Communications for Statistical Applications and Methods
2023, Vol. 30, No. 3, 291–309

https://doi.org/10.29220/CSAM.2023.30.3.291
Print ISSN 2287-7843 / Online ISSN 2383-4757

Bayesian Conway-Maxwell-Poisson (CMP) regression for longitudinal count data

Morshed Alam [a], Yeongjin Gwon [1,a], Jane Meza [a]

[a]Department of Biostatistics, University of Nebraska Medical Center, USA

Abstract

Longitudinal count data has been widely collected in biomedical research, public health, and clinical trials. These repeated measurements over time on the same subjects need to account for an appropriate dependency. The Poisson regression model is the first choice to model the expected count of interest, however, this may not be an appropriate when data exhibit over-dispersion or under-dispersion. Recently, Conway-Maxwell-Poisson (CMP) distribution is popularly used as the distribution offers a flexibility to capture a wide range of dispersion in the data. In this article, we propose a Bayesian CMP regression model to accommodate over and under-dispersion in modeling longitudinal count data. Specifically, we develop a regression model with random intercept and slope to capture subject heterogeneity and estimate covariate effects to be different across subjects. We implement a Bayesian computation via Hamiltonian MCMC (HMCMC) algorithm for posterior sampling. We then compute Bayesian model assessment measures for model comparison. Simulation studies are conducted to assess the accuracy and effectiveness of our methodology. The usefulness of the proposed methodology is demonstrated by a well-known example of epilepsy data.

Keywords: CMP, DIC, LPML, LKJ prior, Hamiltonian MCMC, normalizing constant

1. Introduction

Longitudinal count data has been widely used in biomedical and public health research. Researchers are interested in studying or examing change in a response variable over time. In such case, observations are typically measured over time on the same subjects and these tend to be intercorrelated. Thus, this correlation structure should be properly taken into account in statistical models. Dispersion is an additional variability that needs to be taken into consideration. It is associated with the level of the spreadness for the data, and it leads to incorrect statistical inference if it is not properly addressed.

There are a large body of literature to model longitudinal count data. Poisson regression model is commonly used as Poisson distribution belongs to the exponential family, and modeling in the generalized linear model (GLM) framework under the Poisson distributional assumption is quite trivial (Albert, 1992; Fitzmaurice *et al*., 2008; Leppik *et al*., 1987; Wu *et al*., 2019). However, the Poisson distribution is not a suitable choice in the presence of over-dispersion or under-dispersion of the data as dispersed data easily violates equal mean and variance assumption. As an alternative, negative binomial (NB) regression model is considered appropriate for over-dispersion data, but not for under-dispersion data. These days, negative binomial regression model is applied to various research area including RNAseq data (Zhang *et al*., 2018), microbiome data (Tsonaka and Spittle, 2020), and

[1] Corresponding author: Department of Biostatistics, University of Nebraska Medical Center, 984375 Nebraska Medical Center, Omaha, NE, USA. E-mail: yeongjin.gwon@unmc.edu

Published 31 May 2023 / journal homepage: http://csam.or.kr

correlated longitudinal counts (Neelon, 2019). For under-dispersion data, generalized Poisson, restricted generalized Poisson were proposed by multiple authors (Consul and Jain, 2004; del Castillo and P´erez-Casany, 2005; Famoye, 1993; Famoye *et al*., 2004; Ridout and Besbeas, 2004). However, the downside of generalized Poisson models is an inability to capture some level of dispersion due to truncation of the dispersion parameter under certain conditions discussed in Famoye (1993). Recently, Conway-Maxwell-Poisson (CMP) regression model has been widely adopted in modeling count data. The distribution of CMP proposed by Conway and Maxwell (1962) is an extension of Poisson distribution, but offers a flexibility with incorporating a wide spectrum of dispersion in the data. In a theoretical point of view, the CMP distribution belongs to an exponential family and the sufficient statistics along with other properties under the exponential family are elegantly derived. Sellers and Morris (2017) and Morris *et al*. (2017) applied the CMP regression to under-dispersed and over-dispersed data, respectively. In particular, Morris *et al*. (2017) considered the random intercept model to examine performance of three different models: Poisson, negative binomial, and CMP models and showed a better fit of the CMP regression over the rest of models using PROC NLMIXED procedure. In Bayesian context, random intercept model was introduced by Guikema and Goffelt (2008) to assess electric power system reliability using PROC MCMC procedure. However, both approaches are restricted to only random intercept model.

In this article, we propose the CMP regression model to overcome the drawbacks mentioned earlier in Bayesian paradigm. The proposed model is attractive in multiple aspects. First, the proposed model accommodates subject heterogeneity and the variability in covariate effects. Second, it is computationally feasible. The model is readily implemented via standard statistical software STAN. The package 'rstan' in R software allows users to run STAN code from R for Bayesian computation. The posterior sampling is conducted via Hamiltonian Markov chain Monte Carlo (HMCMC) algorithm with No-U-Turn (NUT) sampler (Hoffman and Gelman, 2014). Third, a LKJ prior is used for sampling correlation through Cholesky decomposition. This method assures the positive definiteness of the generated correlation matrix. We also conduct a Bayesian model comparison using deviance information criteria (DIC) and logarithm of pseudo marginal likelihood (LPML) among candidate models.

The remainder of the article is organized as follows. We introduce the CMP distribution and the proposed CMP regression model in Section 2. In Section 3, we conduct a Bayesian inference and a feature of LKJ prior is also discussed. In Section 4, we perform a simulation study to evaluate an empirical performance of the proposed model. The proposed model applied to the analysis of epilepsy data in Section 5, and we close with a brief discussion in Section 6.

2. Statistical method

2.1. Conway-Maxwell-Poisson (CMP) distribution

Let y_i be the count response for subject i for $i = 1, \ldots, N$. Then the probability mass function of the CMP distribution with positive shape parameter θ_i and dispersion parameter ϕ is written as

$$\Pr(Y_i = y_i) = \frac{\theta_i^{y_i}}{(y_i!)^{\phi} Z(\theta_i, \phi)}, \tag{2.1}$$

where $Z(\theta_i, \phi) = \sum_{k=0}^{\infty} (\theta_i^k/(k!)^{\phi})$ is the normalizing constant for $i = 1, \ldots, N$ and not being a closed form expression. The dispersion parameter ϕ in (2.1) controls different types of the spreadness in the data: Equi-dispersion ($\phi = 1$), over-dispersion ($0 \leq \phi < 1$), and under-dispersion ($\phi > 1$), respectively.

The CMP is a large class of the discrete distribution including Poisson, geometric, and Bernoulli distributions as special cases. For example, it reduced to Poisson if $\phi = 1$, geometric if $\phi = 0$, and

Bernoulli if $\phi \to \infty$ and $p_i = \theta_i/(1+\theta_i)$, where p_i is the probability of success. One feature of this distribution is that there is a functional relationship between the dispersion parameter ϕ and shape parameter θ_i. Sellers *et al*. (2012) show that the θ_i is expressed as the form of the moment of Y, that is, $\theta_i = \mathbb{E}(Y^{\phi})$, while mean and variance are approximately calculated as $\mathbb{E}(Y) \approx \theta_i^{1/\phi} - (\phi-1)/2\phi$ and $V(Y) \approx (\theta_i^{1/\phi})/\phi$, respectively. However, the approximation is especially accurate for $\phi \leq 1$ or $\theta_i^{1/\phi} > 10$ as discussed in Shmueli *et al*. (2005). As ϕ is close to 1, θ_i approximates the mean, while θ_i deviates substantially from the mean in the cases where over-dispersion ($\phi < 1$) or under-dispersion ($\phi > 1$).

2.2. CMP regression model

Suppose we have a total of N subjects that are longitudinally measured at time j for $j = 1, \ldots, n_i$. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^{\mathrm{T}}$ be the count response vector of subject i for $i = 1, \ldots, N$. We also let $\boldsymbol{x}_{ij}^{\mathrm{T}} = (x_{ij1}, \ldots, x_{ijp})$ is a p-dimensional covariate vector for subject i including an intercept, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ be the corresponding vector of regression coefficients. We further let $\boldsymbol{z}_{ij}^{\mathrm{T}} = (z_{ij1}, \ldots, z_{ijq})$ is a q-dimensional covariate vector for the random effects, $\boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{iq})^{\mathrm{T}}$ is its corresponding vector of random effects for subject i. Then, the CMP regression model is written as

$$\log\left(\theta_{ij}\right) = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i, \tag{2.2}$$

where $\boldsymbol{\zeta}_i \sim N_q(\mathbf{0}, \Sigma)$ with Σ is a $q \times q$ covariance matrix. The diagonal elements $\sigma_1^2, \ldots, \sigma_q^2$ of Σ are the variance of the random effects, while off-diagonal elements $\rho_{kl}\sigma_k\sigma_\ell$ represent covariance between two $(k,\ell)^{th}$ random effects for $k \neq \ell$, $k = 1, \ldots, q$ and $\ell = 1, \ldots, q$. The between subject variability in the population is captured by the variance of random intercept σ_1^2, while the variability of covariate effect is quantified by σ_k^2 for $k = 2, \ldots, q$. Note that the assumption of correlated random effects is common in longitudinal studies, for example, study on Hamilton depression rating scale (HDRS) discussed in Hedeker and Gibbons (2006).

2.3. The likelihood function

Let $\mathcal{D} = \{D_1, \ldots, D_N\}$ be the data for the entire subjects, where $D_i = \{\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i\}$ is the observed data for subject i. Note that $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}^{\mathrm{T}}, \ldots, \boldsymbol{x}_{in_i}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}^{\mathrm{T}}, \ldots, \boldsymbol{z}_{in_i}^{\mathrm{T}})^{\mathrm{T}}$ are covariate matrices for the subject i. Under the proposed model in (2.2), the observed likelihood function for subject i is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \phi \mid D_i, \boldsymbol{\zeta}_i, \Sigma) &= \prod_{j=1}^{n_i} f\left(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{\beta}, \phi, \boldsymbol{\zeta}_i\right) \\ &\propto \prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi}\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^{y_{ij}}\left(\sum_{k=0}^{\infty}\frac{\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^{k}}{(k!)^{\phi}}\right)^{-1}.\end{aligned} \tag{2.3}$$

Then, complete-data likelihood for the entire data $\mathcal{D}$ is written as

$$\mathcal{L}(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}, \Sigma \mid \mathcal{D}) = \prod_{i=1}^{N}\prod_{j=1}^{n_i} f\left(y_{ij} \mid \boldsymbol{x}_{ij}, \boldsymbol{\beta}, \phi, \boldsymbol{\zeta}_i\right) f\left(\boldsymbol{\zeta}_i \mid \Sigma\right) \tag{2.4}$$

$$\propto \prod_{i=1}^{N}\left[\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi}\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^{y_{ij}}\left(\sum_{k=0}^{\infty}\frac{\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^{k}}{(k!)^{\phi}}\right)^{-1}|\Sigma|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\zeta}_i^{T}\Sigma^{-1}\boldsymbol{\zeta}_i\right)\right].$$

As shown in (2.1) and (2.4), there is an infinite sum of normalizing constant in the CMP distribution. This does not have a closed form expression and results in computational burden in parameter estimation both frequentist and Bayesian method.

Remark 1. For computational convenience, it follows from Morris *et al*. (2017) and Choo-Wosoba *et al*. (2018) that we consider a truncation of the summation up to a certain order. However, we use a criteria below to determine the order to be summed up, rather than using a pre-determined fixed order.

$$\left|\sum_{k=0}^{K}\frac{\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^{k}}{(k!)^{\phi}}-\sum_{k'=0}^{K-1}\frac{\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^{k'}}{(k'!)^{\phi}}\right|<\epsilon. \tag{2.5}$$

In our analysis, we choose a cut-off $\epsilon = 0.001$ in Bayesian computation.

3. Bayesian inference

3.1. Prior and posterior

We assume that $\boldsymbol{\beta}, \phi$, and Σ are independent apriori. Thus, the joint prior distributions are of the form $\pi(\boldsymbol{\beta}, \phi, \Sigma) = \pi(\boldsymbol{\beta})\pi(\phi)\pi(\Sigma)$. To sample parameters from covariance matrix Σ, we first consider the separation strategy proposed by Barnard *et al*. (2000). The method allows users to decompose the covariance matrix Σ as $\Sigma = \mathrm{V\ R\ V^{T}}$, where $V = \mathrm{diag}(\sigma_1, \ldots, \sigma_q)$ and R is a $q \times q$ correlation matrix.

Thus, the joint prior distribution is

$$\pi\left(\boldsymbol{\beta}, \phi, \sigma_1, \sigma_2, \ldots, \sigma_q, R\right) = \pi\left(\boldsymbol{\beta}\right)\pi\left(\phi\right)\pi\left(\sigma_1\right)\pi\left(\sigma_2\right)\cdots\pi\left(\sigma_q\right)\pi\left(R\right).$$

We assign prior distributions for $\boldsymbol{\beta} \sim N(\mathbf{0}, \Omega)$, $\phi \sim \mathrm{LN}(a, b)$, $\sigma_\ell \sim U(0, c)$ for $\ell = 1, \ldots, q$ and $R \sim \mathrm{LKJ}(\eta)$, where Ω, a, b, c, and η are pre-determined hyperparameters. We may simply consider a uniform prior on $\rho_{k\ell}$, i.e., $\pi(\rho_{k\ell}) \propto 1$, however, this is usually not efficient in posterior sampling. Hence, we consider LKJ prior on the correlation matrix itself, which we discuss the next section. Then, the resulting joint posterior distribution is given by

$$\begin{aligned}\pi\left(\boldsymbol{\beta}, \boldsymbol{\zeta}, \phi, \sigma_1, \ldots, \sigma_q, R \mid \mathcal{D}\right) &= \mathcal{L}\left(\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}, \Sigma \mid \mathcal{D}\right)\pi\left(\boldsymbol{\beta}, \phi, \sigma_1, \ldots, \sigma_q, R\right) \\ &\propto \prod_{i=1}^{N}\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi}\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^{y_{ij}}\left(\sum_{k=0}^{\infty}\frac{\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^{k}}{(k!)^{\phi}}\right)^{-1} \\ &\quad\times|\mathrm{VRV}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\zeta}_i^{\mathrm{T}}\left(\mathrm{VRV}\right)^{-1}\boldsymbol{\zeta}_i\right) \\ &\quad\times\exp\left(-\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\Omega^{-1}\boldsymbol{\beta}\right)\frac{1}{ab}\exp\left[-\frac{1}{2}\left(\frac{\log\phi-a}{b}\right)^{2}\right]\pi\left(R\right).\end{aligned} \tag{3.1}$$

We note that there are different approaches to sample Σ and ϕ in (3.1). Since ϕ is non-negative, simply we consider lognormal, uniform, or hierarchical half-t prior by (Huang and Wand, 2013) for

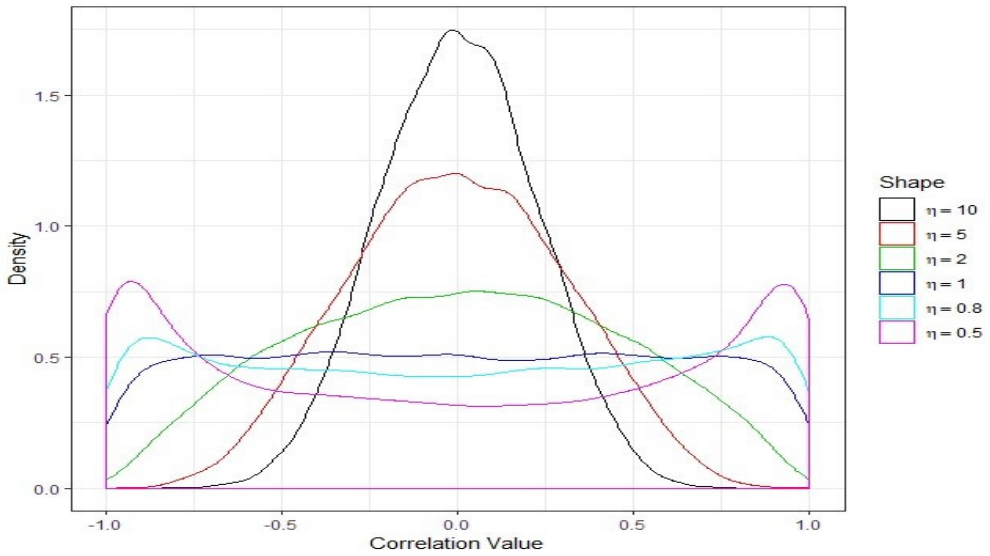

Figure 1: *LKJ density plot for* 2×2 *correlation matrix* R.

$\pi(\phi)$. Inverse-Wishart (IW) distribution, scaled inverse-Wishart (O'Malley and Zaslavsky, 2008), restricted Wishart distribution (Wang *et al*., 2018) can be used for $\pi(\Sigma)$. Recently, some of limitations in using IW distribution were pointed out by Alvarez *et al*. (2014). These are including (a) a single degree of freedom parameter ν in IW distribution, $\text{IW}(\nu, \Gamma)$ controls uncertainty in all variance parameters and this results in less flexibility to incorporate various amount of prior knowledge from different variance components (Gelman *et al*., 2013); (b) if $\nu > 1$, the marginal density of variance parameter is close to the zero region and leads to biased in the posterior estimate of variance (Gelman, 2006); and (c) the larger variances are more likely to be associated with the boundary of correlations (nearly to $+1$ or -1) while smaller variances are associated with correlations near to zero (Tokuda *et al*., 2011).

3.2. LKJ prior

Sampling correlation ρ is one of the major task as it has restricted parameter space $(-1, 1)$. Fisher's z-transformation or de-constraint transformation can be used, however, we use different approach proposed by Lewandowski *et al*. (2009). They use LKJ prior to sample ρ from a space of positive definite correlation matrix. Assume that the correlation matrix $R \sim \text{LKJ}(\eta)$. Then the density of R is defined as

$$\pi(R) \propto \det(R)^{\eta-1},$$

where $\det(R)$ is the determinant of the matrix R. To better understand this density, we generate a plot for the density of R under different η values. As shown in Figure 1, the distribution is symmetric about 0 and has different shapes depending on the value of η. The value of ρ is almost uniformly distributed over the parameter space if $\eta = 1$. As η increases, ρ is more concentrated at the value of zero, while the boundary values ρ have higher density as η is close to zero.

Rather than directly sampling correlation matrix R, an implicit parameterization based on Cholesky decomposition of R is introduced by Carpenter *et al*. (2017) to maintain positive definiteness of the correlation matrix. For $\eta > 0$, the Cholesky decomposition of R is given by $R = LL^{\intercal}$ where L is a

Table 1: Simulation study: Equi-Dispersed data ($\phi = 1.0$)

Model	Param	True	Bias	MSE	LPML	DIC
Poisson	β_1	1.5	−0.029	0.007	−503.81	1007.55
	β_2	−0.05	−0.009	0.005		
	β_3	−0.3	0.018	0.008		
	σ_1	0.02	0.064	0.005		
	σ_2	0.03	0.068	0.006		
	ρ	−0.5	0.340	0.120		
NB	β_1	1.5	−0.028	0.007	−503.92	1008.63
	β_2	−0.05	−0.009	0.005		
	β_3	−0.3	0.018	0.008		
	σ_1	0.02	0.060	0.004		
	σ_2	0.03	0.064	0.005		
	ρ	−0.5	0.350	0.123		
CMP	β_1	1.5	0.118	0.057	−504.68	1008.27
	β_2	−0.05	−0.013	0.006		
	β_3	−0.3	−0.005	0.010		
	σ_1	0.02	0.080	0.008		
	σ_2	0.03	0.088	0.010		
	ρ	−0.5	0.324	0.109		

$q \times q$ lower triangular matrix with positive diagonal entries $l_{kk} > 0$ for $k = 1, \ldots, q$. Then we have $\pi(R) \propto |J| \det(LL^T)^{\eta-1}$, where $|J|$ is the Jacobian for the transformation from R to L and η has the same interpretation as above. Note that this is a standard sampling algorithm in STAN.

3.3. Hamiltonian MCMC (HMCMC) and posterior sampling

HMCMC is a MCMC method that generate efficient transitions covering the posterior distribution by using derivatives of the density function being sampled. It begins with a specific set of initial parameters for $\boldsymbol{\beta}, \phi, \sigma_1, \ldots, \sigma_q, \boldsymbol{\rho}$, and $\boldsymbol{\zeta}$ where $\boldsymbol{\rho} = (\rho_{k\ell})$. These values are either pre-determined or to be generated randomly in STAN. Then, a new vector of momentum is drawn and the current value of parameters is being updated using a numerical integration technique based on the Hamiltonian dynamics.The details on HMCMC can be found in Betancourt *et al*. (2017) and Neal (2011).

We provide the posterior sampling scheme for the proposed model. Although the analytical form of the posterior distribution of $\pi(\boldsymbol{\beta}, \boldsymbol{\zeta}, \phi, \sigma_1, \ldots, \sigma_q, R|\mathcal{D})$ in (3.1) is not available, the proposed model enables us to sample efficiently via Hamiltonian MCMC sampling algorithm. The MCMC sampling is achieved via the following full conditional distributions: (i) $[\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}|\sigma_1, \ldots, \sigma_q, \boldsymbol{\rho}, \mathcal{D}]$ and (ii) $[\sigma_1, \ldots, \sigma_q, \boldsymbol{\rho}|\boldsymbol{\beta}, \phi, \boldsymbol{\zeta}, \mathcal{D}]$.

For (i), the samplings are achieved through (ia) $[\boldsymbol{\beta}|\phi, \boldsymbol{\zeta}, \sigma_1, \ldots, \sigma_q, \boldsymbol{\rho}, \mathcal{D}]$; (ib) $[\phi|\boldsymbol{\beta}, \boldsymbol{\zeta}, \sigma_1, \ldots, \sigma_q, \boldsymbol{\rho}, \mathcal{D}]$; and (ic) $[\boldsymbol{\zeta}|\phi, \boldsymbol{\beta}, \sigma_1, \ldots, \sigma_q, \boldsymbol{\rho}, \mathcal{D}]$. Sampling for (ii) are through conditional distributions:

(iia) $[\sigma_k|\sigma_{-k}, \boldsymbol{\rho}, \boldsymbol{\beta}, \phi, \boldsymbol{\zeta}, \mathcal{D}]$, where $\sigma_{-k} = (\sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_q)$ and (iib) $[\boldsymbol{\rho}|\sigma_1, \ldots, \sigma_q, \boldsymbol{\beta}, \phi, \boldsymbol{\zeta}, \mathcal{D}$]. The form of full conditional distributions are given in **Appendix A**.

3.4. Bayesian model assessment

We conduct Bayesian model comparison using two assessment measures: Deviance information criteria (DIC) (Spiegelhalter *et al*., 2002) and logarithm of pseudo-marginal likelihood (LPML) (Ibarhim *et al*., 2001). We note that the normalizing constants are required to compute these assessment measures when comparing different models. Let $\Psi = (\boldsymbol{\beta}, \phi, \sigma_1, \ldots, \sigma_q, \boldsymbol{\rho}, \boldsymbol{\zeta})$ be the collection of all parameters.

Table 2: Simulation study: Over-Dispersed data ($\phi = 0.6$)

Model	Param	True	Bias	MSE	LPML	DIC
Poisson	β_1	1.5	0.998	0.999	−690.09	1362.39
	β_2	−0.05	−0.039	0.005		
	β_3	−0.3	−0.163	0.031		
	σ_1	0.02	0.139	0.024		
	σ_2	0.03	0.162	0.032		
	ρ	−0.5	0.042	0.055		
NB	β_1	1.5	1.011	1.026	−675.84	1350.67
	β_2	−0.05	−0.038	0.005		
	β_3	−0.3	−0.170	0.033		
	σ_1	0.02	0.047	0.003		
	σ_2	0.03	0.056	0.004		
	ρ	−0.5	0.353	0.128		
CMP	β_1	1.5	0.107	0.041	−674.98	1348.87
	β_2	−0.05	−0.009	0.002		
	β_3	−0.3	−0.014	0.003		
	σ_1	0.02	0.029	0.001		
	σ_2	0.03	0.030	0.001		
	ρ	−0.5	0.329	0.113		

Then, the deviance function is defined as

$$\text{Dev}\left(\Psi\right) = -2\sum_{i=1}^{N}\sum_{j=1}^{n_i}\left[-\phi\log\left(y_{ij}!\right) + y_{ij}\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right) - \log\left(\sum_{k=0}^{\infty}\frac{\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right)\right)^k}{(k!)^{\phi}}\right)\right].$$

Then, the overall DIC is given by

$$\text{DIC} = \text{Dev}\left(\bar{\Psi}\right) + 2pD, \tag{3.2}$$

where $pD = \overline{\text{Dev}(\Psi)} - \text{Dev}(\bar{\Psi})$, $\overline{\text{Dev}(\Psi)} = \mathbb{E}[\text{Dev}(\Psi)|\mathcal{D}]$, and $\bar{\Psi} = \mathbb{E}(\Psi|\mathcal{D})$. In (3.2), the first term measures the goodness-of-fit, while pD is the effective number of model parameters. The smaller value of DIC indicates a better model fit.

The conditional predictive ordinate (CPO) by Geisser and Eddy (1979) measures the contribution of the subject i to the fitted model, which is defined as

$$\text{CPO}_i = \int f\left(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{\theta}\right)\pi\left(\boldsymbol{\theta} \mid \mathcal{D}^{(-i)}\right)d\boldsymbol{\theta}, \tag{3.3}$$

where $\mathcal{D}^{(-i)}$ is the observed data deleting the subject i. We use the approach proposed by Zhang *et al.* (2017) to estimate the CPO_i, which is given by

$$\widehat{\text{CPO}}_i^{-1} = \frac{1}{B}\sum_{b=1}^{B}\frac{w_i\left(\boldsymbol{\zeta}_i^{(b)}\right)}{f\left(\boldsymbol{y}_i, \boldsymbol{\zeta}_i^{(b)} \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}^{(b)}\right)},$$

where $\{(\boldsymbol{\zeta}_i^{(b)}, \boldsymbol{\theta}^{(b)}), b = 1, \ldots, B\}$ is a Gibbs sample and $w_i(\boldsymbol{\zeta}_i)$ is a weight function with $\int w_i(\boldsymbol{\zeta}_i)d\boldsymbol{\zeta}_i = 1$ (See the calculation in **Appendix B**). Then the LPML is readily constructed using the estimated CPO as

$$\text{LPML} = \sum_{i=1}^{N}\log\left(\text{CPO}_i\right). \tag{3.4}$$

Unlike the DIC, a model with a larger LPML fits the data better.

4. Simulation study

We conduct a simulation study to evaluate the performance of our methodology. The simulation design was based on the analysis of epilepsy data in Section 5. We generate 100 simulated datasets, each with 250 observations in total from $n = 50$ subjects with 5 measurements at 5 different time points ($t = 0, 1, 2, 3, 4$) where 0 indicates a baseline. The time variable is redefined as a binary variable (0 if baseline and 1 if post-baseline). The treatment assignment is from $x_i \sim$ Bernoulli(0.5) for placebo vs. treatment group. We assume that the random effects $\zeta_{1i} \sim N(0, \sigma_1)$ and $\zeta_{2i} \sim N(0, \sigma_2)$ where $\sigma_1 = 0.02$ and $\sigma_2 = 0.03$ with $\rho = -0.5$. True values for regression coefficients $\beta_1 = 1.5, \beta_2 = -0.05$, and $\beta_3 = -0.3$. This represents the beneficial effects of treatment and time, respectively. The true dispersion parameters are $\phi = (0.6, 1.0)$. Thus, the outcome is over-dispersion and equi-dispersion at each time point. Note that we do not consider under-dispersed scenario in our simulation study as our real data exhibits over-dispersed.

The simulated datasets are based on the model framework: $y_{ij} \sim \text{CMP}(\theta_{ij}, \phi)$ where $\log(\theta_{ij}) = (\beta_1 + \zeta_{1i}) + \beta_2 x_{ij} + (\beta_3 + \zeta_{2i}) t_{ij}$. We run four independent Markov chain for the three models. Each chain has 1,000 burn-in and the following 3,500 samples are collected for simulation study. This results in a total of 10,000 posterior samples for each model. To evaluate performance, we consider three quantities: (i) Bias: Total bias $= \sum_{m=1}^{100} (\hat{\theta}_m - \theta_0)/100$ where θ_0 is the true parameter value for θ and $\hat{\theta}_m$ is the posterior estimate of θ from the m^{th} data; (ii) mean square error (MSE): MSE $= \sum_{m=1}^{100} (\hat{\theta}_m - \theta_0)^2/100$; (iii) $\text{DIC}_M = \text{Median}\{\text{DIC}_1, \ldots, \text{DIC}_{100}\}$ and $\text{LPML}_M = \text{Median}\{\text{LPML}_1, \ldots, \text{LPML}_{100}\}$, where DIC_m and LPML_m are DIC and LPML from the m^{th} data.

Table 1 presents simulation results for equi-dispersion data. We see that Poisson model has the least DIC (1007.55) and the greatest LPML (−503.81) values among three models, indicating the best fit in terms of goodness-of-fit. However, the CMP model has both smaller DIC and LPML compared to the negative binomial model, meaning that the CMP model does not outperform the NB model when LPML is considered for model comparison. This inconsistent finding may be because the dispersion parameter ϕ in the CMP model is not being generated $\phi = 1$ all the time in posterior sampling. This results in having other posterior samples from three different types of data: Under-Dispersed, equi-dispersed, and over-dispersed data. Interestingly, the estimation of fixed effects and random effects show similar performance in all three models in terms of the bias and MSE evaluation. We see slightly inflated MSE of the intercept under the CMP model, but the difference is 0.057 after rounding up to the three decimal points. These simulation results illustrate that the Poisson model has the advantage in fitting equi-dispersed count data and this is supported by both DIC and LPML.

In Table 2, we present simulation results for over-dispersed data. We obviously see that both Poisson and NB regression model have particularly high bias and MSE on the intercept when the data is over-dispersed. Surprisingly, the Poisson model has much smaller bias and MSE for the correlation coefficient ρ. Moreover, the Poisson model has relatively higher bias 0.139 and 0.162 for the variability compared to NB and CMP models. The CMP model shows a better overall performance than other two models with the smallest DIC (1348.87) and largest LPML (−674.98) quantities. This simulation study demonstrates that the CMP model is a better choice in over-dispersed data.

5. Analysis of epilepsy data

5.1. Data description

Epileptic seizure data obtained from a multi-center, double-blind, placebo-controlled clinical trial (Leppik *et al*., 1987). The analysis for this study was originally analyzed by Thall and Vail (1990),

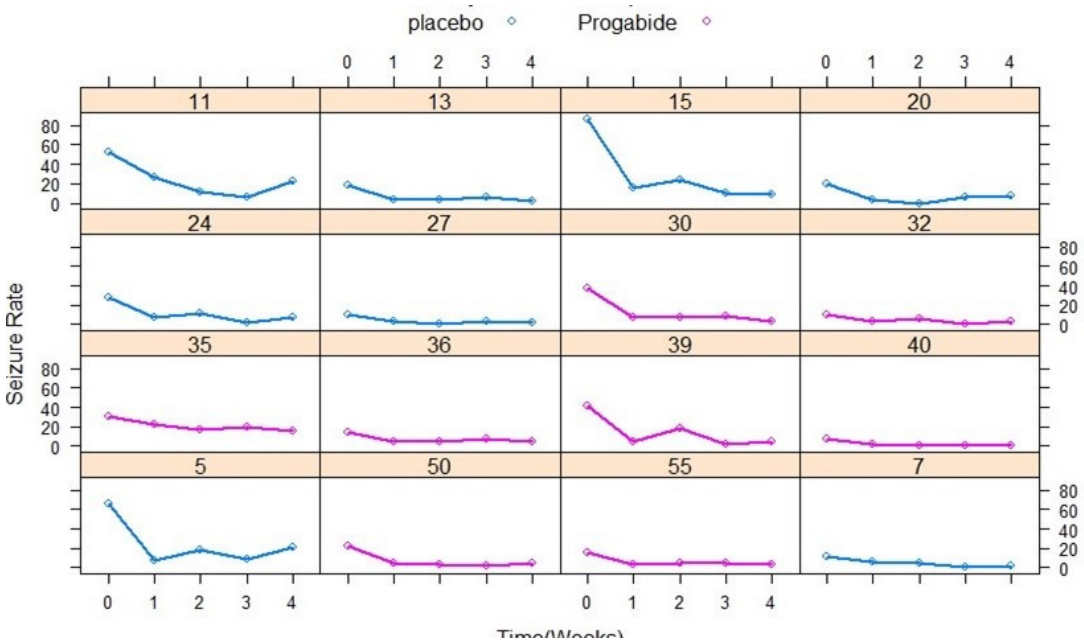

Figure 2: *Trajectory of the seizure rate over study period.*

Table 3: Posterior summary and model comparison under the three models

Parameter	Poisson			NB			CMP		
	Mean	SD	95% CR	Mean	SD	95% CR	Mean	SD	95% CR
β_1	1.07	0.15	(0.78,1.36)	1.11	0.15	(0.81,1.41)	−0.77	0.18	(−1.11,−0.39)
β_2	0.05	0.20	(−0.35,0.44)	0.06	0.21	(−0.36,0.47)	0.01	0.09	(−0.15,0.19)
β_3	0.00	0.12	(−0.22,0.23)	−0.03	0.12	(−0.27,0.20)	0.73	0.09	(0.54,0.89)
β_4	−0.31	0.16	(−0.62,0.01)	−0.32	0.17	(−0.64,0.01)	−0.13	0.08	(−0.30,0.03)
σ_1	0.75	0.08	(0.61,0.92)	0.68	0.09	(0.52,0.87)	0.31	0.05	(0.21,0.42)
σ_2	0.52	0.07	(0.39,0.67)	0.36	0.09	(0.18,0.55)	0.22	0.04	(0.15,0.32)
ρ	0.14	0.16	(−0.17,0.43)	0.56	0.22	(0.09,0.92)	0.44	0.18	(0.08,0.78)
ϕ		-	-	0.13	0.03	(0.09,0.19)	0.42	0.06	(0.32,0.53)
DIC	1678.49			1674.59			1538.17		
LPML	−889.60			−842.10			−815.14		

and the data was revised by Morris *et al*. (2017) with PROC NLMIXED procedure using statistical analysis system (SAS). However, both authors only consider random intercept model to capture subject variability. With regard to this, we revisit the data to implement Bayesian analysis based on our proposed model that is an extension of previous approach. Epilepsy is a chronic neurological disorder that may cause from genetic abnormality, head trauma, stroke, brain tumor, infection, developmental malfunction, etc. Most of the cases causes are unknown. According to international league against epilepsy (ILAE), a patient with two or more unprovoked seizures occurring in 24 hours apart is termed as epileptic.

The datasets consists of the number of seizures for 59 patients suffering from epilepsy, 31 of them are assigned to the treatment group, as an adjuvant to the standard anti-epileptic chemotherapy (phenytoin and carbamazepine), and the rest in the placebo group. The primary objective of the analysis is to compare the changes in the average rates of seizures in two treatment arms. Seizure

rates are longitudinally measured in an initial eight weeks before baseline and then in every two weeks in four consecutive treatment periods.

A partial view of the selected individual profiles in Figure 2 depicts that both baseline and post seizure rates differ by subjects over time across treatment arms. Using the observed data, we compute the mean and variance. The variance (349.17) is way higher than the mean (12.85), clearly indicating that there is an over-dispersion in the data. Therefore, the CMP mixed effects model would be a better choice to accommodate the over-dispersion as well as between and within subjects variability. In the data, time is defined as an indicator variable of a period after baseline. It is 0 if baseline and 1 if after baseline. Treatment (trt) is also used as a binary value, that is, 1 if a patient receives an anti-epileptic drug (progabide) and 0 if placebo. The length of time period measured in weeks and its log scale is used as the offset. By looking at the profiles for the selected individuals in Figure 2, it is reasonable to assume that there is a heterogeneity among subjects both in their baseline level and in the changes of counts over time. Hence, we use the CMP regression model with random effects (random intercept and intercept) for our analysis. This leads to

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \text{VRV}.$$

5.2. Results

We consider three different models based on the distributions of the response count vectors: (i) Poisson regression; (ii) negative binomial regression; and (iii) CMP regression. Under each model, we use the covariate vector $\boldsymbol{x}_{ij}^{\mathrm{T}} = (1, \mathrm{trt}_{ij}, \mathrm{time}_{ij}, \mathrm{trt}_{ij} \times \mathrm{time}_{ij})$ and $\boldsymbol{z}_{ij}^{\mathrm{T}} = (1, \mathrm{time}_{ij})$. For Bayesian analysis, we take priors as $\pi(\phi) = \mathrm{LN}(0, 0.5)$, $\pi(\boldsymbol{\beta}) = N(\mathbf{0}, 100I)$, $\pi(\sigma_k) = U(0, 10)$ for $k = 1, 2$, and $\pi(R) = \mathrm{LKJ}(2)$. For over-dispersion parameter ϕ, the choice of hyper-parameter provides that with 95% probability ϕ is a priori between 0.38 to 2.66, covering a reasonable range of dispersion and it was also discussed in Choo-Wosoba *et al*. (2018). We could use inverse-gamma (IG) distribution for σ_1^2 and σ_2^2 to maintain the conjugacy for computational benefits, however, we simply use uniform distribution as this is typically considered in STAN users. Based on our experience, the posterior mean and corresponding 95% credible intervals are quite robust in using IG(0.01,0.01) vs. $U(0, 10)$. Also, weakly non-informative priors are placed on the regression coefficients $\boldsymbol{\beta}$. We considered the prior with a large variance 10,000, but the posterior estimates are quite robust (not reported here). We assume that the random effects are weakly correlated, so take $\eta = 2$ in L, which the correlation mostly lies in between -0.5 and 0.5 with a higher probability compared to outside of this range as shown in Figure 1.

Table 3 presents the posterior estimates including posterior means, standard deviations (SDs), the 95% credible interval (CR), and model assessment measures under three different models: Poisson, NB, and CMP regression models. Based on assessment measures, we select the model with the smallest value of DIC and the largest value of LPML. As expected, the proposed CMP regression model provides the least DIC (1538.17) and the greatest LPML (-815.14), implying that CMP model is a better fit compared to other two candidate models, Poisson and NB. The result clearly shows that CMP regression is more appropriate and a better fit in modeling for over-dispersed data and these are well-supported by DIC and LPML.

Statistical significance is determined whether the 95% credible interval (CR) of the posterior mean does include the value of zero or not. We see from Table 3 that between-subject and time covariate variabilities are reflected through the significantly greater than zero estimates for both random intercept and slopes in the three models. Correlation is also positively estimated, indicating that higher

between-subject variability results in higher variability of time effect. Furthermore, the estimates for over-dispersion are 0.13 and 0.42 for negative binomial and the CMP regressions, respectively. This represents that two models account for over-dispersion in addition to two variabilities with its correlation. Note that NB model underestimates the over-dispersion effect ϕ compared to the CMP regression model. The negative sign for the interaction between progabide and time (β_4) in all three models shows that the beneficial effect of study drug in reducing number of seizures for epilepsy patients, but these are not statistically significant. However, the estimated regression coefficients of $\boldsymbol{\beta}$ is not directly comparable across the models. This is because θ_i in the CMP distribution does not represent the mean of the distribution as it does for Poisson and NB regression model. Hence, an adjustment β_j/ϕ in the CMP model for $j = 1, \ldots, 4$ using posterior samples provides a crude comparison with β_j in Poisson and NB regression models as discussed in Sellers and Shmueli (2010).

We perform Bayesian analysis by using 'rstan' package in R running MCMC from STAN. A total of 20,000 MCMC samples were generated from the conditional posterior distributions of parameters, and 10,000 posterior samples were used for posterior summary after 10,000 burn-in. MCMC convergence is checked through trace plots and autocorrelation plots in Figures 3 to 4 in Appendix C, and the diagnostic procedure discussed in Chen *et al*. (2000). The STAN code for posterior sampling is postponed in **Appendix D**.

6. Concluding remark

We have proposed a Bayesian CMP regression model for longitudinal counts data. In particular, we incorporate random intercept and slope as well as inherent dispersion in the model. The proposed model is applied to the Epilepsy data to evaluate the efficacy of the treatment effect and examine overall time trend. We have found that the model using the CMP distribution has been substantially improved in the goodness-of-fit, and this is supported by two Bayesian model comparison criteria: DIC and LPML. Although we implemented the model with lower dimension of the random effects, the approach is feasible to more than two dimensional random effects under the proposed model.

As discussed in Sellers and Shmueli (2010), the CMP regression model outperforms NB and Poisson models in model fitting and predicting power by accounting for a wide spectrum of the dispersion. This may because the CMP distribution offers a flexible in modeling count data as the level of dispersion is directly estimated from the model. Although the method is developed using well-known existing clinical trial dataset, the same method can be equally applied to other clinical trials or biomedical studies where the observations are repeatedly measured over time based on count type data. For example, it is potentially applicable to estimate the number of taxa counts in microbiome study and the number of mortality in public health research. We implement the proposed model in user-friendly statistical software R using "rstan" package to conduct a Bayesian computation. The HMCMC algorithm is used for generating samples from the posterior distribution and this has been demonstrated to improve efficient search of parameter spaces in the target posterior distribution. The algorithm is also the default sampling algorithm in STAN.

While the MCMC sampling via HMCMC is successfully performed in our current setting, we are aware that there is a room for improvement to overcome computational challenge. As shown in Section 2.1, the normalizing constant of the CMP distribution is associated with calculating infinite sum, which is not possible to have the closed expression. This is also addressed by Choo-Wosoba *et al*. (2018) to determine the order for the normalizing constant. In our current research, the truncation level is chosen as $K = 50$ for epilepsy data to compute Bayesian model assessment measure as the tail probability after this term is negligible. Our approach has a feasible computational speed and time,

however, a better methodology needs to be developed for a high level of accuracy in statistical inference. We also remind the readers that the estimated regression coefficients under the three models, Poisson, NB, and CMP, may not be directly comparable. The parameter θ does not represent the mean in the CMP distribution as shown in Section 2.1, and this results in providing difficult comparisons and ambiguous interpretation for the estimated parameters. However, a simple adjustment or conversion β/ϕ allows us to approximate comparison of the estimated parameters between models as discussed in Sellers and Shmueli (2010).

Notice that we only consider fully observed counts and lower level of zero counts for all subjects in the current analysis. However, we often see many-zero counts or missing counts in longitudinal study. A natural extension of the proposed model is to develop either a Bayesian zero-inflated CMP regression model or Bayesian CMP model with missing observations. These models, however, will be computationally expensive as more random effects are associated with zero count component and missing component. Additionally, there is a growing attention to develop the joint modeling of longitudinal count data and survival outcome. The joint modeling offers an improved efficacy assessment of the treatment effects directly or indirectly, however, we need to deal with additional variability led by this model framework. The feasibility of such extensions and future research is currently under consideration.

Acknowledge

We would like to thank the Editor, the Associate Editor, and the two anonymous reviewers for their very helpful and constructive comments along with suggestions. This has led to a substantial improved version of the article.

Appendix A. Full conditional distributions

The full conditional distributions to implement MCMC algorithm are given as follows. The conditional posterior distribution of $\boldsymbol{\beta}$ given others is of the form

$$\pi(\boldsymbol{\beta} \mid V, R, \boldsymbol{\zeta}, \phi, \mathcal{D}) \propto \left\{ \prod_{i=1}^{N} \prod_{j=1}^{n_i} \left(\frac{1}{y_{ij}!} \right)^{\phi} \left(\exp(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i) \right)^{y_{ij}} \left(\sum_{k=0}^{\infty} \frac{\left(\exp(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i) \right)^{k}}{(k!)^{\phi}} \right)^{-1} \right\} \times \exp\left(-\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\Omega^{-1}\boldsymbol{\beta} \right).$$

The conditional posterior distribution of ϕ given others is written as

$$\pi(\phi \mid \boldsymbol{\beta}, V, R, \boldsymbol{\zeta}, \mathcal{D}) \propto \left\{ \prod_{i=1}^{N} \prod_{j=1}^{n_i} \left(\frac{1}{y_{ij}!} \right)^{\phi} \left(\exp(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i) \right)^{y_{ij}} \left(\sum_{k=0}^{\infty} \frac{\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}}\boldsymbol{\zeta}_i\right) \right)^{k}}{(k!)^{\phi}} \right)^{-1} \right\} \times \frac{1}{ab} \exp\left[-\frac{1}{2} \left(\frac{\log\phi - a}{b} \right)^2 \right].$$

The conditional posterior distribution of $\boldsymbol{\zeta}$ given others is given by

$$\pi(\boldsymbol{\zeta} \mid \boldsymbol{\beta}, V, R, \phi, \mathcal{D}) \propto \left\{ \prod_{i=1}^{N} \prod_{j=1}^{n_i} \left(\frac{1}{y_{ij}!} \right)^{\phi} \left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}} \boldsymbol{\zeta}_i \right) \right)^{y_{ij}} \left(\sum_{k=0}^{\infty} \frac{\left(\exp\left(\boldsymbol{x}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathrm{T}} \boldsymbol{\zeta}_i \right) \right)^{k}}{(k!)^{\phi}} \right)^{-1} \right\} \times \prod_{i=1}^{N} \exp\left(-\frac{1}{2} \boldsymbol{\zeta}_i^{\mathrm{T}} (\mathrm{VRV})^{-1} \boldsymbol{\zeta}_i \right).$$

The conditional posterior distribution of $V = \mathrm{diag}(\sigma_0, \sigma_1)$ given others is written as

$$\pi(V \mid \boldsymbol{\beta}, \boldsymbol{\zeta}, R, \phi, \mathcal{D}) \propto \prod_{i=1}^{N} |V|^{-1} \exp\left(-\frac{1}{2} \boldsymbol{\zeta}_i^{\mathrm{T}} (\mathrm{VRV})^{-1} \boldsymbol{\zeta}_i \right).$$

The conditional posterior distribution of R is written as

$$\pi(R \mid \boldsymbol{\beta}, \boldsymbol{\zeta}, V, \phi, \mathcal{D}) \propto \prod_{i=1}^{N} |R|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{\zeta}_i^{\mathrm{T}} (\mathrm{VRV})^{-1} \boldsymbol{\zeta}_i \right).$$

Appendix B. Calculation of the LPML

The logarithm pseudo marginal likelihood (LPML) is defined as the sum of the individual logarithm CPO values. As shown in the main article, original definition of the CPO is given by

$$\mathrm{CPO}_i = \int f(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}) \, \pi\left(\boldsymbol{\theta} \mid \mathcal{D}^{(-i)} \right) d\boldsymbol{\theta},$$

where $\mathcal{D}^{(-i)}$ is the observed data deleting the subject i. Following Chen *et al.* (2000), the CPO is rewritten as

$$\mathrm{CPO}_i = \left\{ \int \frac{1}{f(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})} \pi(\boldsymbol{\theta} \mid \mathcal{D}) \, d\boldsymbol{\theta} \right\}^{-1}.$$

The proof of this identity can be found from Chapter 10 of Chen *et al.* (2000). The CPO above can be obtained using Monte Carlo approach using the posterior samples, which is given by

$$\widehat{\mathrm{CPO}}_i^{-1} = \frac{1}{B} \sum_{b=1}^{B} \frac{1}{f\left(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}^{(b)} \right)}.$$

However, Zhang *et al.* (2017) developed a weighted Monte Carlo estimation to compute CPO_i in the presence of random effects. The approach is required to have a normalized weight function $w_i(\boldsymbol{\zeta}_i)$ where $\int w_i(\boldsymbol{\zeta}_i) d\boldsymbol{\zeta}_i = 1$.

$$\widehat{\mathrm{CPO}}_i^{-1} = \frac{1}{B} \sum_{b=1}^{B} \frac{w_i\left(\boldsymbol{\zeta}_i^{(b)} \right)}{f\left(\boldsymbol{y}_i, \boldsymbol{\zeta}_i^{(b)} \mid \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}^{(b)} \right)},$$

where $\{(\boldsymbol{\zeta}_i^{(b)}, \boldsymbol{\theta}^{(b)}), b = 1, \ldots, B\}$ is a Gibbs sample. As shown in Zhang *et al.* (2017), the approach is more efficient and weight function can be obtained by the conditional density of the random effects $\boldsymbol{\zeta}_i$ given other parameters and data, $w_i(\boldsymbol{\zeta}_i) = f(\boldsymbol{\zeta}_i | \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{\theta})$.

Appendix C. Trace plots and autocorrelation plots

We present trace and autocorrelation plots for all model parameters.

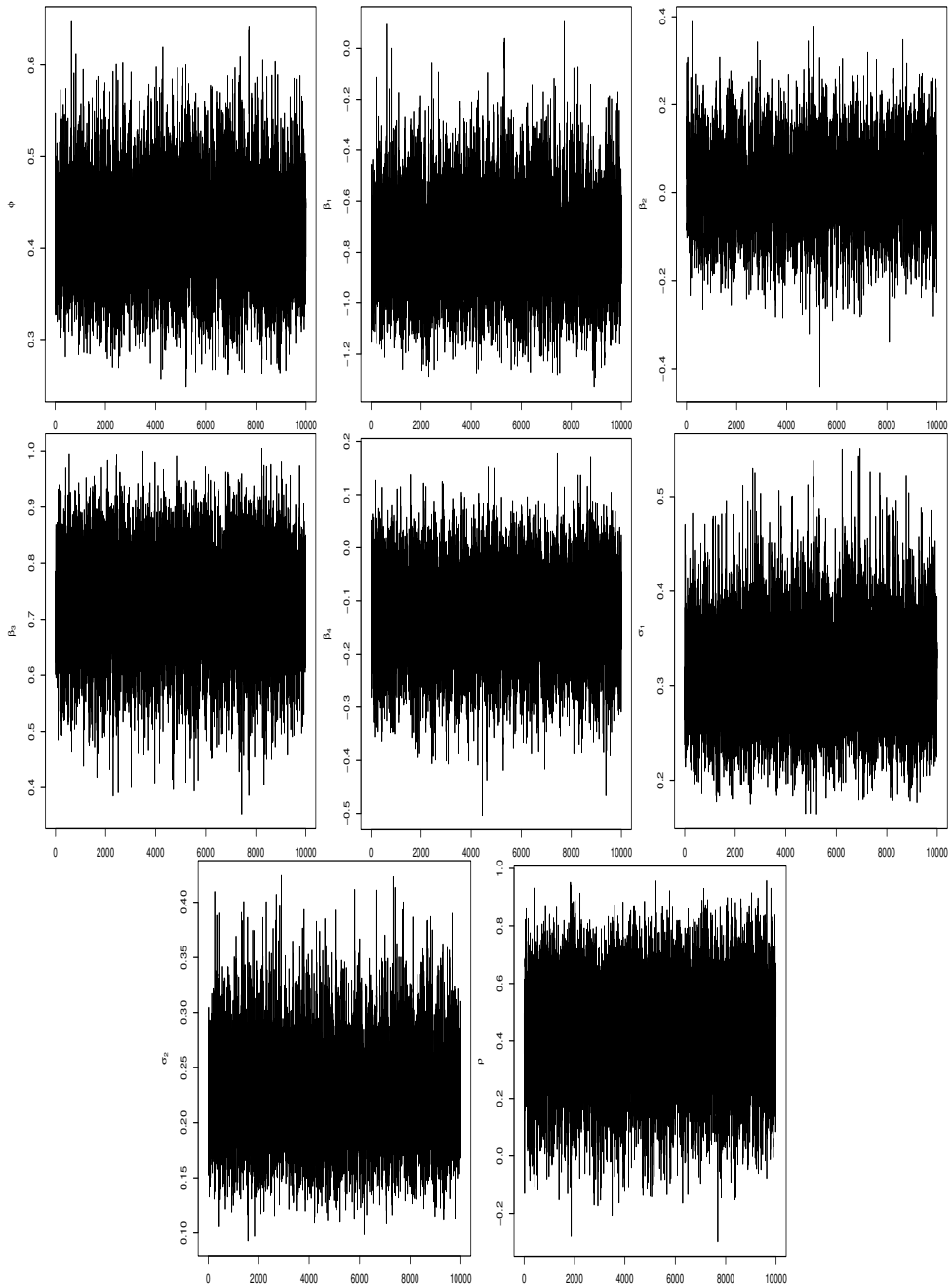

Figure 3: *Trace plots for all parameters under the proposed model.*

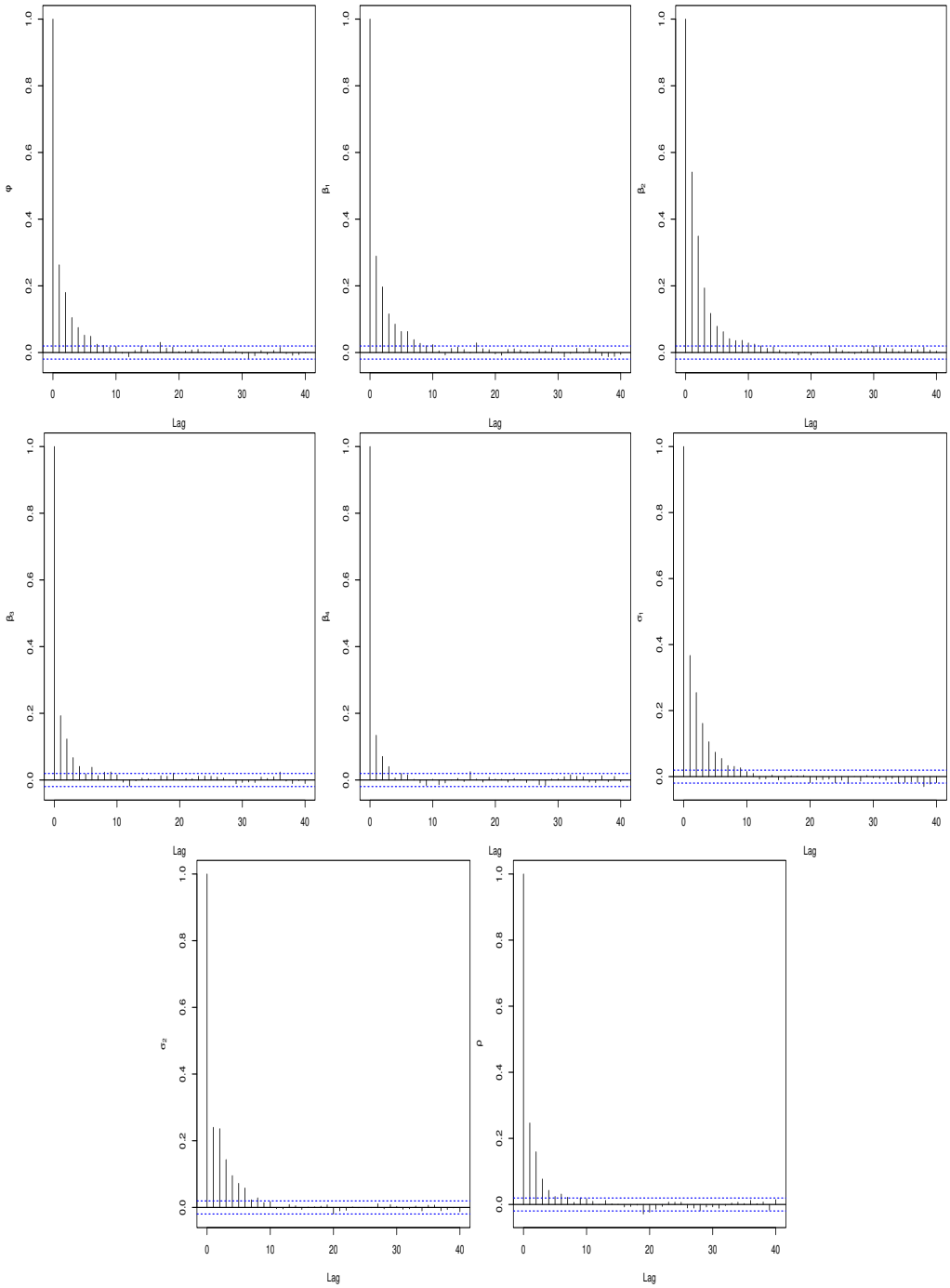

Figure 4: *Autocorrelation plots for all parameters under the proposed model.*

Appendix D. STAN code for Bayesian computation

```
functions{
  real Z(real lambda, real nu){
        real sm;
        real sm_prev;
        int k;
        real diff;
        sm=0;
        k=0;
        diff=1;
        while (diff > 0.001){
          sm_prev=sm;
          sm=sm+exp(k*log(lambda)-nu*lgamma(k+1));
          diff=sm-sm_prev;
          k=k+1;
        }
        return(sm);
  }

  vector Zv(vector theta, real phi){
        int N = rows(theta);
        vector[N] zs;
        for (i in 1:N){
          zs[i] = Z(theta[i], phi);
        }
        return(zs);
  }

  real com_poisson_lpdf(vector y, vector theta, real phi){
     int N = rows(y);
     return sum(y .*log(theta))-sum(log(Zv(theta, phi)))-phi*sum(lgamma(y+1));
  }
}

data {
     int<lower=0> N;
     int<lower=0> P;
     int<lower=0> J;
     vector[N] y;
     matrix[N,P] X;
     vector[N] trt;
     vector[N] time;
     vector[N] ltime;
     int<lower=1,upper=J>subj[N];
}

parameters {
     real<lower=0.0> phi;
     vector[P] beta;
     vector<lower=0>[2] sigma_u;
     cholesky_factor_corr[2] L_u;
     matrix[2,J] z_u;
}

transformed parameters {
     matrix[2,J] u;
     u=diag_pre_multiply(sigma_u,L_u)*z_u;
}

model {
  vector[N] yhat;
  L_u~\text{LKJ}_corr_cholesky(2.0);
  to_vector(z_u)~normal(0,1);
  sigma_u~uniform(0,10);
  phi~lognormal(0,0.5);

  for (i in 1:N)
```

```
    yhat[i]=exp(X[i]*beta+u[1,subj[i]]+u[2,subj[i]]*time[i]+ltime[i]);
    y ~ com_poisson(yhat,phi);
}

generated quantities {
  vector[N] yhat;
  for (i in 1:N){
    yhat[i]=exp(X[i]*beta+u[1,subj[i]]+u[2,subj[i]]*time[i]+ltime[i]);
  }
}
```

References

Albert J (1992). A Bayesian Bayesian analysis of a poisson random effects model for home run hitters, *The American Statistician*, **46**, 246–253.

Alvarez I, Niemi J, and Simpson M (2014). Bayesian inference for a covariance matrix, *Conference on Applied Statistics in Agriculture 2014*, **26**, 71–82, Available from: *arXiv preprint arXiv:1408.4050*

Barnard J, McCulloch R, and Meng XL (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage, *Statistica Sinica*, **10**, 1281–1311.

Betancourt M, Byrne S, Livingstone S, and Girolami M (2017). The geometric foundations of hamiltonian monte carlo, *Bernoulli*, **23**, 2257–2298.

Carpenter B, Gelman A, Hoffman MD *et al.* (2017). Stan: A probabilistic programming language, *Journal of Statistical Software*, **76**, 1–32.

Chen MH, Shao QM, and Ibrahim JG (2000). *Monte Carlo Methods in Bayesian Computation*, Springer, New York.

Choo-Wosoba H, Gaskins J, Levy S, and Datta S (2018). A Bayesian approach for analyzing zero-inflated clustered count data with dispersion, *Statistics in Medicine*, **37**, 801–812.

Consul PC and Jain GC (2004). A generalization of the poisson distribution, *Technometrics*, **15**, 791–799.

Conway RW and Maxwell WL (1962). A queuing model with state dependent service rates, *Journal of Industrial Engineering*, **12**, 132–136.

del Castillo J and P´erez-Casany M (2005). Overdispersed and underdispersed poisson generalizations, *Journal of Statistical Planning and Inference*, **134**, 486–500.

Famoye F (1993). Restricted generalized poisson regression model, *Communications in Statistics (Theory and Methods)*, **22**, 1335–1354.

Famoye F, Wulu JT, and Singh KP (2004). On the generalized poisson regression model with an application to accident data, *Journal of Data Science*, **2**, 287–295.

Fitzmaurice G, Davidian M, Verbeke G, and Molenberghs G (2008). *Longitudinal Data Analysis*, Columbia University, New York.

Geisser S and Eddy WF (1979). A predictive approach to model selection, *Journal of the American Statistical Association*, **74**, 153–160.

Gelman A (2006). Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis*, **1**, 515–533.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, and Rubin DB (2013). *Bayesian Data Analysis*, Chapman and Hall/CRC, Boca Raton, Florida, USA.

Guikema S and Goffelt J (2008). A flexible count data regression model for risk analysis, *Risk Analysis*, **28**, 213–223.

Hedeker D and Gibbons RD (2006). *Longitudinal Data Analysis, volume 451*, JohnWiley & Sons,

Hoboken, New Jersey, USA.

Hoffman M and Gelman A (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, *Journal of Machine Learning Research*, **15**, 1593–1623.

Huang A and Wand MP (2013). Simple marginally noninformative prior distributions for covariance matrices, *International Society for Bayesian Analysis*, **8**, 439–452.

Ibarhim JG, Chen MH, and Shiha D (2001). *Bayesian Survival Analysis*, Springer, New York.

Leppik I, Dreifuss F, Porter R *et al.* (1987). A controlled study of progabide in partial seizures: Methodology and results, *Neurology*, **37**, 963–968.

Lewandowski D, Kurowicka D, and Joe H (2009). Generating random correlation matrices based on vines and extended onion method, *Journal of Multivariate Analysis*, **100**, 1989–2001.

Morris D, Sellers K, and Menger A (2017). Fitting a flexible model for longitudinal count data using the NLMIXED procedure, *SAS Global Forum Paper*, **202**, 1–6.

Neal R (2011). *MCMC Using Hamiltonian Dynamics (Handbook of Markov Chain Monte Carlo)*, CRC Press, Boca Raton, Florida.

Neelon B (2019). Bayesian zero-inflated negative binomial regression based on Pölya-Gamma mixtures, *Bayesian Analysis*, **14**, 829–855.

O'Malley AJ and Zaslavsky AM (2008). Domain-Level covariance analysis for multilevel survey data with structured nonresponse, *Journal of the American Statistical Association*, **103**, 1405–1418.

Ridout MS and Besbeas P (2004). An empirical model for underdispersed count data, *Statistical Modelling*, **4**, 77–89.

Sellers KF, Borle S, and Shmueli G (2012). The com-poisson model for count data: A survey of methods and applications, *Applied Stochastic Models in Business and Industry*, **28**, 104–116.

Sellers KF and Morris DS (2017). Underdispersion models: Models that are "under the radar", *Communications in Statistics (Theory and Methods)*, **46**, 12075–12086.

Sellers KF and Shmueli G (2010). A flexible regression model for count data, *The Annals of Applied Statistics*, **4**, 943–961.

Shmueli G, Minka TP, Kadane JB, Borle S, and Boatwright P (2005). A useful distribution for fitting discrete data: Revival of the Conway–Maxwell–Poisson distribution, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 127–142.

Spiegelhalter DJ, Best NG, Carlin BP, and Van Der Linde A (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B (statistical methodology)*, **64**, 583–639.

Thall P and Vail S (1990). Some covariance models for longitudinal count data with overdispersion, *Biometrics*, **46**, 657–671.

Tokuda T, Goodrich B, Van Mechelen I, Gelman A, and Tuerlinckx F (2011). Visualizing distributions of covariance matrices, *Columbia University, New York, USA*, **1**, 1–30.

Tsonaka R and Spittle P (2020). Negative binomial mixed models estimated with the maximum likelihood method can be used for longitudinal rnaseq data, *Bioinformatics*, **22**, Available from: http://doi.org/10.1093/bib/bbaa264

Wang Z, Wu Y, and Chu H (2018). On equivalence of the LKJ distribution and the restricted Wishart distribution, *arXiv: Computation*, Available from: arXiv preprint arXiv:1809.04746

Wu J, Chen MH, Schifano E, Ibrahim JG, and Fisher J (2019). A new Bayesian joint model for longitudinal count data with many zeros, intermittent missingness, and dropout with applications to HIV prevention trials, *Statistics in Medicine*, **38**, 5565–5586.

Zhang D, Chen MH, Ibrahim JG, Boye ME, and Shen W (2017). Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials, *Journal of*

Computational and Graphical Statistics, **26**, 121–133.

Zhang X, Pei Y, Zhang L, Gun B, Pendegraft A, Zhuang W, and Yi N (2018). Negative binomial mixed models for analyzing longitudinal microbiome data, *Frontiers in Microbiology*, **9**, Available from: http://doi.org/10.3389/fmicb.2018.01683

Received October 3, 2022; Revised March 22, 2023; Accepted March 22, 2023