Analysis of Zero-Inflated Clustered Count Data: a Marginalized Model Approach

Keunbaik Lee^a, Yongsung Joo^b, Joon Jin Song^c^{*}, Dee Wood Harper^d

^aBiostatistics Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA 70112, U.S.A.

^bStatistics Department, Dongguk University-Seoul, Seoul 100-715, Korea

^cDepartment of Mathematical Sciences, University of Arkansas, Fayetteville, AR 72701, U. S. A.

^dDepartment of Criminal Justice, Loyola University, New Orleans, LA 70118, U. S. A.

Min and Agresti (2005) proposed random effect hurdle models for zero-inflated clustered count data with two-part random effects for a binary component and a truncated count component. In this paper, we propose new marginalized models for zero-inflated clustered count data using random effects. The marginalized models are similar to Dobbie and Welsh's (2001) model in which generalized estimating equations were exploited to find estimates. However, our proposed models are based on likelihood-based approach. Quasi-Newton algorithm is developed for estimation. We use these methods to carefully analyze two real datasets.

Keywords: Hurdle models; ZIP models; Random effects; Quasi-Newton.

^{*}Corresponding author. Tel.:+1 479 575-6319; fax: +1 479 575-8630. E-mail address: jjsong@uark.edu

1. Introduction

Count data with excess zeros are often encountered in a wide range of applications including medical, public health and social studies, particularly when the event of interest is rare. This type of count data is typically assumed to be from two-component models, which are often built using mixture and conditional models. The former is well known as the zero-inflated Poisson (ZIP) model (Lambert, 1992), mixing a discrete point mass and a Poisson distribution. The latter is a two-part conditional model, known as hurdle model (Mullahy, 1986), using a zero mass, so-called 'hurdle', and truncated Poisson distribution. The ZIP regression model is mixtures of logistic and Poisson regression models where the logistic portion contributes to the probability of a count of zero and the Poisson portion contributes to the frequency of utilization conditional on use. The hurdle model is also a two-part mixture model. One part is a binary model for whether the response outcome is zero or positive. The second part uses a truncated Poisson model for positive outcome.

Although the ZIP model is commonly used, the ZIP model is suitable only for handling zero inflation. However, when a data set is zero deflated at a level of a factor, the estimate of the corresponding parameter in the binary part of the ZIP model can be infinity (Min and Agresti, 2005). In addition, the ZIP model is more complex to fit than the hurdle model. In contrast, the hurdle model is suitable for modeling both zero inflation and zero deflation (Min and Agresti, 2005). In this paper, we propose new models based on the idea of hurdle model using a marginalized model approach in order to analyze clustered count data with extra zeros.

To analyze clustered count data with extra zeros, high positive correlation between responses must be considered. In this situation, random effects are commonly used to explain the within-subject dependence. Hall (2000) proposed zero-inflated Poisson models with random effects to account for the within-subject dependence in the Poisson state. However, the random effect was not used for the part of the model for the zero inflation. Yau and Lee (2001) used random effects to explain the within-subject dependence in both components in the hurdle model. However, the random effects in both components are independent. Min and Agresti (2005) extended Yau and Lee's (2001) model for repeated zero-inflated count data with two-part random effects for a binary component and a truncated count component and considered correlated random effects models unlike the models in Yau and Lee (2001). However, the models using random effects cannot have marginal mean directly which can be of interest in longitudinal or clustered data. As an alternative to the inclusion of random effects, several authors considered marginal models for countable data with extra zeros. The approach is to incorporate generalized estimating equations (GEEs) with a dependence working correlation matrix into the fitting algorithm (Dobbie and Welsh, 2001). Therefore, Dobbie and Welsh (2001) have marginal relationship between mean of response and covariates directly. Hall and Zhang (2001) proposed marginal model with generalized estimating equations (GEE) approach using EM algorithm. However, the GEE approach cannot be used directly under missing at random (MAR) which is common in longitudinal studies and semiparametric GEE approaches require explicit specification of the missing data mechanism (mdm). In addition, the re-weighting based approaches (based on the mdm) to handle MAR in GEEs only 'impute' missing values at the observed data points. Likelihood based approaches

do not have this restriction. In this paper, we propose a marginalized random effects model (MREM) which is one of the likelihood-based approaches and has the advantages of random effects and marginal models.

The MREMs were proposed by Heagerty (1999). The MREMs are commonly used to analyze longitudinal categorical data. Important features of these models are that marginal means are modeled directly and that the correlation among responses from same subject is accounted using random effects (Heagerty, 1999; Lee and Daniels, 2008; Lee et al., 2009; Lee et al., 2010). There are several advantages of the MREMs. First, the interpretation of regression coefficients does not depend on specification of the dependence in the model unlike in conditional models. Second, they can be much less susceptible to bias resulting from random effects model mis-specification (Heagerty and Zeger, 2000; Heagerty and Kurland, 2001; Lee and Daniels, 2008; Lee et al., 2009; Lee et al., 2010). Third, likelihood based inference is valid under ignorability and the missing data mechanism need not be explicitly specified. The use of a likelihood based approach will have advantages for longitudinal data that is missing at random (MAR) and in particular, ignorable (Lee and Daniels, 2008; Lee et al., 2009).

We propose likelihood-based marginalized models using the idea of MREMs for clustered count data with excess zeros which have marginal relationship of mean of response and covariates as in Dobbie and Welsh (2001). They also explain clustered dependence of responses using random effects as in Min and Agresti (2005). The proposed models are a useful class of models for zero-inflated clustered count data. The nonzero counts necessarily follow a zero-truncated Poisson distribution. In practice, nonzero count data are often overdispersed and alternative distributions, such as the zero-truncated negative binomial distribution, may be more appropriate more than Poisson hurdle model (Ridout et al., 2001). Thus, we also consider a zero-inflated negative binomial hurdle model.

The paper is organized as follows. In Section 2, we propose marginalized Poisson hurdle (MPH) and marginalized negative binomial hurdle (MNBH) models for the zero-inflated clustered count data. In Section 3, we conduct a simulation study to examine bias and efficiency of estimation of marginal mean parameters. In Section 4, we will analyze two real data sets using our proposed models. Finally, a brief summary is included in Section 5.

2. Proposed Models

Denote the response vector for the *i*th subject by $y_i = (y_{i1}, \dots, y_{in_i})^T$ where y_{it} is count response at time *t* for $t = 1, \dots, n_i$. We assume that y_{it} is conditionally independent given $b_i = (b_{i1}, b_{i2})^T$ and the responses on different subjects are independent. Let x_{it} be covariates corresponding to y_{it} .

2.1. Marginalized Poisson Hurdle Models

As we described in Section 1, we first develop the marginalized Poisson hurdle (MPH) model to accommodate clustered count data with excess zeros. We assume the marginal

probability of response for countable responses is given by

$$P(y_{it}|x_{it}) = \begin{cases} 1 - p_{it}^{M}, & \text{if } y_{it} = 0; \\ p_{it}^{M} \frac{g(y_{it};\lambda_{it}^{M})}{\left(1 - e^{-\lambda_{it}^{M}}\right)}, & \text{if } y_{it} = 1, 2, \cdots, \end{cases}$$

where

$$logit p_{it}^{M} = x_{it}^{T} \gamma,$$

$$g(y_{it}; \lambda_{it}^{M}) = \frac{e^{-\lambda_{it}^{M}} \left(\lambda_{it}^{M}\right)^{y_{it}}}{y_{it}!}$$

$$\lambda_{it}^{M} = \exp(x_{it}^{T} \beta),$$
(1)
(2)

 β and γ are $p \times 1$ dimensional unknown marginal parameter vectors, x_{it} is $p \times 1$ dimensional vector of covariates. To explain clustered association of responses we now use random effects from the idea of the MREMs (Heagerty, 1999). We assume a conditional hurdle model corresponding to (1) and (2). Given $b_i = (b_{i1}, b_{i2})^T$,

$$P(y_{it}; b_i) = \begin{cases} 1 - p_{it}^c(b_{i1}), & \text{if } y_{it} = 0; \\ p_{it}^c(b_{i1}) \frac{g(y_{it}; \lambda_{it}(b_{i2}))}{(1 - e^{-\lambda_{it}^c(b_{i2})})}, & \text{if } y_{it} = 1, 2, \cdots, \end{cases}$$
(3)

where

$$logit p_{it}^{c}(b_{i1}) = \Delta_{it1} + z_{it1}^{T} b_{i1},$$

$$e^{-\lambda_{it}^{c}(b_{i2})} \lambda_{it}^{c}(b_{i2})^{y_{it}}$$
(4)

$$g(y_{it}; \lambda_{it}^{c}(b_{i2})) = \frac{u(z_{i2})}{y_{it}!}$$
$$\log \lambda^{c}(b_{i2}) = \Delta_{it2} + z_{it2}^{T} b_{i2},$$
(5)

$$b_{i} = \begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} N(0, \Sigma), \qquad (6)$$

with z_{it1} and z_{it2} being subsets of x_{it} , b_{i1} and b_{i2} being corresponding random vectors, and

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12} & \Sigma_2 \end{pmatrix}.$$
(7)

where Σ , Σ_1 , and Σ_2 are unknown positive-definite matrices. Note that Δ_{it1} and Δ_{it2} are the subject-specific intercept and function of both the marginal parameters (γ, β) and the dependence parameters (Σ) . Detailed description is given later. As the simple random intercept form of models is often adequate in practice, we only discuss the case with b_{i1} and b_{i2} being univariate and $z_{it1} = z_{it2} = 1$. Then, we have

$$\Sigma = \left(\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array}\right).$$

Note that the two random effects, b_{i1} and b_{i2} , are likely correlated and account for the within-subject dependence and Σ represents the covariance matrix between b_{i1} and b_{i2}

like Min and Agresti's (2005) model. From the marginal and conditional distributions, we respectively have the marginal and conditional expectations

$$E(Y_{it}|x_{it}) = p_{it}^{M} \frac{\lambda^{M}}{1 - e^{-\lambda^{M}}},$$

$$E(Y_{it}|b_{i}) = p_{it}^{c}(b_{i1}) \frac{\lambda^{c}(b_{i2})}{1 - e^{-\lambda^{c}(b_{i2})}}.$$

 Δ_{it1} and Δ_{it2} in (4) and (5) are determined implicitly by γ and σ_1 , and β and Σ , respectively. The parameters Δ_{it1} are a function of both the marginal mean parameters, γ , in (4) and the random effects variances σ_1 , in (7) and can be obtained using the following identity

$$p_{it}^{M} = \int p_{it}^{c}(b_{i1})f(b_{i1})db_{i1},$$
(8)

where $f(b_{i1})$ is a univariate normal distribution with mean 0 and variance σ_1^2 . Similarly, Δ_{it2} in (5) is a function of γ , β , Σ in (1), (2), (4), and (5) and can be obtained using the following identity

$$E(Y_{it}|x_{it}) = E\{E(Y_{it}|b_i)\}, \Leftrightarrow \quad p_{it}^{M} \frac{\lambda_{it}^{M}}{1 - e^{-\lambda_{it}^{M}}} = \int p_{it}^{c}(b_{i1}) \frac{\lambda_{it}^{c}(b_{i2})}{1 - e^{-\lambda_{it}^{c}(b_{i2})}} f(b_{i}) db_{i},$$
(9)

where $f(b_i)$ is a bivariate normal distribution with mean 0 and variance $var(b_i)$. Given γ , β , and σ , Δ_{it1} and Δ_{it2} are deterministic functions of the parameters. Thus, we use (8) and (9) to solve Δ_{it1} and Δ_{it2} using using a Newton-Raphson. For the explicit forms of the terms in the Newton-Raphson algorithm, see the Appendix. This technique was also used in MREMs for longitudinal categorical data (Heagerty, 1999; Lee and Daniels, 2008; Lee et al., 2009).

The marginal mean models, (1) and (2), are modeled using logistic regression and Poisson regression, while clustered association are captured by the random effects in (4) and (5). In contrast to Min and Agresti's (2005) models, the regression parameters (γ and β) have marginal interpretations. Dobbie and Welsh (2001) also proposed marginal model using GEE approach. However, the GEE approach cannot be used directly under missing at random (MAR) which is common in longitudinal studies. Because our proposed models are combined models of Min and Agresti's (2005) and Dobbie and Welsh's (2001), there are several advantages of these models. First, clustered association is explained using conditional models while the marginal mean as a function of covariates still structures directly. As a result, the interpretation of the regression coefficients, γ and β , does not depend on the specification of the dependence model. Second, likelihood based inference is valid under ignorability and the missing data mechanism need not be explicitly specified unlike the GEE approach in Dobbie and Welsh (2001). The use of a likelihood based approach will have advantages for longitudinal data that is missing at random (MAR) and in particular, ignorable.

2.2. Marginalized Negative Binomial Hurdle models

Now we propose marginalized negative binomial hurdle models to account for nonzero overdispersed data. The marginal probability of responses are given by

$$P(y_{it}; x_{it}) = \begin{cases} 1 - q_{it}^{M}, & \text{if } y_{it} = 0; \\ q_{it}^{M} \frac{k(y_{it}; \mu_{it}^{M})}{1 - \left(\frac{1}{1 + \nu \mu_{it}^{M}}\right)^{\nu^{-1}}, & \text{if } y_{it} = 1, 2, \cdots, \end{cases}$$

where

$$\begin{aligned} \log & \text{i} t q_{it}^{M} = x_{it}^{T} \gamma, \\ & k(y_{it}; \mu_{it}^{M}) = \frac{\Gamma(\nu^{-1} + y_{it})}{\Gamma(\nu^{-1}) y_{it}!} \left(\frac{1}{1 + \nu \mu_{it}^{M}}\right)^{\nu^{-1}} \left(\frac{\nu \mu_{it}^{M}}{1 + \nu \mu_{it}^{M}}\right)^{y_{it}}, \\ & \log \mu_{it}^{M} = x_{it}^{T} \gamma. \end{aligned}$$

The conditional probabilities of responses given random effects are given by

$$P(y_{it}; b_i) = \begin{cases} 1 - q_{it}^c(b_{i1}), & \text{if } y_{it} = 0; \\ q_{it}^c(b_{i1}) \frac{k(y_{it}; \mu_{it}^c(b_{i2}))}{1 - \left(\frac{1}{1 + \nu \mu_{it}^c(b_{i2})}\right)^{\nu^{-1}}, & \text{if } y_{it} = 1, 2, \cdots, \end{cases}$$

where

$$\begin{aligned} \log i q_{it}^{c}(b_{i1}) &= \Delta_{it1} + z_{i1}^{T} b_{i1}, \\ k(y_{it}; \mu_{it}^{c}(b_{i2})) &= \frac{\Gamma(\nu^{-1} + y_{it})}{\Gamma(\nu^{-1}) y_{it}!} \left(\frac{1}{1 + \nu \mu_{it}^{c}(b_{i2})}\right)^{\nu^{-1}} \left(\frac{\nu \mu_{it}^{c}(b_{i2})}{1 + \nu \mu_{it}^{c}(b_{i2})}\right)^{y_{it}}, \\ \log \mu_{it}^{c}(b_{i2}) &= \Delta_{it2} + z_{i2}^{T} b_{i2}, \end{aligned}$$

where b_i is random effects which are jointly normal and possibly correlated and is given in (6). We also consider the simple random intercept form here.

Similar to the identities, (8) and (9), for the MPH, we have the following relationships

$$q_{it}^{M} = \int q_{it}^{c}(b_{i1})f(b_{i1})db_{i1},$$
(10)

$$E(Y_{it}|x_{it}) = E\left\{E(Y_{it}|b_{i2})\right\},$$

$$\Leftrightarrow \quad q_{it}^{M} \frac{\mu_{it}^{M}}{1 - \left(\frac{1}{1 + \nu \mu_{it}^{M}}\right)^{\nu^{-1}}} = \int q_{it}^{c}(b_{i1}) \frac{\mu_{it}^{c}(b_{i2})}{1 - \left(\frac{1}{1 + \nu \mu_{it}^{c}(b_{i2})}\right)^{\nu^{-1}}} f(b_{i}) db_{i}.$$
(11)

We also solve Δ_{it1} and Δ_{it2} using (10) and (11) given γ , β and σ using a Newton-Raphson. For the explicit forms of the terms in the Newton-Raphson algorithm, also see the Appendix.

2.3. Reparametrization of the random effects and their covariance matrix

From a computational perspective, it is convenient to orthogonalize the random effects by setting $b_i = \Sigma^{\frac{1}{2}} z_i$, where $\Sigma^{\frac{1}{2}}$ is the Cholesky factor of the 2 × 2 matrix Σ (Gibbons and Bock, 1987), which are given by

$$\Sigma^{\frac{1}{2}} = \begin{pmatrix} s_{11} & 0 \\ s_{21} & s_{22} \end{pmatrix},$$

and z_i is a 2×1 vector of independent standard normals. The reparameterized conditional model for MPH model is then given by

$$p_{it}^{c}(b_{i1}) = \frac{\exp(\triangle_{it1} + s_{11}z_{i1})}{1 + \exp(\triangle_{it1} + s_{11}z_{i1})} \stackrel{\text{let}}{=} p_{it}^{c}(z_{i}),$$
$$\lambda_{it}^{c}(b_{i2}) = \exp(\triangle_{it2} + s_{21}z_{i1} + s_{22}z_{i2}) \stackrel{\text{let}}{=} \lambda_{it}^{c}(z_{i})$$

Similarly, we reexpress (10) in MNBH model as given by,

$$\mu_{it}^{c}(b_{i2}) = \exp(\triangle_{it2} + s_{21}z_{i1} + s_{22}z_{i2}) \stackrel{\text{let}}{=} \mu_{it}^{c}(z_{i}).$$

This transformation allows us to estimate the Cholesky factor $\Sigma^{\frac{1}{2}}$ instead of the covariance matrix Σ . Since the Cholesky factor is the square root of the covariance matrix, it allows a more stable estimation of near-zero variance terms (Hedeker and Gibbons, 1994).

2.4. Estimation

The likelihood function for the MPH model, which is the integral over random effects of a product of two distributions, is given by

$$L(\theta; y) = \prod_{i=1}^{N} \int \prod_{t=1}^{n_i} (1 - p_{it}^c(z_i))^{I_{(y_{it}=0)}} \left\{ p_{it}^c(z_i) \frac{g(y_{it}; \lambda_{it}^c(z_i))}{1 - e^{-\lambda_{it}^c(z_i)}} \right\}^{1 - I_{(y_{it}=0)}} \phi(z_i) dz_i, \quad (12)$$

where $\phi(\cdot)$ is a multivariate standard normal density with mean vector 0 and variancecovariance matrix I, $g(\cdot; \lambda_{it}^c(z_i))$ is a Poisson probability mass function with mean $\lambda_{it}^c(z_i)$, $\theta^T = (\gamma^T, \beta^T, s^T)$, and $s^T = (s_{11}, s_{21}, s_{22})$. Since the marginalized likelihood in (12) is not available in closed form, we use Gauss-Hermite quadrature to (numerically) integrate out the random effects.

Maximizing the log-likelihood with respect to θ yields the likelihood equation

$$\sum_{i=1}^{N} \frac{\partial \log L(\theta; y_i)}{\partial \theta} = \sum_{i=1}^{N} L^{-1}(\theta; y_i) \int \frac{\partial L(\theta, z_i; y_i)}{\partial \theta} \phi(z_i) dz_i = 0,$$

where

$$L(\theta; y_i) = \int \prod_{t=1}^{n_i} (1 - p_{it}^c(z_i))^{I_{(y_{it}=0)}} \left\{ p_{it}^c(z_i) \frac{g(y_{it}; \lambda_{it}^c(z_i))}{1 - e^{-\lambda_{it}^c(z_i)}} \right\}^{1 - I_{(y_{it}=0)}} \phi(z_i) dz_i,$$

$$L(\theta, z_i; y_i) = \prod_{t=1}^{n_i} (1 - p_{it}^c(z_i))^{I_{(y_{it}=0)}} \left\{ p_{it}^c(z_i) \frac{g(y_{it}; \lambda_{it}^c(z_i))}{1 - e^{-\lambda_{it}^c(z_i)}} \right\}^{1 - I_{(y_{it}=0)}}.$$
(13)

The (2p+3) dimensional likelihood equations are given in the Appendix.

The matrix of second derivatives of the observed data log-likelihood has a very complex form. Fortunately, the sample empirical covariance matrix of the individual scores in any correctly specified model is a consistent estimator of the information and involves only the first derivatives. So the Quasi-Newton method can be used to solve the likelihood equations using

$$\theta^{(c+1)} = \theta^{(c)} + \left[I_e\left(\theta^{(g)}; y\right)\right]^{-1} \frac{\partial \log L}{\partial \theta^{(c)}},$$

where $I_e(\theta)$ is an empirical and consistent estimator of the information matrix at step g and is given by

$$I_{e}\left(\theta;y\right) = \sum_{i=1}^{N} \frac{\partial L(\theta;y_{i})}{\partial \theta} \frac{\partial L(\theta;y_{i})}{\partial \theta^{T}}.$$

At convergence, the large-sample variance-covariance matrix of the parameter estimates is then obtained as the inverse of $I_e(\hat{\theta}; y)$. For the explicit forms of the terms in the Quasi-Newton algorithm, see the Appendix.

Similarly, we have the log likelihood function, score function, and information matrix for the MNBH models and detailed calculations in the Appendix.

3. Simulation Study

We conducted a simulation study to examine the bias in estimating the parameters in the MPH model. We simulated clustered zero-inflated data under an MPH model. Covariates were group (2 levels) and time. The marginal models from (1) and (2) were specified as

$$\begin{aligned} \operatorname{logit}(p_{it}^{M}) &= \gamma_{0} + \gamma_{1}\operatorname{group}_{i} + \gamma_{2}\operatorname{time}_{it}, \\ \operatorname{log}(\lambda) &= \beta_{0} + \beta_{1}\operatorname{group}_{i} + \beta_{2}\operatorname{time}_{it}, \\ \gamma &= (\gamma_{0}, \gamma_{1}, \gamma_{2}) = (-2.0, 0.3, 0.1), \\ \beta &= (\beta_{0}, \beta_{1}, \beta_{2}) = (0.2, 0.4, 0.5), \end{aligned}$$

where time_{it} = (t - 1)/10 for $t = 1, \dots, 6$ and group_i = 0 or 1 such that the sample size per group was approximately equal for all times t. The dependence models were using from (4) and (5) with

$$\Sigma = \left(\begin{array}{cc} 2.5 & 2\\ 2 & 3 \end{array}\right).$$

We generated 200 simulated data sets each with sample size of 400. We fit two models, the true model (Model 1) and an independent hurdle model ignoring the temporal dependence (Model 2). Based on this specification, the observed zero rates were approximately 60 percent.

Table 1 presents mean estimates, the absolute values of biases (AVB), and 95% Monte Carlo error intervals of the parameters. The estimates were essentially unbiased with very similar AVB's for Models 1 and 2 except β_0 . The 95% Monte Carlo error intervals contained the true values for the parameters. However, the coefficient of β_0 in Model 2 had a big bias. Therefore, we conclude that the estimated marginal parameters in our proposed model (Model 1) are on average closer to the true parameter value.

4. Example

4.1. A Pharmaceutical Study

This data set first described by Min and Agresti (2005) to illustrate hurdle models with random effects is used here to demonstrate the use of our marginalized hurdle models. One hundred and eighteen patients were randomly assigned to one of two treatment groups,

Bias of maximum likelihood estimators. Displayed are the average regression coefficient estimates and the absolute value of biases $(AVB=|\bar{\theta}-\theta|)$ where M is the number of simulated data sets. 95% Monte Carlo error intervals $(\bar{\theta}\pm 1.96\sqrt{var(\bar{\theta})/M})$ in parentheses.

| | | Moc | lel 1 | Moc | lel 2 |
|------------|------|---------|--------|---------|---------|
| para. | true | Mean | AVB | Mean | AVB |
| γ_0 | -2.0 | -2.01 | 0.01 | -2.01 | 0.01 |
| | | (-2.03, | -1.99) | (-2.03) | ,-1.99) |
| γ_1 | 0.3 | 0.30 | 0.00 | 0.30 | 0.00 |
| | | (0.27, | (0.32) | (0.27 | ,0.32) |
| γ_2 | 0.1 | 0.12 | 0.02 | 0.12 | 0.02 |
| | | (0.08, | (0.17) | (0.08 | ,0.17) |
| β_0 | 0.2 | 0.21 | 0.01 | 0.03 | 0.17 |
| | | (0.17, | (0.25) | (-0.03 | ,0.09) |
| β_1 | 0.4 | 0.43 | 0.03 | 0.45 | 0.05 |
| | | (0.39, | (0.47) | (0.37) | ,0.52) |
| β_2 | 0.5 | 0.46 | 0.04 | 0.46 | 0.04 |
| | | (0.42) | (0.51) | (0.39) | ,0.54) |

Table 2

Sample means and variances of side effect number and nonzero side effect number, and proportions of nonzero side effect number for each year.

| Visit | 1 | 2 | 3 | 4 | 5 | 6 |
|----------------------|-------|-------|-------|-------|-------|-------|
| Mean | 0.093 | 0.229 | 0.229 | 0.314 | 0.424 | 0.432 |
| variance | 0.102 | 0.366 | 0.315 | 0.696 | 1.033 | 1.153 |
| Mean (Nonzero) | 1.100 | 1.421 | 1.350 | 1.762 | 2.000 | 2.217 |
| variance | 0.100 | 0.591 | 0.345 | 1.390 | 1.750 | 1.996 |
| Proportion (Nonzero) | 0.085 | 0.161 | 0.169 | 0.178 | 0.212 | 0.195 |

treatment A (TRT 1, 59 patients) and treatment B (TRT 2, 59 patients). Response variable was the number of side effect episodes which was measured at each of six visits. As Min and Agresti (2005) indicated, about 83% of the observations were zeros and there was a variety of count number as visit number increased. Therefore, time was incorporated as a covariate in the model.

Table 2 presents sample means and variances of side effect numbers and nonzero side effect numbers, and proportions of nonzero side effect numbers for each visit. Since the sample means of side effects numbers were smaller than the corresponding sample variances, we fitted the negative binomial GLMM. The sample means of nonzero side effects numbers is larger than the corresponding variances. This indicates that overdispersion did not happen in nonzero count data. Therefore, we fitted the MPH model. We also fitted the Poisson GLMM, Dobbie and Welsh's model (2001), and Min and Agresti's Poisson hurdle random effects model (2005), respectively.

Comparison of maximized loglikelihoods and AICs

| | MPH | M&A | Poisson GMM | NB GMM |
|--------------|--------|--------|-------------|--------|
| Max. loglik. | -409.1 | -409.2 | -426.1 | -416.6 |
| AIC | 836.2 | 836.4 | 860.1 | 843.1 |

The Quasi-Newton algorithm is not computationally trivial because of numerical integration. We simultaneously used R 2.6.0 software (http://www.r-project.org) and FOR-TRAN 77. The R software was used to compute complicated calculation of matrices and the Quasi-Newton iteration and the FORTRAN 77 was used to make subroutines (.dll files) to implement the calculations of Δ using the Newton-Raphson and of derivatives of Δ . Each Quasi-Newton step (in which all the Δ_{itj} need to be computed) on a PC with 1.86 GHz processor took about 1.5 minutes for the MPH model with 40 point Gauss-Hermite quadrature. To reduce the number of iterations until convergence, we used initial values that are obtained by fitting an independent hurdle model in R resulting in a minimal number of iterations until convergence. In our analysis below, we obtained convergence in 40 iterations using a fairly strict convergence criterion, $\sqrt{(\hat{\theta}^{old} - \hat{\theta}^{new})^T(\hat{\theta}^{old} - \hat{\theta}^{new})} \leq 10^{-5}$,

where $\hat{\theta}^{new}$ and $\hat{\theta}^{old}$ are current and previous fitted values of parameters, respectively.

Table 3 presents maximized loglikelihood and AIC for all models except Dobbie and Welsh model. The MPH and Min and Agresti model fit better than GLMMs. The maximized loglikelihood and AIC for the MPH and M&A model were very similar. However, the MPH and M&A was based on marginal model and conditional models, respectively. Here we focus on the marginal relationship between mean of side effects numbers and covariates.

Table 4 presents maximum likelihood estimates of marginal mean parameters for MPH and D & W. The estimated values for the two models were very similar. Because there was no missing data in this data set, GEE and maximum likelihood estimates were similar (Diggle et al., 2002). The estimated parameter comparing the treatment in the logistic regression was significant with an estimate of 0.665 (SE=0.310). This indicated that the estimated marginal probability of side effect occurrence for treatment B was higher than for treatment A. In the estimates of marginal mean of side effect parameters, the coefficients of Trt2 (0.868, SE=0.415) and log(Time) (0.540, SE=0.205) were significant, respectively. This indicated that the marginal mean of side effect number for treatment B was higher than for treatment A controlling at fixed time and the marginal mean increased as time increased, respectively. The ML estimate for Σ for the MPH was

$$\hat{\Sigma} = \left(\begin{array}{cc} 2.719 & 0.999 \\ 0.999 & 0.509 \end{array} \right).$$

This presented large subject-to-subject variation in the odds of probability of side effect occurrence (2.179) compared with the variation of mean of side effect number (0.509). The estimated correlation of two random effects, b_{i1} and b_{i2} , was 0.894 and was very positive correlated. It is very similar value to Min and Agresti's result (0.848).

Maximum likelihood estimates and standard errors for MPH and D & W models.

| | MP | H | | | | | |
|---------------------|--------------|---------|--------------|-------|--|--|--|
| Term | Est. | S.E. | Est. | S.E. | | | |
| Logistic | e Part | | | | | | |
| Int. | -2.038^{*} | 0.406 | -2.033^{*} | 0.471 | | | |
| Trt2 | 0.665^{*} | 0.310 | 0.709^{*} | 0.210 | | | |
| $\log(\text{Time})$ | 0.018 | 0.117 | 0.010 | 0.160 | | | |
| Poisson | Part | | | | | | |
| Int. | -1.949^{*} | 0.598 | -1.999^{*} | 0.593 | | | |
| Trt2 | 0.868^{*} | 0.415 | 0.904^{*} | 0.278 | | | |
| $\log(\text{Time})$ | 0.540^{*} | 0.205 | 0.527^{*} | 0.182 | | | |
| * 1 1 | · · · · · | 11 OFOT | C 1 | 1 1 | | | |

* indicates significance with 95% confidence level.

4.2. New Orleans Murder Rate Study

The data set for this study includes 37 census tract/police zones for each decennial census year from 1940 to 2000 ($7 \times 37 = 259$). While the selection of these 37 census tract/police zones creates a bias toward the older tracts in the city, these tracts were the only tracts that were also police zones, thus providing consistent murder counts and census data for the 60 year period. Examining murder over this 60 year period is theoretically relevant because the rate of murder in the city changed dramatically in the early 1970s. For example, the average number of murders in the 37 census tract/polices zones for the decennial years 1940 to 1970 was 14.5. However, the average from 1980 to 2000 was 66.6. This occurred in the context of a population decline in the 37 tracts from 128,204 in 1970 to 115,209 in 1980. The question is: how do we account for this dramatic change in the volume of murder? About 63% of all 259 count values were zero. Therefore, hurdle models could be considered to fit.

For simplicity, we assume that the murder numbers in the district zones are independent (Mears and Bhati, 2006). It is justified from the perspective of criminological theory in which murders in contiguous zones could have spillover and human behavior is not constrained by artificial or arbitrary boundaries. The zones, however, when they were established were drawn along census tract lines. These census tracts were thought to be relatively homogeneous neighborhoods in a sociological sense. (Green and Truesdell, 1934).

To find the demographic factors that elevate murder rate, longitudinally measured count response of murders of district zones in New Orleans were analyzed with our proposed models. As predictors, we included the proportion of African American - a proxy for racial isolation (Black), income quantiles (Pnilf), the proportion of people in the age group 15 to 24 (Pro15to24), and decennial years (Year= 0.0, 0..., 0.6). Two covariates, Black and Pro15to24, were two structural covariates of neighborhoods selected because racial isolation as measured by percent African American and the proportion of the population 15 to 24 years of age are consistent with the criminological literature (Wilson, 1987, 1994; Cohen and Tita, 1999; Cork, 1999; Messner et al., 1999; Lee, 2000; Baller et al., 2001;

| ıα | and standard deviation of independent variables. | | | | | | | | |
|----|--|------|-------|-------|-------|-------|-------|-------|-------|
| | | | 1940 | 1950 | 1960 | 1970 | 1980 | 1990 | 2000 |
| | Black | Mean | 0.333 | 0.277 | 0.422 | 0.490 | 0.439 | 0.542 | 0.575 |
| | | Std. | 0.229 | 0.235 | 0.398 | 0.377 | 0.360 | 0.350 | 0.364 |
| | Pnilf | Mean | 0.327 | 0.341 | 0.361 | 0.318 | 0.436 | 0.352 | 0.341 |
| | | Std. | 0.058 | 0.056 | 0.047 | 0.066 | 0.117 | 0.113 | 0.118 |
| | Pro15to24 | Mean | 0.097 | 0.094 | 0.063 | 0.085 | 0.181 | 0.064 | 0.076 |
| | | Std. | 0.039 | 0.039 | 0.014 | 0.022 | 0.038 | 0.023 | 0.038 |

Mean and standard deviation of independent variables

Table 5

Sample means and variances of murder numbers and nonzero murder numbers, and proportions of nonzero murder numbers for each year.

| Year | 1940 | 1950 | 1960 | 1970 | 1980 | 1990 | 2000 |
|-----------------------------|-------|-------|-------|-------|-------|--------|-------|
| Mean (murder no.) | 0.297 | 0.432 | 0.324 | 0.676 | 1.594 | 2.270 | 1.568 |
| variance | 0.326 | 0.808 | 0.503 | 1.836 | 5.359 | 12.758 | 3.530 |
| Mean (Nonzero murder) | 1.222 | 1.778 | 1.500 | 2.273 | 3.278 | 4.200 | 2.900 |
| variance | 0.194 | 0.944 | 0.571 | 2.618 | 5.507 | 15.642 | 2.621 |
| Proportion (Nonzero murder) | 0.243 | 0.243 | 0.216 | 0.297 | 0.486 | 0.541 | 0.541 |

Griffiths and Chavez, 2004; Mears and Bhati, 2006).

Table 5 indicates means and standard deviation for the predictors. Population at each year was used as offset and Pop = Population/10,000 was used for numerical reasons. Table 5 shows the means and standard deviations of the independent variables for the each census year, indicating as follows: 1) the proportion of African Americans was likely to increase in the census year, 2) income quantiles did not change from year to year, and 3) the proportion of 15 to 24 years old did not change from year to year except 1980.

Year effect was also important. Table 6 indicates sample means and variances of murder numbers and nonzero murder numbers, and proportions of nonzero murder numbers for each year. The means of murder and nonzero murder numbers had cubic trend over years. A similar pattern occurred in the proportions of nonzero murder numbers. We considered linear, quadratic, and cubic terms of covariate year. We also considered the fourth power of year, but it was not significant

Using marginal models may allow us to clarify some of the inconsistencies in the predictive strength of covariates of murder over this 60 year period. We fit two marginalized hurdle models proposed in this paper, two hurdle models with random effects which were proposed in Min and Agresti (2005), and two generalized mixed models with Poisson and negative binomial distributions, respectively. Let Model 1 be an MPH and Model 2 be an MNBH. Both had variance-covariance structures for random effects. Models 3 and 4 are Min and Agresti's hurdle models using Poisson and negative binomial distributions, respectively. Finally, Models 5 and 6 are the Poisson regression model and negative binomial regression models with random effects, respectively.

Comparison of maximized loglikelihoods and AICs

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|--------------|--------------|---------|-----------|---------|---------|---------|
| Max. loglik. | -308.9 | -299.7 | -308.4 | -299.8 | -311.9 | -299.9 |
| AIC | 651.8 | 635.4 | 650.7 | 635.6 | 639.8 | 617.7 |
| * | in lineton o | .::c | : LL OF 0 | 7C.l | ll | |

indicates significance with 95% confidence level.

Table 8

Maximum likelihood estimates and standard errors for negative binomial GLMM.

| Term | Int. | PerBlack | Per15to24 | \mathbf{Pnilf} | Year | $Year^2$ | $Year^3$ | ν | σ |
|-----------|--------------|----------|-------------|------------------|--------------|--------------|----------|-------|-------------|
| Estimate | -1.807^{*} | 0.810 | 5.540^{*} | 0.096 | -9.678^{*} | 56.825^{*} | -58.001 | 0.445 | 1.666^{*} |
| Std. err. | 0.633 | 0.487 | 1.959 | 1.151 | 4.656 | 17.668 | 18.360 | 0.157 | 0.279 |
| * :1: + | : : C | | -1 - 0507 | C.] | 11 | | | | |

* indicates significance with 95% confidence level.

Table 7 presents maximized loglikelihood and AIC for all models. We first considered marginalized and marginal models (Models 3, 4, 5, 6). Since Models 3 and 5, and 4 and 6 are nested models, respectively, the likelihood ratio tests were conducted and the comparisons indicated that we cannot claim Model 3 fit better than Model 5 ($\Delta D_{35} = 2 \times (311.9 - 308.4) = 7$, p-value= 0.637 on 9 d.f.). Similarly, we cannot claim that Model 4 fit better than Model 6 ($\Delta D_{46} = 0.4$, p-value= 0.9999 on 9 d.f.). Since Models 1 and 2 were not nested, we compared them using AIC. The AIC for Model 1 was 651.8 and for Model 2 635.4, indicating that Model 2 had the better fit. Similar comparison was conducted for conditional Models 2 and 6. The comparison of AIC for Model 2 and 6 indicates that Model 6 fit better than Model 2 (617.7 for Model 6). From model comparison, the hurdle models (Models 1-4) did not find significant evidence of zero inflation. An ordinary negative binomial GLMM seems sufficient for this data set.

Table 8 presents maximum likelihood estimates of marginal mean parameters for negative binomial GLMM. The overdispersion parameter ($\hat{\nu} = 0.445$, p-value= 0.008) was significant. This indicated that the New Orleans murder data were overdispersed. The estimate for σ was $\hat{\sigma} = 1.666$ (p-value< 0.001). This indicated the census tract/police zone variation in the mean rate of murder. The coefficient of Pro15to24 (5.540, pvalue= 0.0076) in the regression was significant. This indicated that the estimated rate of murder increased by the proportion of people in the age group 15 to 24. The result provides support for life-course criminology. This perspective asserts that offending and anti-social behavior varies by age with peaking around age 20 (Farrington, 2003). The linear (-9.678, p-value= 0.045), quadratic (56.825, p-value= 0.0027), and cubic (-58.001, p-value= 0.0032) terms of year were also significant. This indicated that murder rate trend was cubic in years and explained the dramatic change in murder rates in the 1970s.

5. Conclusion

In this paper, we develop two new marginalized models to accommodate zero-inflated clustered count data. An advantage of these models is the ability to use conditional models for serial association of responses while still structuring the marginal mean as a function of covariates directly. As a result, the interpretation of the regression coefficients in two marginal components does not depend on the specification of the dependence model. Two-part random effects, which have a binary component and a truncated Poisson or negative binomial component, are used to explain the serial dependence of responses and covariance of responses at the same time. Parameter estimation was based on maximum likelihood using a Quasi-Newton algorithm. To evaluate the marginalized likelihood, we used Gauss-Hermite quadrature to evaluate integrations. Gauss-Hermite quadrature is commonly used for low dimensional random effects models such as random intercept models. However, a major disadvantage of the Gauss-Hermite quadrature is that the number of quadrature points increases as an exponential function of the number of dimensions. As an alternative, Monte Carlo methods are often used for models with higher-dimensional integrals and use randomly sampled points to approximate the integrals.

Although two data sets had many zeros in Section 4, hurdle models worked in the pharmaceutical data but not in the New Orleans murder data. We know that a high percentage of zeros does not disqualify Poisson or negative binomial models. The pharmaceutical data analysis provided that the estimated probability of side effects for occurrence increased in treatment 2 and the estimated mean of side effect number increased by the time. In the New Orleans murder data, our analysis provided sociologically meaningful contributions that predict murder rates in New Orleans. The estimated rate of murder increased by the proportion of people in the age group 15 to 24.

We can extend marginalized hurdle models to allow multivariate zero-inflated clustered count data. These are on-going works.

Acknowledgments

We would like to thank Kelly Frailing, Institute of Criminology, Cambridge University, U.K. for comments on an earlier draft of the paper. The authors gratefully acknowledge the editor and two referees for insightful comments that have greatly improved the manuscript. This research was supported by the KIEST (Grant No. 091-091089) and Arkansas Biosciences Institute (ABI).

APPENDIX A: Calculations for Estimation of MPH Models

Detailed Calculations of Δ

Given γ , β , and σ , we calculate Δ_{it1} and Δ_{it2} from the relationship (8) and (9) using a Newton-Raphson. Let

$$h_1(\Delta_{it1}) = \int P_{it}^c(b_{it1})f(b_{it1})db_{it1} - P_{it}^M,$$

$$h_2(\Delta_{it}) = \int p_{it}^c(b_{i1})\frac{\lambda_{it}^c(b_{i2})}{1 - e^{-\lambda_{it}^c(b_{i2})}}f(b_i)db_i - p_{it}^M\frac{\lambda_{it}^M}{1 - e^{-\lambda_{it}^M}}$$

Estimates of Δ_{it1} and Δ_{it2} satisfy $(h_1(\Delta_{it1}), h_2(\Delta_{it})) = (0, 0)$ and can be obtained using Newton-Raphson as follows

$$\begin{pmatrix} \Delta_{it1}^{(c+1)} \\ \Delta_{it2}^{(c+1)} \end{pmatrix} = \begin{pmatrix} \Delta_{it1}^{(c)} \\ \Delta_{it2}^{(c)} \end{pmatrix} - \begin{pmatrix} \frac{\partial h_1(\Delta_{it1})}{\partial \Delta_{it1}} & \frac{\partial h_1(\Delta_{it1})}{\partial \Delta_{it2}} \\ \frac{\partial h_2(\Delta_{it})}{\partial \Delta_{it1}} & \frac{\partial h_2(\Delta_{it})}{\partial \Delta_{it2}} \end{pmatrix}^{-1}_{\Delta_{it} = \Delta_{it}^{(c)}} \begin{pmatrix} h_1(\Delta_{it1}^{(c)}) \\ h_2(\Delta_{it1}^{(c)}) \end{pmatrix}$$

where

$$\frac{\partial h_1(\Delta_{it1})}{\partial \Delta_{it1}} = \int P_{it}^c(b_{it1})(1 - P_{it}^c(b_{it1}))f(b_{it1})db_{it1},$$

$$\frac{\partial h_1(\Delta_{it1})}{\partial \Delta_{it2}} = 0,$$
(14)

$$\frac{\partial h_{2}(\Delta_{it})}{\partial \Delta_{it1}} = \int p_{it}^{c}(b_{i1})(1 - p_{it}^{c}(b_{i1})) \frac{\lambda_{it}^{c}(b_{i2})}{1 - e^{-\lambda_{it}^{c}(b_{i2})}} f(b_{i}) db_{i},$$
(15)

$$\frac{\partial h_2(\Delta_{it})}{\partial \Delta_{it2}} = \int p_{it}^c \frac{\lambda_{it}^c(b_{i2})}{1 - e^{-\lambda_{it}^c(b_{i2})}} \left(1 - \frac{\lambda_{it}^c(b_{i2})e^{-\lambda_{it}^c(b_{i2})}}{1 - e^{-\lambda_{it}^c(b_{i2})}}\right) f(b_i)db_i,\tag{16}$$

To calculate the integrals in (14)-(16), we use Gauss-Hermite quadrature to evaluate this integrals.

Detailed Calculations of Quasi-Newton

The contribution of subject i to the log likelihood is given by

$$\log L(\theta; y_i) = \log \int L(\theta, z_i; y_i) \phi(z_i) dz_i,$$

where

$$L(\theta, z_i; y_i) = \exp\left[\sum_{t=1}^{n_i} (1 - I_{(y_{it}=0)}) \left(\Delta_{it1} + s_{11}z_{i1}\right) - \sum_{t=1}^{n_i} \log\left(1 + e^{\Delta_{it1} + s_{11}z_{i1}}\right) + \sum_{t=1}^{n_i} (1 - I_{(y_{it}=0)}) \left\{-\lambda_{it}(z_i) + y_{it}\log\lambda_{it}(z_i) - \log(y_{it}!) - \log\left(1 - e^{-\lambda_{it}(z_i)}\right)\right\}\right].$$

The forms of the derivatives for Quasi-Newton algorithm are

$$\begin{split} \frac{\partial \log L}{\partial \gamma} &= \sum_{i=1}^{N} L(\theta; y_i)^{-1} \int L(\theta, z_i; y_i) \left\{ \sum_{t=1}^{n_i} \left(u_{it} - p_{it}^c(z_i) \right) \frac{\partial \Delta_{it1}}{\partial \gamma} + \sum_{t=1}^{n_i} u_{it} \left(y_{it} - \frac{\lambda_{it}^c(z_i)}{1 - e^{-\lambda_{it}^c(z_i)}} \right) \frac{\partial \Delta_{it1}}{\partial \gamma} \right\} \phi(z_i) dz_i \\ \frac{\partial \log L}{\partial \beta} &= \sum_{i=1}^{N} L(\theta; y_i)^{-1} \int L(\theta, z_i; y_i) \left\{ \sum_{t=1}^{n_i} u_{it} \left(y_{it2} - \frac{\lambda_{it}^c(z_i)}{1 - e^{-\lambda_{it}^c(z_i)}} \right) \frac{\partial \Delta_{it2}}{\partial \beta} \right\} \phi(z_i) dz_i, \\ \frac{\partial \log L}{\partial s_{11}} &= \sum_{i=1}^{N} L(\theta; y_i)^{-1} \int L(\theta, z_i; y_i) \left\{ \sum_{t=1}^{n_i} \left(u_{it} - p_{it}^c(z_i) \right) \left(\frac{\partial \Delta_{it1}}{\partial s_{11}} + z_{i1} \right) \right. \\ &+ \sum_{t=1}^{n_i} u_{it} \left(y_{it} - \frac{\lambda_{it}^c(z_i)}{1 - e^{-\lambda_{it}^c(z_i)}} \right) \frac{\partial \Delta_{it2}}{\partial s_{11}} \right\} \phi(z_i) dz_i, \\ \frac{\partial \log L}{\partial s_{21}} &= \sum_{i=1}^{N} L(\theta; y_i)^{-1} \int L(\theta, z_i; y_i) \left\{ \sum_{t=1}^{n_i} u_{it} \left(y_{it} - \frac{\lambda_{it}^c(z_i)}{1 - e^{-\lambda_{it}(z_i)}} \right) \left(\frac{\partial \Delta_{i2}}{\partial s_{21}} + \frac{b_{i2}}{\partial s_{21}} \right) \right\} \phi(z_i) dz_i, \\ \frac{\partial \log L}{\partial s_{22}} &= \sum_{i=1}^{N} L(\theta; y_i)^{-1} \int L(\theta, z_i; y_i) \left\{ \sum_{t=1}^{n_i} u_{it} \left(y_{it} - \frac{\lambda_{it}^c(z_i)}{1 - e^{-\lambda_{it}(z_i)}} \right) \left(\frac{\partial \Delta_{i2}}{\partial s_{22}} + \frac{b_{i2}}{\partial s_{22}} \right) \right\} \phi(z_i) dz_i, \end{aligned}$$

where $u_{it} = 1$ if $y_{it} > 0$; = 0 if $y_{it} = 0$, $L(\theta, z_i | y_i)$ is given by (13). The integrals are estimated using Gauss-Hermite quadrature.

To make the derivatives simpler, (8) and (9) can be reexpressed as

$$p_{it}^{M} = \int p_{it}^{c}(s_{11}z_{i1})\phi(z_{i1})dz_{i1},$$
(17)

$$p_{it}^{M} \frac{\lambda_{it}^{M}}{1 - e^{-\lambda_{it}^{M}}} = \int p_{it}^{c}(z_{i1}) \frac{\lambda_{it}^{c}(s_{2}^{T}z_{i})}{1 - e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})}} \phi(z_{i}) dz_{i},$$
(18)

where $s_2^T = (s_{21}, s_{22}), z_i^T = (z_{i1}, z_{i2})$ is a random vector with the standard normal distribution, and $\phi(\cdot)$ is the standard normal density function. Note that the integrals in (17) and (18) are one dimensional. To compute the score vector and information matrix, we also need derivatives of Δ_{it} with respect to γ , β , and s. They can be obtained from the relationships (17) and (18)

$$\frac{\partial p_{it}^{M}}{\partial \gamma} = \int \frac{\partial p_{it}^{c}(s_{11}z_{i1})}{\partial \Delta_{it1}} \frac{\partial \Delta_{it1}}{\partial \gamma} \phi(z_{i1}) dz_{i1},$$

$$\Rightarrow \quad \frac{\partial \Delta_{it1}}{\partial \gamma} = \frac{\frac{\partial p_{it}^{M}}{\partial \gamma}}{\int \frac{\partial p_{it}^{c}(s_{11}z_{i1})}{\partial \Delta_{it1}} \phi(z_{i1}) dz_{i1}}.$$
(19)

Similarly, we have

$$\frac{\partial \Delta_{it2}}{\partial \gamma} = \frac{\frac{\partial p_{it}^{M}}{\partial \gamma} \nu_{it}^{M} - \frac{\partial \Delta_{it1}}{\partial \gamma} \int p_{it}^{c}(s_{11}z_{i1})(1 - p_{it}^{c}(s_{11}z_{i1}))\nu_{it}^{c}(s_{2}^{T}z_{i})\phi(z_{i})dz_{i}}{\int p_{it}^{c}(s_{11}z_{i1})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})}\right)\phi(z_{i})dz_{i}}, \\
\frac{\partial \Delta_{it2}}{\partial \beta} = \frac{p_{it}^{M}\nu_{it}^{M}\left(1 - \nu_{it}^{M}e^{-\lambda_{it}^{M}}\right)x_{it}}{\int p_{it}^{c}(s_{11}z_{i1})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})}\right)\phi(z_{i})dz_{i}}, \\
\frac{\partial \Delta_{it1}}{\partial s_{11}} = -\frac{\int p_{it}^{c}(s_{11}z_{i1})(1 - p_{it1}^{c}(s_{11}z_{i1}))z_{i1}\phi(z_{i1})dz_{i1}}{\int \frac{\partial p_{it1}^{c}(s_{11}z_{i1})}{\partial \Delta_{it1}}\phi(z_{i1})dz_{i1}}, \quad (20)$$

$$\frac{\partial \Delta_{it2}}{\partial s_{11}} = -\frac{\int p_{it}^{c}(s_{11}z_{i1})(1 - p_{it1}^{c}(s_{11}z_{i1}))\left(\frac{\partial \Delta_{it1}}{\partial \Delta_{i1}} + z_{i1}\right)\nu_{it}^{c}(s_{2}^{T}z_{i})\phi(z_{i})dz_{i}}, \\
\frac{\partial \Delta_{it2}}{\partial s_{21}} = -\frac{\int p_{it}^{c}(s_{11}z_{i1})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})}\right)\phi(z_{i})dz_{i}}{\int p_{it}^{c}(s_{11}z_{i})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})\phi(z_{i})dz_{i}}, \\
\frac{\partial \Delta_{it2}}{\partial s_{22}} = -\frac{\int p_{it}^{c}(s_{11}z_{i})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})\right)\phi(z_{i})dz_{i}}{\int p_{it}^{c}(s_{11}z_{i})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})\right)\phi(z_{i})dz_{i}}, \\
\frac{\partial \Delta_{it2}}{\partial s_{22}} = -\frac{\int p_{it}^{c}(s_{11}z_{i})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})\right)\phi(z_{i})dz_{i}}, \\
\frac{\partial \Delta_{it2}}{\int p_{it}^{c}(s_{11}z_{i})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})\right)\phi(z_{i})dz_{i}}, \\
\frac{\partial \Delta_{it2}}{\int p_{it}^{c}(s_{11}z_{i})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})\right)\phi(z_{i})dz_{i}}, \\
\frac{\partial \Delta_{it2}}{\partial s_{22}} = \frac{\int p_{it}^{c}(s_{11}z_{i})\nu_{it}^{c}(s_{2}^{T}z_{i})\left(1 - \nu_{it}^{c}(s_{2}^{T}z_{i})e^{-\lambda_{it}^{c}(s_{$$

where

$$\nu_{it}^{M} = \frac{\lambda_{it}^{M}}{1 - e^{-\lambda_{it}^{M}}},$$

$$\nu_{it}^{c}(s_{2}^{T}z_{i}) = \frac{\lambda_{it}^{c}(s_{2}^{T}z_{i})}{1 - e^{-\lambda_{it}^{c}(s_{2}^{T}z_{i})}}.$$

APPENDIX B: Calculations for Estimation of MNBH Models

Detailed Calculations of Δ_{it}

Given γ , β and σ , we calculate Δ_{it1} and Δ_{it2} from the relationships (10) and (11) using a Newton-Raphson. Let

$$h_{3}(\Delta_{it1}) = \int q_{it}^{c}(b_{i1})f(b_{i1})db_{i1} - q_{it}^{M},$$

$$h_{4}(\Delta_{it2}) = \int q_{it}^{c}(b_{i1})\frac{\mu_{it}^{c}(b_{i2})}{1 - \left(\frac{1}{1 + \nu \mu_{it}^{c}(b_{i2})}\right)^{\nu^{-1}}}f(b_{i})db_{i2} - q_{it}^{M}\frac{\mu_{it}^{M}}{1 - \left(\frac{1}{1 + \nu \mu_{it}^{M}}\right)^{\nu^{-1}}}.$$

Estimates of Δ_{it1} and Δ_{it2} satisfies $(h_3(\Delta_{it1}), h_4(\Delta_{it})) = (0, 0)$ and can be obtained using Newton-Raphson as follows

$$\begin{pmatrix} \Delta_{it1}^{(c+1)} \\ \Delta_{it2}^{(c+1)} \end{pmatrix} = \begin{pmatrix} \Delta_{it1}^{(c)} \\ \Delta_{it2}^{(c)} \end{pmatrix} - \begin{pmatrix} \frac{\partial h_3(\Delta_{it1})}{\partial \Delta_{it1}} & \frac{\partial h_3(\Delta_{it1})}{\partial \Delta_{it2}} \\ \frac{\partial h_4(\Delta_{it})}{\partial \Delta_{it1}} & \frac{\partial h_4(\Delta_{it})}{\partial \Delta_{it2}} \end{pmatrix}^{-1}_{\Delta_{it} = \Delta_{it}^{(c)}} \begin{pmatrix} h_3(\Delta_{it1}^{(c)}) \\ h_4(\Delta_{it}^{(c)}) \end{pmatrix}$$

where

$$\begin{aligned} \frac{\partial h_3(\Delta_{it1})}{\partial \Delta_{it1}} &= \int q_{it}^c(b_{i1})(1 - q_{it}^c(b_{i1}))f(b_{i1})db_{i1}, \\ \frac{\partial h_3(\Delta_{it1})}{\partial \Delta_{it2}} &= 0, \\ \frac{\partial h_4(\Delta_{it})}{\partial \Delta_{it1}} &= \int q_{it}^c(b_{i1})(1 - q_{it}^c(b_{i1}))\frac{\mu_{it}^c(b_{i2})}{1 - (1 + \nu\mu_{it}^c(b_{i2}))^{-\nu^{-1}}}f(b_i)db_i, \\ \frac{\partial h_4(\Delta_{it})}{\partial \Delta_{it2}} &= \int q_{it}^c(b_{i1})\frac{\mu_{it}^c(b_{i2})}{1 - (1 + \nu\mu_{it}^c(b_{i2}))^{-\nu^{-1}}} \left\{ 1 - \frac{\mu_{it}^c(b_{i2})(1 + \nu\mu_{it}^c(b_{i2}))^{-\nu^{-1}}}{1 - (1 + \nu\mu_{it}^c(b_{i2}))^{-\nu^{-1}}} \right\} f(b_{i2})db_{i2}. \end{aligned}$$

Detailed Calculations of Quasi-Newton

Let $\psi = (\gamma, \beta, s, \nu)$. Then the contribution of subject *i* to the log likelihood is given by

$$\log L(\psi; y_i) = \log \int L(\psi, z_i; y_i) \phi(z_i) dz_i,$$

where

$$L(\psi, z_i; y_i) = \exp\left[\sum_{t=1}^{n_i} (1 - I_{(y_{it1}=0)}) \left(\Delta_{it1} + s_{11}z_{i1}\right) - \sum_{t=1}^{n_i} \log\left(1 + e^{\Delta_{it1} + s_{11}z_{i1}}\right) + \sum_{t=1}^{n_i} (1 - I_{(y_{it1}=0)}) \left\{\sum_{l=0}^{y_{it}-1} \log(1 + \nu l) + y_{it} \log\mu_{it}^c(b_{i2}) - \left(y_{it} + \nu^{-1}\right) \log\left(1 + \nu\mu_{it}^c(b_{i2})\right) - \log(y_{it}!) - \log\left(1 - (1 + \nu\mu_{it}^c(b_{i2}))^{-\nu^{-1}}\right)\right\}\right].$$

The forms of the derivatives for Quasi-Newton algorithm are

$$\begin{split} \frac{\partial \log L}{\partial \gamma} &= \sum_{i=1}^{N} L(\psi; y_{i})^{-1} \int L(\psi, z_{i}; y_{i}) \left[\sum_{t=1}^{n_{i}} (u_{it} - q_{it}^{c}(z_{i})) \frac{\partial \Delta_{it1}}{\partial \gamma} + \sum_{t=1}^{n_{i}} u_{it} \left\{ y_{it} - \frac{\mu_{it}^{c}(z_{i})}{1 - (1 + \nu \mu_{it}^{c}(z_{i})^{-\nu^{-1}})} \right\} \frac{1}{1 + \nu \mu_{it}^{c}(z_{i})} \frac{\partial \Delta_{it2}}{\partial \gamma} \right] \phi(z_{i}) dz_{i}, \\ \frac{\partial \log L}{\partial \beta} &= \sum_{i=1}^{N} L(\psi; y_{i})^{-1} \int L(\psi, z_{i}; y_{i}) \sum_{t=1}^{n_{i}} u_{it} \left(y_{it2} - \frac{\mu_{it}^{c}(z_{i})}{1 - (1 + \nu \mu_{it}^{c}(z_{i}))^{-\nu^{-1}}} \right) \frac{1}{1 + \nu \mu_{it}^{c}(z_{i})} \frac{\partial \Delta_{it2}}{\partial \beta} \phi(z_{i}) dz_{i}, \\ \frac{\partial \log L}{\partial s_{11}} &= \sum_{i=1}^{N} L(\psi; y_{i})^{-1} \int L(\theta, z_{i}; y_{i}) \left\{ \sum_{t=1}^{n_{i}} (u_{it} - q_{it}^{c}(z_{i})) \frac{\partial \Delta_{it1}}{\partial s_{11}} + z_{i1} \right) \\ &+ \sum_{t=1}^{n_{i}} u_{it} \left(y_{it} - \frac{\mu_{it}^{c}(z_{i})}{1 - (1 + \nu \mu_{it}^{c}(z_{i}))^{-\nu^{-1}}} \right) \frac{1}{1 + \nu \mu_{it}^{c}(z_{i})} \frac{\partial \Delta_{it2}}{\partial s_{11}} \right\} \phi(z_{i}) dz_{i}, \\ \frac{\partial \log L}{\partial s_{21}} &= \sum_{i=1}^{N} L(\psi; y_{i})^{-1} \int L(\psi, z_{i}; y_{i}) \sum_{t=1}^{n_{i}} u_{it} \left(y_{it} - \frac{\mu_{it}^{c}(z_{i})}{1 - (1 + \nu \mu_{it}^{c}(z_{i}))^{-\nu^{-1}}} \right) \frac{1}{1 + \nu \mu_{it}^{c}(z_{i})} \frac{\partial \Delta_{it2}}{\partial s_{21}} + z_{i1} \right) \phi(z_{i}) dz_{i}, \\ \frac{\partial \log L}{\partial s_{22}} &= \sum_{i=1}^{N} L(\psi; y_{i})^{-1} \int L(\psi, z_{i}; y_{i}) \sum_{t=1}^{n_{i}} u_{it} \left(y_{it} - \frac{\mu_{it}^{c}(z_{i})}{1 - (1 + \nu \mu_{it}^{c}(z_{i}))^{-\nu^{-1}}} \right) \frac{1}{1 + \nu \mu_{it}^{c}(z_{i})} \left(\frac{\partial \Delta_{it2}}{\partial s_{22}} + z_{i1} \right) \phi(z_{i}) dz_{i}, \\ \frac{\partial \log L}{\partial s_{22}} &= \sum_{i=1}^{N} L(\psi; y_{i})^{-1} \int L(\psi, z_{i}; y_{i}) \sum_{t=1}^{n_{i}} u_{it} \left[- \sum_{l=0}^{y_{it}-1} \frac{1}{1 - (1 + \nu \mu_{it}^{c}(z_{i}))^{-\nu^{-1}}} \right) \frac{1}{\nu(1 + \nu \mu_{it}^{c}(z_{i}))} + (y_{it} - \mu_{it}^{c}(z_{i})) \frac{\partial \Delta_{it2}}{\partial \nu} \\ + \frac{1}{\nu^{2}} \log (1 + \nu \mu_{it}^{c}(z_{i})) + \frac{1}{1 - (1 + \nu \mu_{it}^{c}(z_{i}))^{-\nu^{-1}}} \left\{ \nu^{-2} (1 + \nu \mu_{it}^{c}(z_{i}))^{-\nu^{-1}} \log (1 + \nu \mu_{it}^{c}(z_{i})) \frac{\partial \Delta_{it2}}{\partial \nu} \right\} \right\} \phi(z_{i}) dz_{i}. \end{split}$$

where $u_{it} = 1$ if $y_{it} > 0$; = 0 if $y_{it} = 0$.

To make the derivatives simpler, (8) and (9) can be reexpressed as

$$q_{it}^{M} = \int q_{it}^{c}(s_{11}z_{i1})\phi(z_{i1})dz_{i1},$$
(22)

$$q_{it}^{M} \frac{\mu_{it}^{M}}{1 - \left(\frac{1}{1 + \nu \mu_{it}^{M}}\right)^{\nu^{-1}}} = \int q_{it}^{c}(s_{11}z_{i1}) \frac{\mu_{it}^{c}(s_{2}^{T}z_{i})}{1 - \left(\frac{1}{1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})}\right)^{\nu^{-1}}} \phi(z_{i}) dz_{i},$$
(23)

where z_i is a random variable with the standard normal distribution, and $\phi(\cdot)$ s in (22) and (23) are the univariate and bivariate standard normal density functions, respectively. Note that the integral in (22) and (23) are respectively one and two dimensional. To compute the score vector and information matrix, we also need derivatives of Δ_{it1} with respect to γ and s_{11} , and Δ_{it2} with respect to γ , β , ν , s_{11} , s_{21} , and s_{22} . The derivatives of Δ_{it1} are calculated in (19) and (20) by replacing $p_{it}^c(\cdot)$ and p_{it}^M with $q_{it}^c(\cdot)$ and q_{it}^M . The derivatives of Δ_{it2} can be obtained from the relationship (23)

$$\begin{split} \frac{\partial \Delta_{it2}}{\partial \beta} &= \frac{q_{it}^{M} \tau_{it}^{M} \left(1 - \tau_{it}^{M} \left\{1 + \nu \mu_{it}^{M}\right\}^{-\nu^{-1}-1}\right) x_{it}}{\int q_{it}^{c}(s_{11}z_{11}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right)\right\} \phi(z_{i}) dz_{i}},\\ \frac{\partial \Delta_{it2}}{\partial \gamma} &= \frac{\frac{\partial q_{it}^{M}}{\partial \gamma} \tau_{it}^{M} - \frac{\partial \Delta_{it1}}{\partial \gamma} \int q_{it}^{c}(s_{11}z_{1}) \left(1 - q_{it}^{c}(s_{11}z_{1})\right) \tau_{it}^{c}(s_{2}^{T}z_{i}) \phi(z_{i}) dz_{i}}{\int q_{it}^{c}(s_{11}z_{1}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right)\right\} \phi(z_{i}) dz_{i}},\\ \frac{\partial \Delta_{it2}}{\partial \nu} &= \frac{A - B}{\int q_{it}^{c}(s_{11}z_{1}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right)\right\} \phi(z_{i}) dz_{i}},\\ \frac{\partial \Delta_{it2}}{\partial s_{11}} &= -\frac{\int q_{it}^{c}(s_{11}z_{i}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right) \phi(z_{i}) dz_{i}},\\ \frac{\partial \Delta_{it2}}{\partial s_{21}} &= -\frac{\int q_{it}^{c}(s_{11}z_{i}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right) - \nu^{-1-1}\right\} \phi(z_{i}) dz_{i}},\\ \frac{\partial \Delta_{it2}}{\partial s_{22}} &= -\frac{\int q_{it}^{c}(s_{11}z_{i}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right) - \nu^{-1-1}\right\} \phi(z_{i}) dz_{i}},\\ \frac{\partial \Delta_{it2}}{\int q_{it}^{c}(s_{11}z_{i}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right) - \nu^{-1-1}\right\} \phi(z_{i}) dz_{i}},\\ \frac{\partial \Delta_{it2}}{\partial s_{22}} &= -\frac{\int q_{it}^{c}(s_{11}z_{i}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right) - \nu^{-1-1}\right\} \phi(z_{i}) dz_{i}}{\int q_{it}^{c}(s_{11}z_{i}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left\{1 - \tau_{it}^{c}(s_{2}^{T}z_{i}\right) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right) - \nu^{-1-1}\right\} \phi(z_{i}) dz_{i}}. \end{split}$$

where

$$\begin{split} \tau_{it}^{M} &= \frac{\mu_{it}^{M}}{1 - (1 + \nu \mu_{it}^{M})^{-\nu^{-1}}}, \\ \tau_{it}^{c}(s_{2}^{T}z_{i}) &= \frac{\mu_{it}^{c}(s_{2}^{T}z_{i})}{1 - (1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i}))^{-\nu^{-1}}}, \\ A &= q_{it}^{M} \tau_{it}^{M} \left(1 + \nu \mu_{it}^{M}\right)^{-\nu^{-1}} \frac{1}{\nu} \left\{ \frac{\frac{1}{\nu} \log \left(1 + \nu \mu_{it}^{M}\right)}{1 - (1 + \nu \mu_{it}^{M})^{-\nu^{-1}}} - \tau_{it}^{M} \left(1 + \nu \mu_{it}^{M}\right)^{-1} \right\}, \\ B &= \int q_{it}^{c}(s_{11}z_{i1}) \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right)^{-\nu^{-1}} \frac{1}{\nu} \left\{ \frac{\frac{1}{\nu} \log(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i}))}{1 - (1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i}))^{-\nu^{-1}}} - \tau_{it}^{c}(s_{2}^{T}z_{i}) \left(1 + \nu \mu_{it}^{c}(s_{2}^{T}z_{i})\right)^{-1} \right\} \phi(z_{i}) dz_{i}. \end{split}$$

References

- Baller, R. D., Anselin, L., Messner, S. F., Deane, G., and Hawkins, D. F. (2001). Structural Covariates of U.S. county homicide rates: Incorporating Spatial Effects. *Criminology.* 39, 561-90.
- Cohen, J. and Tita, G. (1999). Diffusion in Homicide: Exploring a general method for detecting spatial diffusion processes. *Journal of Quantitative Criminology*. 15, 451-93.

- Cork, D. (1999). Examining space-time interaction in city level homicide data: Crack markets and the diffusion of guns among youth. *Journal of Quantitative Criminol*ogy. 15, 379-406.
- Diggle, P.J., Heagerty, P, Liang, K.Y., and Zeger, S. (2002). Analysis of Longitudinal Data. Oxford University Press.
- Dobbie, M. J. and Welsh, A. H. (2001). Modeling correlated zero-inflated count data. Australian and New Zealand Journal of Statistics. 43, 431-444.
- Farrington, D. P. (2003) Developmental and life course criminology: Key theoretical and empirical issues. The 2002 Sutherland Award Address. *Criminology*. 41, 221-55.
- Gibbons, R. and Bock, R. (1987). Trend in correlated proportions. *Psychometrika*. 52, 113-124.
- Green, H. W. and Truesdell, L. E. (1934). Census Tracts in American Cities. *Bureau of the Census*, Washington D.C.
- Griffiths, E. and Chavez, J. M. (2004). Communities, street guns, and homicide trajectories in Chicago, 1980-1995: Merging methods for examining homicide trends across space and time. *Criminology*. 42, 941-78.
- Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics.* 56, 1030-1039.
- Hall, D. B. and Zhang, Z. (2004). Marginal models for zero inflated clustered data. Statistical Modelling. 4, 161-180.
- Heagerty, P. J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics.* 55, 688–698.
- Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics.* 50, 933-944.
- Gallant, A. R. and Nychka, D. W. (1987). Semiparametric maximum likelihood estimation. *Econometrica*. 55, 363-390.
- Gupta, P. L., Gupta, R. C., and Tripathi, R. C. (1996). Analysis of zero-adjusted count data. Computational Statistics & Data Analysis. 23, 207-218.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics.* 34, 1-14.
- Lee, K. and Daniels, M. (2008). Marginalized Models for Longitudinal Ordinal Data with Application to Quality of Life Studies. *Statistics in Medicine*. 27, 4359-4380.
- Lee, K., Joo, Y., Yoo, J. K., and Lee, J. (2009). Marginalized random effects models for multivariate longitudinal binary data. *Statistics in Medicine*. **28**, 1284-1300.

- Lee, K., Kang, S., Liu, X., and Seo, D. (2010). Likelihood-based approach for analysis of longitudinal nominal data using marginalized random effects models. *Journal of Applied Statistics*. Resubmitted.
- Lee, M. (2006). Concentrated poverty, race and homicide. *The Sociological Quarterly*. 41, 189-206.
- Mears, D. P. and Bhati, A. S. (2006). No Community is an island: the effects of resource deprivation on urban Violence in spatially and socially proximate communities. *Criminology*. 44, 509-47.
- Messner, S. F., Anselin, L., Baller, R. D., Hawkins, D. F., Deane, G. and Tolnay, S. E. (1999). The spatial patterning of county homicide rates: An application of exploratory spatial data analysis. *Journal of Quantitative Criminology*. 15, 423-50.
- Min, Y. and Agresti, A. (2005). Random effect models for repeated measures of zeroinflated count data. *Statistical Modelling.* 5, 1-19.
- Mullahy, J. (1986). Specification and testing of some modified count data models. Journal of Econometrics. 33, 341-365.
- Ridout, M., Hinde, J., and Demetrio, C. G. B. (2001). A score test for testing a zeroinflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics.* 57, 219-223.
- Wilson, W. J. (1987). The Truly Disadvantaged: The Inner City, The Underclass, and Public Policy. Chicago: University of Chicago Press.
- Wilson, W. J. (1997). When Work Disappears: The World of the New Urban Poor. New York: Alfred A. Knopf.
- Yau, K. K. W. and Lee, A. H. (2001). Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. *Statistics in Medicine*. 20, 2907-2920.