A Note on Estimating Safety Performance Functions with a Flexible Specification of Traffic Volume

Naveen Eluru*

Professor

Department of Civil, Environmental & Construction Engineering

University of Central Florida

Tel: 407-823-4815

Email: naveen.eluru@ucf.edu

Vikash V. Gayah

Associate Professor

Department of Civil and Environmental Engineering,

The Pennsylvania State University

Tel: 814-865-4014

E-mail: gayah@engr.psu.edu

*Corresponding Author

January 3, 2022

Abstract

In this note, a flexible approach to allow for variation in the impact of traffic volume in the estimation of Safety Performance Functions (SPFs) is proposed. The approach generalizes a recently proposed approach by Gayah and Donnell (2021) (GD) titled "Estimating safety performance functions for two-lane rural roads using an alternative functional form for traffic volume". GD approach proposes a multiple regime structure for AADT impact while explicitly constraining the impact at the regime threshold to be the same. While the GD approach provides a flexible structure, the framework as proposed calls for careful judgement for threshold selection and additional model estimation complexity for the AADT constraint. The current note establishes the equivalence of the proposed approach with the GD approach and subsequently presents a more flexible model structure that improves on the GD approach. Subsequently, we document the advantages of our proposed approach in terms of model estimation, parameter significance testing, flexibility to consider multiple traffic volume ranges and ease of accommodating random parameters for analysis. Finally, we present potential directions for future research.

Keywords: Safety Performance Function, traffic volume, and piece-wise linear regression

Background

Traffic volume — typically represented using Average Annual Daily Traffic (AADT) — are considered in Safety Performance Function (SPF) estimation employing a natural logarithm transformation. The natural logarithm of the expected number of crashes on a facility i (μ_i) is specified as a function of AADT (in log form) and other explanatory variables using a log-link function as follows:

$$ln\ (\mu_i) = \beta_{AADT}\ ln(\mathrm{AADT}_i) + \gamma \boldsymbol{z}_i + \varepsilon_i \quad (1)$$

where β_{AADT} represents the parameter for the natural logarithm of AADT for facility i, $\boldsymbol{z}_i$ is a vector of other explanatory variables associated with facility i including a constant, γ is a vector of coefficients to be estimated, and $\exp(\varepsilon_i)$ is a gamma distributed error term with mean 1 and variance α.

Traditional SPFs restrict the impact of AADT on crash frequency to remain the same across the range of AADT[1]. However, as highlighted in several research efforts, it is possible that the influence of AADT on crash frequency might vary with AADT. Towards addressing this limitation, Gayah and Donnell (2021) – referred to as GD in the rest of the note – recently proposed an alternative functional form for considering AADT in estimating SPFs. The GD functional form facilitates different elasticities of traffic volume for different traffic volume ranges and addresses the limitations of approaches documented in prior research (Shankar et al., 1998; Ulfarsson and Shankar, 2003; Anastasopoulos and Mannering, 2009; Venkataraman et al., 2011; Hauer, 2015; see Gayah and Donnell (2021) for more details). It provides a continuous relationship between crash frequency and AADT, which more realistically reflects how crash frequencies should vary as traffic volume increases. The use of discrete traffic volume ranges also provides natural breakpoints to consider how the effects of other features might vary on low- or high-volume roads.

Proposed Approach

The GD approach assumes that two regimes exist in the crash frequency model. Regime 1 is applied when AADT is below the pre-defined AADT threshold, while Regime 2 is applied otherwise. To ensure continuity in AADT impact across the regimes at the threshold value, an explicit constraint is added to the model estimation process. This process can be extended to three or more regimes to accommodate a continuous relationship between safety performance and AADT, while allowing the impact of traffic volume to vary across the range of AADT values. However, this requires a priori determination of the AADT threshold(s) used to separate the regimes. This is far from trivial and involves an iterative procedure. Further, the constrained maximization approach places substantial burden for estimating random parameters. Overall, while the GD approach provides a flexible structure, the framework as proposed calls for careful

[1] The reader would note that several alternative approaches such as a semi-parametric approach or adoption of generalized additive models have been proposed to allow for varying impact of AADT in safety literature (for example, see Kononov et al., 2011 and Zhang et al., 2012).

judgement for threshold selection, additional model estimation complexity due to the presence of the AADT regime constraint (in particular for random parameter estimation).

In this note, we propose an approach that offers a simpler process to achieve the objectives sought by Gayah and Donnell (2021). The proposed approach subsumes the GD method without any need for constrained optimization. Traditional SPFs consider a linear representation of the impact of ln (AADT) on crash propensity ($ln\ (\mu_i)$) as presented in Equation 1. The assumption of restricting the parameter (β_{AADT}) to be the same implies a constant marginal impact or slope for the ln (AADT) variable irrespective of AADT value. To address this restriction, in the proposed approach, we consider a piece-wise linear representation of the impact our variable of interest (natural logarithm of AADT). To elaborate, we consider that the influence of ln (AADT) will stay the same within a specified range and the slope of the line can increase or decrease at pre-defined intervals selected by the analyst. Consider the following illustrative formulation

$$ln\ (\mu_i) = \beta_{AADT}\ ln(\mathrm{AADT}_i) + \delta_1 AADT_{inc1} + \delta_2 AADT_{inc2}\ + \gamma \boldsymbol{z}_i + \varepsilon_i \qquad (2)$$

where, the newly added independent variables $AADT_{i\_inc1}$ and $AADT_{i\_inc2}$ are defined as follows

$$AADT_{i\_inc1} = \ Max[0, ln(AADT_i) - ln(T_1)] \qquad (3)$$

$$AADT_{i\_inc2} = \ Max[0, ln(AADT_i) - ln(T_2)] \qquad (4)$$

and δ_1 and δ_2 represent the parameters for the newly created independent variable to capture the changes to the impact of AADT variable ($ln(\mathrm{AADT}_i)$); T_1 and T_2 represent the $ln(AADT_i)$ threshold points where the slope is expected to change.

To facilitate the reader's understanding, the traditional and proposed model forms are shown in Figure 1. The x-axis represents the $ln(AADT_i)$ while the y-axis represents crash propensity ($ln\ (\mu_i)$). The figure on the left presents the traditional model where the marginal effect (or slope) of the relationship is constant. In the figure on the right, we present the more flexible model structure. Specifically, at $ln(AADT_i)$ value = T_1 the slope changes from β_{AADT} to $\beta_{AADT} + \delta_1$ and is associated with $AADT_{i\_inc1}$. A positive value for δ_1 would imply an increase in the marginal effect of AADT (as shown in the figure for illustrative purposes) while a negative value would represent a reduction in marginal effect of AADT. A second shift in marginal effect is considered at T_2. The overall slope beyond T_2 is ($\beta_1 + \delta_1 + \delta_2$). The approach can accommodate additional changes in slope in a similar manner. The reader would note that the newly created independent variables (such as $AADT_{i\_inc1}$ and $AADT_{i\_inc2}$) are non-negative only after the corresponding threshold is crossed and 0 otherwise. The proposed approach has been employed for variables in their linear form for relaxing the constant slope assumption for the contribution of travel time to mode choice utility (see Pinjari and Bhat, 2006). The change in slope at the thresholds (such as T_1 and T_2) can be attributed to differences in driver behavior across facilities with different volumes (increased caution exercised in locations with increasing volumes), potential presence of custom design features closely associated with traffic volumes and other AADT associated factors.

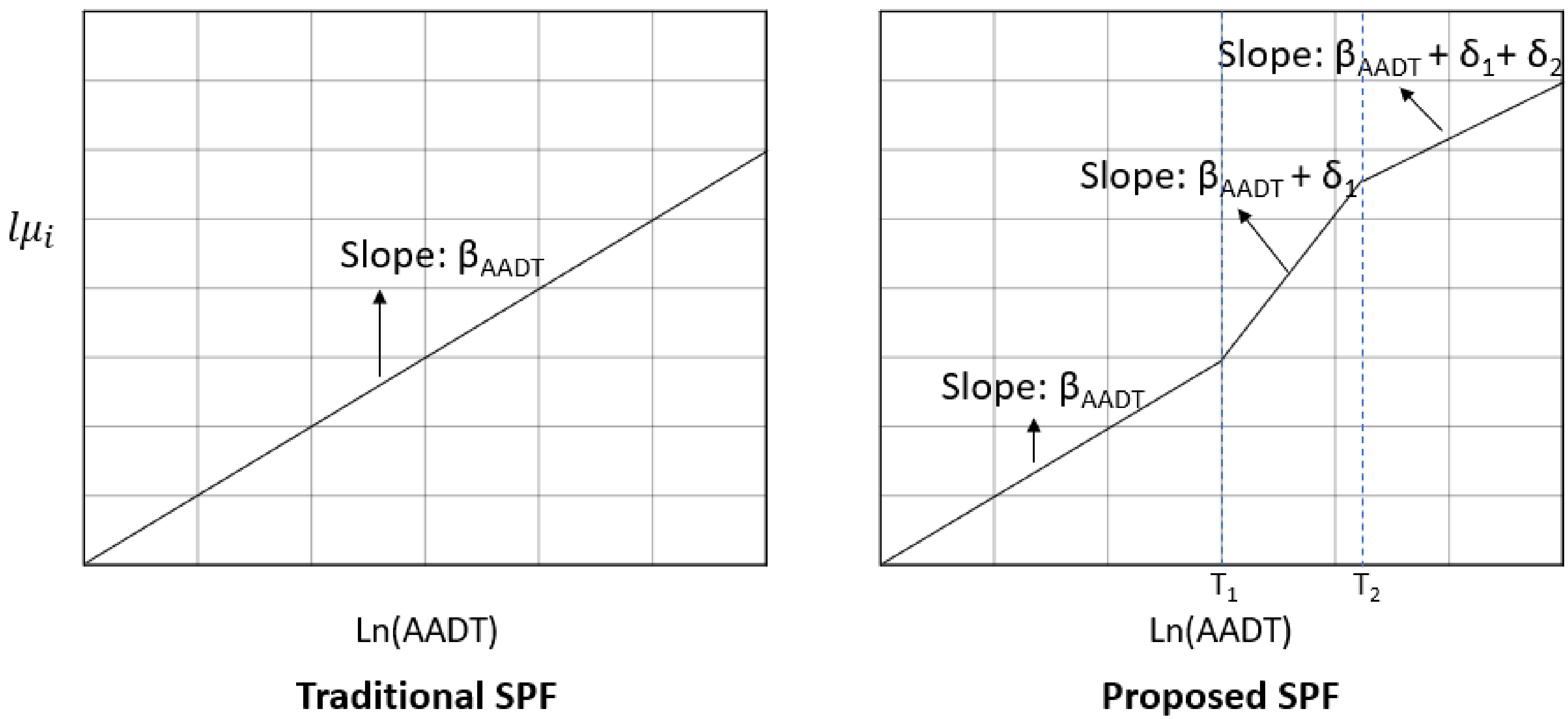

Figure 1: Illustration of the traditional SPF and the proposed SPF

Empirical Analysis

In this section, we present the results of the model estimated employing the proposed functional form. For the sake of brevity, data preparation and sample characteristics are not presented (interested readers can see Gayah and Donnell (2021)). We start with illustrating how the proposed model can represent the GD functional form and subsequently present a more flexible model with our proposed system.

GD Approach and Proposed Equivalent Model

The GD approach considers two ranges of AADT employing a single threshold to differentiate the impact at AADT = 1900. We illustrate how our proposed framework can replicate this one threshold model in our framework. The model results are presented in Table 1. The second column panel presents the results from Gayah and Donnell (2021) and the third column panel presents the One Threshold framework for our proposed alternative form. The propensity equation and the AADT variable employed in the estimation for our proposed form is defined as follows

$$ln\ (\mu_i) = \beta_{AADT}\ ln(\mathrm{AADT}_i) + \delta_1 AADT_{i\_inc\_1900} + \gamma \boldsymbol{z}_i + \varepsilon_i \quad (5)$$

where

$$AADT_{i\_inc\_1900} = Max[0, ln(\mathrm{AADT}_i) - ln(1900)] \quad (6)$$

The following observations can be made from the review of the results in Table 1. *First*, apart from the second constant in the GD approach and second AADT related variable in the two systems, all other variables are identical across the two systems. The model fit for the two systems is also identical. Further, it is easy to illustrate that the other parameters also offer the same mathematical relationship across the regimes. For example, the reduced constant for the higher traffic volume regime is -4.7634 in the GD approach (1.485 reduction from -6.2488). The drop compensates the drop in AADT parameter at an AADT value of 1900 (i.e., (0.7723-0.5756)*ln(1900) = 1.485). Also, the new incremental coefficient in the proposed model -0.1967 is the same as the change in AADT parameter from the GD system (0.5756 – 0.7723 = -0.1967). The comparison clearly illustrates that the two models are identical in how impact of AADT and all other variables are accommodated. *Second*, the number of parameters in Gayah and Donnell (2021) is higher by 1. Based on what we have shown above, it indicates that the second constant in GD approach is an artifact of the specification and not an additional degree of freedom. *Finally*, given the nature of the separate regime model in the GD approach, the AADT coefficient for both regimes will be significant and quite likely different in most cases. However, this does not establish that the difference is statistically significant. In our proposed framework, the comparison is more straightforward. If the parameter for $AADT_{i\_inc\_1900}$ is statistically insignificant then one can conclude there is no reason for using the different AADT ranges. The ease of undertaking this statistical test is of value to analysts testing if AADT impact is indeed different across different ranges.

Flexible version of the Proposed Model

The proposed system can be employed to test several AADT ranges. For this exercise, we considered AADT values of 500, 1000, 2000, 5000 and 10,000 as thresholds. The exact independent variables generated and considered in our model estimation include:

$$AADT_{i\_inc\_500} = Max[0, ln(AADT_i) - ln(500)] \quad (7)$$

$$AADT_{i\_inc\_1000} = Max[0, ln(AADT_i) - ln(1000)] \quad (8)$$

$$AADT_{i\_inc\_2000} = Max[0, ln(AADT) - ln(2000)] \quad (9)$$

$$AADT_{i\_inc\_5000} = Max[0, ln(AADT) - ln(5000)] \quad (10)$$

$$AADT_{i\_inc\_10000} = Max[0, ln(AADT) - ln(10000)] \quad (11)$$

The final model results with this specification are presented in Table 2. From the results, we can observe that only two of the five tested variables are statistically significant. The results indicate that there exist three ranges of AADT impact: 0-2000, 2000-5000 and above 5000. The coefficients associated with AADT within these ranges can be computed as 0.7998, 0.4522 (0.7998 - 0.3476) and 0.7799 (0.7998 - 0.3476 + 0.3277). The flexible system significantly improved the log-likelihood compared to the one threshold system. The log-likelihood test value comparing the two model systems is 2 * (-19525.262+19531.932) = 13.34; a value higher than the corresponding ψ^2 test-statistic for any level of significance for one degree of freedom.

Advantages of the Proposed Framework

The proposed approach offers several advantages. *First*, the proposed approach does not require any additional infrastructure to estimate the model. Unlike the GD approach, which requires a constraint to be included in the model estimation, the proposed model can be estimated with currently employed software for SPF estimation without any modification. The approach requires only an addition of independent variables to allow for $ln(AADT_i)$ impact to vary across its range as shown in equations 3 and 4. *Second*, the determination of whether the impact changes beyond a threshold value is relatively straightforward. The t-statistic of the parameters (such as δ_1 and δ_2 in Equation 2) for the new independent variables (such as $AADT_{i\_inc1}$ and $AADT_{i\_inc2}$) provide a clear indication of the parameter significance. Thus, the analyst can easily evaluate if the proposed additional parameter results represent a significant change in AADT impact. The t-statistic based evaluation can also be augmented with a likelihood ratio test[2] to confirm the findings. In most empirical cases, the t-statistic evaluation and likelihood ratio test are closely aligned. The computation of such comparisons for the GD approach will be far from straight forward and would

[2] The likelihood ratio test statistic is computed as -2 [L_R – L_{UR}] where L_R represents the restricted model and L_{UR} represents the unrestricted model (model with higher number of parameters). The computed statistic is compared to χ^2 distribution with *K* degrees of freedom where *K* is the difference in the number of parameters in the two models (see Wilks, 1938).

require complex procedures for hypothesis testing. *Third*, as opposed to the GD approach, consideration of multiple AADT ranges entails only small incremental computation in the form of independent variable generation. To elaborate, the analyst will only need to generate additional independent variables as illustrated in Equations 7-11 to examine changes to AADT parameter in their empirical context. *Finally*, the flexibility to accommodate multiple thresholds might increase analyst burden in determining the appropriate number of thresholds and the actual threshold values. The analyst will need to draw on their experience to identify the number of thresholds and cut-off points. A useful approach would be to plot the AADT frequency distribution and identify natural cut-off points in the data. We recommend considering up to 4 cut-off points for typical data [resulting in 5 ranges of AADT]. After determining the cut-off points, the model can be estimated relatively easily with additional independent variables (following equations 7-11). The variables included should be evaluated for statistical significance using the t-statistic or the likelihood ratio test. At the same time, given the flexible nature of the proposed approach, differences in the number and cut-offs are likely to result in only smaller overall changes to AADT impact across the study region.

Consideration of Unobserved Heterogeneity

The proposed framework is focused on simplifying the GD paper framework for estimating AADT parameters for different ranges. The current formulation of the GD approach employs a constraint maximization framework for model estimation. Within this framework, estimating random parameters that satisfy the constraint across regimes is far from straightforward. For example, let us consider random parameters for each of the AADT parameter across the two regimes. To estimate the model, the constraint that AADT parameters from different regimes need to be the same at the threshold value needs to be satisfied. Given the usual unbounded nature of the random parameters (such as those that arise from a multivariate normal distributional assumption), it is not possible to test the constraint for every realization of the random parameter. Traditional estimation approaches for RP will estimate the constraint maximization framework over a set of realizations (say 500-1000 draws). However, matching the constraint for these draws does not ensure that the constraint will be met for all possible realizations. To elaborate, it is not possible to estimate the simulated probability for all realizations of the random parameters to ensure that the constraint is met. Hence, overlaying random parameters could be analytically infeasible in the GD approach. The proposed approach in our note by eliminating the need for constrained optimization, circumvents the problem and allows for employing the approaches routinely employed for estimating random parameters.

The reader would note that recent research in accommodating for unobserved heterogeneity can be overlaid on the proposed model specification to accommodate for the influence of unobserved factors. The methods to be considered include random parameters approach, random parameters with heterogeneity in means and variances approach, latent segmentation methods, and latent segmentation methods with random parameters in segments (see for example Anastasopoulos and Mannering, 2009; Xiong and Mannering, 2013; Yasmin et al., 2014; Mannering et al., 2016; Behnood and Mannering, 2016; for a recent review of literature on unobserved heterogeneity see Bhowmik et al., 2021). More recently, there is also growing recognition of the importance of temporal stability of parameters in safety literature as documented in Mannering (2018). The approaches accounting for temporal instability include Markov switching models, subsample-based model comparisons and scaled model frameworks (for example see Malyshkina et al. 2009; Marcoux et al., 2018; Tirtha et al., 2020; Islam and Mannering, 2021; Yan et al., 2021). The approaches employed in these studies can also be embedded within our proposed approach easily. Finally, analysts considering advanced model structures are encouraged to consider recent research on prediction with random parameter model systems (Hou et al., 2021a; Xu et al., 2021; Hou et al., 2021b).

While estimating random parameters is beneficial, the analyst should recognize the inherent challenges with the estimation process. The estimation of fixed parameters model is easy and accurate. The process of estimating random parameters requires simulated maximum likelihood estimation as the probability functions are no longer analytically tractable. Hence, even with an adequate number of random draws for estimation, the likelihood function is still an approximation. The simulation-based estimation can run into challenges particularly in gradient and hessian computation when many random parameters are simultaneously considered or in cases where the multidimensional log-likelihood function is flat near the optimal region (see Bhat, 2011 for a detailed discussion). In such scenarios, pinning down the value of the random parameter is not easy and sometimes can lead to non-global optima. Hence, it is beneficial to estimate the best fixed parameter model and then augment the model with random parameters (drawing on earlier work cited above). This approach reduces the potential complexity involved by reducing the number of random parameters being estimated simultaneously.

Conclusion

The brief note presented a flexible approach to allow for variation in the impact of AADT across different ranges of AADT. The approach generalizes a recently proposed approach by Gayah and Donnell (2021). The note establishes the equivalence of the proposed approach with Gayah and Donnell (2021) and subsequently presents a more flexible model structure that improves on the Gayah and Donnell (2021) approach. We document the advantages of our proposed approach in terms of model estimation, parameter significance testing and the flexibility to consider multiple AADT ranges for analysis. Subsequently, we also discuss how the proposed methodology addresses the limitation of GD approach in accommodating for unobserved heterogeneity and thus can be easily extended to accommodate for unobserved heterogeneity and temporal instability. Finally, it would be interesting for future research efforts to compare the proposed approach with semi-parametric approaches or generalized additive models that offer alternative approaches that allow for varying impact of AADT. For the comparison, it might be beneficial to employ a comprehensive experimental design framework for simulating data generated using a wide range of independent variable distributions (see Eluru, 2013; Bhowmik et al., 2021; Xu et al., 2021 for example studies employing simulation for model comparison).

Author contribution statement

Conceptualization: Eluru and Gayah; Methodology: Eluru and Gayah; Data curation: Gayah and Eluru; Writing: Eluru and Gayah

Declaration of Competing Interest

The authors have no conflicts of interest to declare

References

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.

Bhat CR., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transportation Research Part B: Methodological Aug 1;45(7):923-39.

Bhowmik T., Yasmin S., and Eluru N., 2021. "A New Econometric Approach for Modeling Several Count Variables: A Case Study of Crash Frequency Analysis by Crash Type and Severity", Transportation Research Part B Volume 153, November 2021, Pages 172-203

Hauer, E. (Ezra), 2015. The Art of Regression Modeling in Road Safety. Springer.

Hou, Q., Huo, X., Tarko, A., Leng, J., 2021a. Comparative analysis of alternative random parameters count data models in highway safety. Analytic Methods in Accident Research 30, 100158.

Hou, Q., Huo, X., Leng, J. and Mannering, F., 2021b. A note on out-of-sample prediction, marginal effects computations, and temporal testing with random parameters crash-injury severity models. Analytic Methods in Accident Research, p.100191.

Eluru N. (2013), "Evaluating Alternate Discrete Choice Frameworks for Modeling Ordinal Discrete Variables," Accident Analysis & Prevention, 55 (1), pp. 1-11

Gayah, V.V. and Donnell, E.T., 2021. Estimating safety performance functions for two-lane rural roads using an alternative functional form for traffic volume. Accident Analysis & Prevention, 157, p.106173.

Islam, M. and Mannering, F., 2021. The role of gender and temporal instability in driver-injury severities in crashes caused by speeds too fast for conditions. Accident Analysis & Prevention, 153, p.106039.

Kononov, J., Lyon, C. and Allery, B.K., 2011. Relation of flow, speed, and density of urban freeways to functional form of a safety performance function. Transportation research record, 2236(1), pp.11-19.

Malyshkina, N.V., Mannering, F.L. and Tarko, A.P., 2009. Markov switching negative binomial models: an application to vehicle accident frequencies. Accident Analysis & Prevention, 41(2), pp.217-226.

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analytic methods in accident research, 17, pp.1-13.

Marcoux. R., Yasmin S., Eluru N., and Rahman M., 2018. "Evaluating Temporal Variability of Exogenous Variable Impacts Over 25 Years: An Application of Scaled Generalized Ordered Logit Model for Driver Injury Severity", Analytic Methods in Accident Research Volume 20, December 2018, Pages 15-29

Pinjari, A.R. and Bhat, C., 2006. Nonlinearity of response to level-of-service variables in travel mode choice models. Transportation research record, 1977(1), pp.67-74.

Tirtha S.D., Yasmin S., and Eluru N., 2020, "Modeling of Incident Type and Incident Duration Using Data from Multiple Years" Analytic Methods in Accident Research Volume 28, December 2020, 100132

Shankar, V.N., Albin, R.B., Milton, J.C., Mannering, F.L., 1998. Evaluating median crossover likelihoods with clustered accident counts an empirical inquiry using the random effects negative binomial model. Transp. Res. Rec. J. Transp. Res. Board 1635, 44–48.

Ulfarsson, G.F., Shankar, V.N., 2003. Accident count model based on multiyear cross-sectional roadway data with serial correlation. Transp. Res. Rec. J. Transp. Res. Board 1840, 193–197.

Venkataraman, N.S., Ulfarsson, G.F., Shankar, V., Oh, J., Park, M., 2011. Model of relationship between interstate crash occurrence and geometrics. Transp. Res. Rec. J. Transp. Res. Board 2236, 41–48.

Wilks, S.S., 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. The annals of mathematical statistics, 9(1), pp.60-62.

Xu, P., Zhou, H., Wong, S.C., 2021. On random-parameter count models for out-of-sample crash prediction: Accounting for the variances of random-parameter distributions. Accident Analysis and Prevention 159, 106237.

Yan, X., He, J., Zhang, C., Liu, Z., Wang, C. and Qiao, B., 2021. Spatiotemporal instability analysis considering unobserved heterogeneity of crash-injury severities in adverse weather. Analytic Methods in Accident Research, 32, p.100182.

Zhang, Y., Xie, Y. and Li, L., 2012. Crash frequency analysis of different types of urban roadway segments using generalized additive model. Journal of safety research, 43(2), pp.107-114.

Table 1: Model Estimation Results for One Threshold Systems

Coefficient	Gayah and Donnell (2021)			One Threshold		
	Coefficient	Std. error	p-value	Coefficient	Std. error	p-value
Constant when traffic volume lower than threshold	-6.2488	0.2369	<0.001	---	---	---
Constant when traffic volume higher than threshold	-4.7634	0.2039	<0.001	---	---	---
Constant	---	---	---	-6.2488	0.2369	<0.001
Natural log of traffic volume (lower than threshold) [veh/day]	0.7723	0.0323	<0.001	---	---	---
Natural log of traffic volume (greater than threshold) [veh/day]	0.5756	0.0248	<0.001	---	---	---
$\ln(AADT_i)$	---	---	---	0.7723	0.0323	<0.001
$AADT_{i\_inc\_1900}$	---	---	---	-0.1967	0.0471	<0.001
Natural log of segment length [mi]	0.8032	0.0366	<0.001	0.8032	0.0366	<0.001
Presence of a passing zone [1 if present, 0 otherwise]	-0.1274	0.0244	<0.001	-0.1274	0.0244	<0.001
Presence of shoulder rumble strips [1 if present, 0 otherwise]	-0.1096	0.0540	0.042	-0.1096	0.0540	0.042
Access density [access points/mi]	0.0108	0.0009	<0.001	0.0108	0.0009	<0.001
Curve density [curves/mi]	0.0398	0.0059	<0.001	0.0398	0.0059	<0.001
Total degree of curvature per mile along the segment [deg/100 ft/mile]	0.0014	0.0003	<0.001	0.0014	0.0003	<0.001
Roadway segment in Bradford county [1 if yes, 0 otherwise]	0.1043	0.0286	<0.001	0.1043	0.0286	<0.001
Roadway segment in Lycoming or Montour counties [1 if yes, 0 otherwise]	0.0985	0.0313	<0.001	0.0985	0.0313	<0.001
Roadway segment in Sullivan or Union counties [1 if yes, 0 otherwise]	-0.1317	0.0382	<0.001	-0.1317	0.0382	<0.001
Over-dispersion parameter	0.4773	0.0262	<0.001	0.4773	0.0262	<0.001
AADT threshold	1900			1900		
Log-likelihood at convergence	-19531.932			-19531.932		

Table 2: Flexible Model Estimation Results for Multiple Thresholds

Coefficient	Multiple Thresholds		
	Coefficient estimate	Std. error	p-value
Constant	-6.4219	0.2373	<0.001
Natural log of segment length [mi]	0.7999	0.0366	<0.001
$\ln(AADT_i)$	0.7998	0.0325	<0.001
$AADT_{i\_inc\_2000}$	-0.3476	0.0626	<0.001
$AADT_{i\_inc\_5000}$	0.3277	0.0884	<0.001
Presence of a passing zone [1 if present, 0 otherwise]	-0.1224	0.0244	<0.001
Presence of shoulder rumble strips [1 if present, 0 otherwise]	-0.1275	0.0544	0.019
Access density [access points/mi]	0.0106	0.0009	<0.001
Curve density [curves/mi]	0.0404	0.0059	<0.001
Total degree of curvature per mile along the segment [deg/100 ft/mile]	0.0014	0.0003	<0.001
Roadway segment in Bradford county [1 if yes, 0 otherwise]	0.0993	0.0287	0.001
Roadway segment in Lycoming or Montour counties [1 if yes, 0 otherwise]	0.0910	0.0313	0.004
Roadway segment in Sullivan or Union counties [1 if yes, 0 otherwise]	-0.1504	0.0386	<0.001
Over-dispersion parameter	0.4733	0.0262	<0.001
Log-likelihood at convergence	-19525.262		