Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Application of finite mixture models for vehicle crash data analysis

Byung-Jung Park[1], Dominique Lord*

Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States

ARTICLE INFO

Article history:
Received 16 October 2008
Received in revised form 2 February 2009
Accepted 19 March 2009

Keywords:
Highway safety
Over-dispersion
Finite mixture
Negative binomial regression model
Latent class model

ABSTRACT

Developing sound or reliable statistical models for analyzing motor vehicle crashes is very important in highway safety studies. However, a significant difficulty associated with the model development is related to the fact that crash data often exhibit over-dispersion. Sources of dispersion can be varied and are usually unknown to the transportation analysts. These sources could potentially affect the development of negative binomial (NB) regression models, which are often the model of choice in highway safety. To help in this endeavor, this paper documents an alternative formulation that could be used for capturing heterogeneity in crash count models through the use of finite mixture regression models. The finite mixtures of Poisson or NB regression models are especially useful where count data were drawn from heterogeneous populations. These models can help determine sub-populations or groups in the data among others. To evaluate these models, Poisson and NB mixture models were estimated using data collected in Toronto, Ontario. These models were compared to standard NB regression model estimated using the same data. The results of this study show that the dataset seemed to be generated from two distinct sub-populations, each having different regression coefficients and degrees of over-dispersion. Although over-dispersion in crash data can be dealt with in a variety of ways, the mixture model can help provide the nature of the over-dispersion in the data. It is therefore recommended that transportation safety analysts use this type of model before the traditional NB model, especially when the data are suspected to belong to different groups.

1. Introduction

Developing sound or reliable statistical models for analyzing motor vehicle crashes is important in highway safety studies. The models can be used for identifying hazardous sites, predicting motor vehicle collisions, and developing accident modification factors (Lord and Park, 2008). What makes the analysis difficult in modeling crash data is that this kind of data often exhibit over-dispersion. The level of complexity becomes more significant when the data are characterized by small sample sizes and low sample mean values, which are commonly observed with crash data (Lord, 2006). Hauer (2001) reported that over-dispersion observed in crash data can be described in terms of "*represented traits*" and *unrepresented traits*"; the root cause of over-dispersion is that entities with the same represented traits have different means because of the unrepresented traits (measured or unmeasured) not included in the model. On the other hand, Lord et al. (2005) provided a more fundamental definition in which the over-dispersion arises from the actual nature of the crash process. This process dictates that the over-dispersion is the result of Bernoulli trials with unequal probability of independent events (known as Poisson trials) and all distributions, such as the Poisson-gamma (or negative binomial or NB) or Poisson-lognormal, are used as approximation to capture the over-dispersion observed in crash data.

Since the NB model can effectively approximate this underlying crash process, and reduce the unrepresented uncertainties by introducing a probabilistic error term related to the mean of the Poisson variable, transportation safety analysts have adopted this kind of model for developing crash prediction models. Within the NB regression modeling framework, many studies have been focused on the structure of the inverse dispersion parameter (ϕ, or its inverse $\phi = \alpha^{-1}$) of the NB distribution. Up until early 2000s, most of researchers in traffic safety have developed predictive models using a fixed or common dispersion parameter model (Hauer, 2001). In 2001, Heydecker and Wu suggested that ϕ could be modeled as a function of the covariates of the model (which can be defined as the varying dispersion parameter). Hauer (2001) argued that the inverse dispersion parameter should be modeled as a function of segment, $\phi_i = \delta L_i$, to correct for the unequal variance of the NB regression model. Since then, other researchers have investigated various structures of the dispersion parameter, both spatially and temporally (Lord and Park, 2008; Miaou and Lord, 2003; Miranda-Moreno et al., 2005; El-Basyouny and Sayed, 2006; Geedipally and Lord, 2008). Very recently, it was found that the structure of the dispersion parameter can greatly depend on how the mean function

[*] Corresponding author. Tel.: +1 979 458 3949; fax: +1 979 845 6481.
E-mail addresses: soldie@tamu.edu (B.-J. Park), d-lord@tamu.edu (D. Lord).

[1] Tel.: +1 979 862 8492; fax: +1 979 845 6481.

0001-4575/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2009.03.007

is modeled (Mitra and Washington, 2007). Models with a well-defined mean function may not have a structured variance. On the other hand, Shankar et al. (1998) showed that when spatial and temporal effects are not explicitly included in the NB model, the random effect negative binomial model offered advantages.

Despite the considerable efforts put in place to improve the performance of the NB models, several studies have documented important limitations associated with these models. As described above, research activities regarding the development of NB models using a fixed versus a varying dispersion parameter, and random parameter models with a fixed dispersion parameter are still on going. Varying dispersion parameter models may be preferred because one can determine sources influencing over-dispersion (Hilbe, 2007). However, finding appropriate covariates that influence the over-dispersion can be problematic if it is partly caused by unobserved variables or conditions. If the fixed dispersion model is preferred in terms of parameter parsimony, it may not tell us about the nature of the over-dispersion in the data. It only confirms that evidence of over-dispersion has been found and that this has been taken into account in the NB model (Land et al., 1996). Furthermore, if the datasets are characterized by small sample sizes and low mean values, the performance of the NB models can be significantly affected in terms of parameter estimation (Lord, 2006) as well as goodness-of-fit (Maher and Summersgill, 1996; Wood, 2002; Park and Lord, 2008). On the other hand, some have reported that NB regression models have difficulties handling the heavily over-dispersed data with a very long-tail and relatively high mean value because a negligible probability is usually assigned to high counts (Guo and Trivedi, 2002). Given these important limitations, Lord et al. (2008) suggested using the Conway–Maxwell–Poisson (COM–Poisson) model for crash data analysis. They showed that the COM–Poisson model performed as good as the NB model in terms of goodness-of-fit statistics and predictive performance.

As an alternative approach to address aforementioned problems, this paper documents an alternative formulation that could be used for capturing heterogeneity in crash count models through the use of finite mixture regression models (both for Poisson mixtures and NB mixtures). These models are compared to the results produced from the standard NB regression model. Modeling based on finite mixture distributions has a long history, and with the advancement of computing power and technology, it has continued to receive increasing attention in many areas, such as biometrics, genetics, medicine, and marketing (Frühwirth-Schnatter, 2006). The finite mixtures of Poisson regression models or NB regression models (abbreviated as FMP and FMNB, respectively, hereafter) are especially useful where count data were drawn from heterogeneous populations. There are many reasons to expect the existence of different sub-populations since the crash data are generally collected from various geographic, environmental and geometric design contexts over some fixed time periods. In such cases, it may be inappropriate to apply one aggregate NB regression model and the interpretation of the model could be misleading. Therefore, it would be reasonable to hypothesize that the individual crashes on highway entities (intersections, segments, etc.) are generated from a certain number (K) of hidden subgroups, or components that are unknown to the transportation safety analyst. The final outputs of FMNB regression models will be the number of components, component proportions, component-specific regression coefficients, and the degree of over-dispersion within each component.

2. Background

This section describes the characteristics of NB models and the finite mixture models.

2.1. Negative binomial distribution as a continuous mixture

Since the conventional Poisson model does not provide flexibility to accommodate frequently observed over-dispersion in crash data, several different mixed Poisson distributions have been applied by assuming a particular distribution in the Poisson mean rate (m_i):

$$m_i = \exp(\mathbf{x}_i\boldsymbol{\beta}) \cdot \varepsilon_i \tag{1}$$

where m_i is the Poisson mean rate for site i; $\mathbf{x}_i$ the covariate vector for site i where the first element is 1, $(1 \times d)$ vector; $\boldsymbol{\beta}$ a vector of unknown coefficients, $(d \times 1)$ vector; ε a model error independent of all the covariates; and, d is the number of covariates including an intercept.

Depending upon the parametric form imposed on $g(\varepsilon_i)$, various mixed Poisson regression models can be derived (e.g. Poisson-gamma, Poisson-lognormal, Poisson-inverse Gaussian, etc.). The negative binomial (NB) regression model arises if one assumes that $g(\varepsilon_i)$, or equivalently the distribution of m_i, $f(m_i)$, follows a gamma distribution. The Poisson-gamma distribution is the most common distribution used for modeling crash data because its marginal distribution has a closed form and this mixture results in a conjugate model (Hauer, 1997). The interpretation and derivation of the negative binomial as a Poisson-gamma mixture is well described in Cameron and Trivedi (1998). Despite its reported limitations (see Lord, 2006 and references herein), the NB regression model is still very popular, especially since all statistical software programs have built-in functions that can handle such models (Hilbe, 2007).

Unfortunately, the continuous parametric mixing distributions assumed in the Poisson mean rate may pose limitations in fitting the data, especially, with a small sample size or low sample mean value (Lord, 2006; Park and Lord, 2008). Furthermore, the choice of a particular distribution imposes a restrictive functional form between the mean and variance (e.g. quadratic relationship in the NB model), and is difficult to justify in practice because there is no *a priori* reason why the empirical frequency of crash data should be well approximated by that particular distribution, as discussed above.

In addition, the continuous mixed Poisson models usually estimate a common parameter vector ($\boldsymbol{\beta}$) and inverse dispersion parameter (ϕ) for all the cross-sections (Ramawamy et al., 1994). In other words, they are estimated at the aggregate level with one standard probability distribution function, which can mask the possibility of heterogeneity in the coefficients of the covariates across the sites. As discussed above, since the crash data are generally collected from various geographic, environmental and geometric design contexts, there are many reasons to expect the different effects of each variable on the crash occurrence. To capture unobserved heterogeneity in these parameters, one can first classify the data based on some criteria, and then apply several Poisson or negative binomial regression models at a disaggregate level. However, there may be some arbitrariness involved in the criteria dividing the groups. Gelman et al. (2004, p. 467) warned that this type of crude analysis completely ignores the uncertainty in the dividing indicators and thus can overestimate the differences between each model.

2.2. Finite mixture model

The finite mixture model allows for extremely flexible modeling of heterogeneous data because it incorporates a combination of discrete and continuous representation of population heterogeneity. Accordingly, finite mixture models have been widely applied in many areas, such as biology, biometrics, genetics, medicine, and marketing. For a comprehensive list of the applications and numerical derivations of finite mixture models, readers are referred to McLachlan and Peel (2000) and Frühwirth-Schnatter (2006).

The random vector $\mathbf{y} = (y_1, y_2, \ldots, y_N)'$ is said to arise from a finite mixture distribution, if the probability density function $p(\mathbf{y})$ of this distribution has the following form:

$$p(\mathbf{y}|\mathbf{\Theta}) = w_1 f_1(\mathbf{y}|\boldsymbol{\theta}_1) + w_2 f_2(\mathbf{y}|\boldsymbol{\theta}_2) + \cdots + w_K f_K(\mathbf{y}|\boldsymbol{\theta}_K) \tag{2}$$

where $\mathbf{\Theta} = ((\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K)', \mathbf{w})$ denotes the vector of all parameters, and $\mathbf{w} = (w_1, w_2, \ldots, w_K)'$ is called a weight distribution whose elements are restricted to be positive and sum to unity $\left(w_k > 0 \text{ and } \sum w_k = 1\right)$. A single density $f(\cdot|\boldsymbol{\theta}_k)$ is referred to as the component distribution for component k ($k = 1, 2, \ldots, K$), and K is the number of components. In most applications, it is assumed that all component distributions arise from the same parametric distribution family, $f(\cdot|\boldsymbol{\theta}_k)$. In our case, it is a Poisson or a NB distribution.

The mean and the variance are given, respectively, by

$$\mu = E(\mathbf{y}|\mathbf{\Theta}) = \sum_{k=1}^{K} \mu_k w_k \tag{3}$$

$$\sigma^2 = Var(\mathbf{y}|\mathbf{\Theta}) = \sum_{k=1}^{K} (\mu_k^2 + \sigma_k^2) w_k - \mu^2 \tag{4}$$

provided that the component moments $\mu_k = E(\mathbf{y}|\boldsymbol{\theta}_k)$ and $\sigma_k^2 = Var(\mathbf{y}|\boldsymbol{\theta}_k)$ exist (Frühwirth-Schnatter, 2006).

With this formulation, the heterogeneity in the data can be accounted for in two ways. First, it accounts for the population heterogeneity by choosing a finite number of unobserved latent components, each of which may be regarded as a sub-population. This is a discrete representation of heterogeneity in the data since the mean event rate is approximated by a finite number of support points. In this respect, the finite mixture model assumes that there is more than one component in the dataset. If such a distinct difference is not observed from modeling using Eq. (2), in other words, if the probability density function does not take the stated mixture density, the resulting parameter estimates would be very unstable and inaccurate. In such cases, it is possible to choose the traditional regression model (Poisson or NB regression model) that does not account for the heterogeneity due to the existence of different sub-populations. This can be done by setting $K = 1$ in Eq. (2). Second, depending on the choice of the component distribution, $f(\cdot|\boldsymbol{\theta}_k)$, it can also accommodate heterogeneity within each component. For example, for FMP-K and FMNB-K regression models, the heterogeneity within each component is accounted for by including the explanatory variables in the mean event rate function. Using the NB distribution as a component distribution would explain additional over-dispersion within component not captured by those explanatory variables. Thus, the formulation is flexible enough to allow for both between-component and within-component variations. It should be noted that the finite mixture approach does not require any distributional assumptions for the mixing variable.

The general setups for FMP-K and FMNB-K regression models can be extended from Eq. (2) and their means and variances are obtained from Eqs. (3) and (4).

The FMP-K regression model assumes that the marginal distribution of y_i follows a mixture of Poisson distributions:

$$p(y_i|\mathbf{x}_i, \mathbf{\Theta}) = \sum_{k=1}^{K} w_k \, Pois(\mu_{k,i}) = \sum_{k=1}^{K} w_k \left(\frac{e^{-\mu_{k,i}} (\mu_{k,i})^{y_i}}{y_i!} \right) \tag{5}$$

$$E(y_i|\mathbf{x}_i, \mathbf{\Theta}) = \sum_{k=1}^{K} \mu_{k,i} w_{k,i} \tag{6}$$

$$\mathrm{Var}(y_i|\mathbf{x}_i, \mathbf{\Theta}) = E(y_i|\mathbf{\Theta}) + \left(\sum_{k=1}^{K} w_k \mu_{k,i}^2 - E(y_i|\boldsymbol{\Theta})^2 \right) \tag{7}$$

where $\mu_{k,i} = \exp(\mathbf{x}_i \boldsymbol{\beta}_k)$ and $\mathbf{\Theta} = \{(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K), \mathbf{w}\}$. It can be readily seen that unless all the component's means are the same ($\mu_{1,i} = \cdots = \mu_{K,i}$), the variance is always greater than the mean.

For the FMNB-K regression model, it is assumed that the marginal distribution of y_i follows a mixture of negative binomial distributions:

$$\begin{aligned} p(y_i|\mathbf{x}_i, \mathbf{\Theta}) &= \sum_{k=1}^{k} w_k NB(\mu_{k,i}, \phi_k) \\ &= \sum_{k=1}^{K} w_k \left[\frac{\Gamma(y_i + \phi_k)}{\Gamma(y_i + 1)\Gamma(\phi_k)} \left(\frac{\mu_{k,i}}{\mu_{k,i} + \phi_k} \right)^{y_i} \right. \\ &\quad \left. \times \left(\frac{\phi_k}{\mu_{k,i} + \phi_k} \right)^{\phi_k} \right] \end{aligned} \tag{8}$$

$$E(y_i|x_i, \mathbf{\Theta}) = \sum_{k=1}^{K} \mu_{k,i} w_{k,i} \tag{9}$$

$$\begin{aligned} Var(y_i|\mathbf{x}_i, \mathbf{\Theta}) &= E(y_i|\mathbf{x}_i, \mathbf{\Theta}) \\ &\quad + \left(\sum_{k=1}^{K} w_k \mu_{k,i}^2 \left(1 + \frac{1}{\phi_k} \right) - E(y_i|\mathbf{x}_i, \mathbf{\Theta})^2 \right) \end{aligned} \tag{10}$$

where $\mu_{k,i} = \exp(\mathbf{x}_i \boldsymbol{\beta}_k)$ and $\mathbf{\Theta} = \{(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K), (\phi_1, \ldots, \phi_K), \mathbf{w}\}$. In this case, even if all the component's means are the same, the variance of y_i is always greater than the mean. When ϕ_k in each component goes to infinity, the FMNB-K model is reduced to the FMP-K model. Thus, the FMNB-K models allow for additional over-dispersion within components not captured by the explanatory variables. If additional heterogeneity is present within components, the Poisson mixture model is misspecified. An implication of such additional heterogeneity is that the standard errors are underestimated (Cameron and Trivedi, 1998).

It should be noted that the weight distribution (**w**) used in both FMP-K and FMNB-K was treated as a constant variable. The constant weight model can be extended to a more generalized model by parameterizing the weight distribution as a function of covariates (Wang et al., 1998; Scaccia and Green, 2003; Frühwirth-Schnatter and Kaufmann, 2006). This parameterization allows each observation to have a different weight that is dependent on the covariates, similar to the application of the varying dispersion parameter for the standard negative binomial model (see, e.g. Miaou and Lord, 2003; Lord and Park, 2008). Unfortunately, the use of varying weight factors was beyond the scope of this study, since the estimation process is very complex (i.e. there are various link functions that can be used to define the varying weights), and may not always provide the best modeling approach (Frühwirth-Schnatter, 2006).

It is noteworthy that the finite mixture of regression models as defined in Eq. (5) or (8) embrace the zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB) regression models as a special case (Cameron and Trivedi, 1998); see Lord et al. (2005, 2007) for a discussion about their use in highway safety. This can be obtained by setting $K = 2$ and $\mu_{1,i} = 0$ for all i. However, the generalized two-component mixture model does not make this somewhat strict dual-state process assumption and allows mixing with respect to both zeroes and positives. The group separation is characterized by low mean with low variance and high mean with high variance. Recently, Malyshkina et al. (2009) demonstrated a superior statistical fit of two-state Markov switching negative binomial models (with fixed weights) using time series crash data in Indiana interstate highway segments. Therefore, the FMP or FMNB models are expected to improve the goodness-of-fit relative to the conventional one-component NB model even when the sample mean is very low although this still needs to be verified in the future.

3. Model estimation method

Estimating a mixture models is not an easy task. Traditionally, the expectation-maximization (EM) algorithm has been most commonly applied based on the work of Dempster et al. (1977), who realized that a finite mixture model may always be expressed in terms of an incomplete data problem by introducing the allocations as missing data. However, there are several drawbacks in using likelihood approach (McLachlan and Peel, 2000; Frühwirth-Schnatter, 2006). The EM algorithm tends to lead to a local maximum and thus a grid of many different starting points is needed for finding the global maximum. It is also well known that the sample size has to be very large because the maximum likelihood method is based on the asymptotic theory. Furthermore, the calculation of standard errors is not straightforward when the likelihood function has unusual features. For the discussion of comparing various estimation methods for the finite mixture models, see Frühwirth-Schnatter (2006, pp. 49–56).

In this paper, a Bayesian sampling approach was adopted which provides much richer inference than the maximum likelihood approach in that it can address the issue of parameter uncertainty via full posterior distribution. Following the work of Diebolt and Robert (1994) (data augmentation and Gibbs sampling), Bayesian mixture models can be applied routinely when the number of components is assumed to be known. According to Richardson and Green (1997), Bayesian method is the only sensible way if the number of mixture components is allowed to vary. However, this paper does not intend to address the mixture models with a varying number of components. Estimating the number of components (K) is a special kind of model choice problem, for which there is a number of possible solutions: using information-based criteria; Bayes factors; reversible jump MCMC; and birth-and-death process, etc. The first two methods were adopted in this paper because they are relatively easy to implement and therefore widely used. The last two methods assume that K is not fixed but variable, and is estimated within the modeling process, which was beyond the scope of this paper. Interested readers are referred to Richardson and Green (1997) and Stephens (2000a). Instead, to determine the number of components in the mixture, a series of models with increasing numbers of components are fitted and then the most plausible model is selected by various model selection criteria which will be described later.

3.1. Data augmentation and Gibbs sampling

The algorithm for the data augmentation and Gibbs sampling consists of three steps: first, the data are augmented with a latent random variable $\mathbf{z}_i = (z_{1,i}, z_{2,i}, \ldots, z_{K,i})'$ which indicates the component membership of site i; second, conditional on $\mathbf{z}_i$, the component parameters are drawn sequentially from the full conditional posterior distributions; third, conditional on knowing the component parameters, each component indicator vector $\mathbf{z}_i$ is drawn from a multinomial distribution, satisfying $\sum_{k=1}^{K} z_{k,i} = 1$. For details about the data augmentation and Gibbs sampling, see Dempster et al. (1977) and Diebolt and Robert (1994).

For the FMNB-K regression model, it is assumed that the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \phi_1, \ldots, \phi_K$ and $\mathbf{w}$ are, *a priori*, mutually independent. As priors for the regression coefficient $\boldsymbol{\beta}_k$ and for the inverse dispersion parameter ϕ_k, multivariate normal distribution and gamma distribution were specified, respectively. For the weight distribution $\mathbf{w}$, the *Dirichlet* $(e_0, \ldots, e_0)$ was used as a prior. With these prior settings, the posterior distribution of $\mathbf{w}$ follows the *Dirichlet* $(e_0 + n_1, \ldots, e_0 + n_K)$, where $n_k = \sum_{i=1}^{N} z_{k,i}$ denotes the number of observations allocated to component k. However, the conditional distributions for $\boldsymbol{\beta}_k$ and ϕ_k do not belong to any standard distribution family. This paper used the Random-Walk Metropolis algorithm with a normal distribution as a proposal density. The acceptance rates were tuned to lie between 25% and 45%. This range of acceptance rates is generally accepted for Metropolis algorithms (Roberts, 1996). The Software R (R Development Core Team, 2006) was used for coding the algorithm.

3.2. Model selection

To determine the appropriate model and the number of components, various model selection criteria were examined: information-based criteria (AIC, BIC, and DIC) and Bayes factor via marginal likelihoods. The Akaike information criterion, or AIC is defined as $-2LL + 2p$, where p is the number of parameters in the model. It penalizes the models by the number of parameters included. Smaller values represent better overall fit. The Bayesian information criterion, or BIC is uses a penalty term of $p\log(n)$, where n is the total number of observation (868 in this case study). The BIC is more conservative than the AIC by requiring a greater improvement in fit before it will accept a more complex model (Burnham and Anderson, 2002). The Deviance information criterion, or DIC is defined as $\hat{D} + 2(\bar{D} - \hat{D})$, where $\bar{D}$ is the average of the deviance $(-2LL)$ over the posterior distribution, and $\hat{D}$ is the deviance calculated at the posterior mean parameters. As with AIC and BIC, DIC uses $p_D = \bar{D} - \hat{D}$ (effective number of parameters) as a penalty term on the goodness-of-fit. Differences in DIC from 5 to 10 indicate that one model is clearly better (Spiegelhalter et al., 2002).

Formal Bayesian model assessment is based on the Bayes factor, B_{12}, for comparing model M_1 to model M_2 after observing the data (Lewis and Raftery, 1997). The Bayes factor is the ratio of the marginal likelihoods of the two models being compared $(B_{12} = p(\mathbf{y}|M_1)/p(\mathbf{y}|M_2))$. However, in practice, computing Bayes factors for a particular set of models can be demanding because it requires either complicated multidimensional integrals or some kind of stochastic sampling from the prior distribution. For calculating the marginal likelihood, we adopted the method developed by Lewis and Raftery (1997), who suggested using the posterior simulation output for the computation of the marginal likelihoods (so-called Laplace-Metropolis estimator). The approximation of the marginal likelihood is carried out on the logarithmic scale such as

$$\log\{p(\mathbf{y}|M)\} \approx \frac{d}{2}\log(2\pi) + \frac{1}{2}\log\{|\mathbf{H}^*|\} + \log\{f(\mathbf{y}|\boldsymbol{\Theta}^*)\} + \log\{\pi(\boldsymbol{\Theta}^*)\} \tag{11}$$

where d is the number of parameters, $\log\{f(\mathbf{y}|\boldsymbol{\Theta}^*)\}$ is the log-likelihood of data at $\boldsymbol{\Theta}^*$, and $\log\{\pi(\boldsymbol{\Theta}^*)\}$ is the log-likelihood of prior distribution at $\boldsymbol{\Theta}^*$. One way of estimating $\boldsymbol{\Theta}^*$ is to find the value of $\boldsymbol{\Theta}$ at which $\log\{f(y|\boldsymbol{\Theta}^*)\} + \log\{\pi(\boldsymbol{\Theta}^*)\}$ achieves its maximum from the posterior simulation output. $|\mathbf{H}^*|$ is the determinant of the variance–covariance matrix estimated from the Hessian at the posterior mode, and it is asymptotically equal to the posterior variance–covariance matrix. This can be estimated from the sample variance–covariance matrix of the posterior simulation output. Assuming that the prior probabilities for the competing models are equal, B_{12} is expressed as follows:

$$\log(B_{12}) = \log\{p(\mathbf{y}|M_1)\} - \log\{p(\mathbf{y}|M_2)\} \tag{12}$$

According to Kass and Raftery (1995), the values between 3 and 20 are positive and the values between 20 and 150 are strong in support of model 1.

4. Data and mean functional form

To test the applicability of the proposed model, data collected in 1995 at urban 4-legged signalized intersections in Toronto, Canada were used. There are two main reasons for selecting this dataset for this study. First, the data have been used extensively for various study purposes and has been found to be of relatively good quality (Miaou and Lord, 2003; Lord et al., 2008; Lord, 2000; Persaud et al., 2002). Second, more importantly, despite many factors that may have influenced crash occurrences around and within intersections, many transportation safety analysts have often favored using traffic flow-only models over models with covariates, even though the former models may be affected by the omitted variables bias (Hauer, 1997; Persaud et al., 2001). They are often preferred over models that include several covariates because they can be easily re-calibrated when they are developed in one jurisdiction and applied to another (Persaud et al., 2002; Lord and Bonneson, 2005).

As initially discussed by Miaou and Lord (2003) and later confirmed by Mitra and Washington (2007), the un-modeled heterogeneity across sites might be structured spatially in some way, especially when a limited number of covariates are used in the model. This study speculates that part of the heterogeneity could come from the existence of the several different sub-populations. There are many evidences to support this speculation in the data. For example, the data were collected across different business environment (e.g. shopping centers, schools, office compounds, etc.). The data contain a mix of fixed and actuated traffic signals with permissive, semi-protected, and protected left turns. It also includes divided and undivided approaches with different speed limits and different number of approaching lanes. Therefore, once the mixture model is estimated, one can go back to the data and see if there are common traits among the different observations that have separated the dataset (if they are known).

The summary statistics are provided in Table 1. It contains 868 intersections, which have a total of 10,030 reported crashes. Individual intersections experienced crashes from 0 to 54 crashes. Entering traffic volumes vary widely from intersections to intersections: from about 5469 to 72,178 vehicles/day for major approaches and from 53 to about 42,644 vehicles/day for minor approaches. For a detailed description of the dataset, the readers are referred to Lord (2000).

Similar to previous research that made use of this dataset, the mean functional form for each component was the following:

$$\mu_{k,i} = \beta_{k,0} F_{1i}^{\beta_{k,1}} F_{2i}^{\beta_{k,2}} \tag{13}$$

where $\mu_{k,i}$ is the kth component's estimated number of crashes for intersection i; F_{1i} the entering flows in veh/day from the major approaches at intersection i; F_{2i} the entering flows in veh/day from the minor approaches at intersection i; and, $\boldsymbol{\beta}_k = (\beta_{k,0}, \beta_{k,1}, \beta_{k,2})'$ the estimated regression coefficients for component k.

Table 1
Summary statistics for application dataset.

	Min.	Max.	Average	Standard deviation
Crashes	0	54	11.56	10.02
Major-approach AADT	5,469	72,178	28044.81	10660.39
Minor-approach AADT	53	42,644	11010.18	8599.40

Table 2
Posterior means, standard deviations, 95% credible intervals for Poisson mixtures.

Poisson mixtures	w	$\mathrm{Ln}(\beta_0)$	β_1	β_2	ϕ
Single component					
Estimate	1.0	−10.2294	0.6023	0.7038	–
(Std. Dev.)		(0.2837)	(0.0287)	(0.0139)	
(2.5%)		(−10.7860)	(0.5471)	(0.6765)	
(97.5%)		(−9.6770)	(0.6589)	(0.7311)	
FMP-2					
Component 1					
Estimate	0.492	−9.3231	0.6107	0.6286	–
(Std. Dev.)	(0.038)	(0.4564)	(0.0452)	(0.0219)	
(2.5%)	(0.418)	(−10.2046)	(0.5230)	(0.5842)	
(97.5%)	(0.566)	(−8.4356)	(0.6993)	(0.6713)	
Component 2					
Estimate	0.508	−11.4200	0.6031	0.7871	–
(Std. Dev.)	(0.038)	(0.6679)	(0.0639)	(0.0337)	
(2.5%)	(0.434)	(−12.7555)	(0.4780)	(0.7213)	
(97.5%)	(0.581)	(−10.1314)	(0.7309)	(0.8552)	
FMP-3					
Component 1					
Estimate	0.114	−5.3116*	0.1021*	0.6238	–
(Std. Dev.)	(0.028)	(2.9760)	(0.2516)	(0.1930)	
(2.5%)	(0.066)	(−11.0566)	(−0.4148)	(0.1764)	
(97.5%)	(0.178)	(0.4982)	(0.5684)	(0.9731)	
Component 2					
Estimate	0.329	−9.0683	0.6190	0.6018	–
(Std. Dev.)	(0.049)	(0.5763)	(0.0559)	(0.0279)	
(2.5%)	(0.229)	(−10.1907)	(0.5114)	(0.5465)	
(97.5%)	(0.424)	(−7.9200)	(0.7300)	(0.6556)	
Component 3					
Estimate	0.557	−11.0158	0.6139	0.7558	–
(Std. Dev.)	(0.044)	(0.6530)	(0.0625)	(0.0352)	
(2.5%)	(0.467)	(−12.3389)	(0.4905)	(0.6920)	
(97.5%)	(0.640)	(−9.7712)	(0.7372)	(0.8308)	

* Not significant at 5% significance level.

5. Model results

This section describes the results of the analysis. We first fitted the data with increasing number of components; for Poisson mixtures (FMP), models with K = 2, 3, 4 were estimated, and for NB mixtures (FMNB), models with K = 2, 3 were fitted. For each model, total 3×10^5 MCMC iterations were used, keeping every 10th samples, and half the iterations were discarded (burn-in period). From the remained 15,000 samples, the posterior means, standard deviations and 95% credible intervals were calculated.

Table 3
Posterior means, standard deviations, 95% credible intervals for NB mixtures.

NB mixtures	w	$\mathrm{Ln}(\beta_0)$	β_1	β_2	ϕ
Single component					
Estimate	1.0	−10.2300	0.6190	0.6854	7.0894
(Std. dev.)		(0.4659)	(0.0459)	(0.0216)	(0.6156)
(2.5%)		(−11.1707)	(0.5296)	(0.6428)	(5.9590)
(97.5%)		(−9.3241)	(0.7118)	(0.7273)	(8.3760)
Two component					
Component 1					
Estimate	0.430	−10.9407	0.8588	0.5056	9.3692
(Std. dev.)	(0.153)	(1.3641)	(0.1595)	(0.0812)	(1.6220)
(2.5%)	(0.150)	(−13.8766)	(0.5991)	(0.3199)	(6.8739)
(97.5%)	(0.731)	(−8.3865)	(1.2297)	(0.6384)	(12.9768)
Component 2					
Estimate	0.570	−9.7842	0.3987	0.8703	8.2437
(Std. dev.)	(0.153)	(1.0447)	(0.1289)	(0.0782)	(1.3502)
(2.5%)	(0.268)	(−11.8434)	(0.1116)	(0.7445)	(6.0746)
(97.5%)	(0.849)	(−7.6601)	(0.6181)	(1.0497)	(11.2873)

Table 4
Model selection criteria.

Models	No. of parameters	LL	AIC	BIC	DIC	$\log p(\mathbf{y}\|M_k)$
Standard Poisson	3	−2791.5	5587.1	5596.6	5589.3	−2811.494
FMP-2	7	−2560.4	5134.7	5168.1	5134.6	−2598.153
FMP-3	11	−2529.5	5081.1	5133.5	5080.0	−2582.033
FMP-4	15	−2519.7	5069.4	5141.3	5060.4	−2583.296
Standard NB	4	−2534.6	5077.3	5096.3	5077.3	−2559.311
FMNB-2	9	−2525.9	5069.8	5112.6	5068.7	−2569.193
FMNB-3	14	−2518.6	5065.2	5131.9	5060.6	−2570.023

Convergence was checked by monitoring the trace plots of the samples, marginal posterior distributions of model parameters and the autocorrelations.

The results for the Poisson mixtures (for $K = 2$ and 3) are given in Table 2 along with the standard Poisson regression model (one single distribution). The results for the NB mixtures ($K = 2$) and the standard NB regression model are provided in Table 3. The results for FMP-4 and FMNB-3 are not presented in Tables 2 and 3, because it was noticed that the marginal densities for all parameters were very wide with multiple modes and non-vanishing high autocorrelations. This indicated difficulties with convergence and the nonidentifiability of the models. Nonidentifiability of finite mixtures of regression models is caused not only by the invariance of a finite mixture distribution to relabeling the components (known as "label switching problem"), but also by potential over-fitting. When the models are unidentified, it is meaningless to draw inference directly from MCMC output using ergodic averaging (Frühwirth-Schnatter, 2006).

Table 4 shows the computed values of log-likelihood, AIC, BIC and DIC for each model. Log-likelihood, indicated as LL, was obtained from the greatest values from the posterior simulation output for the individual models. With the exception of the AIC criteria, none of the evaluation measures showed that Poisson mixtures performed better than the standard negative binomial regression.

Based on the DIC criteria, FMP-4 or FMNB-3 appears to be the best model. However, for these two models, the MCMC output displayed non-vanishing autocorrelation with extremely high lags and the marginal distributions of the model parameters showed multiple modes, indicating difficulties with convergence and label switching problems. Thus, the DIC values from such models are not reliable and should not be used to compare models (Spiegelhalter et al., 2002). Instead, FMNB-2 model could be selected as the best model. As shown in the trace plots in Fig. 1 (left-hand side), although there was evidence of label switching at the initial part of the simulation, after some period the samples appear to be performing well. It can be seen that the component-wise regression parameters are well separated. The marginal posterior distributions of regression parameters (after a burn-in period) in Fig. 1 (right-hand side) are very close to a normal distribution. For the ϕ parameters, it was observed that the occasional very large samples skewed the marginal posterior distribution to the right with a relatively long tail.

Using the BIC and Bayes factor (marginal likelihood) criteria, the standard NB regression model could be selected as the best model. This suggests that FMNB-2 model did not make a significant improvement in terms of fit by adding additional parameters. Actually, for this particular dataset, the standard NB regression model itself produced a very satisfactory goodness-of-fit. This is why these criteria favored the simpler model. When the predicted crash mean values for each intersection were compared between the two models, there were little differences by showing almost same predictive capabilities. However, when the variances were compared, they were significantly reduced in FMNB-2 model, especially for those intersections with high mean values.

Table 5
Summary statistics for each group.

	Min.	Max.	Average	Standard deviation
Component 1				
Crashes	1	48	11.61	7.85
Major-approach AADT	5,967	56,623	28724.45	10464.09
Minor-approach AADT	53	30,824	7147.26	6173.03
Component 2				
Crashes	0	54	11.54	10.53
Major-approach AADT	5,469	72,178	27861.98	10712.77
Minor-approach AADT	823	42,644	12049.33	8863.41

In summary, the four model selection criteria used in this research produced conflicting conclusions, especially, between the FMNB-2 and the standard NB models (if we exclude FMNB-3). As described above, it was confirmed that the BIC and Bayes factor are more conservative than AIC and DIC criteria (see Burnham and Anderson, 2002). Even though the FMNB-2 was not supported by all selection criteria, the parameter estimate results (Fig. 1 and Table 3) suggest that there is something going on in the dataset that remains unexplained by the standard NB model. Thus, in this case, the FMNB-2 should be the recommended model, since it provides additional information about the data. This is described in the next paragraph.

Based on the FMNB-2 model, to study the effects of the traffic flow, the data were classified into two groups by assigning each site to the component with the highest posterior probability. When the component proportions $\mathbf{w}$ is given, the posterior probability that the observation y_i belongs to one component is given by the following equation by Bayes' rule (Frühwirth-Schnatter, 2006):

$$\Pr(z_{k,i}|\boldsymbol{\Theta}, \mathbf{x}_i, y_i) = \frac{p(y_i|\mathbf{x}_i, \boldsymbol{\beta}_k, \phi_k) \cdot w_k}{\sum_{j=1}^{K} p(y_i|\mathbf{x}_i, \boldsymbol{\beta}_j, \phi_j) \cdot w_j} \propto p(y_i|\mathbf{x}_i, \boldsymbol{\beta}_k, \phi_k) \cdot w_k, \quad k = 1, \ldots, K \tag{14}$$

The calculated posterior probabilities of components 1 and 2 were 21.2% (184 obs.) and 78.8% (684 obs.), respectively, and they were grouped accordingly. Table 5 shows the summary statistics for each group. Compared with Table 1, there is a striking difference in the average value of minor road approaching AADT. Most of low minor-approach AADT was assigned to component 1 resulting in a low average value, and many of high minor-approach AADT is associated with component 2. Therefore, it is obvious that the variability in minor-approach AADT is one source of over-dispersion along with other unobserved variables.[2] The next step would be to examine each group more closely to see if they have common characteristics, such as traffic light phasing scheme, intersection geometry, lane configuration or unique geographical locations among others. Unfortunately, due to the

[2] Another possible explanation is that if we had formulated the FMNB-2 model with a variable weight by including the minor-approach AADT as a covariate, it might have improved the model. We thank a reviewer for this suggestion.

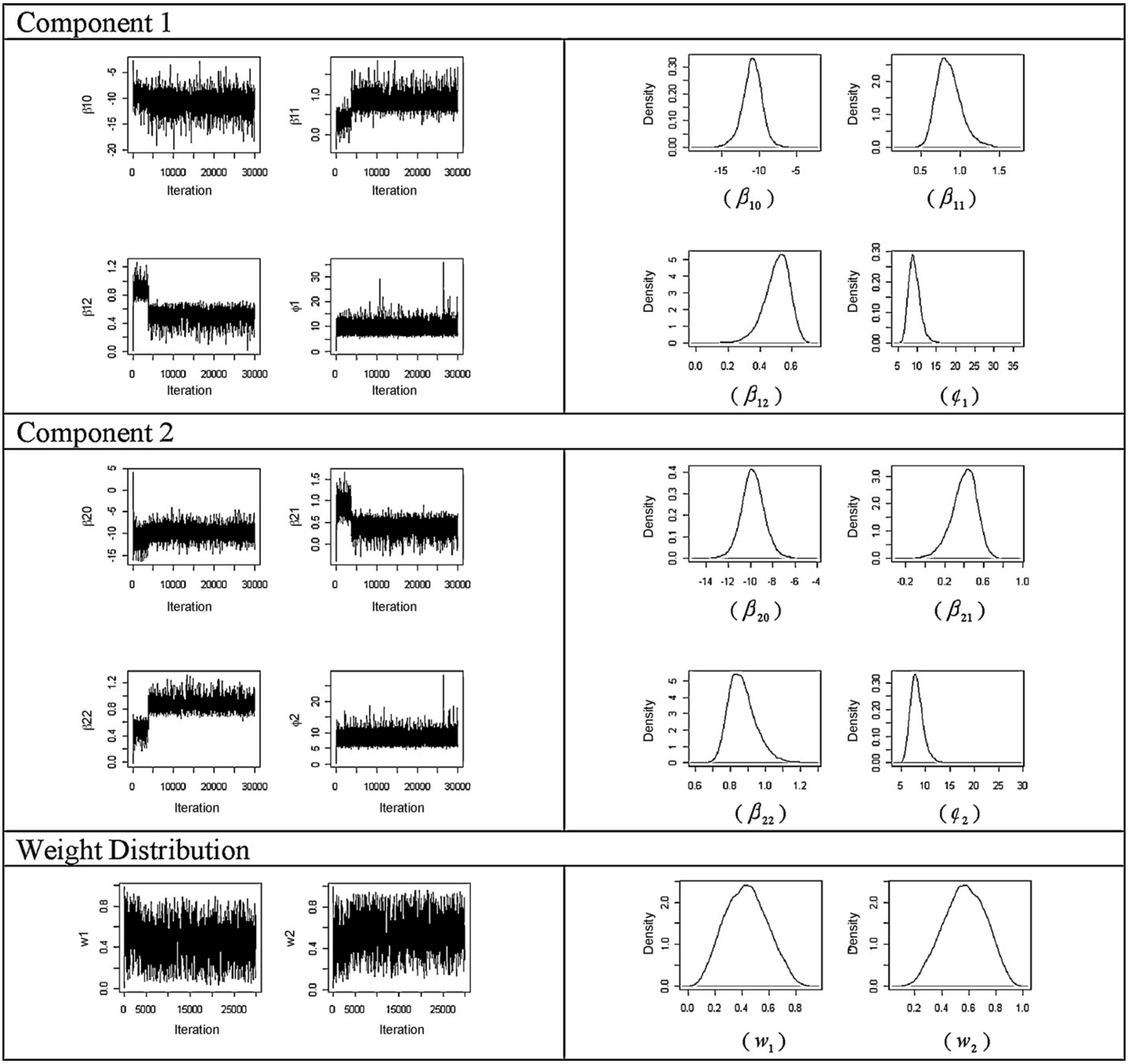

Fig. 1. MCMC trace plots and marginal posterior distributions.

age of the dataset, we are not able to get access to this kind of information.

6. Summary and discussion

This paper has proposed an alternative formulation that could capture the heterogeneity in crash count models through the use of finite mixture regression models. Seven models were estimated using data collected at signalized 4-legged intersections in Toronto, Ont. Overall, the results show that the standard NB and the FMNB-2 regression models performed almost similarly; each one being favored over the other based on some of the four evaluation criteria. Despite this mixed assessment, the parameter estimates of the FMNB-2 clearly show that important characteristics observed in the data are not captured by the standard NB model.

Given this modeling outcome, two points needs to be discussed. First, it seems that the dataset may have been generated from two distinct sub-populations, with each population having its own regression coefficients and degrees of over-dispersion. Although over-dispersion in crash data can be dealt with in a variety of ways, the mixture model provided the nature of the over-dispersion in the data. This result partly supports those found by Miaou and Song (2005) who used the same data. They showed that the inclusion of a spatial effect (induced by omitted variables) in the model could significantly improve the overall goodness-of-fit of the model. Second, from the post analysis of the mixture model, we could partly associate the source of over-dispersion with the minor-approach AADT. This is important because those intersections with different characteristics of minor approach flows may warrant different interventions to improve their safety. If additional descriptive data such as signal types, intersection geometry or number of approaching lanes are available, this finding also makes it possible for us to go back to the data and relate those variables with the separation of the dataset. This approach would be advantageous over the analysis with arbitrarily grouped intersections.

Another point that should be noted is the fact that the standard NB regression model can be misleading when the data were actually generated by two or more component finite mixtures of Poisson regression models. In line with this study, we tested this point with simulated data generated by two-component finite mixture Poisson distribution (not included in this paper). The standard NB regression model accounted for almost all over-dispersion existing in the data and provided a satisfactory goodness-of-fit. However, because of model misspecification, it failed intrinsically to capture the existence of coefficient heterogeneity among components and it had a very poor prediction performance as compared to the FMP-2. This supports the idea that the analysts should consider the application of the finite mixture models when the data are suspected to comprise observations from several sub-populations.

Despite many advantages of the finite mixture models, there are still several unresolved issues especially in terms of model parameter estimation through a Bayesian approach. First, because of the nonidentifiability caused by the invariance to relabeling the components, numerous authors suggested different approaches for relabeling the MCMC draws; imposing constraints on the parameters (Richardson and Green, 1997), clustering methods (Stephens, 2000b), or random permutation (Frühwirth-Schnatter, 2006). Judging from the literature review in this area, it seems that there is no consensus on using a unique method yet. This paper, instead, excluded the unidentified models (i.e. FMP-4 or FMNB-3) and confined the analyses to the models whose label switching problem can be easily corrected by inspecting the MCMC trace plots. This is an obvious limitation of this paper and further work should be carried out.

Future work also includes the examining the performance of the finite mixture models when the data have small sample mean values and small sample sizes. Since the estimate of the dispersion parameter is significantly biased under this situation, the results from the finite mixture models can provide an alternative estimates for the dispersion parameters for each component. Regarding the mean functional form (Eq. (13)), it is quite simplistic and clearly has many important missing variables that are known to influence crash frequencies. In such case, one would expect that a more flexible modeling form suggested in this paper would naturally fit the data better. In this respect, the analysis can be extended to test the model with more varied datasets to examine whether the suggested model would work better in a more fully specified model (both for fixed and varying weights). Finally, developing a COM–Poisson mixture model may prove to be useful for analyzing motor vehicle crashes.

Acknowledgements

The authors would like to thank Dr. Fred L. Mannering and Ms. Nataliya Malyshkina from Purdue University for their helpful comments and suggestions on an earlier version of the paper.

References

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multi-Model Inference: A Practical Information Theoretic Approach. Springer-Verlag, Now York.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B 39, 1–37.

Diebolt, J., Robert, C.P., 1994. Estimation of finite mixture distributions through Bayesian sampling. Journal of the Royal Statistical Society, Series B 56, 363–375.

El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transportation Research Record 1950, 9–16.

Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer Series in Statistics. Springer, New York.

Frühwirth-Schnatter, S., Kaufmann, S., 2006. Model-based Clustering of Multiple Time Series. Research Report IFAS, http://www.ifas.jku.at.

Geedipally, S.R., Lord, D., 2008. Effects of the varying dispersion parameter of Poisson-models on the estimation of confidence interval of crash prediction models. Transportation Research Record 2061, 46–54.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd ed. Chapman & Hall/CRC.

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

Hauer, E., 1997. Observational Before–After Studies in Road Safety. Pergamon Press/Elsevier Science Ltd., Oxford, UK.

Hauer, E., 2001. Overdispersion in modeling accidents on road sections and in empirical Bayes estimation. Accident Analysis and Prevention 33 (6), 799–808.

Heydecker, B.G., Wu, J., 2001. Identification of sites for road accident remedial work by Bayesian statistical methods: an example of uncertain inference. Advances in Engineering Software 32, 859–869.

Hilbe, J.M., 2007. Negative Binomial Regression. Cambridge University Press, UK.

Kass, R.E., Raftery, A.E., 1995. Bayes factors and model uncertainty. Journal of the American Statistical Association 90, 773–795.

Land, K.C., McCall, P.L., Nagi, D.S., 1996. A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models. Sociological Methods & Research 24 (4), 387–442.

Lewis, S.M., Raftery, A.E., 1997. Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. Journal of the American Statistical Association 92, 648–655.

Lord, D., 2000. The Prediction of Accidents on Digital Networks: Characteristics and Issues Related to the Application of Accident Prediction Models. PhD Dissertation. Department of Civil Engineering, University of Toronto, Toronto.

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention 38 (4), 751–766.

Lord, D., Bonneson, J.A., 2005. Calibration of predictive models for estimating the safety of ramp design configurations. Transportation Research Record 1908, 88–95.

Lord, D., Guikema, S.D., Geedipally, S., 2008. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes. Accident Analysis and Prevention 40 (3), 1123–1134.

Lord, D., Park, P.Y.-J., 2008. Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. Accident Analysis and Prevention 40 (4), 1441–1457.

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37 (1), 35–46.

Lord, D., Washington, S.P., Ivan, J.N., 2007. Further notes on the application of zero inflated models in highway safety. Accident Analysis and Prevention 39 (1), 53–57.

Maher, M., Summersgill, I., 1996. A comprehensive methodology for the fitting of predictive accident models. Accident Analysis and Prevention 28 (6), 281–296.

Malyshkina, N.V., Mannering, F.L., Tarko, A.P., 2009. Markov switching negative binomial models: An application to vehicle accident frequencies. Accident Analysis and Prevention 41 (2), 217–226.

McLachlan, G., Peel, D., 2000. Finite Mixture Models. John Wiley & Sons Inc.

Miaou, S.-P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. Transportation Research Record 1840, 31–40.

Miaou, S.-P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accident Analysis and Prevention 37 (4), 699–720.

Miranda-Moreno, L.F., Fu, F.F., Saccomanno, L., Labbe, A., 2005. Alternative risk models for ranking locations for safety improvement. Transportation Research Record 1908, 1–8.

Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accident Analysis and Prevention 39 (3), 459–468.

Park, B.-J., Lord, D., 2008. Adjustment for the maximum likelihood estimate of the negative binomial dispersion parameter. Transportation Research Record 2061, 9–19.

Persaud, B.N., Lord, D., Palminaso, J., 2002. Issues of calibration and transferability in developing accident prediction models for urban intersections. Transportation Research Record 1784, 57–64.

Persaud, B.N., Retting, R.A., Gårder, P.E., Lord, D., 2001. Safety effect of roundabout conversions in the United States: empirical Bayes observational before–after study. Transportation Research Record 1751, 1–8.

Ramawamy, V., Anderson, E.W., DeSarbo, W.S., 1994. A disaggregate negative binomial regression procedure for count data analysis. Management Science 40 (3), 405–417.

R Development Core Team, 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0; http://www.R-project.org (accessed January 2008).

Richardson, S., Green, P.J., 1997. On Bayesian analysis of mixtures with an unknown number of components (with discussion). Journal of the Royal Statistical Society, Series B 59, 731–792.

Roberts, G.O., 1996. Markov chain concepts related to sampling algorithms. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 45–57.

Scaccia, L., Green, P.J., 2003. Bayesian growth curves using normal mixtures with nonparametric weights. Journal of Computational and Graphical Statistics 12, 308–331.

Shankar, V.N., Albin, R.B., Milton, J.C., Mannering, F.L., 1998. Evaluating median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model. Transportation Research Record 1635, 44–48.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B 64, 583–639.

Stephens, M., 2000a. Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. The Annals of Statistics 28, 40–74.

Stephens, M., 2000b. Dealing with label switching in mixture models. Journal of the Royal Statistical Society, Series B 62, 795–809.

Wang, P.M., Cockburn, I.M., Puterman, M.L., 1998. Analysis of patent data—a mixed Poisson regression model. Journal of Business and Economic Statistics 16 (1), 27–41.

Wood, G.R., 2002. Generalized linear accident models and goodness of fit testing. Accident Analysis and Prevention 34 (1), 417–427.