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

# Abstract

This study proposes a method to identify and evaluate driving risk as the primary step to calculate premiums in the newly emerging context of usage-based insurance. Telematics data obtained from Internet of vehicles contains a large number of near-miss events, which can be regarded as a supplement to claims or accidents, to estimate a driving risk score for each vehicle. Poisson regression and Negative binomial regression are applied to a summary data set with one record per vehicle and to a panel data set of daily vehicle data containing near-miss events; i.e. counts of excess speed, high speed brake, harsh acceleration or deceleration, and additional driving behavior parameters not including accidents. Negative binomial regression performs better than Poisson regression. Vehicles are classified with a driving risk score computed from individual effects of the 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

usage-based insurance; driving risk assessment; driving risk classification; telematics; near-miss event; driving behavior; panel data regression; Poisson; Negative binomial

# Introduction

Near-miss events are incidents that denote the existence of danger even if no accident occurs. In motor insurance, near-miss events provide support for actuarial premium caculations in the auto insurance industry(Guillen et al. 2020). Reporting of near-miss event as an established error reduction technique has been used by many industries to manage risk and reduce accidents. In the auto insurance industry, the insurer traditionally calculate the premium by knowing how many claims the insured person made in the past and rewards with discounts for those drivers that do not report accidents. However, this may be a loose calculation, when the insured has suffered accidents but has not claimed. In other words, it may seem not accurate to use only claims or accidents alone to calculate premiums. Fortunately, the advent of internet of vehicles has given an improved solution to this problem, using near-miss events to identify driving risk and calculate premiums.

This study purposes to explore how to evaluate driving risks and score drivers in the short term without claims and accidents. The model obtained in this study has important significance for driving risk identification. Not only can the model reflect the risk factors of each near-miss event, but the coefficients obtained by the model can also help us evaluate drivers’ risks and rank them according to their risks. The modeling method and results are valuable for insurance companies to develop usage-based insurance (UBI) personalizing premiums and for traffic regulatory authorities to manage drivers and prevent accidents.

Near-miss events need to be extracted from the original data according to the actual requirements for further processing and analysis. Because the original telematics data in this study does not contain claims or accidents, the extraction of near-miss events that are highly relevant to driving patterns is critical. This study was carried out on a summary data set and a panel data set containing four newly extracted near-miss events and several key parameters after data processing. Over speed, high speed braking, harsh acceleration and harsh deceleration, have been defined based on actual driving conditions and local laws and regulations. Since extracted near-miss events were unbounded non-negative integers, Poisson regression and Negative binomial regression are suitable for modeling that dependent variable conforms to this distribution law.

Poisson regression, Negative binomial regression, Zero-inflated Poisson regression and Zero-inflated Negative binomial regression were respectively applied to summary data set. The parameters in the data set, such as average speed, brake times, accelerator pedal position, engine fuel rate etc., were selected as independent variables. In particular, mileage or fuel consumption was chose as exposure variables to offset the impact of non-unit capacity on model accuracy. In order to have a clear understanding of risky factors of different near-miss events, each near-miss event was individually used as dependent variables. However, no matter which one was selected as the dependent variable, Negative binomial regression is the method more suitable than the others for the summary data in this study.

Negative binomial regression also performed better than Poisson regression on panel data sets. Individual effects and time effects were estimated using panel Poisson regression and panel Negative binomial regression on short panel data set of 5 days in length. The regression results not only confirmed the existence of individual effect and time effect but also rated the driving risk of each vehicle. Then, according to these scores, the driving risk level of vehicles can be classified, 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 described the data and introduced the key parameters used in modeling. Section 4 listed the model expression of Poisson regression and Negative binomial regression in this study. Negative binomial regression results on summary data set and panel data set were reported and analysed in Section 5. The results were discussed and the conclusions were presented in Section 6.

# Literature Review

The auto insurance industry has never stopped pursuing new ways to calculate more accurate actuarial premiums. Traditional auto insurance business has been limited by the difficulty of obtaining information of policy holder, which mainly focused on the utilization of basic information of drivers, vehicles and driving sections(Litman 2007). With the continuous progress of information technology, a new type of insurance business, UBI, based on multi-source data and personalized premium calculation is becoming the mainstream. It has experienced the pay-as-you-drive (PAYD) mode of charging premiums according to the driving mileage or fuel consumption which indicates the probability of accidents(Tselentis, Yannis, and Vlahogianni 2016). Then it is transiting to the pay-how-you-drive (PHYD) mode of calculating premiums based on multiple sources of data including driving behavior data that reflect driving risk(Paefgen, Staake, and Thiesse 2013). As the development of 5G communication technology, it may be possible to realize the manage-how-you-drive (MHYD) mode i.e. real-time calculation of premiums based on multi-source data and providing real-time information to drivers to restrain bad driving behavior(Litman 2007; Tselentis, Yannis, and Vlahogianni 2017). However, due to various reasons such as technologies, policies, regulations and privacy(Troncoso et al. 2010), there is even no mature PHYD product on the market at present(Pesantez-Narvaez, Guillen, and Alcañiz 2019; Guillen et al. 2019). As the core issue of UBI, driving risk needs to be further studied to produce products that are more suitable for demands(Sun et al. 2020).

Traffic accidents all over the world cause a large number of casualties every year, and high risk driving is one of the main factors that cause traffic accidents(Litman 2007). Therefore, the research on driving risk has been a hot topic in recent years. Fundamentally, there were simulation experiments in the laboratory setting designed to identify driving risk factors and predict driving risks(Diego et al. 2013; Siordia et al. 2014; Charlton et al. 2014; Peng and Shao 2018). Further, there were real vehicle experiments on real road environment conducted to evaluate driving risk(J. Wang et al. 2015; L. Yan et al. 2016; Liao et al. 2018; K. Jiang et al. 2019; Y. Yan et al. 2019). In addition, there were studies to make a questionnaire survey for driving risk assessment(Lu, Xie, and Zhang 2013; J. Wang et al. 2020). In fact, the naturalistic driving data collected by the Internet of vehicles or smart phones, telematics data, can effectively reduce the influence of subjective factors and unreasonable assumptions, so as to obtain more objective and meaningful research results(Handel et al. 2014; Joubert, De Beer, and De Koker 2016; Verbelen, Antonio, and Claeskens 2018; Ma et al. 2018; Y. Jiang et al. 2020).

In the research of driving risk assessment in the auto insurance industry, machine learning and generalized linear model coexist. With strong ability to process big data efficiently, machine learning is increasingly explored to auto insurance business due to increasing amounts of data. Logistic regression(Jin et al. 2018), cluster analysis(Carfora et al. 2019), decision tree(Paefgen, Staake, and Thiesse 2013), support vector machine(Burton et al. 2016), neural network(Baecke and Bocca 2017) and other machine learning models(Guelman 2012; Bian et al. 2018; Jafarnejad, Castignani, and Engel 2017) have been widely studied in the field of driving risk assessment, and the results also show that machine learning has a good effect in that(Paefgen, Staake, and Fleisch 2014). However, since most machine learning algorithms as black box algorithms do not have good interpretability and stability, they cannot completely replace the conventional generalized linear models in the auto insurance industry(Pesantez-Narvaez, Guillen, and Alcañiz 2019).

Generally, the conventional generalized linear models paid more attention to the correlation between influencing factors and claims or accidents in the study of driving risk(Boucher, Pérez-Marín, and Santolino 2013; Verbelen, Antonio, and Claeskens 2018; Ma et al. 2018; Guillen et al. 2019). Most of the databases used in the study contain claims and accidents. But the study of near-miss events due to the lack of claims and accidents should not be ignored(J. Wang et al. 2015; Guillen et al. 2020), on the contrary, they could be good complements to previous studies and have important significance for driving risk assessment.

To sum up, this study put forward a novel approach of driving risk assessment. In the absence of claims and accidents, the Negative binomial regression results on the panel data set of naturalistic driving data could not only reflect the causal relationship between independent variable driving behavior factors and dependent variable near-miss events, but also provided individual effects as driving risk scores for each observation. This approach gave consideration to both interpretability and practicability, and had reference value for the promotion of UBI.

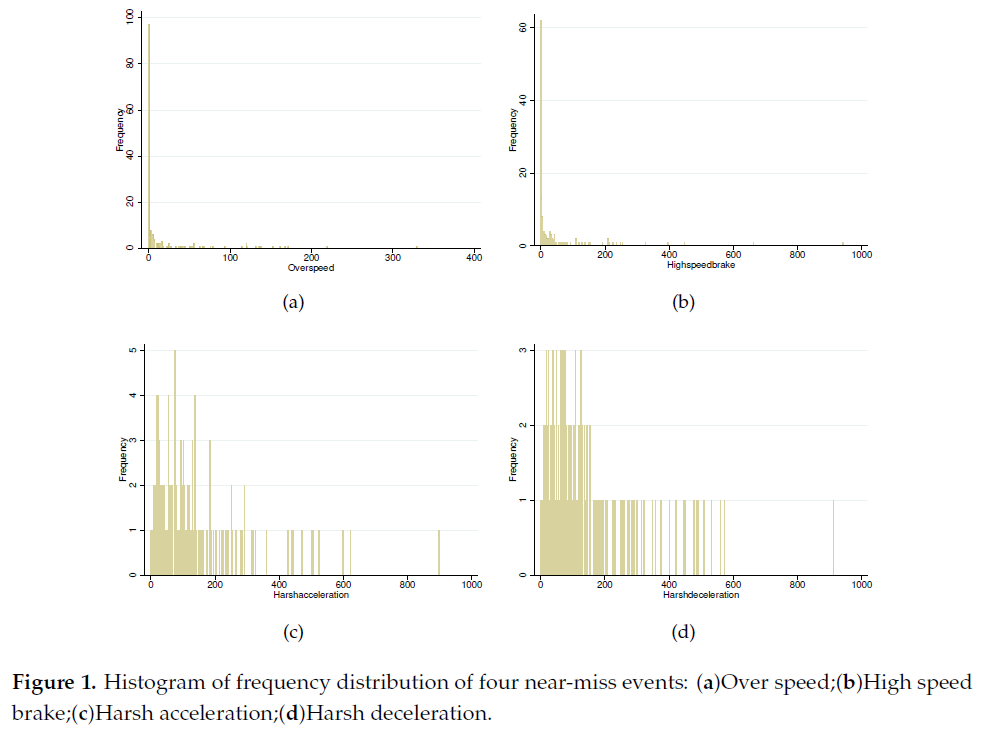
# Data Description

The telematics data used in this study is collected from an internet-of-vehicle information service provider in China. The original data set contains 182 data files representing sensor data for 182 vehicles runs from July 3, 2018 to July 8, 2018(Sun et al. 2020). Each data file contains 62 parameters, but after data processing(Sun, Bi, and Ding 2019), less than one-third of the available parameters related to this study remain. Since the original data cannot be directly used for modeling, the summary data set is aggregated according to the custom statistical rules. It is basically that taking the difference value of the accumulated type, the mean value of the continuous type and the sum value of the discrete type. A summary of the processed data can be found in Table 1. Among them, , , , , , and are obtained according to the above basic statistical rules for the original parameters, while is calculated from the longitude and latitude numerical values(Sun et al. 2020).

**Table 1.** Descriptive statistics of summary data set.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Mean | Standard Deviation | Minimum | Maximum | Defination |
| overspeed | 19.19 | 45.37 | 0 | 330 | Cumulative times of driving speed greater than 90km/h |
| highspeedbrake | 44.23 | 108.3 | 0 | 942 | Cumulative times of braking when the driving speed is greater than 60km/h |
| harshacceleration | 139.0 | 134.7 | 0 | 899 | Cumulative times of the acceleration is greater than 6*m*/*s*2 |
| harshdeceleration | 141.9 | 137.8 | 1 | 913 | Cumulative times of the acceleration is less than 6*m*/*s*2 |
| kilo | 2,223 | 1,674 | 3.73 | 7,164 | Cumulative driving distance (km) |
| fuel | 621.7 | 470.9 | 10.25 | 2,018 | Cumulative fuel consumption (L) |
| brakes | 1,588 | 1,426 | 6 | 9,243 | Cumulative number of brakes |
| range | 5.201 | 5.021 | 0.027 | 26.78 | Range of driving (geographical units) |
| speed | 36.88 | 16.37 | 0.297 | 67.84 | Mean of speed (km/h) |
| rpm | 1,028 | 188.3 | 233.1 | 1,620 | Mean of revolutions per minute (r/min) |
| acceleratorpedalposition | 21.05 | 7.110 | 0.187 | 39.29 | Mean of acceleration pedal position (%) |
| enginefuelrate | 11.52 | 4.464 | 1.868 | 22.01 | Mean of engine fuel rate (%) |
| *a* The number of each parameter is 182. |  |  |  |  |  |

In particular, , , and are individually filtered by combining the rules of traffic law and driving code. Firstly, previous studies have confirmed that over speed is a dangerous driving behavior that is likely to cause traffic accidents(Litman 2007). And China’s traffic safety regulations stipulate a maximum speed for the each type of vehicles on all types of roads. Secondly, if the emergency braking of a car running at a high speed is operated improperly or subjected to lateral force, it is prone to side-slip or even cartwheel, thus high-speed braking is a risky near-miss event worthy of study. Thirdly, both harsh acceleration and harsh deceleration are near-miss events that need to be avoided in terms of driving safety and fuel economy. Previous studies have been referred to in defining the threshold value of harsh acceleration and harsh deceleration(Guillen et al. 2020). It can be seen from Figure [frequency] that near-miss events are all non-negative integers with no upper limit of frequency. Combined with the relationship between expectation and variance shown in Table [summary], the four near-miss events are suitable as dependent variables of Poisson regression and Negative binomial regression.



The panel data set is based on the summary data set. Statistical rules and processing methods are similar to summary data sets, except that each observed value is divided into six observed values in days. The statistics of panel data set are shown in the Table 2.

**Table 2.** Descriptive statistics of panel data set.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | N | Mean | Standard Deviation | Minimum | Maximum |
| overspeed | 1,092 | 3.199 | 14.37 | 0 | 315 |
| highspeedbrake | 1,092 | 7.435 | 21.74 | 0 | 215 |
| harshacceleration | 1,092 | 23.37 | 29.78 | 0 | 223 |
| harshdeceleration | 1,092 | 23.86 | 30.16 | 0 | 233 |
| kilo | 1,092 | 372.6 | 373.2 | 0 | 1,739 |
| fuel | 1,092 | 104.1 | 105.7 | 0 | 565.8 |
| brakes | 1,092 | 264.7 | 291.0 | 0 | 1,940 |
| range | 1,092 | 2.406 | 2.963 | 0 | 14.07 |
| speed | 1,092 | 31.96 | 21.58 | 0 | 77.74 |
| rpm | 1,092 | 894.3 | 346.9 | 0 | 1,731 |
| acceleratorpedalposition | 1,092 | 17.51 | 10.19 | 0 | 45.74 |
| enginefuelrate | 1,092 | 9.794 | 5.835 | 0 | 26.18 |

# Methods

Poisson model and Negative binomial model are both generalized linear models. Generally speaking, Negative binomial regression can be considered as a special case of Poisson regression with over-dispersion of explained variables . The probability density function of the Poisson distribution is:

where is the Poisson arrival rate determined by explanatory variable to represent the average number of events, which is equal to the expectation and variance of the explained variable .

Negative binomial is a mixture of a Poisson() and a Gamma(,). The probability density function of the Negative binomial distribution follows the Poisson distribution:

where and .

The conditional expectation function of Negative binomial distribution, which is similar to Poisson regression, is:

where i is the serial number of the observation, k depends on the number of independent variables, j depends on the existence of time effect and individual effect, takes or as the exposure variable, … represent the independent variables such as , , , , and , dummy variable … represent the time effect and individual effect, … and … are unknown parameters that need to be estimated, random variables represents the heterogeneity of individuals in the conditional expectation function. Regression to Poisson model and Negative binomial model can be accomplished with STATA or Python.

# Results

Both Poisson regression and Negative binomial regression were applicable to this study, and the Zero-inflated model was taken as a consideration for the large number of zero values of dependent variables. In order to determine the most suitable model for this study, the performance of models on different dependent variables was compared. Furthermore, the fittest model on different data performed different.

## Results of summary data set

In summary data set, four near-miss events were respectively treated as dependent variables. Parameters , , , and were taken as independent variables. And parameter was chosen as exposure variable. Poisson regression, Zero-inflated Poisson regression, Negative binomial regression and Zero-inflated Negative binomial regression were conducted. The regression effect of the four models on each dependent variable is shown in Table 3. It indicates that no matter which near-miss event is the dependent variable, Negative binomial regression has minimum log-likelihood value, AIC value and BIC value. That is, Negative binomial regression has the best performance and is most suitable for modeling in this data set.

**Table 3.** 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 |
| POS | 182 | -3518.92 | 7 | 7051.846 | 7074.274 |
| overspeed 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 |
| POS | 182 | -2830.75 | 7 | 5675.498 | 5697.926 |
| highspeedbrake 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 |
| POS | 182 | -5857.26 | 7 | 11728.51 | 11750.94 |
| harshacceleration ZIP | 182 | -5857.26 | 8 | 11730.51 | 11756.14 |
| NB | 182 | -1032.81 | 8 | 2081.623 | 2107.255 |
| ZINB | 182 | -1032.81 | 9 | 2083.623 | 2112.459 |
| POS | 182 | -6269.47 | 7 | 12552.93 | 12575.36 |
| harshdeceleration ZIP | 182 | -6269.47 | 8 | 12554.93 | 12580.56 |
| NB | 182 | -1037.14 | 8 | 2090.285 | 2115.917 |
| ZINB | 182 | -1037.14 | 9 | 2092.285 | 2121.121 |

According to the results of Negative binomial regression in different dependent variables (seeing Table 4), different near-miss events are affected by different driving risk factors with different influences. Relatively speaking, the number of braking has the most obvious influence on near-miss events, it has a significant positive effect on high speed braking(0.000191), harsh acceleration(0.000133) and harsh deceleration(0.000126). The impact of average speed on near-miss events is also significant. The higher the average driving speed, the less rapid acceleration(-0.0474) and rapid deceleration(-0.0402) occur. In addition, average RPM is positively correlated with harsh acceleration(0.000947), and average accelerator pedal position is positively correlated with harsh acceleration(0.0214) and harsh deceleration(0.0330). Interestingly, some influencing factors have opposite effects on different dependent variables. Range of driving has positive effect on high speed brake(0.0541) but negative effect on harsh deceleration(-0.0305). And average engine fuel rate has a significant positive effect on high speed braking(0.158) but a negative effect on sharp deceleration(-0.0351). What’s more, 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.

**Table 4.** Negative binomial regression results for four near-miss events.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | overspeed | highspeedbrake | harshacceleration | harshdeceleration |
| Constant | -7.536\*\*\* | -8.456\*\*\* | -2.101\*\*\* | -1.903\*\*\* |
|  | (-3.363) | (-7.526) | (-4.006) | (-3.933) |
| brakes | 0.000185 | 0.000191\*\*\* | 0.000133\*\*\* | 0.000126\*\*\* |
|  | (1.293) | (2.601) | (3.384) | (3.450) |
| range | 0.0369 | 0.0541\*\* | -0.0200 | -0.0305\* |
|  | (0.791) | (2.052) | (-1.287) | (-1.942) |
| speed | -0.00690 | 0.0152 | -0.0474\*\*\* | -0.0402\*\*\* |
|  | (-0.200) | (1.277) | (-8.810) | (-7.201) |
| rpm | 0.000666 | -0.000128 | 0.000947\* | 0.000515 |
|  | (0.431) | (-0.113) | (1.896) | (1.072) |
| acceleratorpedalposition | 0.0407 | 0.0241 | 0.0214\* | 0.0330\*\*\* |
|  | (1.130) | (1.028) | (1.872) | (2.815) |
| enginefuelrate | 0.0508 | 0.158\*\*\* | -0.0198 | -0.0351\*\* |
|  | (0.987) | (4.493) | (-1.116) | (-2.073) |
| log-likelihood | -490.5169 | -627.4217 | -1032.811 | -1037.142 |
| AIC | 997.0338 | 1270.843 | 2081.623 | 2090.285 |
| BIC | 1022.666 | 1296.476 | 2107.255 | 2115.917 |
| Observations | 182 | 182 | 182 | 182 |
| *a* Robust z-statistics in parentheses | *b* \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 |  |  |  |

## Results of panel data set

As shown in Table [comparison\_xt], the evaluation index(log-likelihood, AIC and BIC) of Negative binomial regression is lower than that of Poisson regression under each dependent variable. Therefore, Negative binomial regression is better than Poisson regression on panel data.

**Table 5.** Model performances of Poisson and Negative binomial in panel data set.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 | 12610.89 | 13550.09 |
| XTNB | 1092 | -3526.09 | 189 | 7430.186 | 8374.386 |
| harshdeceleration XTPOS | 1092 | -6042.02 | 188 | 12460.03 | 13399.24 |
| XTNB | 1092 | -3547.66 | 189 | 7473.311 | 8417.51 |

The panel Negative binomial regression was 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 it (seeing Table A1) are not all the same as the results of summary data. What remains is that rapid acceleration and rapid deceleration are positively affected by the number of brakes(0.000845&0.000869) and average accelerator pedal position(0.0244&0.0265) but negatively affected by the average speed(-0.0299&-0.0272) and average engine fuel rate(-0.0323&-0.0392). However, RPM which is not significant in the summary data is significantly positive for over speed(0.00485) and high speed braking(0.00371).

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

# Discussions and Conclusions

In this study, driving risks can be evaluated by the regression coefficient of Negative binomial models on panel data. Four near-miss events have been used as dependent variables to obtain four sets of regression coefficients. Given the influencing factors and generating mechanisms of different near-miss events are different, so it is not recommended to combine the four groups of regression coefficients into one group. Within a group, a higher coefficient means a higher probability of the near-miss event.

In order to more intuitively reflect the risk relationship between observations, it is suggested to carry out driving risk grading. Firstly, winsorization could be done to avoid the influence of possibly spurious outliers (the double tail was winsorized with the threshold 0.01 in this study). Secondly, the regression coefficient need to be compressed to the interval of [0,1] through normalization. Then, the corresponding grade will be obtained by enlarging corresponding multiple according to actual demands. For example, we mapped each group of coefficients into an interval of [0,5] (seeing Table A2), each observation got a driving risk level from 1 to 5, i.e. excellent, good, medium, bad and terrible (seeing Figure 2). To be clear, the values of 0 and 5 are because the corresponding observations are the minimum and maximum values in their group and are Min-Max scaled. In and groups, two types of observations with high risk or low risk can be clearly seen. It indicates that these two near-miss events are more sensitive to driving behavior than and and can be considered with 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 will be charged individually according to the driving risk level of the insured one.

The number and type of dependent variables and independent variables selected in this study are limited by the size and quality of 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, near-miss events as dependent variables could be increased or decreased flexibly according to needs. For example, it is recommended to include sharp turn as a near-miss event if possible, because sharp turn is highly studied and accident-proven patterns of high driving risk. For the same, more driving behavior parameters 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 factor, 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.

**Figure 2.** Driving risk rank of four near-miss events.

In practical applications, near-miss events can be combined with claims and accidents to accurately evaluate driving risks. The study proves that near-miss events can be used as driving risk score when there is no claims and accidents. However, when claims or accidents exist, it is recommended to adopt the driving risk evaluation strategy as follows. 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 general, near-miss events can provide insurers with effective risk information in the absence of claims and accident data. Negative binomial regression is the most suitable modeling method for near-miss events as dependent variables. Negative binomial regression can score and rate the driving risk for the insured, and the evaluation result can help the insurer to actualize the auto insurance premium. This study provides a technical reference for the promotion and development of PHYD mode.

# Appendix A

**Table A1.** Panel Negative binomial regression results for four near-miss events.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | overspeed | highspeedbrake | harshacceleration | harshdeceleration |
| Constant | -8.820\*\*\* | -8.364\*\*\* | -2.029\*\*\* | -2.035\*\*\* |
| brakes | -0.000138 | 0.000174 | 0.000845\*\*\* | 0.000869\*\*\* |
| range | -0.0215 | -0.00365 | -0.00139 | -0.0201 |
| speed | -0.00188 | -0.00447 | -0.0299\*\*\* | -0.0272\*\*\* |
| rpm | 0.00485\*\* | 0.00371\*\*\* | 0.000412 | 0.000417 |
| acceleratorpedalposition | 0.0384 | 0.0172 | 0.0244\*\* | 0.0265\*\* |
| enginefuelrate | 0.0193 | 0.0505 | -0.0323\* | -0.0392\*\* |
| 2018-07-04 | 0.273 | 0.216\* | -0.111\*\* | -0.216\*\*\* |
| 2018-07-05 | -0.168 | -0.0572 | -0.206\*\*\* | -0.317\*\*\* |
| 2018-07-06 | -0.00718 | -0.228\*\* | -0.257\*\*\* | -0.370\*\*\* |
| 2018-07-07 | -0.477\*\* | -0.200\* | -0.485\*\*\* | -0.600\*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2018-07-08 | 0.206 | 0.117 | -0.694\*\*\* | -0.784\*\*\* |
| id2 | -29.70\*\*\* | -2.001\*\* | 1.266\*\*\* | 1.342\*\*\* |
| id3 | -19.47\*\*\* | -17.47\*\*\* | 2.004\*\* | 1.740\*\*\* |
| id4 | -18.62\*\*\* | -16.69\*\*\* | 1.891\*\*\* | 1.960\*\*\* |
| id5 | -30.60\*\*\* | -4.956\*\*\* | -1.193\*\*\* | -1.072\*\*\* |
| id6 | -1.478\*\* | -0.554\* | 1.067\*\*\* | 0.935\*\*\* |
| id7 | -3.237\*\*\* | -0.645 | 0.656\*\*\* | 0.835\*\*\* |
| id8 | -21.39\*\*\* | -2.368\*\*\* | -0.190 | 0.124 |
| id9 | -1.156 | -0.0679 | -0.251 | -0.109 |
| id10 | -3.110\*\*\* | -1.527\*\*\* | -0.345\*\* | -0.256 |
| id11 | -2.026\*\* | -1.163\*\*\* | -0.162 | -0.272 |
| id12 | -1.342\*\* | -0.772\*\* | 0.0781 | 0.0981 |
| id13 | -2.344\*\*\* | -0.808\*\* | -0.138 | -0.129 |
| id14 | -3.178\*\*\* | 0.442 | -0.629\*\*\* | -0.365\*\* |
| id15 | -1.254\*\* | 0.167 | -0.0894 | 0.0270 |
| id16 | -23.00\*\*\* | -20.31\*\*\* | 0.271 | 0.439\*\* |
| id17 | -22.41\*\*\* | -2.102\*\*\* | -0.200 | 0.0983 |
| id18 | -21.61\*\*\* | -0.805 | -1.124\*\*\* | -1.267\*\*\* |
| id19 | -0.998 | 0.380 | 0.587\*\*\* | 0.586\*\*\* |
| id20 | -24.78\*\*\* | -3.749\*\*\* | 0.292 | 0.0926 |
| id21 | -22.39\*\*\* | -2.577\*\*\* | 0.322 | 0.458\*\*\* |
| id22 | -2.642\*\*\* | -0.229 | 0.496\*\*\* | 0.538\*\*\* |
| id23 | -0.792 | 0.00111 | -0.474 | -0.409\* |
| id24 | -24.05\*\*\* | -21.10\*\*\* | -0.329 | -0.103 |
| id25 | -21.70\*\*\* | -19.47\*\*\* | -0.882\*\*\* | -0.731\*\* |
| id26 | -2.739\*\*\* | -1.000\*\*\* | -0.440\* | -0.667\*\*\* |
| id27 | -23.77\*\*\* | -20.99\*\*\* | -0.0464 | 0.0656 |
| id28 | -18.27\*\*\* | -17.04\*\*\* | 0.0432 | 0.309 |
| id29 | -1.137 | -0.872\*\* | 0.591\*\*\* | 0.625\*\*\* |
| id30 | -21.14\*\*\* | -18.81\*\*\* | -0.223 | -0.102 |
| id31 | -0.407 | -0.632\*\* | -1.148\*\*\* | -0.949\*\*\* |
| id32 | -3.255\*\*\* | -2.923\*\* | -0.110 | 0.143 |
| id33 | -19.47\*\*\* | -18.27\*\*\* | -0.177 | -0.153 |
| id34 | -2.431\*\*\* | -1.547\*\*\* | -0.00573 | -0.0439 |
| id35 | -3.832\*\*\* | -1.041\*\* | -0.607\*\*\* | -0.552\*\*\* |
| id36 | -4.135\*\*\* | -2.411\*\*\* | -0.285 | -0.343\* |
| id37 | -39.82\*\*\* | -1.232\* | -0.480 | -0.218 |
| id38 | -20.79\*\*\* | -1.364\*\* | -1.484\*\*\* | -1.121\*\*\* |
| id39 | -39.80\*\*\* | 10.89\*\*\* | 11.65\*\*\* | 11.77\*\*\* |
| id40 | -1.325 | -0.416 | -0.278 | 0.0791 |
| id41 | -2.443\*\*\* | -1.020\*\* | 0.180 | 0.155 |
| id42 | -0.467 | 0.442 | 0.607 | 0.398 |
| id43 | -2.164\*\* | 0.219 | -0.0359 | 0.0900 |
| id44 | -2.465\*\*\* | -0.156 | 0.336 | 0.468\* |
| id45 | -2.110\*\*\* | -1.315\*\*\* | 0.105 | 0.282 |
| id46 | 0.132 | -0.480 | -0.312\*\*\* | -0.235\* |
| id47 | -2.957\*\*\* | -0.975 | -0.853\*\*\* | -0.656\*\*\* |
| id48 | 0.486 | 1.381\*\*\* | 0.829\*\*\* | 0.787\*\*\* |
| id49 | -26.13\*\*\* | -1.575\*\*\* | -0.568\*\*\* | -0.353\* |
| id50 | -2.556\*\*\* | -1.907\*\*\* | -0.413\*\* | -0.331\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id51 | -21.19\*\*\* | -19.04\*\*\* | 1.123\*\*\* | 1.140\*\*\* |
| id52 | -21.82\*\*\* | -20.71\*\*\* | -0.354 | -0.952\*\*\* |
| id53 | -21.26\*\*\* | -19.27\*\*\* | -0.133 | 0.200 |
| id54 | -4.881\*\*\* | -1.082\*\*\* | -0.686\*\*\* | -0.639\*\*\* |
| id55 | -4.290\*\*\* | -1.731\*\*\* | 0.472\* | 0.476 |
| id56 | -2.462\*\*\* | -0.0866 | 0.119 | 0.377 |
| id57 | -21.86\*\*\* | -0.700 | 0.110 | 0.719\*\* |
| id58 | -1.877\* | -0.692 | -0.344 | 0.0660 |
| id59 | -40.00\*\*\* | -0.0709 | -0.726\*\* | -0.587\* |
| id60 | -3.117\*\*\* | -3.813\*\*\* | -0.711\*\* | -0.565\* |
| id61 | 0.821 | 1.078\* | -1.288\*\*\* | -1.076\*\* |
| id62 | -0.465 | 0.546 | -0.670 | -0.473 |
| id63 | -22.05\*\*\* | -19.52\*\*\* | 1.393\*\*\* | 1.513\*\*\* |
| id64 | -2.529 | -1.707 | 1.334\*\*\* | 1.339\*\*\* |
| id65 | -22.00\*\*\* | -19.36\*\*\* | -1.923\*\*\* | -1.288\*\*\* |
| id66 | -1.389 | -1.510\*\*\* | 0.504\*\*\* | 0.971\*\*\* |
| id67 | -26.39\*\*\* | -3.400\*\*\* | -0.371\*\* | -0.304\* |
| id68 | -19.61\*\*\* | -17.60\*\*\* | -1.286\*\*\* | -1.660\*\*\* |
| id69 | -25.19\*\*\* | -20.76\*\*\* | -0.589\*\*\* | -0.625\*\* |
| id70 | -21.81\*\*\* | -3.693\*\*\* | -1.489\*\*\* | -1.501\*\*\* |
| id71 | -32.23\*\*\* | -28.28\*\*\* | 0.587\*\*\* | 1.212\*\*\* |
| id72 | -5.534\*\*\* | -1.058\* | -0.516 | -0.643\* |
| id73 | -4.323\*\*\* | -2.863\*\*\* | -1.527\*\*\* | -1.523\*\*\* |
| id74 | -31.88\*\*\* | -27.94\*\*\* | 0.299 | 0.765\*\*\* |
| id75 | -2.868\*\*\* | -1.677\*\*\* | -0.267 | -0.0911 |
| id76 | -21.77\*\*\* | -22.16\*\*\* | -1.646\*\*\* | -1.903\*\*\* |
| id77 | -20.38\*\*\* | -18.72\*\*\* | 0.835\*\*\* | 0.729\*\*\* |
| id78 | -24.70\*\*\* | -3.260\*\*\* | -2.855\*\*\* | -2.759\*\*\* |
| id79 | -3.449\*\*\* | -0.618 | -0.232 | -0.110 |
| id80 | -22.34\*\*\* | -20.24\*\*\* | -0.0149 | 0.0509 |
| id81 | -35.06\*\*\* | -1.132\*\* | -0.341\*\* | -0.336\*\* |
| id82 | -1.391 | -0.541 | -0.312 | -0.326 |
| id83 | -1.516\*\*\* | 0.157 | -0.123 | -0.242 |
| id84 | -24.88\*\*\* | -1.866\*\* | -0.750\*\*\* | -0.855\*\*\* |
| id85 | -22.91\*\*\* | -3.843\*\*\* | -1.430\*\*\* | -1.318\*\*\* |
| id86 | -29.96\*\*\* | -2.036\*\*\* | -1.272\*\*\* | -1.111\*\*\* |
| id87 | -1.851\*\* | 1.034\*\*\* | 0.196 | 0.425\*\* |
| id88 | -20.59\*\*\* | -18.45\*\*\* | -0.208 | -0.165 |
| id89 | -26.39\*\*\* | -22.35\*\*\* | 1.100\*\* | 1.135\*\* |
| id90 | -4.008\*\*\* | -0.841 | -0.972\*\*\* | -0.982\*\*\* |
| id91 | -20.05\*\*\* | -19.02\*\*\* | 0.676\*\*\* | 0.818\*\*\* |
| id92 | -26.97\*\*\* | -22.96\*\*\* | 0.848\*\* | 0.663\*\* |
| id93 | -24.47\*\*\* | -21.29\*\*\* | -0.290 | -0.300 |
| id94 | -2.684\*\*\* | -1.034\*\*\* | -0.157 | 0.0139 |
| id95 | -25.60\*\*\* | -21.99\*\*\* | -0.503 | -0.670 |
| id96 | -23.42\*\*\* | -20.46\*\*\* | 1.374\*\*\* | 1.343\*\*\* |
| id97 | -21.44\*\*\* | -19.45\*\*\* | -0.464\*\* | -0.282 |
| id98 | -19.05\*\*\* | -17.65\*\*\* | -1.405\*\*\* | -0.887\*\* |
| id99 | -18.58\*\*\* | -17.09\*\*\* | -1.774\*\*\* | -1.369\*\*\* |
| id100 | -4.226\*\*\* | -20.22\*\*\* | 0.802\*\* | 0.824\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id101 | -23.30\*\*\* | -20.24\*\*\* | 0.955\*\* | 0.814\* |
| id102 | -25.66\*\*\* | -22.49\*\*\* | 0.0308 | -0.0294 |
| id103 | -18.28\*\*\* | -16.93\*\*\* | 0.542\*\* | 0.606\*\*\* |
| id104 | -20.60\*\*\* | -18.55\*\*\* | 0.131 | 0.262\*\* |
| id105 | -3.426\*\*\* | -0.430 | -0.464 | -0.925\*\* |
| id106 | -25.81\*\*\* | -22.46\*\*\* | 0.317\* | 0.252 |
| id107 | -21.63\*\*\* | -19.42\*\*\* | 0.0144 | 0.147 |
| id108 | -24.10\*\*\* | -2.647\*\*\* | -0.532\*\* | -0.635\*\*\* |
| id109 | -21.46\*\*\* | -19.64\*\*\* | -0.347\*\*\* | -0.782\*\*\* |
| id110 | -21.03\*\*\* | -20.38\*\*\* | -1.801\*\*\* | -1.044\*\*\* |
| id111 | -3.405\*\*\* | -1.277\*\*\* | 0.173 | 0.198 |
| id112 | -20.18\*\*\* | -18.15\*\*\* | -1.453\*\*\* | -0.831\*\*\* |
| id113 | -30.49\*\*\* | -2.997\*\*\* | -1.703\*\*\* | -1.296\*\*\* |
| id114 | -24.22\*\*\* | -21.07\*\*\* | 0.637\*\*\* | 0.537\*\*\* |
| id115 | -22.90\*\*\* | -20.24\*\*\* | -0.0179 | -0.109 |
| id116 | -22.43\*\*\* | -3.753\*\*\* | -1.349\*\*\* | -1.135\*\*\* |
| id117 | -21.32\*\*\* | -19.05\*\*\* | -0.156 | -0.273\*\* |
| id118 | -19.53\*\*\* | -0.705 | 0.116 | 0.00337 |
| id119 | -20.89\*\*\* | -18.80\*\*\* | -0.145 | -0.143 |
| id120 | -28.10\*\*\* | -24.40\*\*\* | -0.0170 | 0.133 |
| id121 | -29.59\*\*\* | -0.687\* | 0.239 | 0.387\*\* |
| id122 | -22.31\*\*\* | -2.515\*\* | 0.623\*\* | 0.653\*\* |
| id123 | -23.52\*\*\* | -3.892\*\*\* | 0.698 | 0.886\*\* |
| id124 | -4.268\*\* | -2.612\*\* | 0.698\*\*\* | 0.361 |
| id125 | -3.828\*\* | -21.08\*\*\* | 0.296 | 0.619\* |
| id126 | -2.023 | -2.183\*\* | 0.576\* | 0.539\* |
| id127 | -22.73\*\*\* | -19.80\*\*\* | 1.158\*\*\* | 1.010\*\*\* |
| id128 | -21.55\*\*\* | -19.82\*\*\* | 0.762\*\*\* | 0.618\*\* |
| id129 | -1.540\*\* | 0.777\*\* | 0.0280 | 0.165 |
| id130 | -25.56\*\*\* | -22.20\*\*\* | -1.578\*\*\* | -1.635\*\*\* |
| id131 | -1.659\*\* | -0.403 | -0.980\*\*\* | -0.794\*\*\* |
| id132 | -19.92\*\*\* | -17.86\*\*\* | -0.863\*\*\* | -0.435\* |
| id133 | -27.00\*\*\* | -2.904\*\*\* | -0.622\*\*\* | -0.691\*\*\* |
| id134 | -32.40\*\*\* | -2.618\*\*\* | 0.488\*\* | 1.176\*\*\* |
| id135 | -24.19\*\*\* | -20.95\*\*\* | 0.930\*\*\* | 1.350\*\*\* |
| id136 | 3.358\*\*\* | 4.212\*\*\* | 2.661\*\*\* | 2.709\*\*\* |
| id137 | -24.30\*\*\* | -2.508\*\*\* | 0.0440 | 0.804\*\*\* |
| id138 | -19.39\*\*\* | -17.42\*\*\* | -0.827\*\*\* | -0.890\*\*\* |
| id139 | -4.105\*\*\* | -1.187\*\* | -0.922\*\*\* | -0.677 |
| id140 | -2.970\*\*\* | -0.615 | -1.276\*\*\* | -1.035\*\*\* |
| id141 | -25.41\*\*\* | -22.20\*\*\* | -1.071\*\*\* | -1.100\*\*\* |
| id142 | -38.18\*\*\* | -0.873\* | 0.0500 | 0.175 |
| id143 | -38.58\*\*\* | -0.397 | -0.368 | -0.157 |
| id144 | -0.585 | 0.551 | 0.0261 | 0.167 |
| id145 | -2.485\*\* | -1.273\*\* | -0.750 | -0.631 |
| id146 | -23.85\*\*\* | -1.250\*\*\* | -1.130\*\*\* | -0.758\*\* |
| id147 | -2.851\*\*\* | -0.0796 | -1.021\*\*\* | -0.896\*\*\* |
| id148 | -3.737\*\*\* | 1.617\*\*\* | -0.0483 | -0.0520 |
| id149 | -3.202\*\*\* | -1.184\*\* | -0.554\*\*\* | -0.343 |
| id150 | -3.616\*\*\* | -0.905\*\* | -1.167\*\*\* | -0.905\*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id151 | -0.362 | -1.167\*\* | -1.654\*\*\* | -1.677\*\*\* |
| id152 | -33.99\*\*\* | -3.751\*\*\* | -1.421\*\*\* | -1.382\*\*\* |
| id153 | -1.598\*\* | -0.169 | -2.936\*\*\* | -3.067\*\*\* |
| id154 | -22.42\*\*\* | -2.716\*\*\* | -1.703\*\*\* | -1.483\*\*\* |
| id155 | -4.238\*\*\* | -2.441\*\* | -0.759\*\*\* | -0.814\*\*\* |
| id156 | -44.40\*\*\* | -1.456\*\*\* | -0.590\*\*\* | -0.429\*\* |
| id157 | -1.868\*\* | 0.337 | -0.753\*\*\* | -0.502\*\* |
| id158 | -19.82\*\*\* | -17.98\*\*\* | 0.678\*\*\* | 0.744\*\*\* |
| id159 | -19.82\*\*\* | -18.42\*\*\* | 0.827\*\*\* | 0.715\*\*\* |
| id160 | -3.790\*\*\* | 0.550 | 0.148 | 0.337\* |
| id161 | -22.82\*\*\* | -19.90\*\*\* | 0.608\*\*\* | 0.494\*\*\* |
| id162 | -20.72\*\*\* | -18.61\*\*\* | 0.431\* | 0.176 |
| id163 | -23.06\*\*\* | -20.45\*\*\* | 1.844\*\*\* | 1.656\*\*\* |
| id164 | -2.923\*\*\* | -2.557\*\*\* | -0.245\* | -0.301\* |
| id165 | -21.48\*\*\* | -19.51\*\*\* | 1.341\*\*\* | 1.447\*\*\* |
| id166 | -26.53\*\*\* | -2.439\*\*\* | -0.158 | -0.0534 |
| id167 | -2.696\*\* | -4.124\*\*\* | 1.119\*\*\* | 1.089\*\*\* |
| id168 | -5.731\*\*\* | -1.940\*\*\* | 0.0447 | 0.0115 |
| id169 | -26.92\*\*\* | -23.39\*\*\* | 0.473\*\* | 0.422\* |
| id170 | -15.44\*\*\* | -13.89\*\*\* | -19.55\*\*\* | -0.684 |
| id171 | -3.650\*\*\* | -1.497\*\*\* | -0.344 | -0.313 |
| id172 | -3.659\*\*\* | -1.951\*\*\* | -0.427\* | -0.367 |
| id173 | -3.036\*\*\* | -3.500\*\*\* | -0.874\*\*\* | -0.888\*\*\* |
| id174 | 1.453 | 0.361 | -1.484\*\*\* | -1.288\*\*\* |
| id175 | -0.688 | 1.615\*\*\* | 0.114 | 0.333 |
| id176 | -1.666\* | -0.313 | 0.530 | -0.0614 |
| id177 | -2.576\*\*\* | -1.675\*\*\* | -0.245 | 0.187 |
| id178 | -0.823 | 0.510 | 0.213 | 0.0436 |
| id179 | -20.05\*\*\* | -1.071 | -1.386\*\*\* | -1.021\* |
| id180 | -4.457\*\*\* | -2.934\*\*\* | -0.402\*\* | -0.277 |
| id181 | -1.850\*\* | -0.909\*\* | -0.573\* | -0.354 |
| id182 | -4.755\*\*\* | -2.082\*\*\* | 0.387 | 0.409 |
| 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 |
| Observations | 1,092 | 1,092 | 1,092 | 1,092 |
| *a* \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 |  |  |  |  |

**Table A2.** Driving risk scores for four near-miss events after winsorizing and Min-Max scaling on regression coefficients.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| VARIABLES | 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 |
| 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 |
| 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 |
| 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 |
| 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

Baecke, Philippe, and Lorenzo Bocca. 2017. “The Value of Vehicle Telematics Data in Insurance Risk Selection Processes.” *Decision Support Systems* 98. Elsevier: 69–79.

Bian, Yiyang, Chen Yang, J Leon Zhao, and Liang Liang. 2018. “Good Drivers Pay Less: A Study of Usage-Based Vehicle Insurance Models.” *Transportation Research Part A: Policy and Practice* 107. Elsevier: 20–34.

Boucher, Jean-Philippe, Ana Maria Pérez-Marín, and Miguel Santolino. 2013. “Pay-as-You-Drive Insurance: The Effect of the Kilometers on the Risk of Accident.” In *Anales Del Instituto de Actuarios Españoles*, 19:135–54. Instituto de Actuarios Españoles.

Burton, Angela, Tapan Parikh, Shannon Mascarenhas, Jue Zhang, Jonathan Voris, N Sertac Artan, and Wenjia Li. 2016. “Driver Identification and Authentication with Active Behavior Modeling.” In *2016 12th International Conference on Network and Service Management (Cnsm)*, 388–93. IEEE.

Carfora, Maria Francesca, Fabio Martinelli, Francesco Mercaldo, Vittoria Nardone, Albina Orlando, Antonella Santone, and Gigliola Vaglini. 2019. “A ‘Pay-How-You-Drive’ Car Insurance Approach Through Cluster Analysis.” *Soft Computing* 23 (9). Springer: 2863–75.

Charlton, Samuel G, Nicola J Starkey, John A Perrone, and Robert B Isler. 2014. “What’s the Risk? A Comparison of Actual and Perceived Driving Risk.” *Transportation Research Part F: Traffic Psychology and Behaviour* 25. Elsevier: 50–64.

Diego, Isaac Martín de, Oscar S Siordia, Raúl Crespo, Cristina Conde, and Enrique Cabello. 2013. “Analysis of Hands Activity for Automatic Driving Risk Detection.” *Transportation Research Part C: Emerging Technologies* 26. Elsevier: 380–95.

Guelman, Leo. 2012. “Gradient Boosting Trees for Auto Insurance Loss Cost Modeling and Prediction.” *Expert Systems with Applications* 39 (3). Elsevier: 3659–67.

Guillen, Montserrat, Jens Perch Nielsen, Mercedes Ayuso, and Ana M Pérez-Marín. 2019. “The Use of Telematics Devices to Improve Automobile Insurance Rates.” *Risk Analysis* 39 (3). Wiley Online Library: 662–72.

Guillen, Montserrat, Jens Perch Nielsen, Ana M Pérez-Marín, and Valandis Elpidorou. 2020. “Can Automobile Insurance Telematics Predict the Risk of Near-Miss Events?” *North American Actuarial Journal* 24 (1). Taylor & Francis: 141–52.

Handel, Peter, Isaac Skog, Johan Wahlstrom, Farid Bonawiede, Richard Welch, Jens Ohlsson, and Martin Ohlsson. 2014. “Insurance Telematics: Opportunities and Challenges with the Smartphone Solution.” *IEEE Intelligent Transportation Systems Magazine* 6 (4). IEEE: 57–70.

Jafarnejad, Sasan, German Castignani, and Thomas Engel. 2017. “Towards a Real-Time Driver Identification Mechanism Based on Driving Sensing Data.” In *2017 Ieee 20th International Conference on Intelligent Transportation Systems (Itsc)*, 1–7. IEEE.

Jiang, Kun, Diange Yang, Shichao Xie, Zhongyang Xiao, Alessandro Corrêa Victorino, and Ali Charara. 2019. “Real-Time Estimation and Prediction of Tire Forces Using Digital Map for Driving Risk Assessment.” *Transportation Research Part C: Emerging Technologies* 107. Elsevier: 463–89.

Jiang, Ying, Junyi Zhang, Yinhai Wang, and Wenyuan Wang. 2020. “Drivers’ Behavioral Responses to Driving Risk Diagnosis and Real-Time Warning Information Provision on Expressways: A Smartphone App–based Driving Experiment.” *Journal of Transportation Safety & Security* 12 (3). Taylor & Francis: 329–57.

Jin, Wen, Yinglu Deng, Hai Jiang, Qianyan Xie, Wei Shen, and Weijian Han. 2018. “Latent Class Analysis of Accident Risks in Usage-Based Insurance: Evidence from Beijing.” *Accident Analysis & Prevention* 115. Elsevier: 79–88.

Joubert, Johan W, Dirk De Beer, and Nico De Koker. 2016. “Combining Accelerometer Data and Contextual Variables to Evaluate the Risk of Driver Behaviour.” *Transportation Research Part F: Traffic Psychology and Behaviour* 41. Elsevier: 80–96.

Liao, Yuan, Minjuan Wang, Lian Duan, and Fang Chen. 2018. “Cross-Regional Driver–vehicle Interaction Design: An Interview Study on Driving Risk Perceptions, Decisions, and Adas Function Preferences.” *IET Intelligent Transport Systems* 12 (8). IET: 801–8.

Litman, Todd. 2007. “Distance-Based Vehicle Insurance Feasibility, Costs and Benefits.” *Victoria* 11.

Lu, Jingyi, Xiaofei Xie, and Ruogu Zhang. 2013. “Focusing on Appraisals: How and Why Anger and Fear Influence Driving Risk Perception.” *Journal of Safety Research* 45. Elsevier: 65–73.

Ma, Yu-Luen, Xiaoyu Zhu, Xianbiao Hu, and Yi-Chang Chiu. 2018. “The Use of Context-Sensitive Insurance Telematics Data in Auto Insurance Rate Making.” *Transportation Research Part A: Policy and Practice* 113. Elsevier: 243–58.

Paefgen, Johannes, Thorsten Staake, and Elgar Fleisch. 2014. “Multivariate Exposure Modeling of Accident Risk: Insights from Pay-as-You-Drive Insurance Data.” *Transportation Research Part A: Policy and Practice* 61. Elsevier: 27–40.

Paefgen, Johannes, Thorsten Staake, and Frédéric Thiesse. 2013. “Evaluation and Aggregation of Pay-as-You-Drive Insurance Rate Factors: A Classification Analysis Approach.” *Decision Support Systems* 56. Elsevier: 192–201.

Peng, Jinshuan, and Yiming Shao. 2018. “Intelligent Method for Identifying Driving Risk Based on V2v Multisource Big Data.” *Complexity* 2018. Hindawi.

Pesantez-Narvaez, Jessica, Montserrat Guillen, and Manuela Alcañiz. 2019. “Predicting Motor Insurance Claims Using Telematics Data—XGBoost Versus Logistic Regression.” *Risks* 7 (2). Multidisciplinary Digital Publishing Institute: 70.

Siordia, Oscar S, Isaac Martín de Diego, Cristina Conde, and Enrique Cabello. 2014. “Subjective Traffic Safety Experts’ Knowledge for Driving-Risk Definition.” *IEEE Transactions on Intelligent Transportation Systems* 15 (4). IEEE: 1823–34.

Sun, Shuai, Jun Bi, and Cong Ding. 2019. “Cleaning and Processing on the Electric Vehicle Telematics Data.” In *INFORMS International Conference on Service Science*, 1–6. Springer.

Sun, Shuai, Jun Bi, Montserrat Guillen, and Ana M Pérez-Marín. 2020. “Assessing Driving Risk Using Internet of Vehicles Data: An Analysis Based on Generalized Linear Models.” *Sensors* 20 (9). Multidisciplinary Digital Publishing Institute: 2712.

Troncoso, Carmela, George Danezis, Eleni Kosta, Josep Balasch, and Bart Preneel. 2010. “Pripayd: Privacy-Friendly Pay-as-You-Drive Insurance.” *IEEE Transactions on Dependable and Secure Computing* 8 (5). IEEE: 742–55.

Tselentis, Dimitrios I, George Yannis, and Eleni I Vlahogianni. 2016. “Innovative Insurance Schemes: Pay as/How You Drive.” *Transportation Research Procedia* 14. Elsevier: 362–71.

———. 2017. “Innovative Motor Insurance Schemes: A Review of Current Practices and Emerging Challenges.” *Accident Analysis & Prevention* 98. Elsevier: 139–48.

Verbelen, Roel, Katrien Antonio, and Gerda Claeskens. 2018. “Unravelling the Predictive Power of Telematics Data in Car Insurance Pricing.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 67 (5). Wiley Online Library: 1275–1304.

Wang, Jianqiang, Heye Huang, Yang Li, Hanchu Zhou, Jinxin Liu, and Qing Xu. 2020. “Driving Risk Assessment Based on Naturalistic Driving Study and Driver Attitude Questionnaire Analysis.” *Accident Analysis & Prevention* 145. Elsevier: 105680.

Wang, Jianqiang, Yang Zheng, Xiaofei Li, Chenfei Yu, Kenji Kodaka, and Keqiang Li. 2015. “Driving Risk Assessment Using Near-Crash Database Through Data Mining of Tree-Based Model.” *Accident Analysis & Prevention* 84. Elsevier: 54–64.

Yan, Lixin, Yishi Zhang, Yi He, Song Gao, Dunyao Zhu, Bin Ran, and Qing Wu. 2016. “Hazardous Traffic Event Detection Using Markov Blanket and Sequential Minimal Optimization (Mb-Smo).” *Sensors* 16 (7). Multidisciplinary Digital Publishing Institute: 1084.

Yan, Ying, Youhua Dai, Xiaodong Li, Jinjun Tang, and Zhongyin Guo. 2019. “Driving Risk Assessment Using Driving Behavior Data Under Continuous Tunnel Environment.” *Traffic Injury Prevention* 20 (8). Taylor & Francis: 807–12.