HHS Public Access

Author manuscript

Stat Med. Author manuscript; available in PMC 2019 February 28.

Published in final edited form as:

Stat Med. 2018 February 28; 37(5): 801–812. doi:10.1002/sim.7541.

A Bayesian Approach for Analyzing Zero-Inflated Clustered Count Data with Dispersion

Hyoyoung Choo-Wosoba[1,a], **Jeremy Gaskins**[2,a], **Steven Levy**[3], and **Somnath Datta**[4,*]

[1]Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Rockville, Maryland 20850, U.S.A

[2]Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, Kentucky 40202, U.S.A

[3]Department of Preventive & Community Dentistry, Department of Epidemiology, University of Iowa, Iowa City, Iowa 52242, U.S.A

[4]Department of Biostatistics, University of Florida, Gainesville, Florida 32610, U.S.A

SUMMARY

In practice, count data may exhibit varying dispersion patterns and excessive zero values; additionally, they may appear in groups or clusters sharing a common source of variation. We present a novel Bayesian approach for analyzing such data. In order to model these features, we combine the Conway-Maxwell-Poisson distribution which allows both over- and under-dispersion with a hurdle component for the zeros and random effects for clustering. We propose an efficient Markov chain Monte Carlo sampling scheme to obtain posterior inference from our model. Through simulation studies, we compare our hurdle CMP model with a hurdle Poisson model to demonstrate the effectiveness of our CMP approach. Furthermore, we apply our model to analyze an illustrative dataset containing information on the number and types of carious lesions on each tooth in a population of 9-year-olds from the Iowa Fluoride Study, which is an ongoing longitudinal study on a cohort of Iowa children that began in 1991.

Keywords

Bayesian modeling; clustering; Conway-Maxwell-Poisson distribution; count data; zero inflation

1. Introduction

Zero inflation arises when the parametric model for the counts underestimates the proportion of zeros in the data.[1–4] A new component is introduced to the model so that data are drawn from a mixture model containing the count distribution and a binary component providing additional zeros. An alternative model structure is the hurdle model. The hurdle model is a conditional model that first determines if the count will be zero or non-zero. Conditional on

* somnath.datta@ufl.edu.

[a]These authors contributed equally to this paper.

the observation being non-zero, the count is drawn from a distribution with support on the positive integers.[5–7]

In practice, we often face situations where such count data are not independent and exhibit dependence between observations within clusters. The failure to appropriately account for dependence in the responses can lead to inefficient parameter estimates and invalid hypothesis tests. For analyzing this type of data, zero-inflated modelings have been developed to incorporate correlations in the marginal regression (i.e., GEE) framework.[8–12] There are also mixed effects models for zero-inflated clustered data that use random effects to introduce dependence.[13–19] However, these methods are typically limited to the equidispersed or overdispersed cases due to the properties of Poisson and negative binomial distributions.

On the other hand, the Conway-Maxwell-Poisson (CMP) distribution, introduced by Conway and Maxwell,[20] can model a wide range of dispersion from underdispersion to overdispersion and includes the usual Poisson distribution as a special case. Motivated by this versatility of the CMP distribution, Barriga and Louzada[21] introduced a Bayesian approach to zero-inflated dispersed data based on a CMP distribution. However, their work only considers independent data and not clustered data. On the other hand, Choo-Wosoba and Datta[22] developed statistical methodology with a CMP distribution for analyzing clustered data with excessive zero counts. In their paper, a mixed effects model approach is applied to handle the correlations within clusters. However, this frequentist approach typically is limited to equicorrelation through a single random component due to the difficulty of the Laplace approximation to the likelihood.[23]

The development of methodology in this paper is partially motivated by the Iowa Fluoride Study.[24] The Iowa Fluoride Study is a longitudinal study with the goal of identifying risk and protective factors for dental health in children. Information about the study is available at http://www.dentistry.uiowa.edu/preventive-fluoride-study. This data contains the caries experience score (CES) for each of the patient's teeth; the CES is a count variable with a higher scores indicating a more damaged tooth. Obviously, teeth within a child's mouth share the same dental environment, which implies that the CESs will be correlated within patient. Furthermore, most teeth are healthy (no cavities) leading to excessive zero counts in CES (Figure 1).

To deal with these issues, we propose a Bayesian CMP model that will simultaneously accommodate the common challenges mentioned previously: zero-inflation, clustering, and both over- and underdispersion. Our paper is unique in that we can handle all three in a fully likelihood-based approach. In our previous work, we considered a frequentist approach by using a marginal GEE specification[25] and a mixed effects model.[22] For the mixed effects model, the Laplacian/quadrature methods used for estimation struggle to estimate correlation structure more complex than equicorrelation. By relying on a Bayesian estimation scheme, we have more flexibility in specifying the dependence across teeth; in particular, we allow differing levels of correlation between different classes of teeth (molar and non-molar). While Barriga and Louzada[21] does use a Bayesian zero-inflated CMP distribution, they do not model dependence/clustering which is clearly needed in this application. Additionally,

we develop our model using a hurdle framework to account for the excess zeros, instead of a zero-inflation model. This provides more natural interpretations as there are separate models for those factors that cause/prevent cavities and those that lead to more/less severe cavities when they appear.

We introduce our Bayesian model in Section 2. Section 3 describes the Markov chain Monte Carlo (MCMC) sampling scheme used for inference. The application to the Iowa Fluoride Study data is undertaken and discussed in Section 4. We perform two simulations and compare the results with a Bayesian Poisson model in Section 5. The paper ends with a discussion in Section 6.

2. The Bayesian Model

In this section, we describe our model structure. We begin with the following general notation which will be used throughout the manuscript. Let N denote the total number of clusters (i.e., children for our illustrative data example), $i = 1, \ldots, N$ is the cluster level index, n_i is the sample size (number of teeth) in the i^{th} cluster, j indexes the observations (teeth) within the i^{th} cluster ($j = 1, \ldots, n_i$), and Y_{ij} represents the response (CES score) of the j^{th} observation in the i^{th} cluster.

To describe the role of the zeros in the data, we use a hurdle model instead of a zero-inflation component, due to its more natural interpretation. This Bayesian hurdle model consists of two different parts, the presence model and the severity model. The presence model considers a binary random variable for the non-zero outcome, i.e., whether there is any caries present on the tooth. The severity model, based on the CMP, describes the positive counts, that is, how much caries is on an involved tooth (how much decay there is quantified by the number of tooth surfaces affected). While we use the terms presence and severity for interpretation of the dental application, these can be considered generally as the zero model and the positive count model.

The presence model is based on a probit regression with both fixed and random effects terms. The probability that an outcome is positive (non-zero) is modeled through $P(Y_{ij}>0) = \Phi\left(X_{ij,\beta}^T \beta + U_{ij,\delta}^T \delta_i\right)$. This model is associated with fixed effects covariates X_{ij} and coefficient vector β. Clustering comes from the random effects vector δ_i with corresponding design matrix $U_{i,\delta}$. The choice of the design matrix $U_{i,\delta}$ determines the form of the clustering, allowing many common choices such as equicorrelation, multiple classes, etc.

The Conway-Maxwell-Poisson distribution[20] is given by

$$P(Y=y) = \frac{\lambda^y}{(y!)^\upsilon \mathscr{Z}(\lambda,\upsilon)},\ y=0,1,2,\cdots,$$

where $\mathscr{Z}(\lambda,\upsilon) = \sum_{s=0}^{\infty} \frac{\lambda^s}{(s!)^\upsilon}$ is the normalizing constant, λ is a positive shape parameter, and υ is a non-negative dispersion parameter. The parameter υ yields underdispersion if $\upsilon > 1$,

overdispersion if $0 < \upsilon < 1$, or equidispersion if $\upsilon = 1$. When $\upsilon = 1$, $\mathscr{Z}(\lambda, 1) = e^{\lambda}$, which implies that CMP distribution is the same as the Poisson distribution with mean λ. Interpretations from CMP modeling are guided by the result[26] that $E(Y^{\upsilon}) = \lambda$.

As the hurdle model is specified through $P(Y = y | Y > 0)$, we use a CMP restricted to the positive integers,

$$P(Y_{ij}=y|Y_{ij}>0) = \frac{(\lambda_{ij})^{y}}{(y!)^{\upsilon}\mathscr{Z}^{*}(\lambda_{ij},\upsilon)},\ y=1,2,\cdots, \quad (1)$$

where the new normalizing constant is $\mathscr{Z}^{*}(\lambda_{ij},\upsilon) = \mathscr{Z}(\lambda_{ij},\upsilon) - 1$. Thus, our severity model uses a truncated CMP distribution which excludes zero. The response-specific shape parameter λ_{ij} is modeled through $log\ \lambda_{ij} = X_{ij,\alpha}^{T}\alpha + U_{ij,\gamma}^{T}\gamma_i$. The regression coefficients α describe the fixed effects, whereas the random effect vector γ_i accounts for clustering. The severity model may or may not use the same random effects design matrix $U_{i,\delta}$ as the presence model. The full distribution of Y_{ij} (conditional on random effects) can be written as

$$P(Y_{ij}=y) = \begin{cases} 1 - \Phi\left(X_{ij,\beta}^{T}\beta + U_{ij,\delta}^{T}\delta_i\right), & y=0 \\ \Phi\left(X_{ij,\beta}^{T}\beta + U_{ij,\delta}^{T}\delta_i\right) \times \frac{exp\left\{y\left(X_{ij,\alpha}^{T}\alpha + U_{ij,\gamma}^{T}\gamma_i\right)\right\}}{(y!)^{\upsilon}\mathscr{Z}^{*}(\lambda_{ij},\upsilon)}, & y=1,2,\cdots \end{cases} \quad (2)$$

where $\mathscr{Z}^{*}(\lambda_{ij},\upsilon) = \sum_{s=1}^{\infty} exp\left\{s\left(X_{ij,\alpha}^{T}\alpha + U_{ij,\gamma}^{T}\gamma_i\right)\right\} / (s!)^{\upsilon}$.

In Equation 2, clustering across outcomes is induced by the random effects δ_i from the presence model and γ_i from the severity model. We assume they jointly follow a multivariate normal distribution: $(\delta_i, \gamma_i)^{T} \sim MVN(0, \Sigma)$. As previously noted, the form of the dependence in U is determined by the choice of the random effect design matrices U_{δ} and U_{γ}.

For the regression coefficients, we use proper and disperse priors $\beta \sim MVN(0, \Omega_{\beta})$ and $\alpha \sim MVN(0, \Omega_{\alpha})$. For the covariance matrix of the random effects, we use the conjugate inverse Wishart prior:

$$\sum = \begin{bmatrix} \sum_{\delta\delta} & \sum_{\delta,\gamma} \\ \sum_{\gamma,\delta} & \sum_{\gamma\gamma} \end{bmatrix} \sim IW(c, \Psi).$$

As the $\mathscr{Z}$ function of a CMP distribution is not available in a closed form, a conjugate prior for the dispersion is not available. We recommend the prior distribution of the dispersion parameter υ be a lognormal distribution (LN, hereafter) with a median of υ at 1, so that our model is centered at equidispersion. We choose the variance of $\log(\upsilon)$ to be 0.5^2 so that, with 95% probability, υ is a priori between 0.38 to 2.66, representing a reasonable range of dispersions. That is, $\upsilon \sim LN(0, 0.5^2)$. Values for the hyperparameters can be taken based on

subject matter experience. When none is available, we use proper, weakly-informative priors determined by $\Omega_\alpha = \Omega_\beta = 10 \times I_q$, $\sigma_\upsilon = 0.5$; $c = p + 2$, and $\Psi = I_p$, where q is the number of fixed effect variables, p is the number of random effect variables and I_p is the p dimensional identity matrix. In the prior on Σ, we choose c such that $E(\Sigma) = I_p$. Additionally, we perform a brief sensitivity analysis and find that our results are robust to these prior choices (see Web Appendix A).

3. Details of Markov chain Monte Carlo Sampling

As inference is not available in a closed-form, we use iterative MCMC sampling methods. To improve mixing through the sampling process, we introduce continuous latent variables to correspond with whether Y_{ij} is zero or positive. To that end, we can equivalently express our probit model through the conditionally independent latent variables $Z_{ij} \sim N\left(X_{ij}^T\beta + U_{ij}^T\delta_i, 1\right)$ where $Y_{ij} > 0$ if $Z_{ij} > 0$ and $Y_{ij} = 0$ if $Z_{ij} < 0$. This data augmentation scheme provides conjugate sampling in the presence model and speeds mixing for this component.[27]

The sampling algorithm iterates between the following steps.

1. Presence model latent variable Z_{ij}: If $Y_{ij} = 0$, then we sample Z_{ij} from a normal with mean $X_{ij}^T\beta + U_{ij}^T\delta_i$ and unit variance, truncated to support $(-\infty, 0]$. If $Y_{ij} > 0$, then we sample from the previous normal distribution truncated to $(0, \infty)$.
2. Presence model regression coefficient β and random effects δ: A naïve Gibbs sampler that samples from $p(z|\beta, \delta, \cdots)$, $p(\beta|z, \delta, \cdots)$, and $p(\delta|z, \beta, \cdots)$ will demonstrate high autocorrelation and slow mixing. To alleviate this, we jointly update β and δ through $p(\beta, \delta_i| z, \cdots) = p(\beta|z, \gamma) \times p(\delta_i|\beta, \gamma)$. This is a partially collapsed Gibbs sampler.[28]

   Note that the sampling distribution for β (marginal over the random effects δ) is

$$\beta|z,\gamma \sim MV\,N\left(\left[\left\{\sum_{i=1}^{N} X_i^T\left(I(n_i) + U_i\sum\nolimits_{\delta|\gamma} U_i^T\right)^{-1} X_i\right\} + \Omega_\beta^{-1}\right]^{-1} \times \left\{\sum_{i=1}^{N} X_i^T\left(I(n_i) + U_i\sum\nolimits_{\delta|\gamma} U_i^T\right)^{-1}\left(z_i - U_i\sum\nolimits_{\delta,\gamma}\sum\nolimits_{\gamma\gamma}^{-1}\gamma_i\right.\right.$$

   where $I(a)$ denotes the a-dimensional identity matrix and $\sum_{\delta|\gamma} = \sum_{\delta\delta} - \sum_{\delta,\gamma}\sum_{\gamma\gamma}^{-1}\sum_{\gamma,\delta}$ is the conditional variance of δ_i given γ_i.

   After updating β, we sample δ_i given β, Z, γ_i, and Σ for each i.

$$\delta_i|\beta,\gamma_i \sim MV\,N\left(\xi_i, \left\{\sum\nolimits_{\delta|\gamma}^{-1} + U_i^T U_i\right\}^{-1}\right),$$

   where $\xi_i = \left\{\sum_{\delta|\gamma}^{-1} + U_i^T U_i\right\}^{-1}\left(\sum_{\delta|\gamma}^{-1}\sum_{\delta,\gamma}\sum_{\gamma\gamma}^{-1}\gamma_i + U_i^T z_i - U_i^T X_i\beta\right)$.

3. 3. Severity model regression coefficients α: As conjugacy is not available for the CMP model, we use a pair of Metropolis-Hastings updates. We perform a global step that seeks to update the full α vector, as well as a local step that attempts to update each component of α one at a time.

For the global step, we propose the candidate value $\alpha^c \sim MVN(\alpha^t, \Omega_{q,\alpha})$, where α^t is the current value of α at the t^{th} iteration. For ease of explanation, we use P_{ij}^+ to denote the probability mass function of the truncated CMP distribution of Y_{ij} in (1) and Π to be the prior density. We accept the candidate value α^c with probability A_α where

$$A_\alpha = min\left(1, \left\{\prod_{i,j:y_{ij}>0} P_{ij}^+(y_{ij}|\alpha^c, \upsilon, \gamma_i)\right\}\prod(\alpha^c) \Big/ \left\{\prod_{i,j:y_{ij}>0} P_{ij}^+\left(y_{ij}|\alpha^t, \upsilon, \gamma_i\right)\right\}\prod\left(\alpha^t\right)\right);$$

otherwise, we remain at α^t.

For the local step that seeks to update the k^{th} element of the α, we propose α_k^c from $N\left(\alpha_k^t, \sigma_{q,\alpha_k}^2\right)$, and for all other components of α, we keep the current value, i.e., $\alpha_j^c = \alpha_j^t\ (j \neq k)$. We accept the proposed α^c with probability A_{α_k} where

$$A_{\alpha_k} = min\left(1, \frac{\left\{\prod_{i,j:y_{ij}>0} P_{ij}^+(y_{ij}|\alpha^c, \upsilon, \gamma_i)\right\}\prod(\alpha_k^c)}{\left\{\prod_{i,j:y_{ij}>0} P_{ij}^+\left(y_{ij}|\alpha^t, \upsilon, \gamma_i\right)\right\}\prod(\alpha_k^t)}\right).$$

Otherwise, we remain at the current value α^t. We repeat this for all components α_k.

4. Severity model dispersion parameter υ: We update υ using the Metropolis-Hastings algorithm. For the proposal distribution, we use a pseudo-random walk $\upsilon^c \sim LN\left(log\left(\upsilon^t\right), \sigma_{q,\upsilon}^2\right)$, and accept the move with probability A_υ,

$$A_\upsilon = min\left(1, \frac{\left\{\prod_{i,j:y_{ij}>0} P_{ij}^+(y_{ij}|\alpha, \upsilon^c, \gamma_i)\right\}\prod(\upsilon^c)\, q(\upsilon^t|\upsilon^c)}{\left\{\prod_{i,j:y_{ij}>0} P_{ij}^+\left(y_{ij}|\alpha, \upsilon^t, \gamma_i\right)\right\}\prod(\upsilon^t)\, q(\upsilon^c|\upsilon^t)}\right),$$

where $q(\upsilon^c|\upsilon^t)$ is the log-normal $LN\left(log\left(\upsilon^t\right), \sigma_{q,\upsilon}^2\right)$ proposal density.

5. Severity model random effects γ: As with the regression coefficients α, there is no conjugacy, and generally, the Metropolis-Hastings algorithm is needed to

update γ_i for each cluster i. However, in many situations this can be simplified by using only a MVN-Gibbs step for γ_i or a Metropolis step for part of γ_i with a MVN draw for the rest. This partially collapsed Gibbs step will lead to move efficient computation than a naive Metropolis-Hastings approach.

Recall that only the non-zero Y_{ij}s appear in the CMP part, and hence, only the γ_i that contribute to the distribution of these Y_{ij}s are identified by the CMP distribution. The sampling distribution for γ_i is proportional to

$$p\left(\gamma_i|\delta_i,\alpha,\upsilon,\sum,y_i\right)\propto p\left(\gamma_i,\delta_i|\sum\right)\prod_{j:y_{ij}>0}P_{ij}^{+}\left(y_{ij}|\alpha,\upsilon,\gamma_i\right).$$

In particular, for a cluster where all counts are zero, the data provide no information about γ_i, as the product over y_{ij} is empty. Hence, the sampling distribution for γ_i is

$p\left(\gamma_i|\delta_i.\sum\right)=N\left(\sum_{\gamma\delta}\sum_{\delta\delta}^{-1}\delta_i,\sum_{\gamma\gamma}-\sum_{\gamma\delta}\sum_{\delta\delta}^{-1}\sum_{\delta\gamma}\right)$, which we can draw from exactly.

More generally, we let $\tilde{\gamma}_i$ denote the portion of γ_i identified in $\prod_{j:y_{ij}>0}P_{ij}^{+}\left(y_{ij}|\alpha,\upsilon,\gamma_i\right)$, and γ_i^* be the unidentified portion of γ_i. We will update $\tilde{\gamma}_i$ by Metropolis-Hastings, marginalized over γ_i^*, and then, sample γ_i^* from MVN conditional on $\tilde{\gamma}_i$ and δ_i. To that end, we propose the candidate $\tilde{\gamma}_i^c\sim N\left(\tilde{\gamma}_i,\Omega_{\gamma,q}\right)$, with a random walk around the current value $\tilde{\gamma}_i$. We accept the move with probability

$$A_{\tilde{\gamma}_i}=min\left(1,\frac{\left\{\prod_{j:y_{ij}>0}P_{ij}^{+}\left(y_{ij}|\tilde{\gamma}_i^c,\alpha,\upsilon\right)\right\}\prod\left(\tilde{\gamma}_i^c,\delta_i|\sum\right)}{\left\{\prod_{j:y_{ij}>0}P_{ij}^{+}\left(y_{ij}|\tilde{\gamma}_i^t,\alpha,\upsilon\right)\right\}\prod\left(\tilde{\gamma}_i^t,\delta_i|\sum\right)}\right),$$

where $\Pi(\gamma_i, \delta_i|\Sigma)$ represents the density of the $MVN(0, \Sigma)$ random effects distribution after marginalizing over the unidentified components in γ_i^*. We then sample the remaining γ_i^* from the conditional $p\left(\gamma_i^*|\delta_i,\tilde{\gamma}_i\right)$ based on $MVN(0, \Sigma)$.

6. Random effects covariance matrix Σ : The covariance matrix can be updated through conjugacy:

$$\sum|\delta,\gamma\sim IW\left(c+N,\sum_{i=1}^{N}\begin{pmatrix}\delta_i\\ \gamma_i\end{pmatrix}\begin{pmatrix}\delta_i\\ \gamma_i\end{pmatrix}^T+\Psi\right).$$

Variance parameters in the Metropolis proposal distributions are chosen by trial and error so that acceptance rate is about 25% for multivariate steps and between

25% and 40% for univariate steps.[29] We further discuss the selection of these variance parameters in Web Appendix B.

4. Analysis of Iowa Fluoride Study

The Iowa Fluoride Study (IFS) is a longitudinal study which in 1991 began collecting dental examination data on children in Iowa starting at age 5. These children were followed up at 9, 13 and 17 years. In this paper, we focus on the 9-year-old children's dataset which is particularly interesting due to the mixed dentition (composed of some primary teeth and some permanent teeth, which does not pose any complications with our approach).

This dataset consists of caries experience scores (CES) as a response variable and seven potential risk/protective factors for caries. Each surface of a tooth is scored 0 (sound), 1 (non-cavitated), or 2 (cavitated) depending on the level of caries involvement, and the CES is the sum over the five surfaces of the tooth. Thus, the response variable takes integer values from 0 to 10. A larger CES indicates more severe decay on the tooth. The dataset includes $N = 464$ clusters (representing 464 children) with cluster size n_i lying between 16 and 24 for the number of teeth per person. Altogether, we have 10,838 observations, with 9,616 zero counts, which is almost 89% of the dataset (Figure 1). Thus, this dataset appears to be zero-inflated.

In the dental field, the location of a tooth inside the mouth is known to have a great effect on the likelihood of dental carries, or cavities. In fact, cavities are more likely to occur on the molars than non-molars (incisors, canines, and premolars) because molars have irregular occlusal surfaces which more easily retain food. Additionally, the mesial and distal surfaces between the teeth of molars also are more likely to retain bacteria and have increased caries risk. To account for this, both the presence and severity models include a covariate for non-molar (relative to molar). Each model includes eight covariates which presumably affect cavities or caries as protective/risk factors (Table 1).

To define the dependence structure across teeth, we again focus on the two types of teeth: molar and non-molar. We introduce the clustering effect by tooth location as we expect that teeth within each class are more closely related than those across classes. We parameterize this by letting δ_i be a length 2 vector, where the first component is the overall cluster effect (random intercept) and the second represents the effect of non-molar teeth (relative to molars). For the severity model, we only include a cluster-specific random effect (random intercept), as there are fewer responses to inform the severity model and its random components (less than 12% of observations have positive Y). In addition, most cavities occur on molars (more than 95%), so there would be insufficient information to identify a non-molar term in this severity model. In total, we have three random effects (δ_{i1}, δ_{i2}, γ_i).

As discussed in Section 2, the hyperparameters for the priors are chosen to be $\Omega_\alpha = \Omega_\beta = 10 \cdot I_{11}$, $c = 5$, $\Psi = I_3$, and $\sigma_\upsilon = 0.5$, yielding relatively disperse priors. The posterior samples of the model parameters are obtained by running the MCMC algorithm (Section 3) for 65,000 iterations. The samples are collected after first 25,000 burn-in iterations, yielding 40,000 samples to be used for inference.

MCMC convergence is assessed through trace plots and Geweke tests[30] for the individual parameters, as well as the log-likelihoods for the presence ($\log L_1$) and severity ($\log L_2$) models, where

$$L_1=\prod_{i=1}^{N}\prod_{j=1}^{n_i}P(y_{ij}=0)^{I(y_{ij}=0)}P(y_{ij}>0)^{I(y_{ij}>0)},\ L_2=\prod_{i=1}^{N}\prod_{j=1}^{n_i}P\left(Y_{ij}=y_{ij}|y_{ij}>0\right).$$

The trace plots indicate adequate mixing (Figure 1 in Web Appendix B) and most parameters pass the Geweke test (Table 2 in Web Appendix B). The effective sample sizes[31] for the log-likelihood functions of the presence and severity models are both found to be greater than 1,000. We also analyze the same dataset with a hurdle mixed Poisson model (i.e., fixing $\upsilon = 1$) to compare with our hurdle CMP model. The posterior means and 95% equal-tailed credible intervals are given in Table 2.

Coinciding with our expectations, we find the molar/non-molar effect to be highly impactful in both the presence and severity models with both the CMP and Poisson frameworks. Non-molars are much less likely to develop caries, and of those teeth that do develop caries, they tend to have lower scores relative to a corresponding molar.

Considering predictors whose credible intervals (CI) exclude zero to be important factors, we find the same set of important covariates in the presence (zero) model for the CMP and the Poisson choices. This is reasonable and expected as the differences in the two models occur in the distribution of the positive counts. Daily fluoride intake (*FlIntake*) and tooth brushing frequency (*ToothBrush*) are protective factors, while greater soda pop intake (*SodaPop*) increases the risk of developing caries.

In the severity model of the CMP and Poisson frameworks, tooth brushing frequency (*ToothBrush*) is predictive of less caries. In the hurdle mixed CMP there is evidence that professional fluoride treatment (*FlTrt*) is associated with higher CES score but the effect size is small; in the Poisson model the CI contains zero. The estimated CMP dispersion parameter is $\hat{\upsilon}=0.888$ with $P(\upsilon < 1|y) = 0.969$, indicating the (positive) CES scores are overdispersed.

To evaluate the resulting dependence structure of our model, we consider the posterior mean of the random effect covariance matrix:

$$\hat{\sum}_{\text{CMP}}=\begin{bmatrix} 0.937 & -0.195 & 0.258 \\ -0.222 & 0.830 & -0.063 \\ 0.565 & -0.147 & 0.222 \end{bmatrix}, \quad (4)$$

and

$$\hat{\sum}_{\text{POISSON}} = \begin{bmatrix} 0.936 & -0.219 & 0.277 \\ -0.262 & 0.749 & -0.082 \\ 0.558 & -0.185 & 0.263 \end{bmatrix}, \quad (5)$$

where the variance-covariance components are displayed in the upper triangular part and the correlation coefficients in the lower off-diagonal part. Based on the CMP-estimated $\hat{\sum}$, we can find the correlations between the latent variables Z to describe the dependence within cluster. The correlation between the Z_is of two molars is 0.484, between two non-molars is 0.579, and between a molar and non-molar is 0.346. Clearly, teeth within a location class are more highly correlated than across class, but there remains positive correlation across all observations in the cluster.

$Var(\gamma_i)$, is the variance component corresponding to the severity model in Equation 5, roughly 18% larger than the corresponding component in the CMP model. In addition to clustering, random effects are used in GLMs to induce overdispersion. Here, $Var(\gamma_i)$ the variance component is inflated to compensate for the overdispersion that the restrictive Poisson model cannot explain. Thus, the Poisson choice conflates overdispersion and the clustering effect. That is, it may overstate the correlation between teeth to account for true overdispersion. However, the mixed CMP is flexible enough to distinguish between the contributions of these two.

To determine the sensitivity of our conclusions to our prior choice, we perform a brief sensitivity study in Web Appendix A. We consider a more informative, strong prior choice using the hyperparameters: $\Omega_\alpha = \Omega_\beta = 1 \times I_{11}$ with $c = 25$, $\Psi = 5 \times I_3$, and $\sigma_\upsilon = 0.2$. Under this choice of prior on Σ, we have $\left[E\left(\sum^{-1}\right)\right]^{-1} = \frac{1}{5} \times I$ as in the original prior. Additionally, we use a less-informative, weak prior using the values: $\Omega_\alpha = \Omega_\beta = 100 \times I_{11}$, and $\sigma_\upsilon = 0.8$. Here, we use an improper prior for Σ, $\Pi(\Sigma) \propto |\Sigma|^{\{-(p+1)/2\}}$. We find our parameter estimates and conclusions are consistent across these choices. Detailed results are contained in Web Appendix A.

5. Simulation Studies

To validate our method and better understand its operating characteristics, we consider two simulation experiments, one with overdispersion ($\upsilon = 0.7$) and the other with equidispersion ($\upsilon = 1$, leading to a Poisson distribution). For each setting, we consider $N = 200$ clusters with cluster size of $n_i = 20$. Similar to the role of molars and non-molars, each cluster contains two classes with 10 observations each. This setting enables us to compare our hurdle CMP mixed model and the hurdle Poisson mixed model. The simulation set-up of this section is guided by the Iowa data for the nine-year-old children.

For each scenario, we use the same design matrices for fixed effects in both binary and positive count parts. The design matrix for the fixed effects consists of four different covariates with an intercept. The first two are the intercept and a binary class indicator, 0 for

the first 10 observations and 1 for the rest. Then, we consider three continuous factors corresponding to *FlIntake*, *SodaPop*, and *ToothBrush* chosen by sampling with replacement from the IFS data, respectively.

For the dependence, we take the first two columns from the fixed effects design matrix as random effects design matrix in both the binary and positive count parts. Hence, the random effect is (δ_{i1}, δ_{i2}, γ_{i1}, γ_{i2}), corresponding to an intercept and binary class effect in the presence model, and an intercept and binary class effect in the severity model. The true parameter values can be found in Tables 3–6.

We generate 200 simulated datasets under the two choices of υ, 0.7 and 1.0. For each data set, we run MCMC to obtain 50,000 posterior iterations to use for inference. When the data are drawn from CMP ($\upsilon = 0.7$), we first sample 5,000 burn-in iterations (55,000 total), and when the true data are conditionally Poisson ($\upsilon = 1.0$), the MCMC chain takes longer to reach the stationary distribution and 15,000 burn-in iterations are used (65,000 total). Based on these samples over 200 datasets, bias and mean squared error (MSE) are calculated for all the fixed effect estimators and individual variance components for the random effects. We also calculate the sum of squared errors (SSE) for a block of the random effect variance matrix by summing the MSE over the appropriate (i, j) parameters corresponding to the relevant block of Σ.

For the overdispersion case ($\upsilon = 0.7$), estimation of the β parameters in the binary component from the CMP model is similar to estimation in the Poisson model (Table 3). As noted in the previous section, this is expected, as the models are the same for this component. However, the positive part estimators α behave noticeably different for those two models. Table 3 shows that the CMP model estimates have smaller bias and smaller MSE than those estimators from the Poisson model.

As we use the 95% credible intervals to determine important factors, we also consider the coverage rate (CR) in Table 3. As we use only 200 datasets to evaluate these proportions, all are found to be within the margin of error of the nominal 0.95 rate with one exception (the theoretical Monte Carlo standard error is 0.015). In the hurdle Poisson model, the intercept α_0 displays very poor coverage due to the bias from model misspecification.

Estimation of the variance components is assessed in Table 4. As with the regression coefficients, estimation of the upper block of Σ corresponding to δ (σ_{11}, σ_{12}, σ_{22}), the random effects in the binary model, are roughly equivalent between the CMP and Poisson models; SSE in this block is 0.0416 for the CMP model and 0.0415 for the Poisson. However, differences between the models are apparent in the γ block of Σ (σ_{33}, σ_{34}, σ_{44}) describing the positive count model. SSE for this block is 0.0375 for CMP, compared to 0.0545 for Poisson. In particular, the variance terms for γ tend to be biased high in the hurdle Poisson model to recover overdispersion, as discussed in the previous section. Overall, we find the hurdle CMP mixed effects model performs significantly better for zero-inflated, dispersed data than the simpler Poisson choice.

The $\upsilon = 1$ scenario is also considered, and the results are given in Tables 5 and 6. Overall, estimation of both the fixed effects and random effects variances perform similarly in both

the CMP and Poisson models. We do see slightly inflated MSE of the regression coefficients in the severity component under the hurdle CMP model, but the difference is minor and the bias is still negligible. The coverage rates (CR) of 95% CI are also close to the nominal rate. In addition, a low bias in υ indicates the flexibility of our model. Thus, we conclude that even for $\upsilon = 1$ case, the hurdle CMP mixed effects model is comparable to the hurdle Poisson mixed effects model.

In conclusion, our simulation studies demonstrate that the hurdle CMP mixed effects model fits significantly better for dispersed data and is comparable to the hurdle Poisson mixed effects model, when the data are equidispersed.

As suggested by a reviewer, we conducted another simulation study where data have no zero inflation to see how our hurdle mixed CMP model behaves compared with the true model, a mixed effects CMP. It turns out that most of all the corresponding estimators are reasonably close to the true values even under the no zero-inflation setting. For full detail, see details in Web Appendix C.

6. Discussion

We have proposed a new Bayesian approach for modeling dependent, zero-inflated count data by combining a probit component with the Conway-Maxwell-Poisson regression using shared random effects. Our approach is flexible in terms of allowing various types of dispersions from under- to over-dispersion (unlike negative binomial and many other count models), and the structure of the dependence across counts is adaptable to many correlation forms. Most frequentist approaches to this problem experience difficulties in approximating the likelihood with Laplacian/quadrature methods. However, our Bayesian approach avoids this by relying on an iterative sampling scheme that draws the values of the random effects each iteration. Consequently, such a sampling scheme-based framework gains the flexibility to choose a more versatile form of the random effects design matrix. The data application and simulation studies provide clear-cut evidence that our approach is superior to a more standard, random effects, hurdle Poisson model.

While the MCMC scheme we describe in Section 3 is found to perform successfully in the experiments we consider, one of the outstanding challenges is further improving and speeding up the computation. By using collapsed Gibbs steps, our sampler is designed to minimize auto-correlation when possible, but as with any Gibbs sampler in a complex modeling framework, mixing can be slow. One relevant contributor is that evaluation of the mass function requires the normalizing constant $\mathscr{Z}(\lambda, \upsilon)$, which is an infinite sum with no closed form representation; Gillispie and Christopher[32] suggest an approximation of $\mathscr{Z}$, but it only works under certain ranges of the parameters, λ and υ. As all of the observed counts are less than or equal to 10, we instead choose to truncate the sum at 100 for computation, providing a high level of accuracy since $P(Y > 100)$ is negligible under the υ and λ_{ij}s we encounter.

In the analysis of the IFS data, we include a tooth-location factor in the model to control the dependence across teeth. Another possibility might be to consider a spatial structure to

define the dependence across teeth, perhaps along the lines of some previous work.[17,19] Intuitively, we may expect that adjacent teeth are highly correlated, and that the dependence decays the farther away the teeth are located. In principle, our model can handle this approach by defining a γ_{ij} and δ_{ij} term for every tooth, and the full vector is drawn from MVN with covariance matrix Σ_i that provides a (low-dimension) spatial structure based on tooth location. In practice, there are some challenges. First, as our data is so sparse (almost 89% zeros), we have very little information regarding the random effects for the count components. Secondly, the expansion of the random effect structure will require greater MCMC computational time. The continued use the partially collapsed steps for δ (see Step 5 in Section 3) may help manage this issue. Third, in the 9-year-old children's dataset, there is a mix of primary and permanent teeth, and their locations do not exact align. Specification of the dependence for these terms will require additional considerations. Thus, with respect to these challenges, the random effects structure we chose is a more reasonable and flexible choice than equicorrelation as typically used. Additionally, we only focus on the data of the 9-year-old children. Further methodology can be developed to leverage the longitudinal information across the several study visits to understand the factors leading to improving or declining dental health across childhood and adolescence. We leave these as possibilities for future extensions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by National Institutes of Health grants 1R03DE020839-01A1, 1R03DE022538-01, R01-DE09551, R01-DE12101, and M01-RR00059. We thank two anonymous reviewers for their constructive comments.

References

1. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992; 34(1):1–14.
2. Greene, WH. Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models. NYU Working Paper No EC-94-10. 1994. https://ssrn.com/abstract=1293115
3. Böhning D. Zero-inflated Poisson models and C.A.MAN: a tutorial collection of evidence. Biometrical Journal. 1998; 40(7):833–843.
4. Long DL, Preisser JS, Herring AH, Golin CE. A marginalized zero-inflated Poisson regression model with overall exposure effects. Statistics in Medicine. 2014; 33(29):5151–5165. [PubMed: 25220537]
5. Rose CE, Martin SW, Wannemuehler KA, Plikaytis BD. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. Journal of Biopharmaceutical Statistics. 2006; 16(4):463–481. [PubMed: 16892908]
6. Zuur, A., Ieno, EN., Walker, N., Saveliev, AA., Smith, GM. Mixed Effects Models and Extensions in Ecology with R. Springer; New York: 2009. Zero-truncated and zero-inflated models for count data; p. 261-293.
7. Hu M, Pavlicova M, Nunes EV. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. American Journal of Drug and Alcohol Abuse. 2011; 37(5):367–375. [PubMed: 21854279]
8. Rosen O, Jiang W, Tanner MA. Mixtures of marginal models. Biometrika. 2000; 87(2):391–404.

9. Hall DB, Zhang Z. Marginal models for zero-inflated clustered data. Statistical Modelling. 2004; 4(3):161–180.
10. Lee K, Joo Y, Song JJ, Harper D. Analysis of longitudinal zero-inflated count data using marginalized models. Computational Statistics & Data Analysis. 2011; 55:824–837.
11. Iddi S, Molenberghs G. A marginalized model for zero-inflated, overdispersed and correlated count data. Electronic Journal of Applied Statistical Analysis. 2013; 6(2):149–165.
12. Kong M, Xu S, Levy SM, Datta S. GEE type inference for clustered zero-inflated negative binomial regression with application to dental caries. Computational Statistics and Data Analysis. 2015; 85:54–66. [PubMed: 25620827]
13. Hall DB. Zero-inflated Poisson and binomial regression with random effects: A case study. Biometrics. 2000; 56(4):1030–1039. [PubMed: 11129458]
14. Yau KKW, Wang K, Lee AH. Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. Biometrics. 2003; 45(4):437–452.
15. Rodrigue-Motta M, Gianola D, Heringstad B. A mixed effects model for overdispersed zero inflated Poisson data with an application in animal breeding. Journal of Data Science. 2010; 8(3): 379–396.
16. Fulton KA, Liu D, Haynie DL, Albert PS. Mixed model and estimating equation approaches for zero inflation in clustered binary response data with application to a dating violence study. Annals of Applied Statistics. 2015; 9(1):275–299. [PubMed: 26937263]
17. Aktekin T, Musal M. Analysis of income inequality measures on Human Immunode-Ficiency Virus mortality: A spatiotemporal Bayesian perspective. Journal of the Royal Statistical Society: Series A. 2015; 178(2):383–403.
18. Long DL, Preisser JS, Herring AH, Golin CE. A marginalized zero-inflated Poisson regression model with random effects. Journal of the Royal Statistical Society: Series C. 2015; 64(5):815–830.
19. Musal M, Aktekin T. Bayesian spatial modeling of HIV mortality via zero-inflated Poisson models. Statistics in Medicine. 2013; 32(2):267–281. [PubMed: 22807006]
20. Conway RW, Maxwell WL. A queuing model with state dependent service rates. Journal of Industrial Engineering. 1962; 12(3):132–136.
21. Barriga GDC, Louzada F. The zero-inflated Conway–Maxwell–Poisson distribution: Bayesian inference, regression modeling and influence diagnostic. Statistical Methodology. 2014; 21:23–34.
22. Choo-Wosoba, H., Datta, S. Analyzing clustered count data with a cluster-specific random effect zero-inflated Conway–Maxwell–Poisson distribution; Journal of Applied Statistics. 2017. p. 1-16.http://dx.doi.org/10.1080/02664763.2017.1312299
23. Lee Y, Nelder JA. Likelihood for random-effects (with discussion). Statistical and Operational Research Transactions. 2005; 29:141–182.
24. Levy SM, Warren JJ, Broffitt BA, Hillis SL, Kanellis MJ. Fluoride, beverages and dental caries in the primary dentition. Caries Research. 2003; 37(3):157–165. [PubMed: 12740537]
25. Choo-Wosoba H, Levy SM, Datta S. Marginal regression models for clustered count data based on zero-inflated Conway–Maxwell–Poisson distribution with applications. Biometrics. 2016; 72(2): 606–618. [PubMed: 26575079]
26. Sellers KF, Shmueli G. A flexible regression model for count data. Annals of Applied Statistics. 2010; 4(2):943–961.
27. Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. American Statistician. 1995; 49(4):327–335.
28. Van Dyk DA, Park T. Partially collapsed Gibbs samplers: Theory and methods. Journal of the American Statistical Association. 2008; 103(482):790–6.
29. Robert, CP. Discretization and MCMC convergence assessment. Springer; New York: 1998.
30. Geweke, J. Bayesian Statistics 4. Clarendon Press; Oxford, UK: 1992. Evaluating the accuracy of sampling-based approached to calculating posterior moment.
31. Gelman, A., Carlin, JB., Stern, HS., Dunson, DB., Vehtari, A., Rubin, DB. Bayesian Data Analysis. Third. Chapman and Hall/CRC; 2013.

32. Gillispie SB, Christopher GC. Approximating the Conway-Maxwell-Poisson distribution normalization constant. Statistics. 2015; 49(5):1062–1073.

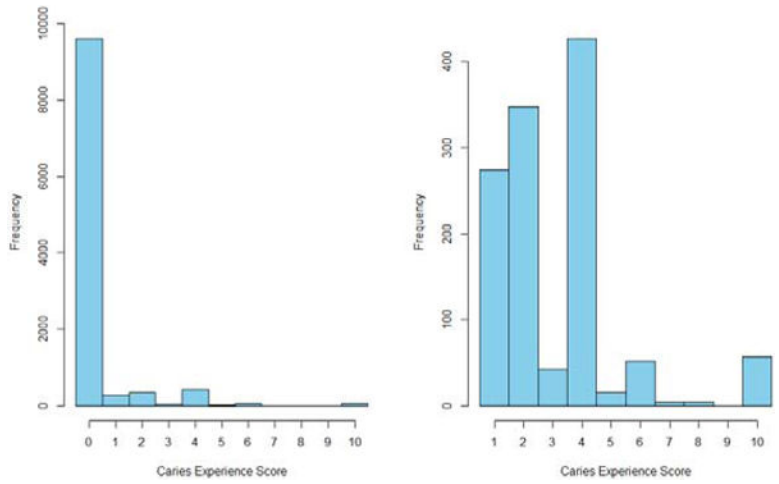

Figure 1.

Frequency histograms of the entire caries experience scores (CES) (left panel) and the non-zero CES(right panel) of the nine-year-old children's dataset from the Iowa Fluoride Study.

Table 1

Description of potential risk/protective factors from the Iowa Fluoride Study (IFS)

Non-molar	Non-molar effect; non-molar is coded as 1.
Sex	Sex of the child; male is coded as 1.
ExamAge	Age in years at the time of the dental examination (centered at 9 years).
FlIntake	Daily fluoride intake (mg) from water, other beverages and selected foods, ingested dentifrice and fluoride supplements. Computed by AUC trapezoidal method from all available data within the time span 5 to 9 years.
SodaPop	Daily soda pop intake (oz.) computed with AUC trapezoidal method using all available data within the time span 5 to 9 years.
ToothBrush	Average of all tooth brushing frequencies reported for the period 5 to 9 years.
DentalVisit	Proportion of times a dental visit was reported with each individual point assessing the previous 6 months.
FlTrt	Average proportion of times a professional dental fluoride treatment was received with each individual point assessing the previous 6 months.
FlHome	Average home tap water fluoride level for all returned questionnaires for the period 5 to 9 years.

Table 2

Posterior means and 95% credible intervals (CI) for both presence and severity models as applied to the Iowa data

	hurdle mixed CMP		**hurdle mixed Poisson**	
	Presence Model			
	posterior mean	CI	posterior mean	CI
Intercept	−0.555	(−1.130, 0.015)	−0.560	(−1.124, −0.007)
Non-molars	−2.608	(−3.314, −2.082)	−2.525	(−3.133, −2.026)
Sex	−0.191	(−0.404, 0.019)	−0.186	(−0.397, 0.024)
ExamAge	0.132	(−0.017, 0.283)	0.133	(−0.017, 0.281)
FlIntake	−0.423	(−0.774, −0.074)	−0.417	(−0.764, −0.080)
SodaPop	0.073	(0.029, 0.117)	0.074	(0.029, 0.118)
ToothBrush	−0.566	(−0.796, −0.339)	−0.567	(−0.796, −0.342)
DentalVisit	0.222	(−0.342, 0.785)	0.219	(−0.330, 0.775)
FlTrt	0.334	(−0.039, 0.710)	0.340	(−0.027, 0.710)
FlHome	0.122	(−0.128, 0.378)	0.116	(−0.131, 0.363)
	Severity Model			
	posterior mean	CI	posterior mean	CI
Intercept	0.904	(0.422, 1.406)	1.038	(0.560, 1.453)
Non-molars	−0.627	(−0.897, −0.373)	−0.668	(−0.951, −0.400)
Sex	−0.102	(−0.240, 0.039)	−0.108	(−0.257, 0.050)
ExamAge	0.103	(−0.001, 0.207)	0.112	(−0.001, 0.230)
FlIntake	−0.104	(−0.343, 0.136)	−0.115	(−0.359, 0.131)
SodaPop	0.015	(−0.014, 0.044)	0.018	(−0.014, 0.050)
ToothBrush	−0.189	(−0.360, −0.036)	−0.202	(−0.382, −0.024)
DentalVisit	−0.071	(−0.474, 0.357)	−0.058	(−0.457, 0.344)
FlTrt	0.259	(0.002, 0.554)	0.272	(−0.004, 0.538)
FlHome	−0.117	(−0.308, 0.064)	−0.123	(−0.306, 0.051)
υ	0.888	(0.772, 1.005)	1	

Table 3

Summary of fixed effect estimation in a simulation study described in Section 5; here the model is overdispersed, $\upsilon = 0.7$.

		hurdle mixed CMP			hurdle mixed Poisson		
	True	**Bias**	**MSE**	**CR**	**Bias**	**MSE**	**CR**
$\hat{\beta}_0$	−0.40	0.0178	0.1141	0.975	0.0188	0.1139	0.975
$\hat{\beta}_1$	−1.00	−0.0014	0.0136	0.940	−0.0017	0.0135	0.940
$\hat{\beta}_2$	−0.20	−0.0036	0.0429	0.955	−0.0042	0.0428	0.960
$\hat{\beta}_3$	0.20	−0.0002	0.0009	0.970	<0.0001	0.0009	0.970
$\hat{\beta}_4$	−0.60	−0.0122	0.0309	0.945	−0.0129	0.0309	0.930
$\hat{\alpha}_0$	0.20	0.0697	0.1032	0.955	0.3555	0.2127	0.635
$\hat{\alpha}_1$	−2.00	−0.0157	0.1116	0.980	−0.1121	0.1245	0.965
$\hat{\alpha}_2$	−0.20	−0.0151	0.0308	0.945	−0.0407	0.0390	0.925
$\hat{\alpha}_3$	0.05	0.0005	0.0006	0.965	0.0069	0.0008	0.945
$\hat{\alpha}_4$	−0.30	−0.0186	0.0212	0.955	−0.0557	0.0278	0.930
$\hat{\upsilon}$	0.70	0.0490	0.0234	0.940	0.3000	N/A	N/A

Table 4

Summary of variance component estimation in a simulation study described in Section 5; here the model is overdispersed, $\upsilon = 0.7$.

		hurdle CMP		hurdle Poisson	
	True	**Bias**	**MSE**	**Bias**	**MSE**
$\hat{\sigma}_{11}$	0.80	−0.0063	0.0141	−0.0066	0.0142
$\hat{\sigma}_{12}$	−0.20	0.0172	0.0087	0.0180	0.0087
$\hat{\sigma}_{13}$	0.20	0.0095	0.0044	0.0400	0.0072
$\hat{\sigma}_{14}$	0.20	−0.0425	0.0182	−0.0566	0.0208
$\hat{\sigma}_{22}$	0.40	−0.0163	0.0101	−0.0171	0.0100
$\hat{\sigma}_{23}$	−0.10	−0.0002	0.0026	−0.0165	0.0037
$\hat{\sigma}_{24}$	−0.05	0.0043	0.0064	0.0158	0.0072
$\hat{\sigma}_{33}$	0.20	0.0206	0.0031	0.0862	0.0115
$\hat{\sigma}_{34}$	0.10	−0.0324	0.0046	−0.0390	0.0073
$\hat{\sigma}_{44}$	0.20	0.1220	0.0252	0.1327	0.0284

Table 5

Summary of fixed effect estimation in a simulation study described in Section 5; here the model is equidispersed, $\upsilon = 1$.

		hurdle mixed CMP			hurdle mixed Poisson		
	True	**Bias**	**MSE**	**CR**	**Bias**	**MSE**	**CR**
$\hat{\beta}_0$	−0.40	0.0242	0.1249	0.940	0.0250	0.1246	0.935
$\hat{\beta}_1$	−1.00	−0.0041	0.0134	0.960	−0.0041	0.0134	0.950
$\hat{\beta}_2$	−0.20	−0.0116	0.0442	0.940	−0.0116	0.0442	0.950
$\hat{\beta}_3$	0.20	−0.0009	0.0009	0.965	−0.0009	0.0009	0.965
$\hat{\beta}_4$	−0.60	−0.0065	0.0308	0.945	−0.0068	0.0308	0.945
$\hat{\alpha}_0$	0.20	0.0344	0.1504	0.940	−0.0064	0.0982	0.930
$\hat{\alpha}_1$	−2.00	−0.0118	0.1401	0.960	0.0014	0.1330	0.965
$\hat{\alpha}_2$	−0.20	−0.0101	0.0360	0.950	−0.0080	0.0348	0.950
$\hat{\alpha}_3$	0.05	0.0022	0.0007	0.965	0.0015	0.0007	0.975
$\hat{\alpha}_4$	−0.30	−0.0051	0.0271	0.945	−0.0027	0.0264	0.935
$\hat{\upsilon}$	1.00	0.0461	0.0548	0.955	0.0000	N/A	N/A

Table 6

Summary of variance component estimation in simulation study described in Section 5; here the model is equidispersed, $\upsilon = 1$.

		hurdle mixed CMP		hurdle mixed Poisson	
	True	**Bias**	**MSE**	**Bias**	**MSE**
$\hat{\sigma}_{11}$	0.80	−0.0171	0.0159	−0.0172	0.0159
$\hat{\sigma}_{12}$	−0.20	0.0245	0.0109	0.0245	0.0108
$\hat{\sigma}_{13}$	0.20	−0.0038	0.0051	−0.0070	0.0049
$\hat{\sigma}_{14}$	0.20	−0.0657	0.0262	−0.0658	0.0260
$\hat{\sigma}_{22}$	0.40	−0.0146	0.0120	−0.0147	0.0120
$\hat{\sigma}_{23}$	−0.10	0.0023	0.0032	0.0042	0.0031
$\hat{\sigma}_{24}$	−0.05	0.0144	0.0062	0.0131	0.0061
$\hat{\sigma}_{33}$	0.20	0.0124	0.0034	0.0042	0.0025
$\hat{\sigma}_{34}$	0.10	−0.0468	0.0066	−0.0463	0.0062
$\hat{\sigma}_{44}$	0.20	0.1483	0.0363	0.1444	0.0345