Using GPS data to analyze the distance traveled to the first accident at fault in pay-as-you-drive insurance

Mercedes Ayuso[a,*], Montserrat Guillén[a], Ana María Pérez Marín[a]

[a]Riskcenter, University of Barcelona, Av. Diagonal, 690, 08034 Barcelona (Spain)

Abstract

In this paper we employ survival analysis methods to analyze the impact of driving patterns on distance travelled before a first claim is made by young drivers underwriting a pay-as-you-drive insurance scheme. An empirical application is presented in which we analyze real data collected by a GPS system from a leading Spanish insurer. We show that men have riskier driving patterns than women and, moreover, that there are gender differences in the impact driving patterns have on the risk of being involved in an accident. The implications of these results are discussed in terms of the 'no- gender' discrimination regulation.

Keywords: motor insurance, pricing, telematics, gender discrimination, survival analysis.

1. Introduction

Traditionally, men have paid more than women for their automobile insurance. Indeed it was a recognized exception to the EU's so-called Gender Directive (officially Council Directive 2004/113/EC of 13 December 2004) implementing the principle of equal treatment between men and women in the access to and supply of goods and services. However, since December 2012, insurance companies have no longer been allowed to charge different rates according to the driver's gender following the ruling of the European Court of Justice (ECJ), issued on 1 March 2011, invalidating the use of gender as a rating factor in insurance, although based on relevant and accurate actuarial and statistical data (Aseervatham et al., 2016; Sass and Seifried, 2014; Schmeiser et al., 2014).

In this new legal framework, insurance companies have had to tackle the problem of establishing a unisex rating system in which the proportion of men and women in the portfolio acquires considerable importance. The task is a challenging one, especially in the case of life insurance where it is not easy to find alternative risk factors that can explain the probability of a claim, once gender has been excluded from the rating system.

In the case of traditional automobile insurance, the Gender Directive has also had important repercussions. Thus, in a similar way, insurance companies have had to fix their prices by taking into account the composition of their portfolios, while bearing in mind that different risk classes do exist (Guillen, 2012). However, here it is relatively easy to identify additional risk factors to compensate the elimination of gender from the calculation of the premium, particularly usage-based systems such as pay-as-you-drive (PAYD) (Paefgen et al, 2014).

* Corresponding author at Riskcenter, University of Barcelona, Av. Diagonal, 690, 08034 Barcelona (Spain).
Tel: +34 93 402 14 09. E-mail address: mayuso@ub.edu (M. Ayuso).

In PAYD automobile insurance, the premium is calculated on the basis of vehicle usage. In this way, premiums are personalized, that is, occasional drivers pay less than frequent users. In addition, the policyholder's driving pattern is also taken into consideration in calculating the premium. Thus, drivers' speed profiles, the type of roads they most frequently take and the time of day they are typically on the roads are taken into account in the rating system, since these factors have been shown to explain the likelihood of being involved in an accident (Litman, 2005; Sivak et al., 2007; Langford et al., 2008; Ayuso et al., 2010; Jun et al., 2007 and 2011; Ellison et al, 2015). All this information is normally collected by a GPS system that the insurance company installs in the insured vehicle. The insured person allows the insurer to place this device in the vehicle in return for a bonus in the premium or any other discount (such as providing vouchers for fuel).

Today, many insurance companies around the world sell PAYD contracts, especially to young drivers; however, given the only recent introduction of this system, little is known about it. Actuaries typically fix the premium in line with the number of kilometers travelled and the frequency of use of the car. Thus, a company will usually offer a discount depending on the distance travelled during the year and the driving patterns. Furthermore, the company establishes certain thresholds for the number of yearly kilometers driven, the percentage of urban and nighttime driving, and the percentage of kilometers travelled above the mandatory speed limit. In this way, policyholders that respect these thresholds are entitled to a discount while the others suffer a penalization. The pricing difficulty then is in determining where to fix these threshold values and what their corresponding discounts should be. In order to do this, actuaries need to know how a driver's accident risk is influenced by the distance travelled and their driving pattern.

However, while common sense dictates that the greater the exposure to the risk of an accident, the greater the probability of actually having an accident, drivers clearly acquire experience as they drive more and so their risk of being involved in an accident diminishes (McCartt et al., 2003). However, at a certain juncture, it is believed that drivers reach a stable level of driving skills and patterns (Underwood, 2013), and from that moment onwards the risk of having an accident is proportional to distance travelled, with this proportion being dependent on rating factors that determine the price of insurance coverage.

In this paper, we seek to estimate the number of kilometers driven before the first accident occurs during coverage by a PAYD policy, as a function of the policyholder's driving pattern. In this way, we can determine the impact of driving patterns on the risk of accident, and we should obtain an estimation of the expected number of kilometers travelled before the first accident occurs. To do this, we employ a survival analysis technique, namely, a Weibull regression model that can explain the distance travelled as a function of driving patterns. We present an empirical application with real data from a leading Spanish insurance company in which gender differences can be seen in the impact of driving patterns on the risk of accident. Examples for different types of driver are presented. Finally, the implications of these results in terms of the 'no gender' discrimination regulation are discussed.

The paper is organized as follows. In section 2, we present a bibliographical review of PAYD insurance in which we describe the main outcomes of its implementation. In section 3 the survival methodology used to estimate the expected distance travelled before the first accident is presented. In section 4, we present the results of the empirical application and, finally, in section 5 we conclude.

2. Background

Vickrey (1968) was one of the first authors to promote the implementation of distance-based insurance pricing, and to criticize the lump-sum pricing of auto insurance on the grounds of inefficiency. Yet, the relationship between the distance traveled by a vehicle and the risk of accident has been questioned by many authors, with most concluding that it is not proportional (Langford et al., 2008; Litman, 2005). More recently, Boucher et al. (2013) have shown that the association between the number of kilometers traveled and the claim frequency is not properly captured by a linear relationship, and they discuss other possibilities.[1]

Pay-at-the-pump (PATP) insurance was one of the first distance-based pricing systems to be introduced. Under this system, the driver paid for his coverage as he bought fuel for the vehicle. 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). The main criticism leveled at these systems was that the use of the car was measured in terms of fuel consumption or tire wear, instead of the actual distance covered by the vehicle. Additionally, these systems failed to distinguish between good and bad drivers when fixing the insurance charges (Khazzoom, 2000; Guensler et al., 2003). An alternative scheme involved measuring the distance driven by the car using an odometer auditing system. In this case, however, there were concerns regarding potential fraudulent practices. Today, technological advances mean the use of the car can be measured objectively employing a GPS system and associated sophisticated PAYD pricing systems, which is the pricing option offered by most insurance companies. Note, however, that the permission of the driver is always required before installing the GPS equipment. Once in place the price can be fixed in relation not only to the distance driven, but also to the speed (typically in terms of the percentage of kilometers travelled above the mandatory speed limit), time (that is, in terms of daily/nighttime driving, with nights being more expensive) and location (with a distinction being drawn between urban and non-urban driving, with the former being more expensive).

Analysis has generally been made without considering correlation between drivers in the insurance portfolio. However, from our point of view, geo-demographic profiles of the drivers could also be of great interest in the context of PAYD insurance, and even become a line for future research. For example, Quddus (2015) demonstrates that the severity of injuries of urban drivers involved in crashes increased if they traveled to rural areas (one of the main variables that is included in our analysis without considering correlation). In a similar way, Lee et al. (2014) show that the crash occurrence is not only affected by roadway/traffic factors but also by several demographic and socioeconomic characteristics of residence zones where drivers at fault live.

[1] Three main reasons can justify the existence of a non lineal relationship between the number of accidents and the distance travelled by the driver according to the literature (Litman, 2005; Boucher et al., 2013), namely: i) Presence of more driving skills in those drivers who use the car more than others; ii) A more frequent use of highways and other safer roads by those with more kilometers per year than the average; iii) Newer and safer vehicles used by those who drive more than the others (more frequent change of vehicle along time to this kind of drivers).

The advantages of commercializing PAYD contracts for the insurance company and the driver alike have also been examined (Bolderdijk et al., 2011; Litman, 2011). For insurance companies, one advantage is that each driver's exposure to the risk of being involved in an accident can be measured more accurately, thus enhancing the actuarial fairness of premiums. Moreover, the insurer can also obtain a more sophisticated segmentation of the market compared to the traditional risk classes. In this regard, Hultkrantz et al. (2012) claim that PAYD helps the insurance industry to target risk classes more effectively. Likewise, Ayuso et al. (2014) report that the driving patterns and accident rates of young drivers are heterogeneous and so these policyholders should not be included in homogeneous risk groups, as frequently occurs in the insurance industry. Additionally, offering PAYD policies can help the company improve their corporate image ensuring they are perceived as customer-oriented, proactive and environmentally responsible (since PAYD schemes provide incentives to reduce vehicle use). For customers, the advantages are clear: they pay a lower premium if they drive fewer kilometers and drive safely. In short, many authors claim that PAYD makes insurance more affordable while rewarding careful driving.

In this regard, the literature reports evidence of drivers modifying their driving patterns so as to obtain a better premium under a PAYD system. For example, Bolderdijk et al. (2011) observed a significant impact on the reduction in speed violations among young drivers with a PAYD policy. Additionally, Lahrmann et al. (2012) and Toledo et al. (2008) also found evidence of the positive effect of in-vehicle data recorders and monitoring equipment on speed reduction. However, note that none of the aforementioned studies considered gender differences in driving patterns.

All these contributions show that PAYD policies represent a new approach to automobile insurance with potential advantages for customers, insurers and society as a whole.

3. Method

We are interested in explaining the distance travelled by drivers underwriting a PAYD policy until their first claim at fault as a function of their driving patterns. We use information collected by a GPS system installed in the insured vehicle. We use a Weibull regression model (Klein and Moeschberger, 2003). Although different methodological approaches have been used in the literature to capture influence of kilometers in the number of claims suffered by the insured (for example, Poisson regression models by Boucher et al., 2013), survival analysis can be more appropriate when we are interested in distance until the first accident with existence of censored observations. The Weibull regression allows us to consider the large variability in the distance actually driven before the first accident occurs. Previous use of this methodological approach can be found in Ayuso et al. (2014).

Let T_i be the accumulated number of kilometers until the first accident involving individual $i = 1,\ldots, n$, where n is the total number of individuals. A linear model can then be assumed for the logarithmic transformation of T_i, $Y_i = \ln T_i$, namely

$$Y_i = X_i'\beta + \sigma w_i$$

where β is a p-dimensional column vector of unknown regression parameters (usually including an intercept term), X_i is a p-dimensional column vector of explanatory covariates, σ is an unknown scale parameter, and w_i is an error term that is assumed to have an extreme value distribution and so it has a density function equal to $f(w) = \exp(w - e^w)$. The model can then be estimated by maximum likelihood, where the log-likelihood function is given by

$$L = \sum_{i\in\mathcal{U}} log\left(\frac{f(u_i)}{\sigma}\right) + \sum_{i\in\mathcal{R}} log\big(S(u_i)\big) + \sum_{i\in\mathcal{L}} log\big(F(u_i)\big) + \\ + \sum_{i\in\mathcal{J}} log\big(F(u_i) - F(v_i)\big) \tag{1}$$

where $F(\cdot)$ and $S(\cdot)$ are the cumulative distribution function and survival function of the error term w, respectively, $\mathcal{U}$ is the set of uncensored observations, $\mathcal{R}$ is the set of right-censored observations, $\mathcal{L}$ is the set of left-censored observations and $\mathcal{J}$ is the set of interval-censored observations. Additionally, $u_i = \frac{1}{\sigma}(y_i - x_i'\beta)$ and $v_i = \frac{1}{\sigma}(z_i - x_i'\beta)$, where z_i is the lower end of the censoring interval. In practice, the Weibull regression model can be easily estimated with SAS using the LIFEREG procedure (see SAS, 2014).

4. Empirical application

Here we analyze a sample of 8,198 drivers that underwrote a PAYD policy in 2009 with one of Spain's leading insurance companies. Their driving patterns were registered using a GPS, while the follow-up period was concluded on 31 December 2011. All drivers were under the age of thirty at the time of underwriting the policy.

The policyholders' driving patterns include the total number of kilometers travelled, the respective percentages of urban and nighttime driving and the percentage of kilometers travelled in excess of the speed limit.[2] This information is gathered for different time periods during each year, identified by the corresponding beginning/end date. In each time period, the number of kilometers travelled by each driver is recorded as well as the number of claims at fault. Thus, information regarding the number of kilometers to the first accident is interval-censored, given that for each driver at fault we only know the interval of the accumulated kilometers travelled in which the accident happened (time interval windows are, on average, equal to 151 days, while distance interval windows are, on average, equal to 4.6 thousand km). For some individuals, this information is right-censored. This occurs when the driver reached the end of the follow-up period (31 December 2011) or decided not to renew the PAYD policy prior to this date, without having being involved in an accident. Table 1 shows the variable descriptions.

Table 1. Variable descriptions.

Variable	Label
Sex	Binary variable (= 1 male, = 0 female)
Age	Age of the driver when their driving patterns began to be recorded (measured in years)
Age Vehicle	Age of the vehicle when the driving patterns began to be recorded (measured in years)
Experience	Driving experience, measured by the time elapsed since obtaining driving license until the moment when the driving patterns began to be recorded (measured in years)
Urban	% of urban driving (% of total kilometers travelled in urban areas)
Night	% of nighttime driving (% of total kilometers travelled at night – between midnight and 6 am)
Speed	% of the total kilometers travelled above the mandatory speed limits

[2] Note that some other driving patterns indicators as acceleration or heavy braking have not been included in this study due to lack of information in the database.

Note that in some cases we have considered "*Night > 3%*" representing a binary variable which is equal to 1 when nighttime driving is higher than 3% and 0 otherwise, similarly "*Speed > 7%*" is equal to 1 when the percentage of kilometers travelled above the speed limit is higher than 7% and 0 otherwise. As we explain in section 4.2 these threshold values have been chosen in order to produce significant associated parameters in the model.

4.1. Dataset

In our sample of drivers that underwrote a PAYD policy in 2009, 45.32% are women and 54.68% men. In Table 2 the means and standard deviations of *Age*, *Age Vehicle*, *Experience*, *Urban*, *Night* and *Speed* are presented for all drivers, as well as for men and women separately.

The mean age for all drivers is 23.67 years (standard deviation 3.06), while the mean ages by gender are almost identical. Recall, the product was offered to young drivers, which accounts for this low average age. The mean vehicle age is not as high for women as it is for men (5.66 and 6.55 years, respectively), while women have, on average, 3.35 years of driving experience while men have 3.82 years.

Women are found to do slightly less urban driving than men (27.11% vs. 27.81%), travel a lower percentage of kilometers above the speed limit (7.09% vs. 9.08%) and to do less nighttime driving than men (6.08% vs. 8.41%). We conducted a Kruskal-Wallis test to determine whether the above differences between women and men are statistically significant or not (note that the normality hypothesis for the variables in Table 2 is rejected when using the Kolmogorov-Smirnov test). The results of the test indicate that the differences between women and men are statistically significant for *Age vehicle*, *Experience*, *Urban*, *Night* and *Speed* (p-values < 0.01). In the case of *Age*, the differences are not statistically significant (p-value = 0.4724). Thus, we conclude that men in general present riskier driving patterns than women.

Table 2. Descriptive statistics for the age of the driver (*age*), the age of the vehicle (*age vehicle*), years of experience (*experience*), percentage of urban driving (*urban*), percentage of nighttime driving (*night*) and speed limit violations (*speed*).

	Men		Women		All	
	Mean	SD	Mean	SD	Mean	SD
Age	23.66	3.09	23.67	3.03	23.67	3.06
Age vehicle	6.55	4.48	5.66	4.37	6.15	4.45
Experience	3.82	2.99	3.35	2.78	3.61	2.91
Urban (%)	27.81	14.01	27.11	14.38	27.49	14.18
Night (%)	8.41	6.23	6.08	5.32	7.35	5.95
Speed(%)	9.08	8.14	7.09	7.06	8.18	7.73

SD is standard deviation. Variables *age*, *age vehicle* and *experience* are measured in years. *Urban* indicates percentage of total kilometers travelled in urban areas. *Night* indicates percentage of total kilometers travelled between midnight and 6am. *Speed* indicates percentage of total kilometers travelled above the mandatory speed limits.

4.2. Results

A Weibull model was estimated with interval-censored data for all drivers, as well as for men and women separately. We focus on the explanatory features so that we are interested in detecting what characteristics influence significantly the average distance driven before the first accident occurs. The results are shown in Table 3.

In the case of men (see Table 3), urban driving can be seen to have a significant effect and to reduce the distance travelled to the first accident. Additionally, speed limit violations have a significant effect, reducing the number of kilometers to the first accident (as well as having a quadratic effect). Nighttime driving has not been included in the final model as it has no significant effect in explaining the risk of accident among men. Driving experience and age of the vehicle both have a significant effect on the risk of accident. Thus, the more driving experience a man has, the longer the distance to the first accident; whereas the older the vehicle the policyholder drives, the shorter the distance to the first accident. Finally, the age of the driver is not included in the final model as it has no significant effect in explaining the risk of accident, in all likelihood because of its strong correlation with driving experience.

This model can now be used to estimate the survival curve and distance traveled to the first accident for different types of driver. Examples are provided in Figure 1. Thus, in Figure 1(a) the survival curves are presented for men according to their driving experience with the following driving pattern variables held equal: 30% urban driving, 15% excess speed and driving a vehicle that is 12 years old. It can be seen that the expected distance to the first accident is 50.4 thousand kilometers if the driver has only one year of experience, rising to 69.5 when the driver has eight years of experience when his PAYD policy is underwritten.

In Figure 1(b), the survival curves are presented for men according to the percentage of speed limit violations with the following driving pattern variables held equal: 30% urban driving, one year of experience and driving a vehicle that is 12 years old. It can be seen that the expected distance to the first accident is 54.9 thousand kilometers when the speed limits are exceeded 5% of the time, falling to 47.9 when speed limits are exceeded 20% of the time.

In the case of women (see Table 3), urban driving can be seen to have a significant effect in reducing the distance travelled to the first accident (as well as having a quadratic effect). Likewise, nighttime driving has a significant effect in reducing the distance to the first accident. Speed limit violations have not been included in the final model as they have no significant effect in explaining the risk of accident among women. Driving experience and age of the vehicle have the same impacts as those reported above for men: the more driving experience a woman has, the longer the distance to the first accident; whereas the older the vehicle the policyholder drives, the shorter the distance to the first accident. Here again the age of the driver has no significant effect in explaining the risk of accident, probably because of its high correlation with driving experience.

In Figure 1(c), the survival curves are presented for women according to the percentage of urban driving with the following driving pattern variables held equal: 6% nighttime driving, one year of experience and driving a vehicle that is 12 years old. It can be seen that the expected distance to the first accident is 55.8 thousand kilometers if the urban driving level is equal to 25%, falling to 39.1 if it rises to 40%.

Table 3. Weibull model estimations for men, women and all drivers.

	Men		Women		All	
	Parameter	p-value	Parameter	p-value	Parameter	p-value
Intercept	12.2226	<.0001	12.2429	<.0001	12.4345	<.0001
Urban	-0.0251	<.0001	-0.0497	<.0001	-0.0488	<.0001
$Urban^2$	-	-	0.0004	0.0060	0.0004	0.0003
Night	-	-	-0.0116	0.0332	-	-
Night > 3%	-	-	-	-	-0.1315	0.0321
Speed	-0.0291	0.0114	-	-	-	-
$Speed^2$	0.0008	0.0191	-	-	-	-
Speed > 7%	-	-	-	-	-0.0989	0.0436
Experience	0.0460	<.0001	0.0788	<.0001	0.0587	<.0001
Age vehicle	-0.0329	<.0001	-0.0209	0.0097	-0.0264	<.0001
Scale	0.7578	-	0.7681	-	0.7659	-
Shape	1.3196	-	1.3020	-	1.3057	-
-2logL	4847.038		4017.05		8863.154	
AIC[a]	4861.038		4031.05		8879.154	
BIC[b]	4905.433		4074.30		8934.807	

Note: Only variables with significant parameters are included in the three models. The likelihood ratio statistic has a p-value smaller than 0.001, so we reject the hypothesis that all parameters except the intercept are zero. [a]The AIC (Akaike Information Criterion) statistic and the [b]BIC (Bayesian Information Criterion) statistic are measures of the goodness of fit, and should be interpreted as "the smaller the better".

Finally, Figure 1(d) presents the survival curves for women according to the percentage of nighttime driving with the following driving pattern variables held equal: 25% urban driving, one year of driving experience and driving a vehicle that is 12 years old. It can be seen that the expected distance to the first accident is as low as 42.2 thousand kilometers with 30% of nighttime driving, rising to 53.2 if the driver has a 10% level of nighttime driving.

If the model is estimated for the whole sample (see Table 3), urban driving has a significant effect in reducing the distance traveled to the first accident (as well as having a quadratic effect). Nighttime driving and speed limit violations have a significant effect only when introduced in the model using binary variables, namely "Night > 3%" and "speed > 7%" (these threshold values have been chosen in order to produce significant associated parameters in the model).[3] Thus, nighttime driving above 3% is associated with a lower distance travelled to the first accident, while speed limit violations above 7% are also associated with higher risk of accident. Finally, driving experience and age of the vehicle have the same effect as those observed for men and women in the corresponding models.

[3] To select the threshold values we have previously made a descriptive analysis between the average distance travelled to the first accident at fault and values for percentages in each variable *Night* and *Speed*. We have observed significant differences between average distances to the first accident for percentages above and below these thresholds.

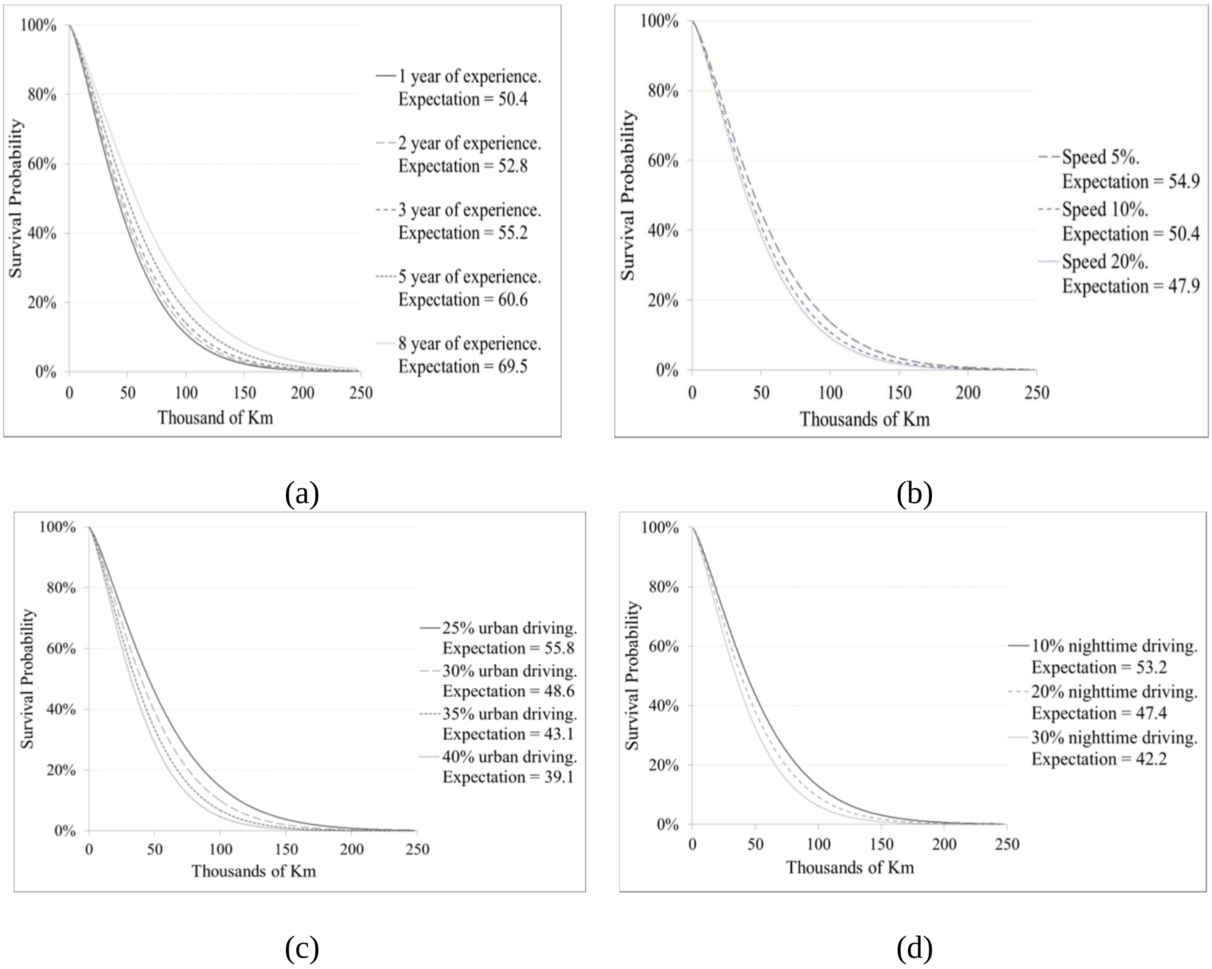

Figure 1. Survival curves for different types of drivers. (a) Men with 30% of urban driving, 15% of speed, depending on driving experience. (b) Men with 30% of urban driving, 1 year of experience, depending on speed. (c) Women with 6% of nighttime driving, 1 year of experience, depending on urban driving. (d) Women with 25% of urban driving, 1 year of experience, depending on nighttime driving. In all cases, vehicle age is assumed to be 12 years.

By way of example, we plot the survival curve for a policyholder with 35% urban driving, 5% speed limit violations, 10% nighttime driving, 1 year of driving experience and driving a car that is 12 years old (see Figure 2). The curves are obtained for men and women (using the corresponding Weibull model estimates) and without making a gender distinction (using the Weibull model estimated for the whole sample). It can be seen that the distance travelled to the first accident is 41.2 thousand kilometers for women, 48.4 for men and 46.5 if we do not make any gender distinction.

Finally, in Table 4 we compare the expected distances to the first accident for different driving patterns. The results are shown for men, women and all drivers separately. Driving experience and vehicle age are assumed to be equal to 1 and 12 years, respectively. It can be seen that the lowest estimation is 41.1 thousand kilometers, corresponding to women drivers with 35% urban driving and 10% nighttime driving. In contrast, the highest estimation is 62.2 thousand kilometers, corresponding to men with 25% urban driving and 5% in excess of the speed limit. It can also be seen that the expected distance to the first accident can differ substantially for men and women with the same driving patterns – see, for example, the case corresponding to 25% urban driving, 5% in excess of the speed limit and 10% nighttime driving. In this case, the difference in the expected distance travelled for men and women is almost 10 thousand kilometers (53.2 for women vs. 62.2 for men).

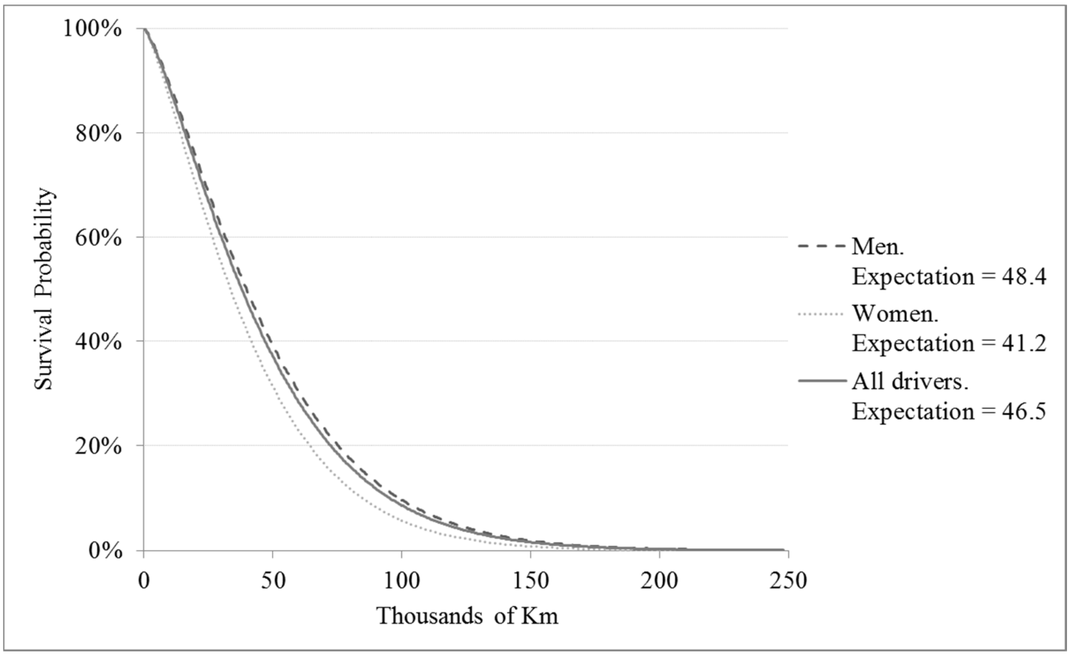

Figure 2. Survival curve for a policyholder with 35% urban driving, 5% speed limit violations, 10% nighttime driving, 1 year of experience and driving a car that is 12 years old. The curves are shown for men, women and without distinguishing between genders.

Table 4. Expected distance travelled to the first accident (in thousands of km) for different driving patterns. Driving experience and age of the vehicle are assumed to be equal to 1 and 12 years, respectively.

Driving pattern			Men	Women	All
Urban 25%	Speed 5%	Night 5%	62.2	56.4	59.5
		Night 10%	62.2	53.2	59.5
	Speed 10%	Night 5%	57.1	56.4	53.9
		Night 10%	57.1	53.2	53.9
Urban 35%	Speed 5%	Night 5%	48.4	43.6	46.5
		Night 10%	48.4	41.1	46.5
	Speed 10%	Night 5%	44.4	43.6	42.1
		Night 10%	44.4	41.1	42.1

5. Conclusions

Our findings allow us to draw a number of highly relevant conclusions. First and foremost, we have shown that men present riskier driving patterns than women, which accounts for the fact that they have traditionally had to pay more for automobile insurance than women. However, this distinction has been invalidated by the ECJ. Yet, if a PAYD pricing system is adopted, drivers of different gender can be charged different premiums if they present different driving patterns. Clearly, this distinction is not based on gender, but rather on driving behavior. This said, the conditions for obtaining a premium discount under a PAYD pricing system must treat both genders equally, even though we have shown here that the vehicle usage and driving patterns of men and women are not the same. For example, in our application we have seen that men can be expected to travel further than women before suffering the first accident, but owing to the EU's Gender Directive insurance companies are not permitted to establish a different threshold for the annual accumulation of kilometers by gender so that drivers can receive a PAYD premium discount. Additionally, we have shown that speed (number of kilometers above the limit) reduces the distance travelled to the first crash in the case of men, although not for women; and, that nighttime driving reduces the distance travelled to the first crash in the case of women, although not for men. Yet, here again, the premium discount for not speeding (or not driving during the night) cannot be different for men and women under a PAYD pricing system.

In conclusion, therefore, a PAYD system incorporates variables that go some way to compensating the effect of having to eliminate gender as a variable from the rating system. Thus, in the new context imposed by the Gender Directive, the concept of usage-based insurance may, in some cases, contribute to the maintenance of actuarial fairness. Insurance companies should consider different driving patters for pricing automobile insurance policies. Insurers must select the correct thresholds which allow applying discounts or surcharges that are actuarially fair.

Acknowledgements

This study was supported by ICREA Academia, the Spanish Ministry of Economy and Competitiveness and the ERDF under grants ECO2013-48326-C2-1-P and ECO2015-66314-R.

References

Aseervatham, V., Lex, Ch. and Spindler, M. (2016). How Do Unisex Rating Regulations Affect Gender Differences in Insurance Premiums? *The Geneva Papers on Risk and Insurance: Issues and Practice*, 41, 128–160.

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

Ayuso, M., Guillen, 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 and Prevention*, 73, 125-131.

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*, 43, 3, 1181-1186.

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.

Ellison, A.B., Bliemer, M.C.J. and Greaves, S.P. (2015). Evaluating changes in driver behaviour: a risk profiling approach, *Accident Analysis and Prevention*, 75, 298-309.

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.

Guillen, M. (2012). Sexless and beautiful data: from quantity to quality, *Annals of Actuarial Science*, 6, 2, 231-234.

Hultkrantz, L., Nilsson, J. E. and Arvidsson, S. (2012). Voluntary internalization of speeding externalities with vehicle insurance, *Transportation Research Part A*, 46, 926–937.

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*, 19, 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.

Khazzoom, J.D. (2000). Pay-at-the-Pump Auto Insurance: Review of criticisms and proposed modifications, *Journal of Insurance Regulation*, 18, 448-496.

Klein, J. P. and Moeschberger, M. L. 2003. Survival Analysis Techniques for Censored and Truncated Data, Springer-Verlag, New York.

Lahrmann, H., Agerholm, N., Tradisauskas, N., Berthelsen, K. K. and Harms, L. (2012). Pay as You Speed, ISA with incentives for not speeding: Results and interpretation of speed data, *Accident Analysis and Prevention*, 48, 17-28.

Langford, J., Koppel, S., McCarthy, D. and Srinivasan, S. (2008). In defence of the ‘low-mileage bias’, *Accident Analysis and Prevention*, 40, 6, 1996–1999.

Lee, J., Abdel-Aty, M. and Choi, K. (2014). Analysis of Residence Characteristics of At-Fault Drivers in Traffic Crashes, *Safety Science*, 68, 6–13.

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). Distance-based vehicle insurance feasibility, benefits and costs: Comprehensive Technical Report. Victoria Transport Policy Institute.

McCartt, A. T., Shabanova, V. I. and Leaf, W. A. (2003). Driving experience, crashes and traffic citations of teenage beginning drivers, *Accident Analysis and Prevention*, 35, 3, 311-320.

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

Quddus, M.A. (2015). Effects of geodemographic profiles of drivers on their injury severity from traffic crashes using multilevel mixed-effects ordered logit model, *Transportation Research Record: Journal of the Transportation Research Board*, 2514, 149-157.

SAS (2015). SAS/STAT 9.2 User’s Guide. The LIFEREG Procedure. Available at https://support.sas.com/documentation/cdl/en/statuglifereg/61799/PDF/default/statuglifereg.pdf.

Sass. J. and Seifried, F.T. (2014). Insurance markets and unisex tariffs: is the European Court of Justice improving or destroying welfare? *Scandinavian Actuarial Journal*, 3, 228-254.

Schmeiser, H., Störmer, T. and Wagner, J. (2014). Unisex insurance pricing: Consumers’ perception and market implications, *The Geneva Papers on Risk and Insurance: Issues and Practice*, 39(2), 322–350.

Sivak, M., Luoma, J., Flannagan, M.J., Bingham, C.R., Eby, D.W. and Shope, J.T. (2007). Traffic safety in the U.S.: Re-examining major opportunities, *Journal of Safety Research*, 38, 3, 337-355.

Toledo, T., Musicant, O. and Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers’ behavior, *Transportation Research Part C*, 16, 320–331.

Underwood, G. (2013). On-road behaviour of younger and older novices during the first six months of driving, *Accident Analysis and Prevention*, 58, 235– 243.

Vickrey, W. (1968). Automobile accidents, tort law, externalities, and insurance: an economist's critique, *Law and Contemporary Problems,* 33, 464-487.