Received May 22, 2020, accepted June 4, 2020, date of publication June 10, 2020, date of current version June 25, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3001165*

# A Zero-Inflated Ordered Probit Model to Analyze Hazmat Truck Drivers' Violation Behavior and Associated Risk Factors

**JINZHONG WU[1,2], WENJI FAN[2], AND WENCHENG WANG[3]**
[1]School of Information Engineering, Chang'an University, Xi'an 710064, China
[2]Research Institute of Highway, Ministry of Transport, Beijing 100088, China
[3]Beijing Municipal Institute of City Planning and Design, Beijing 100045, China

Corresponding authors: Wenji Fan (wj.fan@rioh.cn) and Wencheng Wang (wangwencheng@bjtu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0804800, and in part by the National Natural Science Foundation of China under Grant 91746201.

**ABSTRACT** There are few studies on the violation of truck drivers, especially the hazmat truck driver, although truck driver's violation may cause serious casualties. This paper aims to investigate hazmat truck drivers' violation behavior and identify associated risk factors. Different data sources in intelligent transportation system (ITS) including hazmat transportation management system and traffic safety management system are extracted and emerged together. Three years (2016–2018) of violation data that comprised 11612 trip record in China are employed in this research. Based on Bayesian theory, this study proposes zero-inflated ordered probit (ZIOP) model and three alternative models to exploring the relationship between hazmat truck drivers' violation frequency and the key risk factors. The results show that ZIOP model can handle excessive zero observation problem of violation data properly and differentiate between 'always-zero group' drivers and drivers who did not violate the traffic rules during research period but would do so in different surroundings. The results also indicate that the violation probability and the violation frequency level of hazmat truck drivers are influenced by driver characteristics, freight order attributes, and drivers' violation records. This research provides guidance for driving training and safety education of hazmat truck drivers, and will be helpful in building better driving simulation models.

**INDEX TERMS** Traffic data analysis, road traffic safety, Hazmat truck violation, zero-inflated ordered probit.

## I. INTRODUCTION

Transporting by trucks is more efficient and flexible in terms of time and cost for short distance freight as compared to using modes such as air, railway, or sea. Therefore, trucking plays an irreplaceable role in social and economic development, and truck safety is an area worthy of in-depth study. Although, the number of trucks is often much smaller than the number of cars even in developing countries. For example, according to the latest China statistics available, in 2016 there are 21.72 million civil freight trucks, lower than civil passenger vehicles (162.78 million) [1]. However, in order to ensure the profit of the enterprise, the average annual distance travelled by commercial truck far exceeds that travelled by cars [2]. Greater exposure for each truck might give trucks a greater probability of crash involvement than private cars [2]. Furthermore, trucks are more likely to be involved in a crash that results in fatalities due to the weight and relative size of the vehicle compared to cars, as well as increased length of stopping distances [3]–[6]. For instance, in Korea, crashes involved trucks (15,011) occurred on freeways accounted for approximately 30% of all freeway traffic crashes (48,593) during 2012-2016 [7]. In addition, crashes involving truck are normally associated with greater property damage and more severe injury such as a fatality [4], [8], [9]. According to the statistics provided by Federal Motor Carrier Safety Administration, there were 14 fatalities in large truck crashes per 100 million vehicle miles traveled by large trucks, greater than passenger vehicle crashes in 2014 in the USA (10.5 fatalities) [10]. Particularly, hazmat truck crash can cause

The associate editor coordinating the review of this manuscript and approving it for publication was J. D. Zhao.

disastrous socioeconomic and environmental damages, even if the number of hazmat truck crash is very small in comparison to total traffic crashes.

Drivers' Inappropriate driving behavior is a key factor that have great influence on traffic safety conditions, and according to previous research, over 90% of motor vehicle accidents were more or less related to drivers' violation behavior [11]–[13]. There have been many studies examining the relationship between drivers' violation behavior and crash involvement: non-use of seat belts [14], [15], speeding [9], [13], [16]–[19], drunk driving [8], [19], [20], and other violation behavior [21], [22]. The results of those researches have proven that driver's violation behavior is a key risk factor affecting traffic crash.

Many existing traffic safety studies emphasize the use of crash data to examine the relationship between crash and risk factors. However, these research might lose sight of the importance of proactive safety countermeasures [9]. In recent years, a few researchers have particularly investigated the effects of risk factors on drivers' violation behavior. For example, under the premise of control for the effect of age and gender, González-Iglesias *et al*. [23] explored the contribution of anger-related variables to driver's violations. Tseng [24] analyzed the relationships between self-reported speeding violations and social-economic attributes using logistic regression model. Zhang *et al*. [19] developed stepwise logistic regression to identify significant risk factors associated with speeding and drunk driving. Rosenbloom and Perlman [25] examined whether the presence of other persons in the car have effects on driver's violation tendency. However, the study concerning analysis and prediction of the drivers' violation frequency, in particular, truck drivers' violation frequency is very limited. Hosseinlou *et al*. [26] validated several risk factors that have influences on drivers' violations in freeways and developed zero-truncated Poisson model to quantify the influence of those factors. Akaateba *et al*. [27] investigated the influence of three individual driving attributes on drivers' self-reported attitudes towards the frequency of commission of violations in Kumasi, Ghana. Compared with drivers' violation probability, violation frequency is more appealing because it is more intuitive for traffic safety managers to judge drivers' violation level and to develop targeted countermeasures.

Given the importance of violation frequency, this research aims to analyze truck drivers' violation frequency and identify its key risk factors. Violation frequency is divided into five levels (no violations, 1 violation, 2 violations, 3 violations, and more than 3 violations). To account for the ordinal nature of violation frequency levels, ordered probit models were used in this paper. However, traditional ordered probit model have difficulty fitting the abundance of no violations (zero value) which is a key feature of violation frequency data. To deal with the extensive zero problem, the zero-inflated model was proposed to model violation frequency in this paper. The zero-inflated model was first applied to the field of traffic safety by Shankar *et al* [28]. The combined models of zero-inflated model and traditional models has attracted considerable interest in recent traffic safety studies, including zero-inflated Poisson (ZIP) model [29], [30], zero-inflated negative binomial model [30]–[32], zero-inflated ordered probit (ZIOP) model [33], [34].

Violation data are traditionally obtained from police crash reports, which include information such as date and time, driver information, traffic condition, road geometry, and severity of crash [35]. Most existing analyses of drivers' violation behavior have relied on driving simulators [36], [37], questionnaire surveys [2], [12], [35], [38], [39], and field observation [25]. The prevalence of mobile and GPS technologies have made it possible to collect a massive amount of data [9]. Effective analysis of mobile and GPS data contributes to understanding road safety problems [40], [41]. Unlike data obtained via driving simulators and self-reported questionnaire surveys, most mobile and GPS data are more accurate and practical. Currently, this type of data have yet to be fully utilized in drivers' violation behavior research. To overcome the limitations of traditional data collection approaches, the present research used commercial truck digital tachograph data and police reported violation data collected from a freight transport company in Jiangsu, China in three years (2016-2018). The data used in this research is based on individual drivers, which is helpful to identify different driving features and develop comprehensive road safety measures tailored to the driver population [9].

In this context, the objective of this research is to investigate hazmat truck drivers' violation behavior and associated risk factors by building a zero-inflated ordered probit model at the micro level. Three year period (2016–2018) violation data from 170 drivers, 11612 record in total, are used for the analysis. Three candidate models, i.e. probit model and ordered probit (OP) model were also estimated and compared with the zero-inflated ordered probit model under the Bayesian framework. In summary, there are three main innovations in this paper.

First, many researches have been conducted to explore the violation behavior of car drivers. Few studies focus on the violation of truck drivers, especially the hazmat truck driver, although the crashes involved hazmat truck are often more serious than those of passenger cars due to possible leakage of dangerous goods. To fill this gap, the present research was conducted to analyze the violation behavior of hazmat truck driver.

Second, this paper provides an effective and practical methodology, zero-inflated ordered probit model instead of traditional count model, for investigating drivers' violation hazards and risk factors, and this model can better deal with excessive zero observation problem in violation data.

Third, this paper extracts and merges different data sources from intelligent transportation system (ITS), including hazmat transportation management system (freight order attributes and driver characteristics) and traffic safety management system (driver violation information), which can reflect the driving characteristics of the truck driver group more accurately than the traditional questionnaire data. In addition, the consideration of several key risk factors such as historical violation records in the model is also the advantage of this paper over other articles.

## II. LITERATURE REVIEW

Drivers' violation behavior usually affected by multiple concurrent factors. Prior studies have demonstrated that individual attributes play a critical role in determining driving performance. Drivers' individual factors include age [27], [37], gender [19], [23], [27], [37], educational attainment [24], [27], income level [24], and driving experience [27]. Driving environment have direct influences on driver' violation behavior. Environmental factors include traffic condition [42], road geometry features [26], road type/grades [19], visibility [19], [43], weather condition [43], [44], presence of passenger [25]. Furthermore, in addition to the individual attributes and environment factors, the effects of traffic law and enforcement system cannot be ignored in the analysis of drivers' violation behavior, such as exemption from paying toll [26], Penalty point or license sanctions system [45], [46], and traffic enforcement program [47].

Although driver's violation behavior is a key risk factor affecting traffic crash, it has received little attention concerning analysis and prediction of the truck driver's violation frequency. On the one hand, some existing researches focused on the likelihood or time distribution of car drivers' violation behavior using various approaches such as Logistic regression model [24], proportional hazard model [45]. In order to quantify the effects of variables, such as age, education, income, driving exposure, on the likelihood of drivers' speeding violation, Tseng [24] employed logistic model and explored driver's compliance/or non-compliance in the context of speed limit. Lee *et al.* [45] used Cox's proportional hazard model to estimate different driver's duration of compliance from acquisition of a driving license to conviction or from conviction to reconviction. On the other hand, there are limited number of researches focused on evaluating the effects of risk factors on car drivers' violation behavior frequency. And the methods include One-Way ANOVA analysis [27], count regression model [26], and ordered probit model [48]. One-Way ANOVA and the Bonferroni Post Hoc analysis were applied in the work of Akaateba *et al.* [27] to examine whether there are any differences in the mean drivers' violation frequencies among car drivers with different educational attainments, driving experience and forms of training. Hosseinlou *et al.* [26] employed zero-truncated Poisson model and Poisson model to validate the relationship of risk factors with the number of drivers' violation and to determine the best model for predicting the intended violations. Alver *et al.* [48] developed ordered probit models for hypothetical scenarios to identify the relationship between socio-demographic characteristics, traffic rule violations among young drivers using face-to-face questionnaire data.

As aforementioned, excess zero is a key feature for violation frequency data and crash frequency data. Unlike traditional regression model, zero-inflated can deal with the problem of excess zeros. In traffic safety studies, the zero-inflated model was first used to examine frequencies roadway section accident and marine accident [49], and it was found the zero-inflated structure models are promising and have great flexibility in uncovering processes affecting accident frequencies on roadway sections observed with zero accidents and those with observed accident occurrences [28]. To investigate the effects of curbed outside shoulders on traffic-related injury severity, Jiang *et al.* [33] applied the zero-inflated ordered probit model in their work on the injury severity of single-vehicle crashes. Later, Fountas and Anastasopoulos [34] adopted zero-inflated hierarchical ordered probit approach with correlated disturbances in the analysis of single-vehicle accidents injury-severity. The results of those researches indicated that zero-inflated models were superior to traditional models in terms of goodness of fit or prediction accuracy.

The body of literature review shows no research was conducted to analyze the relationship between hazmat truck drivers' violation frequency and risk factors, which are crucial to reduce the occurrence of road crashes. To the authors' knowledge, this research first employ zero-inflated ordered probit model in this area, which could enhance and provide a more comprehensive understanding in the mechanism of truck drivers' violation risk.

## III. DATA

To illustrate the application of zero-inflated models, the data of commercial freight truck drivers (total of 170 drivers) in Nanjing Sansheng Logistics Company, China, collected from 2016 to 2018, are used.

An initial database were obtained together from different data sources in ITS including hazmat transportation management system (freight order attributes and driver characteristics) and traffic safety management system (driver violation information). Then, the data were supplemented with a data collection effort performed. The original data is freight order information including: order data, order completion time, origin location, destination location, cargo type, cargo weight, and truck drivers' identification (include age, gender, driving experience and so on). The supplemental data are drivers' information including the following: age, gender, driving license type, driving experience and so on. Travel distance (minimum freeway distance between origin location and destination location) was also calculated for each order. Annual violation data for each driver during the period 2016–2018 were also acquired from traffic safety management system. These data included all police-reported violation from all type roads. For each violation behavior, vehicle plate number, report date and time, violation type, and driver identification were all provided. Finally, different data sources were merged according to drivers' identification.

In China, violations are citations for infractions that reduce points to an individual's driving record and include speeding, running at the red-light, and going in the wrong direction and so on while driving. Each driver have 12 points in one cycle (one year), and the violation records will be cleared when a new cycle starts. In general, there is a high probability that the score will be deducted by more than 12 points for drivers who have more than 3 serious violation records in one cycle. So, a violation count indicator was added to designate each driver as a risk-taking driver or conservative driver in 2016 or 2017. Drivers were given a value of 1 for having more than 3 violation record (risk-taking driver) in 2016 or 2017, otherwise (conservative drivers), received a value of 0.

The violation database contains data collected on the police violation report for each police-reported violation occurring during 2016-2018. Order information database and violation database were linked via drivers' identification. In total, 40% of drivers who work for Nanjing Sansheng Logistics Company, China matched to police reported violation record, which means 102 drivers (60%) didn't have violation records in 2018.

According to the number of violations, the present study divides drivers' violation count into five levels coded as 0, 1, 2, 3, 4. The definitions of these five levels are: 0 for "no violations", 1 for "1 violation", 2 for "2 violations", 3 for "3 violations", and 4 for "more than 3 violations".

Descriptive summary statistics are provided for the analysis database in Table 1. Variances of all three years' violation count were larger than their means, as shown in Table 1, which implies that over dispersion existed for drivers' violation count. Combined with Fig. 1, we can conclude that excess zeros existed for drivers' violation counts in 2016, 2017, and 2018.

## IV. METHODS

### A. POISSON MODEL

Statistically, driver violation behavior is rationally assumed to be a Poisson process. The model is specified as follows

$$Y_i|\lambda_i \sim Poisson(\lambda_i) \quad (1)$$
$$\log(\lambda_i) = \beta_0 + \sum\nolimits_k \beta_k x_{ik} \quad (2)$$

where $Y_i$ is the observed violation behavior count of driver $i$ during research period, which is Poisson distributed with the underlying mean $\lambda_i$. In the present research, $\lambda_i$ denotes the expected violation frequency for driver $i$. $\beta_0$ is the intercept. $x_{ik}$ and $\beta_k$ denote the $k$th risk factor value of driver $i$. and the coefficient of covariate $k$, respectively.

### B. ZERO-INFLATED POISSON MODEL

The prime justification for the use of zero-inflated models has rested on improved statistical fit compared with traditional Poisson model. Zero-inflated models provides improved fit for modeling violation data characterized by a preponderance of zeros.

The ZIP model is a two-component mixture model consisting of a Bernoulli distribution at zero mixed with a Poisson distribution [50]. For discrete random responses $Y_i$ which are independent identically distributed, the zero-inflated Poisson model is given by

$$P(Y_i = 0) = p_i + (1 - p_i)e^{-\lambda_i}, \quad 0 \leq p \leq 1 \quad (3)$$
$$P(Y_i = y_i) = (1 - p_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad y_i = 1, 2, \ldots, 0 < \lambda_i < \infty \quad (4)$$

where $Y_i$ denotes the violation behavior count for driver $i = 1, \ldots, n$. Suppose that drivers with zeros violation behavior are divided into two categories corresponding to distinct underlying states. The first category occurs with probability p, while the other category occurs with probability 1-p and follows a standard Poisson distribution with mean $\lambda_i$ [51]. In drivers' violation behavior research, p is known as the zero violation probability—i.e. the probability of having no violation behavior. Typically, one assumes that p is strictly between 0 and 1, so that all subjects have a non-zero probability of violation and are therefore considered 'potential' violators even if they do not actually have violation behavior during the study period. The parameter $\lambda_i$ is the mean of Poisson distribution measuring the frequency of violations; as $\lambda_i$ increases, the average number of violations among drivers also increases.

Because the 0s and non-zero counts are modelled uniquely, the zero-inflated Poisson model can accommodate both the large proportion of 0s and a right-skewed distribution for the positive counts. By comparison, a standard Poisson regression would have to compromise between these two competing features of the data, since the large proportion of 0s would tend to lower the Poisson mean whereas large non-zero values would tend to increase it.

The mean and variance under the zero-inflated Poisson model is given by

$$E(Y_i) = (1 - p_i)\lambda_i \quad (5)$$
$$Var(Y_i) = (1 - p_i)\lambda_i + p_i(1 - \omega)\lambda_i^2 \quad (6)$$

The link functions are given by

$$\text{logit}(p_i) = \log(\frac{p_i}{1 - p_i}) = \gamma_0 + \sum\nolimits_k \gamma_k z_{ik} \quad (7)$$
$$\log(\lambda_i) = \beta_0 + \sum\nolimits_k \beta_k x_{ik} \quad (8)$$

where $\gamma_k$ and $\beta_k$ are the coefficients for covariates $z_{ik}$ and $x_{ik}$ under the zero state and Poisson process, respectively. Zero-inflated models allow different sets of predictors in the zero state and Poisson process. In the present research, only two variable (Violation count indicator of 2016 and 2017) are selected as covariates for the zero state.

### C. ZERO-INFLATED ORDERED PROBIT MODEL

The prime justification for the use of zero-inflated models has rested on improved statistical fit compared with standard

**TABLE 1.** Descriptive statistics of key variables.

| Variable | Description | Mean | Max. | Min. | Std. err. |
|---|---|---|---|---|---|
| Viola_18 | Police reported violation count for the driver in 2018 | 0.84 | 13 | 0 | 1.52 |
| N_order | Number of order for the driver in 2018 | 68.31 | 283 | 1 | 59.40 |
| Ln_n_order | Logarithm of number of order for the driver in 2018 | 3.58 | 5.65 | 0 | 1.46 |
| W_cargo | Average cargo weight for each order | 28.00 | 31.52 | 0 | 3.86 |
| Ln_trav_dist | Logarithm of total travel distance for the driver in 2018 | 9.94 | 12.93 | 4.08 | 1.70 |
| Driv_expe | Driving experience for driver(years) | 18.35 | 35 | 8 | 5.92 |
| Age | Driver's age | 44.80 | 58 | 30 | 5.65 |
| Licen_type | Type of driver's license (1 driver is allowed to drive the bus, 0 driver is not allowed to drive the bus) | 0.13 | 1 | 0 | 0.34 |
| Licen_agen | Certification authority of driver's license indicator (1 if local traffic management bureau, 0 otherwise) | 0.79 | 1 | 0 | 0.41 |
| Viola_16 | Police reported violation count in 2016 | 1.15 | 11 | 0 | 2.39 |
| Viola_17 | Police reported violation count in 2017 | 1.87 | 17 | 0 | 3.19 |
| Ind_viola_16 | Violation count indicator of 2016 (1 if driver had 4 or more violation records in 2016, 0 otherwise) | 0.18 | 1 | 0 | 0.38 |
| Ind_viola_17 | Violation count indicator of 2017 (1 if driver had 4 or more violation records in 2017, 0 otherwise) | 0.28 | 1 | 0 | 0.45 |

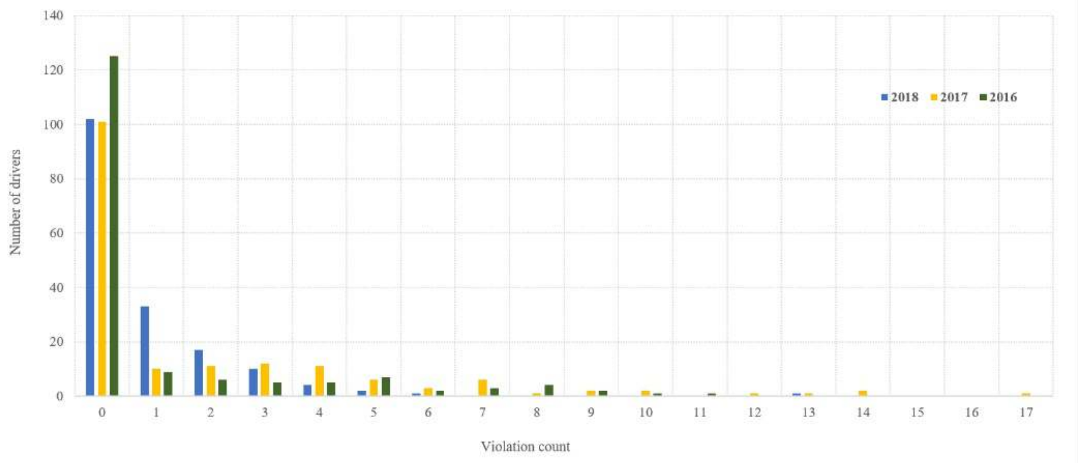


**FIGURE 1.** Frequency distribution of violation counts in 2016, 2017, and 2018.

OP model, which provides improved fit for modeling drivers' violation data characterized by a preponderance of zeros.

Consider a discrete ordered response variable $y_i$ denoting the observed violation level of driver $i$ during research period. As aforementioned, $y_i$ have five levels coded as 0, 1, 2, 3, 4. The statistical description in Table 1 shows that the lowest violation level, zero (drivers without violation), is inflated. As zero-inflated Poisson model, we suspect that these drivers with no violation belong to one of two groups in the ZIOP model. Individuals in the first group have a strong sense of safety (always-zero group) who almost have no violations in his driving career and will never active violation. Individuals in the second group are drivers with no violation only during research period, or drivers didn't have violation because of monetary punishment and who may have violation, say, if the punishment system is cancelled or their income increases. To some extent, these two types of zeros are drove by different patterns of driving psychology and a ZIOP model is a good alternative to handle excessive zero in this case.

The zero-inflated ordered probit model is a two-component mixture model consisting of a binary probit regression at zero mixed with an ordered probit regression [52], [53]. The part of binary probit regression is used to identify the violation-prone state from the non-violation state. And the ordered probit regression is used to model the level of violation count in the violation-prone state. Let $s_i = 1$ if the $i$th driver belongs to the violation group or let $s_i = 0$ otherwise. $s_i$ is related to a latent variable $s_i^*$ through the criteria: $s_i = 0$ for $s_i^* < 0$ and $s_i = 1$ for $s_i^* > 0$. The latent variable $s^*$ represents the propensity of drivers' violation and is given by

$$s_i^* = \boldsymbol{x}_i^T \beta + \varepsilon_i \quad i = 1, 2, \ldots, n \tag{9}$$

where $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ij})^{\mathrm{T}}$ represents covariates that determines drivers' violation propensity, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_j)^{\mathrm{T}}$ is the corresponding vector of coefficients to be estimated. The error term $\varepsilon_i$ are independent identically distributed following a standard normal distribution. The probability of a driver having violation is given by Maddala [54].

$$\Pr(s_i = 1|\boldsymbol{x}_i) = \Pr(s_i^* > 0|\boldsymbol{x}_i) = \Phi(\boldsymbol{x}_i^T \beta) \tag{10}$$

$\Phi(\cdot)$ is the standard normal distribution function. Next, conditioning on $s_i = 1$, the observed violation levels $\tilde{y}_i$ are modeled using an ordered probit model; these levels may also include 0. $\tilde{y}_i$ can be connected to a latent variable $y_i^*$ through a function $g(y_i^*)$. $y_i^*$ is given by

$$y_i^* = z_i^T \gamma + \theta_i \quad i = 1, 2, \ldots, n \tag{11}$$

where $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{ik})^{\mathrm{T}}$, covariates in the ordered probit regression process, could be different from $x_i$; $\gamma = (\gamma_1, \ldots, \gamma_k)^{\mathrm{T}}$ is associated vector of parameters to be estimated; $\theta_i$ is a random error term following an independent and identically distributed standard normal distribution. The mapping between $\tilde{y}_i$ and $y_i^*$ is obtained by

$$y_i = g(y_i^*) = \begin{cases} 0 & \text{if } y_i^* \leq u_0 \\ 1 & \text{if } u_0 < y_i^* \leq u_1 \\ 2 & \text{if } u_1 < y_i^* \leq u_2 \\ 3 & \text{if } u_2 < y_i^* \leq u_3 \\ 4 & \text{if } u_3 < y_i^* \end{cases} \tag{12}$$

where $u_0$, $u_1$, $u_2$, $u_3$ are the threshold values for all five violation levels which need to be estimated in addition to the coefficients vector $\boldsymbol{\beta}$. In the present research, we assume that $u_0 = 0$. Conditional on $s_i = 1$, the probability of five violation levels in the ordered probit regression process is expressed as follow

$$\Pr(\tilde{y}_i) = \begin{cases} \Pr(\tilde{y}_i = 0) = \Phi(-z_i^T \gamma) \\ \Pr(\tilde{y}_i = 1) = \Phi(u_1 - z_i^T \gamma) - \Phi(-z_i^T \gamma) \\ \Pr(\tilde{y}_i = 2) = \Phi(u_2 - z_i^T \gamma) - \Phi(u_1 - z_i^T \gamma) \\ \Pr(\tilde{y}_i = 3) = \Phi(u_3 - z_i^T \gamma) - \Phi(u_2 - z_i^T \gamma) \\ \Pr(\tilde{y}_i = 4) = 1 - \Phi(u_3 - z_i^T \gamma) \end{cases} \tag{13}$$

Note that $s_i$ and $\tilde{y}_i$ are both unobservable in terms of the zeros. The observed response variable is $y_i = s_i \times \tilde{y}_i$. Thus, the zero observation occurs when $s_i = 0$ (the driver has a strong sense of safety, belongs to always-zero group) or occurs when $s_i = 1$ and $\tilde{y}_i = 0$ (the driver has no violation only during research period). To observe a positive $y_i$, it is a joint requirement that $s_i = 1$ and $\tilde{y}_i > 0$. This can be illustrated by

$$y_i = s_i \times \tilde{y}_i = \begin{cases} 0 & \text{if } s_i = 0 \text{ or } \tilde{y}_i = 0 \\ 1 & \text{if } s_i = 1 \text{ and } \tilde{y}_i = 1 \\ 2 & \text{if } s_i = 1 \text{ and } \tilde{y}_i = 2 \\ 3 & \text{if } s_i = 1 \text{ and } \tilde{y}_i = 3 \\ 4 & \text{if } s_i = 1 \text{ and } \tilde{y}_i = 4 \end{cases}$$

$$= \begin{cases} 0 & \text{if } s_i = 0 \text{ or } y_i^* \leq u_0 \\ 1 & \text{if } s_i = 1 \text{ and } u_0 < y_i^* \leq u_1 \\ 2 & \text{if } s_i = 1 \text{ and } u_0 < y_i^* \leq u_1 \\ 3 & \text{if } s_i = 1 \text{ and } u_0 < y_i^* \leq u_1 \\ 4 & \text{if } s_i = 1 \text{ and } u_3 < y_i^* \end{cases} \tag{14}$$

The distribution of $y_i$ is given by

$$\Pr(y_i) = \begin{cases} \Pr(y_i = 0|\boldsymbol{x}_i, z_i) \\ \Pr(y_i = 1|\boldsymbol{x}_i, z_i) \\ \Pr(y_i = 2|\boldsymbol{x}_i, z_i) \\ \Pr(y_i = 3|\boldsymbol{x}_i, z_i) \\ \Pr(y_i = 4|\boldsymbol{x}_i, z_i) \end{cases}$$

$$= \begin{cases} \Pr(s_i = 0|\boldsymbol{x}_i) + \Pr(s_i = 1|x_i) \\ \qquad \Pr(\tilde{y}_i = 0|z_i, s_i = 1) \\ \Pr(s_i = 1|x_i)\Pr(\tilde{y}_i = 1|z_i, s_i = 1) \\ \Pr(s_i = 1|x_i)\Pr(\tilde{y}_i = 2|z_i, s_i = 1) \\ \Pr(s_i = 1|x_i)\Pr(\tilde{y}_i = 3|z_i, s_i = 1) \\ \Pr(s_i = 1|x_i)\Pr(\tilde{y}_i = 4|z_i, s_i = 1) \end{cases} \tag{15}$$

Substituting (10) and (13) in (15), we get

$$\Pr(y_i) = \begin{cases} \Pr(y_i = 0|\boldsymbol{x}_i, z_i) \\ \Pr(y_i = 1|\boldsymbol{x}_i, z_i) \\ \Pr(y_i = 2|\boldsymbol{x}_i, z_i) \\ \Pr(y_i = 3|\boldsymbol{x}_i, z_i) \\ \Pr(y_i = 4|\boldsymbol{x}_i, z_i) \end{cases}$$

$$= \begin{cases} [1 - \Phi(\boldsymbol{x}_i^T \beta)] + \Phi(\boldsymbol{x}_i^T \beta)\Phi(-z_i^T \gamma) \\ \Phi(\boldsymbol{x}_i^T \beta)[\Phi(u_1 - z_i^T \gamma) - \Phi(-z_i^T \gamma)] \\ \Phi(\boldsymbol{x}_i^T \beta)[\Phi(u_2 - z_i^T \gamma) - \Phi(u_1 - z_i^T \gamma)] \\ \Phi(\boldsymbol{x}_i^T \beta)[\Phi(u_3 - z_i^T \gamma) - \Phi(u_2 - z_i^T \gamma)] \\ \Phi(\boldsymbol{x}_i^T \beta)[1 - \Phi(u_3 - z_i^T \gamma)] \end{cases} \tag{16}$$

Because zero outcomes are the sum of zero from the ordered probit model and zero from the probit model, the zero-inflated ordered probit model can accommodate the large proportion of zeros.

### D. MODEL ESTIMATION

Four models (Poisson model, ZIP model, OP model, and ZIOP model) are estimated using the Bayesian theory, and Markov chain Monte Carlo (MCMC) is applied to sample the posterior distribution of parameters. Estimations of the parameters can be obtained by specifying prior distributions for those parameters. Without credible prior information, uninformative priors were assumed for parameters to be estimated which are normally distributed.

The marginal effects of ZIOP model are estimated using the maximum likelihood method. The log-likelihood function is

$$\ln L = \sum\nolimits_{i=1}^{N} w_i \sum\nolimits_{h=0}^{H} I(y_i = h) \ln[\Pr(y_i = h|\boldsymbol{x}_i, z_i)] \tag{17}$$

where weight of the $i$th driver $w_i$ is optional and the indicator function $I(y_i = h)$ is given by

$$I(y_j = h) = \begin{cases} 1 & \text{if } y_j = h \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

As with conventional ordered probit models, the estimated coefficients of the ZIOP model are not particularly informative [55]. In order to evaluate the amount of change in the probability of certain violation level with the change in one of the variables, marginal effects were computed in the present research. For more information about marginal effects, the readers are referred to the works of Jiang *et al.* [33], Harris and Zhao [53], and Downward *et al.* [55].

### E. MODEL COMPARISON

#### 1) VUONG'S TEST

The choice between the zero-inflated models and the conventional models cannot be made using a likelihood-ratio test because the two hypotheses are not nested in the usual sense of parameter restrictions. However, the Vuong's test is commonly used to compare the fitness of non-nested models [52]. So, the Vuong's test was

$$V = \frac{\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} m_i)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \overline{m})^2}} \tag{19}$$

$$m_i = \log(\frac{p_1(y_i)}{p_2(y_i)}) \tag{20}$$

where $p_1$ and $p_2$ are the estimated probabilities of the level $yi$ using model 1 (zero-inflated models) and 2 (conventional models), respectively; and n is the sample size. If $-1.96 < V < 1.96$, the two models do not have significantly different. If $V > 1.96$, the model 1 is preferred, and if $V < 1.96$, model 2 has a better model fitness [33], [56].

#### 2) DIC

The deviance information criterion (DIC) [57] can be used to determine the optimal model. The DIC is considered the Bayesian equivalent of the Akaike Information Criterion (AIC). The DIC is defined as

$$\text{DIC} = \overline{D}(\boldsymbol{\theta}) + p_D \tag{21}$$

where

$$\overline{D}(\boldsymbol{\theta}) = \text{E}[D(\boldsymbol{\theta})|\boldsymbol{y}] \tag{22}$$

where $D(\theta)$ is the Bayesian deviance of the estimated parameter $\theta$. $\bar{D}(\boldsymbol{\theta})$ can also be taken as a measure of model fitting. $p_D$ is the effective number of parameters and indicates complexity of models. As a rule of thumb, if two models differ in DIC by seven or more, the one with the smaller DIC is preferred [57].

## V. RESULTS

### A. PARAMETER ESTIMATES

This research established the ZIOP model of hazmat truck driver's violation and influencing factors. For comparison, a standard OP model, Poisson model, ZIP model with the same variables were also estimated.

Intuitively, there is a certain correlation between Driv_expe and Age, Viola_16 and Viola_17, Ln_n_order and Ln_trav_dist. To test the correlation and multi-collinearity among those risk factors, Pearson test is conducted. The results show that Driv_expe and Age (0.525), Viola_16 and Viola_17 (0.527), Ln_n_order and Ln_trav_dist (0.855) are significantly correlated with p-value lower than 0.01. To reduce the model complexity, Driv_expe, Viola_17, and Ln_trav_dist are excluded from the models. Ind_viola_17 is selected as the only Explanatory variable in the binary probit regression for ZIOP. The result of each model is presented in Table 2.

Table 2 shows that the Vuong's test value of ZIP model and ZIOP model are 2.17, 4.95 and 2.36 respectively, which are much greater than the critical value 1.96, indicating that zero-inflated models are necessary for the data in this research. The results show that DIC values of ZIP model (339.4) and ZIOP model (306.0) are significantly less than standard Poisson model (360.4) and OP model (331.6) respectively, also suggesting that the zero-inflated models outperformed the standard models. Furthermore, the ZIOP model has the lowest DIC value among the four models indicating that the ZIOP model is superior. Considering the results of Vuong's test and DIC, the ZIOP model appears to offer a clear improvement in the overall fit performance for drivers' violation in comparison with other alternative models.

Table 2 also lists parameter estimates of the ZIOP model and other 5 models. Positive parameter estimates indicate that a particular cluster of a variable leads to more violation number or a higher violation level, and vice versa. Since the fitting performance of ZIOP model is optimal, the following parameter interpretations are mainly based on the results of ZIOP model.

The result reveals significant effects of W_cargo on driver's violation level with positive coefficient at 95% confidence interval, which indicates that drivers were more likely to have violation behavior with increasing W_cargo. Compared with light truck, heavy truck tend to be less maneuverable because of its multi-unit configurations, large sizes, and high centers of gravity. The fact that heavy trucks require wide roads and large radii of path curvature for evasive maneuvers and frequently display unstable motion modes [58] supports this argument. Therefore, heavy truck drivers cannot make timely corrective actions in case of emergency, and they are more likely to have violation behaviors.

Logarithm of number of order (Ln_n_order) is statistically significant and positively associated with violation level. This is intuitive as number of order is regarded as the main risk exposure to drivers' violation. The more orders a truck driver delivered, the greater the likelihood of violation behavior for him/her.

Viola_16 is remarkably positive, which means the probability of driver's violation in present is consistent with the number of his violations in the past. And a higher number of

**TABLE 2. The results of Parameter estimates and model performance for four models.**

| | Poisson | | | OP | | | ZIP | | | ZIOP | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | [95% Cred. Interval] | | Mean | [95% Cred. Interval] | | Mean | [95% Cred. Interval] | | Mean | [95% Cred. Interval] | |
| W_cargo | 0.22 | 0.01 | 0.50 | 0.21 | 0.10 | 0.33 | 0.26 | 0.10 | 0.39 | 0.28 | 0.18 | 0.39 |
| Ln_n_order | 0.65 | 0.41 | 0.91 | 0.50 | 0.30 | 0.72 | 0.63 | 0.35 | 0.93 | 1.00 | 0.59 | 1.42 |
| Viola_16 | 0.22 | 0.17 | 0.26 | 0.26 | 0.18 | 0.34 | 0.15 | 0.10 | 0.20 | 0.20 | 0.11 | 0.30 |
| Age | -0.03 | -0.06 | 0.00 | -0.02 | -0.06 | 0.01 | -0.03 | -0.06 | 0.01 | -0.06 | -0.11 | 0.01 |
| Licen_type | 0.46 | -0.05 | 0.94 | 0.48 | -0.05 | 1.00 | 0.42 | -0.12 | 0.94 | 1.78 | 1.10 | 2.69 |
| Licen_agen | -0.85 | -1.24 | -0.45 | -0.55 | -1.03 | -0.10 | -0.65 | -1.02 | -0.24 | -0.78 | -1.35 | -0.25 |
| Constant | -7.70 | -16.07 | -1.26 | - | - | - | -8.47 | -12.31 | -3.98 | - | - | - |
| **Inflate** | | | | | | | | | | | | |
| Ind_viola_17 | - | - | - | - | - | - | -82.74 | -239.65 | -4.20 | 1.66 | 1.17 | 2.08 |
| Constant | - | - | - | - | - | - | 0.02 | -0.76 | 0.68 | -0.33 | -0.63 | 0.05 |
| cut1 | - | - | - | 7.17 | 4.43 | 10.16 | - | - | - | 7.84 | 6.30 | 9.70 |
| cut2 | - | - | - | 7.98 | 5.28 | 10.94 | - | - | - | 9.84 | 7.89 | 12.30 |
| cut3 | - | - | - | 8.60 | 5.99 | 11.58 | - | - | - | 10.68 | 8.72 | 13.06 |
| cut4 | - | - | - | 9.24 | 6.73 | 12.16 | - | - | - | 11.50 | 9.48 | 13.82 |
| Vuong's test | | - | | | - | | | 2.17 | | | 2.36 | |
| Log-likelihood | | -223.82858 | | | -205.92826 | | | -218.44973 | | | -204.98955 | |
| DIC | | 360.4006 | | | 331.6435 | | | 339.4348 | | | 305.9825 | |

violation record is more likely to be associated with a higher level of violation. In general, drivers' violation record can reflect their driving habits. Drivers with more violation record are more aggressive, and they are more likely to have traffic violations in the future.

The effect of Age on the drivers' violation behavior is significantly negative, which means older drivers have fewer traffic violations than younger. This may be due to the rich driving experience and conservative driving habits of the elderly. The results of previous studies that young cyclists and e-cyclists are more likely to run a red light [59]–[62] supports this argument. This is also consistent with intuition that older people act more cautiously than younger ones.

The positive and statistically significant Licen_type variable shows that drivers qualified to drive buses are more likely to have violation behavior than drivers not qualified to drive buses. These effect patterns may be attributed to different training strategies for truck drivers and bus drivers in China. For the sake of safety, bus drivers undergo more driving trainings and responsibility educations than truck drivers, which makes bus drivers more confident in their driving ability and more aggressive when driving, which leads to a higher probability of traffic violations.

The negative and statistically significant Licen_agen variable shows that drivers with license managed by Nanjing traffic police are less likely to have violation behavior. The requirements of punishments to drivers' violation behavior and the implementation of the national road traffic safety law in different place are different. Drivers with license managed by Nanjing traffic police are familiar with the local traffic rules and can avoid traffic violations at the black spots. Therefore, local drivers have less traffic violations.

**TABLE 3. Marginal effects of Ind_viola_17 in the binary probit regression process.**

| | dy/dx | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| $\Pr(y_i = 0, s_i = 1)$ | 0.186 | 0.064 | 2.91 | 0.004 | 0.061 | 0.312 |
| $\Pr(s_i = 0)$ | -0.573 | 0.080 | -7.13 | 0.000 | -0.731 | -0.416 |

In the binary probability process, significantly positive parameter of Ind_viola_17 indicates that risk-taking drivers (having more than 3 violation record in 2017) are more likely to violate the traffic rules in 2018 than conservative drivers (having no more than 3 violation record in 2017). This result is in line with the effect of Viola_16 in ordered probit regression process. Therefore, to some extent, the truck drivers' historical violation records can reflect their driving habits in the future.

### B. MARGINAL EFFECTS

The estimation results of OP model cannot be interpreted directly because of its nonlinear characteristics. In order to quantify the effect of each factor more accurately, this paper estimates the marginal effects of each parameter based on the moment estimation method. Table 3 lists the marginal effects of Ind_viola_17 on $P(y = 0)$ in the binary probit regression process, and this paper decompose the overall marginal effect into two parts: the marginal effect on non-violation state, $P(s = 0)$, and the effect on violation-prone state,

**TABLE 4. Marginal effects of each factor in ordered probit regression process.**

| Variable | y | dy/dx | Std. Err. | z | P>\|z\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|---|
| W_cargo | | | | | | | |
| | 0 | -0.022 | 0.014 | -1.55 | 0.122 | -0.049 | 0.006 |
| | 1 | -0.018 | 0.016 | -1.13 | 0.258 | -0.049 | 0.013 |
| | 2 | 0.009 | 0.006 | 1.50 | 0.134 | -0.003 | 0.021 |
| | 3 | 0.013 | 0.009 | 1.43 | 0.152 | -0.005 | 0.030 |
| | 4 | 0.018 | 0.013 | 1.42 | 0.156 | -0.007 | 0.043 |
| Ln_n_order | | | | | | | |
| | 0 | -0.064 | 0.025 | -2.56 | 0.011 | -0.114 | -0.015 |
| | 1 | -0.054 | 0.029 | -1.87 | 0.061 | -0.109 | 0.002 |
| | 2 | 0.027 | 0.010 | 2.81 | 0.005 | 0.008 | 0.047 |
| | 3 | 0.038 | 0.013 | 2.95 | 0.003 | 0.013 | 0.062 |
| | 4 | 0.053 | 0.017 | 3.14 | 0.002 | 0.020 | 0.086 |
| Viola_16 | | | | | | | |
| | 0 | -0.013 | 0.008 | -1.66 | 0.098 | -0.028 | 0.002 |
| | 1 | -0.011 | 0.004 | -2.90 | 0.004 | -0.018 | -0.003 |
| | 2 | 0.006 | 0.003 | 1.94 | 0.052 | -0.000 | 0.011 |
| | 3 | 0.008 | 0.003 | 2.70 | 0.007 | 0.002 | 0.013 |
| | 4 | 0.011 | 0.003 | 3.09 | 0.002 | 0.004 | 0.017 |
| Age | | | | | | | |
| | 0 | 0.003 | 0.002 | 1.44 | 0.151 | -0.001 | 0.008 |
| | 1 | 0.003 | 0.002 | 1.42 | 0.157 | -0.001 | 0.006 |
| | 2 | -0.001 | 0.001 | -1.52 | 0.130 | -0.003 | 0.000 |
| | 3 | -0.002 | 0.001 | -1.62 | 0.105 | -0.004 | 0.000 |
| | 4 | -0.003 | 0.002 | -1.64 | 0.101 | -0.006 | 0.001 |
| Licen_type | | | | | | | |
| | 0 | -0.121 | 0.063 | -1.93 | 0.054 | -0.244 | 0.002 |
| | 1 | -0.101 | 0.048 | -2.07 | 0.038 | -0.196 | -0.006 |
| | 2 | 0.052 | 0.027 | 1.94 | 0.053 | -0.001 | 0.104 |
| | 3 | 0.071 | 0.026 | 2.67 | 0.008 | 0.019 | 0.122 |
| | 4 | 0.100 | 0.031 | 3.17 | 0.002 | 0.038 | 0.161 |
| Licen_agen | | | | | | | |
| | 0 | 0.041 | 0.032 | 1.29 | 0.196 | -0.021 | 0.104 |
| | 1 | 0.034 | 0.022 | 1.54 | 0.125 | -0.009 | 0.078 |
| | 2 | -0.018 | 0.013 | -1.40 | 0.163 | -0.042 | 0.007 |
| | 3 | -0.024 | 0.016 | -1.52 | 0.128 | -0.055 | 0.007 |
| | 4 | -0.034 | 0.020 | -1.67 | 0.095 | -0.074 | 0.006 |

$P(y = 0, s = 1)$. Table 4 present marginal effects of each factor on the unconditional probabilities of all five levels of violation in ordered probit regression process.

Table 3 reveals to us that driver with violation record in 2017 as opposed to driver without violation record during that period, were associated with 57.3% decrease in the probability of non-violation state in 2018, but 18.6% increase in the joint probability of 0 violation and violation-prone state. The former marginal effect shows drivers that have violation record in 2017 were more likely to be non-violation state in 2018 as compared to drivers without violation record in 2017. The latter marginal effect indicates that drivers that have violation record in 2017 were associated with a relatively higher likelihood of having violation behavior in 2018 for drivers in violation-prone state. As illustrated above, the ZIOP model assumes that zero violation comes from two distinct sources:

non-violation state and violation-prone state, so the comprehensive marginal effect of Ind_viola_17 should combine its opposing effects in the two state, which can result in some offsetting.

Table 4 shows that Ln_n_order is not significant for y = 1 in the ordered probit process, but is significantly associated with other four violation levels. One unit increase in Ln_n_order results in a 6.4% fall in the probability of having 0 violation, but 3.8% and 5.3% rises in the probability of being 3 and 4 violation levels respectively.

Viola_16 and Licen_type have similar effects that they are not significant for y = 1 in the ordered probit process, but are significantly associated with other four violation levels. An increase of one unit in Viola_16 brought about a 1.3% reduction in the probability of having 1 violation, but 0.8% and 1.1% increases in the probability of being 3 and 4 violation levels respectively. Drivers qualified to drive buses were associated with 10.1% decrease in the probability of having only one violation behavior in 2018, but 7.1% and 10.0% increases in the probability of being 3 and 4 violation levels respectively, when compared to drivers not qualified to drive buses.

Among other factors, the increase of W_cargo, Age and Licen_agen are not significant in violation-prone state for all five levels, but are significantly associated with higher, lower and lower violation level in the Bayesian estimation results. This finding highlights interesting differences between traditional method and Bayesian method from alternative models for some of the explanatory factors, and deserves further study.

## VI. CONCLUSION

As many previous researches have been conducted to explore the violation behavior of truck drivers, few similar studies focus on the driver of hazmat truck. To fill this gap, the present research was conducted to analyze hazmat truck drivers' violation behavior and associated risk factors. Using three years (2016–2018) of violation data that comprised 11612 trip record from 170 drivers in China, this study proposes ZIOP model to exploring the relationship between truck drivers' violation frequency and the key risk factors under the Bayesian framework, and the other three models are also developed as comparison. The ZIOP model is able to handle excessive zero observation problem and differentiate between 'always-zero group' and drivers who did not have violation during research period but would do so if the surroundings and conditions were different. The results show that the decision to violate or not and the frequency of violation for truck drivers are driven by different factors.

Factors that have statistically significant influences on truck drivers' violation include:

- Higher cargo weight, the number of order carried by driver and the number of drivers violation are associate with higher frequency of violation given that truck drivers fall in the violation-prone state;
- Young drivers were more likely to violate the traffic rule as opposed to old drivers conditional on being the violation-prone state.
- Truck drivers qualified to drive buses as opposed to drivers not qualified to drive buses were associated with significantly higher likelihood of high frequency violations. On the contrary, local drivers are less likely to have violation behavior than non-local drivers.
- For the binary probability process, drivers with violation record in 2017 were more likely violate the traffic law in 2018 as opposed to drivers without violation record in 2017.

The present research provides a methodological understanding that zero-inflated models (ZIP and ZIOP) are more suitable to analyze hazmat truck drivers' violation data than traditional models (Poisson and OP). The best fitting performance of ZIOP indicates that selecting an appropriate model from probability models and count models is as important as dealing with the excessive zero observations problem in terms of accurately fit driving behaviors. The application of ZIOP model in truck driver' violations will be helpful to build better driving simulation models for ITS [63], [64].

In terms of the empirical contributions, the current findings might be helpful to propose several specific interventions to improve traffic safety. First, the high frequency of violation among truck drivers with violation records suggests the need for an increase in penalties and stricter enforcement for violations of truck driver frequently violate the traffic rules. A system like credit investigation system might be imposed to enhance the enforcement and promote those truck drivers' responsibility for safe driving. Second, safety education and other intervention programs did not achieve expected effects, and traffic safety management department should ameliorate the existing education methods to improve road safety since this research found that truck drivers qualified to drive buses are more likely to have violation behavior than drivers not qualified to drive buses. Third, traffic police can provide more travel information to nonlocal truck drivers through intelligent devices, so as to help them get familiar with local traffic conditions more quickly.

Despite its several merits, the study also has two primary limitations. First, in this study, due to time and cost constrains, data were collected in Nanjing city, China and its surrounding areas. The results may not be exactly the same in other cities or countries due to the difference in culture, driving behaviors, etc. Because of differences in culture or driving habits of drivers, the modeling results may be different in different countries or cities. The outcomes of this paper still have reference significance, and the experimental design method and modeling process can also be used in other researches. Second, there are many factors that may affect truck drivers' violations, but some key factors, such as drivers' gender, education psychological reaction, and environmental conditions are not considered in this paper because of data constraints. The lack of these factors greatly limits the practical application of the model results. Therefore, the future work can focus on the integration of big data and questionnaire of truck drivers to obtain more effective data.

## REFERENCES

[1] China National Bureau of Statistics, *China Transportation Statistical Yearbook 2017*. Beijing, China: Statistics Press, 2017.

[2] D. R. Poulter, P. Chapman, P. A. Bibby, D. D. Clarke, and D. Crundall, "An application of the theory of planned behaviour to truck driving behaviour and compliance with regulations," *Accident Anal. Prevention*, vol. 40, no. 6, pp. 2058–2064, Nov. 2008, doi: 10.1016/j.aap.2008.09.002.

[3] K. L. Campbell, "Fatal accident involvement rates by driver age for large trucks," *Accident Anal. Prevention*, vol. 23, no. 4, pp. 287–295, Aug. 1991, doi: 10.1016/0001-4575(91)90006-Q.

[4] L.-Y. Chang and F. Mannering, "Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents," *Accident Anal. Prevention*, vol. 31, no. 5, pp. 579–592, Sep. 1999, doi: 10.1016/S0001-4575(99)00014-7.

[5] Y.-H. Huang, M. Roetting, J. R. McDevitt, D. Melton, and G. S. Smith, "Feedback by technology: Attitudes and opinions of truck drivers," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 8, no. 4, pp. 277–297, Jul. 2005, doi: 10.1016/j.trf.2005.04.005.

[6] U. Björnstig, J. Björnstig, and A. Eriksson, "Passenger car collision fatalities—With special emphasis on collisions with heavy vehicles," *Accident Anal. Prevention*, vol. 40, no. 1, pp. 158–166, Jan. 2008, doi: 10.1016/j.aap.2007.05.003.

[7] J. Hong, J. Park, G. Lee, and D. Park, "Endogenous commercial driver's traffic violations and freight truck-involved crashes on mainlines of expressway," *Accident Anal. Prevention*, vol. 131, pp. 327–335, Oct. 2019, doi: 10.1016/j.aap.2019.07.026.

[8] X. Zhu and S. Srinivasan, "A comprehensive analysis of factors influencing the injury severity of large-truck crashes," *Accident Anal. Prevention*, vol. 43, no. 1, pp. 49–57, Jan. 2011, doi: 10.1016/j.aap.2010.07.007.

[9] T. Zhou and J. Zhang, "Analysis of commercial truck drivers' potentially dangerous driving behaviors based on 11-month digital tachograph data and multilevel modeling approach," *Accident Anal. Prevention*, vol. 132, Nov. 2019, Art. no. 105256, doi: 10.1016/j.aap.2019.105256.

[10] Z. Zheng, P. Lu, and B. Lantz, "Commercial truck crash injury severity analysis using gradient boosting data mining model," *J. Saf. Res.*, vol. 65, pp. 115–124, Jun. 2018, doi: 10.1016/j.jsr.2018.03.002.

[11] E. Petridou and M. Moustaki, "Human factors in the causation of road traffic crashes," *Eur. J. Epidemiol.*, vol. 16, no. 9, pp. 819–826, Sep. 2000, doi: 10.1023/A:1007649804201.

[12] A. B. Ellison, S. P. Greaves, and M. C. J. Bliemer, "Driver behaviour profiles for road safety analysis," *Accident Anal. Prevention*, vol. 76, pp. 118–132, Mar. 2015, doi: 10.1016/j.aap.2015.01.009.

[13] C. Atombo, C. Wu, M. Zhong, and H. Zhang, "Investigating the motivational factors influencing drivers intentions to unsafe driving behaviours: Speeding and overtaking violations," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 43, pp. 104–121, Nov. 2016, doi: 10.1016/j.trf.2016.09.029.

[14] Ö. im eko lu, and T. Lajunen, "Social psychology of seat belt use: A comparison of theory of planned behavior and health belief model," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 11, no. 3, pp. 181–191, May 2008, doi: 10.1016/j.trf.2007.10.001.

[15] R. J. S. A. Khattak and F. Targa, "Risk factors in large truck rollovers and injury severity: Analysis of single-vehicle collisions," presented at the 82nd Annu. Meeting Transp. Res. Board, 2003.

[16] P. De Pelsmacker and W. Janssens, "The effect of norms, attitudes and habits on speeding behavior: Scale development and model building and estimation," *Accident Anal. Prevention*, vol. 39, no. 1, pp. 6–15, Jan. 2007, doi: 10.1016/j.aap.2006.05.011.

[17] H. W. Warner and L. Åberg, "Drivers' decision to speed: A study inspired by the theory of planned behavior," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 9, no. 6, pp. 427–433, Jan. 2006, doi: 10.1016/j.trf.2006.03.004.

[18] F. Letirand and P. Delhomme, "Speed behaviour as a choice between observing and exceeding the speed limit," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 8, no. 6, pp. 481–492, Jan. 2005, doi: 10.1016/j.trf.2005.06.002.

[19] G. Zhang, K. K. W. Yau, and X. Gong, "Traffic violations in Guangdong province of China: Speeding and drunk driving," *Accident Anal. Prevention*, vol. 64, pp. 30–40, Mar. 2014, doi: 10.1016/j.aap.2013.11.002.

[20] C. J. Armitage, P. Norman, and M. Conner, "Can the theory of planned behaviour mediate the effects of age, gender and multidimensional health locus of control?" *Brit. J. Health Psychol.*, vol. 7, no. 3, pp. 299–316, 2002, doi: 10.1348/135910702760213698.

[21] M. J. M. Sullman, M. L. Meadows, and K. B. Pajo, "Aberrant driving behaviours amongst New Zealand truck drivers," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 5, no. 3, pp. 217–232, Sep. 2002, doi: 10.1016/S1369-8478(02)00019-0.

[22] G. Zhang, K. K. W. Yau, and G. Chen, "Risk factors associated with traffic violations and accident severity in China," *Accident Anal. Prevention*, vol. 59, pp. 18–25, Oct. 2013, doi: 10.1016/j.aap.2013.05.004.

[23] B. González-Iglesias, J. A. Gómez-Fraguela, and M. Á. Luengo-Martín, "Driving anger and traffic violations: Gender differences," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 15, no. 4, pp. 404–412, Jul. 2012, doi: 10.1016/j.trf.2012.03.002.

[24] C.-M. Tseng, "Speeding violations related to a driver's social-economic demographics and the most frequent driving purpose in Taiwan's male population," *Saf. Sci.*, vol. 57, pp. 236–242, Aug. 2013, doi: 10.1016/j.ssci.2013.02.005.

[25] T. Rosenbloom and A. Perlman, "Tendency to commit traffic violations and presence of passengers in the car," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 39, pp. 10–18, May 2016, doi: 10.1016/j.trf.2016.02.008.

[26] M. H. Hosseinlou, A. Mahdavi, and M. J. Nooghabi, "Validation of the influencing factors associated with traffic violations and crashes on freeways of developing countries: A case study of Iran," *Accident Anal. Prevention*, vol. 121, pp. 358–366, Dec. 2018, doi: 10.1016/j.aap.2018.06.009.

[27] M. A. Akaateba, R. Amoh-Gyimah, and O. Amponsah, "Traffic safety violations in relation to drivers' educational attainment, training and experience in Kumasi, Ghana," *Saf. Sci.*, vol. 75, pp. 156–162, Jun. 2015, doi: 10.1016/j.ssci.2015.02.010.

[28] V. Shankar, J. Milton, and F. Mannering, "Modeling accident frequencies as zero-altered probability processes: An empirical inquiry," *Accident Anal. Prevention*, vol. 29, no. 6, pp. 829–837, Nov. 1997, doi: 10.1016/S0001-4575(97)00052-3.

[29] K. Wang, A. H. Lee, K. K. W. Yau, and P. J. W. Carrivick, "A bivariate zero-inflated Poisson regression model to analyze occupational injuries," *Accident Anal. Prevention*, vol. 35, no. 4, pp. 625–629, Jul. 2003, doi: 10.1016/S0001-4575(02)00036-2.

[30] D. Lord, S. P. Washington, and J. N. Ivan, "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory," *Accident Anal. Prevention*, vol. 37, no. 1, pp. 35–46, Jan. 2005, doi: 10.1016/j.aap.2004.02.004.

[31] Z. Li, S. Knight, L. J. Cook, L. K. Hyde, R. Holubkov, and L. M. Olson, "Modeling motor vehicle crashes for street racers using zero-inflated models," *Accident Anal. Prevention*, vol. 40, no. 2, pp. 835–839, Mar. 2008, doi: 10.1016/j.aap.2007.09.022.

[32] H. Jang, S. Lee, and S. W. Kim, "Bayesian analysis for zero-inflated regression models with the power prior: Applications to road safety countermeasures," *Accident Anal. Prevention*, vol. 42, no. 2, pp. 540–547, Mar. 2010, doi: 10.1016/j.aap.2009.08.022.

[33] X. Jiang, B. Huang, R. L. Zaretzki, S. Richards, X. Yan, and H. Zhang, "Investigating the influence of curbs on single-vehicle crash injury severity utilizing zero-inflated ordered probit models," *Accident Anal. Prevention*, vol. 57, pp. 55–66, Aug. 2013, doi: 10.1016/j.aap.2013.03.018.

[34] G. Fountas and P. C. Anastasopoulos, "Analysis of accident injury-severity outcomes: The zero-inflated hierarchical ordered probit model with correlated disturbances," *Anal. Methods Accident Res.*, vol. 20, pp. 30–45, Dec. 2018, doi: 10.1016/j.amar.2018.09.002.

[35] F. L. Mannering and C. R. Bhat, "Analytic methods in accident research: Methodological frontier and future directions," *Anal Methods Accident Res.*, vol. 1, pp. 1–22, Jan. 2014, doi: 10.1016/j.amar.2013.09.001.

[36] C. Pankok and D. Kaber, "The effect of navigation display clutter on performance and attention allocation in presentation- and simulator-based driving experiments," *Appl. Ergonom.*, vol. 69, pp. 136–145, May 2018, doi: 10.1016/j.apergo.2018.01.008.

[37] S. A. Chermahini, D. Stavrinos, C. N. Pope, and A. Behzadnia, "Relations between a laboratory-based sustained attention task and traffic violations in an Iranian sample of drivers," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 58, pp. 177–183, Oct. 2018, doi: 10.1016/j.trf.2018.06.007.

[38] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transp. Res. A, Policy Pract.*, vol. 44, no. 5, pp. 291–305, Jun. 2010, doi: 10.1016/j.tra.2010.02.001.

[39] H. Huang, H. Zhou, J. Wang, F. Chang, and M. Ma, ''A multivariate spatial model of crash frequency by transportation modes for urban intersections,'' *Anal. Methods Accident Res.*, vol. 14, pp. 10–21, Jun. 2017, doi: 10.1016/j.amar.2017.01.001.
[40] C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang, ''The promises of big data and small data for travel behavior aka human mobility analysis,'' *Transp. Res. C, Emerg. Technol.*, vol. 68, pp. 285–299, Jul. 2016, doi: 10.1016/j.trc.2016.04.005.
[41] J. Bao, P. Liu, and S. V. Ukkusuri, ''A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data,'' *Accident Anal. Prevention*, vol. 122, pp. 239–254, Jan. 2019, doi: 10.1016/j.aap.2018.10.015.
[42] S. Li, J. Zhang, P. Li, Y. Wang, and Q. Wang, ''Influencing factors of driving decision-making under the moral dilemma,'' *IEEE Access*, vol. 7, pp. 104132–104142, 2019, doi: 10.1109/ACCESS.2019.2932043.
[43] A. S. Mueller and L. M. Trick, ''Driving in fog: The effects of driving experience and visibility on speed compensation and hazard avoidance,'' *Accident Anal. Prevention*, vol. 48, pp. 472–479, Sep. 2012, doi: 10.1016/j.aap.2012.03.003.
[44] J. O. Brooks, M. C. Crisler, N. Klein, R. Goodenough, R. W. Beeco, C. Guirl, P. J. Tyler, A. Hilpert, Y. Miller, Y. Miller, J. Grygier, B. Burroughs, A. Martin, R. Ray, and C. Palmer, ''Speed choice and driving performance in simulated foggy conditions,'' *Accident Anal. Prevention*, vol. 43, no. 3, pp. 698–705, May 2011, doi: 10.1016/j.aap.2010.10.014.
[45] J. Lee, B.-J. Park, and C. Lee, ''Deterrent effects of demerit points and license sanctions on drivers' traffic law violations using a proportional hazard model,'' *Accident Anal. Prevention*, vol. 113, pp. 279–286, Apr. 2018, doi: 10.1016/j.aap.2018.01.028.
[46] F. Sagberg and R. Ingebrigtsen, ''Effects of a penalty point system on traffic violations,'' *Accident Anal. Prevention*, vol. 110, pp. 71–77, Jan. 2018, doi: 10.1016/j.aap.2017.11.002.
[47] R. Factor, ''Reducing traffic violations in minority localities: Designing a traffic enforcement program through a public participation process,'' *Accident Anal. Prevention*, vol. 121, pp. 71–81, Dec. 2018, doi: 10.1016/j.aap.2018.09.005.
[48] Y. Alver, M. C. Demirel, and M. M. Mutlu, ''Interaction between sociodemographic characteristics: Traffic rule violations and traffic crash history for young drivers,'' *Accident Anal. Prevention*, vol. 72, pp. 95–104, Dec. 2014, doi: 10.1016/j.aap.2014.06.015.
[49] W. Hao, Y. Ya-Dong, and M. Yong, ''Research on the Yangtze River accident casualties using zero-inflated negative binomial regression technique,'' in *Proc. IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Aug. 2016, pp. 72–75, doi: 10.1109/ICITE.2016.7581310.
[50] D. Lambert, ''Zero-inflated Poisson regression, with an application to defects in manufacturing,'' *Technometrics*, vol. 34, no. 1, pp. 1–14, Feb. 1992, doi: 10.1080/00401706.1992.10485228.
[51] N. Jansakul and J. P. Hinde, ''Score tests for zero-inflated Poisson models,'' *Comput. Statist. Data Anal.*, vol. 40, no. 1, pp. 75–96, Jul. 2002, doi: 10.1016/S0167-9473(01)00104-9.
[52] C. Xu, S. Xu, C. Wang, and J. Li, ''Investigating the factors affecting secondary crash frequency caused by one primary crash using zero-inflated ordered probit regression,'' *Phys. A, Stat. Mech. Appl.*, vol. 524, pp. 121–129, Jun. 2019, doi: 10.1016/j.physa.2019.03.036.
[53] M. N. Harris and X. Zhao, ''A zero-inflated ordered probit model, with an application to modelling tobacco consumption,'' *J. Econometrics*, vol. 141, no. 2, pp. 1073–1099, Jun. 2007, doi: 10.1016/j.jeconom.2007.01.002.
[54] G. S. Maddala, *Limited-Dependent and Qualitative Variables in Econometrics Econometric Society Monographs Cambridge*. Cambridge, U.K.: Cambridge Univ. Press, 1983.
[55] P. Downward, F. Lera-Lopez, and S. Rasciute, ''The zero-inflated ordered probit approach to modelling sports participation,'' *Econmic Model.*, vol. 28, no. 6, pp. 2469–2477, Nov. 2011, doi: 10.1016/j.econmod.2011.06.024.
[56] S. P. Washington, M. G. Karlaftis, and F. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*. Boca Raton, FL, USA: CRC Press, 2010.
[57] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde, ''Bayesian measures of model complexity and fit,'' *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 64, no. 4, pp. 583–639, 2002, doi: 10.1111/1467-9868.00353.
[58] J. Brown, Y. He, and H. Lang, ''Quantifying drivers' driving skills using closed-loop dynamic simulations of articulated heavy vehicles,'' *Simulation Model. Pract. Theory*, vol. 99, Feb. 2020, Art. no. 102014, doi: 10.1016/j.simpat.2019.102014.
[59] F. Fraboni, V. M. Puchades, M. D. Angelis, L. Pietrantoni, and G. Prati, ''Red-light running behavior of cyclists in Italy: An observational study,'' *Accident Anal. Prevention*, vol. 120, pp. 219–232, Nov. 2018, doi: 10.1016/j.aap.2018.08.013.
[60] L. Bai, P. Liu, Y. Chen, X. Zhang, and W. Wang, ''Comparative analysis of the safety effects of electric bikes at signalized intersections,'' *Transp. Res. D, Transp. Environ.*, vol. 20, pp. 48–54, May 2013, doi: 10.1016/j.trd.2013.02.001.
[61] Y. Y. Guo, P. Liu, L. Bai, C. C. Xu, and J. Chen, ''Red light running behavior of electric bicycles at signalized intersections in China,'' *Transp. Res. Rec.*, vol. 2468, no. 1, pp. 28–37, 2014, doi: 10.3141/2468-04.
[62] X. Yang, M. Huan, M. Abdel-Aty, Y. Peng, and Z. Gao, ''A hazard-based duration model for analyzing crossing behavior of cyclists and electric bike riders at signalized intersections,'' *Accident Anal. Prevention*, vol. 74, pp. 33–41, Jan. 2015, doi: 10.1016/j.aap.2014.10.014.
[63] X. Qu, Y. Yu, M. Zhou, C.-T. Lin, and X. Wang, ''Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach,'' *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 114030, doi: 10.1016/j.apenergy.2019.114030.
[64] J. Zhao, Y. Gao, Z. Yang, J. Li, Y. Feng, Z. Qin, and Z. Bai, ''Truck traffic speed prediction under non-recurrent congestion: Based on optimized deep learning algorithms and GPS data,'' *IEEE Access*, vol. 7, no. 1, pp. 9116–9127, 2019, doi: 10.1109/ACCESS.2018.2890414.

**JINZHONG WU** is currently pursuing the Ph.D. degree in intelligent transportation and information systems engineering with Chang'an University, Xi'an, China. He is also a Researcher with the Research Institute of Highway Ministry of Transport. He has published ten journal and international conference papers. His research interests include transport of dangerous goods and intelligent transportation.

**WENJI FAN** received the Ph.D. degree in logistics management and engineer from Beijing Jiaotong University, Beijing, China, in 2010. She is currently an Associate Researcher with the Research Institute of Highway Ministry of Transport. She has published about 12 articles in journals and international conferences. Her research interests include transport of dangerous goods and intelligent transportation.

**WENCHENG WANG** received the B.E. degree in traffic engineering from Shandong Jianzhu University, Jinan, China, in 2012. He is currently doing research work with the Beijing Municipal Institute of City Planning and Design, Beijing, China. His current research interests include traffic planning and management, road traffic safety, traffic crash analysis, and drivers' violation.