Zero-inflated Models for Different Severity Types in Rural Two-lane Crashes

Subasish Das, PhD Candidate
University of Louisiana

Xiaoduan Sun, PhD & PE, Professor
University of Louisiana

Research Question

What are the effects of the key geometric features in crash severity? Are ZINB and ZIP models sufficient to get better prediction?

Abstract

This research aims to investigate the application of zero-inflated models for different severity types in rural two-lane highway crashes in Louisiana. These roadways carry one-third of the total vehicle miles traveled (VMT) and have experienced a considerably high percentage of fatal crashes in Louisiana. Crashes recorded from 2004 to 2011, of which 1,780 were fatal, and 36,569 resulted in injuries, were analyzed. It is found that there are a large number of highway segments which contain no crashes under the recorded years. To tackle this issue, zero-inflated models, zero-inflated Poisson (ZIP) models and zero-inflated negative binomial (ZINB) models, have been developed for crash frequencies of different severity types.

Background

The State of Louisiana controls 60,937 miles of public road serving nearly 105,000 vehicle miles a day, and consisting of 46,959 miles of rural roads and 13,941 miles of urban roads. Nearly 58,000 miles of undivided rural roadways are two-lane in nature. Each year, approximately 150,000 crashes occur, over 90,000 of which are on the state-maintained highway system. In 2013, 703 people were killed and 70,658 were injured in highway crashes in Louisiana. Zero-inflated models, zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB), have been developed in this study for crash frequencies of different severity types.

Data Description

Eight years (2004-2011) of Louisiana crash data was used in this study.

- The important roadway factors considered in the study include segment length, pavement type and width, shoulder type and width, and annual average daily traffic (AADT).
- There are a total of 7,779 rural two-lane roadway segments in each year's crash dataset. The key variables available in the current dataset, related to roadway geometrics, were considered here. Louisiana Department of Transportation and Development (LaDOTD) has been maintaining crash data which doesn't have details on other roadway geometrics like vertical and horizontal curve degree, deflection angle, and percentage of gradient.

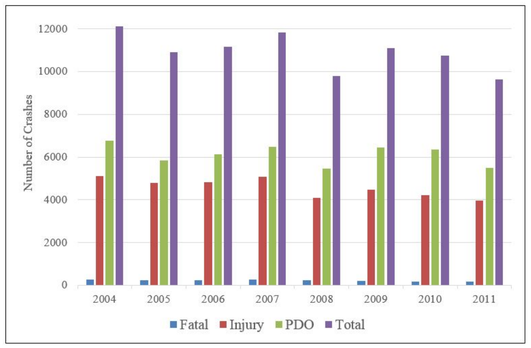

Figure 1: Crash frequencies of different severity types.

Zero Inflated Models

Crash-frequency data are non-negative integers, the application of the standard ordinary least-squares regression (which assumes a continuous dependent variable) is not appropriate.

The ZIP and ZINB regressions directly model the zeroes in the structural portion of the model. These models are generally considered as mixture models in which the complete distribution of the outcome is approximated by mixing two component distributions. The basic idea is to assume a logistic regression model for the **zero, and not zero** aspect of the consequence and either a Poisson or negative binomial distribution for the count portion in the model. ZIP and ZINB are well suited for the models in which there are two procedures and where the factors of the two procedures vary.

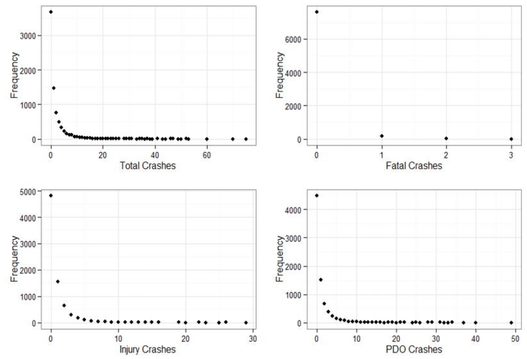

Figure 2: Crash frequency of different count of crashes per segment.

Major Findings

- Wider shoulder and pavement were found to reduce the likelihood of crash occurrence in rural two-lane highways.
- Gravel-top pavements are inclined to crash proneness according to both of the models.
- Lower values of AADT is significant for reducing the likelihood of crashes.
- When the test statistic value > 1.96 (the 95% confidence level for the t-test), the ZINB or ZIP model is more significant than traditional negative binomial or Poisson model.

Model Comparison

		ZIP Model			ZINB Model		
		Total Crashes	Injury Crashes	PDO Crashes	Total Crashes	Injury Crashes	PDO Crashes
Pearson Residuals	Min	-2.832	-2.278	-2.411	-1.114	-1.227	-1.104
	Median	-0.358	-0.338	-0.337	-0.319	-0.363	-0.311
	Max	31.833	23.370	26.111	29.370	26.505	24.689
Iteration (BFGS)		38	40	38	42	64	97
logL		-1.03E+04	-8.32E+03	-1.03E+04	-1.23E+04	-7.95E+03	-9.45E+03
DOF		34	34	34	35	35	35
Theta					1.38	1.82	1.35

Model Validation

Severity Types	ZINB versus negative binomial	Vuong Test-Statistic	p-value	ZIP versus Poisson	Vuong Test-Statistic	p-value
Total Crashes	ZINB > NB	2.6410	0.0041	ZIP > Poisson	15.164	0.0000
Injury Crashes	ZINB > NB	2.1639	0.0152	ZIP > Poisson	8.1841	0.0000
PDO Crashes	ZINB > NB	3.4129	0.0003	ZIP > Poisson	12.7071	0.0000

Limitations

- Although zero-inflated models offer improved statistical fit to crash data in many cases, it is argued that the inherent assumption of a dual state process underlying the development of these models is inconsistent with crash data
- This research aims to utilize ZIP and ZINB models to investigate the significance of the recoded categorical values of the geometric factors for traffic crashes of different severities which has not been done extensively in crash data analysis before.

Conclusions

Based on the test statistic, ZIP and ZINB models provided a better fit than conventional Poisson or negative binomial model for total, injury and PDO crashes. One future scope of this research is to introduce non-parametric statistical methods to the extended dataset to compare the statistical significance.