Accident Analysis and Prevention 131 (2019) 112–121

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Adjusting finite sample bias in traffic safety modeling

Huiying Mao[a], Xinwei Deng[a], Dominique Lord[b], Gerardo Flintsch[c,d], Feng Guo[a,c,*]

[a] Department of Statistics, Virginia Tech, Blacksburg, VA 24061, USA
[b] Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843-3136, USA
[c] Virginia Tech Transportation Institute, Virginia Tech, Blacksburg, VA 24061, USA
[d] Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA

ARTICLE INFO

Keywords:
Finite sample bias
Maximum likelihood estimate
Poisson regression
Negative binomial regression
Traffic safety

ABSTRACT

Poisson and negative binomial regression models are fundamental statistical analysis tools for traffic safety evaluation. The regression parameter estimation could suffer from the finite sample bias when event frequency is low, which is commonly observed in safety research as crashes are rare events. In this study, we apply a bias-correction procedure to the parameter estimation of Poisson and NB regression models. We provide a general bias-correction formulation and illustrate the finite sample bias through a special scenario with a single binary explanatory variable. Several factors affecting the magnitude of bias are identified, including the number of crashes and the balance of the crash counts within strata of a categorical explanatory variable. Simulations are conducted to examine the properties of the bias-corrected coefficient estimators. The results show that the bias-corrected estimators generally provide less bias and smaller variance. The effect is especially pronounced when the crash count in one stratum is between 5 and 50. We apply the proposed method to a case study of infrastructure safety evaluation. Three scenarios were evaluated, all crashes collected in three years, and two hypothetical situations, where crash information was collected for "half-year" and "quarter-year" periods. The case-study results confirm that the magnitude of bias correction is larger for smaller crash counts. This paper demonstrates the finite sample bias associated with the small number of crashes and suggests bias adjustment can provide more accurate estimation when evaluating the impacts of crash risk factors.

1. Introduction

Accurately evaluating the crash risk associated with transportation infrastructure and driver characteristics is essential to improving safety. Traffic safety is usually measured by crash frequency, which can be the number of crashes that occurred in a roadway segment, or the number of crashes a driver experienced, over a specified period (e.g., one year) (AASHTO, 2010). As crashes are rare events, it is not uncommon to see crash-count data with low sample mean and excessive zero responses (Lord and Mannering, 2010; Lord and Geedipally, 2018). Poisson and negative binomial (NB) regression have been the fundamental statistical analysis tools for count data; however, a small number of events brings challenges to parameter estimation and inference. This paper focuses on the finite sample bias for parameter estimation caused by a small number of events when fitting a Poisson or NB regression model.

Poisson and NB regression models are important methods in transportation safety studies. For instance, safety performance function, an essential tool to evaluate crash risk provided by *The Highway Safety Manual*, is based on the Poisson/NB regression (AASHTO, 2010). NB regression assumes that the rate in a Poisson model follows a Gamma distribution and can accommodate over-dispersion, a common issue for crash-count data. The coefficient estimates of the two models can be used to evaluate risk factors associated with crashes.

There are other models developed for transportation safety studies by relaxing certain assumptions in the Poisson and NB regressions. For example, the generalized estimating equation (GEE) and random/mixed effect models can be used when data violate the independence assumption. The GEE model takes into account the within-subject correlation, such as spatial-temporal correlation or observations with repeated measures, by an empiric covariance matrix (Lord and Persaud, 2000; Lord et al., 2005). The random/mixed effect model assumes a distribution for the unobserved effect over subjects, and it can handle multiple sources of correlation (Guo et al., 2010; Guo and Fang, 2013; Chen and Tarko, 2014). Semi-parametric models and generalized additive models relax the linear assumption of Poisson and NB models to accommodate more-complicated relationships between crash rate and risk factors (Xie and Zhang, 2008; Li et al., 2010).

As crashes are rare events, the number of crashes for a specific road

* Corresponding author at: Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA.
E-mail address: feng.guo@vt.edu (F. Guo).

https://doi.org/10.1016/j.aap.2019.05.026
Received 27 August 2018; Received in revised form 22 February 2019; Accepted 29 May 2019
Available online 25 June 2019
0001-4575/ © 2019 Elsevier Ltd. All rights reserved.

segment or a driver is usually limited, leading to a low average occurrence rate and/or a large number of zeros in the crash-count data (Shankar et al., 1997; Kumara and Chin, 2003). The low event frequency can cause biased estimation and inaccurate inference for the count models (Fridstrøm et al., 1995; Lord and Bonneson, 2005; Lord and Miranda-Moreno, 2008). Lord (2006) showed that a low sample mean and small sample size can seriously affect the estimation of the dispersion parameter of the NB model. The commonly used goodness-of-fit test statistics (scaled deviance and Pearson's χ^2) are also inappropriate for the low mean value problem (Wood, 2002). When excessive zero responses exist in the dataset, both the Poisson model and the NB model will produce inaccurate prediction (Lord and Geedipally, 2011). For a thorough review of the low-occurrence problem in transportation safety, please refer to Lord and Mannering (2010) and Lord and Geedipally (2018).

Zero-inflated models, proposed by Lambert (1992), are commonly used in analyzing crash counts with an excessive number of zeros (Miaou, 1994; Shankar et al., 1997, 2003; Carson and Mannering, 2001; Lee and Mannering, 2002; Kumara and Chin, 2003; Qin et al., 2004; Aguero-Valverde, 2013; Songpatanasilp et al., 2015; Anastasopoulos, 2016). Nevertheless, they have also been subject to several criticisms (Lord et al., 2005, 2007; Malyshkina et al., 2009). In fact, Lord et al. (2005) argued that the assumed safe period with an event rate being zero in zero-inflated models does not reflect the actual crash-generating process. Alternatively, many methods have been proposed and applied. For instance, Malyshkina and Mannering (2010) proposed a zero-state Markov switching model and Park and Lord (2009) applied finite mixture models to such datasets. Recently, researchers have proposed more-flexible models, including Sichel (SI), Negative Binomial-Lindley (NB-L), Negative Binomial-Generalized Exponential (NB-GE), and Negative Binomial-Crack (NB-CR). These models incorporate more parameters into the underlying count distribution so that the extra degree of freedom can account for the excessive number of zeros (Zou et al., 2013, 2015; Lord and Geedipally, 2011, 2018; Vangala et al., 2015). These models may better fit the data, but they are hard to estimate and difficult for practitioners to understand.

Poisson and NB regression models are generally estimated using the maximum likelihood method. The resultant estimators for the unknown parameters are called maximum likelihood estimators (MLEs). The MLEs have the consistency property that ensures the MLEs converging to the true values when the sample size is sufficiently large. However, McCullagh and Nelder (1989) showed that the MLEs could be biased and the bias is not negligible for a modest sample size. When the number of events is limited, a bias adjustment to the MLEs can improve the parameter estimation. Generally, there are two types of approaches to bias reduction for MLEs. One approach is based on applying the Jefferys invariant prior to the likelihood function to directly generate an improved estimator (Firth, 1993; Kosmidis and Firth, 2009; Kosmidis et al., 2010). The other approach reduces the bias by subtracting the approximated bias from the regular MLE (McCullagh and Nelder, 1989; Cordeiro and McCullagh, 1991; Lambrecht et al., 1997). McCullagh and Nelder (1989) determined a specific correction formula for the coefficient estimation of generalized linear models (GLMs).

The finite sample bias of Poisson and NB regression models has been sporadically investigated in the literature (Saha and Paul, 2005; Giles and Feng, 2011). Saha and Paul (2005) studied the bias-corrected dispersion parameter estimation of the NB regression, which showed less bias and superior efficiency compared to the MLE, the method of moments estimator, and the maximum extended quasi-likelihood estimators in most instances. Giles and Feng (2011) derived the bias-correction formula for the parameter estimation of Poisson regression from Cox and Snell's (1968) general definition of residuals. Although considerable research has been devoted to reducing the bias of MLE under Poisson and NB models, limited research has been conducted in transportation safety to identify the situations where the bias correction is necessary and factors affecting the magnitude of bias.

This paper aims to study the finite sample bias for the parameter estimation of Poisson and NB regression models in the context of traffic safety modeling. We demonstrate a bias-correction procedure based on the approximated bias provided by McCullagh and Nelder (1989), followed by deriving the explicit bias correction formula for one special scenario, Poisson regression with a single binary explanatory variable. Using a Monte Carlo simulation study, we quantitatively evaluate the magnitude of bias and identify factors affecting the bias. We apply the bias-correction method to an infrastructure safety evaluation, which involves a three-year crash dataset collected from road segments with different pavement types. We also examine the relationship between the bias correction magnitude and crash counts by hypothetically reducing the number of crashes in the pavement data.

The remainder of this article is organized as follows. Starting from the basic setup of the Poisson and NB regressions and the analytic bias results of McCullagh and Nelder (1989), Section 2 derives the explicit bias-correction formula for one illustrative special scenario. Section 3 examines the benefit of bias correction through a Monte Carlo simulation study, which also elucidates when the bias correction starts to make a difference and to what extent the bias-adjustment procedure is beneficial. Section 4 demonstrates the bias correction through a real-case safety application and two hypothetical situations by reducing the number of crashes. Section 5 summarizes this work with some discussion.

2. Method

Poisson and NB regression models assume that the frequency of events Y_i, e.g., the crash count, follows a Poisson distribution,

$$Y_i \sim \text{Poisson}(\lambda_i \cdot E_i), \quad i = 1, 2, \ldots, n, \tag{1}$$

where λ_i is the crash occurrence rate for the ith road segment or the ith driver. The λ_i is a constant in the Poisson regression and a random variable in the NB regression, respectively. In the NB regression, random variable λ_i follows a Gamma distribution, i.e.,

$$\lambda_i \sim \text{Gamma}(k, \mu_i),$$

where μ_i is the mean of the crash occurrence rate λ_i and $1/k$ is the dispersion parameter of the NB regression. The NB regression is more dispersed for smaller k. The Poisson regression is a special case of the NB regression when $k = \infty$. In Eq. (1), the E_i is the corresponding exposure, which could be the length of the observation period or the total vehicle miles traveled.

A logarithm link function is used to link the event rate λ_i in the Poisson regression or the expected event rate μ_i in the NB regression with a linear transformation of p explanatory variables, $X_{i1}, X_{i2}, \ldots, X_{ip}$. That is,

$$\begin{aligned} \log(\lambda_i) = \eta_i \quad \text{or} \quad \log(\mu_i) = \eta_i, \\ \eta_i = \mathbf{X}_i' \boldsymbol{\beta}, \end{aligned} \tag{2}$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$; $\mathbf{X_i}$ is the covariates vector for entity i, $\mathbf{X_i} = (1, X_{i1}, \ldots, X_{ip})'$. The coefficient β_j indicates the impact of the jth variable on crash risk, $j = 1, \ldots, p$. The estimation of $(\beta_1, \ldots, \beta_p)$ is the focus of the safety evaluation.

2.1. Adjusting finite sample bias for regression coefficient estimation

The MLE for the regression coefficient $\boldsymbol{\beta}$ is obtained by maximizing the log-likelihood function

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[-e^{\mathbf{X}_i \boldsymbol{\beta}} \cdot E_i + Y_i (\mathbf{X}_i' \boldsymbol{\beta} + \log(E_i)) - \log(Y_i!) \right]. \tag{3}$$

Denote the MLE as $\hat{\boldsymbol{\beta}}$, which, in general, has no closed-form expression. The estimation is typically based on a numerical method, such as the Newton-Raphson method.

The MLE $\hat{\boldsymbol{\beta}}$ is asymptotically unbiased and normally distributed. However, $\hat{\boldsymbol{\beta}}$ in general is a biased estimator and the difference between $\hat{\boldsymbol{\beta}}$ and the true value $\boldsymbol{\beta}$ might not negligible for a small sample size (McCullagh and Nelder, 1989; Firth, 1993). The approximation of MLEs' bias can be traced at least as far back as Bartlett (1953), being a side-product of when Bartlett studied the confidence interval for one unknown parameter from a random sample. Cox and Snell (1968) extended Bartlett's (1953) result to multidimensional MLEs. McCullagh (1987) derived the bias using a similar procedure but with tensor notation and showed that the bias is of order $O(n^{-1})$. McCullagh and Nelder (1989) approximated the bias of GLMs with a canonical link using the leading $O(n^{-1})$ term. Since Poisson regression and NB regression are GLMs and the log link is a canonical link, we apply the bias approximation for $\hat{\boldsymbol{\beta}}$ provided by McCullagh and Nelder (1989) as

$$\text{bias}(\hat{\boldsymbol{\beta}}) \approx (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\boldsymbol{\xi}, \tag{4}$$

where $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)'$, $\mathbf{W} = \text{cov}(\mathbf{Y})$, and $\boldsymbol{\xi}$ is an n-dimensional vector with the ith element being $\xi_i = -\frac{1}{2}Q_{ii}\frac{\kappa_{3i}}{\kappa_{2i}}$. Q_{ii} is the ith diagonal element of the matrix $\mathbf{Q} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'$, $\kappa_{3i} = \partial^2\lambda_i/\partial\eta_i^2$, and $\kappa_{2i} = \partial\lambda_i/\partial\eta_i$. With log-link, $\kappa_{3i} = e^{\eta_i}$, and $\kappa_{2i} = e^{\eta_i}$. Therefore, the ξ_i is reduced to $\xi_i = -\frac{1}{2}Q_{ii}$. An estimate of the approximated bias in Eq. (4) can be

$$\widehat{\text{bias}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\,\hat{\mathbf{W}}\,\hat{\boldsymbol{\xi}}, \tag{5}$$

where $\hat{\mathbf{W}}$ and $\hat{\boldsymbol{\xi}}$ are obtained by plugging in the regular MLE $\hat{\boldsymbol{\beta}}$. The bias-corrected coefficient estimate $\tilde{\boldsymbol{\beta}}$ can be calculated as

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \widehat{\text{bias}}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\,\hat{\mathbf{W}}\,\hat{\boldsymbol{\xi}}. \tag{6}$$

The bias-correction procedure in (6) is applicable to both the Poisson and NB models. The only difference is that the $\hat{\mathbf{W}}$ of an NB model involves the estimated dispersion parameter while that of a Poisson model does not. The procedure can be applied for both continuous and discrete explanatory variables. In both cases, the covariates matrix $\mathbf{X}$ represents the corresponding explanatory variable data.

To provide a more illustrative understanding for the finite sample bias of $\boldsymbol{\beta}$, we obtain the explicit bias formula for the following special case of the Poisson/NB regression.

2.2. A special case: Negative binomial regression with one binary explanatory variable

Consider a NB model with a single dichotomous exploratory variable, i.e.,

$$\log(\mu_i) = \beta_0 + \beta_1 X_i,$$

where X_i is either 0 or 1. In epidemiology, the observation group with $X_i = 0$ is typically referred to as the reference group and the group with $X_i = 1$ is the treatment group. The coefficient β_1 represents the treatment effect, the impact of $\{X_i\}_{i=1}^n$ on the response variable $\{Y_i\}_{i=1}^n$, and it is the focus of the evaluation.

From Eq. (4), the approximated bias for $\hat{\beta}_1$ is

$$\text{bias}(\hat{\beta}_1) = \frac{1}{2V_0} - \frac{1}{2V_1}, \tag{7}$$

where

$$V_0 = \text{Var}\left(\sum_i Y_i|X_i = 0\right) = \sum_{X_i=0}\left[\mu_0 E_i + \frac{(\mu_0 E_i)^2}{k}\right],$$

$$V_1 = \text{Var}\left(\sum_i Y_i|X_i = 1\right) = \sum_{X_i=1}\left[\mu_1 E_i + \frac{(\mu_1 E_i)^2}{k}\right].$$

A detailed derivation can be found in Appendix A. It is difficult to estimate V_0 and V_1, the variance of crash counts in each stratum. The estimation of V_0 and V_1 depends on the estimation of μ_0, μ_1, and k, which often requires an iterative estimation procedure. Practically, under the finite sample situation, $\mu_0 E_i$ or $\mu_1 E_i$ for one observation is typically small. The magnitude of higher order terms $\frac{(\mu_0 E_i)^2}{k}$ and $\frac{(\mu_1 E_i)^2}{k}$ are much smaller than $\mu_0 E_i$ and $\mu_1 E_i$, respectively. Thus we consider to approximate V_0 and V_1 by C_0 and C_1, the expected number of crashes in the reference group and treatment group.

$$C_0 = \mathbf{E}\left(\sum_i Y_i|X_i = 0\right) = \sum_{X_i=0}\mu_0 E_i \approx V_0,$$

$$C_1 = \mathbf{E}\left(\sum_i Y_i|X_i = 1\right) = \sum_{X_i=1}\mu_1 E_i \approx V_1.$$

Therefore, the bias of $\hat{\beta}_1$ can be approximated by

$$\text{bias}(\hat{\beta}_1) \approx \frac{1}{2C_0} - \frac{1}{2C_1}, \tag{8}$$

where the expected crash counts in each stratum, C_0 and C_1, can be estimated by the observed number of crashes in the reference group and treatment group. The balance of crash counts in different stratum also matters to the magnitude of bias. The magnitude would be larger when the number of crashes in different stratum were more unbalanced.

3. Simulation

We conducted a Monte Carlo simulation study to evaluate the performance of the regular MLE $\hat{\boldsymbol{\beta}}$ and the bias-corrected MLE $\tilde{\boldsymbol{\beta}}$. The objectives were to examine if the bias-correction procedure could lead to a more accurate estimation, to identify the non-negligible finite sample bias situations, the magnitude of bias, and the factors affecting bias.

Without loss of generality, we consider observations generated from an NB regression, whose expected event rate is associated with a binary predictor variable. That is,

$$\begin{aligned} Y &\sim \text{Poisson}(\lambda\cdot E) \\ \lambda &\sim \text{Gamma}(k,\ \mu) \end{aligned} \tag{9}$$

where

$$\mu = e^{\beta_0+\beta_1 X}, \quad X = 0 \text{ or } 1.$$

We generated $n = 5000$ observations from the above NB model. The distribution of the exposure, E_i, is similar to the application described in Section 4. The specific simulation setup for β_0 and β_1 can be found in Table 1, where μ_0 is the expected event rate for the reference group and μ_1 is the expected event rate for the treatment group. The simulation parameters are setup such that the expected number of crashes within each group, C_0 and C_1, range from 4 to 1000. Within which, 2500 observations come from the reference group ($X = 0$) and 2500 observations are from the treatment group ($X = 1$). We enumerate the parameter k in model (9) within {0.5, 50, ∞}. For each scenario of model setting, the simulation is repeated 1000 times. Note that the NB model will get close to a Poisson model when the value of k becomes large. When $k = \infty$, the simulation is generated from the Poisson regression.

Fig. 1 shows the probability of both the reference group and treatment group having at least one events when C_0 and C_1 range from 1 to 1000. When the expected number of total crashes in either group is

Table 1
Simulation setup for β_0 and β_1, and $\mu_0 = \exp(\beta_0)$, $\mu_1 = \exp(\beta_0 + \beta_1)$.

β_0	β_1	μ_0	μ_1
1.5	−1.0	4.48	1.65
1.5	−0.5	4.48	2.72
1.5	−0.1	4.48	4.06
1.5	0.1	4.48	4.95
1.5	0.5	4.48	7.39
1.5	1.0	4.48	12.18

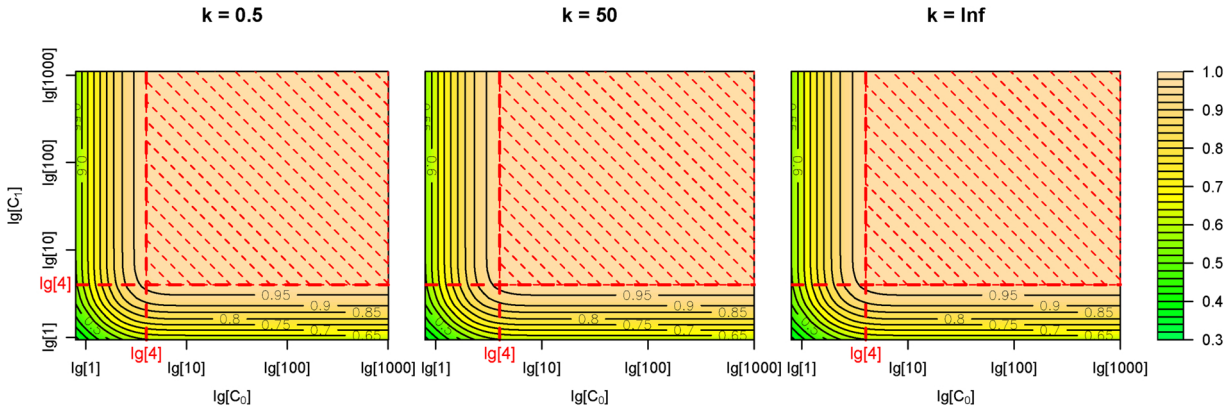

Fig. 1. The heatmap for the probability of at least one events in both the reference group and the treatment group (C_0 is the expected total number of crashes in the reference group and C_1 is that of the treatment group.)

greater than four, with probability approximately equal to one, there would be at least one event within each group. The shaded area indicates the scenarios included in the simulation. *x*-axis is the logarithm of the expected total number of crashes for the reference group C_0 to the base 10; *y*-axis is the logarithm of the expected total number of crashes for the treatment group C_1 to the base 10.

3.1. Simulation results

We present the simulation results in three parts. Recall that the regular MLE of β is denoted as $\hat{\boldsymbol{\beta}}$ and the bias-corrected MLE of β is denoted as $\tilde{\boldsymbol{\beta}}$. For each model setting, the simulation is repeated 1000 times. Firstly, we use the percentage bias of $\hat{\beta}_1$ and $\tilde{\beta}_1$ to show the effect of bias correction using some example cases. The percentage bias of $\hat{\beta}_1$ and $\tilde{\beta}_1$ are calculated as

$$\text{pct bias}(\hat{\beta}_1) = \frac{1}{1000}\sum \frac{\hat{\beta}_1 - \beta_1}{\beta_1} \times 100\%, \quad \text{pct bias}(\tilde{\beta}_1) = \frac{1}{1000}\sum \frac{\tilde{\beta}_1 - \beta_1}{\beta_1} \times 100\%,$$

where β_1 is the underlying true parameter as set in Table 1. Secondly, we use the side-by-side plot to compare the bias of $\hat{\boldsymbol{\beta}}_1$ and the bias of $\tilde{\boldsymbol{\beta}}_1$ for comprehensive simulation scenarios, where the biases are calculated as

$$\text{bias}(\hat{\beta}_1) = \frac{1}{1000}\sum \hat{\beta}_1 - \beta_1, \quad \text{bias}(\tilde{\beta}_1) = \frac{1}{1000}\sum \tilde{\beta}_1 - \beta_1.$$

Lastly, we use the difference between the variance of $\hat{\boldsymbol{\beta}}_1$ and the variance of $\tilde{\boldsymbol{\beta}}_1$ to show that the bias-corrected MLE has smaller variance than the regular MLE. The variance difference of the two estimators are calculated as

$$\text{var}(\hat{\beta}_1) - \text{var}(\tilde{\beta}_1) = \frac{1}{1000}\sum (\hat{\beta}_1 - \text{ave}(\hat{\beta}_1))^2 - \frac{1}{1000}\sum (\tilde{\beta}_1 - \text{ave}(\tilde{\beta}_1))^2,$$

where $\text{ave}(\hat{\beta}_1) = \frac{1}{1000}\sum \hat{\beta}_1$ and $\text{ave}(\tilde{\beta}_1) = \frac{1}{1000}\sum \tilde{\beta}_1$.

Fig. 2 plots the percent bias of $\hat{\beta}_1$ and $\tilde{\beta}_1$ when $\beta_1 = 0.1$ (event rate in treatment group is 10% higher than the reference group), $k = \infty$, and the expected number of crashes in the treatment group (C_1) is 9, 28, and 89. It shows that the bias-corrected estimator is more close to the true parameter value than the regular MLE for the majority cases. The percent bias of the bias-corrected estimator $\tilde{\beta}_1$ are around 0% except for some unlikely scenarios that the expected number of crashes in the reference group (C_0) is smaller than five, while the percent bias of the uncorrected MLE $\hat{\beta}_1$ are further away from zero percent, ranging from -60% to 100%. The left plot also shows that the bias of $\hat{\beta}_1$ decreases when the expected number of crashes in the reference group increases, which testifies our approximation of bias($\hat{\beta}_1$) in Eq. (7).

Fig. 3 comprehensively shows the bias of $\hat{\boldsymbol{\beta}}$ and the bias of $\tilde{\boldsymbol{\beta}}$ side-by-side for all the simulation scenarios. From the plot, we can see the regular MLE $\hat{\beta}_1$ underestimates β_1 when the expected number of events in the reference group is larger than the expected number of events in the treatment group ($C_0 > C_1$), and it overestimates β_1 when the expected number of events in the reference group is smaller than the expected number of events in the treatment group ($C_0 < C_1$). The bias is more severe when the difference between the expected numbers of events in the two groups is larger. In other words, the bias is smaller when the expected event counts in the two groups are more balanced.

For the bias-corrected MLE $\tilde{\beta}_1$, the bias is relatively small for a "larger area" in the plot, which again shows the effectiveness of bias-correction procedure for the majority scenarios. The procedure over-corrects for small expected number of crashes in either the reference group or the treatment group (the bottom and left "edges" in the right plots). Higher-order correction can help the estimation when the number of events for one group is small. In addition, the two "edges" get narrower as the *k* increases, which means the benefit of bias correction is more prominent when the *k* becomes larger.

Fig. 4 shows the difference between the sample variance of $\hat{\beta}_1$ and the sample variance of $\tilde{\beta}_1$. The difference are non-negative for all the scenarios plotted. Non-negative means that the bias-corrected coefficient estimate $\tilde{\beta}_1$ has smaller variance compared to the regular MLE $\hat{\beta}_1$, especially when the expected number of events in either group is limited. (The bottom-left corner has darker yellow.) A smaller sample variance means a more stable estimate, so the bias-corrected estimator $\tilde{\beta}_1$ is better.

For a short summary, the simulation study shows that (1) bias-corrected MLE $\tilde{\boldsymbol{\beta}}$ is less biased and has smaller variance compared to the regular MLE $\hat{\boldsymbol{\beta}}$; (2) the benefit is more substantial as the *k* increases; (3) the effect of bias correction is more pronounced when the number of crashes in one stratum of a categorical explanatory variable is less than 50; (4) the bias-corrected MLE $\tilde{\boldsymbol{\beta}}$, as well as the regular MLE, is unstable with too few events in one stratum of a categorical explanatory variable (i.e., when the number of crashes is less than five for Poisson and less than seven for Negative Binomial); (5) the balance of the crash counts within strata of a categorical explanatory variable also matters to the magnitude of bias.

4. Case study

To illustrate the benefit of bias correction and examine the magnitude of bias, we applied the bias-correction procedure to an infrastructure safety evaluation dataset. This dataset includes information from 5238 short road segments, which are collected from 2012 to 2014 in the State of Washington. The length for each segment is 0.1 mile. The number of crashes with property damage is the safety response. A total

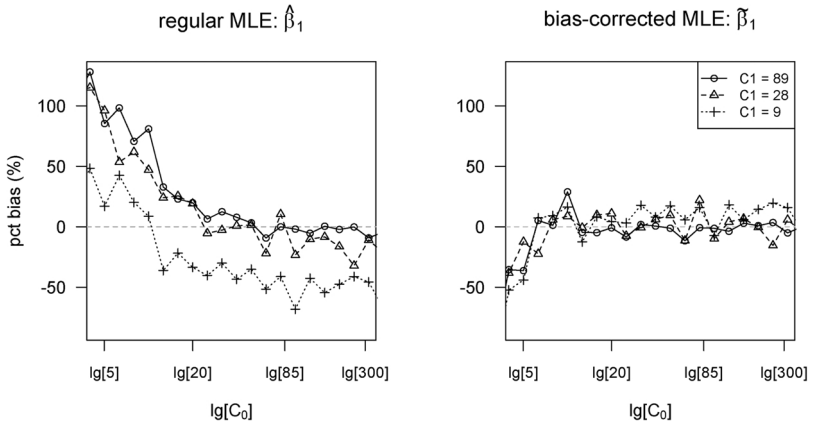

Fig. 2. The comparison between the percent bias of $\hat{\beta}_1$ (left) and the percent bias of $\tilde{\beta}_1$ (right) when $\beta_1 = 0.1$ and $k = \infty$ (C_0 is the expected total number of crashes in the reference group and C_1 is that in the reference group.)

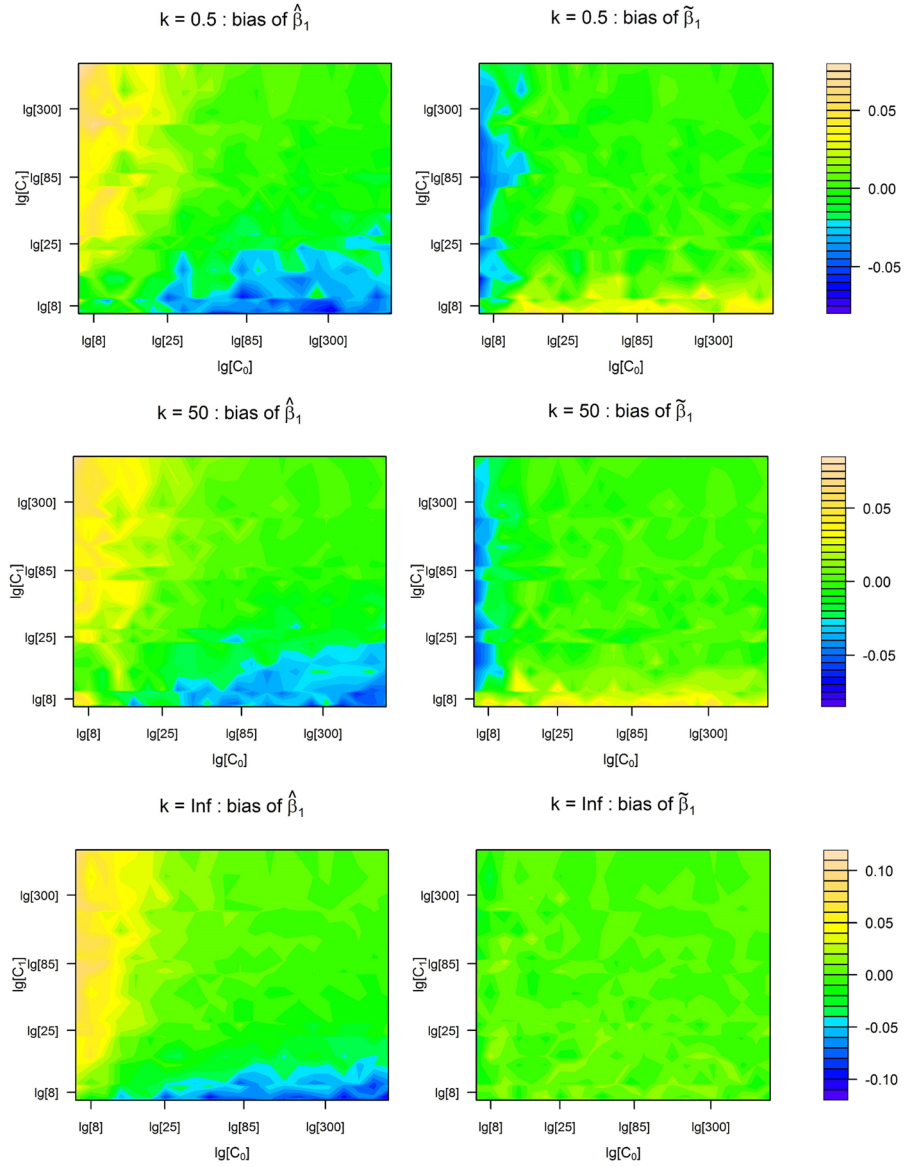

Fig. 3. The contour plots of the bias of $\hat{\beta}_1$ (left) and the bias of $\tilde{\beta}_1$ (right) (C_0 and C_1 is the expected total number of crashes in the reference and treatment group, respectively.). Here "yellow" represents positive bias (overestimation), "blue" represents negative bias (underestimation), and "green" means relatively small bias. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

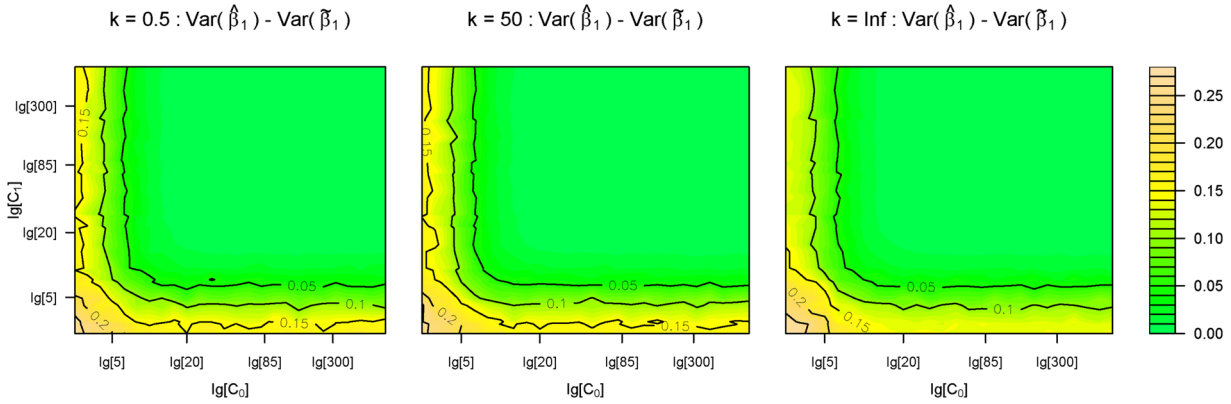

Fig. 4. The heatmap for the difference between the sample variance of $\hat{\beta}_1$ and the sample variance of $\tilde{\beta}_1$.

of 32,298 crashes was observed during the study period. There is 59.9 percent of zero responses in the dataset. Million vehicle miles traveled is used as the exposure. For the 5238 short road segments, the overall average crash rate is 2.7 per million vehicle miles traveled.

Table 2 lists the 12 covariates used in the analysis, including route type, whether the road segment is an entrance/exit, whether it is an intersection, whether it is a ramp, whether it is a wye connection, whether it is a divided highway, rural/urban, number of lanes, pavement type, friction, gradient, and horizontal curvature. For categorical variables, we list their number of observations and percentage in each stratum. For continuous variables, we present their mean and standard deviation.

The NB regression was implemented because of the existence of overdispersion. The estimated dispersion parameter is 2.06. The difference between bias-corrected coefficient estimates $\tilde{\boldsymbol{\beta}}$ and regular MLEs $\hat{\boldsymbol{\beta}}$, as well as the percentage change $\frac{\tilde{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}} \times 100\%$, are given in Table 3. The first stratum of each categorical variable is treated as the reference level, so there is no coefficient estimation and hence no bias correction for it.

Comparing the magnitude of bias correction along with the number of crashes (two columns under the "original dataset" tab), the bias correction is generally larger for a stratum having a smaller number of events. For example, the bias correction is the largest for the coefficient of 'BST' pavement type (14.8×10^{-3}, $-2.5\,\%$), which only has 20 crashes in its stratum. In addition, it is the number of events rather than the sample size that matters to the bias magnitude. For instance, there is only one observation for five lanes, but it has 103 crashes. The bias correction magnitude for this stratum is trivial.

To test if the number of crashes affects the magnitude of bias, we also conducted bias correction for two hypothetical pavement datasets where the crash count and exposure of each road segment is reduced to only 1/6 and 1/12, respectively, of the original pavement dataset. The covariates used in the two hypothetical pavement datasets are the same as the original dataset. The original data were collected for three years. Reducing the crash count and exposure to 1/6 is thus like the data being collected for only half a year, resulting in the total number of crashes being 5192. The number of crashes in each stratum of the categorical covariates and the bias correction magnitude can be found under the "'half-year' dataset" tab in Table 3. Similarly, reducing the crash count and exposure to 1/12 is like the data collection only lasting for a quarter of a year, resulting in a total crash count of 2502. Its results are under "'quarter-year' dataset." After reducing the number of crashes, the percentage of zero responses are 76.7 percent and 82.5 percent for the "half-year" dataset and the "quarter-year" dataset, respectively. There is no longer a crash for the 'BST' pavement type, so "NA" (not available) appears in the corresponding bias magnitude places. By comparing the results from the original dataset and the two hypothetical datasets, we find that the magnitude of the correction gets larger when the number of crashes decreases. This testifies that the number of crashes is the factor that influences the magnitude of bias rather than the number of observations.

It is seen that the balance of event counts in one stratum compared to the reference stratum also matters to the magnitude of bias correction. For example, the magnitude for the coefficient of five lanes is smaller than that for three lanes, even though the number of crashes for five-lane road segments is rarer. The reason is that the number of crashes happening on five-lane roads is more comparable to the number of events occurring within the reference level. It is worth pointing out that how the bias correction will change the significance of certain covariate is case-dependent. The significance can be directly related to the confidence interval, which is automatically adjusted based on the bias-

Table 2
Descriptive statistics of pavement data.

Categorical variable	Levels	Frequency	Percentage
Route type	Interstate	2236	42.7%
	State route	1160	22.1%
	United States route	1842	35.2%
Entrance/Exit	No	5163	98.6%
	Yes	75	1.4%
Intersection	No	4906	93.7%
	Yes	332	6.3%
Ramp	No	4759	90.9%
	Yes	479	9.1%
Wye connection	No	5211	99.5%
	Yes	27	0.5%
Divided highway	No	1883	35.9%
	Yes	3355	64.1%
Rural/Urban	Rural	3467	66.2%
	Urban	1771	33.8%
No. of lanes	1	1752	33.4%
	2	2481	47.4%
	3	624	11.9%
	4	38	7.3%
	5	1	0.0%
Pavement type	Asphalt Concrete (ACP)	3846	73.4%
	Portland Cement Concrete (PCCP)	1229	23.5%
	Bituminous Surface (BST)	119	2.3%
	ACP/PCCP[a]	44	0.8%

Continuous variable	Range	Mean	Std. deviation
Friction	5.0–85.2	51.8	13.4
Gradient	0.0–6.4	1.4	1.4
Horizontal curvature	0.0–10.5	0.9	1.1

[a] Half of the road segment is ACP and the other half is PCCP.

Table 3
Bias magnitude for the explanatory variables of pavement data.

Variables	original dataset		"half-year" dataset		"quarter-year" dataset	
	No. of crashes	$\tilde{\beta} - \hat{\beta}$ ($\times 10^{-3}$) (pct change)	No. of crashes	$\tilde{\beta} - \hat{\beta}$ ($\times 10^{-3}$) (pct change)	No. of crashes	$\tilde{\beta} - \hat{\beta}$ ($\times 10^{-3}$) (pct change)
Route type						
I	29543		4851		2377	
SR	1053	0.2 (0.0%)	134	8.0 (-0.8%)	51	39.1 (-2.1%)
US RTE	978	0.2 (-0.1%)	110	9.0 (-1.0%)	35	43.4 (-2.5%)
Entrance/Exit						
0	31047		5006		2423	
1	527	0.0 (0.0%)	89	0.6 (0.1%)	40	4.0 (1.3%)
Intersection						
0	30836		4986		2410	
1	738	0.0 (0.0%)	109	0.8 (0.1%)	53	−3.0 (-0.3%)
Ramp						
0	22641		3610		1738	
1	8933	0.0 (0.0%)	1485	0.0 (0.0%)	725	0.1 (0.1%)
Wye Connection						
0	31381		5063		2447	
1	193	0.0 (0.0%)	32	−1.2 (-0.1%)	16	−3.0 (-0.2%)
Divided highway						
0	1384		174		69	
1	30190	0.0 (0.0%)	4921	3.6 (-0.2%)	2394	19.7 (-0.9%)
Rural/Urban						
Rural	2422		291		101	
Urban	29152	0.0 (0.0%)	4804	−1.4 (-0.1%)	2362	−4.7 (-0.3%)
No. of lanes						
1	910		100		33	
2	6270	0.0 (0.0%)	943	−1.6 (-0.5%)	433	−6.7 (-1.2%)
3	12904	0.0 (0.0%)	2132	−1.8 (-0.3%)	1038	−7.4 (-0.8%)
4	11387	0.0 (0.0%)	1903	−1.8 (-0.5%)	950	−7.4 (-1.4%)
5	103	0.0 (0.0%)	17	0.0 (0.0%)	9	1.3 (0.1%)
Pavement type						
ACP	9911		1519		692	
ACP/PCCP	611	0.0 (0.0%)	101	0.7 (0.4%)	51	3.0 (1.2%)
BST	20	14.8 (-2.5%)	0	NA	0	NA
PCCP	21032	0.0 (0.0%)	3475	0.0 (-0.1%)	1720	−0.1 (-0.1%)
Friction		0.0 (0.0%)		0.0 (0.0%)		0.0 (0.0%)
Gradient		0.0 (0.0%)		0.0 (0.0%)		0.1 (0.1%)
Horizontal Curvature		0.0 (0.0%)		0.1 (0.2%)		0.2 (0.8%)

corrected point estimator.

To sum up, the number of events is a key statistic affecting the magnitude of bias rather than the number of observations collected. The balance of event counts within different strata also plays a significant role to the bias magnitude.

5. Discussion

This study introduces a finite sample bias correction for Poisson and NB regression models in traffic safety modeling where the number of crashes is small. We applied the bias approximation formula, provided by McCullagh and Nelder (1989), to correct the MLE of the regression coefficients, and conducted a Monte Carlo simulation to examine the properties of this bias-corrected MLE. The simulation results show that the bias-corrected MLEs not only have less bias but also smaller variance. The benefit is more pronounced when the number of events in one stratum of a categorical explanatory variable is between 5 and 50.

Based on McCullagh and Nelder's (1989) general bias approximation, we derived the explicit bias formulas for one special Negative Binomial regression scenario, namely a model with one single binary explanatory variable. Through the explicit bias formula, we demonstrated that the leading factors affecting the magnitude of bias are the number of events and the balance of event counts in different strata of a categorical explanatory variable. Generally, the coefficient estimation bias is higher for a smaller number of events and highly unbalanced event counts among strata.

We also demonstrated a bias correction procedure using a real safety application, which evaluated the crash risk factors for road infrastructure. The comparison between the original data and reduced data shows that the magnitude of bias is directly related to the number of events.

As crashes are rare events and the number of crashes in a stratum could be small, we suggest that a robust approach would be using bias-corrected estimation when evaluating the effects of crash risk factors. In practice, it is not possible to control the number of crashes in every stratum of a risk factor. Using bias correction would provide a more-accurate and robust estimation.

On a related topic, the bias correction for variance and significance level of estimated parameters is also of key interest. A common practice for variance estimation is to still use the original MLE (King and Zeng, 2001). There is limited literature on whether bias in variance estimation exists. The significant level is directly related to the upper and lower boundary of the confidence interval, which is affected by both the point estimate and variance of the estimator. How the bias correction would affect significance of the results is worth further investigation.

We would like to remark that the bias-correction procedure is not restricted to the Poisson and NB regression models. It can be applied to other regression models of which their parameters are obtained by MLE, and up to the third (partial) derivatives exist for their maximum likelihood functions. For future study, it will be interesting to investigate bias correction for random/mixed effect models and other count-frequency models, such as the random parameters model (Anastasopoulos and Mannering, 2009; Chen and Tarko, 2014; Mannering et al., 2016; Shaon et al., 2018).

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgements

This study is supported by the grant “Big Data Methodologies for Simplifying Traffic Safety Analyses” from the Safe-D University Transportation Center.

Appendix A

For the NB regression with only one binary exploratory variable, suppose there are n data points available for model estimation, $\{X_i, Y_i, E_i\}_{i=1}^{n}$, with X_i being either 0 or 1, and Y_i and E_i are the corresponding event count and exposure, respectively. Assume there are n_1 data points with $X_i = 1$ and n_0 data points with $X_i = 0$, and $n_0 + n_1 = n$.

Without the loss of generality, we arrange the data $\{X_i, Y_i, E_i\}_{i=1}^{n}$ with $X_i = 1$ instances before $X_i = 0$ instances. That is,

$$X_i = \begin{cases} 1 & \text{if } i = 1, \ldots, n_1, \\ 0 & \text{if } i = n_1 + 1, \ldots, n. \end{cases}$$

For the $X_i = 1$ group, denote μ_1 as the mean of event rate and V_1 as the variance for the sum of crash counts, $V_1 = \mathrm{Var}(\sum_i Y_i | X_i = 1)$.

$$V_1 = \mathrm{Var}\left(\sum_i Y_i | X_i = 1\right) = \sum_{i=1}^{n_1} \left[\mu_1 E_i + \frac{(\mu_1 E_i)^2}{k}\right].$$

Similarly, denote the mean of event rate as μ_0 and V_0 as the variance for the sum of crash counts, $V_0 = \mathrm{Var}(\sum_i Y_i | X_i = 0)$, for the $X_i = 0$ group.

$$V_0 = \mathrm{Var}\left(\sum_i Y_i | X_i = 0\right) = \sum_{i=n_1+1}^{n} \left[\mu_0 E_i + \frac{(\mu_0 E_i)^2}{k}\right].$$

The data matrix $\mathbf{X}$ and covariance matrix $\mathbf{W}$ in Eq. (4) are

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ \hdashline 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\\vdots\\1\end{matrix}}\right\} n_1 \text{ rows} \\ \left.\vphantom{\begin{matrix}1\\\vdots\\1\end{matrix}}\right\} n_0 \text{ rows} \end{matrix},$$

$$\mathbf{W} = \left[\begin{array}{ccc:ccc} \mu_1 E_1 + \frac{(\mu_1 E_1)^2}{k} & & & & & \\ & \ddots & & & & \\ & & \mu_1 E_{n_1} + \frac{(\mu_1 E_{n_1})^2}{k} & & & \\ \hdashline & & & \mu_0 E_{n_1+1} + \frac{(\mu_0 E_{n_1+1})^2}{k} & & \\ & & & & \ddots & \\ & & & & & \mu_0 E_n + \frac{(\mu_0 E_n)^2}{k} \end{array}\right] \begin{matrix} \left.\vphantom{\begin{matrix}1\\\vdots\\1\end{matrix}}\right\} n_1 \text{ rows} \\ \left.\vphantom{\begin{matrix}1\\\vdots\\1\end{matrix}}\right\} n_0 \text{ rows} \end{matrix}$$

Denote

$$V_1 = \mathrm{Var}\left(\sum_i Y_i | X_i = 1\right) = \sum_{i=1}^{n_1} \left[\mu_1 E_i + \frac{(\mu_1 E_i)^2}{k}\right],$$

$$V_0 = \mathrm{Var}\left(\sum_i Y_i | X_i = 0\right) = \sum_{i=n_1+1}^{n} \left[\mu_0 E_i + \frac{(\mu_0 E_i)^2}{k}\right].$$

That is, V_1 is the sum of the upper half of the diagonal elements of the matrix $\mathbf{W}$, and V_0 is the sum of the lower half of the diagonal elements.

Then

$$\mathbf{X}'\mathbf{W}\mathbf{X} = \begin{bmatrix} [c]{cc} V_0 + V_1 & V_1 \\ V_1 & V_1 \end{bmatrix}, \quad (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} = \frac{1}{V_0 V_1} \begin{bmatrix} [c]{cc} V_1 & -V_1 \\ -V_1 & V_1 + V_0 \end{bmatrix},$$

and

$$\mathbf{Q} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}' = \frac{1}{V_0 V_1}\left[\begin{array}{ccc:ccc} V_0 & \cdots & V_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ V_0 & \cdots & V_0 & 0 & \cdots & 0 \\ \hdashline 0 & \cdots & 0 & V_1 & \cdots & V_1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & V_1 & \cdots & V_1 \end{array}\right].$$

$$\Rightarrow \quad \xi_i = -\frac{1}{2}Q_{ii} = \begin{cases} -\dfrac{1}{2V_0} & \text{if } i = 1, \ldots, n_1; \\ -\dfrac{1}{2V_1} & \text{if } i = n_1 + 1, \ldots, n. \end{cases}$$

Therefore, the bias of the MLE $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)$ is

$$\text{bias}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\boldsymbol{\xi} = \begin{bmatrix} -\dfrac{1}{2V_0} \\ \dfrac{1}{2V_0} - \dfrac{1}{2V_1} \end{bmatrix},$$

that is,

$$\text{bias}(\hat{\beta}_1) = \frac{1}{2V_0} - \frac{1}{2V_1}.$$

References

AASHTO, 2010. Highway Safety Manual, vol. 1 American Association of State Highway and Transportation Officials, Washington DC.

Aguero-Valverde, J., 2013. Full bayes poisson gamma, poisson lognormal, and zero inflated random effects models: comparing the precision of crash frequency estimates. Accid. Anal. Prev. 50, 289–297.

Anastasopoulos, P.C., 2016. Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. Anal. Methods Accid. Res. 11, 17–32.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accid. Anal. Prev. 41 (1), 153–159.

Bartlett, M., 1953. Approximate confidence intervals. Biometrika 40 (1/2), 12–19.

Carson, J., Mannering, F., 2001. The effect of ice warning signs on ice-accident frequencies and severities. Accid. Anal. Prev. 33 (1), 99–109.

Chen, E., Tarko, A.P., 2014. Modeling safety of highway work zones with random parameters and random effects models. Anal. Methods Accid. Res. 1, 86–95.

Cordeiro, G.M., McCullagh, P., 1991. Bias correction in generalized linear models. J. R. Stat. Soc. Ser. B (Methodol.) 629–643.

Cox, D.R., Snell, E.J., 1968. A general definition of residuals. J. R. Stat. Soc. Ser. B-Stat. Methodol. 30 (2), 248–275.

Firth, D., 1993. Bias reduction of maximum likelihood estimates. Biometrika 80 (1), 27–38.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., Thomsen, L.K., 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. Accid. Anal. Prev. 27 (1), 1–20.

Giles, D.E., Feng, H., 2011. Reducing the bias of the maximum likelihood estimator for the poisson regression model. Econ. Bull. 31 (4), 2933–2943.

Guo, F., Fang, Y., 2013. Individual driver risk assessment using naturalistic driving data. Accid. Anal. Prev. 61, 3–9.

Guo, F., Wang, X., Abdel-Aty, M.A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accid. Anal. Prev. 42 (1), 84–92.

King, G., Zeng, L., 2001. Logistic regression in rare events data. Polit. Anal. 9 (2), 137–163.

Kosmidis, I., Firth, D., 2009. Bias reduction in exponential family nonlinear models. Biometrika 96 (4), 793–804.

Kosmidis, I., Firth, D., et al., 2010. A generic algorithm for reducing bias in parametric estimation. Electron. J. Stat. 4, 1097–1112.

Kumara, S., Chin, H.C., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. Traffic Injury Prev. 4 (1), 53–57.

Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 34 (1), 1–14.

Lambrecht, B., Perraudin, W., Satchell, S., 1997. Approximating the finite sample bias for maximum likelihood estimators using the score-solution. Econom. Theory 13 (2), 310–312.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accid. Anal. Prev. 34 (2), 149–161.

Li, X., Lord, D., Zhang, Y., 2010. Development of accident modification factors for rural frontage road segments in texas using generalized additive models. J. Transp. Eng. 137 (1), 74–83.

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. Accid. Anal. Prev. 38 (4), 751–766.

Lord, D., Bonneson, J., 2005. Calibration of predictive models for estimating safety of ramp design configurations. Transp. Res. Rec.: J. Transp. Res. Board (1908) 88–95.

Lord, D., Geedipally, S.R., 2011. The negative binomial-lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros. Accid. Anal. Prev. 43 (5), 1738–1742.

Lord, D., Geedipally, S.R., 2018. Safety prediction with datasets characterised with excess zero responses and long tails. Safe Mobility: Challenges Methodology and Solutions. Emerald Publishing Limited, pp. 297–323.

Lord, D., Manar, A., Vizioli, A., 2005a. Modeling crash-flow-density and crash-flow-v/c ratio relationships for rural and urban freeway segments. Accid. Anal. Prev. 37 (1), 185–199.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A: Policy Pract. 44 (5), 291–305.

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. Saf. Sci. 46 (5), 751–770.

Lord, D., Persaud, B., 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transp. Res. Rec.: J. Transp. Res. Board (1717), 102–108.

Lord, D., Washington, S., Ivan, J.N., 2007. Further notes on the application of zero-inflated models in highway safety. Accid. Anal. Prev. 39 (1), 53–57.

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

Malyshkina, N.V., Mannering, F.L., 2010. Zero-state markov switching count-data models: an empirical assessment. Accid. Anal. Prev. 42 (1), 122–130.

Malyshkina, N.V., Mannering, F.L., Tarko, A.P., 2009. Markov switching negative binomial models: an application to vehicle accident frequencies. Accid. Anal. Prev. 41 (2),

217–226.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.
McCullagh, P., 1987. Tensor Methods in Statistics, vol. 161 Chapman and Hall, London.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, vol. 37 CRC Press.
Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 (4), 471–482.
Park, B.-J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accid. Anal. Prev. 41 (4), 683–691.
Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accid. Anal. Prev. 36 (2), 183–191.
Saha, K., Paul, S., 2005. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. Biometrics 61 (1), 179–185.
Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accid. Anal. Prev. 29 (6), 829–837.
Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., Nebergall, M.B., 2003. Modeling crashes involving pedestrians and motorized traffic. Saf. Sci. 41 (7), 627–640.
Shaon, M.R.R., Qin, X., Shirazi, M., Lord, D., Geedipally, S.R., 2018. Developing a random parameters negative binomial-lindley model to analyze highly over-dispersed crash count data. Anal. Methods Accid. Res. 18, 33–44.
Songpatanasilp, P., Yamada, H., Horanont, T., Shibasaki, R., 2015. Traffic accidents risk analysis based on road and land use factors using glms and zero-inflated models. Proceedings of 14th International Conference on Computers in Urban Planning and Urban Management (CUPUM 2015) 7–10.
Vangala, P., Lord, D., Geedipally, S.R., 2015. Exploring the application of the negative binomial-generalized exponential model for analyzing traffic crash data with excess zeros. Anal. Methods Accid. Res. 7, 29–36.
Wood, G., 2002. Generalised linear accident models and goodness of fit testing. Accid. Anal. Prev. 34 (4), 417–427.
Xie, Y., Zhang, Y., 2008. Crash frequency analysis with generalized additive models. Transp. Res. Rec.: J. Transp. Res. Board (2061), 39–45.
Zou, Y., Lord, D., Zhang, Y., Peng, Y., 2013. Comparison of sichel and negative binomial models in estimating empirical bayes estimates. Transp. Res. Rec.: J. Transp. Res. Board (2392), 11–21.
Zou, Y., Wu, L., Lord, D., 2015. Modeling over-dispersed crash data with a long tail: examining the accuracy of the dispersion parameter in negative binomial models. Anal. Methods Accid. Res. 5, 1–16.