Accident Analysis and Prevention 42 (2010) 741–749

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Bias properties of Bayesian statistics in finite mixture of negative binomial regression models in crash data analysis

Byung-Jung Park[a,1], Dominique Lord[a,*], Jeffrey D. Hart[b,2]

[a] Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States
[b] Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, United States

ARTICLE INFO

Article history:
Received 18 September 2009
Received in revised form 27 October 2009
Accepted 3 November 2009

Keywords:
Bias
Bayesian summary statistics
Heterogeneity
Finite mixture model
Negative binomial

ABSTRACT

Factors that cause heterogeneity in crash data are often unknown to researchers and failure to accommodate such heterogeneity in statistical models can undermine the validity of empirical results. A recently proposed finite mixture for the negative binomial regression model has shown a potential advantage in addressing the unobserved heterogeneity as well as providing useful information about features of the population under study. Despite its usefulness, however, no study has been found to examine the performance of this finite mixture under various conditions of sample sizes and sample-mean values that are common in crash data analysis. This study investigated the bias associated with the Bayesian summary statistics (posterior mean and median) of dispersion parameters in the two-component finite mixture of negative binomial regression models. A simulation study was conducted using various sample sizes under different sample-mean values. Two prior specifications (non-informative and weakly-informative) on the dispersion parameter were also compared. The results showed that the posterior mean using the non-informative prior exhibited a high bias for the dispersion parameter and should be avoided when the dataset contains less than 2,000 observations (even for high sample-mean values). The posterior median showed much better bias properties, particularly at small sample sizes and small sample means. However, as the sample size increases, the posterior median using the non-informative prior also began to exhibit an upward-bias trend. In such cases, the posterior mean or median with the weakly-informative prior provided smaller bias. Based on simulation results, guidelines about the choice of priors and the summary statistics to use are presented for different sample sizes and sample-mean values.

1. Introduction

Over-dispersion caused by unobserved heterogeneity in crash data is a serious problem and has been addressed in a variety of ways within the negative binomial (NB) modeling framework (Hauer, 2001; Heydecker and Wu, 2001; Miaou and Lord, 2003; Geedipally et al., 2009; Anastasopoulos and Mannering, 2009). However, the true factors that affect heterogeneity are often unknown to researchers and failure to accommodate such heterogeneity in the model can undermine the validity of the empirical results. Recently, a finite mixture approach has been proposed as an alternative to address the unobserved heterogeneity problem in vehicle crash data (Park and Lord, 2009). This approach provided useful information on features of the population. A similar approach by Malyshkina et al. (2009) also demonstrated the usefulness of two-state Markov switching models in terms of goodness-of-fit and the capability of capturing unobserved heterogeneity. The finite mixture regression model rests on the assumption that there are a finite number of unobservable categories of observations and the heterogeneity arises from different values of regression coefficients caused by missing variables (Frühwirth-Schnatter, 2006). It is a flexible semi-parametric model that allows one to capture heterogeneity through usually a small number of simple regression models such as Poisson or NB regression models. It has several known advantages over the standard NB regression model. First, it can effectively account for unobserved heterogeneity without imposing strong distributional assumptions on the mixing variable (for example, a continuous Gamma distribution assumed in NB model). Laird (1978) and Heckman and Singer (1984) showed that finite mixture models can provide good numerical approximations even if the underlying mixing distribution is continuous (Cameron and Trivedi, 1998). Second, it allows the vector of regression coefficients to vary from component to component. Most of the traditional models constrain the coefficients to be fixed across observations. Washington et al. (2003) noted that it could lead to

* 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), hart@stat.tamu.edu (J.D. Hart).
[1] Tel.: +1 979 862 8492; fax: +1 979 845 6481.
[2] Tel.: +1 979 845 3141; fax: +1 979 845 3144.

0001-4575/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2009.11.002

inconsistent and biased parameter estimates when the coefficients actually vary across observations. In this context, Anastasopoulos and Mannering (2009) recently applied random-parameter count models to vehicle crash data by employing a normal error term in the coefficients to allow them to vary. The finite mixture regression model is different from the random-parameter models in that the parameter heterogeneity is approximated by a finite number of support points and their probability masses without making a distributional assumption on the regression coefficients or mixing variable. The finite mixture model allows the data to determine the true relationships by choosing a finite number of unobserved latent components.

Given the potential advantages of finite mixture models in describing the heterogeneity in crash data, there is a need to conduct further research on characterizing their performance, preferably within the Bayesian framework. From an application-oriented point of view, it is important to know what sample size is necessary at minimum in order to guarantee the unbiased or bias-reduced estimates of model parameters. Another factor closely related to this problem concerns sample-mean values. Within the standard NB modeling framework, several researchers have found that the dispersion parameter in the NB model is significantly influenced by sample sizes and sample-mean values (Maher and Summersgill, 1996; Lord, 2006; Park and Lord, 2008; Lord and Miranda-Moreno, 2008). This is true for both the Frequentist and Bayesian estimates although the estimate from the latter was found more robust when the data are characterized by small sample size and small sample-mean value. However, within the finite mixture models, the necessary sample size may depend on the data at hand—that is, the sample size need not be large for well-separated data, but it can be huge for a poorly-separated case. Therefore, rather than searching for the minimum sample size, it would be better to focus on the bias and variability properties of an estimator we choose. If we adopt a Bayesian method and use its posterior summary statistics (i.e., posterior mean or median) for making posterior inference, it is important to obtain consistent posterior mean or median estimates of model parameters under repeated random sampling. Up until now no studies have analyzed the performance of the posterior summary statistics in the finite mixture regression models under various conditions of sample sizes and sample-mean values.

In this respect, the objective of this study is to investigate the potential bias and the variability in parameter estimates in the finite mixture models using various combinations of sample sizes under different sample-mean values. Specifically, we mainly investigated the bias associated with the posterior mean and median of dispersion parameters in the two-component finite mixture of negative binomial regression models (termed as a FMNB-2). This is because the FMNB-2 model was considered more useful and parsimonious than the finite mixture of Poisson regression models (termed as a FMP-K models; K represents the number of components). The FMP-K models often produce too many components (Van Duijn and Böckenholt, 1995), making it difficult to apply. For this purpose, a Monte Carlo simulation is conducted using various sample sizes under different sample-mean values. In addition, two different prior specifications for the dispersion parameters are investigated: non-informative and weakly-informative gamma priors. The results are compared in terms of the magnitude of the bias introduced by various sample sizes and sample-mean values.

This paper is divided into four sections. The first section describes a brief background on finite mixture regression models and the Bayesian estimation method. The second section provides two hypothetical examples to illustrate the disadvantage of using the standard NB models when the data were actually drawn from two components Poisson or NB distribution. In the third section, a Monte Carlo simulation study is carried out on the FMNB-2 model for different scenarios. Finally, the last section summarizes the results and concludes with a brief guideline about the choice of priors and posterior summary statistics.

2. Background

This section gives some background on finite mixture regression models and estimation of their parameters. For more details, the reader is referred to Park and Lord (2009) and the references therein.

2.1. Finite mixture regression models

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 $p(\mathbf{y})$ has the following form:

$$p(\boldsymbol{y}|\boldsymbol{\Theta}) = w_1 f_1(\boldsymbol{y}|\boldsymbol{\theta}_1) + w_2 f_2(\boldsymbol{y}|\boldsymbol{\theta}_2) + \cdots + w_K f_K(\boldsymbol{y}|\boldsymbol{\theta}_K) \tag{1}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_K)', \boldsymbol{w})$ denotes the vector of all unknown parameters, and $\boldsymbol{w} = (w_1, w_2, \ldots, w_K)'$ is a weight distribution (or mixing proportion) whose elements are restricted to be positive and sum to unity ($w_k > 0$ and $\sum w_k = 1$). In a typical cross-sectional crash data analysis, y_i represents the number of crashes per given years on roadway segment i. For model simplicity, w_k is usually treated as constant, but it can be further parameterized using observable covariates. A single density $f_k(\cdot|\boldsymbol{\theta}_k)$ is referred to as the component distribution for component $k (k = 1, 2, \ldots, K)$. 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 either a Poisson or a NB distribution.

The FMP-K regression model assumes that the marginal distribution of y_i follows a mixture of Poisson distributions. The probability density function, mean and variance of y_i are as follows:

$$p\left(y_i|\boldsymbol{x}_i, \boldsymbol{\Theta}\right) = \sum_{k=1}^{K} w_k Pois(\mu_{i,k}) = \sum_{k=1}^{K} w_k \left(\frac{e^{-\mu_{i,k}}(\mu_{i,k})^{y_i}}{y_i!}\right) \tag{2}$$

$$E(y_i|\boldsymbol{x}_i, \boldsymbol{\Theta}) = \sum_{k=1}^{K} \mu_{i,k} w_k \tag{3}$$

$$Var(y_i|\boldsymbol{x}_i, \boldsymbol{\Theta}) = E(y_i|\boldsymbol{\Theta}) + \left(\sum_{k=1}^{K} w_k \mu_{i,k}^2 - E(y_i|\boldsymbol{\Theta})^2\right) \tag{4}$$

where $\mu_{i,k} = \exp(\mathbf{x}_i \boldsymbol{\beta}_k)$ and $\boldsymbol{\Theta} = \{(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K), \boldsymbol{w}\}$. It can be readily seen that unless all the components have the same mean ($\mu_{i,1} = \ldots = \mu_{i,K}$), 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,

$$p\left(y_i|\mathbf{x}_i, \boldsymbol{\Theta}\right) = \sum_{k=1}^{K} w_k NB(\mu_{i,k}, \phi_k)$$
$$= \sum_{k=1}^{K} w_k \left[\frac{\Gamma(y_i + \phi_k)}{\Gamma(y_i + 1)\Gamma(\phi_k)} \left(\frac{\mu_{i,k}}{\mu_{i,k} + \phi_k}\right)^{y_i} \left(\frac{\phi_k}{\mu_{i,k} + \phi_k}\right)^{\phi_k}\right] \tag{5}$$

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

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

Table 1
The values for model comparison criteria should be properly positioned right under each model (i.e., NB regression and FMP-2).

Model parameters	True values		NB regression	FMP-2	
	Component 1	Component 2		Component 1	Component 2
$\beta_{0,k}$	2.0	0.0	1.1401 (0.0648)[a]	1.9825 (0.0493)	0.0173[b] (0.0622)
$\beta_{1,k}$	−0.5	0.5	−0.1107[b] (0.0600)	−0.5741 (0.0350)	0.5198 (0.0469)
$\beta_{2,k}$	0.5	−0.5	0.0861[b] (0.0553)	0.4389 (0.0354)	−0.5118 (0.0444)
ϕ_k	–	–	0.6040 (0.049)	–	–
w_k	0.2	0.8	–	0.213 (0.023)	0.787 (0.023)
Model comparison criteria[c]					
−2LL	The smaller the better		2242.3	1918.4	
AIC	The smaller the better		2250.3	1932.4	
BIC	The smaller the better		2267.1	1961.9	
DIC	The smaller the better		2250.4	1932.0	
Log(ML)	The larger the better		−1142.8	−995.5	

[a] The standard deviation of the coefficient.
[b] The coefficient whose 95% credible interval includes zero.
[c] LL represents the Log-Likelihood. AIC, BIC and DIC are the information-based criteria, and Log(ML) represents the log of marginal likelihood.

where $\mu_{i,k} = \exp(\mathbf{x}_i\boldsymbol{\beta}_k)$ and $\boldsymbol{\Theta} = \{(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K), (\phi_1, \ldots, \phi_K), \mathbf{w}\}$. In this case, even if all the components have the same mean, 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 heterogeneity within components not captured by the explanatory variables. If additional heterogeneity is present within components, the Poisson mixture model is a misspecification. A ramification of such additional heterogeneity is that the standard errors are underestimated (Cameron and Trivedi, 1998).

2.2. Parameter estimation method

The estimation of finite mixture models can be done using either maximum likelihood with the EM algorithm or a Bayesian framework coupled with data augmentation and Gibbs sampling techniques. We adopted the second approach. Among many advantages of using a Bayesian method, the most appropriate one for this study is that we can simply choose the best posterior summary statistic that minimizes the bias. This obviates an additional correction process.

The algorithm for the data augmentation and Gibbs sampling consists of three steps: first, the data are augmented with a latent random variable $\mathbf{z}_i = (z_{i,1}, z_{i,2}, \ldots, z_{i,K})'$ which indicates the component membership of site i; second, conditional on knowing $\mathbf{z}_i$, 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 $\sum_{k=1}^{K} z_{i,k} = 1$. For more details about the data augmentation and Gibbs sampling, see Dempster et al. (1977) and Diebolt and Robert (1994). In the second step, prior distributions have to be defined for each component parameter. For example, 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 prior for the regression coefficient $\boldsymbol{\beta}_K$, a non-informative multivariate normal distribution with mean = 0 and large variance is specified. As prior for the inverse dispersion parameter ϕ_k, two choices are examined in the simulation section: a non-informative $\Gamma(0.01, 0.01)$ prior and a weakly-informative $\Gamma(0.5, 0.1)$ prior. The notation $\Gamma(a, b)$ indicates the gamma distribution with mean $E(\phi) = a/b$ and variance $V(\phi) = a/b^2$. For the weight distribution $\mathbf{w}$, the *Dirichlet* $(e_0, \ldots, e_0)$ is used as a prior. Under this prior specification, the posterior distribution of $\mathbf{w}$ is *Dirichlet* $(e_0 + n_1, \ldots, e_0 + n_K)$, where $n_k = \sum_{i=1}^{N} z_{i,k}$ 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. Thus, 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% to induce good mixing behavior. The Software R (R Development Core Team, 2006) was used for coding the algorithm.

The optimal number of components (K) is usually determined in two ways. One approach is to fit a series of models with a fixed number of components, and then the best model is selected based on various model selection criteria such as information-based criteria or a Bayes factor via marginal likelihood. Another approach is to treat K as an unknown parameter and to estimate it within the modeling process. While the second approach appears more appealing, there are some issues with regard to prior selection for K and the sensitivity of its posterior distribution (Aitkin, 2001; Jasra et al., 2005). For this study, we adopted the first approach since the number of components is pre-specified.

3. Hypothetical examples

The purpose of this section is to illustrate the appropriateness of the mixture model specifications when the data were actually drawn from two components Poisson or NB distribution. Working with the hypothetical examples is effective in illustrating the theoretical aspects of the finite mixture models in that we can generate and analyze a random sample with known characteristics. We will show how effectively the finite mixture regression models can capture the sub-populations, and thereby emphasize the disadvantage of using the standard NB regression model in such situations.

3.1. Example 1

A dataset is drawn from a two-component finite mixture of Poisson regression models. This example is meant to show how poor the prediction capability of the standard NB model will be because of model misspecification. Since the standard NB model estimates a single set of regression coefficients, the interpretation of its estimates is misleading if the population is heterogeneous with respect to the impact of explanatory variables. For a comparison with FMP-2 model, the NB model was selected rather than the Poisson model because the latter does not accommodate the over-dispersion intrinsically.

For generating the FMP-2 random variates, two covariates, $\mathbf{x}_i = (1, x_{i1}, x_{i2})$, were introduced which are randomly generated from the standard normal distribution. The Poisson mean rates $\mu_{i,1}$ and $\mu_{i,2}$ were then constructed from the independent variables by assuming a log-linear relationship using known regression coeffi-

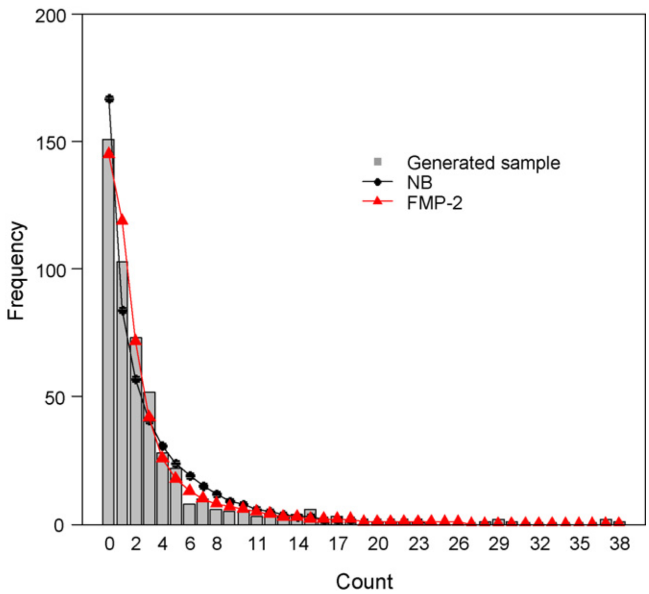

Fig. 1. Goodness-of-fit comparison between NB and FMP-2.

cients $\boldsymbol{\beta}_1=(\beta_{0,1}, \beta_{1,1}, \beta_{2,1})'$ and $\boldsymbol{\beta}_2=(\beta_{0,2}, \beta_{1,2}, \beta_{2,2})'$. This results in each component mean rate $\mu_{i,1}=\exp(\mathbf{x}_i\boldsymbol{\beta}_1)$ and $\mu_{i,2}=\exp(\mathbf{x}_i\boldsymbol{\beta}_2)$, respectively. Furthermore, a binary value (0 or 1) was generated with probability w from the *Binomial* $(1, w)$ distribution for each site i and the values saved as z_i. Then, the FMP-2 random variate y_i can be generated from $z_i \cdot Poisson(\mu_{i,1})+(1-z_i) \cdot Poisson(\mu_{i,2})$. In this manner, five hundred random variates were generated ($N=500$). The assumed true values for parameters are given in Table 1 and the histograms of the generated count data are shown in Fig. 1. The data appear to be highly dispersed and resemble empirical crash frequency plots which are likely to be encountered by highway safety analysts. The sample mean and variance were 3.14 and 27.54, respectively.

In the initial assessment, we first fitted the standard NB regression model using the maximum likelihood method and checked the quality of the fit using the Pearson X^2 statistic. For a well-fitting model the value of X^2 should come from a χ^2 distribution with $(N-p)$ degrees of freedom (McCullagh and Nelder, 1987), where N is the number of observations and p is the number of parameters which have been estimated. The ad-hoc assessment is if $X^2/(N-p)$ is close to 1, we conclude that the model's goodness-of-fit is satisfactory. For this dataset, the Pearson X^2 statistic was 1.09. It seems that the NB model produced a very satisfactory goodness-of-fit and addressed the over-dispersion. However, the NB regression model is in fact misspecified and any inference or prediction from this model can be misleading because it totally ignores the existence of different coefficients.

We then fitted the data with FMP-2 and NB regression models using the Bayesian methodology. For FMP-2 model estimation, the non-informative prior specifications were used for each model parameter: i.e., the prior for the weight distribution $\mathbf{w}$ was defined as *Dirichlet* $(1, 1)$; as prior for the regression coefficient $\boldsymbol{\beta}_k$, the MVN_3 $(\mathbf{b}_0, \mathbf{B}_0)$ distribution was used, where $\mathbf{b}_0=(0, 0, 0)'$ and $\mathbf{B}_0=100\mathbf{I}_3$ ($\mathbf{I}_3$ denotes the three-dimensional identity matrix). For the NB model, the prior for the component regression coefficients was again MVN_3 $(\mathbf{b}_0, \mathbf{B}_0)$, and a non-informative $\Gamma(0.01, 0.01)$ prior was used for the dispersion parameter. A total of 5000 MCMC iterations was run without thinning, and half the iterations were discarded. From the remaining 2500 samples, the posterior means, and standard deviations were calculated.

Table 1 summarizes the estimated parameters and computed values of model comparison criteria for each model. Five model comparison criteria are log-likelihood (LL), Akaike Information Criterion (AIC), Bayesian information criterion (BIC), Deviance information criterion (DIC), and the log of marginal likelihood (log(ML)). The log-likelihood and the information criteria were evaluated at the posterior mean of the parameters in the model. The marginal likelihood was approximated by the Laplace–Metropolis estimator introduced by Lewis and Raftery (1997). As a rule of thumb, a difference greater than 10 between information-based criteria (AIC, BIC, or DIC) indicates very strong evidence in favor of the model with the lower value, while a difference greater than 5 between log-marginal likelihoods is strong evidence for the model with the higher value (Kass and Raftery, 1995; Spiegelhalter et al., 2002; Burnham and Anderson, 2004). For this dataset, as shown in the table, the coefficients estimated from the FMP-2 model are close to the true values and, as expected, all model selection criteria support the choice of FMP-2 model. The NB model, by nature, does not explain the heterogeneous impact of the covariates. Fig. 1 visualizes the goodness-of-fit between generated frequencies and predicted frequencies from each model. The NB model is showing a poor predictive capability, especially at the smaller numbers of counts. This resulted from the fact that the NB model was not able to consider the population heterogeneity by completely ignoring the discrete nature of the data generation process.

3.2. Example 2

This is another example to show how well a two-component finite mixture of negative binomial regression (FMNB-2) model can replicate the data as compared to the standard NB model or FMP-2 model, when the data were originally generated from a FMNB-2 distribution. As in Example 1, this example is meant to show the poor prediction capability of other models because of model misspecification. The results of this example will support the idea of using the FMNB-2 model when the data are suspected to belong to different components and each component exhibits over-dispersion. Note that the FMP-2 model can accommodate the population heterogeneity, but cannot handle the extra-variation within components.

Generating FMNB-2 random variates is very similar to the generation of FMP-2 random variates, except that the negative binomial distribution is used for a component model. The component means, $\mu_{i,1}$ and $\mu_{i,2}$ are constructed in the same manner as in Example 1. Then, with probability w, the binary value (z_i) is generated from the *Binomial* $(1, w)$ distribution for each site i, and then the FMNB-2 random variates are generated from $z_i \cdot NB\ (\mu_{i,1}, \phi_1)+(1-z_i) \cdot NB\ (\mu_{i,2}, \phi_2)$. For this exercise, a sample size $N=500$ was used. The assumed values for parameters are shown in Table 2, and the histograms of the generated count data are shown in Fig. 2. The sample mean and sample variance were 2.84 and 28.85, respectively.

The initial check of the goodness-of-fit of the NB model was carried out with the $X^2/(N-p)$ statistic. For this dataset, it was around 1.19 which indicates that the observations are a little over-dispersed with respect to the NB model. However, all the coefficients turned out to be significant at 5% level. Note that the negative binomial can also be over-dispersed. Hilbe (2007) suggests that if the Pearson X^2 statistic is greater than 1.25 for moderate sized models and 1.05 for large number of observations, a correction for over-dispersion may be warranted. It is obvious in this case that the correction is required and the source of over-dispersion is the population heterogeneity. Thus, the finite mixture model is a good option among others.

For FMNB-2 model estimation, the prior specifications for the weight distribution and the regression coefficients are the same as in Example 1. As prior for the dispersion parameter ϕ_k, the non-informative $\Gamma(0.01, 0.01)$ was used. Table 2 displays the posterior

Table 2
The values for model comparison criteria should be properly positioned right under each model (i.e., NB regression, FMNB-2, and FMP-2).

Model parameters	True values		NB regression	FMNB-2		FMP-2	
	Component 1	Component 2		Component 1	Component 2	Component 1	Component 2
$\beta_{0,k}$	2.0	0.0	1.0221 (0.0634)	1.8333 (0.1133)[a]	−0.0292[b] (0.0674)	2.0692 (0.0550)	−0.0277[b] (0.0576)
$\beta_{1,k}$	−0.5	0.5	−0.1587 (0.0627)	−0.5195 (0.0828)	0.4641 (0.0582)	−0.3690 (0.0382)	0.4344 (0.0471)
$\beta_{2,k}$	0.5	−0.5	0.1409 (0.0614)	0.6351 (0.0907)	−0.5034 (0.0632)	0.5284 (0.0451)	−0.4729 (0.0490)
ϕ_k	5	10	0.575 (0.047)	4.152 (1.634)	19.492 (19.174)	–	–
w_k	0.2	0.8	–	0.218 (0.030)	0.782 (0.030)	0.179 (0.022)	0.821 (0.022)
Model comparison criteria[c]							
−2LL	The smaller the better		2135.6	1891.6		1943.8	
AIC	The smaller the better		2143.6	1909.6		1957.8	
BIC	The smaller the better		2160.5	1947.6		1987.3	
DIC	The smaller the better		2143.7	1911.8		1957.7	
Log(ML)	The larger the better		−1089.3	−987.2		−1007.7	

[a] The standard deviation of the coefficient.
[b] The coefficient whose 95% credible interval includes zero.
[c] LL represents the Log-Likelihood. AIC, BIC and DIC are the information-based criteria, and Log(ML) represents the log of marginal likelihood.

means of parameters and computed values of model comparison criteria for each model. The coefficients estimated from the FMNB-2 model are close to the true values except for the dispersion parameters, and all model selection criteria support the FMNB-2 model. Apparently, the FMP-2 model depicts the true regression parameters quite well. However, because of model misspecification, it could not account for the additional heterogeneity present within components. Such heterogeneity resulted in the underestimation of the standard deviation of the parameters. As shown in the table, the standard deviations for each parameter of FMP-2 model are consistently lower than those of FMNB-2 model. This is typical of the standard Poisson regression model when the over-dispersion is not accounted for. A comparison of the goodness-of-fit of the three models is provided in Fig. 2. It is clear from the figure that the predicted frequencies from the standard NB model do not fit the data very well, especially for smaller counts. As the number of counts increases, the difference between the models becomes minimal because the very low probabilities are usually assigned to high counts.

On the other hand, the posterior mean of ϕ_2 for the smaller-mean component is by no means close to the true value, and its standard deviation is also very large. This is because the posterior distribution included implausibly large values of ϕ_2 during the

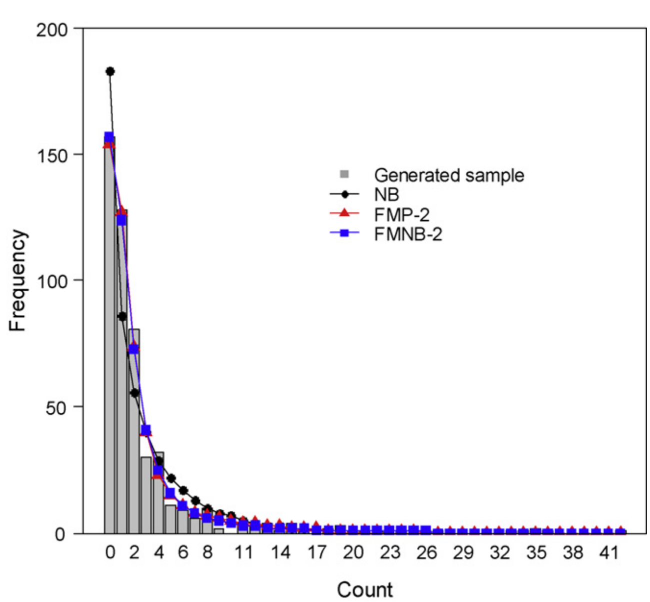

Fig. 2. Goodness-of-fit comparisons between NB, FMP-2, and FMNB-2.

sampling process, which rendered its posterior skewed with a long right tail. One probable reason for this is attributable to the use of a non-informative gamma prior with an extremely large variance. The posterior mean from such skewed distribution is biased and its use as a posterior summary statistic is not a good option. The posterior medians for ϕ_1 and ϕ_2 were 3.893 and 12.715, respectively. In contrast, the maximum likelihood estimates for ϕ_1 and ϕ_2 were 5.103 and 32.676, respectively. Van Dongen (2006) noted that if the posterior is skewed, the mean or median of the posterior will not necessarily be close to the maximum likelihood estimates even if a non-informative prior is used. This prompts the suggestion that different prior specifications should be considered on the dispersion parameter so that its posterior distribution is less skewed and the posterior mean or median is less biased. In the next section, we will investigate the bias and variability associated with the estimates of dispersion parameters using both non-informative and weakly-informative priors under various combinations of sample sizes and sample-mean values.

4. Monte Carlo simulation

In the previous section, we have shown the usefulness of finite mixture of regression models for accommodating the population heterogeneity with a single random sample—that is, the results were evaluated with only one simulated dataset under a specific condition. It is believed that repeating the sampling would produce estimates more clustered around the true values. This point is illustrated by means of a Monte Carlo simulation. As we already observed, the posterior mean of the dispersion parameters in FMNB-2 model is biased upwards, and hence the objective of the simulation is primarily to investigate the bias associated with the estimates of dispersion parameters in the FMNB-2 models. To this end, the simulation is carried out under the various sample sizes under different sample-mean categories. In addition, since the prior specification for the dispersion parameter has a potential influence on the posterior summary statistics, the results from non-informative and weakly-informative prior specifications are compared in terms of the magnitude of the bias introduced by various sample sizes and sample-mean values.

4.1. Simulation design

We first designed the simulation scenarios for generating FMNB-2 random variates. The regression parameters ($\boldsymbol{\beta}_k$), mixing proportions (w_k), and dispersion parameters (ϕ_k) were controlled in order to generate three sample-mean categories: high mean ($\bar{\mathbf{y}} > 5$), moderate mean ($1 < \bar{\mathbf{y}} < 5$), and low mean ($\bar{\mathbf{y}} < 1$). To

Table 3
True values used for generating two-component negative binomial mixture (FMNB-2) random variates.

Parameters	High mean ($\bar{\mathbf{y}} > 5$)		Moderate mean ($1 < \bar{\mathbf{y}} < 5$)		Small mean ($\bar{\mathbf{y}} < 1$)	
	Component 1	Component 2	Component 1	Component 2	Component 1	Component 2
$\beta_{0,k}$	2.5	1.0	2.0	0.0	0.5	−1.0
$\beta_{1,k}$	−0.5	0.5	−0.5	0.5	−0.5	0.5
$\beta_{2,k}$	0.5	−0.5	0.5	−0.5	0.5	−0.5
ϕ_k	5	10	5	10	5	10
w_k	0.4	0.6	0.2	0.8	0.2	0.8
Sample size N	100–1000 (one hundred step) and 2000		100–1000 (one hundred step) and 2000		500–3500 (five hundred step) and 5000	

allow for a high level of heterogeneity the higher-mean component (Component 1) was combined with a lower ϕ value, and the smaller-mean component (Component 2) was combined with a higher ϕ value. Note that to appreciate the sensitivities with respect to different ϕ and w values, it is necessary to test with more combinations of their values. However, since the number of combinations can be prohibitive, the simulation was limited to the one provided in Table 3. It is also desirable to work with more realistic values for the regression parameters assumed in each component model. However, since the main focus of the simulation was to investigate the bias associated with the estimates of dispersion parameters in the FMNB-2 model, we took the same values for the covariates used in Example 2, except for the intercept. This was based on the fact that we assumed the results would not be affected much by different values for the regression parameters.

Second, two prior specifications for the dispersion parameter are compared. The first one is a non-informative gamma prior: $\phi_k \sim \Gamma(0.01, 0.01)$. This is by far the most common prior within the standard NB model (e.g., Miaou and Lord, 2003). This prior has a spike near zero with a mean = 1 and a large variance of 100. However, the automatic assignment of such flat or wide priors can be problematic in some cases (Van Dongen, 2006; Lord and Miranda-Moreno, 2008), so the prior specification should be done with great care. The extremely large variance can create problems especially when the sample size is small or sample mean is low. Recently, less vague priors have been proposed to analyze crash data (Washington and Oh, 2006; Miranda-Moreno et al., 2008). Not to mention, in the FMNB-2 model, it is difficult to assign informative priors on each component model because such information is rarely available. Therefore, as a comparative purpose, we introduced a weakly-informative gamma prior: $\phi_k \sim \Gamma(0.5, 0.1)$, and investigated its performance in terms of the bias. This prior also has a mode near zero, but it has a mean = 5 and a much reduced variance of 50. This prior was suggested in this paper with the hope of reducing the implausibly large values of ϕ_k in the posterior samples, and improving the behavior of the posterior distribution of ϕ_k. Also, it is unlikely that this prior will prevent the chain from exploring the plausible space of ϕ_k since the prior variance is still large. Within the standard NB model, Lord and Miranda-Moreno (2008) found that the priors with very small mean and variance (for example, $\phi_k \sim \Gamma(0.1, 1.0)$) often generated extremely small values for the dispersion parameter, which resulted in a significant underestimation of the true value.

Based on the above-described simulation scenarios, the FMNB-2 random variable generation process was replicated 100 times for each category, and then for each of the datasets, Bayesian esti-

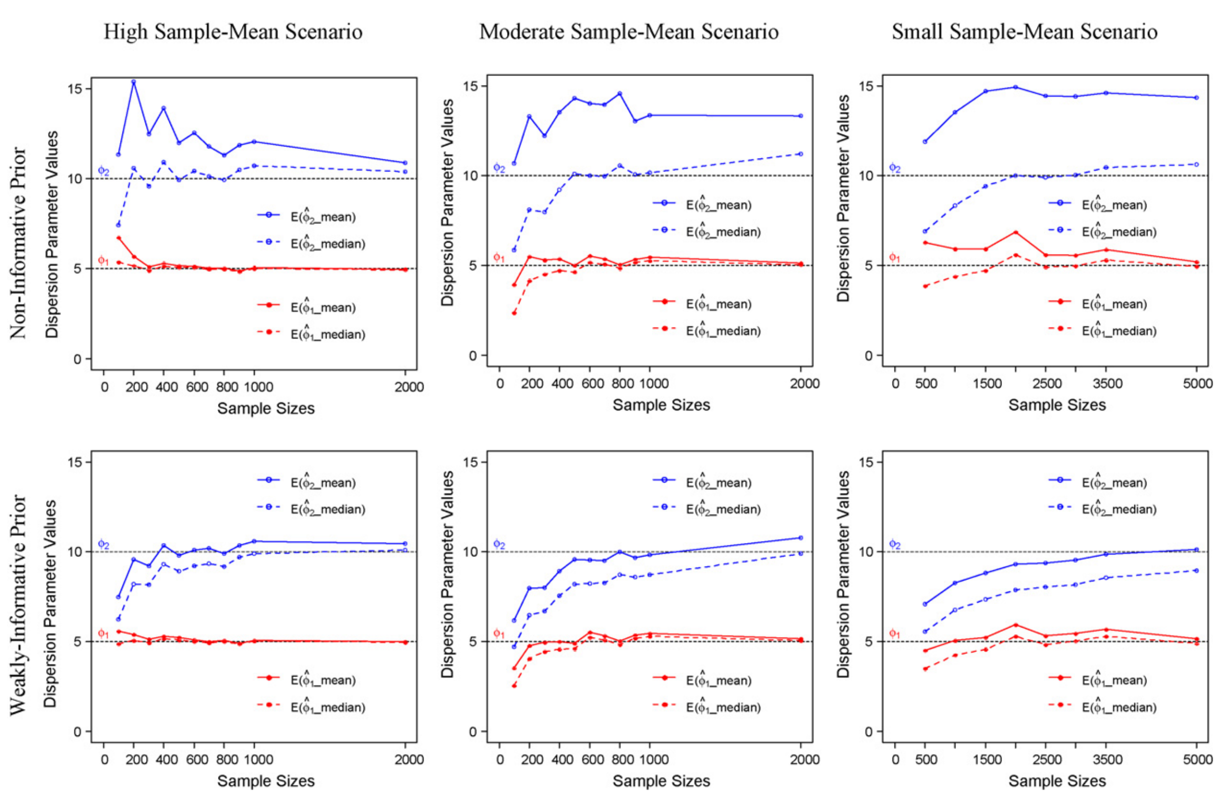

Fig. 3. Bias trends for dispersion parameters by sample size under different scenarios.

Table 4
Mean squared error (MSE) information.

High sample-mean value scenario											
Non-informative prior, $\phi \sim \Gamma(0.01, 0.01)$											
Sample size	100	200	300	400	500	600	700	800	900	1000	2000
MSE($\hat{\phi}_{1,mean}$)	25.64	14.09	2.31	2.44	2.00	0.93	0.71	0.68	0.40	0.26	0.23
MSE($\hat{\phi}_{1,median}$)	11.13	6.90	2.03	1.75	1.73	0.85	0.69	0.67	0.39	0.25	0.23
MSE($\hat{\phi}_{2,mean}$)	42.02	107.06	54.81	68.47	35.54	42.52	25.27	30.34	35.16	30.15	7.59
MSE($\hat{\phi}_{2,median}$)	20.76	34.22	20.40	23.50	15.12	17.16	11.89	16.19	18.23	14.98	5.46
Weakly-informative prior, $\phi \sim \Gamma(0.5, 0.1)$											
MSE($\hat{\phi}_{1,mean}$)	5.12	4.89	2.11	1.97	1.76	1.14	0.72	0.68	0.39	0.26	0.23
MSE($\hat{\phi}_{1,median}$)	3.65	3.72	1.91	1.59	1.61	1.07	0.69	0.66	0.39	0.25	0.23
MSE($\hat{\phi}_{2,mean}$)	14.00	11.82	9.96	10.97	10.24	9.05	7.60	8.40	10.52	9.42	4.81
MSE($\hat{\phi}_{2,median}$)	19.80	11.98	10.19	8.70	9.13	7.55	6.19	7.42	8.35	7.12	4.00
Moderate sample-mean value scenario											
Non-informative prior, $\phi \sim \Gamma(0.01, 0.01)$											
Sample size	100	200	300	400	500	600	700	800	900	1000	2000
MSE($\hat{\phi}_{1,mean}$)	10.59	15.39	8.23	15.39	5.34	4.11	5.78	2.70	2.57	3.61	0.63
MSE($\hat{\phi}_{1,median}$)	9.90	6.78	4.81	8.23	4.01	3.09	3.90	2.38	2.18	3.01	0.57
MSE($\hat{\phi}_{2,mean}$)	18.91	52.88	39.09	63.38	74.18	71.56	71.02	70.51	67.65	53.37	55.19
MSE($\hat{\phi}_{2,median}$)	23.60	20.64	16.61	21.95	23.07	24.01	23.03	20.37	25.91	16.87	24.39
Weakly-informative prior, $\phi \sim \Gamma(0.5, 0.1)$											
MSE($\hat{\phi}_{1,mean}$)	5.27	5.29	3.93	5.48	3.34	3.46	3.65	2.11	2.38	3.39	0.61
MSE($\hat{\phi}_{1,median}$)	7.93	4.58	3.36	4.42	2.93	2.73	3.00	1.91	2.05	2.86	0.56
MSE($\hat{\phi}_{2,mean}$)	18.47	10.85	10.36	9.30	9.63	9.83	9.80	8.82	11.41	8.38	10.08
MSE($\hat{\phi}_{2,median}$)	30.88	17.96	15.56	12.38	10.22	10.34	10.36	8.34	10.61	7.87	7.53

Small sample-mean value scenario								
Non-informative prior, $\phi \sim \Gamma(0.01, 0.01)$								
Sample size	500	1000	1500	2000	2500	3000	3500	5000
MSE($\hat{\phi}_{1,mean}$)	23.22	19.73	17.02	18.73	9.90	5.92	7.99	3.092
MSE($\hat{\phi}_{1,median}$)	8.12	7.74	6.72	6.96	5.05	3.20	4.16	2.145
MSE($\hat{\phi}_{2,mean}$)	30.54	41.56	82.30	83.74	77.77	75.29	77.41	74.869
MSE($\hat{\phi}_{2,median}$)	19.22	13.21	24.93	24.67	24.11	22.54	23.19	22.583
Weakly-informative prior, $\phi \sim \Gamma(0.5, 0.1)$								
MSE($\hat{\phi}_{1,mean}$)	5.31	5.93	4.87	6.54	4.47	3.50	4.57	2.166
MSE($\hat{\phi}_{1,median}$)	5.63	4.57	3.58	4.12	3.21	2.48	3.20	1.748
MSE($\hat{\phi}_{2,mean}$)	12.62	7.62	10.69	9.40	10.02	9.32	9.35	9.743
MSE($\hat{\phi}_{2,median}$)	22.88	14.18	14.05	11.23	11.04	10.13	8.98	8.316

mation was carried out using 2500 draws after a burn-in of 2500 draws. The prior specifications for the weight distribution and the regression parameters were the same as in Example 2. A special consideration was also taken during the simulation to prevent the label switching which is caused by the invariance of a finite mixture distribution to relabeling the components. According to Frühwirth-Schnatter (2006), finding identifiability constraints is not trivial in finite mixture regression models. After trying with several datasets from each mean value category, the order constraint on the weight parameters (i.e., $w_1 > w_2$) was found to be appropriate for the moderate and small mean value scenarios. For the high mean scenario, the constraint on the intercepts (i.e., $\beta_{0,1} > \beta_{0,2}$) was found to be most appropriate.

At the end of each replication, the posterior summary statistics such as posterior mean, median, standard deviation for each parameter estimate were computed. This provides the bias information $[E(\hat{\phi}_r) - \phi_{true}]$, where r is the number of replications. The mean squared error, $MSE = Bias^2 + Var(\hat{\phi}_r)$, is another appropriate measure to check the quality of an estimator since it comprises both bias and variability.

4.2. Simulation results

Fig. 3 shows the bias trends by sample size for the three sample-mean value scenarios. The upper figures are for the non-informative prior and the lower ones are for the weakly-informative prior. If there is no bias, all the points would rest on $\phi_1 = 5$ for component 1 and $\phi_2 = 10$ for component 2. Generally, as the sample size increases, the bias tends to decrease.

For the high sample-mean value scenario, the bias is negligible for the higher-mean component (component 1) unless the sample size is too small (about $N = 300$) for both priors. The bias is more significant in the smaller-mean component (component 2). This is particularly true if we choose the posterior mean as a summary statistic with the non-informative prior. For the non-informative prior case, there is an upward-bias trend for both posterior mean and median in component 2. It is evident that the posterior median has much better bias properties than the posterior mean. On the other hand, for the weakly-informative prior case, there is a slightly upward-bias for the posterior mean, but the bias appears to be small. For the posterior median, there is a downward-bias in component 2 when the sample size is less than 1000, but as the sample size increases this trend disappears and the bias becomes negligible. We can see that even though the prior information is weak, it introduces a bias in posterior median for the smaller-mean value component when the sample size is small or moderate. Table 4 provides mean squared errors (MSE) for each case. Because of the reduced variance in the weakly-informative prior case, its posterior mean and median perform better than those for the non-informative prior case. In sum, for the high mean value scenario, if we use the non-informative prior, the choice of posterior mean should be avoided in terms of bias and MSE. The bias risks for other cases seem to be minimal. However, as the sample size becomes larger (more than $N = 1000$), the posterior median with a weakly-informative prior is preferred.

Table 5
Recommended priors and summary statistics in terms of bias properties.

Sample-mean range	Sample size range	Recommended priors	Recommended summary statistics
High mean ($\bar{\mathbf{y}} > 5$)	300–1000	Non-informative prior	Posterior median
		Weakly-informative prior	Posterior mean
	>1000	Weakly-informative prior	Posterior median
Moderate mean ($1 < \bar{\mathbf{y}} < 5$)	500–1000	Non-informative prior	Posterior median
		Weakly-informative prior	Posterior mean
	1000–2000	Weakly-informative prior	Posterior mean
	>2000	Weakly-informative prior	Posterior median
Small mean ($\bar{\mathbf{y}} < 1$)	1500–3000	Non-informative prior	Posterior median
	>3500	Weakly-informative prior	Posterior mean

Many empirical crash data fall in the moderate sample-mean value scenario. The bias trends for this scenario are similar to the high mean value scenario, but the upward or downward trends are more pronounced, especially for the smaller-mean component. Furthermore, it requires higher sample sizes than the high mean value scenario to obtain the similar amount of bias for components 1 and 2 (about $N = 500$). Obviously the posterior mean with the non-informative prior is not an option. It is interesting to notice that up to $N = 1000$, the bias for the posterior median with the non-informative prior is at its minimum, but as the sample size increases significantly larger (i.e., $N = 2000$), it starts to exhibit the upward-bias trend (for component 2). A similar tendency is also observed in the posterior mean with a weakly-informative prior. On the other hand, the posterior median with the weakly-informative prior is consistently lower than the true value up to sample size $N = 2000$. The MSE information is provided in Table 4. Even though there is a downward-bias in the posterior mean and median when a weakly-informative prior is used, because of the reduced variability in the estimates they are performing better than the posterior median with the non-informative prior in all sample sizes.

For small sample-mean value scenario, a much larger sample size was necessary to examine the bias properties. When the sample size was below 500, the parameter estimates were very unstable and many generated datasets showed the label switching problem even with the order constraints. Fig. 3 shows the bias trends for the small-mean values. The sample size starts from $N = 500$ with a five hundred step. For the non-informative prior case, the posterior mean exhibits a high upward-bias for component 1, not to mention for component 2. The posterior median seems to be working fine for both components after $N = 1500$, but as the sample size grows the upward-bias trend becomes noticeable especially for component 2. For the weakly-informative prior case, the dispersion parameter for component 2 is consistently underestimated for both posterior mean and median. This tendency was already observed in the moderate sample-mean value scenario. We can infer from this result that the weakly-informative prior is exercising much influence when the component mean value is low and its sample size is small. It is pulling down the posterior mean toward the prior mean which is 5, which deteriorates the bias properties of posterior mean or median. However, in terms of MSE (Table 4), in spite of the significant underestimation, the summary statistics for the weakly-informative prior perform better than those of the non-informative prior in most sample sizes.

5. Summary and conclusion

This study showed using two hypothetical examples that the standard NB regression model was not a viable option if the source of the over-dispersion is due to population heterogeneity. The FMP-2 could not handle the extra-variation within components which may often be the case in crash data. In both cases, the interpretation of the model parameters was misleading and the FMNB-2 model was a good candidate model. The simulation study conducted on the FMNB-2 model showed that the posterior mean using the non-informative prior exhibited a high bias for the dispersion parameter especially, in the smaller-mean value component. The posterior median, instead, has much better bias properties than the posterior mean, particularly at small sample sizes and small sample means. However, as the sample size increases significantly for both small to moderate mean value scenarios, the posterior median using the non-informative prior also began to exhibit the upward-bias trend. This is because as the sample size increases the posterior median is getting closer to the posterior mean which exhibits the upward-bias. The use of the weakly-informative prior had the advantage of reducing the variability in the estimates for the posterior mean and median, but it tended to underestimate the true value by pulling the estimates toward its prior mean. As the sample-mean value decreases this tendency was more pronounced.

Based on the results of this study, we suggest guidelines about the selection of priors and the corresponding summary statistics to use in terms of their bias properties. The guidelines are tabulated in Table 5 for different sample sizes and sample-mean values. As indicated in the introduction, since the necessary sample size in FMNB-2 model greatly depends on the current dataset, the sample size ranges suggested in Table 5 are for a relatively well-separated data. The minimum sample sizes in each sample-mean range (i.e., $N = 300$ for high mean, $N = 500$ for moderate mean, and $N = 1500$ for small mean) was basically determined by the bias associated with the dispersion parameters, but they also minimize the biases in the regression coefficients and the mixing proportions. The bias trends in the regression coefficients and the mixing proportions were not provided in this paper due to space constraints, but they were generally negligible as compared to the bias associated with the dispersion parameters.

In terms of future study on this topic, similar simulations may be performed for the recently proposed two-state Markov switching models or random-parameter models and the results could then be compared with those found in this study. Although some guidelines for the development and application of finite mixture models were provided in this study, these guidelines are based on the results based from a limited combination of simulation design values. To fully understand the bias properties, larger scale simulation studies may be required. If the upward-bias in the posterior mean using a non-informative prior is still observed, this could potentially limit the use of finite mixture models. Therefore, further methodological work to better understand and minimize this upward-bias for the dispersion parameter may be necessary.

Acknowledgements

The authors would like to thank two anonymous reviewers for providing useful comments aimed at improving this paper.

References

Aitkin, M., 2001. Likelihood and Bayesian analysis of mixtures. Statistical Modelling 1 (4), 287–304.

Anastasopoulos, P., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41 (1), 153–159.

Burnham, K., Anderson, D., 2004. Multimodal inference: understanding AIC and BIC in model selection. Sociological Methods Research 33 (2), 261–304.

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), 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 (2), 363–375.

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

Geedipally, S.R., Lord, D., Park, B.-J., 2009. Analyzing different parameterizations of the varying dispersion parameter as a function of segment length. Transportation Research Record 2103, 108–118.

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

Heckman, J.J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52 (2), 271–320.

Heydecker, B.G., Wu, J., 2001. Identification of sites for road accident remedial work by Bayesien statistical methods: an example of uncertain inference. Advances in Engineering Software 32 (10), 859–869.

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

Jasra, A., Holmes, C.C., Stephens, D.A., 2005. Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. Statistical Science 20 (1), 50–67.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90 (430), 773–795.

Laird, N., 1978. Nonparametric maximum likelihood estimation of a mixing distribution. Journal of the American Statistical Association 73 (364), 805–811.

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 (438), 648–655.

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., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science 46 (5), 751–770.

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.

McCullagh, P., Nelder, J.A., 1987. Generalized Linear Models. Chapman and Hall, London.

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.

Miranda-Moreno, L.F., Lord, D., Fu, L., 2008. Bayesian road safety analysis: incorporation of past experiences and effect of hyper-prior choice. In: TRB 87th Annual Meeting Compendium of Papers DVD, Paper No. 08-1788.

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

Park, B.-J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accident Analysis and Prevention 41 (4), 683–691.

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

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 (4), 583–639.

Van Dongen, S., 2006. Prior specification in Bayesian statistics: three cautionary tales. Journal of Theoretical Biology 242 (7), 90–100.

Van Duijn, M.A.J., Böckenholt, U., 1995. Mixture models for the analysis of repeated count data. Journal of the Royal Statistical Society Series C 44 (4), 473–485.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC.

Washington, S., Oh, J., 2006. Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: example applied to at grade railroad crossings in Korea. Accident Analysis and Prevention 38 (2), 234–247.