}, w_{is})$ have the same sign and that $|\operatorname{corr}(Y_{it}, Y_{is})| < |\operatorname{corr}(w_{it}, w_{is})|$. Also, negative and positive correlations are allowed by using these mixed models. This fact gives an advantage over other multivariate models for discrete outcomes such as multinomial or negative multinomial models that allow only positive correlation. We specify below two types of the PMSN1 model.

Definition 5 *Let* $\mathbf{u}_i \overset{iid}{\sim} SN_T(\mathbf{D}, \boldsymbol{\delta})$. *We denote the corresponding model as PMSN2.*

The mean vector and covariance matrix of $\mathbf{Y}_i$ are derived to be particular cases of (10) and setting (7) in which the scalar σ_{ts}, for $t, s = 1, 2, \cdots, T$, turns into $\sigma_{ts} + \sigma_\alpha^2$, i.e. elements of $\mathbf{D}$. In this model, the corresponding correlation coefficient may take negative or positive values.

Definition 6 *Let* $\mathbf{u}_i \overset{iid}{\sim} SN_T(\mathbf{D}, \delta \mathbf{1}_T)$. *We denote the corresponding model as PMSN3.*

By utilizing the correlation between w_{it} and w_{is}, the resultant equation is always positive showing that PMSN3 permits only positive correlation between events. For a model with only constant term (no explanatory variable) Equations (10) can be simplified as $\operatorname{var}(Y_{it}) = c\mu_v\mu_\psi + c^2\left(\sigma_\psi^2\sigma_v^2 + \mu_v^2\sigma_\psi^2 + \mu_\psi^2\sigma_v^2\right)$, $\operatorname{cov}(Y_{it}, Y_{is}) = c^2\mu_v^2\sigma_\psi^2$, $t \neq s$ where $c = \exp(\beta_0)$ and parameters μ_v, σ_v^2, μ_ψ and σ_ψ^2 denote, correspondingly,means and variances of $v_{it} = \exp(\varepsilon_{it})$ and $\psi_i = \exp(\alpha_i)$ in $\mathbf{u}_i = \alpha_i \mathbf{1}_T + \boldsymbol{\varepsilon}_i$. It follows that

$$\operatorname{corr}(Y_{it}, Y_{is}) = \frac{c\mu_v^2\sigma_\psi^2}{1 + c\left(\sigma_\psi^2\sigma_v^2 + \mu_v^2\sigma_\psi^2 + \mu_\psi^2\sigma_v^2\right)}, \; t \neq s. \tag{12}$$

If the estimate of (12) is statistically significant then the PMSN model fits better to the data set than the standard Poisson regression model.

3.3 An alternative to deal with the identification issue

In the literature of mixed Poisson models the identification is usually addressed by allowing a restriction to the estimation process in order to make estimable the model parameters. To clarify this, let the count Y_{it}, for subject $i = 1, 2, \cdots, n$ and at time $t = 1, 2, \cdots, T$, follows the PSN distribution in (2), where $\log(E(Y_{it})) = \mathbf{x}_{it}'\boldsymbol{\beta} + \log(\mu_w)$. A common approach used by many researchers (e.g. see Balakrishnan and Peng, 2006) is to reparameterize the mixing distribution such that $\mu_w = 1$ to ensure that the logarithm of the marginal expectation of counts is $\mathbf{x}_{it}'\boldsymbol{\beta}$. This is equivalent to solving the nonlinear equation $2\Phi(\delta) = \exp\left\{-0.5\left(\delta^2+\sigma^2\right)\right\}$ for δ. However, this method does not work well when the expectation of exponentiated unobserved heterogeneity has a complex structure. Also, it may cause difficulties in the process of optimization routines for the reparameterized model. Thus, we use an alternative trick by setting the regression parameter β_0 to be equal to $\log(\mu_w)$. A similar trick can be done for models PMSN1–PMSN3 by setting $\beta_{0t} = \log(\mu_{w_t})$ since μ_{w_t} depends on time t.

4 The computational scheme

This section develops an operational MCMC scheme for the Bayesian analysis of the proposed regression models. We utilize the Bayesian data-augmentation method (e.g., Albert and Chib, 1993), which lets us generate the heterogeneity effects along with other known quantities in the simulation process. We use the hierarchical representation of all models to write down the joint likelihood of responses and heterogeneity effects. This representation is quite useful to estimate parameters by using the MCMC technique. To complete the model specifications from a Bayesian perspective, we assume that all parameters are independent. Then, we assign conditionally semi-conjugate priors to these parameters. This choice simplifies computations since the complete conditional posteriors involved in the Gibbs sampler are mostly closed forms of known distributions and hence easy for simulation. It this section, we let $Z_{it} \overset{iid}{\sim} HN(0,1)$ and $Z_i \overset{iid}{\sim} HN(0,1)$.

4.1 Bayesian computation for independent data

To fit the PSN model, we use the data augmentation to θ_{it} based on the hierarchical representation of the skew-normal distribution given in Properties 1 (ii). The related hierarchical form becomes

$$\begin{aligned} Y_{it}|\theta_{it} &\overset{ind}{\sim} Pois(\exp(\theta_{it})), \\ \theta_{it}|z_{it},\boldsymbol{\beta},\sigma^2,\delta &\overset{ind}{\sim} N\left(\mathbf{x}_{it}'\boldsymbol{\beta}+\delta z_{it},\sigma^2\right), \end{aligned} \tag{13}$$

for subject $i = 1,2,\cdots,n$ and at time $t = 1,2,\cdots,T$. By adopting all parameters to be independent, we assign the priors $\boldsymbol{\beta} \sim N_k\left(\boldsymbol{\beta}_0,\mathbf{V}_{\boldsymbol{\beta}}\right)$, $\delta \sim N\left(\delta_0,\sigma_\delta^2\right)$, and an inverse-Gamma, $IG(\nu_0,\nu_0)$, for σ^2, where all hyperparameters are known. The joint posterior density of $\boldsymbol{\beta}$, σ^2 , δ, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_n)'$ and $\mathbf{z} = (\mathbf{z}_1,\cdots,\mathbf{z}_n)'$ with $\boldsymbol{\theta}_i = (\theta_{i1},\cdots,\theta_{iT})'$ and $\mathbf{z}_i = (z_{i1},\cdots,z_{iT})'$ is then given by

$$\begin{aligned} \pi\left(\boldsymbol{\beta},\sigma^2,\delta,\boldsymbol{\theta},\mathbf{z}\right) &\propto \prod_{i=1}^{n}\prod_{t=1}^{T} f_{Pois}(y_{it}|\theta_i)\,\varphi\left(\theta_{it}|\mathbf{x}_{it}'\boldsymbol{\beta}+\delta z_{it},\sigma^2\right) \\ \varphi(z_{it}|0,1)\,I(z_{it}>0) &\times \varphi_k\left(\boldsymbol{\beta}|\boldsymbol{\beta}_0,\mathbf{V}_{\boldsymbol{\beta}}\right)\varphi\left(\delta|\delta_0,\sigma_\delta^2\right) f_{IG}\left(\sigma^2|\nu_0,\nu_0\right). \end{aligned} \tag{14}$$

The marginal posterior is derived by integrating out $\boldsymbol{\theta}$ and $\mathbf{z}$ from (14). This posterior is analytically intractable and the solution requires implementing advanced numerical integration techniques or utilizing the MCMC procedures, such as Gibbs sampling. The Gibbs sampling algorithm simulates iteratively from the complete conditional posterior distribution of each unknown stochastic parameter, or quantity, conditioned on the remaining parameters and unknown quantities. The complete conditional posterior distributions are given in the supplementary Appendix B.

4.2 Bayesian computation for the correlated data

Here, we use the following hierarchical representation of the defined models. The Gibbs sampling to fit model PMSN1 is implemented as follows.

The PMSN1 model. Consider the hierarchical form

$$\begin{aligned} Y_{it}|\theta_{it} &\overset{iid}{\sim} Pois\left(\exp(\theta_{it})\right), \\ \boldsymbol{\theta}_i|z_i,\boldsymbol{\beta},\mathbf{V},\boldsymbol{\delta} &\overset{ind}{\sim} N_T\left(\mathbf{X}_i\boldsymbol{\beta}+\boldsymbol{\delta}z_i,\mathbf{V}\right), \end{aligned} \tag{15}$$

for subject $i=1,2,\cdots,n$ and at time $t=1,2,\cdots,T$. Assuming the multivariate normal prior for $\boldsymbol{\beta}$, the inverse-Wishart $IW_T\left(\boldsymbol{\Omega},m\right)$ for matrix $\mathbf{V}$, and $N_T\left(\boldsymbol{\delta}_0,\mathbf{V}_{\boldsymbol{\delta}}\right)$ for the vector of skewness parameters $\boldsymbol{\delta}$, where all hyper-parameters are assumed to be known, we derive the related complete conditional posteriors as given in the supplementary Appendix B. The specification of models PMSN2 and PMSN3 are given below by using the multivariate skew-normal mixing prior.

The PMSN2 model. The hierarchical form of PMSN2 is

$$\begin{aligned} Y_{it}|\theta_{it} &\overset{ind}{\sim} Pois\left(\exp(\theta_{it})\right), \\ \boldsymbol{\theta}_i|\alpha_i,z_i,\boldsymbol{\beta},\mathbf{V}_\varepsilon,\boldsymbol{\delta} &\overset{ind}{\sim} N_T\left(\mathbf{X}_i\boldsymbol{\beta}+\alpha_i\mathbf{1}_T+\boldsymbol{\delta}z_i,\mathbf{V}_\varepsilon\right), \\ \alpha_i &\overset{iid}{\sim} N\left(0,\sigma_\alpha^2\right). \end{aligned} \tag{16}$$

The PMSN3 model. The hierarchical form of PMSN3 is

$$\begin{aligned} Y_{it}|\theta_{it} &\overset{ind}{\sim} Pois\left(\exp(\theta_{it})\right), \\ \boldsymbol{\theta}_i|\alpha_i,z_i,\boldsymbol{\beta},\mathbf{V}_\varepsilon,\delta &\overset{ind}{\sim} N_T\left(\mathbf{X}_i\boldsymbol{\beta}+\alpha_i\mathbf{1}_T,\mathbf{V}_\varepsilon\right), \\ \alpha_i|z_i &\overset{ind}{\sim} N\left(\delta z_i,\sigma_\alpha^2\right). \end{aligned} \tag{17}$$

The Bayesian computational details of mixed Poisson models PMSN2 and PMSN3, including priors and complete conditional posteriors, are given in supplementary Appendix B. All complete conditional posteriors, except for $\boldsymbol{\theta}$, appear in closed forms of known distributions and thus random samples can easily be generated. However, drawing samples from the posterior of $\boldsymbol{\theta}$ maybe done by the accept-reject algorithm (Gilks and Wild, 1992) or by the Metropolis-Hastings algorithm within the Gibbs sampler (Chib and Greenberg, 1995). Thus, the Gibbs sampler proceeds by simulating a sequence of samples from the complete conditional posteriors. The sampler simulates iteratively from these posteriors by running a sufficient burn-in period until convergence to stationary distributions occurs. Then, the average of samples for each parameter is used as its Bayes estimate. Convergence is monitored via MCMC chain histories, Gelman-Rubin diagnostic, autocorrelation, and density plots.

5 Comparative studies using simulation

We conduct two simulation studies to highlight the usefulness of proposed models. Specifically, we design Monte Carlo experiments to underline the important role of the skewness parameter and the structure of covariance matrices. We also make comparisons between competing models. To implement the Gibbs sampler, the following independent priors are adopted: $N(0,100)$ for the regression coefficients as well as for δ, Uniform$(-1,1)$ for ρ, and Inverse-Gamma$(0.01,0.01)$ for the variance components. Using the OpenBUGs software version 3.2.3, we run 10,000 samples after removing 5,000 burn-in until the convergence occurs. There was no evidence of lack of convergence based on examinations of histories, Gelman-Rubin diagnostic, kernel density, and autocorrelation plots. Also, by using various values of hyperparameters, we obtained similar results, which implies that posterior estimates are not sensitive to the prior in this Bayesian analysis.

a. The simulated model is PMSN1: We generated 1,000 independent Monte Carlo data sets from model PMSN1 with $n = 100$ sample size. Consider the longitudinal data model

$$Y_{it}|\theta_{it} \overset{ind}{\sim} Pois\left(\exp(\theta_{it})\right) \text{ with } \theta_{it} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{it2} + u_{it}, \tag{18}$$

for subject $i = 1,2,\cdots,100$ and at time $t = 1,2,\cdots,5$. Random counts are generated according to (9), where $\mathbf{u}_i \overset{iid}{\sim} SN_2\left(\mathbf{V},\boldsymbol{\delta}\right)$ with $\boldsymbol{\delta}' = \delta\mathbf{1}_5'$ and $\mathbf{V} = \left\{\sigma^2\rho^{|t-s|}\right\}$ for $t,s = 1,2,\cdots,5$. The time-constant covariate X_{i1} is generated by $Bernoulli\,(0.5)$ and the time-varying covariate X_{it2} by $N\,(0,1)$. For all experiments, θ_{it} was computed by setting $\beta_1 = -1$ and $\beta_2 = 1$. We set $\rho = 0.5$, $\sigma^2 = 0.36$, and $\delta = -0.8$, 0, 0.8. Taking into account the identification issue, we obtain $\beta_0 = -0.36, 0.18$ and 0.96, respectively. Results are reported in Tables 1 and 2 along with the fitted standard Poisson model for comparison. Biases and the mean squared error (MSE $\times 10$) of estimates are computed. Smaller values of the MSE indicate a better fit.

In each generation, the variance of Y differs considerably from the mean of Y, unlike the conventional Poisson density. Figure 3 illustrates this feature for $\delta = 0.8$ and the first 100 generations. This shows strong evidence of over-dispersion. Thus, fitting mixed Poisson models may be more appropriate to this data set.

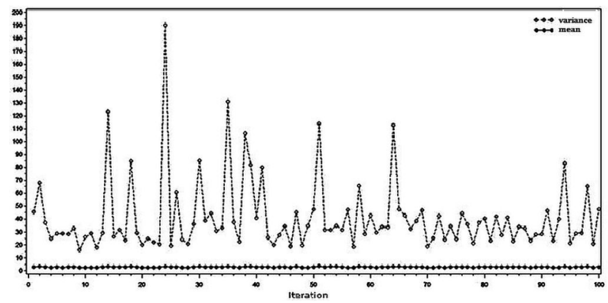

Figure 3: *The first 100 generations of model PMSN1 with* $\delta = 0.8$.

Therefore, we fit the hypothetical models PMN with $\mathbf{V} = \left\{\sigma^2\rho^{|t-s|}\right\}$ and PMSN2 with $\mathbf{D} = \sigma_\alpha^2 \mathbf{1}_5\mathbf{1}_5' + \sigma^2\mathbf{I}_5$, where $\boldsymbol{\delta} = \delta\mathbf{1}_5$.

Table 1: *Biases and MSEs(*$\times 10$*) for the proposed models.*

		Poisson		PMN		PMSN1		PMSN2	
δ		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
-0.8	β_0	-0.017	3.404	0.582	2.629	-0.021	0.005	0.038	0.016
	β_1	-0.117	2.138	-0.512	2.036	-0.013	0.004	-0.140	0.199
	β_2	-0.036	0.713	-0.155	0.160	-0.003	0.001	-0.031	0.010
	σ^2			0.085	0.031	-0.001	0.001	-0.035	0.013
	δ					-0.012	0.004	0.107	0.116
	ρ			0.156	0.244	-0.019	0.007		
0	β_0	0.018	0.904	0.002	0.001	<0.001	0.001	-0.076	0.059
	β_1	-0.012	0.202	-0.002	0.001	<0.001	0.001	-0.010	0.003
	β_2	-0.004	0.002	-0.001	0.001	<0.001	0.001	0.006	0.002
	σ^2			0.004	0.002	0.012	0.002	-0.095	0.090
	δ					-0.014	0.003	-0.042	0.019
	ρ			-0.050	0.029	-0.025	0.008		
0.8	β_0	-0.060	2.037	-0.478	2.284	-0.002	0.001	-0.174	0.304
	β_1	0.032	0.311	0.487	2.374	-0.007	0.002	0.123	0.153
	β_2	-0.018	0.210	0.107	0.115	-0.010	0.001	0.035	0.013
	σ^2			0.287	0.825	-0.040	0.017	0.168	0.284
	δ					-0.017	0.004	-0.233	0.545
	ρ			-0.080	0.064	-0.004	0.006		

Results, after the convergence is achieved, are reported in Table ***1***. Note that, we let $\boldsymbol{\delta} = \delta\mathbf{1}_5$ which implies equivalence of PMSN2 and PMSN3 models. Also, the concern was to illustrate the impact of ignoring dependency between the u_{it}'s for $t = 1, 2, \cdots, T$. Thus, PSN was not fitted. For $\delta = 0$, the PMSN1 performs as well as the PMN model. This finding shows that the PMSN1 is a flexible model since it can cover either symmetric or asymmetric data, depending on the values of its skewness parameter. For $\delta = -0.8, 0.8$ and based on MSEs, the PMSN models, PMSN1 and PMSN2, are better fitted than the conventional Poisson and PMN models. This finding does suggest the importance of identifying the correlation of the heterogeneity effects. To make a further comparison, we compute the relative efficiency $r = MSE_M/MSE_{PMSN1}$, where M denotes the competitive regression model. Efficiency values, shown in Table 2, are remarkably greater than 1, illustrating that the PMSN1 estimates are efficient compared to the parameter estimates in the hypothetical regression models.

Table 2: *Relative efficiencies of estimates in the longitudinal study.*

δ		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma^2}$	$\hat{\delta}$	$\hat{\rho}$
−0.8	Poisson	680.8	534.5	713.0			
	PMN	525.8	509.0	160.0	31.0		34.8
	PMSN2	3.2	49.7	10.0	13.0	29.0	
0	Poisson	904.0	202.0	2.0			
	PMN	1.0	1.0	1.0	2.0		3.6
	PMSN2	59.0	3.0	2.0	45.0	6.3	
0.8	Poisson	2037.0	155.5	210.0			
	PMN	2284.0	1187.0	115.0	48.5		10.667
	PMSN2	304.0	76.5	13.0	16.7	136.2	

b. The simulated model is PMSN2: Here, a simulation study is conducted to distinguish the performance of the PMSN1 model with $\mathbf{V} = \left\{\sigma^2\rho^{t-s}\right\}$, for $t,s = 1,2$, when response values are generated according to the PMSN2 model. We sampled data from (9) where θ_{it} is given in (18) and $\mathbf{u}_i \overset{iid}{\sim} SN_2(\mathbf{D},\boldsymbol{\delta})$ with $\boldsymbol{\delta}' = (\delta_1,\delta_2)$ and $\mathbf{D} = \sigma^2_\alpha \mathbf{1}_T\mathbf{1}'_T + \mathbf{V}_\varepsilon$. Also, the covariates X_{i1} and X_{it2} are generated respectively from a $Bernoulli(0.5)$ and a standard normal distribution. We set $\beta_1 = -1$, $\beta_2 = 1$, $\sigma^2_\alpha = 0.25$, $\delta = -4,1$, and the variance components $\sigma^2_{\varepsilon,1} = \sigma^2_{\varepsilon,2} = 0.25$ and $\sigma_{\varepsilon,12} = 0.75$ for $\mathbf{V}_\varepsilon$. Now, the correlation between observations is negative. Taking into account the identification issue, we obtain $\beta_0 = -0.65$ and 2, respectively. All parameters are estimated using the PMSN1 model. The posterior means (each with standard deviation) of estimates are obtained as $\tilde{\beta}_0 = -0.663(0.031), 2.063(0.033)$, $\tilde{\beta}_1 = -1.0090.030$, $\tilde{\beta}_2 = 1.008(0.016)$, $\tilde{\delta}_1 = -4.144(0.069)$, $\tilde{\delta}_2 = 0.987(0.026)$, $\tilde{\rho} = 0.507(0.012)$, and $\tilde{\sigma}^2 = 2.035(0.056)$. We observe that the bias of each regression coefficient and skewness parameter is small. Thus, the evidence again recommends that the regression model PMSN1 is appropriate to analyse the data.

6 Empirical studies

This section considers two examples taken from the literature that have been previously analysed by several authors. We fit the proposed models using the OpenBugs software. Priors for the regression coefficients and the skewness parameter were assumed to be independent, each distributed normally with zero mean and 0.001 precision, variance components distributed as inverse-Gamma distribution with parameters both equal to 0.1. The model fitting process has carried out for 10,000 iterations after discarding the first 5,000 iterations to ensure us the convergence has occurred. There was no evidence of lack of convergence due to examinations of histories, Gelman-Rubin diagnostic, kernel density, and autocorrelation plots. For illustration, the supplementary Appendix B shows posterior plots of the PMSN2 and β_5 for the health reform data.

A popular model selection in the Bayesian framework is the deviance information criterion (DIC). However, the DIC in OpenBugs is based on the conditional likelihood given the random effects. To compare the fitted models marginally we alternatively compute the Akaike information criterion, $AIC(\boldsymbol{\theta}) = D(\boldsymbol{\theta}) + 2p$, and the Bayesian information criterion, $BIC(\boldsymbol{\theta}) = D(\boldsymbol{\theta}) + p\log(n)$, where the deviance $D(\boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta})$, p and n denote the number of parameters and sample size, respectively. To estimate $D(\boldsymbol{\theta})$, we use the deviance evaluated at the Bayes estimates of parameters $\boldsymbol{\theta}$, where $L(\boldsymbol{\theta})$ is taken as the underlying marginal likelihood. Smaller values of these criteria indicate better fit.

6.1 Polio incidence

The Polio data set is taken from the US Centers for Disease Control and Prevention. The response variable is the monthly number of poliomyelitis cases (Y), over the years 1970 to 1983. The data were previously analysed by several researchers, such as Zeger (1988), Oh and Lim (2001), Davis and Wu (2009), Fokianos and Fried (2012) and Kang and Lee (2014) between others. We fit a similar model as given by Zeger (1988), and Oh and Lim (2001) by noting that the regression model is organized in terms of a re-centred version of time t, such that it can be easily convenient within the usual framework of the cross-sectional data model. The model includes an intercept, a time trend, and some trigonometric components at periods 6 and 12 months. Fitting a Poisson model, the ratio of deviance to degrees-of-freedom was 1.925, illustrating evidence of over-dispersion. Thus, the Poisson model is not suitable to fit the data. We now fit the PSN regression model, for $n = 1$, already specified in Section 2. Specifically, let $Y_t|u_t \overset{ind}{\sim} Pois(\exp(\theta_t))$ for $t = 1, \cdots, 168$, where

$$\begin{aligned}\theta_t &= \beta_0 + \beta_1 t^* \times 10^{-3} + \beta_2 \cos\left(\frac{2\pi t^*}{6}\right) + \beta_3 \sin\left(\frac{2\pi t^*}{6}\right)\\ &+ \beta_4 \cos\left(\frac{2\pi t^*}{12}\right) + \beta_5 \sin\left(\frac{2\pi t^*}{12}\right) + \beta_6 y_{t-1} + u_t,\end{aligned}$$

and $t^* = t - 73$ is used to locate the intercept term at January 1976 as in Zeger's analysis. We also analyse the polio incidence rates using the PN model; i.e. $u_t \overset{iid}{\sim} N\left(0, \sigma^2\right)$. Bayes estimates, standard deviations, 95% confidence intervals, and some information criteria for models comparison are given in Table 3.

Table 3: *Posterior summary statistics for parameters of fitted models.*

Model	Poisson	PN	PSN
	Est(s.d.)	Est(s.d.)	Est(s.d.)
	(95% CI)	(95% CI)	(95% CI)
β_0	0.046(0.088)	0.220(0.068)	0.031(0.084)
	(-0.131,0.215)	(0.108,0.374)	(-0.119,0.217)
β_1	-3.753(1.441)	-4.160(1.816)	-3.508(1.956)
	(-6.554,-0.9409)	(-7.754,-0.663)	(-7.378,0.299)
β_2	0.101(0.103)	0.130(0.129)	0.125(0.140)
	(-0.099,0.304)	(-0.119,0.386)	(-0.144,0.403)
β_3	-0.410(0.102)	-0.348(0.127)	-0.358(0.137)
	(-0.612,-0.211)	(-0.596,-0.099)	(-0.629,-0.090)
β_4	-0.181(0.098)	-0.125(0.125)	-0.140(0.134)
	(-0.376,0.010)	(-0.369,0.122)	(-0.398,0.126)
β_5	-0.464(0.0.111)	-0.443(0.136)	-0.457(0.147)
	(-0.686,-0.250)	(-0.712,-0.185)	(-0.750,-0.164)
β_6	0.092(0.025)	0.059(0.038)	0.104(0.041)
	(0.041,0.140)	(-0.019,0.131)	(0.018,0.181)
σ^2		0.440(0.137)	0.502(0.145)
		(0.217,0.750)	(0.260,0.829)
δ			-0.296(0.069)
			(-0.437,-0.149)
-2logL	531.5	511.9	499.2
AIC	545.5	527.9	517.2
BIC	567.3	552.8	545.3

Results show that the PSN is the best-fitted while the PN is the second one. The parameter δ differs significantly from 0 based on its confidence interval, and a negative direction of the difference exists. It again supports our claim that the PSN model is more appropriate for the polio data. The Bayesian results differ somewhat for the PSN and Poisson models. Standard deviations for the PSN model are larger, up to 14% and 51%, than those for the PN and Poisson models.

One objective in the analysis of polio data is to investigate whether or not the incidence of polio has been decreasing since 1970. This is indicated by the sign of the regression coefficient β_1. Under the PSN model, the negative sign of the trend term indicates that there is a long term decrease in the number of poliomyelitis cases during the observation period. This finding goes along with results achieved by Davis, Dunsmuir and Wang (2000) and Farrell, MacGibbon and Tomberlin (2007). We also note that the state dependence parameter β_6 is significant in the PSN model, which implies the contribution of the lagged response on prediction, while the PN model does not make such a conclusion.

6.2 Health reform data

The health-care reform data is taken from the German Socio-Economic Panel for the years 1995-1999. The main aim of the study was to investigate whether the number of physician visits by patients decreased after the reform. The data were analysed by Winkelmann (2004), who noted that the number of visits dropped by about 10% on average. Rabe-Hesketh and Skrondal (2012) fitted a PLN regression model on the impact of the 1997 health reform on the number of doctor visits. Then, several studies analysed the data for various purposes (e.g., Van Ophem, 2011). Our data consist of a subset taken from Rabe-Hesketh and Skrondal (2012) and are available in Stata and R software packages. We drop all missing values from the data, giving a subsample of 1,418 women who were employed full time the year before and after the reform.

The response variable is the utilization of health services, as measured by the self-reported number of patient visits to a physician's office three months before the interview. Covariates include an indicator variable for the interview being during the year after the reform versus the year before the reform, centred age in years, person education in years, an indicator for being married, a binary variable for self-reported current health, being classified as 'very poor' or 'poor' (versus 'very good'; 'good' or 'fair'), and the centred logarithm of household income.

The standard Poisson model makes the unrealistic assumption that the number of doctor visits before the reform is independent of the number of visits after the reform for the same person, given the included covariates. A fit of this model gives the ratio of deviance to the degrees-of-freedom equals 3.698, illustrating strong evidence of over-dispersion and suggests fitting alternative models. Thus, we propose fitting mixed Poisson regression models with the following specifications. The counts Y_{it}, conditioned on the effects u_{it} for subject $i = 1, 2, \cdots, 709$ and at time $t = 1, 2$, are taken to be independent $Pois(\exp(\theta_{it}))$ where

$$\theta_{it} = \beta_0 + \beta_1 \text{reform}_{it} + \beta_2 \,\text{age}_{it} + \beta_3 \,\text{educ}_{it} + \beta_4 \,\text{married}_{it} \\ + \beta_5 \,\text{badh}_{it} + \beta_6 \,\text{loginc}_{it} + u_{it}.$$

We fit PMSN1-PMSN3, PSN, PMN, and PN as competitive models and let $\mathbf{u}_i' = (u_{i1}, u_{i2})$, $\mathbf{V} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\mathbf{D} = \sigma_\alpha^2 \mathbf{1}_2 \mathbf{1}_2' + \sigma^2 \mathbf{I}_2$. The heterogeneity effects in PN and PSN models are assumed to be independent. These models are inappropriate. It should come as no surprise since no correlation is allowed for the heterogeneity effects, whereas in reality, it exists. The deviance of PN and PSN models are 5917.7 and 5870.8, respectively. Other findings are dropped here to save space. In addition, the estimate of correlation (11) for model PMSN2 was found to be 0.383 (s.d., 0.068; 95% confidence interval, 0.248–0.517) while for model PMSN3 it was 0.264 (s.d., 0.011; 95% confidence interval, 0.243–0.284). Combining these findings with (11) indicates strong evidence of the correlation between the number of patient visits to a physician's office before and after the reform. The posterior means and standard deviations for the conventional and proposed models are given in Table 4. In models PMSN1 and PMSN2, the intercept is replaced by β_{0t} for $t = 1,2$.

Table 4: *Posterior summary statistics for proposed PMSN models.*

Model	PMN	PMSN1	PMSN2	PMSN3
	Est(s.d.)	Est(s.d.)	Est(s.d.)	Est(s.d.)
	(95% CI)	(95% CI)	(95% CI)	(95% CI)
β_{01}	1.488(0.122)	0.634(0.062)	0.419(0.061)	0.807(0.062)
	(1.255,1.733)	(0.510,0.757)	(0.303,0.539)	(0.686,0.932)
β_{02}		-0.005(0.004)	1.175(0.103)	
		(-0.013,0.002)	(0.986,1.386)	
β_1	-0.076(0.050)	-0.708(0.076)	-0.715(0.076)	-0.249(0.048)
	(-0.174,0.021)	(-0.220,0.170)	(-0.865,-0.571)	(-0.348,-0.155)
β_2	-0.005(0.004)	-0.003(0.003)	-0.002(0.004)	-0.001(0.003)
	(-0.013,0.002)	(-0.009,0.006)	(-0.010,0.006)	(-0.007,0.007)
β_3	-0.004(0.017)	-0.006(0.016)	-0.006(0.017)	-0.011(0.017)
	(-0.040,0.029)	(-0.040,0.026)	(-0.038,0.028)	(-0.044,0.021)
β_4	0.318(0.063)	0.128(0.079)	0.108(0.073)	0.070(0.076)
	(0.192,0.443)	(-0.029,0.281)	(-0.033,0.252)	(-0.077,0.219)
β_5	1.127(0.101)	1.040(0.101)	1.024(0.097)	1.017(0.102)
	(0.926,1.324)	(0.841,1.236)	(0.834,1.217)	(0.818,1.217)
β_6	0.087(0.104)	0.114(0.097)	0.141(0.099)	0.116(0.097)
	(-0.111,0.293)	(-0.077,0.305)	(-0.046,0.342)	(-0.074,0.305)
δ_1		0.331(0.085)	0.375(0.075)	0.469(0.076)
		(0.166,0.496)	(0.221,0.515)	(0.325,0.632)
δ_2		1.007(0.087)	1.047(0.082)	
		(0.841,1.177)	(0.881,1.202)	
σ^2	0.979(0.066)	0.693(0.079)	0.183(0.057)	0.445(0.059)
	(0.854,1.122)	(0.545,0.845)	(0.086,0.301)	(0.341,0.570)
σ^2_α			0.465(0.062)	0.329(0.077)
			(0.347,0.591)	(0.172,0.479)
ρ	0.519(0.054)	0.667(0.086)		
	(0.406,0.618)	(0.664,0.824)		
-2logL	5645.9	5630.8	5627.4	5635.5
AIC	5660.9	5650.8	5647.4	5653.5
BIC	5697.4	5696.4	5693.0	5694.6

Table 4 also shows Bayes estimates of σ^2 and ρ with their 95% confidence intervals (CI) for models PMSN1 and PMN. That is, with 95% probability σ^2 lies between (0.854,1.122), for example. These facts reveal that much variability exists for the number of visits after the reform. Similarly, all skewness parameters differ significantly from 0 in a positive direction, showing that the distribution of heterogeneity effects is skewed right. It indicates that models PMSN1, PMSN2, and PMSN3 are more appropriate than the PMN model. Also, according to the deviance and the information criteria values reported in Table *4*, we find that the PMSN2 fits the data better than all competitive models.

Furthermore, in model PMSN2, age, education, married, and loginc are not statistically significant. However, the health care reform is negatively related to the number of visits, meaning that reform makes a decrease in the expected number of visits. Moreover, the badh coefficient is significant and positive, meaning that patients with having bad health make an increase in visits.

7 Concluding remarks

The analysis of correlated counts is challenging since suitable discrete multivariate distributions that can provide appropriate correlation structure are not always available. In longitudinal studies, the problem is addressed by letting the counts be independent Poisson variates conditioned on a vector of correlated heterogeneity effects. The correlation between the count variables is then incorporated in the resulting likelihood functions. In the paper, the correlation was taken into account by adopting that the random effects followed the multivariate skew-normal distribution with various structures for the skewness parameters. The modelling strategy allows for both positive and negative correlations among the subsequent counts. Empirical findings showed that the proposed modelling strategy had many potentials over conventional models. The paper used an accessible technique to compute the AIC and BIC values by plugging in Bayes estimates at the underlying marginal likelihoods. An interesting subject to future work is to use other Bayesian models comparison. Also, an extension of mixed modelling to the multivariate skew-normal random-effects is encouraged for non-Poisson correlated responses when over-dispersion occurs.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 88(422), 669-679.

Azzalini, A. (1985). A class of distributions which includes the normal ones. Scandinavian Journal of Statistics 12(2), 171-178.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. Biometrika 83(4),715-726.

Balakrishnan, N. and Peng, Y. (2006). Generalized gamma frailty model. Statistics in Medicine 25(16),2797-2816.

Bulmer, M.G. (1974). On fitting the Poisson log-normal distribution to species-abundance data. Biometrics 30(1),101-110.

Cameron, A.C. and Trivedi, P.K. (2013). Regression Analysis of Count Data. Cambridge University Press, Cambridge.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. The American Statistician 49(4),327-335.

Chib, S. and Winkelmann, R. (2001). Markov chain Monte Carlo analysis of correlated count data. Journal of Business and Economic Statistics 19(4),428-435.

Davis, R.A., Dunsmuir, W.T. and Wang, Y. (2000). On autocorrelation in a Poisson regression model. Biometrika 87(3),491-505.

Davis, R.A. and Wu, R.A. (2009). A negative binomial model for time series of counts. Biometrika 96(3),735-749.

El-Basyouny, K. and Sayed, T. (2009). Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis and Prevention 41(4),820-828.

Farrell, P.J., MacGibbon, P. and Tomberlin, T.J. (2007). A hierarchical Bayes approach to estimation and prediction for time series of counts. Brazilian Journal of Probability and Statistics 21(2),187-202.

Fokianos, K. and Fried, R. (2012). Interventions in log-linear Poisson autoregression. Statistical Modelling 12(4),299-322.

Gilks, W.R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. Journal of the Royal Statistical Society, Series C(Applied Statistics) 41(2),337-348.

Gonzales-Barron, U. and Butler, F. (2011). A comparison between the discrete Poisson-gamma and Poisson-lognormal distributions to characterise microbial counts in foods. Food Control 22(8),1279-1286.

Guo, J.Q. and Trivedi, P.K. (2002). Flexible parametric models for long-tailed patent count distributions. Oxford Bulletin of Economics and Statistics 64(1),63-82.

Holgate, P. (1970). The modality of some compound Poisson distributions. Biometrika 57(3),666-667.

Izsák, R. (2008). Maximum likelihood fitting of the Poisson lognormal distribution. Environmental and Ecological Statistics 15(2),143-156.

Kang, J. and Lee, S. (2014). Parameter change test for Poisson autoregressive models. Scandinavian Journal of Statistics 41(4),1136-1152.

Karlis, D. and Xekalaki, E. (2005). Mixed Poisson distributions. International Statistical Review 73(1),35-58.

Kuba, M. and Panholzer, A. (2016). On moment sequences and mixed Poisson distributions. Probability Surveys 13,89-155.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project, Evolution, critique and future directions. Statistics in Medicine 28(25),30-49.

Miller, G. (2007). Statistical Modeling of Poisson/Log-Normal Data. Radiation Protection Dosimetry Advance Access Published January 124(2),1-9.

Molenberghs, G., Verbeke, G. and Demétrio, C.G. (2007). An extended random-effects approach to modelling repeated, over-dispersed count data. Lifetime Data Analysis 13(4),513-531.

Montesinos-López, O.A., Montesinos-López, A., Crossa, J., Toledo, F.H., Montesinos-López, J.C., Singh, P., Juliana, p. and Salinas-Ruiz, J. (2017). A Bayesian Poisson-lognormal model for count data for multiple-trait multiple-environment genomic-enabled prediction. G3, Genes, Genomes, Genetics 7(5);1595-1606.

Nadarajah, S. and Kotz, S. (2006a). Compound mixed Poisson distributions I. Scandinavian Actuarial Journal 3,141-162.

Nadarajah, S. and Kotz, S. (2006b). Compound mixed Poisson distributions II. Scandinavian Actuarial Journal 3,163-181.

Oh, M.S. and Lim, Y.B. (2001). Bayesian analysis of time series Poisson data. Journal of Applied Statistics 28(2),259-271.

Rabe-Hesketh, S. and Skrondal, A. (2012). Multilevel and Longitudinal Modeling Using Stata, Volume II: Categorical Responses, Counts, and Survival, 3rd edition, Stata Press, College Station, Texas.

Rizzato, F.B., Leandro, R.A., Demétrio, C.G. and Molenberghs, G. (2016). A Bayesian approach to analyse overdispersed longitudinal count data. Journal of Applied Statistics 43(11),2085-2109.

Sahu, S.K., Dey, D.K. and Branco, M. (2003). A new class of multivariate distributions with applications to Bayesian regression models. The Canadian Journal of Statistics 31(2),129-150.

Van Ophem, H. (2011). The frequency of visiting a doctor, is the decision to go independent of the frequency?. Journal of Applied Econometrics 26(5),872-879.

Williams, M.S. and Ebel, E.D. (2012). Methods for fitting the Poisson-lognormal distribution to microbial testing data. Food Control 27(1),73-80.

Winkelmann, R. (2004). Health care reform and the number of doctor visits an econometric analysis. Journal of Applied Econometrics 19(4),455-472.

Wu, H., Deng, X. and Ramakrishnan, N. (2018). Sparse estimation of multivariate Poisson log-normal models from count data. Statistical Analysis and Data Mining, The ASA Data Science Journal 11(2),66-77.

Zeger, S.L. (1988). A regression model for time series of counts. Biometrika 75(4),621-629.