Semi-autonomous vehicles: usage-based data evidences of what could be expected from eliminating speed limit violations

Ana M. Pérez-Marín[a] and Montserrat Guillen[b]

[a]Riskcenter, Universidad de Barcelona, Av. Diagonal, 690, 08034 Barcelona; amperez@ub.edu. Corresponding author.

[b]Riskcenter, Universidad de Barcelona, Av. Diagonal, 690, 08034 Barcelona; mguillen@ub.edu

Semi-autonomous vehicles: usage-based data evidences of what could be expected from eliminating speed limit violations

Abstract

The use of advanced driver assistance systems and the transition towards semi-autonomous vehicles are expected to contribute to a lower frequency of motor accidents and to have a significant impact for the automobile insurance industry, as rating methods must be revised to ensure that risks are correctly measured. Telematics information and usage-based insurance research are analyzed to identify the effect of driving patterns on the risk of accident. This is used as a starting point for addressing risk quantification and safety for vehicles that can control speed. The effect of excess speed on the risk of accidents is estimated with a real telematics data set. Scenarios for a reduction of speed limit violations and the consequent decrease in the expected number of accident claims are shown. If excess speed could be eliminated, then the expected number of accident claims could be reduced to half of its initial value, applying the average conditions of the data used in this study. As a consequence, insurance premiums also diminish.

Keywords: advanced driver assistance systems, semi-autonomous vehicles, insurance, pay-how-you-drive.

1. Introduction

This paper focuses on usage-based insurance (UBI) schemes and Advanced Driver Assistance Systems (ADAS) as a step before semi-automation. Specifically the effect of speed control systems on the risk of accident is analyzed. Many automobiles nowadays incorporate automatic speed control devices, which allow the driver to keep the vehicle at a predetermined constant speed, and ensure that the speed limit is not going to be violated. At the same time, the driver does not need to look at the speedometer and just needs to concentrate on the road, which contributes to safer driving. What would be the effect of automatic speed controlled driving on the risk of accidents? A revision of the existing literature and an empirical research is carried out based on real UBI information in order to answer this question.

Many insurance companies around the world are currently offering UBI policies. Depending on the level of telematics information accounted for in automobile insurance, UBI can have different forms, such as pay-as-you-drive (PAYD) and pay-how-you-drive (PHYD) insurance. In PAYD insurance, the premium depends on the real distance traveled by the insured party, which is monitored by a telematics device installed in the car. On the other hand, in PHYD insurance the premium calculation also depends on other telematics variables such as the type of road, time, speed, sudden braking events, etc. Therefore, such automobile insurance contracts are a step towards a more personalized concept of motor insurance.

Many recent research articles have analyzed real vehicle usage data in the context of UBI and have determined the effect of driving patterns on the risk of accident. This knowledge can be used as a baseline for approaching risk quantification in insurance policies for vehicles incorporating ADAS as well as for semi-autonomous vehicles.

The contribution of this paper is centered on the role of speed control and it is based on the premise that automated procedures can reduce and eventually eliminate the violation of speed limits on the road. Based on real data the reduction in the frequency of accidents and its impact on safety and insurance premiums are calculated.

Specifically, a real case study is presented where the impact of automatic speed control is measured in different scenarios by using a PHYD insurance database provided by a Spanish company. Thereby, this is a contribution towards the transition to a new model for semi-autonomous vehicle insurance. Additionally, urban driving is also analyzed as a risk factor in the literature on UBI, as the frequency of accidents is higher in urban areas than elsewhere. Therefore, the effect of new devices which make driving easier on urban roads could also be approached, such as the assisted parking systems, proximity sensors, and so on.

The paper is organized as follows. Section two presents the background. In section three, the theory used to assess the impact of speed on the risk of accidents is presented. In section four the material and methods are presented. Section five applies the empirical data and builds scenarios using existing models that emphasize the role of speed limitation and automation for the assessment of accident risk. The results are discussed. Additionally the role of automated speed control on safety from the perspective of traffic authorities and society is analyzed, as well as the role of speed control in insurance premiums. Finally, section six concludes.

2. Background

ADAS support drivers by providing warnings to reduce risk exposure or automating some driving tasks to relieve them from the manual control of the vehicle (Piao and McDonald, 2008). These systems are intended to increase road safety by enhancing driver performance, and include lane maintenance systems, crash-avoidance technologies and systems for keeping safe speed and safe distance (referred to as SASPENCE) among others. Technological advances (Jiménez *et al.*, 2009 and Jiménez and Naranjo, 2011) and the identification of the factors which influence and cause traffic accidents (Staubach, 2009) are the basis for designing and implementing ADAS. There are evidences of the positive effects of such technologies, according to Reagan *et al*. (2018) emergency braking systems reduce rates of insurance claims compared to vehicles that do not have these systems. ADAS afford safety advantages, but also challenge the traditional role of drivers (Rahman *et al*., 2017). This is the reason why there are also potential downsides that may undermine their acceptability. The use of ADAS may also generate false or unnecessary alarms, induce distraction, overload and fatigue (Ruscio *et al*, 2017). Many authors argue that automation has the potential to significantly reduce the number of vehicle crashes and their associated economic burden (Fagnant and Kockelman, 2015), but driver acceptance is a precondition for a successful implementation (Rahman *et al*., 2017). Many authors have analyzed drivers' acceptability of ADAS (Adell *et al*, 2011, Rahman *et al*., 2017, Reagan *et al*., 2018,). Son *et al*. (2015) found that there were significant age and gender differences in the acceptance and effectiveness of the ADAS, and that the roadway environment also affected their effectiveness.

It is widely accepted that speeding is one of the critical factors that has a negative effect on traffic safety. It is well established that speeding is related to the severity of accidents (see, among others, Dissanayake and Lu, 2002; Elvik, 2004 and Jun *et al*., 2007 and 2011). Ayuso *et al*. (2010) found that traffic violations related to excess speed significantly increase the odds of serious or fatal accidents versus small accidents, by using a multinomial logistic regression model. Additionally, Yu and Abdel-Aty (2014) concluded that large variations of speed prior to the crash would increase the likelihood of severe crash occurrence. More recently, Imprialou *et al*. (2016) revisited the crash–speed relationship by creating a new crash data aggregation approach that enables improved representation of the road conditions just before crash occurrences and they found that higher speed is related to increasingly serious crashes.

Many articles have made a contribution to the understanding of speeding by young drivers and its effect on accident risk in the context of UBI. Ayuso *et al*. (2014) concluded that a higher proportion of kilometers traveled at speeds above the limits is associated with a higher risk of accident among young drivers with UBI. The association between gender and risky driving was also stressed by Ayuso *et al*. (2014, 2016a and 2016b), who concluded that, on average, men have riskier driving patterns than women, as men travel more kilometers per day, during the night and at speeds above the limit, than women. All these three factors were found to correlate with a larger expected number of accidents.

Paefgen *et al*. (2014) investigated the differences between vehicles that get involved in crashes and those that do not, by using PAYD insurance data and found that the risk fluctuates throughout the day, and is higher at nightfall, during the weekends, on urban roads and at low-range or higher-range speeds (0-30 km/h and 90-120 km/h, respectively).

Nowadays, semi-autonomous vehicles incorporate automatic speed control devices, which ensure that the speed limit is not going to be violated. This can potentially remove a leading cause of vehicle accidents and therefore may lead to more safety and lower claim rates. Today, drivers face an evolution from manual to semi-autonomous driving with the ultimate aim of introducing driverless vehicles. This transition will progressively reduce accident frequency, resulting in lower losses and lower premiums for motor insurance. Nevertheless, some authors claim that driving performance is safer with lower rather than higher levels of automation, in situations with automation failures (Strand et al., 2014). In that context, the insurance industry should be able to change their rating methods in order to ensure that risks are correctly measured, but most importantly they should be able to contribute to preventive actions and risk mitigating procedures to influence the way drivers perceive their driving performance and to engage them in safer attitudes. As part of that process, telematics information and UBI background are going to play an important role. Tselentis et al. (2017) provided a recent review of UBI schemes and concluded that there is evidence that UBI implementation implies lower insurance costs for less risky and exposed drivers. These authors also provide a strong motivation for drivers to improve their driving behaviors and reduce their degree of exposure by receiving feedback and monitoring their performance, which would result in crash risk reduction. Along those lines, in a study in the Netherlands, Zantema et al. (2008) showed that if PAYD were to be implemented the total crash reduction estimate would be more than 5%, resulting in 60 fewer fatalities and a reduction of over 1,000 people injured by traffic accidents, each year.

Recently, Baecke and Bocca (2017) investigated how driving behavior data can improve the risk selection process in an insurance company. They proved that including standard telematics variables significantly enriches the risk assessment of customers and insurance companies are better able to tailor their products to the customers' risk profile. According to their results, this new type of telematics-based insurance product can be implemented very quickly, since just three months of data is enough to get the best estimations.

Sheehan et al. (2017) proposed using a Bayesian Network statistical approach to estimate aggregate claims losses from a range of risk factors which are based on PAYD and PHYD insurance approaches. They showed the use of this method for a Level 3 Automation vehicle, where the vehicle can perform many aspects of driving such as steering, acceleration/deceleration and monitoring the driving environment, but requires the driver to be ready to intervene, at any moment, at the vehicle's request. These authors considered two scenarios: one where the driver is in control and one where the vehicle is in control. As expected, the automated features remove driver error and reduce accident risks. They found that the aggregate claims loss is one tenth of that where control is by the driver. This question is also analyzed here, as well as the influence of speed control on accident risk.

Payre et al. (2014) investigated the acceptability of fully automated driving (FAD) by using an online questionnaire addressed to French drivers. They found that around 68% of respondents accepted FAD *a priori*, and that preferred uses were on major highways, in traffic congestion and for automatic parking. Jeong et al. (2017) claimed that it is implausible to expect that autonomous driving systems will reach 100% market penetration rate in the near future, therefore, the interaction between equipped and unequipped vehicles must be investigated. More recently, Kyriakidis et al. (2015) also investigated public opinion on automated vehicles in an international study. They found that on average manual driving was rated the most enjoyable mode of driving, with 33% of respondents indicating that fully automated driving would be highly enjoyable. Respondents were found to be most concerned about software hacking/misuse and they were also concerned about legal issues and safety. Recently Guo et al. (2017) stressed the need to explore driver–vehicle cooperation as an opportunity to improve driving performance through human–automation synergy. Harper et al. (2016) investigated the benefits and costs of partially-automated vehicle collision avoidance technologies. These authors considered fleet-wide deployment of blind spot monitoring, lane departure warning, and forward collision warning crash avoidance systems and concluded that this early form of automation has a positive net benefit, suggesting that fleet-wide adoption of such technologies would be beneficial from both an economic and social perspective.

Finally, the advantages for users and their level of acceptance of UBI schemes have also been investigated in the literature. Litman (2011) discussed the advantages of UBI policies compared to the traditional ones. Usage-based insurance reduces accidents, increases insurance affordability and reduces uninsured driving, among others. Tselentis et al. (2017) also argued that UBI policies have potentially a significant impact on traffic safety and congestion. More recently, Tselentis et al. (2018) investigated which factors affect users' willingness to pay for UBI policies. They concluded that women and smartphone owners are more likely to choose UBI schemes. Moreover, the higher the speed reduction imposed by the insurer to the user, the lower the probability to choose UBI schemes. Finally, they also found that people over 40 years old are less likely to choose UBI products than younger drivers.

3. Theory

To assess the impact of automatic speed controlled driving on the risk of accidents the claim frequency is modeled as a function of the proportion of speed violations by using telematics variables. This can easily be done using a Poisson regression model (Boucher and Guillen, 2009). In this case, the classical offset variable that measures exposure time can be changed by a generalized offset variable that introduces into the model the distance traveled during a natural year (as suggested by Boucher et al., 2013 and Lemaire et al., 2016). A generalized offset variable in the context of a Poisson regression model is simply an explanatory variable which is introduced in logarithm scale into the model with an associated parameter which is not constrained to be equal to one. Boucher et al. (2013) proposed this approach to avoid constraining the relationship between the frequency of claims and the distance traveled to be proportional.

The same approach is considered here, all telematics variables are entered into the Poisson regression model in the logarithm scale. The gender will be introduced as a binary variable. Let *N* be the total number of insured users and *K* the total number of explanatory variables (gender and *K-1* telematics variables), then the model for the expected frequency of claims for insured user i = 1,.., *N*, which is denoted as λ_i, can be formulated as:

$$\lambda_i = \exp\left(\beta_0 + \beta_1 x_{1i} + \sum_{k=2}^{K} \beta_k \ln(x_{ki})\right) \tag{1}$$

where β_0 is the intercept term, x_{k1} is the gender of individual *i* and β_1 the corresponding coefficient in the model, and x_{ki} is the telematics variable x_k for individual *i*, and β_k is the corresponding associated parameter. Equation (1) is equivalent to

$$\lambda_i = \exp(\beta_0 + \beta_1 x_{1i}) \cdot \prod_{k=2}^{K} x_{ki}^{\beta_k} \tag{2}$$

which means that effects are combined multiplicatively. Note that according to this formulation β_k for telematics variables measures the elasticity of the frequency of claims with respect to x_k. So, if the value of the variable increases in percentage terms then the frequency changes β_k multiplied by this percent, accordingly. Based on the real data the reduction in the frequency of accidents is calculated and its impact on society in terms of protection and savings in human lives is analyzed.

4. Material and methods

An empirical analysis is carried out by using a data set of PHYD insurance policyholders which was collected by a Spanish insurer. The sample consists of 9,557 young drivers who had a PHYD insurance policy in force during the year 2010. Age ranges from 18 to 35 because this PHYD policy was only offered to young drivers. The temporal exposure to the risk of accident for all of them is one year, as their insurance policies were in force during the entire year 2010. The variables considered in the analysis are summarized in Table 1.

Table 1. Variable description

Variable	Description
km	Distance traveled during the year measured in kilometers
sex	Sex (1 = men, 0 = women)
speed	% of kilometers traveled at speeds above the limit
urban	% of kilometers traveled on urban roads
age	Age of the driver at the beginning of 2010
nsin	Number of “at fault” accident claims during the year

The exogenous variables are *sex*, *km* (which is the total distance traveled during the year in kilometers), *speed* (percentage of kilometers traveled at speeds above the mandatory limits), *urban* (percentage of kilometers traveled on urban roads) and finally the *age* of the driver. The dependent variable is *nsin*, which is the total number of claims occurring during the year 2010 where the driver was at fault. The reason to model only "at fault" claims is that these are a true indicator of accident occurrence that was actually caused by the driver. The existence of other accidents caused by other drivers may be due to hazard or third parties and these are not assumed to be related to the risk of accident directly caused by the insured driver's fault. However, all accidents claimed by the insured user were modeled even if they were caused by third parties, but the main conclusions do not change much. Those results are available upon request from the authors.

Table 2 shows some descriptive statistics. Drivers travel on average 13,031.27 km during the year (standard deviation 7,693.25). They travel on average 8.89% of total kilometers at speeds above the limit. There exists a high heterogeneity regarding speed (the standard error is 8.15 and 5% of them travel more than 26.34% of total kilometers at speeds above the limit). The average level of urban driving is 26.37% (standard deviation 14.18). All drivers are under the age of 35, the average age being 24.78 (standard deviation 2.82). They made on average 0.10 claims during 2010, most of them did not make a claim but some of them made 3 claims. Finally, regarding the variable *sex*, there are 50% men in the sample.

Table 2. Descriptive statistics

	Mean	Standard deviation	Minimum	5% percentile	25% percentile	Median	75% percentile	95% percentile	Maximum
km	13,031.27	7,693.25	0.69	2,921.26	7,517.45	11,676.94	17,304.50	27,249.29	57,756.98
speed	8.89	8.15	0.00	1.01	3.10	6.09	12.16	26.34	44.92
urban	26.37	14.18	0.00	8.58	15.70	23.46	34.40	53.58	100.00
age	24.78	2.82	18.11	20.36	22.66	24.63	26.88	29.46	35.00
nsin	0.10	0.32	0	0	0	0	0	1	3

5. Results and discussion

A Poisson regression model is used to estimate the number of claims (*nsin*) as a function of the independent variables. The independent variables in the model are introduced in logarithms[1] which is denoted by ln. The parameter estimates of the Poisson regression model are shown in Table 3.

[1] Due to the fact that a very small percentage of drivers (0.34%) had *speed* equal to 0% and/or *urban* equal to 0%, it was added to these two variables 0.001 so that the logarithm could be calculated.

Table 3. Parameter estimates for a Poisson regression model

Parameter	Estimate	Standard error	Wald 95% confidence limits		Wald chi square	p-value
Intercept	-3.0760	1.1695	-5.3682	-0.7837	6.92	0.0085
sex	0.0632	0.0657	-0.0656	0.1921	0.92	0.3362
ln(km)	0.3800	0.0598	0.2628	0.4972	40.39	<.0001
ln(speed)	0.0721	0.0351	0.0033	0.1409	4.22	0.0400
ln(urban)	0.4602	0.0697	0.3236	0.5968	43.61	<.0001
ln(age)	-1.3666	0.2826	-1.9206	-0.8127	23.38	<.0001

The model is globally significant (Likelihood Ratio Test statistic equals 111.08, p-value < 0.0001). The Akaike Information Criterion (AIC) equals 6,406.8 and the Bayesian Information Criterion (BIC) 6,449.8. A Negative Binomial Regression was also adjusted to the data, but it resulted in higher values of AIC and BIC. Note that the Poisson parameter estimates are consistent even though there could be overdispersion in the data. A Poisson regression model with random parameters has also been used to explore specifically the speed limit effect based on an individual level. The results suggest that only in the case of the percentage of kilometers travelled in urban areas some level of randomness could be accepted by doing a strained interpretation of the results. Actually, the results of the random parameter model (they are presented in the Appendix) are almost identical to those obtained with the classical Poisson regression model. Therefore, the classical Poisson regression model was chosen for the analysis because it is simpler, but similar conclusions follow from the Poisson regression with random parameters.

According to the results in Table 3, *sex* does not have a significant effect. Vehicle usage (measured by the distance traveled), breaking the speed limits and urban driving are associated with a higher number of claims[2]. Regarding age, the number of claims decreases as the age increases.

It is important to note that if the parameter estimate of the telematics variables is different from one, this means that the relationship between the corresponding variable and the number of claims is not proportional. Regarding the distance traveled (variable km) the parameter equals 0.38 and the Wald 95% confidence limits are 0.26 and 0.50, therefore, it is clearly different from one and the relationship is not proportional (as found by Boucher et al., 2013). The relationship between the distance traveled and the expected number of accidents is represented in Figure 1.

[2] The effect of speed is significant at the 10% level, and almost significant at the 5% level (p-value 0.058).

Figure 1. The expected frequency of claims as a function of the distance traveled. The dots represent the average frequency of claims when the insured users are grouped by intervals of 500 driven km. The line represents the fitted claim frequency as a function of the distance traveled. Dots may represent different number of drivers.

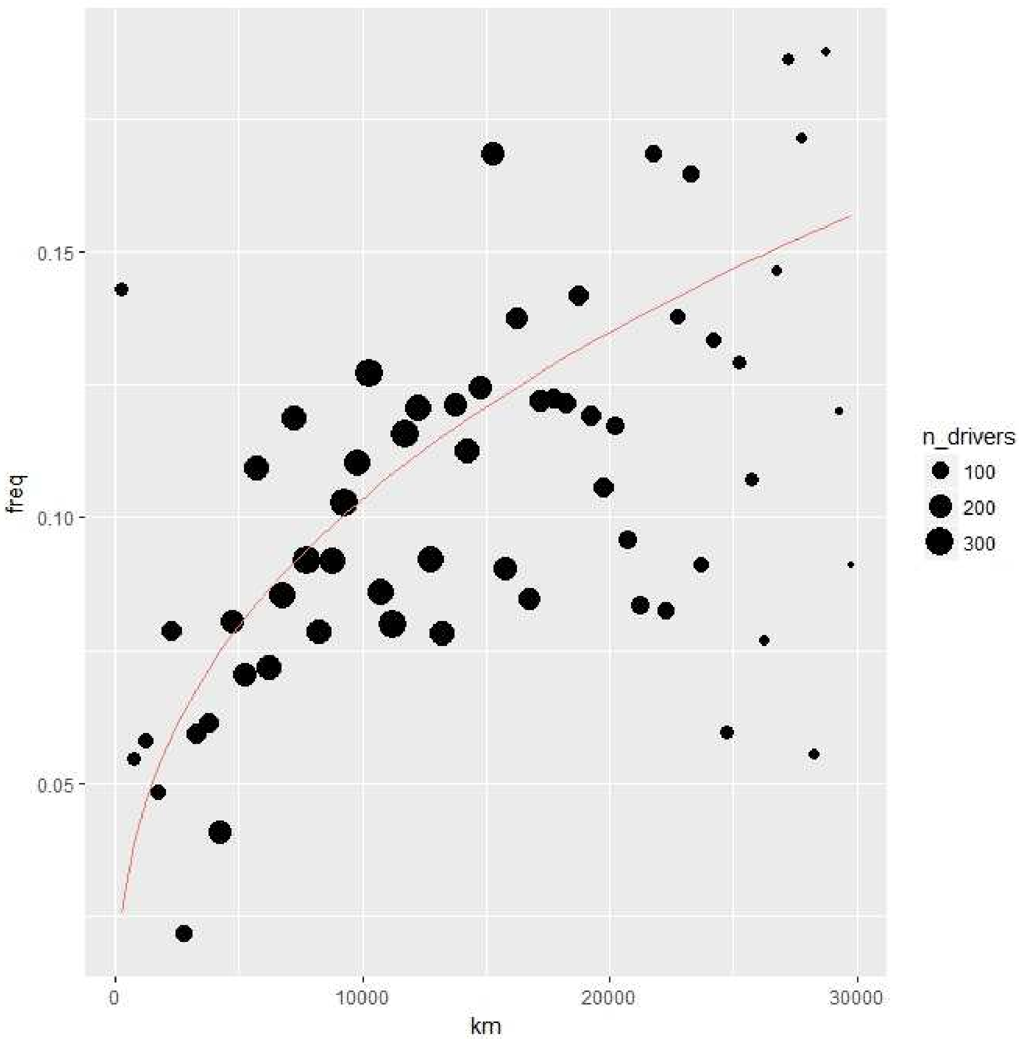
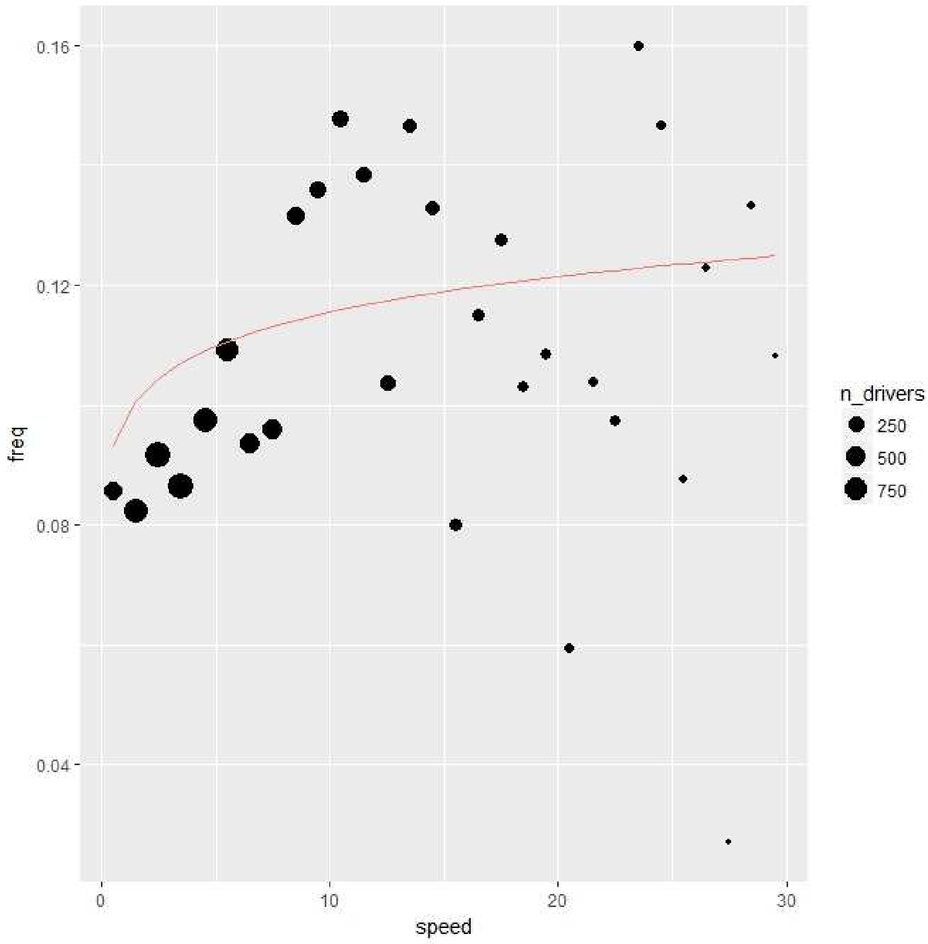


Figure 1 shows the average frequency of claims as a function of the distance traveled. The dots represent the real average frequency when the insured users are grouped by intervals of 500 driven km. Note that each dot represents an average that has been calculated with a different number of drivers. In general, as the total number of kilometers increases, the number of drivers in each interval decreases. Above 20,000 km the data seems to be more heterogeneous, this is due to the fact that there are few insured users with such a large number of traveled kilometers. This is also the reason why the horizontal axis was limited to 30,000 km. The line represents the fitted claim frequency as a function of the distance traveled and it has been calculated by using the Poisson regression model (parameter estimates in Table 3) where the rest of covariates have been taken to be equal to the sample mean (see Table 2). The frequency of claims is far from increasing linearly with the number of kilometers. Instead, a high slope is observed for low values of the distance traveled, and it marginally decreases as more kilometers are driven. This effect is produced by the fact that the parameter associated to the distance traveled in the Poisson regression is lower than one, namely equal to 0.38.

Figure 2. The expected frequency of claims as a function of the percentage of kilometers traveled at speeds above the limit. The dots represent the average frequency of claims when the insured users are grouped according to their speed violations by intervals of 1%. The line represents the fitted claim frequency as a function of the percentage of kilometers traveled at speeds above the limit. Dots may represent different number of drivers.



Similarly, in Figure 2 the frequency of claims is presented as a function of the percentage of kilometers traveled at speeds above the limit. The dots represent the real average frequency when the insured users are grouped by intervals of 1% according to the distance driven at speeds above the limit. As the speed increases there are fewer insured users with such a high level of speed limit violations. Again, the line represents the fitted claim frequency as a function of the percentage of kilometers traveled at speeds above the limit by using the Poisson regression model (the rest of covariates again have been taken to be equal to the sample mean, see Table 2). The frequency of claims increases very sharply for low values of speed violations and further on increases slowly. This effect is again produced by the fact that the associated parameter in the Poisson regression model equals 0.0721, clearly lower than one and therefore far from a proportional relationship.

The same type of analysis was carried out for urban driving and age. The results are plotted in Figures 3 and 4, respectively. In the case of urban driving the dots represent the real average frequency when the insured users are grouped by intervals of 1% according to their urban driving. As the level of urban driving increases there are fewer insured users and the data are more heterogeneous. The line representing the prediction of the frequency of claims increases more sharply for low values of urban driving than further on (the corresponding parameter in the model equals 0.46). On the other hand, in Figure 4, the dots represent the real average frequency of claims when the insured users are grouped according to their age at the beginning of 2010 by using intervals of one month. The relationship is the opposite, as a decrease in claim frequency as age increases is observed.

Figure 3. The expected frequency of claims as a function of the percentage of kilometers traveled on urban roads. The dots represent the average frequency of claims when the insured users are grouped according to their urban driving by intervals of 1%. The line represents the fitted claim frequency as a function of the percentage of kilometers traveled on urban roads. Dots may represent different number of drivers.

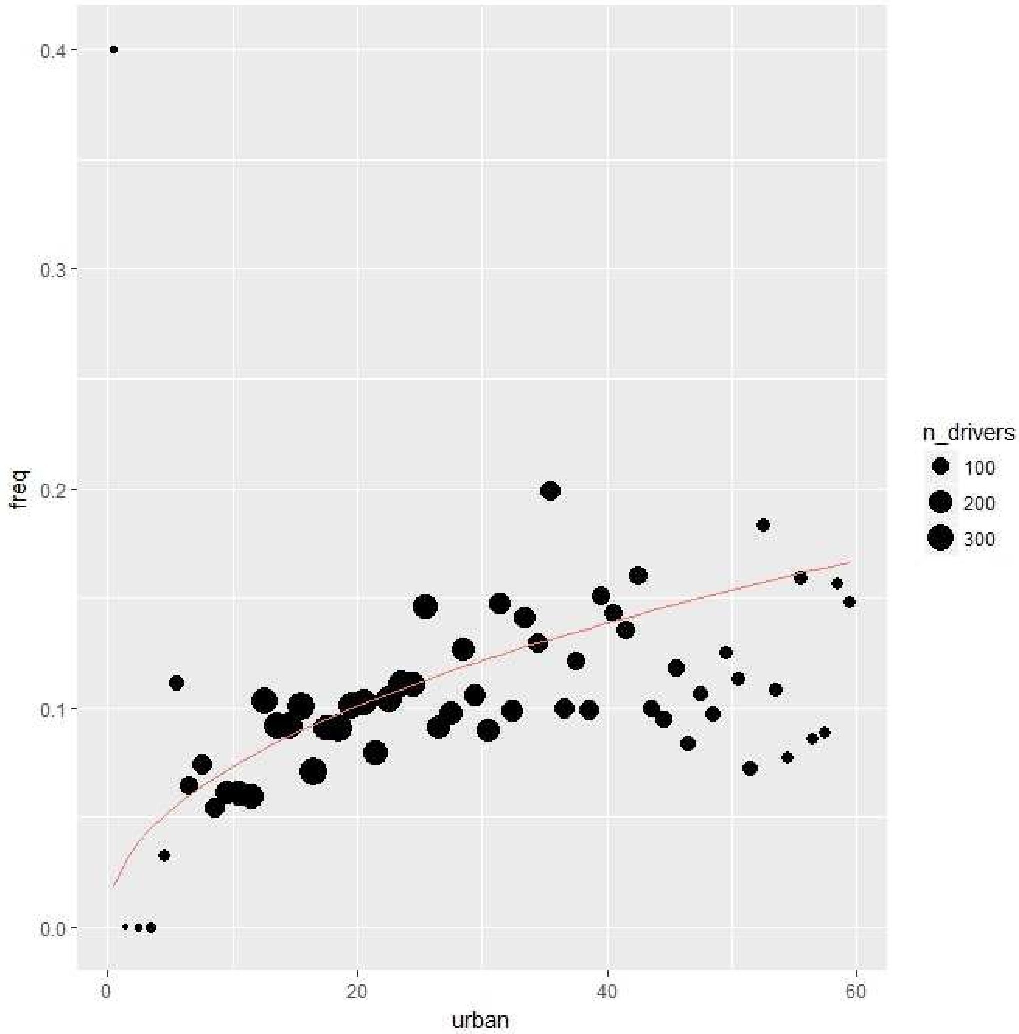
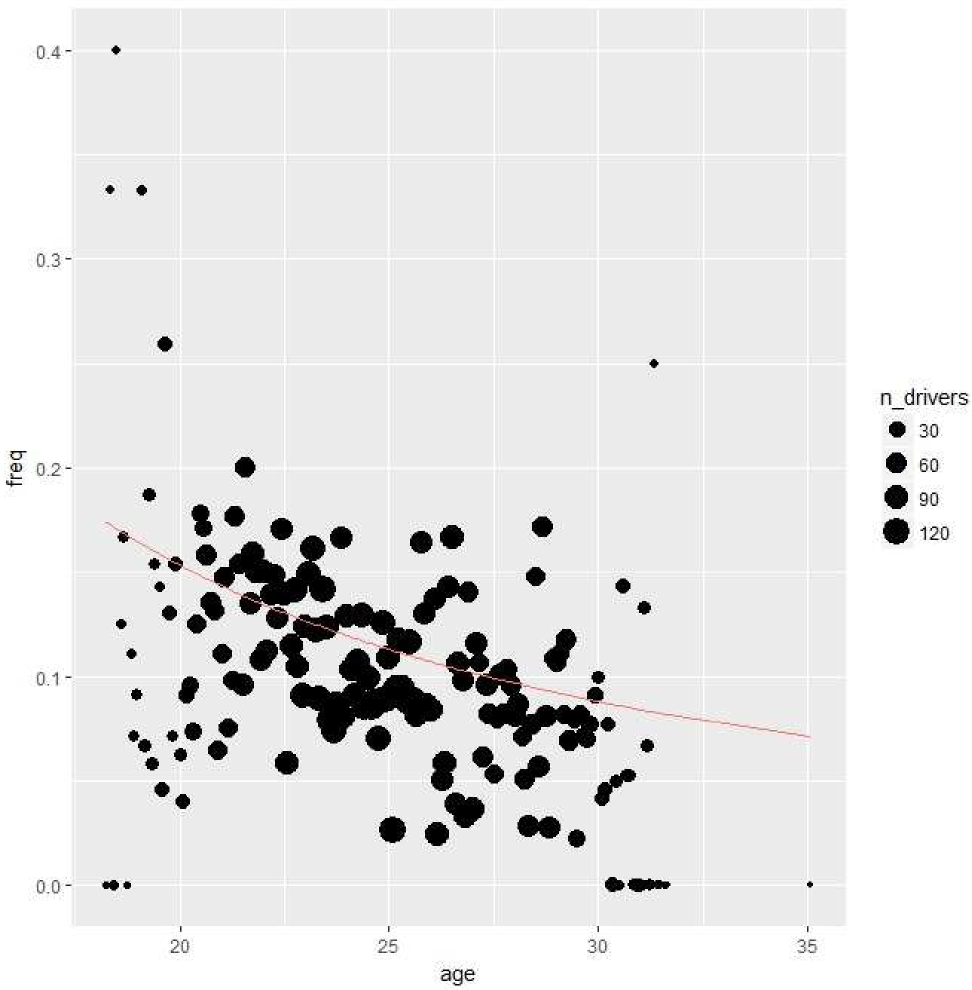


Figure 4. The frequency of claims as a function of the age of the driver. The dots represent the average frequency of claims when the insured users are grouped according to their age by intervals of one month. The line represents the fitted claim frequency as a function of the age of the driver. Dots may represent different number of drivers.



5.1. The role of automated speed control on safety

In this section the impact on safety if vehicles incorporate speed control devices to avoid speed violations is analyzed. Firstly, different scenarios are considered where the number of claims per 1,000 drivers is measured as a function of their level of speed violations (measured by the percentage of kilometers traveled at speeds above the limit) if the rest of the variables are assumed to be equal to the corresponding sample mean. The calculations are done by using the results of the Poisson regression model in Table 3. These results are presented in Table 4.

Table 4. Expected number of claims per 1,000 drivers with different levels of speed violation.

Speed	Expected number of claims
0%	59.46
1%	97.84
2%	102.85
5%	109.87
7%	112.57
9%	114.63
10%	115.50
12%	117.03
15%	118.93
17%	120.01
20%	121.42

If the level of speed violations is reduced, for example from the average sample level of 9% to 0%, then the number of claims per 1,000 drivers would change from 114 to 59, where the rest of the variables are kept constant. This is clearly a significant reduction. This is then the average impact on claim frequency and road safety if vehicles incorporate control devices to avoid speed violations. Additionally, Table 4 also shows the same calculations for different levels of speed violations, and it is very remarkable that if speed violations are reduced from 20% to 0%, then the claim frequency per 1,000 drivers decreases from 121 to 59.

Given that the average level of speed limit violation is around 9%, which means that the expected number of claims per 1,000 drivers is 114 (see Table 4), a complete elimination of the violations, would lead to 0% levels and therefore to an expected number equal to 59. This is more than half of the initial level, i.e. 59/114=52%, therefore the initial level is reduced by approximately 48%, one half.

Finally, Table 5 shows the difference in the expected number of claims per 1,000 drivers and per year due to a change in the level of speed violation. The calculations are done by assuming that the speed level changes from some level *before* (rows in the table) to some level *after* (columns) by keeping the rest of the variables constant and equal to the sample mean. The cells in Table 5 show the number of claims per 1,000 drivers *after* minus *before*. For example, reducing *speed* from 20% to 9% (which is approximately the sample mean) results in 7 fewer claims (per 1,000 drivers). Of course, the largest reduction occurs when speed violations are totally eliminated (by using speed control devices). The reduction equals 38 claims per 1,000 drivers if its initial level is just 1%, and reaches 62 claims per 1,000 drivers if initially the level was 20%.

Table 5. Change in the yearly expected number of claims per 1,000 drivers due to a change in the level of speed violation (from some level *before* to some level *after*). The cells show the number of claims per 1,000 drivers *after* minus *before*.

	After										
Before	0%	1%	2%	5%	7%	9%	10%	12%	15%	17%	20%
0%	0	38.39	43.40	50.42	53.12	55.17	56.05	57.58	59.47	60.55	61.97
1%	-38.39	0	5.01	12.03	14.73	16.79	17.66	19.19	21.09	22.17	23.58
2%	-43.40	-5.01	0	7.02	9.72	11.78	12.65	14.18	16.08	17.15	18.57
5%	-50.42	-12.03	-7.02	0	2.70	4.75	5.63	7.16	9.05	10.13	11.55
7%	-53.12	-14.73	-9.72	-2.70	0	2.06	2.93	4.46	6.36	7.44	8.85
9%	-55.17	-16.79	-11.78	-4.75	-2.06	0	0.87	2.40	4.30	5.38	6.79
10%	-56.05	-17.66	-12.65	-5.63	-2.93	-0.87	0	1.53	3.43	4.50	5.92
12%	-57.58	-19.19	-14.18	-7.16	-4.46	-2.40	-1.53	0	1.90	2.98	4.39
15%	-59.47	-21.09	-16.08	-9.05	-6.36	-4.30	-3.43	-1.90	0	1.08	2.49
17%	-60.55	-22.17	-17.15	-10.13	-7.44	-5.38	-4.50	-2.98	-1.08	0	1.41
20%	-61.97	-23.58	-18.57	-11.55	-8.85	-6.79	-5.92	-4.39	-2.49	-1.41	0

5.2. The role of speed control on insurance premiums

The fundamental principle of insurance is the law of large numbers. In a large group of insured drivers and in a fixed period of time which is usually one year, only a small fraction of those drivers suffers an accident. Here it is assumed that all accidents are reported to the insurance company, but this is not always the case because many insurance companies penalize claims in order to save the cost of handling small claims. At the end of the day customers prefer not to claim a small accident in order to obtain a bonus in the following year and to avoid paying a higher premium due to the penalization.

Based on the idea of pooling the risk of all policyholders, insurance companies calculate the price of the premium as the product of the expected number of claims per contract times the expected cost of each claim plus some general expenses, which cover administration, advertising, claims handling, commissions and legal requirements.

Even if the price of insurance is not directly proportional to the expected number of claims, due to the presence of general expenses of the company, expenses are the smaller part (around 20% of the total price is due to the general expenses and loadings). So, a substantial decrease of the expected number of claims would naturally transmit to the final price. In addition, the impact could differ from one driver to the other due to the influence of some additional factors that are associated to the risk of having an accident such as driving experience, driving patterns in general and the personal driver's characteristics.

Using the scenarios mentioned above and the sample, it has been calculated the reduction of the price of insurance based on the assumption that the expected number of claims is a factor that proportionally to the average cost of claims accounts for 80% of the price of insurance. The results are shown in Table 6. Reducing the percentage of speed violation from 9% to 0% results in a 38.6% reduction in the premium. The highest percentage reduction in the premium is 41.2%, for those decreasing their percentage of speed violation from 20% to 0%.

Table 6. Percentage of variation in the price of insurance due to a change in the level of speed violation (from some level *before* to some level *after*). The cells show the percentage of increase (positive values) or decrease (negative values) according to the formula ((# claims *after* - # claims *before*)/ # claims *before*)*0.8.

	After										
Before	0%	1%	2%	5%	7%	9%	10%	12%	15%	17%	20%
0%	0%	48,7%	55,1%	64,0%	67,4%	70,0%	71,1%	73,1%	75,5%	76,8%	78,6%
1%	-31,0%	0%	4,0%	9,7%	11,9%	13,6%	14,3%	15,5%	17,0%	17,9%	19,1%
2%	-33,5%	-3,9%	0%	5,4%	7,5%	9,1%	9,8%	11,0%	12,4%	13,2%	14,3%
5%	-36,7%	-8,8%	-5,1%	0%	2,0%	3,5%	4,1%	5,2%	6,6%	7,4%	8,4%
7%	-37,8%	-10,5%	-6,9%	-1,9%	0%	1,5%	2,1%	3,2%	4,5%	5,3%	6,3%
9%	-38,6%	-11,8%	-8,2%	-3,3%	-1,4%	0%	0,6%	1,7%	3,0%	3,8%	4,8%
10%	-39,0%	-12,3%	-8,8%	-3,9%	-2,0%	-0,6%	0%	1,1%	2,4%	3,1%	4,1%
12%	-39,6%	-13,2%	-9,7%	-4,9%	-3,1%	-1,6%	-1,1%	0%	1,3%	2,0%	3,0%
15%	-40,3%	-14,3%	-10,9%	-6,1%	-4,3%	-2,9%	-2,3%	-1,3%	0%	0,7%	1,7%
17%	-40,7%	-14,9%	-11,5%	-6,8%	-5,0%	-3,6%	-3,0%	-2,0%	-0,7%	0%	0,9%
20%	-41,2%	-15,7%	-12,3%	-7,7%	-5,9%	-4,5%	-3,9%	-2,9%	-1,7%	-0,9%	0%

6. Conclusions

The transition towards semi-autonomous vehicles is expected to contribute to lowering the frequency of motor accidents and to have a significant impact for the automobile insurance industry, as rating methods must be revised to ensure that risks are correctly measured.

The analysis carried out has some limitations, because the data were not collected in the same conditions for semi-autonomous vehicles, but rather they were collected from manual drivers. The data belong to a group of drivers that are not exactly representative of the general population of drivers. Indeed, they are younger drivers. Authors studying the driving population in Spain report the average age to be older than the age of our sample. Official figures on the age of citizens who have a driving license in Spain indicate that the average is 48.63 years. Alcañiz *et al*. (2014) analyze a sample of random drivers who were stopped at sobriety checkpoints and they report similar results for Catalonia (Spain). Nevertheless, Kyriakidis et al. (2015) carried out a literature review on the public opinion on automated driving and found that several studies (Power, 2012) claimed that vehicle owners with the highest interest in fully autonomous driving are young drivers (between the ages of 18 and 37), which are precisely those that are represented in the sample. Nevertheless, the results of this study should be taken with caution in the context of autonomous or semi-autonomous vehicle insurance, as they provide simply an orientation to the insurer about expected impacts.

Telematics information and UBI research are used as a starting point for addressing risk quantification and safety for semi-autonomous vehicles. The real data used here have produced some scenarios for a reduction of speed limit violations and its impact on the decrease in the expected number of accident claims and premiums. If semi-autonomous vehicles could eliminate driving in excess of speed limits, the expected number of accident claims would be reduced. The benefits of this reduction would translate to a reduction in the number of victims on the road and an increase of overall safety. Specifically, if the percentage of kilometers traveled at speeds above the limit is reduced from the average level of 9% to

0%, then the number of claims is reduced by approximately one half. If all vehicles in Spain are equipped with automated speed control devices, so that this reduction would take place for all drivers, then the number of accidents with victims (bodily injuries and/or death) would be reduced by 1.77 accidents per 1,000 drivers. If only deaths are taken into account, the total number of victims would be reduced by 0.81 deaths per 26,514 drivers[3]. This is a significant reduction that provides relevant information for the insurance industry and the road safety authorities, besides the gains for society as a whole.

Future research in the topic should necessarily be based on the analysis of real claim data of autonomous or semi-autonomous vehicles. The progressive introduction of automatization on driving is expected to reduce human errors, the foremost cause of accidents. Future analysis based on real data could provide a more accurate estimation of the reduction of claim rates due to speed limit violation reduction and other risk factors which could be controlled by the vehicle. Specifically, future lines of research should measure how accident rates will be reduced and the overall impact of autonomous vehicles on road safety. Accident risks will not be eliminated entirely, and circumstances surrounding accidents will be different when technical innovations become available. In this new context, insurance companies should measure how exactly increased vehicle safety will translate into lower claims losses and premiums. Finally, the effect of weather conditions on the severity of claims should also be investigated, as well as the season and hour effects.

Acknowledgements

The study was supported by the Spanish Ministry of Economy and Competitiveness and the ERDF under grant ECO2016-76203-C2-2-P. The authors declare no conflict of interest.

[3] According to the Traffic Authorities in Spain (*Dirección General de Tráfico*, http://www.dgt.es/es/) the total number of drivers in 2016 was 26,514,026, the total number of accidents with victims (bodily injuries and/or death) was 97,756 and the total number of deaths was 1,689. Therefore, by applying the 48% reduction (due to automated speed control devices) to the number of accidents with victims (bodily injuries or death), it results in a reduction of 1.77 accidents (from 3.69 to 1.92) per 1,000 drivers. The same 48% reduction applied to the number of deaths results in a reduction of 0.81 deaths per 26,514 drivers.

Appendix

Table A.1. shows the results of the estimation of alternative Poisson regression models with random parameters (assuming a normal distribution for individual parameters). Model 0 is the basic Poisson regression model with constant parameters for all variables. Model 1 is the Poisson regression model where *ln(km)*, *ln(speed)*, *ln(urban)* and *ln(age)* have random parameters. In Model 2, only *ln(km)*, *ln(speed)* and *ln(urban)* have random parameters. In Model 3, only *ln(speed)* and *ln(urban)* have random parameters. Finally, in Model 4 only *ln(urban)* has a random parameter. For each model with random parameters, the values of the mean and standard deviation (*sd*) for each random parameter are shown. The log-likelihood, BIC and AIC are shown for each model.

Table A.1.Parameter estimates of the Poisson regression model with random parameters (in brackets the standard error is shown).

	Model 0	Model 1	Model 2	Model 3	Model 4
Constant	-3.076**	-3.030*	-3.030*	-3.016*	-2.975*
	(1.170)	(1.177)	(1.177)	(1.178)	(1.181)
Sex	0.063	0.063	0.064	0.063	0.064
	(0.066)	(0.066)	(0.066)	(0.066)	(0.066)
ln(km)	0.380***			0.379***	0.378***
	(0.060)			(0.060)	(0.060)
ln(speed)	0.072*				0.072*
	(0.035)				(0.035)
ln(urban)	0.460***				
	(0.070)				
ln(age)	-1.367***		-1.364***	-1.366***	-1.368***
	(0.283)		(0.284)	(0.285)	(0.285)
mean.ln(km)		0.378***	0.378***		
		(0.060)	(0.060)		
mean.ln(speed)		0.072*	0.072*	0.070	
		(0.035)	(0.035)	(0.037)	
mean.ln(urban)		0.434***	0.434***	0.428***	0.414***
		(0.079)	(0.079)	(0.078)	(0.078)
mean.ln(age)		-1.365***			
		(0.284)			
sd.ln(km)		0.006	0.007		
		(0.038)	(0.040)		
sd.ln(speed)		0.004	0.000	0.042	
		(0.106)	(0.102)	(0.101)	
sd.ln(urban)		0.090	0.090	0.100	0.120**
		(0.064)	(0.064)	(0.054)	(0.045)
sd.ln(age)		0.019			
		(0.060)			
Log-likelihood	-3197.406	-3197.235	-3197.282	-3197.045	-3196.687
BIC	6449.803	6486.120	6477.049	6467.410	6457.529
AIC	6406.813	6414.470	6412.564	6410.090	6407.374

*** p-value < 0.001, ** p–value < 0.01 and * p-value < 0.05.

References

Adell, E., Várhelyi, A. and dalla Fontana, M., 2011. The effects of a driver assistance system for safe speed and safe distance – A real-life field study. *Transportation Research Part C: Emerging Technologies* 19, 1, 145–155.

Alcañiz, M., Guillen, M., Santolino, M., Sánchez-Moscona, D., Llatje, O., Ramón, Ll., 2014. Prevalence of alcohol-impaired drivers based on random breath tests in a roadside survey in Catalonia (Spain). *Accident Analysis & Prevention* 65, 131-141.

Ayuso, M., Guillen, M. and Alcañiz M., 2010. The impact of traffic violations on the estimated cost of traffic accidents with victims. *Accident Analysis & Prevention* 42, 2, 709-717.

Ayuso, M., Guillén, M. and Pérez-Marín, A.M., 2014. Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accident Analysis & Prevention* 73, 125–131.

Ayuso, M.; Guillen, M.; Pérez-Marín, A. M., 2016a. Telematics and gender discrimination: some usage-based evidence on whether men's risk of accident differs from women's. *Risks* 2016, 4, 2, 10.

Ayuso, M., Guillén, M. and Pérez-Marín, A. M., 2016b. Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance. *Transportation Research Part C* 68, 160–167.

Baecke, P. and Bocca, L., 2017. The value of vehicle telematics data in insurance risk selection processes. *Decision Support Systems* 98, 69-79.

Boucher, J.P. and Guillen, M., 2009. A survey on models for panel count data with applications to insurance. *RACSAM*, Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales, Serie A, Matemáticas 103, 2, 277-294.

Boucher, J. P., Pérez-Marín, A. M. and Santolino, M., 2013. Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident. *Anales del Instituto de Actuarios Españoles*, 3ª Época, 19, 135-154.

Dissanayake, S. and Lu, J. J., 2002. Factors influential in making an injury severity difference to older drivers involved in fixed object-passenger car crashes. *Accident Analysis & Prevention* 34, 5, 609–618.

Elvik, R., Christensen, P. and Amundsen, A.H., 2004. Speed and road accidents. An evaluation of the Power Model. TØI Report 740, Institute of Transport Economics, Oslo.

Fagnant, D.J., Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities: barriers and policy recommendations. *Transportation Research Part A: Policy and Pracice* 77, 167–181.

Guo, C., Sentouh, C., Popieul, J.C., Haué, J.B., Langlois, S., Loeillet, J.J., Soualmi, B. and That, T.N., 2017. Cooperation between driver and automated driving system: Implementation and evaluation. *Transportation Research Part F: Traffic Psychology and Behaviour,* in press. DOI: http://dx.doi.org/10.1016/j.trf.2017.04.006.

Harper, C.D., Hendrickson, C.T. and Samaras, C., 2016. Cost and benefit estimates of partially-automated vehicle collision avoidance technologies. *Accident Analysis & Prevention* 95, 104–115.

Imprialou, M-I. M., Quddus, M., Pitfield, D. E. and Lord, D., 2016. Re-visiting crash–speed relationships: A new perspective in crash modelling. *Accident Analysis & Prevention* 86, 173-185.

Jeong, E., Oh, C. and Lee, S., 2017. Is vehicle automation enough to prevent crashes? Role of traffic operations in automated driving environments for traffic safety. *Accident Analysis & Prevention* 104, 115–124.

Jiménez, F., Aparicio, F. and Estrada, G., 2009. Measurement uncertainty determination and curve-fitting algorithms for development of accurate digital maps for advanced driver assistance systems. *Transportation Research Part C: Emerging Technologies* 17, 3, 225–239.

Jiménez, F. and Naranjo, J. E., 2011. Improving the obstacle detection and identification algorithms of a laserscanner-based collision avoidance system. *Transportation Research Part C: Emerging Technologies* 19, 4, 658-672.

Jun, J., Guensler, R. and Ogle, J., 2011. Differences in observed speed patterns between crash-involved and crash-not-involved drivers: Application of in-vehicle monitoring technology. *Transportation Research Part C: Emerging Technologies* 19, 4, 569-578.

Jun, J., Ogle, J. and Guensler, R., 2007. Relationships between crash involvement and temporal-spatial driving behavior activity patterns: use of data for vehicles with global positioning systems. *Transportation Research Record*, 2019, 246-255.

Kyriakidis, M., Happee, R. and de Winter, J.C.F., 2015. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour* 32, 127–140.

Lemaire, J., Park, S. and Wang, K., 2016. The use of annual mileage as a rating variable. *ASTIN Bulletin* 46, 1, 39-69.

Litman, T. (2011). Pay-As-You-Drive Pricing For Insurance Affordability. Victoria Transport Policy Institute, Victoria.

Paefgen, J., Staake, T. and Fleisch, E., 2014. Multivariate exposure modeling of accident risk: insights from pay-as-you-drive insurance data. *Transportation Research Part A: Policy and Practice* 61, 27-40.

Payre, W., Cestac, J. and Delhomme, P., 2014. Intention to use a fully automated car: attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour* 27, 252–263.

Piao, J. and McDonald, M., 2008. Advanced Driver Assistance Systems from Autonomous to Cooperative Approach. *Transport Reviews* 28, 5, 659-684

Power, J. D., 2012. Vehicle Owners Show Willingness to Spend on Automotive Infotainment Features. <http://autos.jdpower.com/content/press-release/gGOwCnW/2012-u-sautomotive-emerging-technologies-study.htm>.

Rahman, M., Lesch, M.F., Horrey, W.J. and Strawderman, L., 2017. Assessing the utility of TAM, TPB, and UTAUT for advanced driver assistance systems. *Accident Analysis & Prevention*, 108, 361–373.

Reagan, I J., Cicchino, J. B., Kerfoot, L. B. and Weast, R.A., 2018. Crash avoidance and driver assistance technologies — are they used? *Transportation Research Part F: Traffic Psychology and Behaviour* 52, 176-190.

Ruscio, D., Bos, A.J. and Ciceri, M.R., 2017. Distraction or cognitive overload? Using modulations of the autonomic nervous system to discriminate the possible negative effects of advanced assistance system. *Accident Analysis & Prevention* 103, 105–111.

Sheehan, B., Murphy, F., Ryan, C., Mullins, M. and Liu, H. Y., 2017. Semi-autonomous vehicle motor insurance: A Bayesian Network risk transfer approach. *Transportation Research Part C: Emerging Technologies* 82, 124–137.

Son, J., Park, M. and Park, B.B., 2015. The effect of age, gender and roadway environment on the acceptance and effectiveness of advanced driver assistance systems. *Transportation Research Part F: Traffic Psychology and Behaviour* 31, 5, 12–24.

Staubach, M., 2009. Factors correlated with traffic accidents as a basis for evaluating Advanced Driver Assistance Systems. *Accident Analysis & Prevention* 41, 1025-1033.

Strand, N., Nilsson, J., Karlsson, I.C.M. and Nilsson, L., 2014. Semi-automated versus highly automated driving in critical situations caused by automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour* 27, 218–228.

Tselentis, D.I., Yannis, G. and Vlahogianni, E.I., 2017. Innovative motor insurance schemes: A review of current practices and emerging challenges. *Accident Analysis & Prevention* 98, 139-148.

Tselentis, D. I., Theofilatos, A., Yannis, G., and Konstantinopoulos, M., 2018. Public opinion on usage-based motor insurance schemes: A stated preference approach. *Travel Behaviour and Society*, 11, 111-118.

Yu, R. and Abdel-Aty, M., 2014. Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. *Accident Analysis & Prevention* 62, 161-167.

Zantema, J., van Amelsfort, D. H., Bliemer, M. C. J. and Bovy, P. H. L., 2008. Pay-as-you-drive strategies: case study of safety and accessibility effects. *Transportation Research Record: Journal of the Transportation Research Board*, 2078, 1, 8-16.