

Article

Driving Risk Assessment Using Near-Miss Events Based on Panel Poisson Regression and Panel Negative Binomial Regression

Shuai Sun ^{1,2}, Jun Bi ^{1,*}, Montserrat Guillen ^{2,*} and Ana M. Pérez-Marín ²

¹ Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; sunshuai@bjtu.edu.cn

² Department of Econometrics, Riskcenter-IREA, Universitat de Barcelona, 08034 Barcelona, Spain; amperez@ub.edu

* Correspondence: bilinghc@163.com (J.B.); mguillen@ub.edu (M.G.); Tel.: +86-13488812321 (J.B.); +34-934037039 (M.G.)

Abstract: This study proposes a method for identifying and evaluating driving risk as a first step towards calculating premiums in the newly emerging context of usage-based insurance. Telematics data gathered by the Internet of Vehicles (IoV) contain a large number of near-miss events which can be regarded as an alternative for modeling claims or accidents for estimating a driving risk score for a particular vehicle and its driver. Poisson regression and negative binomial regression are applied to a summary data set of 182 vehicles with one record per vehicle and to a panel data set of daily vehicle data containing four near-miss events, i.e., counts of excess speed, high speed brake, harsh acceleration or deceleration and additional driving behavior parameters that do not result in accidents. Negative binomial regression ($AIC_{overspeed} = 997.0$, $BIC_{overspeed} = 1022.7$) is seen to perform better than Poisson regression ($AIC_{overspeed} = 7051.8$, $BIC_{overspeed} = 7074.3$). Vehicles are separately classified to five driving risk levels with a driving risk score computed from individual effects of the corresponding panel model. This study provides a research basis for actuarial insurance premium calculations, even if no accident information is available, and enables a precise supervision of dangerous driving behaviors based on driving risk scores.

Keywords: driving risk assessment; usage-based insurance; driving risk score; telematics; near-miss event; driving behavior; panel data analysis; count data model; econometrics; generalized linear model



Citation: Sun, S.; Bi, J.; Guillen, M.; Pérez-Marín, A.M. Driving Risk Assessment Using Near-Miss Events Based on Panel Poisson Regression and Panel Negative Binomial Regression. *Entropy* **2021**, *23*, 829. <https://doi.org/10.3390/e23070829>

Academic Editors: Emilio López Cano and Felipe Ortega

Received: 10 May 2021
Accepted: 24 June 2021
Published: 29 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Near-miss events are incidents that denote the existence of danger, even if no accident occurs. Reporting of near-miss events is an established error reduction technique that has been used by many industries to manage risk and reduce accidents. In the auto insurance industry, insurers traditionally calculate premiums by analyzing past claims reported by the insured policy holders, and reward those drivers that do not report accidents with a no-claims bonus. However, this may be a rather incorrect approach to the assessment of accident risk, especially when the insured has suffered accidents but chooses not to make a claim so as not to lose the no-claims bonus. Fortunately, the advent of the Internet of Vehicles (IoV) offers a better solution to this problem, using near-miss events to identify driving risk. Near-miss events ultimately provide information that can lead to actuarial premium calculations in the auto insurance industry [1,2].

This study explores how to evaluate driving risks, in the short term, and to score drivers without claims and accidents based on information on near-miss counts over a short period of time. One of the main novelties of this approach, in the absence of claims, is to use telematics sensors for observation of drivers over a given period. The model obtained in this study offers an important alternative for driving risk identification. Not only can the model reflect risk factors that influence each near-miss event but it can also help to evaluate

drivers' risks, and fixed-effects panel count data models can be used to rank drivers according to their individual effects. The modeling method and results are invaluable for insurance companies for developing usage-based insurance (UBI) to personalize premiums. They are also of interest to traffic regulatory authorities for promoting safe driving and the prevention of accidents.

Near-miss events are incidents that need to be defined and extracted from the original raw data files for further processing and analysis. By dealing only with near-miss events, and excluding claims or accidents, this study aims to specifically identify driving patterns. This study is carried out both on a per driver summary data set and on a panel data set where a daily summary is shown for each driver. Our data contain counts of the four types of near-miss events in our study. Speeding, high speed braking, harsh acceleration and harsh deceleration have been defined based on actual driving conditions and local laws and regulations. Other high-risk events, e.g., sharp turning, dangerous lane changing and unexpected maneuvers, proved by previous studies to be related to driving risk, are not included in this study due to the dimension and precision limitations of the original data set.

Our interest is to model the frequency of near-miss events given the drivers' characteristics. The simplest statistical model that links a count data dependent variable with explanatory factors is the Poisson model. Essentially, the Poisson model is similar to linear regression, where a response depends on some others inputs. Here we think that distance driven or mean speed among others, influence the expected frequency of near-miss events. A Poisson model, which is also known as a Poisson regression model, is easily interpretable and provides a way to elucidate the significant effects on the conditional expected frequency. Poisson models are constrained by the fact that conditional expectation and conditional variance are equal. Negative binomial regression models are a natural extension that overcomes this restriction. More details on the models are provided in the Methods section below.

Since the extracted frequency of near-miss events is an unbounded non-negative integer, Poisson regression and negative binomial regression are both suitable for modelization. Poisson regression, negative binomial regression, zero-inflated Poisson regression and zero-inflated negative binomial regression are respectively applied to the summary data set. Average speed, brake times, accelerator pedal position, engine fuel rate etc., are selected as independent variables. Either mileage or fuel consumption can be chosen as the exposure variable to offset the model. In order to reach a clear understanding of risk factors of different near-miss events, each near-miss event is individually used as a dependent variable. However, regardless of which one is selected as the dependent variable, negative binomial regression is shown to provide the best fit in the summary data in this study.

Negative binomial regression also performs better than Poisson regression on the panel data sets. Individual effects and time effects are estimated using panel Poisson regression and panel negative binomial regression on a short panel data set of six days in length. The regression results confirm the existence of individual effects and time effects, and also enable the driving risk of each vehicle to be ranked. The driving risk level of vehicles can then be classified by converting the individual effects into scores, thus providing an important reference for further accurate calculation of premiums.

The rest of this article is organized as follows. The development of UBI and previous efforts on driving risk assessment are summarized in Section 2. Section 3 describes the data and introduces the key parameters used in modeling. Section 4 presents the model expression of Poisson regression and negative binomial regression used in the study. The results of negative binomial regression using the summary data set and the panel data set are reported and analyzed in Section 5. The results are discussed and the conclusions are presented in Section 6.

2. Literature Review

The auto insurance industry is continuously pursuing new ways to calculate more accurate actuarial premiums. However, traditional auto insurance calculations are limited by the difficulty of obtaining information on policy holders, so classical ratemaking uses simple information on drivers (age gender,), vehicles (type of car, model and brand) and driving sections [3]. With current advances in information technology, a new type of insurance business, UBI, based on multi-source data and personalized premium calculation is becoming the mainstream. The Pay-as-you-drive (PAYD) mode of charging premiums is based on mileage or fuel consumption, on the premise that mileage or fuel consumption correlates with the probability of suffering an accident [4]. PAYD has evolved into a newer scheme, called the pay-how-you-drive (PHYD) ratemaking mode, which is based on multiple sources of data, including driving behavior data [5]. Following the development of 5G communication technology, it may now be possible to implement an even more sophisticated monitoring and pricing strategy, known as the manage-how-you-drive (MHYD) principle, i.e., real-time calculation of premiums based on multi-source data and providing real-time information to drivers to restrain from bad driving behavior [3,6]. However, due to technological, regulatory and other issues regarding privacy [7], there is still no mature PHYD product on the market at present [8,9] and, in terms of MHYD, further research is necessary on driving risk to produce products that better reflect the driver profile [10].

Traffic accidents all over the world result in a large number of casualties every year, and high-risk driving is one of the main factors behind these incidents [3]. Consequently, research on driving risk has been a topic of interest over recent decades. Simulation experiments to evaluate driving risk have been designed in the laboratory setting to identify driving risk factors [11–14] as well as experiments using actual vehicles on the road [15–19]. Questionnaire surveys for driving risk assessment have also been studied [20,21]. In fact, the naturalistic type of driving data collected by the IoV or smart phones, known as telematics data, can effectively reduce the influence of subjective factors and unreasonable assumptions in producing effective risk-mitigating actions [22–26].

In research related to driving risk assessment in the auto insurance industry, machine learning and generalized linear models feature equally. Machine learning, with its strong ability to process big data efficiently, is increasingly gaining ground in its application in the auto insurance business. Logistic regression [27], cluster analysis [28], decision tree [5], support vector machine [29], neural network [30] and other machine learning models [31–33] have been widely studied in the field of driving risk assessment, and the results have shown machine learning to be a powerful tool [34]. However, since most machine learning procedures, being black box algorithms, do not offer a high degree of interpretability, they cannot completely replace the conventional generalized linear models implemented for decades in the auto insurance industry [8].

Conventional generalized linear models discern the correlation between influencing factors and claims or accidents in frequency and severity models [9,24,25,35]. However, the study of near-miss events even when there is a lack of information on claims and accidents should not be ignored [2,15]; on the contrary, since near-misses are more frequent than accidents and are positively associated with them, they can be considered a good alternative for risk modeling for driving risk assessment [1]. Compared with previous studies, this study not only conducts regression on the summary data set to model and analyze the factors causing near-miss events, but also conducts panel data regression on the panel data set to consider individual effects and time effects. The regression results can not only make more accurate causal inference, but also carry out risk scoring.

3. Data Description

The telematics data used in this study are collected from an IoV information service provider in China. While we cannot obtain more data due to the commercial privacy of the data, the limited data also contains valuable driving risk information, which is worth studying. The original data set contains 182 data files, representing sensor data

for 182 vehicles observed from 3–8 July 2018 [10]. Each data file contains 62 different measurements but, after data processing [36], less than one-third of them can be used due to recording errors and inconsistencies. The original data are transformed for modeling into a summary data set with information on each driver (see details in Table 1).

Table 1. Descriptive statistics of the summary data set for 182 drivers observed from 3–8 July 2018.

Variable	Mean	Standard Deviation	Minimum	Median	Maximum	Definition
overspeed	19.19	45.37	0	0	330	Frequency of driving speed greater than 100 km/h
highspeedbrake	44.23	108.3	0	4	942	Frequency of braking when the driving speed is greater than 90 km/h
harshacceleration	139.0	134.7	0	101	899	Frequency of cases when acceleration is greater than 6 m/s^2
harshdeceleration	141.9	137.8	1	105	913	Frequency of cases when acceleration is less than 6 m/s^2
kilo	2223	1674	3.73	1832.175	7164	Total driving distance (km)
fuel	621.7	470.9	10.25	487.295	2018	Total fuel consumption (L)
brakes	1588	1426	6	1138.5	9243	Total number of brakes
range	5.201	5.021	0.027	3.399	26.78	Range of driving (geographical units)
speed	36.88	16.37	0.297	36.657	67.84	Mean of speed (km/h)
rpm	1028	188.3	233.1	1009.301	1620	Mean of revolutions per minute (r/min)
acceleratorpedalposition	21.05	7.110	0.187	21.26	39.29	Mean of acceleration pedal position (%)
enginefuelrate	11.52	4.464	1.868	11.203	22.01	Mean of engine fuel rate (%)

The number of each parameter is 182.

The variables overspeed, highspeedbrake, harshacceleration and harshdeceleration are individually filtered by combining the rules of traffic law and driving code. Previous studies have confirmed that speeding is a dangerous driving behavior which is likely to cause traffic accidents [3]. In China, traffic safety regulations stipulate a maximum speed for each type of vehicle on all types of roads. The maximum speed limit for the vehicles in this study is 90 km/h; exceeding this by 10% is not deemed to be a traffic offense. Therefore, 100 km/h is taken as the threshold value of the overspeed near-miss event. Another high risk near-miss event that deserves attention is that of emergency braking; at high speed ($>90 \text{ km/h}$), if the brake is not used correctly or is subjected to lateral force, the car is prone to side-slip or even cartwheel. Lastly, both harsh acceleration and harsh deceleration are near-miss events that compromise driving safety and fuel economy. Based on previous research experience [1,2,37] and the filter analysis of the extreme values of this data set by box graph method, 6 m/s^2 is determined as the filtering threshold value of harsh acceleration and harsh deceleration. Figure 1 shows that near-miss events are all non-negative integers. Combined with the relationship between expectation and variance shown in Table 1, the four near-miss events are shown to be suitable as dependent variables of a Poisson regression or a negative binomial regression.

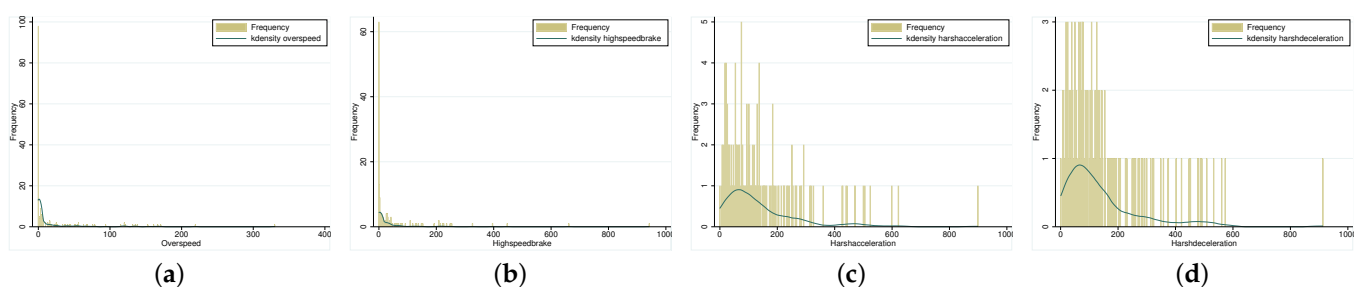


Figure 1. Histogram of frequency distribution of four near-miss events: (a) Over speed; (b) High speed brake; (c) Harsh acceleration; (d) Harsh deceleration.

The panel data set has one summary per day for each driver. The statistics of the panel data set are shown in Table 2.

Table 2. Descriptive statistics of a panel data set for 182 drivers observed over six days (total cases 1092).

Variable	N	Mean	Standard Deviation	Minimum	Median	Maximum
overspeed	1092	3.199	14.37	0	0	315
highspeedbrake	1092	7.435	21.74	0	0	215
harshacceleration	1092	23.37	29.78	0	14	223
harshdeceleration	1092	23.86	30.16	0	13.5	233
kilo	1092	372.6	373.2	0	263.24	1739
fuel	1092	104.1	105.7	0	72.15	565.8
brakes	1092	264.7	291.0	0	178	1940
range	1092	2.406	2.963	0	1.243	14.07
speed	1092	31.96	21.58	0	31.514	77.74
rpm	1092	894.3	346.9	0	973.714	1731
acceleratorpedalposition	1092	17.51	10.19	0	18.613	45.74
enginefuelrate	1092	9.794	5.835	0	10.018	26.18

4. Methods

Poisson regression is a generalized linear model. Negative binomial regression can be considered as a generalization of Poisson regression with overdispersion of the dependent variable Y_i where subindex i refers to the i -th observation in the data set. The probability density function of the Poisson distribution is:

$$P(Y_i = y_i | x_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \tag{1}$$

where λ_i is the Poisson arrival rate and is determined by explanatory variable x_i in Poisson regression to represent the average number of events, which is equal to the expectation and variance of the explained variable $E(Y_i | x_i) = V(Y_i | x_i) = \lambda_i$.

The negative binomial distribution is a mixture of a Poisson (λ) and a Gamma (a, b) distribution. The probability density function of the negative binomial distribution is:

$$f(y | a, b) = \int_0^\infty f(y | \lambda)g(\lambda | a, b)d\lambda = \frac{\Gamma(y + a)}{\Gamma(y + 1)\Gamma(a)} \left(\frac{b}{1 + b}\right)^a \left(\frac{1}{1 + b}\right)^y \tag{2}$$

where λ is the mean and variance of the Poisson distribution, a is the shape parameter of the Gamma distribution, b is the inverse scale parameter of the Gamma distribution, $E(y) = \frac{a}{b} = \bar{\lambda}$ and $V(y) = \frac{a}{b} \left(1 + \frac{1}{b}\right) = \bar{\lambda} \left(1 + \frac{\bar{\lambda}}{a}\right)$.

The zero-inflated model is applicable when the counting data contains a large number of zero values. Theoretically, it is a two-stage decision. First, it decides whether to choose zero or a positive integer, and then it determines which positive integer to choose. Therefore, the probability distribution of Y_i is a mixed distribution:

$$Pr(Y_i = y_i | x_i) = \begin{cases} \theta + (1 - \theta) P(K_i = y_i | x_i) & y_i = 0 \\ (1 - \theta) P(K_i = y_i | x_i) & y_i > 0 \end{cases} \tag{3}$$

where θ is the probability of an extra zero value, K_i can follow a Poisson distribution or a negative binomial distribution depending on the characteristics of the dependent variable.

The conditional expectation function of a negative binomial regression model depends on a vector of explanatory variables x_i and, similar to Poisson, is usually defined by a log-link as:

$$E(y_i | x_i) = \lambda_i = T_i \times \exp(\alpha + \beta_1 x_{1i} + \dots + \beta_k x_{ki}) \tag{4}$$

where i is the number of the observation, k depends on the number of independent variables, T_i denotes the offset variables (so, in our application, $kilo_i$ or $fuel_i$ is the exposure variable), $x_{1i} \dots x_{ki}$ represent the independent variables such as $brakes_i$, $range_i$, $speed_i$, rpm_i ,

$acceleratorpedalposition_i$ and $enginefuelrate_i$, α and $\beta_1 \dots \beta_k$ are unknown parameters that need to be estimated.

The two-way fixed effect model of panel Poisson regression and panel negative binomial regression is specified as:

$$E(y_{it} | x_{it}) = \lambda_{it} = T_{it} \times \exp(\alpha + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + d_i + p_t) \tag{5}$$

where i is the number of the observation, t is of time reference, k depends on the number of independent variables, T_{it} is the offset and equals $kilo_{it}$ or $fuel_{it}$ as the exposure variable of the i th observation at time t , $x_{1it} \dots x_{kit}$ represent the independent variables of the i th observation at time t such as $brakes_{it}$, $range_{it}$, $speed_{it}$, rpm_{it} , $acceleratorpedalposition_{it}$ and $enginefuelrate_{it}$, α and $\beta_1 \dots \beta_k$ are unknown parameters that need to be estimated, d_i represents the individual effect and p_t represents the time effect. To avoid identification problems in the model specification, $d_1 = p_1 = 0$.

The methodology of this study involves data preparation, modeling, risk scoring of driving risk, etc. The whole technical process is shown in Figure 2.

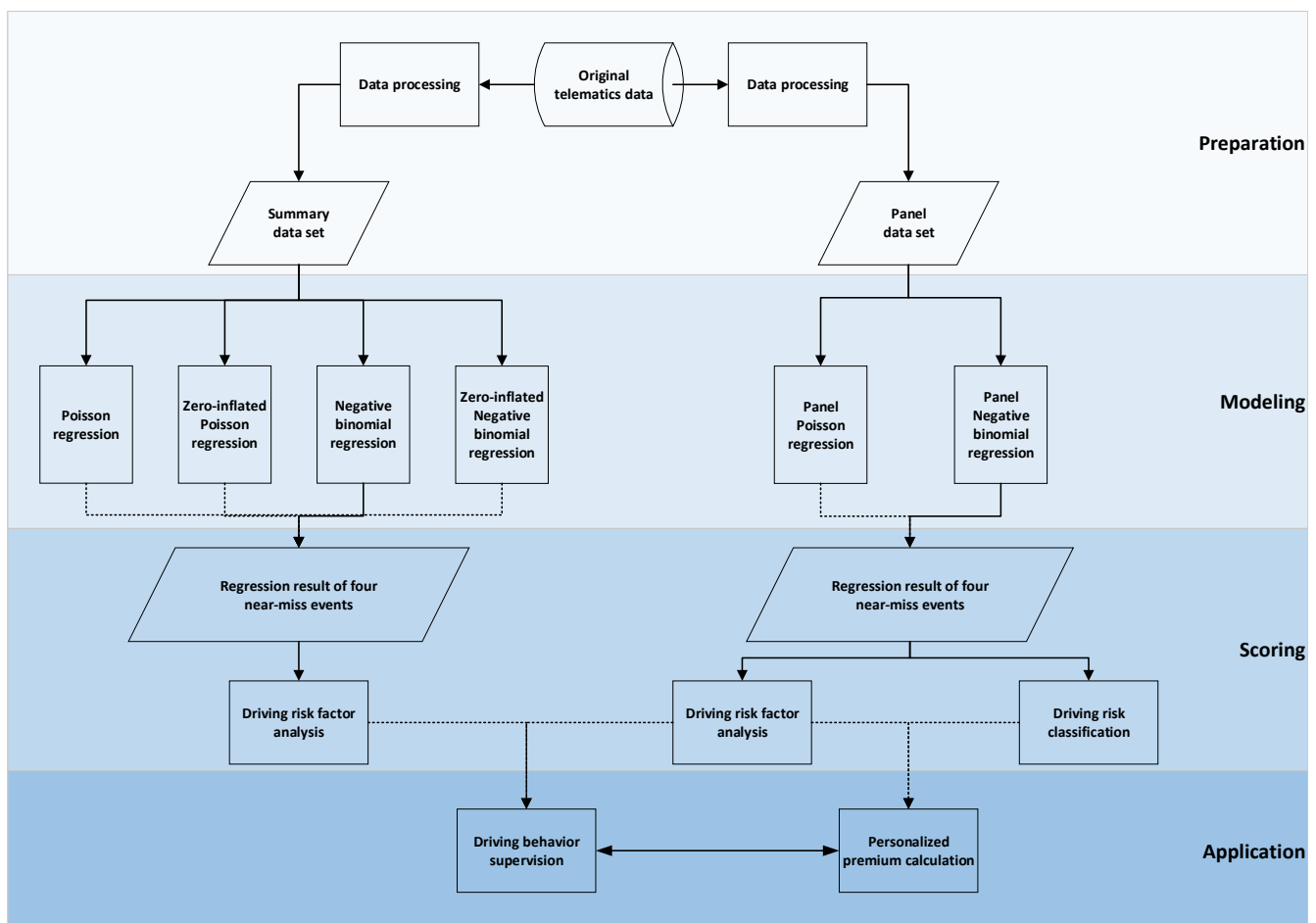


Figure 2. Technical flow chart.

In the data preparation stage, the original data need to be preprocessed, including multi-source data fusion, data cleaning, missing processing, etc. Then the summary data set and panel data set required in this study are obtained through statistical calculation. In the modeling phase, multiple count data models are used on two data sets for regression analysis, which follows certain premises. Our observed drivers can be considered independent of each other. Even if they drive in a similar area, they do not have any apparent relationship between each other. When we observe one driver over time, we have taken care of temporal correlation using the panel model that considers that one individual is

observed repeatedly, here each day. In the scoring stage, the regression results obtained from the regression model most suitable for the data in this study can be used for causal analysis of near-miss events and driving risk scoring and rating. In the research field of telematics data application, the results of this study show this application has potential in, for example, driving behavior supervision and personalized premium calculation. The work at this stage has yet to be completed. Data processing in the preparation and Poisson regression and negative binomial regression on different data sets in the modeling can be implemented with data tools such as Stata, Python, R, etc.

5. Results

Before regression, multicollinearity tests are carried out on all explanatory variables to eliminate the influence of multicollinearity on the model. As shown in Table 3, the variance inflation factors (VIF) of all selected independent variables are less than 5, while the correlation coefficients are generally less than 0.7. This indicates that the multicollinearity among variables is weak, so all of them can be included in the regression equation and robust estimates can be made.

Table 3. Variance inflation factor and correlation of explanatory variables.

Variable	VIF	Brakes	Range	Speed	rpm	Accelerator Pedal Position
brakes	3.07					
range	2.65	0.1213				
speed	2.30	0.0536	0.6262			
rpm	2.13	−0.0254	−0.0203	0.1804		
accelerator pedal position	2.03	0.0154	0.1174	0.3458	0.7695	
engine fuel rate	1.04	0.1687	0.6313	0.6490	0.1075	0.3529

Both Poisson regression and negative binomial regression are applicable to this study, and the zero-inflated model is taken as a consideration for the large number of zero values of dependent variables. In order to determine the regression model which is most suitable for this study, the performance of the two models on different dependent variables is compared. All the estimated results are obtained by regression after standardization of the original values.

5.1. Results of the Summary Data Set

In the summary data set, four near-miss events are respectively treated as dependent variables while the independent variables are brakes, speed, rpm, accelerator pedal position and engine fuel rate, where kilo is chosen as the exposure variable or offset. Poisson regression, zero-inflated Poisson regression, negative binomial regression and zero-inflated negative binomial regression are estimated (see Table 4). Regardless of which near-miss event is the dependent variable, negative binomial regression has maximum log-likelihood value, and minimum AIC value and BIC value. That is, negative binomial regression has the best performance in this data set.

According to the results of negative binomial regression in different dependent variables (see Table 5 and Figure 3a), different near-miss events are affected by different driving risk factors with different influences. Overall, the average speed has the most obvious influence on near-miss events, with a significant negative effect on harsh acceleration (−0.776) and harsh deceleration (−0.658). The impact of braking event number on near-miss events is also positive significant. The higher the number of braking, the more high speed braking (0.272), harsh acceleration (0.189) and harsh deceleration (0.180) occur. In addition, average RPM is positively correlated with harsh acceleration (0.178), and average accelerator pedal position is positively correlated with harsh acceleration (0.152) and harsh deceleration (0.235). Interestingly, some influencing factors have opposite effects on different dependent variables. Range of driving has a positive effect on high speed brake (0.272) but a nega-

tive effect on harsh deceleration (−0.153) while average engine fuel rate has a significant positive effect on high speed braking (0.705) but a negative effect on sharp deceleration (−0.157). Furthermore, the significance of the constant term indicates that, in addition to the factors considered in this study, there are other factors that also influence near-miss events. The results of the other three regression models on the summary data set are shown in Tables A1–A3, and discussed in the Discussion section.

Table 4. Model performances of Poisson, zero-inflated Poisson, negative binomial and zero-inflated negative binomial in summary data set.

Variable	Model	N	Log-Likelihood	df	AIC	BIC
overspeed	POS	182	−3518.92	7	7051.846	7074.274
	ZIP	182	−2369.82	8	4755.64	4781.272
	NB	182	−490.517	8	997.0338	1022.666
	ZINB	182	−490.516	9	999.0315	1027.868
highspeedbrake	POS	182	−2830.75	7	5675.498	5697.926
	ZIP	182	−2667.02	8	5350.034	5375.666
	NB	182	−627.422	8	1270.843	1296.476
	ZINB	182	−627.422	9	1272.843	1301.68
harshacceleration	POS	182	−5857.26	7	11,728.51	11,750.94
	ZIP	182	−5857.26	8	11,730.51	11,756.14
	NB	182	−1032.81	8	2081.623	2107.255
	ZINB	182	−1032.81	9	2083.623	2112.459
harshdeceleration	POS	182	−6269.47	7	12,552.93	12,575.36
	ZIP	182	−6269.47	8	12,554.93	12,580.56
	NB	182	−1037.14	8	2090.285	2115.917
	ZINB	182	−1037.14	9	2092.285	2121.121

Table 5. The results of negative binomial regression for four near-miss events in the summary data set of drivers.

Variable	Overspeed		Highspeedbrake		Harshacceleration		Harshdeceleration	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
constant	−5.175 ***	−15.87	−5.114 ***	−35.36	−2.548 ***	−43.39	−2.525 ***	−42.99
brakes	0.264	1.29	0.272 **	2.60	0.189 ***	3.38	0.180 ***	3.45
range	0.185	0.79	0.272 *	2.05	−0.100	−1.29	−0.153	−1.94
speed	−0.113	−0.20	0.249	1.28	−0.776 ***	−8.81	−0.658 ***	−7.20
rpm	0.125	0.43	−0.0241	−0.11	0.178	1.90	0.0969	1.07
acceleratorpedalposition	0.290	1.13	0.171	1.03	0.152	1.87	0.235 **	2.82
enginefuelrate	0.227	0.99	0.705 ***	4.49	−0.0883	−1.12	−0.157 *	−2.07
log-likelihood	−490.5		−627.4		−1032.8		−1037.1	
AIC	997.0		1270.8		2081.6		2090.3	
BIC	1022.7		1296.5		2107.3		2115.9	
Observation	182		182		182		182	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

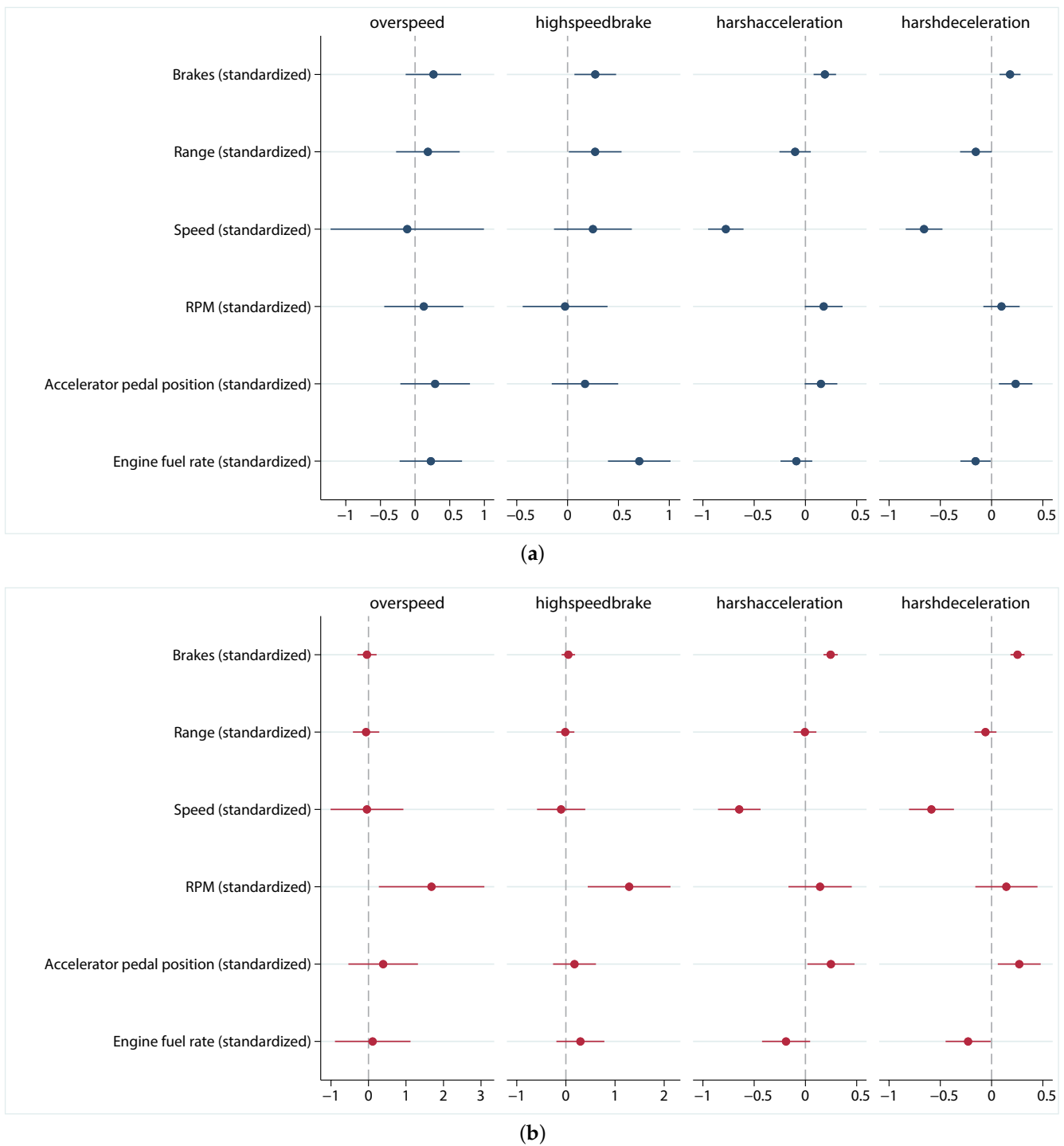


Figure 3. Partial coefficient estimation results of (a) negative binomial regression; (b) Panel negative binomial regression.

5.2. Results of the Panel Data Set

As shown in Table 6, the evaluation index (log-likelihood, AIC and BIC) of negative binomial regression is lower than that of Poisson regression for each dependent variable. Therefore, negative binomial regression is better than Poisson regression on panel data.

The panel negative binomial regression is used to estimate the two-way fixed effect model, considering both individual effect and time effect on four dependent variables. The influencing factors reflected by this (see Table A4 and Figure 3b) differ from those shown in the results of the summary data. For example, harsh acceleration and harsh deceleration are positively affected by the number of brakes (0.246 and 0.253) and average

accelerator pedal position (0.249 and 0.270) but negatively affected by the average speed (−0.645 and −0.586) and average engine fuel rate (−0.188 and −0.229). However, RPM, which is not significant in the summary data, is significantly positive for overspeed (1.683) and high speed braking (1.287). The brakes (0.0505) and engine fuel rate (0.295), which had a significant positive effect on the summary data, become insignificant.

Table 6. Model performances of Poisson and negative binomial in the panel data set of drivers with six observations per driver.

Variable	Model	N	Log-Likelihood	df	AIC	BIC
overspeed	XTPOS	1092	−1926.78	188	4229.559	5168.763
	XTNB	1092	−957.497	189	2292.993	3237.193
highspeedbrake	XTPOS	1092	−2594.37	188	5564.733	6503.937
	XTNB	1092	−1527.05	189	3432.105	4376.305
harshacceleration	XTPOS	1092	−6117.44	188	12,610.89	13,550.09
	XTNB	1092	−3526.09	189	7430.186	8374.386
harshdeceleration	XTPOS	1092	−6042.02	188	12,460.03	13,399.24
	XTNB	1092	−3547.66	189	7473.311	8417.51

The advantage of panel data over summary data is that fixed effects can be estimated and thus individual effects and time effects can be interpreted. The time effect is significant in most cases for high speed braking, harsh acceleration and harsh deceleration, which indicates that these three near-miss events are greatly influenced by time. The time effect on the overspeed event is significant for only one day, suggesting that it is less influenced by time. More importantly, the individual effects of the four near-miss events can be used to score each observation. It should be noted that the first observation has been omitted in the regression to avoid complete multicollinearity, and its value is expected to be zero in the subsequent driving risk score.

6. Discussion

The regression results of Poisson regression (see Table A1), zero-inflated Poisson regression (see Table A2), negative binomial regression (see Table 5) and zero-inflated negative binomial regression (see Table A3) on the summary data set show the importance of driving behavior variables in driving risk. The high significance of two variables, braking times and average speed, in the four regression models indicates that these two factors have a very important impact on the generation of near-miss events. Moreover, the significant performance of specific independent variables in the regression model of specific dependent variables indicates that near-miss events are affected by a variety of driving behavior factors and the formation mechanism of each near-miss event is different. For example, the positive effect of RPM on harsh acceleration events, the positive effect of accelerator pedal position on harsh deceleration events and the positive effect of engine fuel rate on high speed braking events are shown in Tables A1–A3, Table 5.

The results obtained by panel regression are more reliable than those obtained by pooled regression. Tables 5 and A4 and Figure 3 show that some coefficients that are not significant in the pooled negative binomial regression become significant in the panel negative binomial regression, while some significant parameters in the pooled negative binomial regression are not significant in the panel negative binomial regression. This means that the dependent variables are affected by individual effects and time effects. In the panel negative binomial regression, most of the individual and time coefficients are significant, which indicates the suitability of this type of regression analysis.

Driving risks can be evaluated by the regression coefficients of negative binomial models on panel data. The value of the individual coefficients within a regression indicates the individual's deviance from the level of the expected occurrence of a particular near-miss event, given the information on all the other explanatory variables. In other words, the individual effect coefficient can be understood as the effect utility of each vehicle on

the occurrence of the corresponding near-miss event. Geometrically, the effect coefficient of each individual is a change in the intercept.

Four near-miss events are used as dependent variables to obtain four sets of regression coefficients. Given that the influencing factors and generating mechanisms of different near-miss events are different, combining the four groups of regression coefficients into one group is not recommended. However, harsh acceleration and harsh deceleration show very similar characteristics in terms of data description before regression (Tables 1 and 5), after regression (Figure 3 and Table A4) and in distribution of driving risk score (Table A5). Even so, it is not recommended to combine them into a single near-miss event for study, because the occurrence conditions and coping operations of them are different, and it is the most appropriate choice to study each near-miss event separately.

In order to transform individual effect estimates of near-miss models into a driving risk grading, several steps need to be followed. Firstly, winsorization avoids the influence of possibly spurious outliers (the double tail was winsorized with the threshold 0.01 in this study). Secondly, the regression coefficient can be compressed to the interval of [0,1] through normalization. Each group of coefficients is then mapped into an interval of [0,5] (see Table A5), and each observation then is given a driving risk level from 1 to 5, i.e., excellent, good, medium, bad and terrible (see Figure 4). The values of exactly 0 and 5 are included because the corresponding observations are the minimum and the maximum values in their group and are Min-Max scaled. In *overspeed* and *highspeedbrake* groups, two types of observations with high risk or low risk can be clearly seen. This indicates that these two near-miss events are more sensitive to driving behavior than *harshacceleration* and *harshdeceleration* and can be considered as a higher priority and weight in subsequent studies. Note that the same observation (id125) has different risk levels for different near-miss events, which also explains why multiple near-miss events cannot be analyzed together. Ultimately, the premium would be charged individually according to the driving risk level of the insured person.

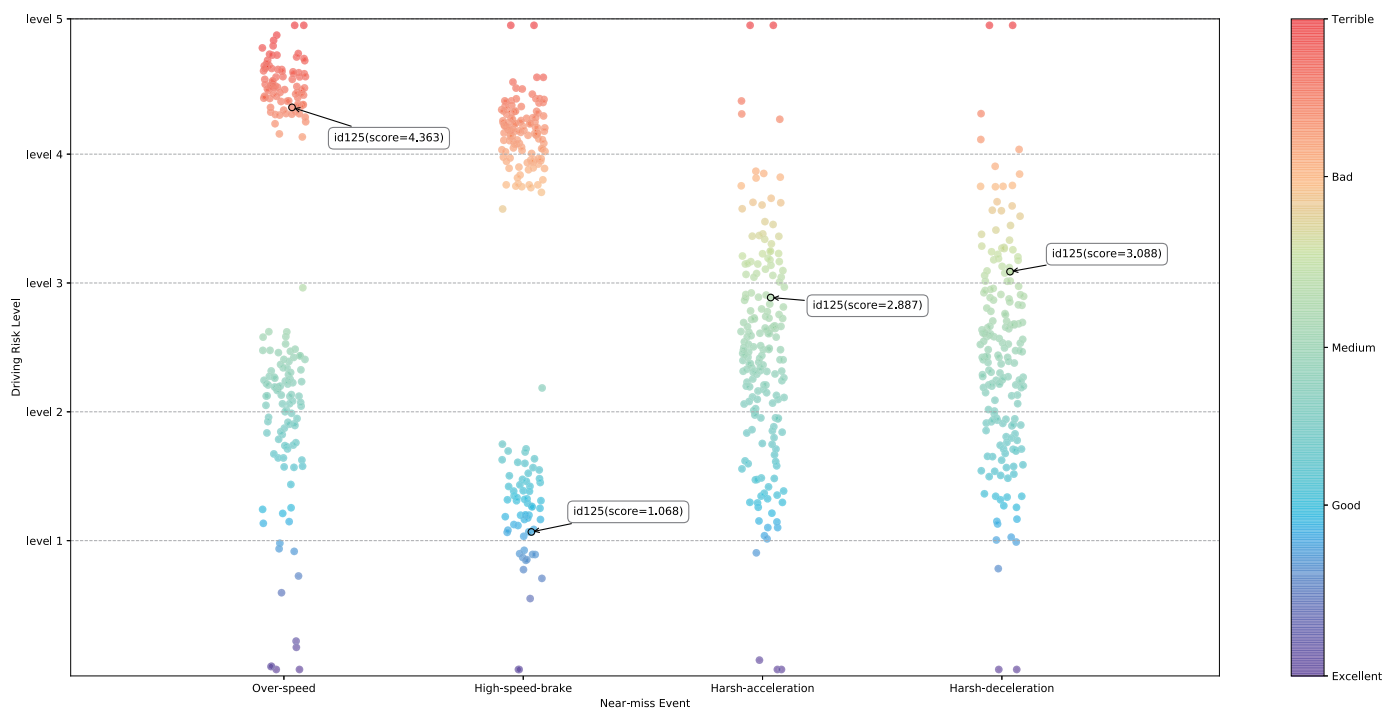


Figure 4. Driving risk ranking of four near-miss events.

7. Conclusions

The number and type of dependent variables and independent variables selected in this study are limited by the size and quality of the original data. With the promotion and innovation of IoV and of new energy vehicles, the amount and dimension of data will be greatly increased. Therefore, application of near-miss events as dependent variables could be easily increased or decreased, according to needs. For example, sharp turn should be included, if possible, as a near-miss event because sharp turn is a highly studied and accident-proven pattern of high driving risk. For the same reason, more driving behavior indicators, such as steering wheel angle speed, brake pedal position, and so on, could be used as independent variables in the regression model. In addition, traditional auto insurance factors, such as driver information, vehicle information, road information, environment information and the health status of batteries (of new energy vehicles) should be considered to provide more optional independent variables for the model.

In practical applications, near-miss events can be combined with claims and accidents to accurately evaluate driving risks. This study proves that near-miss events can be used as driving risk scores when there are no claims or accidents. However, when claims or accidents exist, the driving risk score obtained from claims or accidents can be used as the basis for premium calculation, while the driving risk rating obtained from near-miss events can be used to remind and warn drivers to reduce the corresponding dangerous driving habits.

In this study, the best performing negative binomial regression (see Table 4) was selected as the main method for modeling on our data set. The model is suitable for similar causal analysis of similar data sets. However, in case of risk event prediction or analysis on other data sets, it is necessary to reevaluate the goodness of fit of various models, and even machine learning methods with good prediction performance should be taken into consideration. The optimal method is not fixed, but depends on the data, conditions and purposes.

Econometrics and machine learning complement each other. The generalized linear model established in this study reveals the relationship between driving behavior factors and near-miss events, and gives a driving risk score for each observation. This model has strong explanatory power, but its generalization degree and robustness need to be further tested, especially on larger data volume and data dimension. The successful application of machine learning methods in many fields shows that they are often effective in dealing with big data problems but that their results cannot always be easily interpreted, and this interpretation is exactly what the insurance field values. Therefore, telematics data application offers a new way to help find a balance between econometrics and machine learning so as to have good explainability, good generalization ability, quick response ability, and so on [38,39].

In general, near-miss events can provide insurers with effective risk information in the absence of claims and accident data. In our real case study, negative binomial regression is the most suitable modeling method for near-miss events as dependent variables. This study provides a technical reference for the promotion and development of PHYD ratemaking schemes.

Author Contributions: Conceptualization, S.S. and M.G.; methodology, M.G.; software, S.S.; validation, J.B., M.G. and A.M.P.-M.; formal analysis, S.S.; investigation, S.S.; resources, M.G. and A.M.P.-M.; data curation, J.B., S.S. and M.G.; writing—original draft preparation, S.S.; writing—review and editing, S.S. and M.G.; visualization, S.S.; supervision, J.B. and M.G.; project administration, J.B.; funding acquisition, J.B. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was granted by the Fundamental Research Funds for the Central Universities 2019YJS091.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study is based on the original telematics data from China Satellite Navigation and Communications Co., Ltd., which could not be made public due to confidentiality agreements.

Acknowledgments: M.G. thanks Fundación BBVA and ICREA Academia for providing research funds. S.S. thanks the China Scholarship Council for providing visiting research funds in the second institution. All the authors thank China Satellite Navigation and Communications Co., Ltd. for providing the original telematics data set.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

UBI	usage-based insurance
IoV	Internet of vehicles
PAYD	pay as you drive
PHYD	pay how you drive
MHYD	manage how you drive
VIF	variance inflation factor
POS	Poisson
ZIP	Zero-inflated Poisson
NB	Negative binomial
ZINB	Zero-inflated negative binomial
XTPOS	Panel Poisson
XTNB	Panel negative binomial
AIC	Akaike information criterion
BIC	Bayesian information criterion

Appendix A

Table A1. The results of Poisson regression for four near-miss events in the summary data set of drivers.

Variable	Overspeed		Highspeedbrake		Harshacceleration		Harshdeceleration	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
constant	−5.191 ***	−21.45	−5.194 ***	−29.98	−2.612 ***	−40.34	−2.591 ***	−40.05
brakes	0.279	1.82	0.349 ***	5.93	0.191 ***	3.66	0.186 ***	3.78
range	0.0437	0.21	0.0741	0.78	−0.157	−1.65	−0.208 *	−2.04
speed	−0.175	−0.92	0.489 **	3.22	−0.717 ***	−8.57	−0.601 ***	−6.51
rpm	0.514	1.59	0.202	0.87	0.272 **	2.94	0.183 *	1.98
acceleratorpedalposition	0.0467	0.18	−0.0337	−0.23	0.169	1.91	0.238 **	2.67
enginefuelrate	0.540 *	2.24	0.755 ***	4.32	−0.0499	−0.59	−0.119	−1.48
log-likelihood	−3518.9		−2830.7		−5857.3		−6269.5	
AIC	7051.8		5675.5		11,728.5		12,552.9	
BIC	7074.3		5697.9		11,750.9		12,575.4	
Observation	182		182		182		182	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2. The results of zero-inflated Poisson regression for four near-miss events in the summary data set of drivers.

Variable	Overspeed		Highspeedbrake		Harshacceleration		Harshdeceleration	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
constant	-4.388 ***	-18.08	-5.006 ***	-24.96	-2.612 ***	-40.34	-2.591 ***	-40.05
brakes	0.167	1.08	0.339 ***	5.34	0.191 ***	3.66	0.186 ***	3.78
range	0.0365	0.20	0.0755	0.81	-0.157	-1.65	-0.208 *	-2.04
speed	-0.391 *	-2.12	0.408 *	2.48	-0.717 ***	-8.57	-0.601 ***	-6.51
rpm	0.607 *	2.18	0.274	1.11	0.272 **	2.94	0.183 *	1.98
acceleratorpedalposition	-0.117	-0.52	-0.0563	-0.38	0.169	1.91	0.238 **	2.67
enginefuelrate	0.346	1.59	0.700 ***	4.01	-0.0499	-0.59	-0.119	-1.48
inflate-constant	0.101	0.66	-1.183 ***	-4.72	-27.29 ***	-295.09	-27.00 ***	-363.25
log-likelihood	-2369.8		-2667.0		-5857.3		-6269.5	
AIC	4755.6		5350.0		11,730.5		12,554.9	
BIC	4781.3		5375.7		11,756.1		12,580.6	
Observation	182		182		182		182	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3. The results of zero-inflated negative binomial regression for four near-miss events in the summary data set of drivers.

Variable	Overspeed		Highspeedbrake		Harshacceleration		Harshdeceleration	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
constant	-5.153 ***	-7.37	-5.114 ***	-35.36	-2.548 ***	-43.39	-2.525 ***	-42.99
brakes	0.263	1.25	0.272 **	2.60	0.189 ***	3.38	0.180 ***	3.45
range	0.180	0.73	0.272 *	2.05	-0.100	-1.29	-0.153	-1.94
speed	-0.114	-0.20	0.249	1.28	-0.776 ***	-8.81	-0.658 ***	-7.20
rpm	0.130	0.41	-0.0241	-0.11	0.178	1.90	0.0969	1.07
acceleratorpedalposition	0.284	0.98	0.171	1.03	0.152	1.87	0.235 **	2.82
enginefuelrate	0.227	0.99	0.705 ***	4.49	-0.0883	-1.12	-0.157 *	-2.07
inflate-constant	-3.793	-0.17	-14.62 ***	-6.14	-25.29 ***	-294.22	-23.27 ***	-313.10
log-likelihood	-490.5		-627.4		-1032.8		-1037.1	
AIC	999.0		1272.8		2083.6		2092.3	
BIC	1027.9		1301.7		2112.5		2121.1	
Observation	182		182		182		182	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4. Panel negative binomial regression results for four near-miss events.

Variable	Overspeed		Highspeedbrake		Harshacceleration		Harshdeceleration	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
constant	-3.768 ***	(-8.68)	-4.356 ***	(-17.39)	-2.284 ***	(-25.13)	-2.269 ***	(-20.81)
brakes	-0.0400	(-0.31)	0.0505	(0.73)	0.246 ***	(6.93)	0.253 ***	(7.23)
range	-0.0637	(-0.36)	-0.0108	(-0.12)	-0.00410	(-0.07)	-0.0595	(-1.09)
speed	-0.0405	(-0.08)	-0.0965	(-0.39)	-0.645 ***	(-6.09)	-0.586 ***	(-5.24)
rpm	1.683 *	(2.34)	1.287 **	(3.00)	0.143	(0.91)	0.145	(0.93)
acceleratorpedalposition	0.391	(0.83)	0.175	(0.79)	0.249 *	(2.12)	0.270 *	(2.52)
enginefuelrate	0.113	(0.22)	0.295	(1.18)	-0.188	(-1.58)	-0.229 *	(-2.04)
2018-07-04	0.273	(1.23)	0.216	(1.91)	-0.111 *	(-2.12)	-0.216 ***	(-4.33)
2018-07-05	-0.168	(-0.73)	-0.0572	(-0.52)	-0.206 ***	(-4.34)	-0.317 ***	(-6.72)
2018-07-06	-0.00716	(-0.03)	-0.228 *	(-2.08)	-0.257 ***	(-4.84)	-0.370 ***	(-7.19)
2018-07-07	-0.477 *	(-2.11)	-0.200	(-1.68)	-0.485 ***	(-7.41)	-0.600 ***	(-9.27)
2018-07-08	0.206	(0.90)	0.117	(0.95)	-0.694 ***	(-8.63)	-0.784 ***	(-9.58)
id2	-28.81 ***	(-29.17)	-2.001 *	(-2.25)	1.266 ***	(5.24)	1.342 ***	(4.80)
id3	-19.05 ***	(-14.86)	-18.13 ***	(-16.37)	2.004 *	(2.31)	1.740 ***	(4.75)
id4	-18.29 ***	(-15.94)	-17.91 ***	(-20.04)	1.891 ***	(8.66)	1.960 ***	(8.58)
id5	-29.62 ***	(-41.07)	-4.956 ***	(-7.51)	-1.193 ***	(-3.77)	-1.072 ***	(-3.39)
id6	-1.478 *	(-2.40)	-0.554	(-1.79)	1.067 ***	(3.31)	0.935 ***	(4.27)
id7	-3.236 ***	(-4.11)	-0.645	(-1.58)	0.656 **	(3.15)	0.835 **	(3.05)
id8	-20.79 ***	(-24.48)	-2.368 ***	(-4.01)	-0.190	(-0.61)	0.124	(0.35)
id9	-1.156	(-1.10)	-0.0678	(-0.11)	-0.251	(-0.79)	-0.109	(-0.28)
id10	-3.110 ***	(-5.93)	-1.527 ***	(-4.20)	-0.345 *	(-2.17)	-0.256	(-1.63)
id11	-2.026 *	(-2.42)	-1.163 ***	(-3.45)	-0.162	(-0.88)	-0.272	(-1.22)

Table A4. Cont.

Variable	Overspeed		Highspeedbrake		Harshacceleration		Harshdeceleration	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
id12	-1.342 *	(-2.51)	-0.772 *	(-2.45)	0.0781	(0.34)	0.0981	(0.50)
id13	-2.344 ***	(-5.13)	-0.808 *	(-2.21)	-0.138	(-0.52)	-0.129	(-0.66)
id14	-3.178 ***	(-5.87)	0.442	(1.43)	-0.629 ***	(-3.66)	-0.365 *	(-2.07)
id15	-1.254 *	(-2.31)	0.167	(0.46)	-0.0894	(-0.34)	0.0270	(0.10)
id16	-22.40 ***	(-37.49)	-21.63 ***	(-41.20)	0.271	(1.35)	0.439 *	(2.13)
id17	-21.77 ***	(-25.73)	-2.102 ***	(-4.06)	-0.200	(-0.87)	0.0983	(0.38)
id18	-20.99 ***	(-19.61)	-0.805	(-1.48)	-1.124 ***	(-6.78)	-1.267 ***	(-5.69)
id19	-0.998	(-1.58)	0.380	(1.06)	0.587 **	(2.72)	0.586 ***	(3.89)
id20	-23.99 ***	(-38.51)	-3.749 **	(-3.22)	0.292	(1.19)	0.0926	(0.40)
id21	-21.79 ***	(-35.89)	-2.577 **	(-2.75)	0.322	(1.34)	0.458 **	(2.61)
id22	-2.642 ***	(-3.94)	-0.229	(-0.63)	0.496 **	(3.00)	0.538 **	(2.89)
id23	-0.792	(-1.10)	0.00108	(0.00)	-0.474	(-1.61)	-0.409	(-1.71)
id24	-23.37 ***	(-34.65)	-22.27 ***	(-40.12)	-0.329	(-1.36)	-0.103	(-0.32)
id25	-21.06 ***	(-30.30)	-20.67 ***	(-32.16)	-0.882 ***	(-3.45)	-0.731 *	(-2.56)
id26	-2.739 ***	(-3.82)	-1.000 **	(-2.61)	-0.440	(-1.77)	-0.667 ***	(-3.68)
id27	-23.10 ***	(-41.71)	-22.25 ***	(-45.16)	-0.0464	(-0.15)	0.0656	(0.22)
id28	-17.86 ***	(-17.18)	-17.85 ***	(-18.78)	0.0432	(0.13)	0.309	(1.41)
id29	-1.136	(-1.59)	-0.872 *	(-2.33)	0.591 ***	(3.74)	0.625 ***	(4.09)
id30	-20.56 ***	(-29.51)	-19.84 ***	(-31.89)	-0.223	(-0.53)	-0.102	(-0.25)
id31	-0.407	(-0.91)	-0.633 *	(-2.03)	-1.148 ***	(-3.69)	-0.949 **	(-3.21)
id32	-3.255 **	(-3.24)	-2.923 *	(-2.38)	-0.110	(-0.36)	0.143	(0.49)
id33	-19.05 ***	(-19.59)	-19.50 ***	(-21.06)	-0.177	(-0.37)	-0.153	(-0.32)
id34	-2.431 **	(-3.26)	-1.547 ***	(-3.51)	-0.00573	(-0.02)	-0.0439	(-0.14)
id35	-3.832 ***	(-3.98)	-1.041 *	(-2.18)	-0.607 ***	(-3.58)	-0.552 **	(-3.14)
id36	-4.135 ***	(-4.74)	-2.412 ***	(-4.36)	-0.285	(-1.37)	-0.343	(-1.82)
id37	-38.70 ***	(-32.02)	-1.232	(-1.72)	-0.480	(-1.64)	-0.218	(-0.75)
id38	-20.26 ***	(-29.14)	-1.364 *	(-2.11)	-1.484 ***	(-5.08)	-1.121 ***	(-4.31)
id39	-38.61 ***	(-41.67)	10.89 ***	(14.67)	11.65 ***	(21.85)	11.77 ***	(21.58)
id40	-1.326	(-1.57)	-0.416	(-0.81)	-0.278	(-1.10)	0.0791	(0.30)
id41	-2.443 **	(-3.19)	-1.020 *	(-2.02)	0.180	(0.82)	0.155	(0.61)
id42	-0.467	(-0.57)	0.442	(0.93)	0.607	(1.63)	0.398	(1.45)
id43	-2.164 *	(-2.45)	0.219	(0.39)	-0.0359	(-0.13)	0.0900	(0.35)
id44	-2.465 ***	(-3.52)	-0.156	(-0.35)	0.336	(1.33)	0.468	(1.85)
id45	-2.110 **	(-3.26)	-1.315 **	(-3.20)	0.105	(0.64)	0.282	(1.18)
id46	0.132	(0.14)	-0.480	(-1.01)	-0.312 **	(-2.77)	-0.235	(-1.65)
id47	-2.957 ***	(-5.23)	-0.975	(-1.32)	-0.853 ***	(-4.77)	-0.656 ***	(-4.01)
id48	0.486	(0.78)	1.381 ***	(3.72)	0.829 ***	(5.36)	0.787 ***	(4.63)
id49	-25.28 ***	(-34.70)	-1.575 **	(-3.27)	-0.568 **	(-2.60)	-0.353	(-1.67)
id50	-2.556 ***	(-3.94)	-1.907 ***	(-3.87)	-0.413 *	(-2.43)	-0.331	(-1.80)
id51	-20.62 ***	(-20.07)	-20.01 ***	(-27.73)	1.123 ***	(3.56)	1.140 ***	(3.69)
id52	-21.73 ***	(-16.76)	-20.90 ***	(-19.35)	-0.354	(-1.10)	-0.952 ***	(-3.75)
id53	-20.65 ***	(-21.34)	-20.00 ***	(-31.74)	-0.133	(-0.72)	0.200	(1.01)
id54	-4.881 ***	(-5.39)	-1.082 **	(-2.66)	-0.686 ***	(-3.36)	-0.639 **	(-3.00)
id55	-4.290 ***	(-4.44)	-1.731 ***	(-3.90)	0.472	(1.80)	0.476	(1.37)
id56	-2.462 ***	(-3.72)	-0.0866	(-0.20)	0.119	(0.40)	0.377	(0.77)
id57	-21.96 ***	(-22.99)	-0.700	(-1.51)	0.110	(0.36)	0.719 *	(2.27)
id58	-1.877	(-1.66)	-0.692	(0.83)	-0.344	(-0.77)	0.0660	(0.16)
id59	-38.77 ***	(-43.77)	-0.0709	(-0.10)	-0.726 *	(-2.11)	-0.587	(-1.68)
id60	-3.117 **	(-2.65)	-3.815 ***	(-4.00)	-0.711 *	(-2.07)	-0.565	(-1.69)
id61	0.821	(0.83)	1.078	(1.78)	-1.288 **	(-2.87)	-1.076 *	(-2.42)
id62	-0.465	(-0.61)	0.546	(1.46)	-0.670	(-1.58)	-0.473	(-1.15)
id63	-21.36 ***	(-28.39)	-20.68 ***	(-33.84)	1.393 ***	(10.47)	1.513 ***	(10.22)
id64	-2.529	(-1.39)	-1.707	(-1.18)	1.334 ***	(7.83)	1.339 ***	(6.13)
id65	-21.50 ***	(-34.71)	-20.76 ***	(-42.35)	-1.923 ***	(-5.19)	-1.288 ***	(-5.26)
id66	-1.389	(-1.49)	-1.510 ***	(-3.74)	0.504 **	(2.64)	0.971 ***	(5.69)
id67	-25.65 ***	(-34.83)	-3.400 ***	(-3.49)	-0.371 *	(-1.98)	-0.304	(-1.70)
id68	-19.09 ***	(-27.85)	-18.66 ***	(-35.37)	-1.286 **	(-2.97)	-1.660 ***	(-7.16)
id69	-24.40 ***	(-27.41)	-21.86 ***	(-33.37)	-0.589 **	(-2.94)	-0.625 *	(-2.48)
id70	-21.29 ***	(-22.38)	-3.693 ***	(-4.50)	-1.489 ***	(-3.60)	-1.501 ***	(-6.37)
id71	-31.22 ***	(-36.99)	-29.36 ***	(-53.20)	0.587 ***	(3.95)	1.212 ***	(7.82)
id72	-5.534 ***	(-5.68)	-1.058	(-1.96)	-0.516	(-1.41)	-0.643	(-1.84)
id73	-4.323 ***	(-4.06)	-2.863 ***	(-5.25)	-1.527 ***	(-6.22)	-1.523 ***	(-5.97)
id74	-30.92 ***	(-35.40)	-29.09 ***	(-51.97)	0.299	(1.03)	0.765 **	(3.02)
id75	-2.868 ***	(-3.45)	-1.677 ***	(-4.49)	-0.267	(-0.90)	-0.0911	(-0.31)
id76	-21.21 ***	(-22.83)	-21.18 ***	(-31.54)	-1.646 ***	(-3.80)	-1.903 ***	(-4.47)
id77	-19.83 ***	(-15.23)	-19.23 ***	(-21.15)	0.835 **	(2.84)	0.729 **	(2.71)
id78	-24.02 ***	(-33.62)	-3.260 ***	(-3.47)	-2.855 ***	(-10.67)	-2.759 ***	(-8.49)

Table A4. Cont.

Variable	Overspeed		Highspeedbrake		Harshacceleration		Harshdeceleration	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
id79	-3.449 **	(-2.81)	-0.618	(-1.43)	-0.232	(-1.23)	-0.110	(-0.57)
id80	-21.71 ***	(-15.28)	-20.69 ***	(-20.65)	-0.0149	(-0.05)	0.0509	(0.16)
id81	-33.98 ***	(-60.17)	-1.132 *	(-2.44)	-0.341 *	(-2.34)	-0.336 *	(-2.41)
id82	-1.391	(-1.64)	-0.541	(-0.76)	-0.312	(-1.38)	-0.326	(-1.55)
id83	-1.516 **	(-2.58)	0.157	(0.53)	-0.123	(-0.64)	-0.242	(-1.37)
id84	-23.95 ***	(-25.81)	-1.866 *	(-2.04)	-0.750 **	(-3.19)	-0.855 ***	(-4.50)
id85	-22.55 ***	(-23.68)	-3.844 ***	(-4.49)	-1.430 ***	(-5.83)	-1.318 ***	(-5.43)
id86	-29.06 ***	(-33.35)	-2.036 ***	(-3.87)	-1.272 ***	(-3.40)	-1.111 **	(-2.99)
id87	-1.851 *	(-2.47)	1.034 **	(2.59)	0.196	(1.03)	0.425 *	(2.03)
id88	-20.13 ***	(-31.73)	-19.83 ***	(-38.92)	-0.208	(-1.54)	-0.165	(-1.04)
id89	-25.42 ***	(-13.08)	-23.44 ***	(-19.03)	1.100 *	(2.18)	1.135 *	(2.49)
id90	-4.008 ***	(-4.24)	-0.841	(-1.18)	-0.972 ***	(-4.48)	-0.982 **	(-3.17)
id91	-19.51 ***	(-26.36)	-19.33 ***	(-31.18)	0.676 ***	(5.52)	0.818 ***	(4.54)
id92	-26.04 ***	(-14.88)	-24.17 ***	(-22.56)	0.848 *	(2.54)	0.663 *	(2.00)
id93	-23.79 ***	(-23.97)	-22.53 ***	(-27.69)	-0.290	(-1.41)	-0.300	(-1.26)
id94	-2.684 ***	(-3.53)	-1.034 **	(-3.05)	-0.157	(-0.99)	0.0139	(0.06)
id95	-24.75 ***	(-12.35)	-22.86 ***	(-19.16)	-0.503	(-0.78)	-0.670	(-1.26)
id96	-22.66 ***	(-28.80)	-21.77 ***	(-35.98)	1.374 ***	(7.81)	1.343 ***	(6.60)
id97	-20.83 ***	(-29.33)	-20.42 ***	(-32.43)	-0.464 *	(-2.30)	-0.282	(-1.24)
id98	-18.59 ***	(-24.60)	-18.56 ***	(-27.54)	-1.405 ***	(-4.98)	-0.887 *	(-2.38)
id99	-18.19 ***	(-27.16)	-18.21 ***	(-34.28)	-1.774 ***	(-5.28)	-1.369 ***	(-4.34)
id100	-4.226 **	(-3.13)	-21.52 ***	(-26.23)	0.802 *	(2.30)	0.824 *	(2.37)
id101	-22.60 ***	(-13.70)	-21.29 ***	(-22.32)	0.955 *	(2.21)	0.814	(1.96)
id102	-24.81 ***	(-34.96)	-23.71 ***	(-42.27)	0.0308	(0.14)	-0.0294	(-0.12)
id103	-17.82 ***	(-22.56)	-17.67 ***	(-27.25)	0.542 *	(2.27)	0.606 ***	(3.70)
id104	-20.05 ***	(-32.16)	-19.72 ***	(-37.06)	0.131	(0.73)	0.262 *	(1.98)
id105	-3.426 ***	(-4.29)	-0.430	(-0.94)	-0.464	(-0.90)	-0.925 *	(-2.01)
id106	-24.96 ***	(-26.36)	-23.73 ***	(-35.39)	0.317	(1.84)	0.252	(1.33)
id107	-21.01 ***	(-29.69)	-20.53 ***	(-36.62)	0.0144	(0.09)	0.147	(0.63)
id108	-23.28 ***	(-27.56)	-2.647 ***	(-5.35)	-0.532 *	(-2.40)	-0.635 ***	(-3.35)
id109	-20.89 ***	(-19.32)	-20.46 ***	(-21.38)	-0.347 **	(-2.97)	-0.782 ***	(-5.90)
id110	-20.70 ***	(-17.96)	-20.72 ***	(-20.15)	-1.801 ***	(-5.72)	-1.044 ***	(-5.99)
id111	-3.405 ***	(-4.19)	-1.278 ***	(-3.61)	0.173	(1.27)	0.198	(1.40)
id112	-19.65 ***	(-24.23)	-19.29 ***	(-25.89)	-1.453 ***	(-8.75)	-0.831 ***	(-7.19)
id113	-29.63 ***	(-42.75)	-2.998 ***	(-3.76)	-1.703 ***	(-6.85)	-1.296 ***	(-6.74)
id114	-23.43 ***	(-20.14)	-22.26 ***	(-23.54)	0.637 **	(2.80)	0.537 **	(3.22)
id115	-22.18 ***	(-29.89)	-21.49 ***	(-33.89)	-0.0179	(-0.14)	-0.109	(-0.69)
id116	-21.78 ***	(-24.48)	-3.753 ***	(-7.41)	-1.349 ***	(-4.54)	-1.135 **	(-2.79)
id117	-20.71 ***	(-26.29)	-20.08 ***	(-32.28)	-0.156	(-1.35)	-0.273 *	(-2.17)
id118	-18.97 ***	(-27.81)	-0.705	(-0.91)	0.116	(0.88)	0.00337	(0.02)
id119	-20.31 ***	(-30.19)	-19.89 ***	(-36.33)	-0.145	(-0.79)	-0.143	(-0.92)
id120	-27.27 ***	(-33.83)	-25.62 ***	(-41.44)	-0.0170	(-0.06)	0.133	(0.42)
id121	-28.62 ***	(-36.09)	-0.687	(-1.76)	0.239	(1.20)	0.387 *	(2.01)
id122	-21.69 ***	(-15.37)	-2.515 *	(-2.40)	0.623 *	(2.36)	0.653 *	(2.48)
id123	-23.18 ***	(-12.54)	-3.892 ***	(-4.57)	0.698	(1.51)	0.886 *	(2.13)
id124	-4.268 *	(-2.34)	-2.612 *	(-2.50)	0.698 **	(2.66)	0.361	(1.22)
id125	-3.828 *	(-2.35)	-22.52 ***	(-19.24)	0.296	(0.67)	0.619	(1.69)
id126	-2.023	(-1.24)	-2.183 *	(-2.54)	0.576	(1.84)	0.539	(1.73)
id127	-22.03 ***	(-12.96)	-20.58 ***	(-19.98)	1.158 ***	(4.23)	1.010 ***	(3.82)
id128	-20.90 ***	(-14.97)	-20.01 ***	(-22.10)	0.762 **	(2.84)	0.618 *	(2.18)
id129	-1.540 *	(-2.37)	0.776 *	(2.07)	0.0280	(0.17)	0.165	(0.76)
id130	-24.70 ***	(-30.90)	-23.31 ***	(-35.06)	-1.578 ***	(-5.17)	-1.635 ***	(-4.61)
id131	-1.659 *	(-2.10)	-0.403	(-1.01)	-0.980 ***	(-3.71)	-0.794 ***	(-4.13)
id132	-19.37 ***	(-26.07)	-18.83 ***	(-32.28)	-0.863 ***	(-3.40)	-0.435	(-1.85)
id133	-26.03 ***	(-21.40)	-2.904 ***	(-8.09)	-0.622 ***	(-4.18)	-0.691 ***	(-4.38)
id134	-31.36 ***	(-34.83)	-2.618 ***	(-5.23)	0.488 *	(2.40)	1.176 ***	(6.93)
id135	-23.37 ***	(-24.64)	-22.02 ***	(-35.69)	0.930 ***	(4.58)	1.350 ***	(7.18)
id136	3.358 **	(3.79)	4.212 ***	(6.07)	2.661 ***	(8.38)	2.709 ***	(11.14)
id137	-23.49 ***	(-24.93)	-2.508 **	(-2.65)	0.0440	(0.26)	0.804 ***	(4.16)
id138	-19.02 ***	(-29.14)	-18.74 ***	(-37.30)	-0.827 ***	(-3.86)	-0.890 ***	(-3.54)
id139	-4.105 ***	(-4.05)	-1.187 *	(-2.54)	-0.922 **	(-3.06)	-0.677	(-1.61)
id140	-2.970 **	(-3.06)	-0.615	(-1.10)	-1.276 ***	(-3.79)	-1.035 ***	(-3.61)
id141	-24.65 ***	(-33.01)	-23.40 ***	(-44.51)	-1.071 ***	(-4.15)	-1.100 ***	(-4.52)
id142	-37.04 ***	(-41.13)	-0.873	(-1.69)	0.0500	(0.22)	0.175	(0.76)
id143	-37.41 ***	(-40.91)	-0.397	(-0.82)	-0.368	(-1.15)	-0.157	(-0.52)
id144	-0.585	(-0.65)	0.551	(1.04)	0.0261	(0.09)	0.167	(0.58)
id145	-2.485 *	(-2.36)	-1.273 *	(-2.44)	-0.750	(-1.56)	-0.631	(-1.32)
id146	-22.93 ***	(-26.06)	-1.250 **	(-2.87)	-1.130 **	(-2.83)	-0.758 *	(-1.97)

Table A4. Cont.

Variable	Overspeed		Highspeedbrake		Harshacceleration		Harshdeceleration	
	Coefficient	z	Coefficient	z	Coefficient	z	Coefficient	z
id147	-2.851 ***	(-3.75)	-0.0796	(-0.17)	-1.021 **	(-2.90)	-0.896 **	(-2.59)
id148	-3.737 ***	(-4.05)	1.617 ***	(3.83)	-0.0483	(-0.14)	-0.0520	(-0.16)
id149	-3.202 ***	(-3.53)	-1.184 *	(-2.01)	-0.554 **	(-2.59)	-0.343	(-1.54)
id150	-3.616 **	(-3.24)	-0.905 *	(-2.24)	-1.167 ***	(-4.30)	-0.905 ***	(-3.46)
id151	-0.362	(-0.51)	-1.167 *	(-2.07)	-1.654 ***	(-5.15)	-1.677 ***	(-4.32)
id152	-32.89 ***	(-33.69)	-3.751 ***	(-6.31)	-1.421 ***	(-3.68)	-1.382 ***	(-4.40)
id153	-1.598 *	(-2.14)	-0.169	(-0.46)	-2.936 ***	(-5.34)	-3.067 ***	(-5.01)
id154	-21.84 ***	(-20.17)	-2.716 ***	(-3.85)	-1.703 ***	(-4.38)	-1.483 ***	(-4.26)
id155	-4.238 ***	(-3.67)	-2.441 *	(-2.46)	-0.759 **	(-3.12)	-0.814 **	(-2.63)
id156	-43.11 ***	(-52.10)	-1.456 **	(-3.27)	-0.590 **	(-2.64)	-0.429 *	(-2.03)
id157	-1.868 *	(-2.19)	0.337	(0.62)	-0.753 **	(-2.82)	-0.502 *	(-2.35)
id158	-19.28 ***	(-28.14)	-18.96 ***	(-35.46)	0.678 ***	(4.39)	0.744 ***	(3.87)
id159	-19.26 ***	(-28.40)	-19.02 ***	(-33.84)	0.827 ***	(4.09)	0.715 **	(3.07)
id160	-3.790 ***	(-4.60)	0.550	(1.25)	0.148	(0.72)	0.337	(1.90)
id161	-22.11 ***	(-30.73)	-21.31 ***	(-34.53)	0.608 ***	(4.58)	0.494 ***	(3.29)
id162	-20.15 ***	(-18.45)	-19.53 ***	(-23.73)	0.431	(1.86)	0.176	(0.52)
id163	-22.31 ***	(-22.00)	-21.43 ***	(-29.98)	1.844 ***	(9.58)	1.656 ***	(8.12)
id164	-2.923 **	(-2.69)	-2.557 ***	(-3.68)	-0.245	(-1.66)	-0.301	(-1.85)
id165	-20.82 ***	(-28.93)	-20.41 ***	(-32.12)	1.341 ***	(11.50)	1.447 ***	(10.42)
id166	-25.44 ***	(-20.51)	-2.439 ***	(-4.42)	-0.158	(-0.49)	-0.0534	(-0.15)
id167	-2.696 *	(-2.28)	-4.124 ***	(-4.21)	1.119 ***	(3.43)	1.089 ***	(3.53)
id168	-5.731 ***	(-6.49)	-1.940 ***	(-3.42)	0.0447	(0.23)	0.0115	(0.06)
id169	-26.04 ***	(-24.96)	-24.71 ***	(-35.44)	0.473 *	(2.06)	0.422	(1.82)
id170	-15.15 ***	(-15.82)	-15.11 ***	(-19.48)	-17.48 ***	(-24.79)	-0.684	(-0.99)
id171	-3.650 ***	(-3.49)	-1.497 **	(-2.86)	-0.344	(-1.62)	-0.313	(-1.47)
id172	-3.659 ***	(-3.93)	-1.951 ***	(-4.53)	-0.427	(-1.84)	-0.367	(-1.34)
id173	-3.036 **	(-2.98)	-3.500 ***	(-4.94)	-0.874 ***	(-3.64)	-0.888 ***	(-3.41)
id174	1.453	(1.39)	0.361	(0.38)	-1.484 ***	(-6.68)	-1.288 ***	(-4.64)
id175	-0.688	(-0.99)	1.615 ***	(3.66)	0.114	(0.57)	0.333	(1.64)
id176	-1.666	(-1.82)	-0.313	(-0.81)	0.530	(0.98)	-0.0614	(-0.13)
id177	-2.576 **	(-3.13)	-1.675 ***	(-3.66)	-0.245	(-1.10)	0.187	(0.75)
id178	-0.823	(-0.51)	0.510	(1.04)	0.213	(1.00)	0.0436	(0.18)
id179	-19.54 ***	(-27.56)	-1.071	(-1.13)	-1.386 ***	(-4.24)	-1.021	(-1.91)
id180	-4.457 ***	(-3.85)	-2.934 ***	(-3.90)	-0.402 *	(-2.17)	-0.277	(-1.44)
id181	-1.850 *	(-2.18)	-0.909 *	(-2.21)	-0.573	(-1.78)	-0.354	(-1.41)
id182	-4.754 **	(-3.03)	-2.082 **	(-2.75)	0.387	(1.35)	0.409	(1.60)
log-likelihood	-952.2391		-1519.954		-3479.969		-3488.38	
AIC	2292.478		3427.908		7347.937		7364.76	
BIC	3261.657		4397.086		8317.116		8333.939	
Observation	1092		1092		1092		1092	

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5. Driving risk scores for four near-miss events after winsorizing and Min-Max scaling on regression coefficients.

Variable	Overspeed	Highspeedbrake	Harshacceleration	Harshdeceleration
id1	4.824741	4.344986	2.622834	2.52286
id2	1.242371	4.033808	3.753797	3.75
id3	2.476298	1.628204	4.413078	4.113936
id4	2.578824	1.749502	4.312131	4.315106
id5	1.133814	3.574272	1.557084	1.542612
id6	4.646467	4.258833	3.576023	3.377835
id7	4.434299	4.244682	3.208862	3.286394
id8	2.244711	3.976736	2.4531	2.636247
id9	4.685306	4.334427	2.398606	2.423189
id10	4.449618	4.107521	2.314633	2.288771
id11	4.580368	4.164127	2.478113	2.27414
id12	4.662871	4.224932	2.692603	2.612564
id13	4.542011	4.219333	2.499553	2.404901
id14	4.441416	4.413722	2.060925	2.1891
id15	4.673486	4.370957	2.542969	2.547549

Table A5. Cont.

Variable	Overspeed	Highspeedbrake	Harshacceleration	Harshdeceleration
id16	2.050515	1.186551	2.864928	2.924287
id17	2.12168	4.018102	2.444167	2.612747
id18	2.218175	4.2198	1.618724	1.364301
id19	4.704364	4.404081	3.147222	3.058705
id20	1.835814	3.761974	2.883688	2.607535
id21	2.124092	3.944234	2.910488	2.941661
id22	4.506067	4.309374	3.065928	3.014813
id23	4.729211	4.345159	2.199393	2.148866
id24	1.923866	1.063697	2.328926	2.428676
id25	2.207319	1.317181	1.834912	1.854426
id26	4.494367	4.189475	2.229766	1.912948
id27	1.957639	1.080804	2.581383	2.582846
id28	2.621041	1.695073	2.661426	2.805413
id29	4.687598	4.20938	3.150795	3.094367
id30	2.274866	1.419818	2.42362	2.42959
id31	4.77565	4.246703	1.597284	1.655084
id32	4.432128	3.890427	2.524567	2.653621
id33	2.476298	1.503794	2.464713	2.382955
id34	4.531518	4.10441	2.617715	2.482718
id35	4.362531	4.183099	2.080579	2.018105
id36	4.325984	3.970049	2.368233	2.209217
id37	0.021711	4.153396	2.194033	2.323519
id38	2.317082	4.132869	1.297123	1.497805
id39	0.024124	5	5	5
id40	4.664922	4.280294	2.374486	2.59519
id41	4.53007	4.186365	2.783634	2.664594
id42	4.768412	4.413722	3.165088	2.886796
id43	4.563723	4.379043	2.590763	2.605157
id44	4.527417	4.320727	2.922994	2.950805
id45	4.570236	4.140489	2.716634	2.780724
id46	4.840663	4.270341	2.344113	2.307974
id47	4.468072	4.193363	1.860818	1.923007
id48	4.883362	4.559747	3.363409	3.242502
id49	1.672979	4.100056	2.115419	2.200073
id50	4.51644	4.048426	2.253886	2.22019
id51	2.268835	1.384051	3.62605	3.565289
id52	2.192845	1.124347	2.306593	1.652341
id53	2.260391	1.348283	2.50402	2.705743
id54	4.236002	4.176723	2.010005	1.938552
id55	4.307288	4.075796	3.044488	2.95812
id56	4.527778	4.331519	2.729141	2.867593
id57	2.18802	4.236128	2.721101	3.180322
id58	4.59834	4.237372	2.315526	2.583211
id59	0	4.333961	1.974272	1.986101
id60	4.448773	3.752022	1.987672	2.006218
id61	4.923769	4.512628	1.472217	1.538954
id62	4.768654	4.429895	2.024299	2.090344
id63	2.165103	1.309405	3.86725	3.906364
id64	4.519697	4.079528	3.814544	3.747257
id65	2.171134	1.334287	0.904949	1.345099
id66	4.657202	4.110164	3.073075	3.410753
id67	1.641618	3.816248	2.291406	2.244879

Table A5. Cont.

Variable	Overspeed	Highspeedbrake	Harshacceleration	Harshdeceleration
id68	2.459412	1.607987	1.474004	1.004938
id69	1.78636	1.116571	2.096659	1.951353
id70	2.194051	3.770683	1.292657	1.150329
id71	0.937206	0	3.147222	3.631127
id72	4.157238	4.180455	2.161872	1.934894
id73	4.303307	3.899757	1.25871	1.130212
id74	0.979422	0	2.889941	3.222385
id75	4.478807	4.084194	2.384313	2.439557
id76	2.198876	0.898855	1.152403	0.782736
id77	2.366536	1.433814	3.368769	3.189466
id78	1.845464	3.838019	0.07236	0
id79	4.408728	4.24888	2.41558	2.422275
id80	2.130123	1.197437	2.609523	2.569404
id81	0.595856	4.168947	2.318206	2.215618
id82	4.656961	4.260855	2.344113	2.224762
id83	4.641884	4.369402	2.512953	2.301573
id84	1.823752	4.054802	1.952832	1.741039
id85	2.061371	3.747356	1.345364	1.317666
id86	1.21101	4.028365	1.486511	1.50695
id87	4.601476	4.505785	2.797927	2.911485
id88	2.341206	1.475802	2.43702	2.371982
id89	1.641618	0.869308	3.605503	3.560717
id90	4.341302	4.214201	1.754511	1.624909
id91	2.40634	1.387161	3.226729	3.270849
id92	1.571659	0.774446	3.380382	3.129115
id93	1.873206	1.03415	2.363766	2.248537
id94	4.501001	4.184188	2.48258	2.535571
id95	1.736907	0.925292	2.173486	1.910205
id96	1.999855	1.163225	3.850277	3.750914
id97	2.23868	1.320291	2.208326	2.264996
id98	2.526958	1.600211	1.367697	1.711778
id99	2.583649	1.687298	1.038056	1.271031
id100	4.315007	1.200547	3.339289	3.276335
id101	2.014329	1.197437	3.475969	3.267191
id102	1.72967	0.847537	2.650348	2.495977
id103	2.619835	1.71218	3.107022	3.076993
id104	2.34	1.460251	2.739861	2.762436
id105	4.411502	4.278116	2.208326	1.67703
id106	1.711577	0.852202	2.906021	2.753292
id107	2.215762	1.324956	2.635698	2.657279
id108	1.917835	3.933348	2.147579	1.942209
id109	2.236268	1.290744	2.312846	1.807791
id110	2.288134	1.175666	1.013936	1.568215
id111	4.414035	4.146398	2.777381	2.703914
id112	2.390659	1.522456	1.324817	1.762985
id113	1.147082	3.878919	1.101483	1.337783
id114	1.90336	1.068363	3.191889	3.013899
id115	2.062577	1.197437	2.606843	2.423189
id116	2.119268	3.761352	1.417724	1.485004
id117	2.253154	1.382496	2.483473	2.273226
id118	2.469061	4.235351	2.726461	2.525942

Table A5. Cont.

Variable	Overspeed	Highspeedbrake	Harshacceleration	Harshdeceleration
id119	2.30502	1.421373	2.4933	2.392099
id120	1.435361	0.55051	2.607647	2.644477
id121	1.255639	4.23815	2.836341	2.876737
id122	2.133742	3.953875	3.179382	3.119971
id123	1.987793	3.739736	3.246382	3.333029
id124	4.309941	3.938791	3.246382	2.852963
id125	4.363014	1.066808	2.887261	3.088881
id126	4.58073	4.005505	3.137395	3.015728
id127	2.083082	1.265862	3.657316	3.446416
id128	2.225412	1.262752	3.303555	3.087966
id129	4.638989	4.465819	2.647847	2.673738
id130	1.741732	0.892635	1.21315	1.027798
id131	4.624635	4.282315	1.747365	1.796818
id132	2.42202	1.567554	1.851885	2.125091
id133	1.568041	3.893381	2.067179	1.891002
id134	0.916701	3.937858	3.058781	3.598208
id135	1.906979	1.087024	3.453636	3.757315
id136	5	5	5	5
id137	1.893711	3.954964	2.66214	3.258047
id138	2.485948	1.635979	1.884045	1.709034
id139	4.329602	4.160394	1.799178	1.903804
id140	4.466504	4.249347	1.482937	1.576445
id141	1.759824	0.892635	1.666071	1.517008
id142	0.219526	4.209225	2.6675	2.682882
id143	0.171278	4.283248	2.294086	2.379298
id144	4.754179	4.430673	2.64615	2.675567
id145	4.525004	4.14702	1.952832	1.945867
id146	1.947989	4.150597	1.613364	1.829737
id147	4.480858	4.332608	1.710738	1.703548
id148	4.37399	4.596448	2.579686	2.475311
id149	4.438521	4.160861	2.127926	2.209217
id150	4.388585	4.204249	1.580311	1.695318
id151	4.781077	4.163505	1.145256	0.989393
id152	0.724917	3.761663	1.353404	1.259144
id153	4.631993	4.318705	0	0
id154	2.120474	3.922618	1.101483	1.166789
id155	4.31356	3.965383	1.944792	1.77853
id156	0	4.118562	2.095766	2.130578
id157	4.599426	4.397394	1.950152	2.063826
id158	2.434082	1.548893	3.228515	3.203182
id159	2.434082	1.480468	3.361622	3.176664
id160	4.367597	4.430518	2.755047	2.831017
id161	2.072226	1.250311	3.165982	2.974579
id162	2.325525	1.450921	3.007861	2.683797
id163	2.043278	1.16478	4.270145	4.037125
id164	4.472173	3.947344	2.403966	2.247623
id165	2.233855	1.31096	3.820797	3.846013
id166	1.624732	3.965694	2.481687	2.474031
id167	4.499554	3.703658	3.622476	3.518654
id168	4.133476	4.043294	2.662766	2.533376
id169	1.57769	0.707577	3.045381	2.908742
id170	2.962391	2.184934	0	1.897403
id171	4.384484	4.112186	2.315526	2.23665
id172	4.383398	4.041584	2.241379	2.187271

Table A5. Cont.

Variable	Overspeed	Highspeedbrake	Harshacceleration	Harshdeceleration
id173	4.458543	3.800697	1.842058	1.710863
id174	5	4.401126	1.297123	1.345099
id175	4.741756	4.596137	2.724674	2.827359
id176	4.623791	4.296311	3.096302	2.466715
id177	4.514028	4.084505	2.403966	2.693855
id178	4.725472	4.424297	2.813114	2.562729
id179	2.40634	4.178434	1.38467	1.589247
id180	4.287144	3.888716	2.263713	2.269568
id181	4.601597	4.203627	2.110952	2.199159
id182	4.2512	4.021212	2.968555	2.896854

References

- Guillen, M.; Nielsen, J.P.; Pérez-Marín, A.M. Near-miss telematics in motor insurance. *J. Risk Insur.* **2021**, 1–21. [\[CrossRef\]](#)
- Guillen, M.; Nielsen, J.P.; Pérez-Marín, A.M.; Elpidorou, V. Can automobile insurance telematics predict the risk of near-miss events? *N. Am. Actuar. J.* **2020**, *24*, 141–152. [\[CrossRef\]](#)
- Litman, T. *Distance-Based Vehicle Insurance Feasibility, Costs and Benefits*; Comprehensive Technical Report; Victoria Transport Policy Institute: Victoria, BC, Canada, 2011.
- Tselentis, D.I.; Yannis, G.; Vlahogianni, E.I. Innovative insurance schemes: Pay as/how you drive. *Transp. Res. Procedia* **2016**, *14*, 362–371. [\[CrossRef\]](#)
- Paefgen, J.; Staake, T.; Thiesse, F. Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach. *Decis. Support Syst.* **2013**, *56*, 192–201. [\[CrossRef\]](#)
- Tselentis, D.I.; Yannis, G.; Vlahogianni, E.I. Innovative motor insurance schemes: A review of current practices and emerging challenges. *Accid. Anal. Prev.* **2017**, *98*, 139–148. [\[CrossRef\]](#)
- Troncoso, C.; Danezis, G.; Kosta, E.; Balasch, J.; Preneel, B. Pripayd: Privacy-friendly pay-as-you-drive insurance. *IEEE Trans. Dependable Secur. Comput.* **2010**, *8*, 742–755. [\[CrossRef\]](#)
- Pesantez-Narvaez, J.; Guillen, M.; Alcañiz, M. Predicting Motor Insurance Claims Using Telematics Data—XGBoost versus Logistic Regression. *Risks* **2019**, *7*, 70. [\[CrossRef\]](#)
- Guillen, M.; Nielsen, J.P.; Ayuso, M.; Pérez-Marín, A.M. The use of telematics devices to improve automobile insurance rates. *Risk Anal.* **2019**, *39*, 662–672. [\[CrossRef\]](#)
- Sun, S.; Bi, J.; Guillen, M.; Pérez-Marín, A.M. Assessing driving risk using internet of vehicles data: An analysis based on generalized linear models. *Sensors* **2020**, *20*, 2712. [\[CrossRef\]](#)
- De Diego, I.M.; Siordia, O.S.; Crespo, R.; Conde, C.; Cabello, E. Analysis of hands activity for automatic driving risk detection. *Transp. Res. Part C Emerg. Technol.* **2013**, *26*, 380–395. [\[CrossRef\]](#)
- Siordia, O.S.; de Diego, I.M.; Conde, C.; Cabello, E. Subjective traffic safety experts' knowledge for driving-risk definition. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1823–1834. [\[CrossRef\]](#)
- Charlton, S.G.; Starkey, N.J.; Perrone, J.A.; Isler, R.B. What's the risk? A comparison of actual and perceived driving risk. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *25*, 50–64. [\[CrossRef\]](#)
- Peng, J.; Shao, Y. Intelligent method for identifying driving risk based on V2V multisource big data. *Complexity* **2018**, 2018. [\[CrossRef\]](#)
- Wang, J.; Zheng, Y.; Li, X.; Yu, C.; Kodaka, K.; Li, K. Driving risk assessment using near-crash database through data mining of tree-based model. *Accid. Anal. Prev.* **2015**, *84*, 54–64. [\[CrossRef\]](#)
- Yan, L.; Zhang, Y.; He, Y.; Gao, S.; Zhu, D.; Ran, B.; Wu, Q. Hazardous traffic event detection using Markov Blanket and sequential minimal optimization (MB-SMO). *Sensors* **2016**, *16*, 1084. [\[CrossRef\]](#)
- Liao, Y.; Wang, M.; Duan, L.; Chen, F. Cross-regional driver-vehicle interaction design: An interview study on driving risk perceptions, decisions, and ADAS function preferences. *IET Intell. Transp. Syst.* **2018**, *12*, 801–808. [\[CrossRef\]](#)
- Jiang, K.; Yang, D.; Xie, S.; Xiao, Z.; Victorino, A.C.; Charara, A. Real-time estimation and prediction of tire forces using digital map for driving risk assessment. *Transp. Res. Part C Emerg. Technol.* **2019**, *107*, 463–489. [\[CrossRef\]](#)
- Yan, Y.; Dai, Y.; Li, X.; Tang, J.; Guo, Z. Driving risk assessment using driving behavior data under continuous tunnel environment. *Traffic Inj. Prev.* **2019**, *20*, 807–812. [\[CrossRef\]](#)
- Lu, J.; Xie, X.; Zhang, R. Focusing on appraisals: How and why anger and fear influence driving risk perception. *J. Saf. Res.* **2013**, *45*, 65–73. [\[CrossRef\]](#)
- Wang, J.; Huang, H.; Li, Y.; Zhou, H.; Liu, J.; Xu, Q. Driving risk assessment based on naturalistic driving study and driver attitude questionnaire analysis. *Accid. Anal. Prev.* **2020**, *145*, 105680. [\[CrossRef\]](#)
- Handel, P.; Skog, I.; Wahlstrom, J.; Bonawiede, F.; Welch, R.; Ohlsson, J.; Ohlsson, M. Insurance telematics: Opportunities and challenges with the smartphone solution. *IEEE Intell. Transp. Syst. Mag.* **2014**, *6*, 57–70. [\[CrossRef\]](#)

23. Joubert, J.W.; De Beer, D.; De Koker, N. Combining accelerometer data and contextual variables to evaluate the risk of driver behaviour. *Transp. Res. Part F Traffic Psychol. Behav.* **2016**, *41*, 80–96. [[CrossRef](#)]
24. Verbelen, R.; Antonio, K.; Claeskens, G. Unravelling the predictive power of telematics data in car insurance pricing. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2018**, *67*, 1275–1304. [[CrossRef](#)]
25. Ma, Y.L.; Zhu, X.; Hu, X.; Chiu, Y.C. The use of context-sensitive insurance telematics data in auto insurance rate making. *Transp. Res. Part A Policy Pract.* **2018**, *113*, 243–258. [[CrossRef](#)]
26. Jiang, Y.; Zhang, J.; Wang, Y.; Wang, W. Drivers' behavioral responses to driving risk diagnosis and real-time warning information provision on expressways: A smartphone app-based driving experiment. *J. Transp. Saf. Secur.* **2020**, *12*, 329–357. [[CrossRef](#)]
27. Jin, W.; Deng, Y.; Jiang, H.; Xie, Q.; Shen, W.; Han, W. Latent class analysis of accident risks in usage-based insurance: Evidence from Beijing. *Accid. Anal. Prev.* **2018**, *115*, 79–88. [[CrossRef](#)]
28. Carfora, M.F.; Martinelli, F.; Mercaldo, F.; Nardone, V.; Orlando, A.; Santone, A.; Vaglini, G. A “pay-how-you-drive” car insurance approach through cluster analysis. *Soft Comput.* **2019**, *23*, 2863–2875. [[CrossRef](#)]
29. Burton, A.; Parikh, T.; Mascarenhas, S.; Zhang, J.; Voris, J.; Artan, N.S.; Li, W. Driver identification and authentication with active behavior modeling. In Proceedings of the 2016 12th International Conference on Network and Service Management (CNSM), Montreal, QC, Canada, 31 October–4 November 2016; pp. 388–393.
30. Baecke, P.; Bocca, L. The value of vehicle telematics data in insurance risk selection processes. *Decis. Support Syst.* **2017**, *98*, 69–79. [[CrossRef](#)]
31. Guelman, L. Gradient boosting trees for auto insurance loss cost modeling and prediction. *Expert Syst. Appl.* **2012**, *39*, 3659–3667. [[CrossRef](#)]
32. Bian, Y.; Yang, C.; Zhao, J.L.; Liang, L. Good drivers pay less: A study of usage-based vehicle insurance models. *Transp. Res. Part A Policy Pract.* **2018**, *107*, 20–34. [[CrossRef](#)]
33. Jafarnejad, S.; Castignani, G.; Engel, T. Towards a real-time driver identification mechanism based on driving sensing data. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–7.
34. Paefgen, J.; Staake, T.; Fleisch, E. Multivariate exposure modeling of accident risk: Insights from Pay-as-you-drive insurance data. *Transp. Res. Part A Policy Pract.* **2014**, *61*, 27–40. [[CrossRef](#)]
35. Boucher, J.P.; Pérez-Marín, A.M.; Santolino, M. Pay-as-you-drive insurance: The effect of the kilometers on the risk of accident. In *Anales del Instituto de Actuarios Españoles*; Instituto de Actuarios Españoles: Madrid, Spain, 2013; Volume 19, pp. 135–154.
36. Sun, S.; Bi, J.; Ding, C. Cleaning and Processing on the Electric Vehicle Telematics Data. In Proceedings of the INFORMS International Conference on Service Science, Nanjing, China, 27–29 June 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 1–6.
37. Gao, G.; Wüthrich, M.V.; Yang, H. Evaluation of driving risk at different speeds. *Insur. Math. Econ.* **2019**, *88*, 108–119. [[CrossRef](#)]
38. Gao, G.; Wang, H.; Wüthrich, M.V. Boosting Poisson regression models with telematics car driving data. *Mach. Learn.* **2021**, 1–30. [[CrossRef](#)]
39. So, B.; Boucher, J.P.; Valdez, E.A. Synthetic Dataset Generation of Driver Telematics. *Risks* **2021**, *9*, 58. [[CrossRef](#)]