# Two-Component Poisson Mixture Regression Modelling of Count Data with Bivariate Random Effects

Kui Wang, Kelvin K.W. Yau*

Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong

Andy H. Lee

Department of Epidemiology and Biostatistics, Curtin University of Technology, Perth, WA 6845, Australia

Geoffrey J. McLachlan

Department of Mathematics and the Institute for Molecular Biology, University of Queensland, Brisbane, QLD 4072, Australia

---

* **Correspondence author:**

Professor Kelvin K.W. Yau

Department of Management Sciences

City University of Hong Kong

Tat Chee Avenue, Kowloon

Hong Kong

Phone: +852-2788 8585

Fax: +852-2788 8560

† E-mail: mskyau@cityu.edu.hk

**Two-component Poisson mixture regression modelling of count data with bivariate random effects**

ABSTRACT

Two-component Poisson mixture regression is typically used to model heterogeneous count outcomes that arise from two underlying sub-populations. Furthermore, a random component can be incorporated into the linear predictor to account for the clustering data structure. However, when including random effects in both components of the mixture model, the two random effects are often assumed to be independent for simplicity. A two-component Poisson mixture regression model with bivariate random effects is proposed to deal with the correlated situation. A restricted maximum quasi-likelihood estimation procedure is provided to obtain the parameter estimates of the model. A simulation study shows both fixed effects and variance component estimates perform well under different conditions. An application to childhood gastroenteritis data demonstrates the usefulness of the proposed methodology, and suggests that neglecting the inherent correlation between random effects may lead to incorrect inferences concerning the count outcomes.

Keywords: Bivariate random effects, Clustered data, Mixture regression model, Overdispersion, Variance components

## 1. Introduction

The analysis of count data that are overdispersed relative to the Poisson distribution (i.e., variance > mean) has been paid much attention. It is well known that application of the standard Poisson model may result in misleading inferences. Alternative methods of analysis have been proposed to deal with the overdispersion problem. Finite mixture models are appropriate when the extra variability comes from the unobserved heterogeneity of the population, which composes of two or more subgroups mixed in various proportions [1, 2]. In particular, Poisson mixture regression can be fitted to such overdispersed count data in the presence of covariate information. The methodology of Poisson mixture regression, including properties of estimators, adequacy of fit, and model selection procedure, have been extensively investigated in the literature [3].

In the manner of generalized linear mixed models (GLMM) [4], when the count outcomes are clustered or represent repeated measurements, random effects can be incorporated within the regression model to account for the inherent correlation between observations. Typically, fixed effect parameters and variance component estimation can be implemented using numerical procedures to compute restricted maximum likelihood estimates [5]. Recently, Chan *et al*. [6] proposed a Monte Carlo approximation method to achieve maximum likelihood estimation in GLMM. A two-component Poisson mixture regression model was developed for analysing maternity length of hospital stay, where random effects were included in the linear predictor of both mixture components to accommodate the hierarchical nesting of observations [7]. Similarly, a normal mixture regression model with random hospital effects was presented by Ng *et al*. [8] to model the clustered neonatal length of stay. Both approaches assume the random effects in different mixture components are independent.

In this study, the two-component Poisson mixture regression is extended to the bivariate random effects setting for the mixture components. Specifically, the random effect vectors in the two components are assumed to be bivariate normally distributed with a correlation parameter. Such an extension allows the estimation of the correlation between the two components via the random effects, and enables the assessment of the relationship between the first and second components. Ignoring the correlation between random effects may result in misleading inferences.

In Section 2, a numerical estimation procedure is developed to obtain the fixed effect parameter estimates. The estimation equations for random effect parameters are derived in Section 3. A simulation study is conducted in Section 4 to assess the performance of the estimators, followed by a practical example on childhood gastroenteritis in Section 5 to demonstrate the applicability of the method. Discussion on computational issues and further extensions are given in Section 6.

## 2. Poisson Mixture Regression Model with Correlated Random Effects

Let $Y_{ij}$ ($i$ = 1,2,…, $m$; $j$ = 1,2,…, $n_i$) be the count variable of the $j^{th}$ observation in the $i^{th}$ cluster, where $m$ is the number of clusters and $\sum_{i=1}^{m} n_i = n$ is the sample size. Suppose $Y_{ij}$ comes from a mixture of two Poisson populations with mixing probability $p$, then the probability distribution function is

$$P(Y_{ij} = y_{ij}) = p\frac{\lambda_{1,ij}^{y_{ij}}}{y_{ij}!}\exp(-\lambda_{1,ij}) + (1-p)\frac{\lambda_{2,ij}^{y_{ij}}}{y_{ij}!}\exp(-\lambda_{2,ij}),$$

where $\lambda_1$ and $\lambda_2$ are the means of the two components. In the regression setting, both $\log(\lambda_{1,ij})$ and $\log(\lambda_{2,ij})$ are assumed to be a linear function of covariates. The covariates

appearing in these two parts are not necessarily the same. With random effects, the linear predictors $\eta_{1,ij}$ and $\eta_{2,ij}$ are defined as:

$$\log(\lambda_{1,ij}) = \eta_{1,ij} = x_{1,ij}^T \beta_1 + v_{1,i}, \ \log(\lambda_{2,ij}) = \eta_{2,ij} = x_{2,ij}^T \beta_2 + v_{2,i},$$

where $x_{1,ij}$ and $x_{2,ij}$ are vectors of covariates with corresponding regression coefficients $\beta_1$ and $\beta_2$. Let $v_1 = (v_{1,1},...,v_{1,m})^T$ and $v_2 = (v_{2,1},...,v_{2,m})^T$ represent random effects for the first and second component of a two-component mixture model, respectively. The pair $v_1$ and $v_2$ are likely to be correlated if they arise from the same hospital or cluster. Therefore, $u_i^T = (v_{1i}, v_{2i})^T$ is assumed to be distributed as $N(0, A_i)$, where the matrix

$$A_i = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, i = 1, 2,\ldots, m.$$

When the correlation coefficient $\rho$ is zero, $v_1$ and $v_2$ become independent, so that the model degenerates into that of Wang *et al*. [7]. Following the GLMM formulation [4, 5], the best linear unbiased prediction (BLUP) type log-likelihood is given by $l = l_1 + l_2$, where

$$l_1 = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log\left( p\frac{e^{-\lambda_{1,ij}}(\lambda_{1,ij})^{y_{ij}}}{y_{ij}!} + (1-p)\frac{e^{-\lambda_{2,ij}}(\lambda_{2,ij})^{y_{ij}}}{y_{ij}!} \right),$$

$$l_2 = -\frac{1}{2}\left[ m\log(2\pi \mid A_i \mid) + \sum_{i=1}^{m} u_i^T A_i^{-1} u_i \right],$$

with $l_1$ being the log-likelihood function when the random effects are conditionally fixed, while $l_2$ can be regarded as the penalty function for the conditional log-likelihood. Estimation may be performed iteratively. In the initial step, coefficients in the linear predictors are estimated, for fixed variance components, by maximizing the above BLUP log-likelihood. Estimation of the variance component parameters is then achieved using restricted maximum quasi-likelihood (REMQL) estimating equations; details of which are presented in the next section.

Let $\eta_1 = X_1\beta_1 + Rv_1$, $\eta_2 = X_2\beta_2 + Rv_2$, $\log\left(\frac{p}{1-p}\right) = \xi = W\alpha$, where $W = (1,1,...,1)^T$. Denote $u = (u_1^T, u_2^T, ..., u_m^T)^T$ and $\Phi = (\alpha, \beta_1^T, \beta_2^T, u^T)^T$ is the vector of unknown parameters of interest. Given initial values $\Phi_0$, the estimating equation is thus

$$\hat{\Phi} = \Phi_0 + V^{-1}\frac{\partial l}{\partial \Phi}, \text{ where } V = -\frac{\partial^2 l}{\partial \Phi \partial \Phi^T}.$$

Details of the derivatives involved are provided in the Appendix. The asymptotic standard errors of the parameter estimates are obtained from the approximate variance-covariance matrix $V^{-1}$.

### 3. Estimation of Variance Components

When performing model estimation via the Newton-Raphson procedure in the previous section, the parameters of the variance component are assumed to be known. In practice, they are updated and estimated in each cycle of the procedure using the approximate REMQL estimators of the variance components [7, 8], which are obtained by solving the equations of the first-order derivatives of the REMQL log-likelihood with respect to $\phi = (\sigma_1, \sigma_2, \rho)$, viz.

$$tr\left(A^{-1}\frac{\partial A}{\partial \phi}\right) + tr\left((uu^T + S)\frac{\partial A^{-1}}{\partial \phi}\right) = 0, \quad (1)$$

where $A = [A_1, A_2, ..., A_m]$ denotes a diagonal block matrix, and $S$ is the block matrix portion of $V^{-1}$ corresponding to $u$. The following simplification is performed to obtain the explicit equations for the three parameters. Note that

$$\frac{\partial A_i^{-1}}{\partial \sigma_1} = \frac{1}{\sigma_1^3\sigma_2^2(1-\rho^2)}\begin{pmatrix} -2\sigma_2^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & 0 \end{pmatrix}, \frac{\partial A_i^{-1}}{\partial \sigma_2} = \frac{1}{\sigma_1^2\sigma_2^3(1-\rho^2)}\begin{pmatrix} 0 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & -2\sigma_1^2 \end{pmatrix},$$

$$\frac{\partial A_i^{-1}}{\partial \rho} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}\begin{pmatrix} 2\rho\sigma_2^2 & -(1+\rho^2)\sigma_1\sigma_2 \\ -(1+\rho^2)\sigma_1\sigma_2 & 2\rho\sigma_1^2 \end{pmatrix}, i = 1,2,\ldots,m,$$

and the parameters $\sigma_1$, $\sigma_2$, and $\rho$ satisfy

$$tr\left( A^{-1}\frac{\partial A}{\partial \sigma_1}\right)=\frac{2m}{\sigma_1},\ tr\left( A^{-1}\frac{\partial A}{\partial \sigma_2}\right)=\frac{2m}{\sigma_2}\ \text{and}\ tr\left( A^{-1}\frac{\partial A}{\partial \rho}\right)=-\frac{2m\rho}{(1-\rho^2)},$$

which can be used to replace the first term of the estimation equation (1). Similarly, the following relations can be used to replace the second term of equation (1):

$$tr\left( (uu^T+S)\frac{\partial A^{-1}}{\partial \sigma_1}\right)=\frac{1}{\sigma_1^3\sigma_2^2(1-\rho^2)}(-2\sigma_2^2L_1+2\rho\sigma_1\sigma_2L_2),$$

$$tr\left( (uu^T+S)\frac{\partial A^{-1}}{\partial \sigma_2}\right)=\frac{1}{\sigma_1^2\sigma_2^3(1-\rho^2)}(-2\sigma_1^2L_3+2\rho\sigma_1\sigma_2L_2),$$

$$tr\left( (uu^T+S)\frac{\partial A^{-1}}{\partial \rho}\right)=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}[2\rho\sigma_2^2L_1-2(1+\rho^2)\sigma_1\sigma_2L_2+2\rho\sigma_1^2L_3],$$

where $L_1=tr\left(I(S+u^Tu)\right)$, $L_2=tr\left(J(S+u^Tu)\right)/2$, $L_3=tr(K(S+u^Tu)$ and $I$, $J$, and $K$ are $2m\times 2m$ diagonal block matrices with their component matrices being $\begin{pmatrix}1&0\\0&0\end{pmatrix}$, $\begin{pmatrix}0&1\\1&0\end{pmatrix}$, and $\begin{pmatrix}0&0\\0&1\end{pmatrix}$, respectively. Consequently, three estimation equations are established so that a standard numerical algorithm such as the Newton-Raphson method can be used to facilitate the estimation of random components. The model estimation procedure was implemented as a S-Plus computer program.

## 4. Simulation Study

A simulation study is conducted to assess the performance of the proposed method. Data are simulated under a two-component Poisson mixture regression model whereby the random effects of the two mixture components are correlated. In each simulated data set, $m$ = 20 clusters and $n_i$ = 60 observations are fixed within each cluster. This means 20 pairs of random effects from a bivariate normal distribution are generated. Each pair of these random effects is

used to generate 60 pairs of Poisson data, and the mixed responses are then randomly selected from the first component or the second component distribution with mixing probability *p*. The covariate vector $x_{ij}$ consists of the constant 1 and values randomly generated from the uniform (0, 1) distribution, with associated regression coefficients (1, -1) for the first mixture component and (2, 0.5) for the second mixture component. The mixing probability *p* is chosen to be 0.2, 0.5, and 0.8. For each *p*, values -0.8, -0.5, -0.3, 0, 0.3, 0.5, 0.8 are considered for the correlation parameter $\rho$, while moderate variations of 0.4 and 0.7 are assumed respectively for $\sigma_1$ and $\sigma_2$. The simulation study is thus designed to evaluate the performance of the estimators with respect to a plausible range of bivariate correlation and mixing probability values. The number of replications is 1000 for each setting considered.

Results of the simulation study are presented in Tables 1, 2 and 3, which report the average bias of the parameter estimates and standard error of the estimates over the 1000 replications. From the tables, it is evident that the REMQL estimates of the regression coefficients and the mixture proportion have negligible biases relative to their corresponding standard errors. For variance component parameters, the REMQL estimates of $\sigma_1$, $\sigma_2$ and $\rho$ also perform reasonably well in each of the settings considered. Simulation results ascertain the applicability of the proposed algorithm for parameter estimation in the two-component Poisson mixture regression model with correlated random effects.

**5. Example**

Gastroenteritis is an infectious disease prevalent among Aboriginal infants and children in Australia who live in overcrowded and unhygienic conditions [9]. In this study, the proposed two-component Poisson mixture regression model with bivariate random effects is applied to investigate whether demographic and clinical factors affect the number of diagnoses at each

hospital admission of childhood gastroenteritis. A two-component Poisson mixture model is considered appropriate because of heterogeneity in patient outcome, especially for those infants who are severely dehydrated and have developed other complications. To achieve our study objective, acute hospital discharge data on age (in months), gender, indigenous status (Aboriginal, non-Aboriginal), admission status (elective, emergency), place of residence (rural, metropolitan) and principal diagnoses were extracted from the Western Australia hospital morbidity database for each episode of gastroenteritis between 1995 and 2002. All infants born in 1995 in Western Australia who had index gastroenteritis admission during their first year of birth were selected [9].

In this study, number of diagnoses at each episode is the outcome variable of interest. As evident from its empirical distribution given in Table 4, the count variable is clearly overdispersed with mean 1.366 and variance 2.397. Moreover, chi-square goodness-of-fit tests (without covariates) for the Poisson ($p$-value < 0.001), negative binomial ($p$-value = 0.034), zero-inflated Poisson ($p$-value < 0.001), and zero-inflated negative binomial ($p$-value = 0.017) distributions are all significant at the 5% level, whereas the two-component Poisson distribution appears to fit the count data reasonably well. The chi-square test statistic for the mixture distribution is 3.728 (4 degrees of freedom) with $p$-value = 0.444. Although the fit of other models will improve after incorporating concomitant information, the two-component Poisson mixture model is chosen for subsequent analysis.

There are altogether $m$ = 59 hospitals with a total of $n$ = 683 admissions identified during the study period. Most of the hospitalisations were due to “unspecified diarrhea” (55.8%), while 17.7% were attributed to “bacterial or viral diarrhea”. Other types of gastroenteritis, such as “Salmonellosis”, “Shigellosis” and “Protozoal intestinal diseases”, made up the minority of

hospital admissions. An inspection of the data shows that 91% of the cases were emergency admitted. Male infants contributed 56.8% of the admissions, and 41% of them were found to be of Aboriginal decent. The mean number of diagnosis of Aboriginals (1.94, SD 1.82) was significantly higher than their non-Aboriginal counterparts (0.97, SD 1.17), reflecting the severity of the disease and additional comorbidities experienced by this subgroup of patients.

Table 5 compares the results from fitting the two-component mixture regression model to the data for (a) $\rho = 0$ and (b) $\rho \neq 0$. The fixed effect regression coefficients are generally similar in terms of sign and magnitude between the two models. However, the random component estimates of model (b) are slightly larger, with the bivariate correlation coefficient estimated to be -0.539, indicating the hospital effect on the two mixture components are not independent. As expected, Aboriginality is positively associated with the number of diagnoses for both mixture components. Aboriginal infants, who often live in overcrowded and unhygienic conditions, tend to have additional complications and comorbidities when they are admitted for gastroenteritis [9]. The effects of other significant factors, namely principal diagnoses (first component) and admission status (second component), are also comparable with and without allowing for correlation between the random effects. It should be remarked that gender becomes marginally significant under the more flexible bivariate random effects model. Although the interpretation of its clinical significance is not clear, especially for gastroenteritis episodes exhibiting lesser number of diagnoses, the finding suggests that neglecting the correlation between the random effects may lead to incorrect inferences concerning the count outcome.

## 6. Discussion

A two-component Poisson mixture regression model with bivariate random effects is proposed to accommodate the inherent correlation between random effects. The simulation study confirms that the estimators perform reasonably well with varying correlation and different mixing probabilities. The empirical application demonstrates the usefulness of the proposed model, and provides further evidence that ignoring the correlation between random effects may result in misleading conclusions.

Similar to the Bayesian analysis framework, the posterior probabilities with respect to the components of the Poisson mixture model can be obtained. In particular, the posterior probability $p_{1,ij}$ of the observation $Y_{ij}$ with respect to the first component is given by

$$p_{1,ij} = p\frac{\lambda_{1,ij}^{y_{ij}}}{y_{ij}!}\exp(-\lambda_{1,ij}) \Big/ p\frac{\lambda_{1,ij}^{y_{ij}}}{y_{ij}!}\exp(-\lambda_{1,ij}) + (1-p)\frac{\lambda_{2,ij}^{y_{ij}}}{y_{ij}!}\exp(-\lambda_{2,ij}).$$

Such posterior probabilities may assist in the development of patient-specific programs for improving future outcomes.

The methodology can be further extended to cope with the general *k*-component Poisson mixture situation. Without loss of generality and adopting the extended set of notations, for a three-component Poisson mixture model, the probability distribution function may be specified as

$$P(Y_{ij} = y_{ij}) = p_1\frac{\lambda_{1,ij}^{y_{ij}}}{y_{ij}!}\exp(-\lambda_{1,ij}) + p_2\frac{\lambda_{2,ij}^{y_{ij}}}{y_{ij}!}\exp(-\lambda_{2,ij}) + (1-p_1-p_2)\frac{\lambda_{3,ij}^{y_{ij}}}{y_{ij}!}\exp(-\lambda_{3,ij}),$$

and the BLUP type log-likelihood $l = l_1 + l_2$ can be constructed accordingly. Subsequent derivations are rather tedious, nevertheless, the estimation procedure is in principle analogous to that of a two-component model. For the determination of the number of mixture

components, the Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) can be used in conjunction with the Pearson statistics [7].

In the process of REMQL estimation, good initial values are needed to guarantee convergence. Clustering methods such as *k*-means can be used for separating the response counts into *k* groups. Consequently, the mixing proportions of each group and the standard Poisson regression coefficients are computed for each component. Thereafter, these estimates serve as initial values for the two-component Poisson mixture regression model with independent random effects, following the recommendation of Wang *et al*. [7]. These initial values are the estimates obtained from an EM algorithm, in which convergence is guaranteed. Finally, parameter estimates from the reduced mixture model, together with zero as the initial value of $\rho$, provide a basis to start the iterative procedure for full model REMQL estimation. Based on our experience, this process of setting initial values works well as the divergence rate is less than 1% in the simulation study. The computation time is also acceptable with an average of less than 6 seconds for each converging case in the simulation study.

Comparable approaches for estimation in the GLMM exist in the literature, which include the Bayesian and Monte Carlo approximation methods. In general, the Bayesian approach using the Gibbs sampler is computationally intensive when the full conditional density is not in a standard form. The Monte Carlo estimation methods do not always converge to the global maximum and often impose difficulties in the estimation when the importance function used is far away from the true function. Based on the Gibbs output and adopting a Monte Carlo approximation to the relative marginal likelihood function, recent work by Chan *et al.* [6] bridges the two approaches. For the proposed REMQL estimation using the penalized quasi-likelihood, although it has a potential bias problem in the estimation of variance component

parameters, bias-correction procedures are available to improve the asymptotic performance of the estimates, and the REMQL estimation procedure is in general computationally efficient.

**Acknowledgements**

The authors are grateful to the reviewers for helpful comments. This research is supported by grants from the Research Grants Council of Hong Kong and the Australian Research Council.

**References**


1. D. Bohning, *Computer-Assisted Analysis of Mixture and Applications*, Chapman and Hall, New York, (1999).
2. G.J. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, (2000).
3. P. Wang, M.L. Puterman, I. Cockburn and N. Le, Mixed Poisson regression models with covariate dependent rates, *Biometrics* **52** (2), 381-400, (1996).
4. C.A. McGilchrist, Estimation in generalized mixed models, *Journal of the Royal Statistical Society B* **56** (1), 61-69, (1994).
5. C.A. McGilchrist and K.K.W. Yau, The derivation of BLUP, ML, REML estimation methods for generalized linear mixed models, *Communications in Statistics-Theory and Methods* **24** (12), 2963-2980, (1995).
6. J.S.K. Chan, A.Y.C. Kuk and C.H.K. Yam, Monte Carlo approximation through Gibbs output in generalized linear mixed models, *Journal of Multivariate Analysis* **94** (2), 300-312, (2005).
7. K. Wang, K.K.W. Yau and A.H. Lee, A hierarchical Poisson mixture regression model to analyse maternity length of hospital stay, *Statistics in Medicine* **21** (23), 3639-3654, (2002).
8. S.K. Ng, K.K.W. Yau and A.H. Lee, Modelling inpatient length of stay by a hierarchical mixture regression via the EM algorithm, *Mathematical and Computer Modelling* **37** (3-4), 365-375, (2003).
9. A.H. Lee, J. Flexman, K. Wang and K.K.W. Yau, Recurrent gastroenteritis among infants in Western Australia: A seven-year hospital-based cohort study, *Annals of Epidemiology* **14** (2), 137-142, (2004).

**Appendix. First and Second Derivatives of the BLUP log-likelihood**

From the BLUP log-likelihood $l = l_1 + l_2$, we obtain:

$$\frac{\partial l}{\partial \alpha} = W^T \frac{\partial l_1}{\partial \xi},\ \frac{\partial l}{\partial \beta_1} = X_1^T \frac{\partial l_1}{\partial \eta_1},\ \frac{\partial l}{\partial \beta_2} = X_2^T \frac{\partial l_1}{\partial \eta_2},\ \frac{\partial l_1}{\partial v_1} = R^T \frac{\partial l_1}{\partial \eta_1},\ \frac{\partial l_1}{\partial v_2} = R^T \frac{\partial l_1}{\partial \eta_2},$$

$$\frac{\partial l}{\partial u_i} = \frac{\partial l_1}{\partial u_i} + \frac{\partial l_2}{\partial u_i} = \begin{pmatrix} R_i^T \dfrac{\partial l_1}{\partial \eta_1} \\ R_i^T \dfrac{\partial l_1}{\partial \eta_2} \end{pmatrix} - A_i^{-1} u_i .$$

To get $\frac{\partial l}{\partial u}$, let $H$ be a simple $2m \times 2m$ matrix which satisfies:

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} H = u \text{, then } \frac{\partial l}{\partial u} = \begin{pmatrix} \dfrac{\partial l_1}{\partial v_1} \\ \dfrac{\partial l_1}{\partial v_2} \end{pmatrix} H - A^{-1} u .$$

Based on the above, second-order derivatives of the log-likelihood are obtained as follows:

$$\frac{\partial^2 l}{\partial \alpha \partial \alpha^T} = W^T \frac{\partial^2 l_1}{\partial \xi \partial \xi^T} W,\ \frac{\partial^2 l}{\partial \beta_1 \partial \beta_1^T} = X_1^T \frac{\partial^2 l_1}{\partial \eta_1 \partial \eta_1^T} X_1, \frac{\partial^2 l}{\partial \beta_2 \partial \beta_2^T} = X_2^T \frac{\partial^2 l_1}{\partial \eta_2 \partial \eta_2^T} X_2,$$

$$\frac{\partial^2 l}{\partial \beta_1 \partial \beta_2^T} = X_1^T \frac{\partial^2 l_1}{\partial \eta_1 \partial \eta_2^T} X_2,\ \frac{\partial^2 l}{\partial \beta_1 \partial \alpha^T} = X_1^T \frac{\partial^2 l_1}{\partial \eta_1 \partial \xi^T} W,\ \frac{\partial^2 l}{\partial \beta_2 \partial \alpha^T} = X_2^T \frac{\partial^2 l_1}{\partial \eta_2 \partial \xi^T} W,$$

$$\frac{\partial^2 l}{\partial \beta_1 \partial u^T} = \left( X_1^T \frac{\partial^2 l_1}{\partial \eta_1 \partial \eta_1^T} R, X_1^T \frac{\partial^2 l_1}{\partial \eta_1 \partial \eta_2^T} R \right) H ,$$

$$\frac{\partial^2 l}{\partial \beta_2 \partial u^T} = \left( X_2^T \frac{\partial^2 l_1}{\partial \eta_2 \partial \eta_1^T} R, X_2^T \frac{\partial^2 l_1}{\partial \eta_2 \partial \eta_2^T} R \right) H ,$$

$$\frac{\partial^2 l}{\partial \alpha \partial u^T} = \left( W^T \frac{\partial^2 l_1}{\partial \xi \partial \eta_1^T} R, W^T \frac{\partial^2 l_1}{\partial \xi \partial \eta_2^T} R \right) H ,$$

$$\frac{\partial^2 l}{\partial u \partial u^T} = H^T \begin{pmatrix} R^T \dfrac{\partial^2 l_1}{\partial \eta_1 \partial \eta_1^T} R & R^T \dfrac{\partial^2 l_1}{\partial \eta_1 \partial \eta_2^T} R \\ R^T \dfrac{\partial^2 l_1}{\partial \eta_2 \partial \eta_1^T} R & R^T \dfrac{\partial^2 l_1}{\partial \eta_2 \partial \eta_2^T} R \end{pmatrix} H - A^{-1} .$$

Therefore, as long as the first- and second-order derivatives of $l_1$ with respect to $\xi$, $\eta_1$, and $\eta_2$ are obtained, we can derive the estimation equations presented in Section 2.

Table 1. Bias and standard error of REMQL estimators based on 1000 replications of the two-component Poisson mixture regression model with $m = 20$, $n_i = 60$, mixing proportion $p = 0.2$ and varying bivariate correlations.

| Parameter | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | $p$ | $\sigma_1$ | $\sigma_2$ | $\rho$ |
|---|---|---|---|---|---|---|---|---|
| True value | 1 | -1 | 2 | 0.5 | 0.2 | 0.4 | 0.7 | |
| Simulation 1 | | | | | | | | -0.8 |
| Bias | -0.003 | -0.013 | 0.013 | -0.003 | -0.003 | -0.010 | -0.011 | 0.073 |
| Standard Error | 0.176 | 0.251 | 0.162 | 0.035 | 0.052 | 0.103 | 0.111 | 0.212 |
| Simulation 2 | | | | | | | | -0.5 |
| Bias | 0.003 | -0.010 | 0.000 | -0.002 | 0.002 | -0.024 | -0.018 | -0.066 |
| Standard Error | 0.175 | 0.252 | 0.160 | 0.036 | 0.054 | 0.109 | 0.112 | 0.291 |
| Simulation 3 | | | | | | | | -0.3 |
| Bias | 0.014 | -0.007 | 0.013 | -0.002 | 0.002 | -0.020 | -0.017 | 0.039 |
| Standard Error | 0.171 | 0.251 | 0.161 | 0.033 | 0.053 | 0.106 | 0.112 | 0.294 |
| Simulation 4 | | | | | | | | 0 |
| Bias | 0.012 | -0.011 | 0.009 | -0.002 | 0.002 | -0.023 | -0.008 | 0.034 |
| Standard Error | 0.174 | 0.248 | 0.166 | 0.034 | 0.049 | 0.106 | 0.116 | 0.332 |
| Simulation 5 | | | | | | | | 0.3 |
| Bias | 0.028 | -0.031 | 0.005 | 0.000 | -0.002 | -0.020 | -0.005 | 0.000 |
| Standard Error | 0.172 | 0.254 | 0.154 | 0.034 | 0.052 | 0.106 | 0.115 | 0.316 |
| Simulation 6 | | | | | | | | 0.5 |
| Bias | 0.024 | -0.011 | 0.004 | -0.001 | 0.001 | -0.020 | -0.011 | -0.018 |
| Standard Error | 0.175 | 0.229 | 0.151 | 0.035 | 0.052 | 0.106 | 0.114 | 0.289 |
| Simulation 7 | | | | | | | | 0.8 |
| Bias | 0.021 | -0.021 | 0.004 | -0.001 | 0.001 | -0.013 | -0.010 | 0.000 |
| Standard Error | 0.166 | 0.232 | 0.157 | 0.035 | 0.053 | 0.103 | 0.109 | 0.160 |

Table 2. Bias and standard error of REMQL estimators based on 1000 replications of the two-component Poisson mixture regression model with $m = 20$, $n_i = 60$, mixing proportion $p = 0.5$ and varying bivariate correlations.

| Parameter | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | $p$ | $\sigma_1$ | $\sigma_2$ | $\rho$ |
|---|---|---|---|---|---|---|---|---|
| True value | 1 | -1 | 2 | 0.5 | 0.5 | 0.4 | 0.7 | |
| Simulation 1 | | | | | | | | -0.8 |
| Bias | -0.008 | 0.005 | 0.019 | -0.007 | 0.002 | -0.016 | -0.023 | 0.029 |
| Standard Error | 0.119 | 0.137 | 0.161 | 0.049 | 0.065 | 0.080 | 0.113 | 0.128 |
| Simulation 2 | | | | | | | | -0.5 |
| Bias | 0.005 | -0.006 | 0.006 | -0.004 | 0.000 | -0.016 | -0.024 | 0.039 |
| Standard Error | 0.117 | 0.138 | 0.161 | 0.049 | 0.065 | 0.079 | 0.113 | 0.224 |
| Simulation 3 | | | | | | | | -0.3 |
| Bias | 0.060 | 0.001 | 0.018 | -0.004 | -0.001 | -0.011 | -0.022 | 0.037 |
| Standard Error | 0.115 | 0.134 | 0.161 | 0.045 | 0.066 | 0.075 | 0.113 | 0.239 |
| Simulation 4 | | | | | | | | 0 |
| Bias | 0.010 | -0.012 | 0.013 | -0.003 | 0.001 | -0.013 | -0.013 | 0.017 |
| Standard Error | 0.115 | 0.130 | 0.166 | 0.046 | 0.064 | 0.076 | 0.117 | 0.258 |
| Simulation 5 | | | | | | | | 0.3 |
| Bias | 0.008 | -0.003 | 0.009 | -0.001 | 0.001 | -0.013 | -0.008 | -0.006 |
| Standard Error | 0.115 | 0.129 | 0.154 | 0.046 | 0.064 | 0.079 | 0.117 | 0.247 |
| Simulation 6 | | | | | | | | 0.5 |
| Bias | 0.011 | -0.001 | 0.006 | 0.000 | 0.000 | -0.014 | -0.013 | -0.017 |
| Standard Error | 0.116 | 0.125 | 0.153 | 0.045 | 0.065 | 0.080 | 0.116 | 0.217 |
| Simulation 7 | | | | | | | | 0.8 |
| Bias | 0.013 | -0.006 | 0.007 | -0.002 | 0.002 | -0.013 | -0.012 | 0.001 |
| Standard Error | 0.117 | 0.129 | 0.160 | 0.046 | 0.062 | 0.077 | 0.111 | 0.112 |

Table 3. Bias and standard error of REMQL estimators based on 1000 replications of the two-component Poisson mixture regression model with $m = 20$, $n_i = 60$, mixing proportion $p = 0.8$ and varying bivariate correlations.

| Parameter | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | $p$ | $\sigma_1$ | $\sigma_2$ | $\rho$ |
|---|---|---|---|---|---|---|---|---|
| True value | 1 | -1 | 2 | 0.5 | 0.2 | 0.4 | 0.7 | |
| Simulation 1 | | | | | | | | -0.8 |
| Bias | -0.002 | 0.004 | 0.029 | -0.005 | 0.005 | -0.008 | -0.028 | 0.014 |
| Standard Error | 0.105 | 0.095 | 0.169 | 0.080 | 0.052 | 0.073 | 0.124 | 0.114 |
| Simulation 2 | | | | | | | | -0.5 |
| Bias | 0.000 | -0.002 | 0.029 | 0.000 | 0.004 | -0.014 | -0.036 | 0.035 |
| Standard Error | 0.104 | 0.097 | 0.168 | 0.078 | 0.053 | 0.071 | 0.123 | 0.206 |
| Simulation 3 | | | | | | | | -0.3 |
| Bias | 0.005 | -0.002 | 0.031 | -0.005 | 0.001 | -0.008 | -0.034 | 0.033 |
| Standard Error | 0.102 | 0.096 | 0.167 | 0.080 | 0.052 | 0.069 | 0.120 | 0.229 |
| Simulation 4 | | | | | | | | 0 |
| Bias | 0.003 | -0.003 | 0.028 | -0.005 | 0.001 | -0.007 | -0.027 | 0.021 |
| Standard Error | 0.102 | 0.096 | 0.171 | 0.080 | 0.050 | 0.072 | 0.121 | 0.250 |
| Simulation 5 | | | | | | | | 0.3 |
| Bias | 0.003 | 0.000 | 0.019 | 0.002 | 0.005 | -0.008 | -0.017 | 0.000 |
| Standard Error | 0.102 | 0.094 | 0.156 | 0.080 | 0.052 | 0.073 | 0.118 | 0.232 |
| Simulation 6 | | | | | | | | 0.5 |
| Bias | 0.002 | 0.004 | 0.012 | -0.001 | 0.001 | -0.011 | -0.024 | 0.003 |
| Standard Error | 0.103 | 0.093 | 0.164 | 0.083 | 0.052 | 0.073 | 0.119 | 0.187 |
| Simulation 7 | | | | | | | | 0.8 |
| Bias | 0.009 | 0.000 | 0.010 | 0.005 | 0.001 | -0.009 | -0.017 | -0.002 |
| Standard Error | 0.102 | 0.096 | 0.164 | 0.082 | 0.053 | 0.069 | 0.115 | 0.101 |

Table 4. Frequency distribution of number of diagnoses of gastroenteritis admissions for infants in Western Australia.

| Number of diagnoses | Observed frequency |
|---|---|
| 0 | 234 |
| 1 | 221 |
| 2 | 104 |
| 3 | 58 |
| 4 | 27 |
| 5 | 23 |
| 6 | 11 |
| 7 | 2 |
| 8 | 2 |
| 9 | 0 |
| 10 | 0 |
| 11 | 1 |

Table 5. Parameter estimates (standard error) from fitting the two-component Poisson mixture regression model to the childhood gastroenteritis data.

| | (a) $\rho = 0$ | | (b) $\rho \neq 0$ | |
|---|---|---|---|---|
| | First component | Second component | First component | Second component |
| $p$ | 0.598 (0.086) | | 0.559 (0.081) | |
| Intercept | 0.264 (0.293) | -0.685 (0.370) | 0.386 (0.286) | -0.678 (0.329) |
| Male gender | 0.339 (0.177) | -0.254 (0.172) | 0.365 (0.171)* | -0.236 (0.148) |
| Age | -0.009 (0.008) | -0.002 (0.010) | -0.006 (0.008) | -0.006 (0.009) |
| Aboriginal | 0.668 (0.199)* | 1.004 (0.209)* | 0.642 (0.196)* | 0.983 (0.179)* |
| Elective admission | -0.407 (0.259) | 0.571 (0.266)* | -0.447 (0.270) | 0.577 (0.253)* |
| Rural residence | -0.184 (0.267) | 0.122 (0.214) | -0.275 (0.254) | 0.119 (0.202) |
| Unspecified diarrhea | -0.521 (0.186)* | 0.232 (0.195) | -0.529 (0.171)* | 0.187 (0.165) |
| Bacterial or viral diarrhea | -1.051 (0.278)* | 0.205 (0.288) | -1.038 (0.306)* | 0.145 (0.261) |
| | $\sigma_1 = 0.382$ | | $\sigma_1 = 0.402$ | |
| | $\sigma_2 = 0.622$ | | $\sigma_2 = 0.669$ | |
| | | | $\rho = -0.539$ | |

*$p$-value < 0.05