Analytic Methods in Accident Research 1 (2014) 39–52

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models

Yajie Zou [1], Yunlong Zhang *,[2], Dominique Lord [3]

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

ARTICLE INFO

Article history:
Received 27 June 2013
Received in revised form
1 November 2013
Accepted 1 November 2013
Available online 9 December 2013

Keywords:
Negative binomial
Finite mixture model
Dispersion
Weight parameter
Crash data
Heterogeneity

ABSTRACT

Factors that cause heterogeneity found in motor vehicle crash data are often unknown to transportation safety researchers and failure to capture this heterogeneity in statistical models can weaken the validity of modeling results. A finite mixture of regression models has been proposed to address the unobserved heterogeneity in crash data, and a fixed weight parameter for these models (i.e., the weight parameter is invariant of the characteristics of the observations under study) is commonly assumed. Recent studies have found that the weight parameter of the finite mixture of negative binomial (NB) models can be dependent upon the functional form of the attributes of the sites, and the selection of the functional form for weight parameter has a significant impact on the classification results.

This study investigates the effect of different functional forms on the estimation of the weight parameter as well as the group classification of the finite mixture of NB regression models, using crash data collected on rural roadway sections in Indiana. A total of 11 different functional forms for the varying weight parameter were estimated; these functional forms include various combinations of traffic flow and segment length as covariates. The results suggest that the modeling of the weight parameter (which essentially helps in improving the group classification) is generally necessary when using the finite mixture of NB regression models to analyze the crash data, even in the presence of a well-defined mean function. This study also confirms that the selection of the functional form for weight parameter will affect the classification results significantly. Among 11 different functional forms, one functional form, which uses the linear combination of different explanatory variables to model the classification, stands out based on both the goodness-of-fit statistics and the classification results, and is recommended for describing the weight parameter when using the finite mixture of NB regression models with varying weight parameters to analyze crash data.

1. Introduction

Research on understanding the factors that affect the probability of vehicle crashes has been of great interest to transportation safety analysts for many years. A number of factors known to influence traffic safety are driving-related

* Corresponding author.
E-mail addresses: zouyajie@gmail.com (Y. Zou), yzhang@civil.tamu.edu (Y. Zhang), d-lord@tamu.edu (D. Lord).

[1] Tel.: +1 936 245 5628; fax: +1 979 845 6481.

[2] Tel.: +1 979 845 9902; fax: +1 979 845 6481.

[3] Tel.: +1 979 458 3949; fax: +1 979 845 6481.

2213-6657/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.amar.2013.11.001

factors (acceleration, braking and steering information, driver response to unexpected incidents and so on); factors related to road and vehicle (roadway geometric configurations, surface conditions, vehicle features, etc.); traffic and environment-related factors (traffic flows and composition, traffic control, weather conditions, etc.). Unfortunately, some factors (i.e., driving-related data) are usually not observable. As a result, researchers mainly focus on investigating the observed factors that affect the number of crashes for roadway segments or intersections over some fixed time periods (Lord and Mannering, 2010). For the crash count data, to overcome the over-dispersion problem associated with the Poisson regression, various ways have been proposed within the negative binomial (NB) models (Poch and Mannering, 1996; Hauer, 2001; Miaou and Lord, 2003; El-Basyouny and Sayed, 2006; Mitra and Washington, 2007; Malyshkina et al. 2009; Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2010). Recently, a quantile regression method was introduced by Qin and Reyes (2011) to analyze the heterogeneous crash data. The quantile regression can offer a complete view of how the explanatory variables affect the crash occurrence from the full range of the distribution.

Since the occurrence of crashes is rare (relatively speaking), to ensure the adequacy of sample size for valid and robust statistical estimation and inferences, crash data are often aggregated from a wide range of geographic locations. The aggregated crash data may contain heterogeneity. Heterogeneity implies that crash data may be collected from different sources (i.e., crash data collected at similar locations may share some common characteristics, while crash data collected at different locations may exhibit different characteristics). Thus, it is reasonable to assume that the sites with different combinations of factors (i.e., geometric design features, etc.) can constitute distinct sub-populations (sites are heterogeneous across and homogeneous within the sub-populations). Under this assumption, the commonly used negative binomial (NB) regression model may become inappropriate and the model estimation and inference from an NB modeling framework could be inefficient or misleading. Therefore, as proposed by Park and Lord (2009), it is reasonable to hypothesize that the individual crashes on highway entities (intersections or road segments) are generated from a certain number of hidden subgroups that are unknown to transportation safety researchers.

The concept of finite mixture distribution can date back to 1943 (Frühwirth-Schnatter, 2006). So far, the mixture modeling techniques have been applied in some transportation research areas. For example, to capture the heterogeneity in speed data, the normal mixture model has been used to fit the distribution of speed (Jun, 2010; Park et al., 2010b; Zou and Zhang, 2011; Zou et al., 2012). A finite mixture of regression models has been proposed to address the over-dispersion problem in crash data (Park and Lord, 2009). Xiong and Mannering (2013) developed a finite mixture random-parameters approach to study the heterogeneous effects of guardian supervision on crash-injury severities. For a standard finite mixture of regression models, previous studies (Park and Lord, 2009; Park et al., 2010a; Chang and Kim, 2012) have used a fixed weight parameter that is applied to the entire dataset. Latent segmentation models, which are similar to the finite mixture model, have been developed and applied in various transportation studies (Bhat, 1997; Greene and Hensher, 2003; Eluru et al., 2012; Sobhani et al., in press). These studies used different exogenous variables to estimate the weight parameter in the latent segmentation models. Recently, the finite mixture of NB regression models with varying weight parameters has been introduced by Zou et al. (2013) for analyzing the dispersed crash data. In their study, the weight parameter of the finite mixture models is assumed to be variable and can be dependent upon the attributes of the sites (i.e., covariates), such as segment length, traffic flow, etc. The results suggest that the two-component finite mixture of NB regression models (termed as the FMNB-2 model) with varying weight parameters can provide more reasonable classification results, as well as better statistical fitting performance than the FMNB-2 models with fixed weight parameters. Zou et al. (2013) also noted that the selection of the functional form for weight parameter has a significant impact on the classification results. For the FMNB-2 models with varying weight parameters, suspicious modeling results and erroneous inferences can be obtained if the functional form for weight parameter is mis-specified. Thus, there is a need to investigate how different functional forms affect the estimation of the varying weight parameter and whether there is a common functional form that can be properly used to model the weight parameter for different crash datasets. So far, a few functional forms have been proposed for estimating the weight parameter that varies across observations.

The primary objective of this research is to investigate the effect of different functional forms on estimation of the weight parameter as well as the group classification. Specifically, we mainly examine the modeling results and group classification from the finite mixture of NB regression models with different functional forms for the varying weight parameter. To accomplish the study objectives, 11 different functional forms for the varying weight parameter are estimated using the crash data collected on rural road sections in Indiana.

2. Background

This section describes the characteristics of the finite mixture of NB regression models with fixed and varying weight parameters.

2.1. Finite mixture of NB regression models

This study adopts a finite mixture model to describe heterogeneous crash data. To deal with the unobserved heterogeneity, the occurrence of the random vector, $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ is assumed to follow a finite mixture distribution. The mixture model is very flexible and the probability density function of a g-component mixture distribution can be formulated as follows:

$$f_Y(\mathbf{y}|\boldsymbol{\Theta}) = \sum_{j=1}^{g} w_j f_j(\mathbf{y}|\boldsymbol{\theta}_j) \tag{1}$$

where w_j is the weight of component j (weight parameter), with $w_j>0$, and $\sum_{j=1}^{g} w_j = 1$; $\boldsymbol{\theta}_j$ are vectors of parameters for the component j; $f_j(\mathbf{y}|\boldsymbol{\theta}_j)$ is the component density for component j (j=1, 2, …, g); g is the number of components; and, $\boldsymbol{\Theta} = ((w_1,\ldots,w_g),\ \boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_g)$ is the vector of all unknown parameters.

The mean and variance of a finite mixture distribution can be written as (Frühwirth-Schnatter, 2006)

$$\mu = E(\mathbf{y}|\boldsymbol{\Theta}) = \sum_{j=1}^{g} \mu_j w_j \tag{2}$$

$$\sigma^2 = Var(\mathbf{y}|\boldsymbol{\Theta}) = \sum_{j=1}^{g} (\mu_j^2 + \sigma_j^2) w_j - \mu^2 \tag{3}$$

where $\mu_j = E(\mathbf{y}|\boldsymbol{\theta}_j)$ is the component mean; and, $\sigma_j^2 = Var(\mathbf{y}|\boldsymbol{\theta}_j)$ is the component variance.

The finite mixture distribution (Eq. (1)) can be extended to a more generalized model if more information about the nature of heterogeneity is available. For example, the parameter $\boldsymbol{\theta}_j$ for each component can be parameterized using the observable covariates, leading to a regression model. The weight parameter w_j may also depend on the observable covariates. These extensions are described as follows.

In most studies, the component distributions are assumed to arise from the same parametric distribution family. In this paper, all components are NB distributed. For the g-component finite mixture of negative binomial regression models (termed as the FMNB-g model), it is assumed that the marginal distribution of y_i follows a mixture of NB distributions,

$$f_Y(y_i|\mathbf{x}_i,\boldsymbol{\Theta}) = \sum_{j=1}^{g} w_j NB(\mu_{ij},\phi_j) = \sum_{j=1}^{g} w_j \left[\frac{\Gamma(y_i+\phi_j)}{\Gamma(y_i+1)\Gamma(\phi_j)} \left(\frac{\mu_{ij}}{\mu_{ij}+\phi_j}\right)^{y_i} \left(\frac{\phi_j}{\mu_{ij}+\phi_j}\right)^{\phi_j}\right] \tag{4}$$

$$E(y_i|\mathbf{x}_i,\boldsymbol{\Theta}) = \sum_{j=1}^{g} \mu_{ij} w_j \tag{5}$$

$$Var(y_i|\mathbf{x}_i,\boldsymbol{\Theta}) = E(y_i|\mathbf{x}_i,\boldsymbol{\Theta}) + \left(\sum_{j=1}^{g} w_j \mu_{ij}^2 (1+1/\phi_j) - E(y_i|\mathbf{x}_i,\boldsymbol{\Theta})^2\right) \tag{6}$$

where $\mu_{ij} = \exp(\mathbf{x}_i\boldsymbol{\beta}_j)$; μ_{ij} is the mean rate of component j; $\mathbf{x}_i$ is a vector of covariates; $\boldsymbol{\beta}_j$ is a vector of the regression coefficients for component j; and, $\boldsymbol{\Theta} = \{(\phi_1,\ldots,\phi_g),(\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_g),\mathbf{w}\} = \{(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_g),\mathbf{w}\}$ for $i=1,2,\ \ldots,\ n$.

For the FMNB-g model, the variance of y_i is always greater than the mean. When ϕ_j in each component goes to infinity, the FMNB-g model is reduced to the g-component finite mixture of Poisson regression models. Thus, the FMNB-g models allow for additional heterogeneity within components not captured by the explanatory variables.

The term w_j is defined as the mixture weight of the FMNB-g models. Recent studies show that the weight w_j of the FMNB-g models can potentially be dependent upon the important attributes of the sites. Previously, Park and Lord (2009) also noted that the weight parameter could be modeled as a function of the covariates of the site. However, due to the complexity of the estimation process, the weight parameter was treated as a constant variable in their study. In order to examine different functional forms for the weight parameter, the FMNB-g model with a varying weight w_{ij} is considered (defined as generalized FMNB-g model or GFMNB-g model). The GFMNB-g model has the same probability density function shown in Eq. (4) and estimates the number of crashes of each site, similar to the FMNB-g model (Zou et al., 2013). However, instead of estimating a fixed weight parameter, the varying weight w_{ij} is modeled as a function of covariates:

$$\frac{w_{ij}}{w_{ig}} = \exp(\boldsymbol{\alpha}_j^T \mathbf{z}_i) \tag{7}$$

where $\mathbf{z}_i$ is a vector of secondary covariates that might help classify the sites (not necessarily the same as the covariates in estimating the mean function μ_{ij}); and, $\boldsymbol{\alpha}_j$ is a vector of regression coefficients for component j, and $\boldsymbol{\alpha}_g = \mathbf{0}$.

With Eq. (7), the new parameterization allows each observation to have a different weight that is dependent on the sites' attributes (i.e., covariates), similar to the application of the varying dispersion parameter for the standard negative binomial model (see Lord and Park (2008)). Although this new parameterization can possibly provide more reasonable classification results for the GFMNB-g models, it is found that the selection of the functional form for the weight has an impact on the classification results. Thus, more work needs to be done to examine this finding. Moreover, in the presence of a well-defined mean function, it is useful to compare the group classification results between FMNB-g and GFMNB-g models. If the included secondary covariates fail to improve the resulting classification, then the new parameterization (Eq. (7)) generally becomes unnecessary and the weight parameter will only contain a fixed value (i.e., constant term), resulting in a FMNB-g model.

2.2. Functional forms for estimating the weight parameter

So far, only a few functional forms have been proposed for estimating the weight parameter. Thus, a large number of plausible functional forms are available if we consider the types of covariates included and the forms of variable transformation applied. When selecting the potential functional forms, one important factor transportation safety analysts usually consider is the availability and significance of covariates. For roadway segment crash datasets, two important and basic attributes are traffic flow and segment length. Thus, this study uses traffic flow and segment length as covariates to explore different parameterizations of weight parameter. In addition, the linear combination of different explanatory variables is considered as an alternative functional form. To examine the effect of functional forms on modeling results, 11 different functional forms are used for estimating the weight parameter that allows the weight to vary from site to site based on its characteristics. When selecting the optimal functional form, we evaluate the goodness-of-fit statistics as well as group classification results.

The 11 functional forms are listed below.

Model 1 considers a fixed weight parameter. Eq. (7) can be written as

$$\text{Model 1}: \frac{w_{ij}}{w_{ig}} = e^{\gamma_{0j}} \tag{8}$$

Models 2–5 include segment length as the covariate. They are described as follows:

$$\text{Model 2}: \frac{w_{ij}}{w_{ig}} = e^{\gamma_{0j}} e^{\gamma_{1j} L_i} \tag{9}$$

$$\text{Model 3}: \frac{w_{ij}}{w_{ig}} = e^{\gamma_{0j}} L_i^{\gamma_{1j}} \tag{10}$$

$$\text{Model 4}: \frac{w_{ij}}{w_{ig}} = e^{\gamma_{1j} L_i} \tag{11}$$

$$\text{Model 5}: \frac{w_{ij}}{w_{ig}} = L_i^{\gamma_{1j}} \tag{12}$$

In Models 6–8, traffic flow is considered:

$$\text{Model 6}: \frac{w_{ij}}{w_{ig}} = F_i^{\gamma_{2j}} \tag{13}$$

$$\text{Model 7}: \frac{w_{ij}}{w_{ig}} = e^{\gamma_{0j}} F_i^{\gamma_{2j}} \tag{14}$$

$$\text{Model 8}: \frac{w_{ij}}{w_{ig}} = e^{\gamma_{0j}} e^{\gamma_{2j} F_i} \tag{15}$$

In Models 9 and 10, the effect of both traffic flow and segment length are taken into account:

$$\text{Model 9}: \frac{w_{ij}}{w_{ig}} = e^{\gamma_{0j}} L_i^{\gamma_{1j}} F_i^{\gamma_{2j}} \tag{16}$$

$$\text{Model 10}: \frac{w_{ij}}{w_{ig}} = L_i^{\gamma_{1j}} F_i^{\gamma_{2j}} \tag{17}$$

In Model 11, the linear combination of different available explanatory variables is used:

$$\text{Model 11}: \frac{w_{ij}}{w_{ig}} = e^{\gamma_{0j}} e^{\boldsymbol{\gamma}_j \mathbf{x}_i} \tag{18}$$

where L_i is the segment length for segment i; F_i is the average daily traffic for segment i; w_{ij} is the estimated weight of component j at segment i; $\boldsymbol{\gamma}_j = (\gamma_{1j}, \gamma_{2j}, \ldots, \gamma_{mj})'$ are the estimated coefficients for component j, m is the number of coefficients; and, $\mathbf{x}_i$ is a vector of covariates.

2.3. Parameter estimation method

The estimation of finite mixture models can be done using the maximum likelihood estimation with the Expectation Maximization (EM) algorithm or the Bayesian method. Previously, Park and Lord (2009) adopted the Bayesian framework with data augmentation and Markov Chain Monte Carlo (MCMC) techniques to estimate finite mixture of NB regression models. the maximum likelihood estimation and the Bayesian method both have their advantages and disadvantages (see McLachlan and Peel, 2000; Frühwirth-Schnatter, 2006). For example, the Bayesian method with MCMC techniques is generally computationally demanding and it can be difficult to overcome the label switching problem, while the EM algorithm tends to lead to a local maximum and thus many different starting values are needed for finding the global

Table 1
Summary statistics of characteristics for the Indiana data.

Variable	Minimum	Maximum	Mean (SD[a])	Sum
Number of crashes (5 years)	0	329	16.97 (36.30)	5737
Average daily traffic over the 5 years (*F*)	9442	143,422	30,237.6 (28,776.4)	
Minimum friction reading in the road section over the 5-year period (FR)	15.9	48.2	30.51 (6.67)	
Pavement surface type (PT) (1: asphalt, 0: concrete)	0	1	0.77 (0.42)	
Median width (ft) (MW)	16	194.7	66.98 (34.17)	
Presence of median barrier (BR) (1: present, 0: absent)	0	1	0.16 (0.37)	
Interior rumble strips (RS) (1: present, 0: absent)	0	1	0.72 (0.45)	
Segment length (miles) (*L*)	0.009	11.53	0.89 (1.48)	300.09

[a] Standard deviation.

maximum. In this study, the maximum likelihood estimation with the EM algorithm is used to estimate the model parameters. When estimating the mixture models, we also experienced the problem that the roots of the likelihood equation vary between a few values (each value corresponds to a maximum of the likelihood function). According to McLachlan and Peel (2000), in the scenario that the information of any known consistent estimator of $\mathbf{\Theta}$ is absent, an obvious choice for the root of the likelihood equation is the one corresponding to the largest of the local maxima located. In order to achieve the global (rather than a local) maximum, we repeat the fitting process 20 times using different random starting values and we select the optimal root that corresponds to the largest likelihood value. A detailed discussion about the parameter estimation method and local maximum problem is described in Zou et al. (2013).

3. Data description

To accomplish the objectives of this study, crash data collected on 338 rural interstate road sections in Indiana over a 5-year period from 1995 to 1999 were used. This dataset was selected for two main reasons. First, considering the sufficient explanatory variables included in this dataset, we can develop a well-defined mean functional form. Second, the data have been used in previous studies and are found to have good quality (Anastasopoulos et al., 2008; Geedipally et al., 2012). The summary statistics of characteristics for the Indiana data are provided in Table 1. During the 5-year study period, there were 5737 crashes that happened on 218 of the 338 highway segments, and the other 120 segments (36%) did not have any reported crashes. As shown in Table 1, the observed crash frequency ranges from 0 to 329, and the mean crash frequency is 17.0 with a variance of 1317.7. Note that the variance-to-mean ratio (VMR) is 77.6. For a complete list of variables in this dataset, interested readers can consult Washington et al. (2011). For the heterogeneity observed in this dataset, it is speculated that the heterogeneity can partially come from the existence of two different sub-populations, with each population having distinct degrees of dispersion.

4. Modeling results

The modeling results for the Indiana data are provided in this section. When analyzing the Indiana data, we consider segment length as an offset term (Eq. (19)), which means that the number of crashes is linearly proportional to segment length. The mean functional form for each component is adopted as follows:

$$\mu_{j,i} = \beta_{j,0} L_i F_i^{\beta_{j,1}} e^{\beta_{j,2} FR_i + \beta_{j,3} PT_i + \beta_{j,4} MW_i + \beta_{j,5} BR_i + \beta_{j,6} RS_i} \tag{19}$$

where $\mu_{j,i}$ is the estimated numbers of crashes at segment i for component j; L_i is the segment length in miles for segment i; F_i is the flow (average daily traffic over 5 years) traveling on segment i; FR_i is the minimum friction reading for segment i; PT_i is the pavement surface type for segment i; MW_i is the median width for segment i; BR_i is the presence of median barrier for segment i; RS_i is the presence of interior rumble strips for segment i; and, $\boldsymbol{\beta}_j = (\beta_{j,0}, \beta_{j,1}, \beta_{j,2}, \beta_{j,3}, \beta_{j,4}, \beta_{j,5}, \beta_{j,6})'$ is the vector of estimated coefficients for component j.

The number of components for GFMNB-g models can be determined using two approaches: the first approach is to assume that g is an unknown variable and is estimated within the modeling process; the second approach consists fitting a series of models with an increasing number of components and select the most plausible model using model choice criteria, as suggested by Park et al. (2010b). For GFMNB-g models, the implementation of the first approach is relatively complicated. Hence, we used the second approach for selecting the number of components. As discussed by Eluru et al. (2012), compared to the Akaike information criterion (AIC), the Bayesian information criterion (BIC) imposes a higher penalty on over-fitting with excess parameters. Thus, the BIC is more appropriate to determine the optimal number of components. For the 11 different functional forms, we applied the GFMNB-g models with an increasing number of components $g=2$, 3, 4. As shown in Table 2, the GFMNB models with $g=2$ are preferred for all functional forms except for Model 2, which has the smallest BIC value when $g=3$. Overall, based on reported BIC values, we selected the optimal number of components $g=2$ and use the GFMNB-2 model for the subsequent analyses.

Table 2
BIC values for GFMNB-g models with number of components $g=2$, 3, 4.

Model	Number of components		
	2	3	4
1	1833.90	1845.24	1890.66
2	1794.77	1789.26	1834.82
3	1786.11	1787.05	1826.92
4	1793.53	1794.53	1840.41
5	1813.23	1825.26	1848.16
6	1822.07	1842.35	1887.10
7	1832.87	1846.25	1891.45
8	1834.58	1842.15	1887.20
9	1791.48	1797.24	1851.12
10	1788.05	1791.39	1830.96
11	1807.89	1838.05	1864.28

Given the specific objectives of this study, we considered many possible functional forms to model the weight parameter. The parameter estimation results for the standard NB and Models 1–10 are provided in Table 3 and the estimation results for Model 11 are given in Table 4. From the modeling results for the standard NB, it is clear that over-dispersion exists in the dataset. In terms of the coefficient estimates, the results show some variation between models with different functional forms. Tables 3 and 4 also show that many models include insignificant covariates. Note that for Models 1, 6 and 7, the estimated coefficients for variable traffic flow in some components are counterintuitive and there are two possible reasons to explain this. First, the sample size of this dataset is relatively small (the Indiana data contain 338 road segments); as a result, the assigned number of segments in one component for these models can be far less than 100 (see Table 6 about the group classification results). Second, and more importantly, the parameter estimation results may suggest that for these functional forms, the corresponding GFMNB-2 models provide an unreasonable group classification. Due to the inappropriate grouping, the estimated coefficients are often counterintuitive. Additional discussion about the small sample size problem is described in the discussion section further below. For Model 9, the results indicate that as flow increases and segment length decreases, the probability that the selected site will belong to component 2 increases. Tables 3 and 4 indicate that some functional forms provide possible alternative models based on the statistical significance of the model parameters.

The goodness-of-fit statistics are provided in Table 5. Compared to the standard NB model, Models 1–11 all have significantly smaller deviance, AIC and BIC values. The goodness-of-fit statistics suggest that the crash data may have been generated from two distinct sub-populations, rather than from a single population. When we compare the fitting performance between FMNB-2 model (Model 1) and GFMNB-2 model (Models 2–11), it can be seen that when the mean functional form is well defined, the GFMNB-2 models can still consistently provide a better fit than the FMNB-2 model for the Indiana data, which suggests that varying weight structure can lead to better fitting results in this study. In addition, as the results indicated, for Models 2–5 (with the weight parameter modeled using segment length), the deviance, AIC and BIC values are all smaller than those of Model 1. This indicates that Models 2–5 can improve the goodness-of-fit by using the segment length as the covariate. While for Models 6–8 (with the weight parameter modeled by using traffic flow), the deviance, AIC and BIC values are slightly better than those of Model 1. This suggests that Models 6–8 did not make much improvement in terms of fit by including traffic flow as the covariate in the functional form. Overall, based on the goodness-of-fit statistics, segment length plays a more important role than traffic flow for modeling the weight parameter. Between Models 1 and 11, Model 11 has the smallest deviance and AIC values.

In order to further explore the effect of different functional forms on the classification results, based on Models 1–11, the crash data were classified into two groups by assigning each site to the component with the highest posterior probability. The posterior probability is used to calculate the probability that observation y_i is from component j. In the EM algorithm, at iteration $r+1$, the posterior probability $\hat{\varepsilon}_{ij}^{(r+1)}$ that observation y_i is from component j, given y_i and $\hat{\Theta}^{(r)}$ is defined as (Rigby and Stasinopoulos, 2010)

$$\hat{\varepsilon}_{ij}^{(r+1)} = p(\delta_{ij}=1|y_i,\hat{\Theta}^{(r)},\mathbf{x}_i) = \frac{\hat{w}_{ij}^{(r)} f_j(y_i|\hat{\mathbf{\theta}}_j^{(r)},\mathbf{x}_i)}{\sum_{k=1}^{g}\hat{w}_{ik}^{(r)} f_k(y_i|\hat{\mathbf{\theta}}_k^{(r)},\mathbf{x}_i)} \tag{20}$$

where δ_{ij} is the indicator variable, $\hat{w}_{ij}^{(r)} = p(\delta_{ij}=1|\hat{\Theta}^{(r)})$ is the prior probability that observation y_i is from component j, given $\hat{\Theta}^{(r)}$, which is estimated from iteration r.

The summary statistics of variables for each group are provided in Table 6. Note that the standard deviations and VMRs for crashes are calculated for each group. Since the sites with similar characteristics are generally classified into the same group, it is likely to observe less variation (in other words, lower dispersion) within the same group (Zou et al., 2013). Thus, the standard deviations and VMRs can be used to reasonably measure the homogeneity in each group. For Model 1, the VMRs of components 1 and 2 are 17.81 and 79.59, respectively. Although Model 1 seems to give a good overall VMR, this

Table 3
Parameter estimates for Models 1–10.

GFMNB-2	$\ln(\beta_0)$	β_1	β_2	β_3	β_4	β_5	β_6	ϕ[a]	γ_0	γ_1	γ_2
Standard NB											
Estimate	−4.456	0.688	−0.027	0.430	−0.005	−3.026	−0.398	1.124*			
Std. error	1.292	0.120	0.010	0.185	0.002	0.283	0.180				
Model 1											
Component 1									1.609		
Estimate	0.887*	−0.036*	−0.053*	0.502*	0.004*	2.441	1.579	0.433			
Std. error	4.859	0.433	0.037	0.724	0.007	0.932	0.754				
Component 2											
Estimate	−8.776	1.128	−0.020	0.354	−0.006	−21.006*	−0.514	4.108			
Std. error	1.012	0.096	0.007	0.135	0.002	410.500	0.128				
Model 2											
Component 1									−0.988	5.254	
Estimate	−5.251*	0.627*	−0.058*	−0.246*	−0.003*	1.042*	2.147	0.165			
Std. error	5.664	0.500	0.042	0.841	0.008	0.888	0.919				
Component 2											
Estimate	−8.227	1.057	−0.018	0.303	−0.004	−20.303*	−0.487	4.831			
Std. error	0.965	0.092	0.007	0.127	0.001	312.260	0.121				
Model 3											
Component 1									−3.699	−2.284	
Estimate	−8.061	1.040	−0.018	0.302	−0.003	−20.537*	−0.481	4.735			
Std. error	0.972	0.093	0.007	0.129	0.001	356.200	0.122				
Component 2											
Estimate	−6.321*	0.716*	−0.042*	−0.419*	−0.003*	0.499*	1.995	0.154			
Std. error	5.545	0.492	0.041	0.824	0.008	0.858	0.879				
Model 4											
Component 1										3.314	
Estimate	−4.669*	0.563*	−0.081*	−0.220*	−0.001*	2.166*	2.634	0.144			
Std. error	7.008	0.618	0.052	1.023	0.010	1.110	1.165				
Component 2											
Estimate	−8.299	1.060	−0.017	0.327	−0.004	−20.105*	−0.463	4.415			
Std. error	0.976	0.093	0.007	0.130	0.001	282.200	0.124				
Model 5											
Component 1										1.791	
Estimate	−7.557	0.911	−0.026*	0.938	−0.002*	−20.320*	−0.226*	0.931*			
Std. error	1.804	0.165	0.015	0.294	0.003	418.500	0.269				
Component 2											
Estimate	−6.388	0.905	−0.017	0.076*	−0.007	−0.689	−0.440	8.953			
Std. error	1.078	0.104	0.007	0.124	0.002	0.300	0.138				

Table 3 (*continued*)

GFMNB-2	$\mathrm{Ln}(\beta_0)$	β_1	β_2	β_3	β_4	β_5	β_6	ϕ[a]	γ_0	γ_1	γ_2
Model 6											
Component 1											−0.205
Estimate	−8.861	1.148	−0.026	0.399	−0.006	−20.996*	−0.565	2.779			
Std. error	1.140	0.108	0.008	0.156	0.002	405.600	0.145				
Component 2											
Estimate	−14.223*	−0.211*	−0.015*	−0.423	0.019	18.435*	17.531*	18.338			
Std. error	375.200	0.173	0.011	0.201	0.003	375.200	375.200				
Model 7											
Component 1									−28.251		2.489421
Estimate	−8.642	1.099	−0.022	0.422	−0.005	−20.437*	−0.432	2.795			
Std. error	1.104	0.106	0.008	0.147	0.002	459.500	0.143				
Component 2											
Estimate	69.372	−5.572	−0.324	0.588*	−0.018*	1.629*	−0.085*	0.172			
Std. error	18.512	1.533	0.122	1.636	0.018	1.676	1.288				
Model 8											
Component 1									−2.293		5.67E−05
Estimate	−7.432	1.007	−0.019	0.242*	−0.008	−1.114	−0.554	3.414			
Std. error	1.186	0.112	0.008	0.144	0.002	0.302	0.135				
Component 2											
Estimate	−26.986*	1.138	−0.064	0.661*	0.009	−20.140*	17.840*	1.719			
Std. error	484.302	0.245	0.020	0.514	0.004	456.901	484.294				
Model 9											
Component 1									−24.020	−2.070	2.246
Estimate	−6.999	0.960	−0.019	0.215*	−0.006	−0.734	−0.506	6.366			
Std. error	1.034	0.099	0.007	0.119	0.002	0.300	0.124				
Component 2											
Estimate	−11.940	1.163	−0.046	2.712	0.000*	−19.830*	0.256*	0.654			
Std. error	3.042	0.259	0.022	0.676	0.004	340.800	0.454				
Model 10											
Component 1										−2.281	−0.352
Estimate	−8.148	1.053	−0.018	0.288	−0.003	−20.584*	−0.491	4.923			
Std. error	0.968	0.092	0.007	0.128	0.001	356.800	0.122				
Component 2											
Estimate	−5.576*	0.657*	−0.036*	−0.160*	−0.004*	0.157*	1.474*	0.186			
Std. error	4.923	0.439	0.036	0.739	0.007	0.770	0.772				

[a] Inverse dispersion parameter $\phi = 1/\alpha$.
* Not significant at 5% significance level.

Table 4
Parameter estimates for Model 11.

GFMNB-2	$\text{Ln}(\beta_0)$	β_1	β_2	β_3	β_4	β_5	β_6	ϕ[a]
Component 1								
Estimate	−24.193*	1.502	−0.075	9.253*	0.005*	−20.120*	3.024	0.589
Std. error	34.050	0.369	0.029	33.780	0.005	386.600	1.008	
Component 2								
Estimate	−7.335	0.973	−0.016	0.290	−0.006	−0.439*	−0.472	4.988
Std. error	1.034	0.098	0.007	0.122	0.002	0.337	0.120	
	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
Estimate	4.986	3.102	−4.73E−05	−0.080	−0.392	−0.016	−4.019	−1.065

[a] Inverse dispersion parameter $\phi = 1/\alpha$.
* Not significant at 5% significance level.

Table 5
Goodness-of-fit statistics for the Indiana data.

Model	Functional form	Degrees of freedom	Deviance	AIC	BIC
Standard NB	N/A	8	1884.512	1900.512	1931.096
1 (Fixed)	$\frac{w_{2,i}}{1-w_{2,i}} = e^{\gamma_0}$	17	1734.91	1768.91	1833.9
2	$\frac{w_{2,i}}{1-w_{2,i}} = e^{\gamma_0} e^{\gamma_1 L_i}$	18	1689.95	1725.95	1794.77
3	$\frac{w_{2,i}}{1-w_{2,i}} = e^{\gamma_0} L_i^{\gamma_1}$	18	1681.3	1717.3	1786.11
4	$\frac{w_{2,i}}{1-w_{2,i}} = e^{\gamma_1 L_i}$	17	1694.54	1728.54	1793.54
5	$\frac{w_{2,i}}{1-w_{2,i}} = L_i^{\gamma_1}$	17	1714.24	1748.24	1813.23
6	$\frac{w_{2,i}}{1-w_{2,i}} = F_i^{\gamma_2}$	17	1723.08	1757.08	1822.07
7	$\frac{w_{2,i}}{1-w_{2,i}} = e^{\gamma_0} F_i^{\gamma_2}$	18	1728.05	1764.05	1832.87
8	$\frac{w_{2,i}}{1-w_{2,i}} = e^{\gamma_0} e^{\gamma_2 F_i}$	18	1729.77	1765.77	1834.58
9	$\frac{w_{2,i}}{1-w_{2,i}} = e^{\gamma_0} L_i^{\gamma_1} F_i^{\gamma_2}$	19	1680.84	1718.84	1791.48
10	$\frac{w_{2,i}}{1-w_{2,i}} = L_i^{\gamma_1} F_i^{\gamma_2}$	18	1683.23	1719.23	1788.05
11	$\frac{w_{2,i}}{1-w_{2,i}} = e^{\gamma_0} e^{\boldsymbol{\gamma}\mathbf{x}_i}$ [a]	24	1668.13	1716.13	1807.89

[a] $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7)'$, and $\mathbf{x}_i = (L_i, F_i, FR_i, PT_i, MW_i, BR_i, RS_i)$ is a vector of all available explainable variables.

model provides suspicious grouping results (the number of sites for component 1 is even less than 30). As a result, the estimated coefficients for component 1 are counterintuitive (see parameter results for Model 1 in Table 3). On the other hand, the parameter estimation results in Table 3 indicate that some GFMNB-2 models can classify the sites more appropriately. This result corroborates the finding from Zou et al. (2013), and emphasizes that it is beneficial to model the weight parameter by including explanatory variables as covariates. If we compare the mean functional forms in these two studies, Zou et al (2013) only considered a limited number of covariates in the mean functional form. However, this study investigates the effect of various geometric factors in addition to traffic flow and segment length. Interestingly, the results in Tables 3–6 suggest that in the presence of a well-defined mean function, the difference in group classification between FMNB-2 and GFMNB-2 models is still significant. For the Indiana data, some GFMNB-2 models are favored over the FMNB-2 model based on the reported goodness-of-fit statistics and classification results.

Between Models 2 and 11, Model 11 has smaller overall VMRs than that of other models. For the two components in Model 11, there is a remarkable difference in the mean values of segment length and presence of median barrier. The difference in the mean value of other variables is not so noticeable in Model 11. Fig. 1(b) shows that many of the small segment lengths are assigned to component 1, resulting in a low average value, and most of the long segment lengths are associated with component 2. Fig. 1(f) shows that most of the road segments with median barrier are assigned to component 1, resulting in a high average value, and only three road segments with median barrier are associated with component 2. Fig. 1(a) demonstrates that traffic flow has similar effects on crashes per mile between two components. Overall, since there is no significant difference in the distributions of other explanatory variables between two components, we can thus conclude that the variability in segment length and presence of median barrier are two important sources of dispersion observed in the data. These findings agree with findings from Zou et al. (2013) that segment length plays a more significant role than traffic flow in explaining the unobserved heterogeneity in roadway crash data. Previously, Hauer (2001) and Geedipally et al. (2009) discussed the relationship between segment length and dispersion in crash modeling.

Table 6
Summary statistics of each component for the Indiana data.

GFMNB-2	Crashes	*FR*	*PT*	*MW*	*BR*	*RS*	*L*	*F*
Model 1								
Component 1 (23[a])								
Mean	8.91	31.08	0.74	71.31	0.13	0.52	0.34	46,846.77
SD	12.60	7.85	0.45	37.65	0.34	0.51	0.32	41,531.19
VMR[b]	17.81							
Component 2 (315)								
Mean	17.56	30.47	0.77	66.67	0.16	0.74	0.93	29,024.84
SD	37.39	6.59	0.42	33.94	0.37	0.44	1.53	27,316.81
VMR	79.59							
Model 2								
Component 1 (93)								
Mean	2.37	30.71	0.71	77.42	0.16	0.61	0.17	30,619.90
SD	8.07	7.53	0.46	37.63	0.37	0.49	0.19	28,336.53
VMR	27.53							
Component 2 (245)								
Mean	22.52	30.44	0.79	63.02	0.16	0.77	1.16	30,092.44
SD	41.02	6.33	0.41	31.96	0.37	0.42	1.66	28,997.81
VMR	74.73							
Model 3								
Component 1 (234)								
Mean	23.58	30.37	0.79	62.27	0.15	0.76	1.21	29,990.92
SD	41.69	6.28	0.41	31.03	0.35	0.42	1.68	28,604.85
VMR	73.71							
Component 2 (104)								
Mean	2.12	30.84	0.73	77.59	0.19	0.63	0.16	30,792.53
SD	7.58	7.52	0.45	38.44	0.40	0.48	0.18	29,290.69
VMR	27.13							
Model 4								
Component 1 (77)								
Mean	2.06	32.03	0.66	78.87	0.04	0.66	0.19	25,464.78
SD	7.45	7.29	0.48	38.80	0.19	0.48	0.20	26,820.32
VMR	26.86							
Component 2 (261)								
Mean	21.37	30.07	0.80	63.48	0.20	0.74	1.09	31,645.64
SD	40.08	6.43	0.40	31.92	0.40	0.44	1.63	29,228.76
VMR	75.16							
Model 5[c]								
Component 1 (254)								
Mean	5.51	30.79	0.76	69.10	0.20	0.70	0.36	32,780.92
SD	13.42	6.76	0.43	35.22	0.40	0.46	0.35	32,094.35
VMR	32.69							
Component 2 (84)								
Mean	51.64	29.69	0.81	60.59	0.04	0.81	2.47	22,546.96
SD	56.41	6.39	0.40	30.05	0.19	0.40	2.27	11,949.06
VMR	61.61							
Model 6								
Component 1 (319)								
Mean	17.38	30.43	0.78	66.72	0.16	0.74	0.92	29,002.28
SD	37.21	6.57	0.42	34.04	0.37	0.44	1.52	27,177.43
VMR	79.70							
Component 2 (19)								
Mean	10.21	31.86	0.63	71.41	0.16	0.42	0.37	50,977.48

Table 6 (*continued*)

GFMNB-2	Crashes	*FR*	*PT*	*MW*	*BR*	*RS*	*L*	*F*
SD	12.22	8.30	0.50	37.01	0.37	0.51	0.41	44,425.70
VMR	14.63							
Model 7								
Component 1 (311)								
Mean	18.15	30.46	0.79	68.23	0.12	0.76	0.92	24,190.97
SD	37.52	6.64	0.40	33.98	0.32	0.43	1.53	16,804.49
VMR	77.57							
Component 2 (27)								
Mean	3.44	31.14	0.48	52.62	0.63	0.37	0.52	99,885.48
SD	9.30	7.21	0.51	33.67	0.49	0.49	0.60	43,491.68
VMR	25.13							
Model 8								
Component 1 (258)								
Mean	19.90	30.76	0.80	68.29	0.04	0.81	1.03	19,644.05
SD	39.74	6.55	0.40	32.48	0.20	0.39	1.65	10,048.79
VMR	79.35							
Component 2 (80)								
Mean	7.54	29.73	0.68	62.76	0.54	0.45	0.44	64,401.66
SD	19.12	7.06	0.47	39.05	0.50	0.50	0.50	40,687.81
VMR	48.51							
Model 9								
Component 1 (150)								
Mean	33.79	30.36	0.77	62.76	0.02	0.78	1.61	19,953.42
SD	48.34	6.57	0.42	29.11	0.14	0.42	1.96	10,582.17
VMR	69.17							
Component 2 (188)								
Mean	3.56	30.64	0.77	70.36	0.27	0.68	0.31	38,443.01
SD	10.23	6.77	0.42	37.46	0.45	0.47	0.36	35,363.36
VMR	29.39							
Model 10								
Component 1 (228)								
Mean	24.11	30.34	0.79	61.99	0.15	0.76	1.24	30,541.27
SD	42.10	6.28	0.41	31.07	0.36	0.43	1.69	28,947.35
VMR	73.53							
Component 2 (110)								
Mean	2.19	30.86	0.73	77.33	0.17	0.65	0.16	29,608.09
SD	7.44	7.44	0.45	37.93	0.38	0.48	0.18	28,540.14
VMR	25.25							
Model 11								
Component 1 (152)								
Mean	1.99	31.43	0.76	76.58	0.34	0.65	0.30	41,305.25
SD	6.39	7.02	0.43	40.62	0.47	0.48	0.36	38,256.57
VMR	20.58							
Component 2 (186)								
Mean	29.22	29.77	0.78	59.14	0.02	0.78	1.37	21,193.02
SD	45.07	6.30	0.42	25.35	0.13	0.41	1.84	11,444.57
VMR	69.52							

[a] The number of observations in the component.

[b] Variance to mean ratio.

[c] *Note*: Compared with other models, Model 5 can provide the least overall VMRs and the classification results are reasonable.

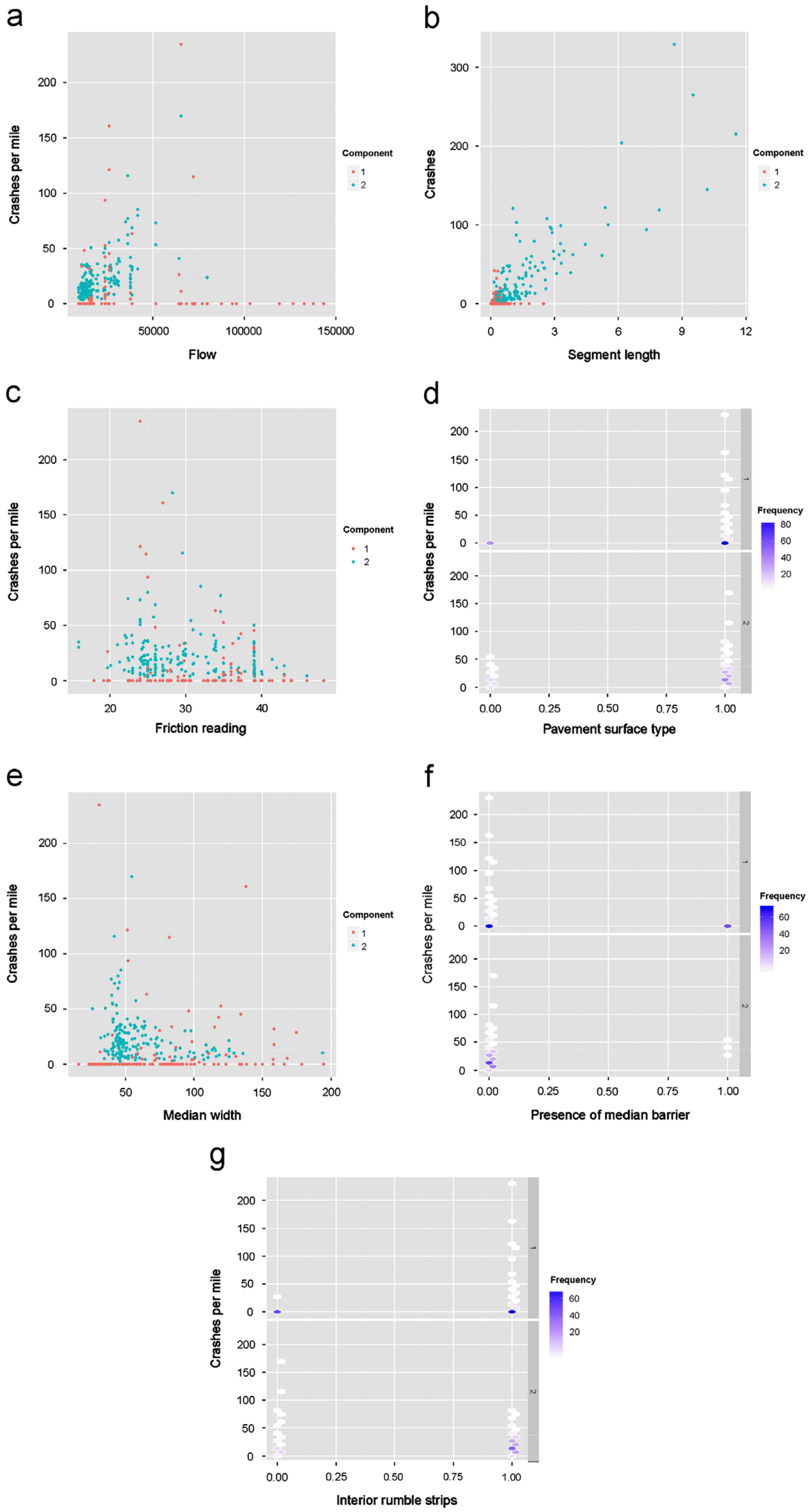

Fig. 1. Scatter plots of the two components for the Indiana data (Model 11). (a) Crashes per mile against the flow, (b) crashes against segment length, (c) crashes per mile against friction reading, (d) crashes per mile against pavement surface type, (e) crashes per mile against median width, (f) crashes per mile against presence of median barrier and (g) crashes per mile against interior rumble strips.

5. Discussion

In this paper, the results are very interesting and deserve further discussions. As discussed by Park and Lord (2009) and later confirmed by Zou et al. (2013), the finite mixture of NB regression models can be used to determine the sources of dispersion observed in crash data. However, the findings in the previous two studies are based on the analysis of data with limited covariates for the mean function. For example, Park and Lord (2009) considered mainly the effect of major and minor road traffic flow for urban intersections of Toronto. In this study, sufficient explanatory variables are used to model the mean, and it shows that GFMNB-2 models are preferred over FMNB-2 models based on the goodness-of-fit statistics and group classification results. If this result is generalizable, it suggests that the modeling of the weight parameter (which essentially helps in improving the resulting classification) is generally necessary when using the finite mixture of NB regression models to analyze the crash data, even in the presence of a well-defined mean function.[4]

Other important conclusions are summarized as follows. First, the results support the work of Zou et al. (2013) who found that the selection of the functional form for weight parameter has a significant impact on the classification results. If the functional form for weight parameter is mis-specified, counterintuitive coefficient estimates and erroneous inferences may be drawn. Thus, it is suggested that transportation safety analysts should evaluate different functional forms describing the weight parameter when using the GFMNB-2 models. Second, models with best fitting performance may not necessarily be the optimal model for grouping. The selection of the best functional form should be based on not only the goodness-of-fit statistics, but also the resulting classification. As discussed by Miaou and Lord (2003), for a given dataset, a large number of plausible functional forms with similar goodness-of-fit statistics are possible. To avoid over-fitting and over-interpretation of the data, some goodness-of-logic measures should be used to determine the appropriate functional form. In our case, when selecting the optimal weight structure, the group classification results from GFMNB-2 models should be emphasized. With a more reasonable grouping result, the nature of the over-dispersion in the data might be better identified. Third, it can be observed that Model 11 with the least overall VMR can be considered as the best functional form. Thus, based on the modeling results from this study, when using the GFMNB-2 model to analyze the crash data, transportation safety analysts are suggested to consider the linear combination of all explanatory variables as an alternative for modeling the weight parameter.

Crash data are often characterized by a small number of observations, and one important issue associated with the small sample size is that the modeling results from FMNB-2 and GFMNB-2 models may suffer from the small sample bias. As documented in Lord (2006), data characterized by small sample size can result in biased estimated coefficients and erroneous inferences within the NB modeling framework. Similar conclusions can be made in this study. As shown in Table 6, the number of observations for some components is even less than 50 (there are 338 observations in the Indiana data). As a result, the estimated coefficients for these components may be counterintuitive (see Models 1, 6 and 7 in Table 3) and the small sample size can affect the estimation of the dispersion parameter as well. This is why although Model 1 seems to provide a good overall VMR, this model provides suspicious coefficient estimates and its functional form should not be considered as a promising structure for the weight parameter in this study. Thus, when applying the FMNB-2 and GFMNB-2 models to crash data analysis, transportation safety modelers should carefully assess whether the size of the data sample are appropriate for parameter estimation and statistical inferences. Previously, Park et al. (2010a) investigated the bias associated with the Bayesian summary statistics in FMNB-2 models for different sample sizes and sample-mean values.

6. Summary and conclusions

This study builds upon the work of Park and Lord (2009) and Zou et al. (2013), and the original motivation is to investigate the effect of different functional forms on estimation of the weight parameter as well as the group classification, using the crash data collected on rural road sections in Indiana. During the course of the study, some interesting findings emerged. First, contrary to the discussion of the varying dispersion parameter for the standard NB model (see Mitra and Washington, 2007; Washington et al., 2011), this research emphasizes that even in the presence of a well-defined mean function, the varying weight structure should be recommended for the finite mixture of NB regression models (also argued in Geedipally et al. (2009)). Second, the transportation safety researchers are cautioned about applying the finite mixture of NB regression models to the crash data with a small sample size. In terms of the effect of functional forms on the modeling results, this study confirms that the selection of the functional form for weight parameter will affect the classification results significantly. Overall, among 11 different functional forms investigated, the results suggest that Model 11 which uses the linear combination of all explanatory variables can be considered as one of the best parameterizations. Therefore, it is suggested that transportation safety analysts should include Model 11 along with other alternative functional forms for describing the weight parameter and select the most appropriate functional form based on not only the goodness-of-fit statistics, but also the classification results. For future work, the methodology adopted in this study can be extended to other types of crash datasets (i.e., urban and rural intersections or road segments) to potentially further corroborate the findings from this study.

[4] Note that this finding is empirical and not theoretical. As a result, it may be possible that the modeling of weight parameter could become insignificant for other datasets. However, the evidence for this possibility is lacking, based on our study.

References

Anastasopoulos, P., Tarko, A., Mannering, F., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention 40 (2), 768–775.

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

Bhat, C., 1997. Endogenous segmentation mode choice model with an application to intercity travel. Transportation Science 31 (1), 34–48.

Chang, I., Kim, S., 2012. Modelling for identifying accident-prone spots: Bayesian approach with a Poisson mixture model. KSCE Journal of Civil Engineering 16 (3), 441–449.

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

El-Basyouny, K., Sayed, T., 2010. Safety performance functions with measurement errors in traffic volume. Safety Science 48 (10), 1339–1344.

Eluru, N., Bagheri, M., Miranda-Moreno, L., Fu, L., 2012. A latent class modelling approach for identifying vehicle driver injury severity factors at highway–railway crossings. Accident Analysis and Prevention 47 (1), 119–127.

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

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

Geedipally, S., Lord, D., Dhavala, S., 2012. The negative binomial-Lindley generalized linear model: characteristics and application using crash data. Accident Analysis and Prevention 45, 258–265.

Greene, W., Hensher, D., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37 (8), 681–698.

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

Jun, J., 2010. Understanding the variability of speed distributions under mixed traffic conditions caused by holiday traffic. Transportation Research Part C 18 (4), 599–610.

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., Park, P., 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., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44 (5), 291–305.

Malyshkina, N., Mannering, F., Tarko, A., 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, New York.

Miaou, S., Lord, D., 2003. Modeling traffic-flow relationships at signalized intersections: dispersion parameter, functional form and Bayes vs Empirical Bayes. Transportation Research Record 1840, 31–40.

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

Park, B., Lord, D., Hart, J., 2010a. Bias properties of Bayesian statistics in finite mixture of negative binomial regression models in crash data analysis. Accident Analysis and Prevention 42 (2), 741–749.

Park, B., Zhang, Y., Lord, D., 2010b. Bayesian mixture modeling approach to account for heterogeneity in speed data. Transportation Research Part B 44 (5), 662–673.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection accident frequency. Journal of Transportation Engineering 122 (2), 105–113.

Qin, X., Reyes, P., 2011. Conditional quantile analysis for crash count data. Journal of Transportation Engineering 137 (9), 601–607.

Rigby, R., Stasinopoulos, D., 2010. A Flexible Regression Approach Using GAMLSS in R. ⟨http://gamlss.org/images/stories/papers/book-2010-Athens.pdf⟩ (accessed April 2012).

Sobhani, A., Eluru, N., Faghih-Imani, A., 2013. A latent segmentation based multiple discrete continuous extreme value model. Transportation Research Part B. http://dx.doi.org/10.1016/j.trb.2013.07.009, in press.

Washington, S., Karlaftis, M., Mannering, F., 2011. Statistical and Econometric Methods for Transportation Data Analysis, second edition Chapman and Hall/CRC, Boca Raton, FL.

Xiong, Y., Mannering, F., 2013. The heteroscedastic effects of guardian supervision on adolescent driver-injury severities: a finite mixture-random parameters approach. Transportation Research Part B 49, 39–54.

Zou, Y., Zhang, Y., 2011. Use of skew-normal and skew-t distributions for mixture modeling of freeway speed data. Transportation Research Record 2260, 67–75.

Zou, Y., Zhang, Y., Lord, D., 2013. Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis. Accident Analysis and Prevention 50, 1042–1051.

Zou, Y., Zhang, Y., Zhu, X., 2012. Constructing a bivariate distribution for freeway speed and headway data. Transportmetrica, 1–18, http://dx.doi.org/10.1080/18128602.2012.745099.