Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models

Karim El-Basyouny and Tarek Sayed

There are several regression techniques to develop accident prediction models. Model development and subsequently the results are affected by the choice of regression technique. The objective of this paper is to compare two types of regression techniques: the traditional negative binomial (TNB) and the modified negative binomial (MNB). The TNB approach assumes that the shape parameter of the negative binomial distribution is fixed for all locations, while the MNB approach assumes that this shape parameter varies with the location's characteristics. The difference between the two approaches in terms of their goodness of fit and the identification and ranking of accident-prone locations is investigated. The study makes use of a sample of accident, volume, and geometric data corresponding to 392 arterial segments in British Columbia, Canada. Both models appear to fit the data well. However, the MNB approach provides a statistically significant improvement in model fit over the TNB approach. A total of 100 locations were identified as accident-prone by both approaches. A comparison between the ranks showed a close agreement in the general trend of ranking between the two models. While the MNB approach appears to fit the data better than the TNB approach, there was little difference in the results of the identification and ranking of accident-prone locations. This is likely due to the nature of the application and the data set used. The difference in results will depend on the extent to which deviant sites exist in the data set.

The success of safety improvement programs in reducing accident occurrence depends on the availability of methods that give reliable estimates of the safety level associated with existing road locations or proposed plans and designs. Several approaches exist for estimating safety. They range from simply using accident rates to accident prediction models that relate the expected accident frequency at a road location to its traffic and geometric characteristics. Several researchers (*1*) have shown that the relationship between accident frequency and exposure is frequently nonlinear, which indicates that accident rates are not appropriate representatives of safety. This finding has led most safety researchers to discard the use of accident rate as a measure of road safety, and accident prediction models are currently the primary tools for estimating road safety.

There are several regression techniques to develop accident prediction models. Model development and subsequently the results are strongly affected by the choice of regression technique. The earlier models were developed by using ordinary or normal linear regression. These models assume a normal error structure for the response variable, a constant variance for the residuals, and the existence of a linear relationship between the response and explanatory variables. Such models were criticized by several researchers (*2–4*) indicating that traffic accidents are discrete, nonnegative, and rare events that could not be adequately modeled with conventional linear regression. Such limitations led to the use of a Poisson or a negative binomial error structure in modeling the occurrence of traffic accidents. The main advantage of the Poisson error structure is the simplicity of the calculations. However, this advantage is also a limitation. It has been shown (*5, 6*) that most accident data are likely to be overdispersed (the variance is greater than the mean), which indicates that the negative binomial distribution is usually the more realistic assumption. In most models, the predicted accident frequency varies as a function of traffic volumes, geometry, and traffic features. The shape parameter of the negative binomial distribution is assumed invariant between locations and is taken as some scalar. Recently, a modification was proposed to this approach under which the shape parameter was allowed to vary from one location to another. This is supposed to lead to better model fit.

The objective of this paper is to compare two types of regression techniques: the traditional negative binomial (TNB) and the modified negative binomial (MNB). The TNB approach assumes that the shape parameter of the negative binomial distribution is fixed for all locations, while the MNB approach assumes that this shape parameter will vary from one location to another. The difference between the two approaches is investigated in terms of their goodness of fit as well as the identification and ranking of accident-prone locations. The study makes use of a sample of accident, traffic volume, and geometric data corresponding to 58 arterials (392 segments) in the cities of Vancouver and Richmond, British Columbia, Canada.

PREVIOUS WORK

TNB Regression

The development of models to predict traffic accidents has been the subject of numerous studies. As mentioned earlier, most accident prediction models are currently developed by using negative binomial regression. The models' development, structure, outlier analysis, and goodness of fit have been discussed by several researchers (*6–10*). The approach has the following theoretical basis. Let Y be the random variable that represents the accident frequency at a given location during a specific time period, and let y be a certain realization of Y.

Department of Civil Engineering, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

Transportation Research Record: Journal of the Transportation Research Board, No. 1950, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 9–16.

The mean of Y, denoted by Λ, is itself a random variable (*3, 6*). For $\Lambda = \lambda$, Y is Poisson distributed with parameter λ:

$$P(Y = y|\Lambda = \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad E(Y|\Lambda = \lambda) = \lambda \qquad \text{Var}(Y|\Lambda = \lambda) = \lambda \tag{1}$$

It is the usual practice to assume that the distribution of Λ can be described by a gamma probability density function (*7*). If Λ is described by a gamma distribution with shape parameter κ and scale parameter κ/μ, its density function is

$$f_\Lambda(\lambda) = \frac{(\kappa/\mu)^\kappa \lambda^{\kappa-1} e^{-(\kappa/\mu)\lambda}}{\Gamma(\kappa)} \qquad E(\Lambda) = \mu \qquad \text{Var}(\Lambda) = \frac{\mu^2}{\kappa} \tag{2}$$

and the distribution of Y around $E(\Lambda) = \mu$ is negative binomial (*7, 11*) with an expected value and variance given by Equations 3*a* and 3*b*, respectively:

$$E(Y) = \mu \tag{3a}$$

$$\text{Var}(Y) = \mu + \frac{\mu^2}{\kappa} \tag{3b}$$

There are several approaches to estimate the shape parameter κ of the negative binomial distribution (*12*), with the method of maximum likelihood being the most widely used. The mathematical form used for the accident prediction model should generally satisfy two conditions (*9*). First, it must yield logical results, meaning that it should not lead to the prediction of a negative number of accidents and that it should ensure a prediction of zero accident frequency for zero values of the exposure variables. Second, there must exist a known link function that can linearize this form for the purpose of coefficient estimation. The following is an example of a model for road segments that satisfies these conditions (*9*):

$$E(\Lambda) = a_0 \times L^{a_1} \times V^{a_2} \times e^{\sum_{j=1}^{m} b_j x_j} \tag{4}$$

where

$E(\Lambda)$ = predicted accident frequency;
L = segment length;
V = segment annual average daily traffic (AADT);
x_j = any of m variables additional to L and V; and
a_0, a_1, a_2, b_j = model parameters.

MNB Regression

In the TNB approach described above, the variance of the predicted accident frequency is taken as

$$\text{Var}(\Lambda) = \frac{E(\Lambda)^2}{\kappa} \tag{5}$$

where κ is assumed a fixed value for all locations. Heydecker and Wu (*13*) proposed an extension to this approach whereby Var(Λ) is allowed to vary with the location characteristics. For example, assuming a mathematical form similar to Equation 4:

$$\text{Var}(\Lambda) = c_0 \times L^{c_1} \times V^{c_2} \times e^{\sum_{j=1}^{m} d_j x_j} \tag{6}$$

where c_0, c_1, c_2, and d_j are the model parameters. Equations 5 and 6 will lead the shape parameter to vary according to the site characteristics. Similar parameterizations that allow observed variability in the shape parameter were also proposed to structure unmeasured heterogeneities (*14, 15*). Such models for the dispersion can increase the flexibility and the precision of accident prediction.

Heydecker and Wu (*13*) used a data set comprising 389 three-leg intersection sites of two-lane, two-way roads, with a stop control on the minor road to develop accident prediction models by using this MNB approach. Of the 389 sites, only 187 had accident records that were nonzero. After some outliers were removed, accident prediction models were developed by using the TNB and MNB approaches. Heydecker and Wu showed that the MNB approach represents a better fit to the data than the TNB approach.

Miaou and Lord (*14*) used a data set of urban four-legged signalized intersections in Toronto, Canada, for the period 1990–1995. They challenged the underlying assumptions adopted in the popular accident prediction models for intersections. In particular, they questioned the assumption that the dispersion parameter is fixed across sites and time periods, and they examined the mathematical limitations of some functional forms used in these models, specifically their properties at the boundaries. They concluded that fixing the dispersion parameter can undermine the goodness of estimates of individual sites by up to about 35%. Furthermore, the authors recommended extending the analysis to other data sets to determine whether varying dispersion parameter value is a common problem, an isolated problem, or specific to certain types of locations.

Miranda-Moreno et al. (*15*) used a data set consisting of 5,094 highway–rail grade crossings with flashing lights as main warning devices. The accidents were collected for a 5-year period (1996–2000). The authors computed six accident estimates, three of which were the conditional means of accident frequency based on the marginal distributions of the TNB approach, the MNB approach, and Poisson-lognormal models, whereas the other three were the posterior means of accident frequency under these models. They found that the MNB approach and Poisson-lognormal fit the data better than the TNB regression model. The models were used to identify and rank accident-prone locations. The Spearman correlation coefficient was used to assess the degree of association between rankings. The results indicated that the ranking based on the posterior means of the TNB approach agreed more with the corresponding ranking of the MNB and less with that of the Poisson-lognormal.

DATA DESCRIPTION

The analysis undertaken in this paper is demonstrated with the same data set described by Sawalha and Sayed (*9*). A total of 58 arterials in the cities of Vancouver and Richmond were investigated to develop accident prediction models relating the safety of urban arterial sections to their traffic and geometric characteristics. Geometric data were collected directly from the field. The approach to geometric data collection consisted of dividing each arterial into segments between consecutive signalized intersections (which resulted in 392 segments) and gathering field information for each segment separately. The data on accident frequencies and traffic volumes along the arterials were obtained from the cities of Vancouver and Richmond and covered the period from 1994 to 1996. Accidents that occurred at signalized intersections were excluded from the accident data used to develop the models. In addition to the traffic accidents and volumes, the following geometric data were collected: section length and number of unsignalized intersections. An outlier analysis was conducted on the basis of the methodology described by Sawalha and Sayed (*9*) by using Cook's distance measure. This measure takes both leverage and influence into account during the identification of outliers. As a result, six segments were identified as outliers and were subsequently removed from further analysis. A statistical summary of the remaining 386 segments is given below.

	Segment Length (km)	*AADT*	*Unsignalized Intersection Density*	*Total Accidents*
Minimum	0.113	4,232	0	1
Maximum	3.608	62,931	20.48	264
Mean	0.820	24,297	5.87	49.35
Standard deviation	0.400	13,092	3.33	45.84

MODEL DEVELOPMENT

The following model form is adopted, the merits of which are discussed in detail by Sawalha and Sayed (*9*):

$$E(\Lambda) = a_0 L^{a_1} V^{a_2} \exp(b_1 \text{USD}) \tag{7}$$

where

$E(\Lambda)$ = predicted accident frequency per 3 years;
L = segment length (km);
V = segment AADT;
USD = unsignalized intersection density; and
a_0, a_1, a_2, b_1 = model parameters.

In addition, for the MNB approach, the variance was assumed to have the following model form:

$$\text{Var}(\Lambda) = c_0 L^{c_1} V^{c_2} \exp(d_1 \text{USD}) \tag{8}$$

where c_0, c_1, c_2, and d_1 are the model parameters. The TNB model can be considered a restricted version of the MNB model where the MNB model has additional parameters.

Several measures can be used to assess the goodness of fit of the models. Two commonly used measures are the scaled deviance (SD) and the Pearson χ^2 statistic. The SD is defined as the likelihood ratio test statistic measuring twice the difference between the log-likelihoods of the studied model and the full or saturated model. The full model has as many parameters as there are observations (each intersection having its own accident parameter) so that the model fits the data perfectly. Therefore, the full model, which possesses the maximum log-likelihood achievable under the given data, provides a baseline for assessing the goodness of fit of an intermediate model with parameters. McCullagh and Nelder (*16*) have shown that for negative binomial error structure the SD is as follows:

$$\text{SD} = 2\sum_{i=1}^{n}\left\{y_i \ln\left[\frac{y_i}{E(\Lambda_i)}\right] - (y_i + \kappa)\ln\left[\frac{y_i + \kappa}{E(\Lambda_i) + \kappa}\right]\right\} \tag{9}$$

In the case of the MNB approach, κ_i will vary for each location, and the SD is calculated as

$$\text{SD} = 2\sum_{i=1}^{n}\left\{y_i \ln\left[\frac{y_i}{E(\Lambda_i)}\right] - (y_i + \kappa_i)\ln\left[\frac{y_i + \kappa_i}{E(\Lambda_i) + \kappa_i}\right]\right\} \tag{10}$$

The Pearson χ^2 is defined as

$$\text{Pearson } \chi^2 = \sum_{i=1}^{n}\frac{[y_i - E(\Lambda_i)]^2}{\text{Var}(Y_i)} \tag{11}$$

where $\text{Var}(Y_i)$ is estimated from Equation 3*b* for both approaches with κ_i varying for each location in the MNB approach.

Both the SD and the Pearson χ^2 have exact χ^2 distributions for normal theory linear models but are asymptotically χ^2 distributed with $n - p$ degrees of freedom for other distributions of the exponential family (*17*).

RESULTS

The statistical software SAS Version 9.1 (*18*) was used to estimate the models' parameters. The parameters a_0, a_1, a_2, b_1, and κ of the TNB model were estimated by using the procedure GENMOD, while the parameters a_0, a_1, a_2, and b_1 and c_0, c_1, c_2, and d_1 of the MNB model were estimated by using the procedure IML along with the nonlinear optimization subroutine NLPTR.

Tables 1 and 2 summarize the model parameter estimates and their associated statistics under the traditional and the modified negative binomial techniques, respectively. An examination of the tables indicates that the model parameter estimates are significant at the 95% confidence level. The coefficients for the segment length, AADT, and USD are all positive, which indicates an increase in the mean accident frequency with each of them. Table 1 shows a significant *t*-ratio of 13.16 for the shape parameter κ of the TNB model, which indicates the presence of overdispersion in the data. Table 2 indicates

TABLE 1 Estimates of Model Parameters Using TNB Regression

	Variable	Estimate	Std. Error	*t*-ratio
Mean	Constant	–5.64	0.54	–10.39
	Length	0.78	0.065	11.91
	AADT	0.90	0.054	16.47
	UNSD	0.11	0.010	10.35
Variance	κ	2.82	0.215	13.16

TABLE 2 Estimates of Model Parameters Using MNB Regression

	Variable	Estimate	Std. Error	*t*-ratio
Mean	Constant	–5.64	0.5300	–10.64
	Length	0.81	0.066	12.44
	AADT	0.90	0.054	16.83
	UNSD	0.10	0.011	9.81
Variance	Constant	–13.79	1.81	–7.63
	Length	1.13	0.21	5.56
	AADT	1.96	0.18	10.78
	UNSD	0.16	0.036	4.41

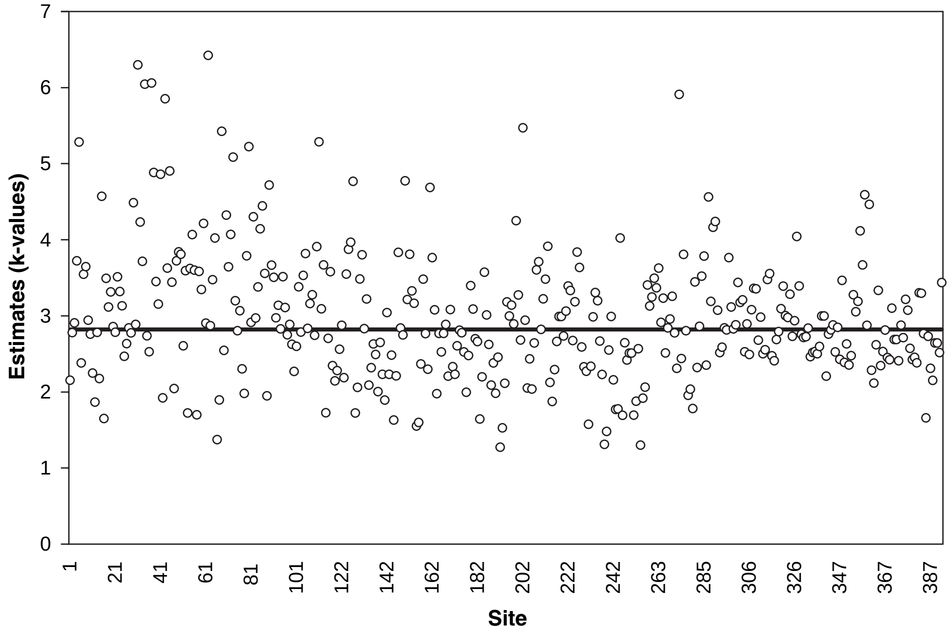


FIGURE 1 Variation in shape parameter by site.

highly significant *t*-ratios for the parameters of the variance function, which signifies the relative advantage of using an invariant shape parameter to account for unobserved heterogeneity.

Figure 1 shows the variation in the estimates of the shape parameter for the MNB model. While the constant shape parameter κ of the TNB approach is estimated at 2.82, the corresponding estimates of the MNB approach range from 1.27 to 6.42, with an average of 2.99. The effect of allowing such variation in the shape parameter is strongly reflected in the variance function. To visualize such effects, Figure 2 shows the variance functions of the two models. The variance function of the TNB approach is estimated from Equation 5, whereas that of the MNB approach is estimated from Equation 8. A second-degree polynomial was fit to the MNB variance for comparison. The figure shows that the variance function of the MNB approach increases at a slower rate than that of the TNB approach as the number of predicted accidents increases. This is due to the MNB approach having more parameters and consequently lower variability. Since the variance of the predicted number of accidents is proportional to the square of the predicted number of accidents, the difference in variability between the two approaches will be more apparent for locations with high predicted accidents (higher means in Figure 2).

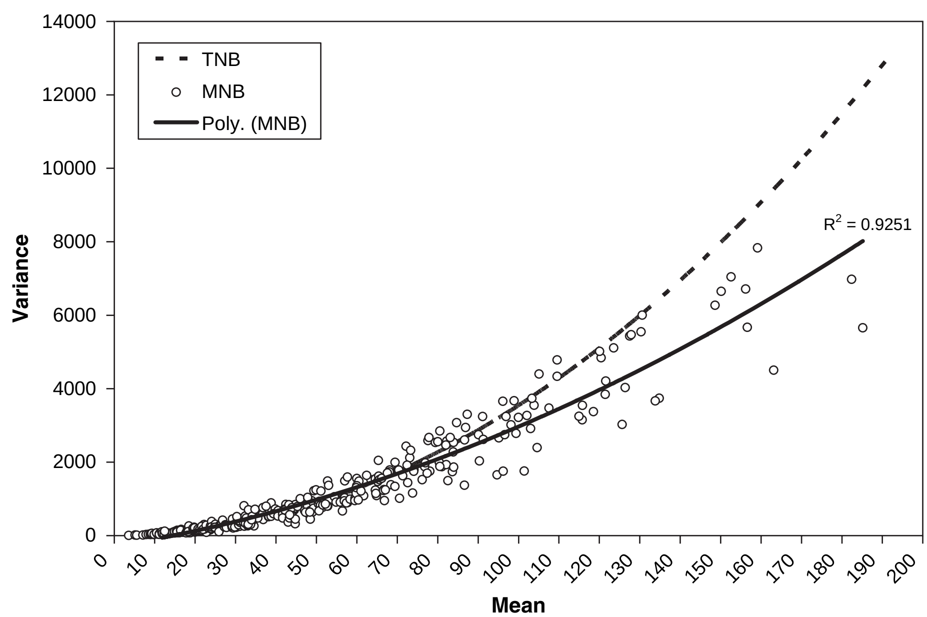


FIGURE 2 Variance versus mean for TNB and MNB models.

The table below shows the SD and Pearson χ^2 statistic of the two models.

Model	*Degrees of Freedom*	*SD*	*Pearson* χ^2	$\chi^2_{.05}$
TNB	381	409.19	392.31	427.51
MNB	378	406.51	397.41	424.33

Both models fit the data relatively well. The MNB model shows a statistically significant improvement in the model fit as indicated by the significant increase in the joint log-likelihood from –1,728.49 for the TNB approach to –1,723.07 for the MNB approach (see the table below).

Model	*Log Likelihood*	*Likelihood Ratio Test*	*Degrees of Freedom*	p	*Significant at 5%?*
TNB	–1,728.49				
MNB	–1,723.07	10.83	3	.013	Yes

To test the (nested) TNB versus MNB approach, the likelihood ratio test was 10.83 with three degrees of freedom, corresponding to a *p*-value of 0.013, which indicates a statistically significant improvement in the goodness of fit when the MNB model was used.

APPLICATIONS OF THE TWO MODELS

The TNB and MNB models will be compared in terms of their applications to the identification and ranking of accident-prone locations. An accident-prone location is defined as any location that exhibits a significantly higher number of accidents compared with a specific normal value. The empirical Bayes (EB) technique improves the location-specific prediction and thus is used to identify accident-prone locations. Two clues are available to determine the safety performance of a location: its traffic and road characteristics and its historical accident frequency (*19*). The EB approach makes use of both of these clues. It refines the estimate of the expected number of accidents at a location by combining the observed number of accidents with the predicted number of accidents obtained from the prediction model. This yields a more accurate, location-specific safety estimate. The EB estimate and its variance can be calculated as follows (*19*):

$$\mathrm{EB} = w \times E(\Lambda) + (1-w) \times \mathrm{count} \qquad (12)$$

$$\mathrm{Var}(\mathrm{EB}) = w(1-w) \times E(\Lambda) + (1-w)^2 \times \mathrm{count} \qquad (13)$$

$$w = \frac{1}{1 + \dfrac{\mathrm{Var}[E(\Lambda)]}{E(\Lambda)}} \qquad (14)$$

where count is the observed number of accidents. The EB method identifies hazardous locations according to the following process:

1. Estimate the predicted number of accidents and its variance for the location. This follows a gamma distribution (the prior distribution) with parameters α and β, where

$$\beta = \frac{E(\Lambda)}{\mathrm{Var}(\Lambda)} \qquad \alpha = \beta \cdot E(\Lambda) = \kappa \qquad (15)$$

2. Calculate the EB safety estimate and the variance from Equations 12 and 13. This is a gamma distribution (the posterior distribution) with parameters α_1 and β_1 defined as follows:

$$\beta_1 = \frac{\mathrm{EB}}{\mathrm{Var}(\mathrm{EB})} \qquad \alpha_1 = \beta_1 \cdot \mathrm{EB} \qquad (16)$$

Then the probability density function of the posterior distribution is given by the following:

$$f_{\mathrm{EB}}(\lambda) = \frac{\beta_1^{\alpha_1} \lambda^{\alpha_1 - 1} e^{-\beta_1 \lambda}}{\Gamma(\alpha_1)} \qquad (17)$$

3. Identify the location as accident-prone if there is a significant probability that the location's safety estimate exceeds $E(\Lambda)$. Thus, the location is accident-prone if

$$\left[1 - \int_0^{E(\Lambda)} \frac{\beta_1^{\alpha_1} \lambda^{\alpha_1 - 1} e^{-\beta_1 \lambda}}{\Gamma(\alpha_1)} d\lambda \right] \geq \delta \qquad (18)$$

where δ is the desired confidence level (usually selected at 0.95).

The ranking of accident problem sites will enable the road authority to establish an effective road safety program and ensure the efficient use of the limited funding available for road safety. Two techniques that reflect different priority objectives for a road authority can be used (*20*). The first ranking criterion, the relative risk, is the ratio between the EB estimate and the predicted accident frequency as obtained from the prediction model (a risk-minimization objective). The ratio represents the level of deviation that the location is away from a "normal" safety performance value, with a higher ratio representing a more hazardous location. The second criterion, the absolute incremental risk, is the difference between the EB estimate and the predicted frequency for each accident-prone location. The difference between these two values is an effective indicator of the potential for improvement (PFI) of each accident-prone location.

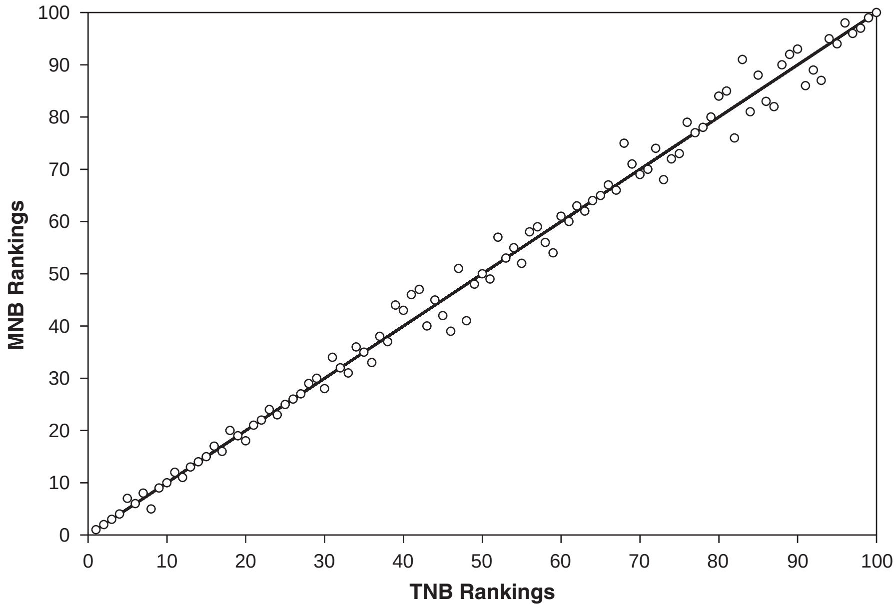
A total of 100 locations were identified as accident-prone by both the TNB and the MNB approaches. A comparison between the ranks obtained from the TNB model and those obtained from the MNB model was undertaken. The results are presented in Figures 3*a* and 3*b*, which indicate a close agreement in the general trend of ranking between the two models. Figure 4 shows that for the majority of accident-prone locations, the difference in ranking between the TNB and the MNB techniques is small (less than 3). In addition, the Spearman rank correlation coefficient was used to determine the level of agreement between the rankings as shown in Equation 19. A score of 1.0 represents perfect correlation, and a score of zero indicates no correlation.

$$\rho_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \qquad (19)$$

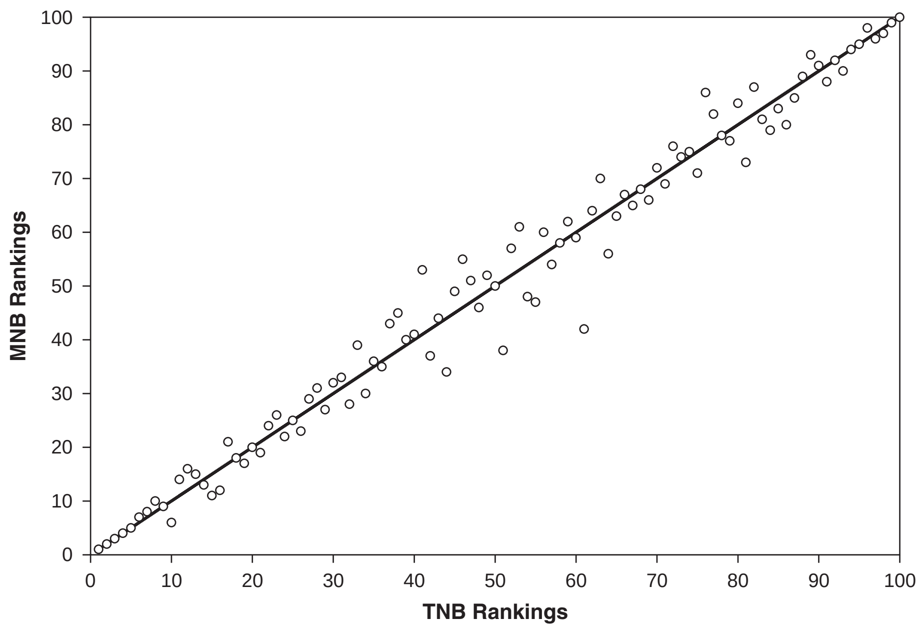
where d is the differences between ranks and n is the number of paired sets.

Under a null hypothesis of no correlation, the ordered data pairs are randomly matched, and thus the sampling distribution of ρ_s has a mean of zero and the standard deviation σ_s as given in Equation 20.

$$\sigma_s = \frac{1}{\sqrt{(n-1)}} \qquad (20)$$



(a)



(b)

FIGURE 3 MNB rankings versus TNB rankings (*a*) using PFI and (*b*) using ratio.

Since this sampling distribution can be approximated with a normal distribution, the null hypothesis can be tested by using the statistic given in Equation 21. This value can be compared with a critical *z*-value of 2.33 representing the 99% level of confidence.

$$z = \rho_s \sqrt{(n-1)} \tag{21}$$

The results from the correlation analysis indicate that with the PFI as the ranking criterion, the Spearman rank correlation coefficient (ρ_s) was estimated to be 0.995 and the 99% confidence interval was (0.992, 0.997), which indicates a strong agreement at the 99% confidence level. With the relative risk as the ranking criterion, the Spearman rank correlation coefficient was estimated to be 0.988 and the 99% confidence interval was (0.980, 0.993), which also indicates a strong agreement at the 99% confidence level.

SUMMARY AND CONCLUSIONS

This paper has presented the results of a comparison between two types of regression techniques: the TNB and the MNB. The TNB approach assumes that the shape parameter of the negative binomial

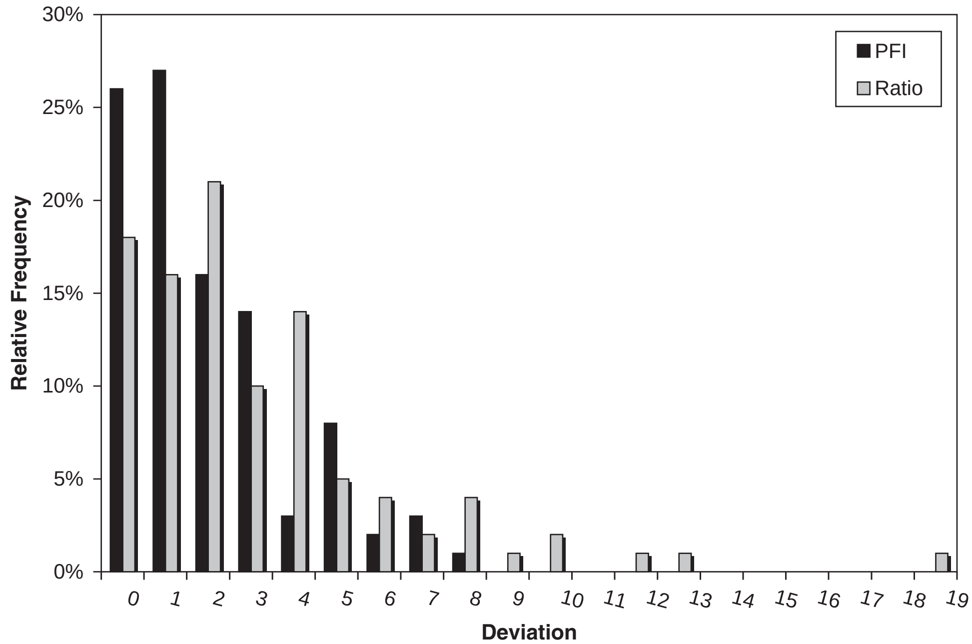


FIGURE 4 Ranking deviations: MNB versus TNB.

distribution is fixed for all locations, while the MNB approach assumes that this shape parameter varies from one location to another. The difference between the two approaches in terms of their goodness of fit as well as the identification and ranking of accident-prone locations was investigated. The research used data for 58 arterials (392 segments) located in the cities of Vancouver and Richmond, British Columbia. Examining the goodness of fit measures of the two models indicated that both models appear to fit the data well. The MNB model showed an improvement in the model fit as indicated by the increase in the joint log-likelihood. The likelihood ratio test indicated a significant improvement (at the 95% confidence level) in the goodness of fit when the MNB model was used. Furthermore, the estimates of the parameters of the variance function of the MNB model were highly significant, justifying the allowance of observed variability in the dispersion parameter.

The effect of allowing such variation in the shape parameter is strongly reflected in the variance function, which increased at a much slower rate than that of the TNB as the number of predicted accidents increased. This indicates the benefit of using the MNB approach for modeling locations with high accident frequencies.

A total of 100 locations were identified as accident-prone by both the TNB and the MNB approaches. Two ranking criteria were used to rank the accident-prone locations: the relative risk criterion, representing the risk-minimization objective, and the PFI, representing the cost-effectiveness objective. A comparison between the ranks indicated a close agreement in the general trend of ranking between the two models. The Spearman rank correlation coefficient was used to determine the agreement level between the rankings of the TNB and the MNB approaches. The Spearman correlation indicated a strong agreement at the 99% confidence level.

For the current data set, use of the MNB approach effectively increased the goodness of fit to the data by allowing for more modeling variability. However, in terms of model application (identification and ranking of accident-prone locations), small or limited differences were observed in the results. There are likely two reasons for this. The first is the nature of the application. The identification of accident-prone locations depends on the existence of deviant sites. The two approaches may give similar results depending on the extent to which deviant sites exist in the data set. The second is the amount of dispersion in the data set. The difference in the results may have been more pronounced if the accident data were more overdispersed. Data sets with higher dispersion are expected to benefit more from the application of the MNB approach.

REFERENCES

1. Hauer, E. On Exposure and Accident Rate. *Traffic Engineering and Control,* Vol. 36, No. 3, 1995, pp. 134–138.
2. Jovanis, P. P., and H.-L. Chang. Modeling the Relationship of Accidents to Miles Traveled. In *Transportation Research Record 1068,* TRB, National Research Council, Washington, D.C., 1986, pp. 42–51.
3. Hauer, E., J. C. N. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185,* TRB, National Research Council, Washington, D.C., 1988, pp. 48–61.
4. Miaou, S., and H. Lum. Modeling Vehicle Accident and Highway Geometric Design Relationships. *Accident Analysis and Prevention,* Vol. 25, No. 6, 1993, pp. 689–709.
5. Kulmala, R., and M. Roine. Accident Prediction Models for Two-Lane Roads in Finland. *Proc., Conference on Traffic Safety Theory and Research Methods,* Session 4: Statistical Analysis and Models, Amsterdam, Netherlands, 1988, pp. 89–103.
6. Kulmala, R. *Safety at Rural Three- and Four-Arm Junctions. Development of Accident Prediction Models.* VTT Publication 233. Technical Research Center of Finland, Espoo, 1995.
7. Hauer, E. *Observational Before–After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety.* Elsevier Science, Ltd., 1997.
8. Sayed, T., P. deLeur, and Z. Sawalha. Evaluating the Insurance Corporation of British Columbia Road-Safety Improvement Program. In *Transportation Research Record: Journal of the Transportation Research*

Board, No. 1865, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 57–63.
9. Sawalha, Z., and T. Sayed. Evaluating Safety of Urban Arterial Roadways. *Journal of Transportation Engineering,* Vol. 127, No. 2, April 2001, pp. 151–158.
10. Maher, M. J., and I. Summersgill. A Comprehensive Methodology for the Fitting of Predictive Accident Models. *Accident Analysis and Prevention,* Vol. 28, No. 3, 1996, pp. 281–296.
11. Hinde, J., and C. G. B. Demetrio. Over-Dispersion: Model and Estimation. *Computational Statistics and Data Analysis*, Vol. 27, No. 2, 1998, pp. 151–170.
12. Lawless, J. F. Negative Binomial and Poisson Regression. *Canadian Journal of Statistics*, Vol. 15, No. 3, 1987, pp. 209–225.
13. Heydecker, B. G., and J. Wu. Identification of Sites for Accident Remedial Work by Bayesian Statistical Methods: An Example of Uncertain Inference. *Advances in Engineering Software,* Vol. 32, 2001, pp. 859–869.
14. Miaou, S.-P., and D. Lord. Modeling Traffic Crash–Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1840,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 31–40.
15. Miranda-Moreno, L. F., L. Fu, F. F. Saccomanno, and A. Labbe. Alternative Risk Models for Ranking Locations for Safety Improvement. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1908,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 1–8.
16. McCullagh, P., and J. A. Nelder. *Generalized Linear Models.* Chapman and Hall, New York, 1989.
17. Aitkin, M., D. Anderson, B. Francis, and J. Hinde. *Statistical Modeling in GLIM.* Oxford University Press, New York, 1989.
18. *GENMOD, IML and NLPTR Functions, Version 9.1.* SAS Institute, Inc., Cary, N.C., 1987.
19. Hauer, E. Empirical Bayes Approach to the Estimation of Unsafety: The Multivariate Regression Method. *Accident Analysis and Prevention,* Vol. 24, No. 5, 1992, pp. 457–477.
20. Sayed, T., and F. Rodriguez. Accident Prediction Models for Urban Unsignalized Intersections in British Columbia. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1665*, TRB, National Research Council, Washington, D.C., 1999, pp. 93–99.

The Statistical Methodology and Statistical Computer Software in Transportation Research Committee sponsored publication of this paper.