Anales del Instituto de Actuarios Españoles, 3ª época, 19, 2013/135-154

PAY-AS-YOU-DRIVE INSURANCE: THE EFFECT OF THE KILOMETERS ON THE RISK OF ACCIDENT

Jean-Philippe Boucher [*], Ana M. Pérez-Marín[*±], Miguel Santolino[*]

Abstract

Pay-As-You-Drive (*PAYD*) motor policies are a new concept of insurance contracts which has started to be commercialized in many countries. It is also called Usage Based Insurance (UBI) because, instead of an annual premium be established, the premium is fixed according to the number of kilometers done by the car, besides other characteristics of the risk traditionally used in pricing. Therefore, those who use more the car are going to pay a higher premium because they are more exposed to the risk of accident. In this article we present a bibliographical review on the most relevant contributions about *PAYD* insurances, which are not still sufficiently known in Spain. An empirical application is carried out in which the influence of the number of driven kilometers in the risk of accident is investigated. Four types of claims are studied: property damages/bodily injuries, at-fault/not-at-fault claims. A generalization of the offset Poisson regression model is applied to identify the shape of the effect of driven kilometers on the risk of accident. We show that the association between the number of kilometers and claim frequency is not properly captured by a linear relationship, and alternative forms of relationship are discussed.

Keywords

Automobile insurance, *PAYD* insurance, risk exposure, vehicle usage.

---

[*] Université du Québec à Montréal, 201, avenue du Président-Kennedy, H2X 3Y7 Montreal (Cánada), E-mail: boucher.jean-philippe@uqam.ca (J-P Boucher); Riskcenter, Universidad de Barcelona, Av. Diagonal, 690, 08034 Barcelona (Spain), E-mail: amperez@ub.edu (A.M. Pérez-Marín), msantolino@ub.edu (Miguel Santolino). This research is sponsored by the Spanish Ministry of Science ECO2010-21787-C03-01 and ECO2012-35584.

[±] Author for correspondence.

Este artículo ha sido recibido en versión revisada el 11 de octubre de 2013.

EL SEGURO *PAY-AS-YOU-DRIVE*: EL EFECTO DE LOS KILÓMETROS EN EL RIESGO DE ACCIDENTE

Resumen

Las pólizas *Pay-As-You-Drive* (*PAYD*) del seguro del automóvil son una nueva modalidad de contratos de seguros que han empezado a ser comercializadas en numerosos países. Otro nombre por el que se las conoce es Seguros Basados en el Uso (SBU) porque, en vez de establecer una prima anual, la prima se fija en base al número de kilómetros realizados por el vehículo, además de otras características tradicionalmente utilizadas en tarificación. Básicamente, aquellos que usan más el vehículo pagarán una mayor prima porque tienen una mayor exposición al riesgo de un accidente. En este artículo, realizamos una revisión bibliográfica de las contribuciones más relevantes en el contexto de seguros *PAYD*, los cuales aún no son totalmente conocidos en España. Se lleva a cabo una aplicación empírica en la que se investiga la influencia de los kilómetros en el riesgo de accidente. Se analizan cuatro tipos de siniestros: daños materiales/daños personales, culpa del asegurado/culpa del contrario. Para identificar la forma en la que afectan los kilómetros en el riesgo de accidente se aplican una generalización del modelo de regresión de Poisson con una *offset* covariable. Observamos que el efecto de los kilómetros sobre el riesgo de accidente no se captura en general adecuadamente restringiéndose a una asociación lineal y discutimos relaciones alternativas.

Palabras clave

Seguro del automóvil, seguro *PAYD*, riesgo de exposición, uso del vehículo

1. Introduction

In the traditional automobile insurance contract the premium depends on a number of variables basically describing the characteristics of the driver and the vehicle (such as the age, driving experience, type of the car, among others). Nevertheless, the use of the car, which can be easily measured by the number of annual kilometers, is ignored even though it is clearly associated with the exposure to the risk of accident.

The *Pay-As-You-Drive* (*PAYD*) automobile insurance is a new concept of contract which takes into account the use of the car in order to calculate the premium. In this way, the premiums are more personalized, as those who use less the car are going to pay less than those who do more kilometers during the year. Other variables describing the use of the can be considered as well, such as the speed, type of road or part of the day when the car is most frequently used by the driver.

There are many insurance companies around the world which already offer this product, especially to young drivers. Nevertheless, this new concept of insurance contract is still very new and not sufficiently known. The first objective of this paper is to present a bibliographical review on *PAYD* insurance where we describe the most relevant contributions and implementation results. The second objective is to do an empirical application on the influence of the use of the car in the risk of accident by using a sample of young drivers with a *PAYD* policy in a Spanish insurance company. Four types of claims are distinguished: property damages and at-fault insured driver, property damages and not-at-fault insured driver, bodily injuries and at-fault insured driver, bodily injuries and not-at-fault insured driver. The effect of the number of kilometers driven in the risk of these types of claims is analyzed by means of a generalization of the offset Poisson regression model. The shape of the effect of kilometers on the risk of claim is not imposed to be linear under this framework. We analyze the relationship between the number of kilometers driven by the insured per year and the number of reported claims.

The paper is organized as follows. In section 2 we describe the first proposals of distance-based insurance pricing systems. In section 3, alternative *PAYD* pricing options which can be found in the literature are discussed. In section 4, we do a summary of the results of implementing *PAYD* systems. In section 5, the methodology for count data is introduced and the empirical application is presented. Finally, in section 6 main conclusions are summarized.

2. Distance-based insurance pricing: first proposals

Vickrey (1968) was one of the first authors to criticize the lump-sum pricing of auto insurance for being inefficient as a result of an inappropriate pattern of premium payment. The same author promoted the implementation of distance-based insurance pricing.

Actually, the relationship between the distance run by a vehicle and its influence on the risk of accident has been discussed by many authors. Some of them consider that this relationship is proportional (Bordoff and Noel, 2008) while others argue that it is not proportional (Langford et al., 2008 and Litman, 2005). According to Litman (2005) those drivers who use more the car have fewer accidents per unit of distance (kilometer) than those who use less the car. The main reasons are:

- Those who use more the car have more driving skills than those who are less used to drive.
- Those with more annual kilometers normally use highways (and other safer roads) more than those who use less the car.
- Those who do more kilometers during the year have normally newer vehicles (therefore, safer vehicles) than those who use less the car.

Therefore, Litman (2005) supports the idea of considering the number of kilometers in order to fix the premium probably by using marginally declining per-kilometer premiums.

One of the first proposals of distance-based pricing systems was the pay-at-the-pump (PATP) insurance, where the driver will pay for his coverage as he buys fuel for the vehicle (Vickrey, 1968). Another proposal was the so-called “insured tires” system, where an associated insurance company identified in some way with the tire itself, would cover the accident caused by the vehicle using these tires (Vickrey, 1968). These systems were criticized because they are measuring the use of the car in terms of fuel consumption or tire wear instead of real distance run by the vehicle. Additionally, these systems do not distinguish between good and bad drivers when charging the cost of the insurance (Khazzoom, 2000 and Guensler et al., 2003). The other possibility was to measure the distance driven by the car by odometer auditing. In that case, there are concerns that fraud could be a problem, but some authors consider that odometers are increasingly tamper-resistant (Litman, 2011) therefore it seems reasonable to propose *PAYD* pricing systems based on odometer audits.

In the next section, we will see that nowadays advanced technologies make it possible to measure the use of the car objectively (by using a GPS system). Therefore, sophisticated *PAYD* pricing systems can be proposed based not only on the distance driven by the vehicle, but also on the speed, type of road and part of the day when the car is most frequently used.

3. *PAYD* pricing options

Many authors have proposed different *PAYD* pricing options. Litman (2011) discussed the following ones:

1. *Mileage Rate Factor (MRF):* The annual number of kilometers is used as a rating factor into premiums. The insurance company offers a discount for customers driving less than a certain level. This system has been criticized (Litman, 2011) because it is normally based on the driver's self-reported estimate and there is no verification or adjustment at the end of the policy term.
2. *Per-Mile-Premium (PMP):* Under this system there is a price per unit of distance (mile, kilometer,…) done by the vehicle also taking into account other rating factors. As mentioned in the previous section, Litman (2011) advocates for declining per-mile (or per-kilometer) premiums. The customer will prepay for the number of kilometers he expects to drive during the policy term (in that sense, the company will require to buy a minimum number of kilometers to insure that transactions costs are covered). There is a verification or adjustment at the end of the policy term based on odometer auditing. There are different proposals to deal with the unused or outstanding kilometers, in some cases requiring some mechanism to ensure that the customer will pay any outstanding fee.
3. *GPS-Based pricing:* This is the *PAYD* pricing option that most insurance companies offer nowadays. In that case, the price can be fixed based not only on the kilometers, but also on the speed, time and location of the vehicle. Speed is normally considered in terms of violation of the limits. Regarding time, the difference is normally established in terms of daily/nightly driving, being the night period more expensive. Finally, regarding location, this system distinguishes between urban and not-urban roads, being urban driving more expensive. It is necessary the permission of the driver for installing the GPS equipment. In that sense, the legal framework and privacy concerns of *PAYD* policies have been discussed by

many authors, especially in the USA (Iqbal and Lim, 2006; Bingham et al., 2009 and Guensler et al., 2003).

4. Results of implementing PAYD systems

The advantages of commercializing *PAYD* contracts, both for the insurance company and the driver, have also been discussed in the literature (Peña, 2007). Regarding the insurance company, the actuarial accuracy of premiums will improve as a result of a better quantification of the exposure to the risk of accident of each driver. In this way, the insurer will also have a better segmentation of the market. Additionally, according to Hagerbaumer (2004) companies that offer *PAYD* will be viewed as customer-oriented, proactive and environmentally responsible (as *PAYD* incentives to reduce the use of the car). Therefore it can help the company to improve their corporate image and potentially increase its market share. The advantages for customers are clear, they will pay a lower premium if they do less kilometers or drive in a safer way. Additionally, it makes insurance more affordable, therefore it seems reasonable to think that the number of cars without insurance will be reduced (Peña, 2007).

There are also many papers focused on the changes in the driving patterns of individuals who want to get a better premium under a *PAYD* system. Namely, Buxbaum (2006) concluded that under this system drivers have an incentive to drive less and quantified that this decline in driving was around 8% as a result of an experiment carried out in Minnesota (USA). The same conclusion was obtained by Bordoff and Noel (2008) in an empirical study carried out in California (USA). Other authors conclude that this reduction could be higher, around 10% or even more (Hagerbaumer, 2004). Moreover, it seems to be clear that the reduction in driving would result in an even higher reduction in terms of accidents according to Edlin and Mandic (2006). One reason for this is that each vehicle removed from traffic reduces both its chances of causing an accident and of being the target of a crash caused by another vehicle, therefore, multi-vehicle accidents are also reduced (Edlin, 2003).

Additionally, under this system drivers have an incentive not only to drive less but also to drive in a safer way; therefore it contributes to increase traffic safety. Namely, Bolderdijk et al. (2011) carried out an empirical study in The Netherlands to test the effects of a *PAYD* insurance fee on driving speed and concluded that has a significant impact on the reduction of speed violations among young drivers.

Finally, we must say that *PAYD* insurance contracts contribute to reduce traffic congestion, road maintenance costs, energy consumption and air pollution (Peña, 2007). Additionally, Parry (2004 y 2005) concluded that under certain conditions, *PAYD* contracts are a better way to reduce gasoline consumption than gasoline taxes.

As a conclusion, we can say that *PAYD* policies are a new concept of insurance contract with potential advantages for customers, insurers and the society as a whole. In nowadays context, insurance companies which want to be customer-orientated should be able to offer personalized products, and *PAYD* insurance is an example which maybe in the future can be exported to other types of contracts.

5. The use of car on the risk of accident: the Spanish experience

An empirical application is shown in which we analyze the effect of the use of the car on the risk of accident by using a sample of young drivers with a *PAYD* policy. The description of the dataset and main statistics are shown in section 5.1. The Poisson regression model with a offset covariate is introduced in section 5.2. Finally, a generalization of the regression model and main results are shown in sections 5.3 and 5.4.

5.1 Data

A dataset of 25,014 motor *PAYD* policies was provided by a Spanish insurance company. All these motor policies were in force in the year 2011. This type of policy was mainly addressed to young drivers according to the strategy of the insurance company. In the dataset all the insured drivers were under 40 years old. Claims were not reported by the 82.4% of the *PAYD* policies. That means, the 17.6% of the total amount of *PAYD* policies reported at least one claim during the year.

Four different categories of claims are distinguished in the dataset: claims involving property damages in which the insured is the at-fault driver (*Nb1*), claims involving property damages in which the insured is not the at-fault driver (*Nb2*), claims involving bodily injuries in which the insured is the at-fault driver (*Nb3*), claims involving bodily injuries in which the insured is not the at-fault driver (*Nb4*). The number of claims of each type reported by the insured was collected by each policy. The empirical distributions of claims involving property damages and claims involving bodily injuries are shown in Table 1a and Table 1b, respectively.

Table 1a. Number of property damage claims reported by insureds

	At-fault insured (*Nb1*)		Not-at-fault insured (*Nb2*)	
Number of claims	Percentage	Aggregated percentage	Percentage	Aggregated percentage
0	91.54	91.54	90.37	90.37
1	8.02	99.56	9.00	99.36
2	0.42	98.98	0.62	99.98
3	0.02	100.00	0.02	100.00

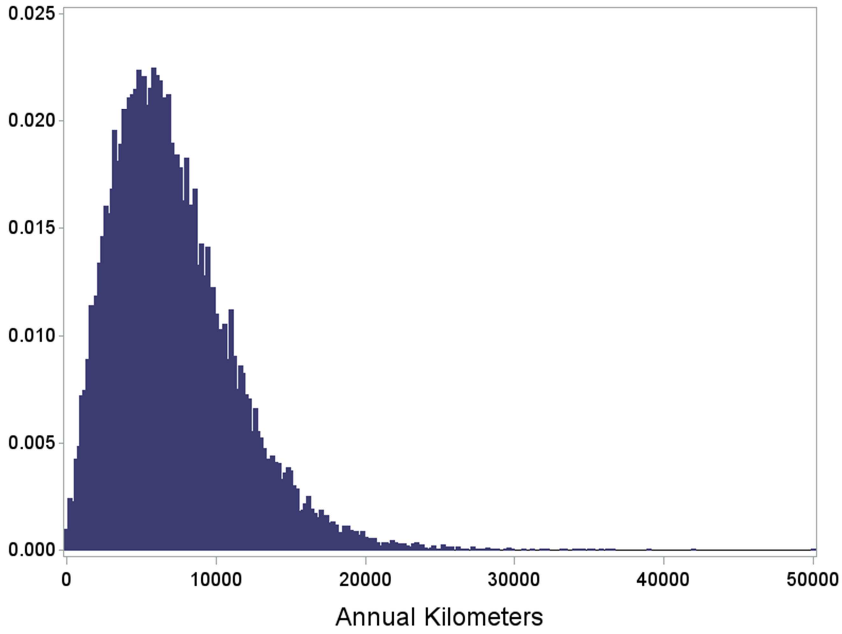
Table 1b. Number of bodily injury claims reported by insureds

	At-fault insured (*Nb3*)		Not-at-fault insured (*Nb4*)	
Number of claims	Percentage	Aggregated percentage	Percentage	Aggregated percentage
0	98.34	98.34	97.94	97.94
1	1.63	99.97	1.98	99.92
2	0.03	100.00	0.08	100.00

Similar percentages are observed whether we compare the empirical distributions of claims in which the insured was the at-fault driver and claims in which the insured was not the at-fault driver. Note that more than the 90% of insured drivers did not report any claim involving property damages. Additionally, most of the policies involving property damage claims have just reported one claim. We emphasize that less than 1% of the policies reported two or more property damage claims (Table 1a). Taking a look on bodily injury claims, differences between the empirical distributions are again not relevant regardless who was the responsible of the accident. Let mention that approximately the 2% of insured drivers reported bodily injury claims (Table 1b). Therefore, the 98% of policies did not report bodily injury claims.

The continuous variable representing the number of driven kilometers for a year (*Km*) is our main interest in the project. Information related to the number of kilometers driven by the insured in the year 2011 was provided by the insurance company. Traditionally, this information is reported by the insured but not verified by the insurance company. In this case, this information was recorded by means of GPS navigation systems. Therefore, we know the number of kilometers effectively driven by the insured. Figure 1 illustrates this distribution for our database. The average annual driven kilometers is 7,160, with a standard error of 4,192. The empirical distribution of the number of driven kilometers per year is skewed to the right.

Figure 1. Distribution of the driven kilometers for a year



Additionally to the number of driven kilometers, characteristics of policyholders traditionally considered by insurance companies in motor ratemaking were collected. Namely, information regarding the age (AGE) of the insured driver and the age of the vehicle was recorded (V_AGE). These two continuous variables were classified into categorical groups. Indeed, the age of the driver (AGE) and age of the vehicle (V_AGE) were categorized in two and three binary variables, respectively. Other information collected was the gender of the insured (X7) and whether the automobile is usually parked inside a private garage (X8). The description of all of these binary variables is presented in Table 2.

Table 2. Binary variables summarizing the information available about each policyholder

Variable	Description
X2	equals 1 if the insured is 25 years old or younger
X3	equals 1 if the insured is between 25 and 30 years old
X4	equals 1 if the insured car is less than 2 years old
X5	equals 1 if the insured car is between 2 and 5 years old
X6	equals 1 if the insured car is between 5 and 10 years old
X7	equals 1 for men and 0 for women
X8	equals 1 if the car is parked inside a private garage

Main descriptive statistics are shown in Table 3. As indicated above, insured drivers drove on average around 7,150 kilometers in the year 2011. However, great differences in the number of driven kilometers are found between insureds as indicated by the sample bounds. The sample insured who made the lowest amount of kilometers with the automobile in 2011 was less than 2 kilometers. On the other extreme, the maximum value observed in the dataset is 50,036 kilometers, which means that an insured drove more than 50,000 kilometers in 2011.

In relation to the age of the driver, the mean age was around 26 years old. Remember that this product was addressed to young drivers. The sample range of driver's age is between 19 and 33 years. Whether we analyze the associated binary variables X2 and X3, we conclude that the 90% of the sample drivers with a *PAYD* policy were 30 years old or younger. The target group to address *PAYD* policies is to young drivers, because this collective of inexperienced drivers traditionally pay higher premiums. A priori, then, these young drivers may be more interested to show to the insurance company how good drivers they are in order to pay lower premiums.

Another factor normally associated with the frequency of claims is the age of the automobile. Automobiles of insured drivers were on average almost 7 years and half old, where the oldest sample vehicle was 19 years. Taking a look on binary variables X4, X5 and X6, we observe than the 8.5% of the automobiles were less than two years old and almost the 40% of the automobiles less than 5 years old. On the opposite, the 23% of the automobiles were more than 10 years old. Regarding the gender of the driver, it is observed almost the same proportion of men and women in the sample (X7). Finally, three-quarters of the sample vehicles are parked in a private garage. More details about the data are found in Alcañiz-Zanón et al. (2013).

Table 3. Descriptive statistics of variables

Variable	Number	Mean	St. Dev.	Min.	Max.
Nb1	25,014	0.089	0.301	0	3
Nb2	25,014	0.103	0.325	0	3
Nb3	25,014	0.017	0.131	0	2
Nb4	25,014	0.021	0.15	0	2
KM	25,014	7,159.51	4,191.75	1.593	50,035.56
AGE	25,014	26.288	3.103	19	33
V_AGE	25,014	7.411	4.217	1	19
X2	25,014	0.424	0.494	0	1
X3	25,014	0.476	0.499	0	1
X4	25,014	0.085	0.279	0	1
X5	25,014	0.308	0.462	0	1
X6	25,014	0.379	0.485	0	1
X7	25,014	0.489	0.5	0	1
X8	25,014	0.769	0.421	0	1

5.2. The Poisson Regression Model: annual kilometers as an offset variable

In the Poisson distribution, the risk exposure (t) is traditionally included into the mean parameter (λ) by working with a Poisson with a mean equal to $t \times \lambda$, where *t* indicates the time period that the driver was insured. In the modeling of the number of claims, such a model means that the insured time should be introduced as an offset in the Poisson regression (e.g. the OFFSET option in the GENMOD procedure of the SAS System). When we include covariates, we then have the following expression for the regression,

$$\lambda_i = t_i \times \exp(x_i^{'} \beta) = \exp(x_i^{'} \beta + \log(t))$$

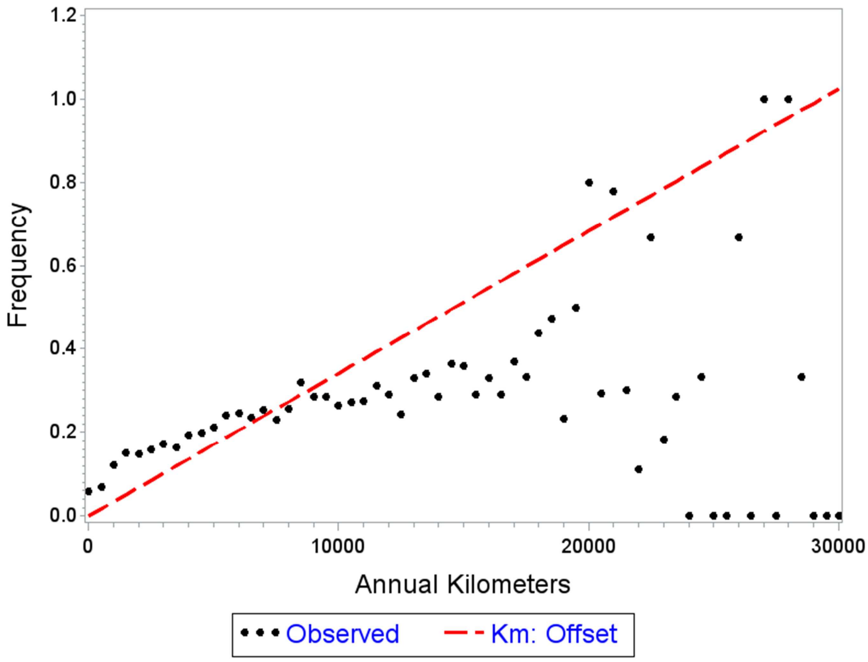
where λ_i is the frequency of claims for the motor policy *i*, *i*=1,…, *I*. The covariates are associated to the mean parameter of the Poisson model via exponential transformations, where $\beta' = (\beta_1, \ldots, \beta_p)$ is a vector of regression parameters for the binary explanatory variables $x_i^{'} = (X_{i,1}, \ldots, X_{i,p})$, with $X_{i,1} = 1$ for all *i*, meaning that β_1 can be considered as the intercept. The reason to this intro is that the risk exposure (when expressed in term of covariate) should be expressed in a logarithmic form. The interpretation is that if someone is insured 6 months, his premium will normally be 50% cheaper that an insured with a one year coverage (for details, see Boucher and Denuit, 2007).

Following the same intuition with the number of driven kilometers, we expect that the drivers who use more the automobile are going to have a higher frequency of claims. The first model an insurer would use to model this covariate is a direct relationship between the premium and the annual number of kilometers. If someone drives 10,000 km in a year, he should normally expect to have a premium that is 50% cheaper than a similar driver who drives 20,000 km. In the modeling of the number of claims by means of a Poisson regression with an offset covariate, such a model should be expressed as,

$$\lambda_i = Km_i \times \exp(x_i^{'} \beta) = \exp(x_i^{'} \beta + \log(Km_i))$$

where Km_i is the number of driven kilometres for the policy *i*, *i*=1,…, *I*. To illustrate the situation, Figure 2 shows the prediction of the frequency of all kind of claims using a Poisson regression without covariates, with the annual number of kilometres as an offset variable. About this graph, the dots represent the average frequency when the insureds are grouped by intervals of 500 driven kilometers. Note that above 20,000 km the data seems more heterogeneous. It is due to the fact that there are few insureds on this area. The sample variance was quite similar for all the kilometers. This result indicates that there is not a relevant heterogeneity in the frequency of claims reported by drivers according to the number of driven kilometers. We can see that the trend line seems to underestimate the frequency of claims for drivers below 10,000 km per year and to overestimate the frequency of claims for drivers with more than 10,000 km per year.

Figure 2. Prediction of the frequency of claims using the annual kilometers as an offset variable



5.3 Generalization of the offset Poisson Regression Model

In the previous section we discussed that a linear association between the frequency of claims and the number of kilometers seems highly restrictive and unrealistic. One way to correct the model is to consider than an offset variable into a Poisson regression model with mean lambda means represent another covariate in the model. Indeed:

$$\lambda_i = Km \times \exp(x_i^{'}\beta) = Km \times \exp(x_i^{'}\beta + \log(Km))$$

The classic offset regression model (Boucher and Denuit, 2007) can then be generalized and expressed with the following form:

$$\lambda_i = \exp(x_i^{'}\beta + c \times \log(Km)) = \exp(x_i^{'}\beta) \times Km^c, \qquad (1)$$

when c=1 then the model collapses to the classic offset regression model. Consequently, model (1) is a particular case of a Poisson regression model (which assumes that there is not overdispersion in the data) where *log(Km)* is introduced as a covariate. The resulting model let us to analyse the effect of driven kilometers on the risk of accident. We can estimate this Poisson model by maximum likelihood and test if c is different from 1. This should be considered as the null hypothesis. This model has been tested on the four different categories of claims in our database. Results of the estimation are shown in Tables 3a and 3b. Covariates with non-significative parameter estimates have been removed from the model. Analysis of the parameter c clearly shows that the annual kilometers per year cannot be directly used as a classical offset variable in the modeling of the number of claims because all c parameters are statistically different than 1.

Table 3a. Parameter Estimates for a Poisson Regression (Property damages)

	Nb1		*Nb2*	
	Est.	Std. Err	Est.	Std. Err
X1	-3.4445	0.0923	-3.5841	0.1131
X2	0.2944	0.0424	0.3452	0.0758
X3	-	-	0.1587	0.0762
X7	0.0859	0.0428	-	-
X8	-	-	0.1086	0.0485
c	0.4188	0.0418	0.4903	0.0389

Table 3b. Parameter Estimates for a Poisson Regression (Bodily injuries)

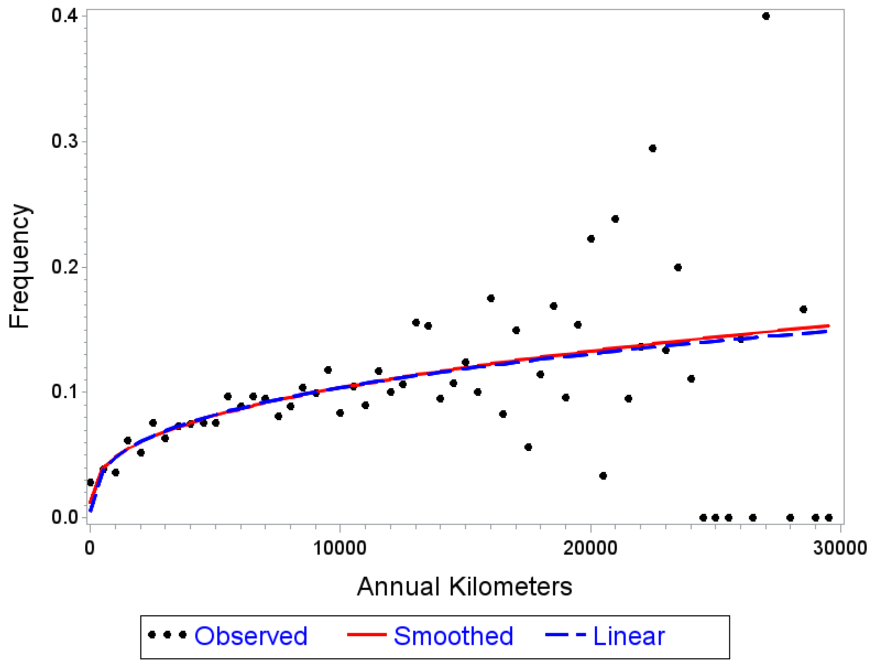
	Nb3		*Nb4*	
	Est.	Std. Err	Est.	Std. Err
X1	-5.4141	0.2170	-4.8246	0.1893
X2	0.4417	0.0975	0.3134	0.0866
X4	-	-	-0.7137	0.2051
X7	-	-	-0.2459	0.0878
c	0.5505	0.0967	0.4908	0.0860

Note that parameter estimates for the binary variable indicating that the driver was younger than 25 years old are significant in the four regressions. Therefore, drivers under 25 years are expected to have a higher number of all types of claims. An interesting result is that the age of the insured automobile was not significant to explain the number of claims in which the insured driver was at-fault. Unlike at-fault claims, *PAYD* policies covering automobiles less than two years old are expected to report less bodily injury claims in which the insured is not at-fault. The variable recording if the automobile is usually parked inside a private garage only showed a significant parameter estimate in the regression of the number of property damage claims without responsibility of the insured. Finally, we expect that males report more property damage claims in which they were the at-fault driver and less bodily injury claims in which they were not responsible of the accident.

5.4 Effect of kilometers on the risk of accident

The effect of kilometers driven by the insured on the risk of accident is plotted in Figure 3. Dots inside the figure represent the observed frequency of claims, with insureds grouped by intervals of 500 kilometers. We fitted a non-parametric curve to data (red line). The shape of the smoothed line reflects that the experience of the driver play a role on the risk of an accident. The frequency of claims is far away to increase linearly with the number of driven kilometers per year. A high slope is observed at the beginning of the line, which marginally decreases as more kilometers are driven by the insured. Although always positive, the slope of the graph reaches the minimum value in the range between 15,000 and 20,000 kilometers. From this point on the slope seems to be almost constant. So, the relationship between the frequency of claims and the annual kilometers would be approximately linear in this upper range.

Figure 3. Prediction of the frequency of claims using the annual kilometers as a generalized offset variable



The decline in the slope of the graph could reflect the driver's expertise and other related safety factors. We interpret this result as drivers who use more the automobile are more experience and skilful drivers because they have more driving hours. Additionally, these drivers often use new (and safer) automobiles to make these kilometers. Other important element that may explain this result is that they usually drive a higher proportion of kilometers on highways which are safer than conventional roads. As a result, the higher risk associated to drive more kilometers is in part balanced by the combination of all of these factors. The positive effect of the higher experience (and associated safety elements) in reducing the frequency of claims is, however, dismissed for drivers who make more than 15,000-20,000 kilometers per year.

In subsection 5.3 we discussed a log-linear association which seems to be adequate and reflects the relationship between the number of kilometers per year and number of claims. However, the parameter in front of the regressor of the log(*Km*) is a way different from 1 (meaning that the value *c* is different than 1). For illustration purposes, in the Figure 3 it is showed the prediction of the frequency of all kind of claims using a Poisson regression without covariates, with the annual number of kilometres as generalization of the offset variable. An almost perfect match is observed between the smooth curve and the expected frequency of claims when the number of driven kilometers is used as generalization of the offset variable in the Poisson model (blue line).

We argue that the parameter *c* collects the effect of the unobserved factors (driver's experience and safety elements) on the decline of the proportionally between the number of driven kilometers and the claim frequency. Note that the expected frequency of claims is multiplied by the annual number of kilometers raised to the power *c*. Whether *c* value tends to one, it indicates that the driver's experience and the other safety factors have not effect on reducing the frequency of claims. In that case, a person who drives 10,000 km in a year would be expected to have on average the half number of claims that a similar driver who drives 20,000 km. Consequently, the premium of the latter should be twice his premium[1]. On the opposite, a value of *c* close to zero means that the risk associated to a higher number of driven kilometers is fully balanced by the larger experience of the driver and the other safety factors. As a result, the expected frequency of claims is not increased by the number of driven kilometers. In the example, the person driving 20,000 km would pay the same premium as who drove 10,000 km. For this plot, the estimated value of *c* was approximately equal to 0.4. That means that, for these two similar divers, who drives 20,000km should pay $2^{0.4}$ times the premium of the insured that drives 10,000km.

[1] Let assume that we observe same values for regressors of two insured drivers. Whether $E(N_1)$ is the expected number of claims of a insured who drives 10,000km and $E(N_2)$ is the expected number of claims of a insured who drives 20,000km, then $E(N_2)/E(N_2)$ is equal to $20{,}000^c/10{,}000^c = (20{,}000/10{,}000)^c = 2^c$.

6. Conclusion

PAYD motor policies have recently gained a lot of attention as a new concept of motor insurance. *PAYD* policies allow insurance companies to offer a *personified* premium to the driver based on the use of the vehicle that he/she actually does. In a highly competition context such as the motor insurance market, this new type of insurance is viewed as a very interesting product by the insurance company. In that sense, insurance companies and drivers are also interested in '*pay how you drive*' insurance policies, where covariates describing the way a person drives are also considered (speed, type of road, time of driving, braking and cornering, etc…). The relationship between these covariates and the risk of accident has been investigated by many authors (Elvik et al., 2004; Jun et al., 2007, Calafat et al., 2008 y Laurie, 2011).

However, *usage-based-insurance* policies involve new challenges that are still not enough known. For a better understanding of the phenomenon, a deep review of literature is provided in this article where the most relevant contributions are presented.

On the study of *PAYD* motor policies, the number of kilometers driven by the insured is a key factor to determine the insurance premium. The exposure measurement of risk of an accident will be a function of the number of driven kilometers. We investigate the effect of the number of kilometers made by the automobile per year on the risk of having an accident. We showed that the effect of kilometers on the risk of claim is far away to be linear. We discuss that alternative forms seem to be more adequate to capture the association between the number of kilometers driven by the insured and the number of claims. It is may be explained because drivers who make more kilometers are more experienced drivers with newer automobiles and who drive more often in safer highways. We believe that this result is highly relevant for motor insurance companies in order to include adequately the information related to the number of driven kilometers on the ratemaking design of the *PAYD* motor policies.

References

Alcañiz-Zanón, M., Ayuso-Gutierrez, M., Pérez-Marín, A.M. (2013). El Seguro Basado en el Uso (Usage Based Insurance), Estudios de la Fundación MAPFRE, en prensa.

Bingham, K., Boyd, P., Greene, M. and Brown, R. (2009). Pay-As-You-Drive: some considerations before you hit the road, *Contingencies*, Jan/Feb. 26-28.

Bolderdijk, J.W., Knockaert, J., Steg, E. M. and Verhoef, .E. T. (2011). Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: results of a Dutch field experiment, *Accident Analysis and Prevention*, 2011, 43, 1181-1186.

Bordoff, J. E. and Noel, P. J. (2008). Pay-As-You-Drive Auto Insurance: A Simple Way to Reduce Driving-Related Harms and Increase Equity, *The Brookings Institution*, Discussion Paper.

Boucher, J.-P. and Denuit, M. (2007), Duration Dependence Models for Claim Counts, Deutsche Gesellschaft fur Versicherungsmathematik (German Actuarial Bulletin), 28-1, 29-45.

Buxbaum, J. (2006). Mileage-Based User Fee Demonstration Project: Potential Public Policy Implications of Pay-As-You-Drive Leasing and Insurance Products. Cambridge Systematics, Inc.Calafat, A., Adrover, D., Juan, M. y Blay, N.T. (2008) Relación del consumo de alcohol y drogas de los jóvenes españoles con la seguridad vial durante la vida recreativa nocturna en tres Comunidades Autónomas en 2007. *Revista Española de Salud Pública*, 82, 323-331.

Edlin, A.S. (2003). Per-mile premiums for auto insurance. In *Economics for an imperfect world: essays in honor of Joseph E. Stiglitz*, edited by Richard Arnott, Bruce Greenwald, Ravi Kanbur and Barry Nalebuff, 53-82. Cambridge, MA: MIT Press. Disponible en: http://works.bepress.com/aaron_edlin/28.

Edlin, A. S. and Mandic, P. K. (2006). The accident externality from driving, *Journal of Political Economy*, 114, 5, 931-955.Elvik, R., Christensen, P. y Amundsen, A. (2004) *Speed and road accidents. An evaluation of the Power Model*. TØI report 740/2004. Institute of Transport Economics TOI, Oslo.

Guensler, R., Amekudzi, A., Williams, J., Mergelsberg, S. and Ogle, J.: (2003). Current state regulatory support for Pay-As-You-Drive automobile insurance options, *Journal of Insurance Regulation*, National Association of Insurance Commissioners, 21, 3.

Hagerbaumer, C. (2004). Drive-By Rates: Can Pay-As-You-Drive Insurance Attract Good Risks And Gain Insurers An Environmentally Responsible Image?, *Best's Review*, April, 68-69.

Iqbal, M. U., and Lim, S. (2006). A privacy preserving GPS-based Pay-as-you-drive Insurance Scheme, *International Global Navigation Satellite System Society 2006 Symposium*.
Jun, J., Ogle, J. y 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.
Khazzoom, J.D. (2000). Pay-at-the-Pump auto Insurance: Review of criticisms and proposed modifications, *Journal of Insurance Regulation*, 18, 448-496.
Langford, J.; Koppel, S.; McCarthy, D. and Srinivasan, S. (2008). In defence of the 'low-mileage bias, *Accident Analysis and Prevention*, 40, 1996–1999.Laurie, A. (2011) *Telematics: the new auto insurance*. En: Emphasis. - New York: Towers Watson. Número 1, 20-25.
Litman, T. (2005). Pay-As-You-Drive pricing and insurance regulatory objectives, *Journal of Insurance Regulation*, National Association of Insurance Commissioners, 2005, 23, 3.
Litman, T. (2011). Pay-As-You-Drive insurance: recommendations for implementation, Victoria Transport Policy Institute.
Parry, I.W.H. (2004). Comparing alternative policies to reduce traffic accidents, *Journal of Urban Economics*, 56, 346-368.
Parry, I.W.H. (2005). Is Pay-As-You-Drive insurance a better way to reduce gasoline than gasoline taxes?, *American Economic Review*, 95, 2, 288-293.
Peña Pérez, L. (2007). Nuevos modelos de gestión de seguros: Mapfre, su proyecto Generación Y, *Carreteras*, noviembre-diciembre, 156, 54-68.
Vickrey, W. (1968). Automobile accidents, tort law, externalities, and insurance: an economist's critique, *Law and Contemporary Problems*, 33, 464-487.