Short Communication

On random-parameter count models for out-of-sample crash prediction: Accounting for the variances of random-parameter distributions

Pengpeng Xu [a,b,*], Hanchu Zhou [c,d], S.C. Wong [a,e]

[a] *Department of Civil Engineering, The University of Hong Kong, Hong Kong, China*
[b] *School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China*
[c] *School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, China*
[d] *School of Data Science, City University of Hong Kong, Hong Kong, China*
[e] *Guangdong – Hong Kong – Macau Joint Laboratory for Smart Cities, Hong Kong, China*

A B S T R A C T

One challenge faced by the random-parameter count models for crash prediction is the unavailability of unique coefficients for out-of-sample observations. The means of the random-parameter distributions are typically used without explicit consideration of the variances. In this study, by virtue of the Taylor series expansion, we proposed a straightforward yet analytic solution to include both the means and variances of random parameters for unbiased prediction. We then theoretically quantified the systematic bias arising from the omission of the variances of random parameters. Our numerical experiment further demonstrated that simply using the means of random parameters to predict the number of crashes for out-of-sample observations is fundamentally incorrect, which necessarily results in the underprediction of crash counts. Given the widespread use and ongoing prevalence of the random-parameter approach in crash analysis, special caution should be taken to avoid this silent pitfall when applying it for predictive purposes.

1. Introduction

Since its introduction by Milton et al. (2008) and Anastasopoulos and Mannering (2009), the random-parameter approach has attracted considerable research interest and has become the benchmark for crash analysis (Mannering and Bhat, 2014). Numerous studies have demonstrated its superiority in interpretability, with a substantial improvement in goodness-of-fit and the ability to account for unobserved heterogeneity (Mannering et al., 2016). Recent studies reported that the approach empirically underperformed in out-of-sample crash prediction compared with its fixed-parameter counterpart (Tang et al., 2019; Hou et al., 2020). However, this underperformance can be readily attributed to using predictions that are based only on the means of the random parameters while ignoring their variances. Our study therefore attempts to demonstrate such a bias by providing a straightforward yet analytic solution to incorporate both the means and variances of random parameters for unbiased prediction.

2. Methods

Let Y_i denote the number of crashes for observation $i (i = 1, 2, ..., n)$ during a certain period. Z_i and X_{ij} represent the exposure and the jth explanatory variable, respectively. When developing the random-parameter approach to model crash frequencies, without loss of generality, we formulate the following functional form given the random, non-negative, and integral nature of crash counts:

$$Y_i \sim \text{Poisson}(\lambda_i)$$
$$\lambda_i = \beta_0 Z_i^{\beta_{1i}} exp\left(\sum_{j=2}^{k} \beta_{ji} X_{ij} + u_i \right) \qquad (1)$$

where λ_i is the parameter of Poisson distribution (i.e., the expected number of crashes), β_0 is the intercept, $\beta_{ji} (j = 1, 2, ..., k)$ refers to the jth regression coefficients, and u_i denotes the unstructured effect. Compared with the negative binomial model (also known as the Poisson-gamma model in which $exp(u_i)$ follows a Gamma distribution with mean 1 and variance α; Lord and Miranda-Moreno, 2008), a more flexible

---

alternative is to assume u_i as a normal distribution as suggested by Lord and Mannering (2010):

$$u_i \sim \text{Normal}\left(0, \sigma_u^2\right) \qquad (2)$$

Unlike the conventional Poisson-lognormal model with fixed parameters, by specifying β_{ji} as a predefined distribution with mean $\overline{\beta}_j$ and variance σ_j^2, each observation now has its individual coefficients. In theory, β_{ji} can be assumed to follow any distributions, but the normal distribution is mostly used, given its better statistical fit (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009; Ukkusuri et al., 2011; Venkataraman et al., 2013; Chen and Tarko, 2014; Xu and Huang, 2015; Amoh-Gyimah et al., 2016; Meng et al., 2017; Cai et al., 2018; Hou et al., 2018, 2021; Tang et al., 2019).

A challenge arises when applying the estimated random-parameter model to predict the number of crashes for out-of-sample observations, as in reality, it appears to be impossible to know the exact value of $\beta_{j[i]}$ for sample $[i]([i] \neq i)$. If $\overline{\beta}_j$ is used as the default to estimate the number of crashes (i.e., $\lambda_{[i]}$; Mannering et al., 2016; Tang et al., 2019; Hou et al., 2020), we then have:

$$\lambda_{[i]} \approx \overline{\lambda}_{[i]} = \beta_0 Z_{[i]}^{\overline{\beta}_1} exp\left(\sum_{j=2}^{k} \overline{\beta}_j X_{[i]j} \right) \qquad (3)$$

However, the expectation of $\lambda_{[i]}$ is not only dependent on $\overline{\beta}_j$ but also on the variances σ_j^2 and σ_u^2. Indeed, if we take a Taylor series expansion near the point $(\beta_{j[i]} = \overline{\beta}_j, u_{[i]} = 0)$, $\lambda_{[i]}$ can be closely approximated using a quadratic polynomial as follows (Wong and Wong, 2019):

5000, $Z_i \sim \text{Uniform}(200, 20000)$, $X_i \sim \text{Uniform}(0, 20)$, $\beta_0 = 0.004$, $\beta_{1i} \sim \text{Normal}(0.70, 0.06^2)$, $\beta_{2i} \sim \text{Normal}(0.03, 0.03^2)$, and $u_i \sim \text{Normal}(0, 0.60^2)$. Only one explanatory variable was considered here for simplicity. We then sequentially sampled the values of $Z_i, X_i, \beta_{1i}, \beta_{2i}$, and u_i from the aforementioned distributions, calculated λ_i, and generated Y_i using the freeware RStudio (R Core Team, 2013). Such a simulated dataset matched well with the empirical data collected by Tang et al. (2019) for rural two-lane roads in Pennsylvania. Detailed descriptive statistics of our data under investigation are listed in Table 1.

To evaluate the out-of-sample predictive performance of the random-parameter count model, we performed a 10-fold cross-validation (James et al., 2013; Hou et al., 2021), in which the original 5000 observations were randomly partitioned into 10 mutually exclusive subsets of equal size. Each time, one subset was left out as the validation while the remaining subsets were combined for model estimation. This process was repeated 10 times. We repeatedly predicted the crash frequencies for validation samples as $\overline{\lambda}_{[i]}$ and $E(\lambda_{[i]})$ using Eq. (3) and Eq. (5), respectively. In each round, in addition to the random-parameter model, we also developed its fixed-parameter counterpart for comparison. To characterize their overall predictive accuracy, since each metric only reflects partial aspects of error characteristics (Chai and Draxler, 2014), we computed three commonly used measures namely the mean bias error (MBE), the mean absolute error (MAE), and the root-mean-square error (RMSE). We then plotted the cumulative residuals (CURE; Elvik et al., 2013; Hauer, 2015; Srinivasan et al., 2016; Tang et al., 2019; Hou et al., 2020) to visualize any anomalies. Models with a CURE curve oscillating around zero within the 95% confidence interval (CI) are regarded as reliable in predictive performance (Hauer, 2015). Consistent

$$\lambda_{[i]} \cong \beta_0 Z_{[i]}^{\overline{\beta}_1} exp\left(\sum_{j=2}^{k} \overline{\beta}_j X_{[i]j} \right) + \sum_{j=1}^{k} \frac{\partial \lambda_{[i]}}{\partial \beta_{j[i]}} \left(\beta_{j[i]} - \overline{\beta}_j \right) + \frac{\partial \lambda_{[i]}}{\partial u_{[i]}} \left(u_{[i]} - 0 \right) + \sum_{j=1}^{k} \frac{1}{2} \frac{\partial^2 \lambda_{[i]}}{\partial \beta_{j[i]}^2} \left(\beta_{j[i]} - \overline{\beta}_j \right)^2 + \frac{1}{2} \frac{\partial^2 \lambda_{[i]}}{\partial u_{[i]}^2} (u_{[i]} - 0)^2 \qquad (4)$$

where $\frac{\partial \lambda_{[i]}}{\partial \beta_{j[i]}}$ and $\frac{\partial \lambda_{[i]}}{\partial u_{[i]}}$ refer to the first-order partial derivative of $\lambda_{[i]}$ to $\beta_{j[i]}$ and $u_{[i]}$, respectively. $\frac{\partial^2 \lambda_{[i]}}{\partial \beta_{j[i]}^2}$ and $\frac{\partial^2 \lambda_{[i]}}{\partial u_{[i]}^2}$ are the corresponding second-order partial derivatives.

Taking the expectation on both sides of Eq. (4), we have:

with Tang et al. (2019) and Hou et al. (2020), we also used the latest econometric software LIMDEP 11 (Greene, 2016) to estimate our models. Reliable parameters for the random-parameter count models could be estimated via 600 Halton draws based on the simulated maximum likelihood method (Tang et al., 2019; Hou et al., 2020). The results are presented in Table 2 and Fig. 1.

$$E(\lambda_{[i]}) = \beta_0 Z_{[i]}^{\overline{\beta}_1} exp\left(\sum_{j=2}^{k} \overline{\beta}_j X_{[i]j} \right) + \sum_{j=1}^{k} \frac{1}{2} \frac{\partial^2 \lambda_{[i]}}{\partial \beta_{j[i]}^2} \sigma_j^2 + \frac{1}{2} \frac{\partial^2 \lambda_{[i]}}{\partial u_{[i]}^2} \sigma_u^2 = \left\{ 1 + \frac{\sigma_1^2}{2} [ln(Z_{[i]})]^2 + \sum_{j=2}^{k} \frac{\sigma_j^2}{2} X_{[i]j}^2 + \frac{\sigma_u^2}{2} \right\} \beta_0 Z_{[i]}^{\overline{\beta}_1} exp\left(\sum_{j=2}^{k} \overline{\beta}_j X_{[i]j} \right) \qquad (5)$$

3. Results

As Table 2 shows, although the predictive accuracy of $E(\lambda_{[i]})$ was slightly lower than that of $\overline{\lambda}_{[i]}$ in terms of the average magnitude of

The difference between Eq. (5) and (3) can then be quantified as:

$$E(\lambda_{[i]}) - \overline{\lambda}_{[i]} = \left(\frac{\sigma_1^2}{2} [ln(E_{[i]})]^2 + \sum_{j=2}^{k} \frac{\sigma_j^2}{2} X_{[i]j}^2 + \frac{\sigma_u^2}{2} \right) \beta_0 E_{[i]}^{\overline{\beta}_1} exp\left(\sum_{j=2}^{k} \overline{\beta}_j X_{[i]j} \right) \qquad (6)$$

That is, using $\overline{\lambda}_{[i]}$ as the prediction will always result in a downward bias, given the convex property of the link function in Eq. (1). This silent pitfall has been ignored by safety analysts when applying random-parameter models for predictive purposes (Tang et al., 2019; Hou et al., 2020).

To validate, we conducted the following numerical experiment, given its advantages in maneuverability and reproducibility. We set $n =$

Table 1
Descriptive statistics for the simulated data.

	Mean	Median	SD	Min	Max
Y_i	5.129	3.000	7.442	0.000	161.000
λ_i	5.047	2.897	6.874	0.031	154.089
Z_i	10234.448	10364.573	5759.323	201.152	19997.846
X_i	9.967	10.080	5.749	0.001	19.999

Table 2
Predictive accuracy metrics for out-of-sample observations based on 10-fold cross-validation.

Round	Fixed-parameter model			Random-parameter model					
	MBE[†]	MAE[†]	RMSE[†]	MBE[‡]	MAE[‡]	RMSE[‡]	MBE[§]	MAE[§]	RMSE[§]
1	0.694	4.079	7.092	1.760	3.933	7.303	0.287	4.185	7.061
2	0.677	3.866	7.035	1.774	3.751	7.225	0.337	3.953	7.017
3	−0.019	3.800	6.167	1.332	3.518	6.352	0.061	3.779	6.173
4	0.388	3.846	6.634	1.786	3.624	6.917	0.440	3.831	6.641
5	1.023	3.968	9.432	2.084	3.917	9.660	0.736	4.029	9.395
6	0.387	4.201	7.397	1.798	3.917	7.676	0.386	4.205	7.390
7	0.188	3.852	6.509	1.413	3.566	6.711	−0.086	3.949	6.499
8	0.461	3.843	6.480	1.352	3.732	6.623	−0.243	4.036	6.484
9	−0.479	4.198	7.475	1.231	3.727	7.607	−0.452	4.186	7.478
10	−0.160	3.287	5.346	1.256	3.044	5.548	−0.057	3.244	5.338
Average	0.316	3.894	6.957	1.579	3.673	7.162	0.141	3.940	6.948

MBE, MAE, and RMSE refer to the mean bias error, mean absolute error, and root-mean-square error, which are calculated as $\frac{1}{500}\sum_i(Y_{[i]}-\widehat{Y}_{[i]})$, $\frac{1}{500}\sum_i\left|Y_{[i]}-\widehat{Y}_{[i]}\right|$, and $\sqrt{\frac{1}{500}\sum_i(Y_{[i]}-\widehat{Y}_{[i]})^2}$, respectively. Here, $Y_{[i]}$ and $\widehat{Y}_{[i]}$ represent the observed and predicted number of crashes for validation sample $[i]$, respectively.

[†] $\widehat{Y}_{[i]}=(1+0.5\sigma_u^2)\beta_0 Z_{[i]}^{\beta_1}exp(\beta_2 X_{[i]})$, where β_0, β_1, β_2 and σ_u are the parameters estimated by fixed-parameter model.

[‡] $\widehat{Y}_{[i]}=\overline{\lambda}_{[i]}=\beta_0 Z_{[i]}^{\overline{\beta}_1}exp(\overline{\beta}_2 X_{[i]})$, where β_0, $\overline{\beta}_1$, and $\overline{\beta}_2$ are the parameters estimated by random-parameter model.

[§] $\widehat{Y}_{[i]}=E(\lambda_{[i]})=(1+0.5\sigma_1^2[ln(Z_{[i]})]^2+0.5\sigma_2^2 X_{[i]}^2+0.5\sigma_u^2)\beta_0 Z_{[i]}^{\overline{\beta}_1}exp(\overline{\beta}_2 X_{[i]})$, where β_0, $\overline{\beta}_1$, $\overline{\beta}_2$, σ_1, σ_2, and σ_u are the parameters estimated by random-parameter model.
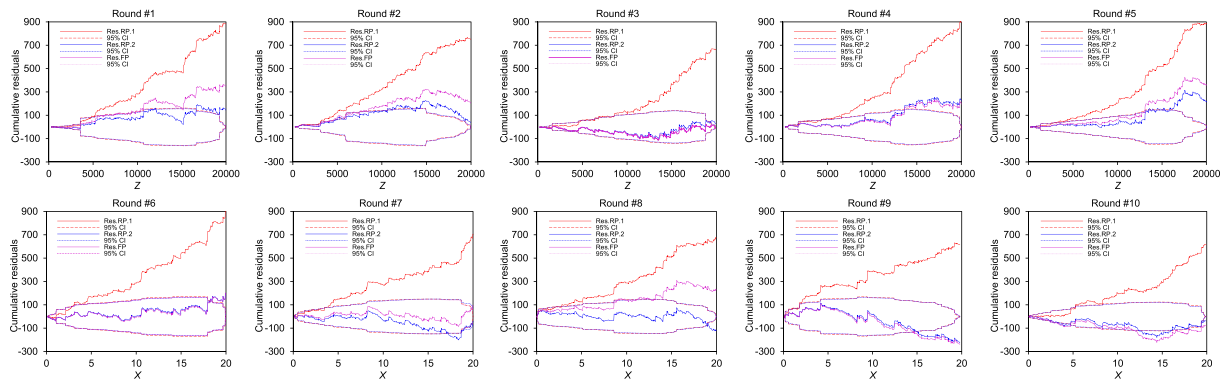
Fig. 1. CURE plots based on 10-fold cross-validation (for fixed-parameter models, Res.FP was computed as $Y_{[i]}-(1+0.5\sigma_u^2)\beta_0 Z_{[i]}^{\beta_1}exp(\beta_2 X_{[i]})$, whereas Res.RP.1 and Res.RP.2 were calculated as $Y_{[i]}-\beta_0 Z_{[i]}^{\overline{\beta}_1}exp(\overline{\beta}_2 X_{[i]})$ and $Y_{[i]}-(1+0.5\sigma_1^2[ln(Z_{[i]})]^2+0.5\sigma_2^2 X_{[i]}^2+0.5\sigma_u^2)\beta_0 Z_{[i]}^{\overline{\beta}_1}exp(\overline{\beta}_2 X_{[i]})$ in random-parameter models, respectively).

absolute values according to the MAE results, $E(\lambda_{[i]})$ was less sensitive and more robust to outliers, given a lower RMSE value. It is also interesting to find that when $E(\lambda_{[i]})$ was used as the predictive value, the predictive performance of the random-parameter model was comparable with that of its fixed-parameter counterpart, as both models produced similar MAE and RMSE values. However, the MBE value of $E(\lambda_{[i]})$ was the closest to zero, suggesting that $E(\lambda_{[i]})$ is least likely to be affected by systematic over- or under-prediction bias. This finding becomes even more plausible when comparing the MBE values of $\overline{\lambda}_{[i]}$ and $E(\lambda_{[i]})$.

More importantly, as illustrated in Fig. 1, when plotting the CURE curves over the entire range of a variable of interest, only $\overline{\lambda}_{[i]}$ consistently exhibited a monotonically increasing trend, resulting in a substantial deviation from zero in the end. By comparison, although the CURE curve for $E(\lambda_{[i]})$ was not always close to zero, it oscillated broadly within the 95% CI. A similar fashion was also observed for the fixed-parameter model.

4. Discussion

A major challenge faced by the random-parameter count models in transferability and extrapolation is the unavailability of unique coefficients for out-of-sample observations. By virtue of the Taylor series expansion, we provided an analytic solution that includes both the

means and variances of the random parameters for unbiased crash prediction. Compared with the simulation method[1] (Alogaili and Mannering, 2020; Islam et al., 2020; Hou et al., 2021), our approach is not only theoretically sound but also readily applicable without much extra computational cost.

Based on a well-designed numerical experiment, we further demonstrated that simply using the means of random parameters to estimate the number of crashes for out-of-sample observations is essentially incorrect, which necessarily results in an underprediction of crash counts in the long run. This finding should be robust, particularly after adjusting for confounders arising from small sample sizes (Lord and Miranda-Moreno, 2008), misspecification of functional forms (Wang et al., 2020), underreporting of crashes (Imprialou and Quddus, 2019; Zeng et al., 2020), excess zeros (Lord et al., 2007; Liu et al., 2018), measurement errors in exposure (Xie et al., 2018; Kamed and Sayed, 2020), omitted variables (Mitra and Washington, 2012), spatial correlation (Xu et al., 2017; Dong et al., 2020; Ziakopoulos and Yannis, 2020; Cui and Xie, 2021), and temporal instability (Mannering, 2018; Alnawmasi and Mannering, 2019; Behnood and Mannering, 2019; Islam et al., 2020), all of which are frequently encountered in empirical studies. Accordingly, the underperformance of random-parameter count models reported in previous studies (Tang et al., 2019; Hou et al., 2020) arises very likely from the convenient use of $\overline{\lambda}_{[i]}$ as the prediction. That is, the conclusion that random-parameter count models were less reliable and inaccurate for out-of-sample crash prediction reached by Tang et al. (2019) and Hou et al. (2020) seems implausible and misleading, which should not be attributed to the modeling technique itself, but rather to the artificial and incorrect use of model estimates by analysts. Given the widespread use and ongoing prevalence of the random-parameter approach in crash analysis, special attention should be paid to this fundamental issue when applying it to predict crash frequencies for out-of-sample observations.

For future research, since the data structures in empirical settings are much more complicated and intractable than anticipated (Mannering et al., 2020), a thorough comparison of the predictive performance of random-parameter count models with that of other techniques, such as the emerging deep learning methods (Bao et al., 2019; Cai et al., 2019), is highly recommended.

Funding

6. Author statement

Pengpeng Xu: Conceptualization, Methodology, Data curation, Software, Visualization, and Writing-original draft; Hanchu Zhou: Writing-review & editing; S.C. Wong: Supervision, Funding acquistion, and Writing-review & editing. We would like to thank Prof. Mannering Fred from University of South Florida and Prof. Huang Helai from Central South University for providing the insightful, thoughtful, and constructive comments that have helped us to substantially improve the quality and readability of our manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alnawmasi, N., Mannering, F., 2019. A statistical assessment of temporal instability in the factors determining motorcyclist injury severities. Anal. Methods Accident Res. 22, 100090.

Alogaili, A., Mannering, F., 2020. Unobserved heterogeneity and the effects of driver nationality on crash injury severities in Saudi Arabia. Accid. Anal. Prev. 144, 105618.

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: a cross-comparison of estimation methods. Accid. Anal. Prev. 93, 147–159.

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

Bao, J., Liu, P., Ukkusuri, S.V., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. Accid. Anal. Prev. 122, 239–254.

Behnood, A., Mannering, F., 2019. Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes. Anal. Methods Accident Res. 23, 100102.

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., Wang, X., 2018. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. Anal. Methods Accident Res. 19, 1–15.

Cai, Q., Abdel-Aty, M., Sun, Y., Lee, J., Yuan, J., 2019. Applying a deep learning approach for transportation safety planning by using high-resolution transportation and land use data. Transport. Res. Part A: Policy and Practice 127, 71–85.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7 (3), 1247–1250.

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

Cui, H., Xie, K., 2021. An accelerated hierarchical Bayesian crash frequency model with accommodation of spatiotemporal interactions. Accid. Anal. Prev. 153, 106018.

Dong, N., Meng, F., Zhang, J., Wong, S.C., Xu, P., 2020. Towards activity-based exposure measures in spatial analysis of pedestrian–motor vehicle crashes. Accid. Anal. Prev. 148, 105777.

El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. Accid. Anal. Prev. 41 (5), 1118–1123.

Elvik, R., Sørensen, M.W.J., Nævestad, T., 2013. Factors influencing safety in a sample of marked pedestrian crossings selected for safety inspections in the city of Oslo. Accid. Anal. Prev. 59, 64–70.

Greene, W., 2016. LIMDEP, Version 11.0. Econometric Software Inc, Plainview, NY.

Hauer, E., 2015. The Art of Regression Modeling in Road Safety. Springer Press, NY.

Hou, Q., Hou, X., Tarko, A.P., Leng, J., 2021. Comparative analysis of alternative random parameters count data models in highway safety. Anal. Methods Accident Res. 30, 100158.

Hou, Q., Tarko, A.P., Meng, X., 2018. Investigating factors of crash frequency with random effects and random parameters models: new insights from Chinese freeway study. Accid. Anal. Prev. 120, 1–20.

Hou, X., Leng, J., Hou, Q., Zheng, L., Zhao, L., 2020. Assessing the explanatory and predictive performance of a random parameters count model with heterogeneity in means and variances. Accid. Anal. Prev. 147, 105759.

Imprialou, M., Quddus, M., 2019. Crash data quality for road safety research: current state and future directions. Accid. Anal. Prev. 130, 84–90.

Islam, M., Alnawmasi, N., Mannering, F., 2020. Unobserved heterogeneity and temporal instability in the analysis of work-zone crash-injury severities. Anal. Methods Accid. Res. 28, 100130.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer Press, NY.

Kamed, M.B., Sayed, T., 2020. Cyclist-vehicle crash modeling with measurement error in traffic exposure. Accid. Anal. Prev. 144, 105612.

Liu, C., Zhao, M., Li, W., Sharma, A., 2018. Multivariate random parameters zero-inflated negative binomial regression for analyzing urban midblock crashes. Anal. Methods Accid. Res. 17, 32–46.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transport. Res. Part A: Policy and Practice 44, 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., 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.

---

[1] For more details, please refer to Hou et al. (2021) who used the Monte Carlo simulation to approximate $E(\lambda_{[i]})$. Similar simulated methods were also applied by Alogaili and Mannering (2020) and Islam et al. (2020) for out-of-sample probability prediction in injury-severity analysis.

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Anal. Methods Accid. Res. 17, 1–13.

Mannering, F., Bhat, C.R., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Anal. Methods Accid. Res. 25, 100113.

Mannering, F.L., Bhat, C., 2014. Analytic methods in accident research: methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22.

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.

Meng, F., Wong, W., Wong, S.C., Pei, X., Li, Y.C., Huang, H., 2017. Gas dynamic analogous exposure approach to interaction intensity in multiple-vehicle crash analysis: case study of crashes involving taxis. Anal. Methods Accid. Res. 16, 90–103.

Milton, J., Shankar, V., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accid. Anal. Prev. 40 (1), 260–266.

Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash modeling. Accid. Anal. Prev. 49, 439–448.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Srinivasan, R., Colety, M., Bahar, G., Crowther, B., Farmen, M., 2016. Estimation of calibration functions for predicting crashes on rural two-lane roads in Arizona. Transp. Res. Rec. 2583, 17–24.

Tang, H., Gayah, V.V., Donnell, E.T., 2019. Evaluating the predictive power of an SPF for two-lane rural roads with random parameters on out-of-sample observations. Accid. Anal. Prev. 132, 105275.

Ukkusuri, S., Hasan, S., Aziz, H.M.A., 2011. Random parameter model used to explain effects of built-environment characteristics on pedestrian crash frequency. Transp. Res. Rec. 2237, 98–106.

Venkataraman, N., Ulfarsson, G.F., Shankar, V.N., 2013. Random parameter models of interstate crash frequencies by severity number of vehicles involved, collision and location type. Accid. Anal. Prev. 59, 309–318.

Wang, K., Zhao, S., Jackson, E., 2020. Investigating exposure measures and functional forms in urban and suburban intersection safety performance functions using generalized negative binomial-P model. Accid. Anal. Prev. 148, 105838.

Wong, W., Wong, S.C., 2019. Unbiased estimation methods of nonlinear transport models based on linearly projected data. Transport. Sci. 53 (3), 665–682.

Xie, S.Q., Dong, N., Wong, S.C., Huang, H., Xu, P., 2018. Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. Accid. Anal. Prev. 121, 285–294.

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16–25.

Xu, P., Huang, H., Dong, N., Wong, S.C., 2017. Revisiting crash spatial heterogeneity: a Bayesian spatially varying coefficients approach. Accid. Anal. Prev. 98, 330–337.

Zeng, Q., Wen, H., Huang, H., Wang, J., Lee, J., 2020. Analysis of crash frequency using a Bayesian underreporting count model with spatial correlation. Physica A 545, 123754.

Ziakopoulos, A., Yannis, G., 2020. A review of spatial approaches in road safety. Accid. Anal. Prev. 135, 105323.